МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Харківський національний автомобільно-дорожній університет

# ЗАГАЛЬНА ТЕОРІЯ СТАТИСТИКИ

*Підручник*

За науковою редакцією
кандидата економічних наук *А. В. Неправа*
доктора економічних наук *І. А. Дмитрієва*

Харків – 2022

**УДК 311.1**

## АВТОРСЬКИЙ КОЛЕКТИВ

Дмитрієв І. А. — гл. 9, Дмитрієва О. І. — гл. 10, Гіржева О. М. — гл. 11,
А. В. Непран — гл. 1, 2, 5, 8, 12, Бірченко Н. О. — гл. 4, 6,
Воронкова А. А. — гл. 3, 4, Чуйко Н. В. — гл. 7.

***Рецензенти:***

***Князєва О.А.***, д-р екон. наук, професор, професор кафедри економіки та цифрового бізнесу Державного університету інтелектуальних технологій і зв'язку;

***Прохорова В.В.***, д-р екон. наук, професор, завідувач кафедри економіки та організації діяльності суб'єктів господарювання Української інженерно-педагогічної академії;

***Третяк В.П.***, д-р екон. наук, доцент, завідувач кафедри управління та адміністрування Навчально-наукового інституту «Каразінська школа бізнесу» Харківського національного університету імені В. Н. Каразіна.

*Рекомендовано до видання рішенням Вченої Ради*
*Харківського національного автомобільно-дорожнього університету*
(*Дозвіл № 42/22/3.2 від 24 травня 2022 р.*)

**Загальна теорія статистики:** Підручник/ За ред. А. В. Непрана, І. А. Дмитрієва. Харків: ПП Іванченка, 2022. 720 с.

У підручнику викладені питання визначення загальних категорій, принципів та методів статистичної науки, організації та проведення статистичного спостереження, зведення первинного матеріалу та його подальшої обробки, правила складання статистичних таблиць та графіків, розрахунки відносних, середніх величин та показників варіації, обчислення та аналіз рядів динаміки, застосування вибіркового методу у статистичному спостереженні, індексного та кореляційного методів в аналізі матеріалів. Звернено увагу на методи аналізу точності та надійності статистичних показників. Поглиблено викладено використання математико-статистичних методів. Виклад ілюструється прикладами з економічної практики.

Підручник призначений для студентів економічних вузів та факультетів, а також для практичних працівників-економістів. Він може бути корисним для студентів інших спеціальностей, які використовують статистику у своїй роботі.

УДК 331.1

ISBN

# ЗМІСТ

## ПЕРЕДМОВА

Практична діяльність економіста будь-якої спеціальності неминуче пов'язана зі збиранням, обробкою та аналізом статистичних даних, вивченням закономірностей, зв'язків та залежностей між соціально-економічними явищами. Нерідко самому економісту доводиться проводити статистичне спостереження, застосовувати різні статистичні методи в аналізі матеріалів. При цьому має місце використання різних областей статистики у взаємодії. Тому вивчення науки статистики під час підготовки економістів має значення у системі вищої економічної освіти.

Підручник написаний переважно відповідно до програми курсу «Статистика» (для спеціальності 051 «Економіка», 071 «Облік та оподаткування», 076 «Підприємництво, торгівля та біржова діяльність»).

В курсі загальної теорії статистики вивчаються загальні категорії статистики як науки, принципи організації та проведення статистичного спостереження та зведення матеріалів, правил складання статистичних таблиць, послідовно розглядаються правила розрахунків відносних показників, середніх величин та показників варіації, дається узагальнення статистичного апарату вивчення взаємозв'язків у рядах динаміки, обґрунтовується застосування вибіркового методу у статистичному спостереженні, індексного та кореляційного методів в аналізі матеріалів. Поглиблено викладено використання математико-статистичних методів. Звернено увагу на методи аналізу точності та надійності статистичних показників.

Автори прагнули максимально використати все цінне, що створено в галузі української статистики і було вже опубліковано у відомих роботах. Даний підручник побудований таким чином, що теоретичний матеріал підкріплювався прикладами, що ілюструють методологію розрахунку показників, що розглядаються. Приклади побудовані на умовних даних. Науковий розгляд багатьох проблем теорії статистики потребує комплексного підходу. Ці обставини і обумовили характер викладу матеріалу в даному підручнику. Деякі питання автори постаралися викласти більш докладно, ніж це прийнято у навчальній літературі. Малюнки та схеми, що містяться в підручнику, будуть, на думку авторів, корисними для викладачів та студентів, можуть бути за допомогою сучасних аудіовізуальних засобів застосовані під час читання лекцій, проведення семінарських та практичних занять.

При написанні книги автори поставили перед собою завдання зробити книгу корисною не тільки для з'ясування логічного сенсу основних понять і формул загальної теорії статистики, але і для конкретного використання цих показників і формул при постановці практичних завдань та їх вирішенні.

Підручник призначений для студентів економічних спеціальностей вишів та практичних працівників-економістів. Він допоможе студентам економічних вузів, які вивчають курс загальної теорії статистики, набути необхідних навичок під час проведення математико-статистичних розрахунків. Книга може бути рекомендована і студентам інших спеціальностей, які у практичній роботі користуються математико-статистичними розрахунками. Наявність у підручнику довідкового матеріалу робить його корисним посібником для всіх, хто так чи інакше стикається зі статистикою.

Підручник з курсу «Загальна теорія статистики» для економічних вищих навчальних закладів та факультетів написаний колективом викладачів Харківського національного автомобільно-дорожнього університету та Державного біотехнологічного університету.

Окремі розділи роботи написані такими викладачами: Дмитрієв І. А. — гл. 9, Дмитрієва О. І. — гл. 10, Гіржева О. М. — гл. 11, Непран А. В. — гл. 1, 2, 5, 8, 12, Бірченко Н. О. — гл. 4, 6, Воронкова А. А. — гл. 3, Чуйко Н. В. — гл. 7.

Висловлюємо подяку д-ру екон. наук, проф. Князєвій О. А., д-ру екон. наук Прохоровій В.В., д-ку екон. наук, доценту Третяк В. П. за уважне прочитання рукопису та цінні зауваження, що сприяють його покращенню.

Автори будуть дуже вдячні своїм читачам за критичні зауваження до цієї книги, які просять надсилати за адресою: м. Харків, вул. Ярослава Мудрого, 25, Харківський національний автомобільно-дорожній університет, кафедра економіка та підприємництва.

## Г Л А В А 1

# ПРЕДМЕТ І МЕТОД СТАТИСТИЧНОЇ НАУКИ

## 1.1. Предмет статистики

Термін «статистика» має в теперішній час декілька значень:

— статистикою називають галузь практичної діяльності, що охоплює збирання, обробку, аналіз і публікацію масових даних про суспільні явища та процеси (статистичний облік);

— статистика — це і сукупність узагальнюючих, підсумкових показників, які якісно характеризують різні сторони суспільного життя — виробництво, розподіл та обмін, політику, культуру, охорону здоров'я, добробут населення (статистика шлюбів, дитячої смертності, чисельності населення в Україні);

— статистикою, нарешті, називають спеціальну наукову дисципліну.

Саме розгляду статистики як спеціальної суспільної науки присвячено даний посібник. Перейдемо тепер до з'ясування сутності статистики як особливої галузі знання, як науки. Вона досліджує кількісну сторону масових суспільних явищ і процесів. Це поєднує статистику з іншими науками, які досліджують суспільство, — історією, демографією, політичною економією, економічною географією, бухгалтерським обліком, галузевими економіками і т. п.

Кожна наука володіє рядом певних специфічних особливостей, що відрізняє її від інших наук. Головна особливість будь-якої науки полягає в специфічному предметі пізнання, що представляє собою сукупність явищ і процесів, або визначених їх характерних рис, суттєвих особливостей.

В чому ж полягає відмінність статистики від інших наук, що досліджують суспільні явища та процеси? Що вона досліджує, що є предметом її пізнання? Чи мають особливі ознаки певні явища суспільного життя, які досліджуються лише статистикою, а не іншими науками?

Об'єктом дослідження статистики виступає все суспільство. В цьому відношенні вона має багато спільного із суспільними науками. Проте вона має об'єкт дослідження, що відрізняє її від інших суспільних наук. В чому заключається специфіка предмета статистики?

*Предметом статистики виступає кількісна характеристика масових суспільних явищ та процесів, взята в нерозривному зв'язку з якісним їх змістом і відображена у відповідних статистичних показниках.* Інакше кажучи, предметом дослідження є групи якісно однорідних об'єктів. Вивчаючи сукупності якісно однорідних об'єктів, статистика встановлює ознаки даної сукупності, що відрізняють її від інших подібних сукупностей (наприклад, відмінність цієї групи робітників від іншої).

Суспільні явища різноманітні, складні та мінливі. Розглядаючи предмет статистики щодо суспільних явищ і процесів, слід передусім пам'ятати про наявність у них двох сторін: кількісної і якісної.

Характерною особливістю явищ суспільного життя є кількісна визначеність. Вона виражається в тому, що в кожний даний момент явищам суспільного життя присутні кількісні характеристики, які відображають сутність та їх розвиток. В процесі розвитку постійно змінюються кількісні характеристики явищ і процесів. Змінюються розміри явищ, змінюється їх кількісний склад, структура, якісна визначеність, співвідношення між окремими явищами, а також між ознаками одного явища. Так, у нашій країні змінюється чисельність населення, його віковий склад, структура, змінюється обсяг виробництва продукції, співвідношення між окремими видами; змінюється персонал підприємств, співвідношення між основними та допоміжними робітниками, між виробленою продукцією та основними засобами і т. п.

Разом про те слід зазначити, що з масовим явищем не можна розуміти будь-яку сукупність, що складається з безлічі одиниць. Наприклад, якщо взяти воду і розлити її в різні ємності, її склад буде одним і тим же. В цьому випадку не можна говорити про масове явище, хоча ця вода і налита в безліч судин.

Таким чином, уявлення про масове явище не повинно пов'язуватися лише з будь-якою сукупністю одиниць будь-яких об'єктів. Для утворення масового явища обов'язковою є варіація тієї ознаки, що досліджується. Варіація ознаки може змінюватись як від однієї одиниці будь-якого об'єкта до іншої, так і змінюватися в часі для одиничного об'єкта.

Вся ця кількісна сторона масових суспільних явищ є складає предмет пізнання статистичної науки. Кількісну сторону суспільних явищ та процесів статистика досліджує в конкретних просторових та часових межах — в різних районах, галузях, у визначені періоди часу. Так, в кожній окремій країні або на визначені дати різні чисельність населення, його віковий та національний склад, обсяги виробництва та споживання продуктів, розмір доходів і т. п. Всі ці та інші об'єкти та кількісні сторони масових суспільних явищ відображаються у статистичних показниках. Проте статистика досліджує не тільки кількісні характеристики явищ і процесів самі по собі, але й у нерозривному зв'язку із якісним змістом.

Якісна сторона явищ і процесів полягає у пізнанні їхньої сутності та закономірностей розвитку. Специфіку предмету статистики визначає те, що кількісна сторона масових суспільних явищ тісно пов'язана із якісною їх стороною. Рахівництво, кількісний облік в елементарному вигляді існують уже сотні років. Але кількісний аналіз передбачає не просто облік, а виявлення та встановлення тісноти зв'язку між явищами, виявлення закономірностей та тенденцій, яким ці явища підкоряються. Ця задача може бути вирішена лише при розкритті сутності явищ або процесів, їх внутрішніх та зовнішніх взаємозв'язків та характеристик.

Наукове пізнання та практична діяльність мають встановити, насамперед, якість або сутність суспільного явища, внутрішню визначеність явища, завдяки якій вони відрізняються один від одного (наприклад, відмінність населення за рівнем освіти). Якість — властивість, специфіка речі; позначає початкове і справжнє єдність системи різноманіття реальності, яке ще передбачає просторового і навіть уявного розчленування, здійснення його швидше наочно. Лише на основі якісного аналізу явищ та процесів можна проводити статистичний вимір закономірностей та взаємозв'язків між ними. Отже, під час проведення кількісного дослідження необхідно попередньо виявити реальну, конкретну природу і особливість розвитку досліджуваного явища або процесу. У цьому провідна роль належить економічної теорії (політичної економії), допоміжна роль — вимірювальним методам математичної статистики.

Для успішного проникнення в сутність явищ і процесів якісний аналіз досліджуваних процесів повинен бути доповнений кількісним виміром і аналізом, що дозволяє встановити механізм функціонування цих процесів. Результати виміру дозволяють більш глибше зрозуміти якісну сторону явищ і процесів, успішно керувати ними та прогнозувати їх розвиток на майбутнє. Важливо наголосити, що, не з'ясувавши якості предметів, неможливо встановити кількісні закономірності їх розвитку.

Слід пам'ятати, що неможливо усвідомити якісні особливості суспільних явищ без кількісного вираження їх взаємозв'язків і виміру закономірностей у суспільному розвиткові. «Кількість є така визначеність речі, завдяки якій (реально чи подумки) її можна поділити на однорідні частини та зібрати ці частини воєдино. Однорідність (подібність) частин чи предметів — відмітна ознака якості. Відмінності між предметами, не подібними один до одного, мають якісний, а відмінності між предметами подібними — кількісний характер»[1].

При вивченні масового явища можна виявити закономірність, яка ховається за варіацією індивідуальних значень. Наприклад, недостатньо взяти два або три підприємства для того, щоб скласти уявлення про стан цієї галузі, оскільки показники цих підприємств можуть бути значно вищими або нижчими, ніж у середньому по галузі. Щоб визначити результати роботи підприємств, потрібно обстежити набагато більше підприємств. Тільки в цьому випадку можна погасити вплив випадкових подій. Так само неможливо на прикладі кількох сімей визначити співвідношення між народженням хлопчиків та дівчаток. Лише масове спостереження, що охоплює безліч випадків народжень, свідчить, що хлопчиків зазвичай народжується більше, ніж дівчаток (106–107 на 100[2]).

Отже, статистика оперує масовими явищами з ознаками, що варіюють; це правильно, але не це найголовніше у визначенні статистики. Далі, статистика оперує сукупністю елементів, які належать до одного і того ж типу явищ і мають схожість між елементами по суттєвим для даного дослідження ознакам; це теж правильно, але цим визначається лише передумова, необхідна для застосування статистики. Більш істотною, найбільш характерною рисою статистики є те, що вона за допомогою специфічно властивих лише їй методів *виявляє за сукупністю великої кількості випадкових процесів внутрішньо необхідний зв'язок з її кількісною стороною*. Кількісна сторона масових

---

[1] Філософський словник. Київ, 1980, с. 150–151.

[2] За даними Державної служби статистики України, у 2020 р. на 100 дівчат народилося 107 хлопчиків).

явищ виявляється за допомогою специфічних прийомів та способів, які складають методологічну основу дослідження в статистиці.

Самі по собі статистичні методи та прийоми не розкривають якісної сторони процесу та його закономірності розвитку; проте коли це зроблено за допомогою якісного аналізу, статистичний метод дає можливість розкрити якісну його сторону, і тим самим конкретніше і точніше відобразити закономірність розвитку даних явищ та процесів.

Саме тому якісний аналіз явищ завжди має передувати кількісному, оскільки другий передбачає подальше відволікання від їх якісної сторони. Якщо якісний аналіз не буде проведено, то в подальшому при відволіканні від сутності та характеристики даних явищ можна допустити грубі помилки, а аналіз може бути безрезультатним і навіть заплутати питання. Так, наприклад, якщо віднести цифрові дані до різних типів явищ, то в результаті можуть бути отримані характеристики окремих груп, які відсутні в дійсності.

Ось чому необхідно, щоб застосуванню статистичних прийомів при дослідження соціально-економічних явищ завжди передував якісний аналіз. Тільки після цього, як встановлені сутність та основні групи явищ, можуть бути застосовані статистичні прийоми.

Проте важливе значення має й третій етап — відносна самостійність кількісного аналізу. В цьому полягає специфіка статистики соціально-економічних процесів на відміну від математичної статистики В математичній статистиці об’єктом виступає кількісна сторона абстрактних математичних сукупностей. Об’єктом соціально-економічної статистики є кількісна сторона суспільних явищ з конкретним якісним змістом. Однак у рамках якісно певної сукупності для статистичного аналізу значення набуває вже кількісна сторона як відносно самостійна.

Враховуючи вищевикладене, сформулюємо визначення статистики. *Статистика — наука, що досліджує кількісну сторону масових суспільних явищ в нерозривному зв’язку з їх якісною стороною, кількісний вираз закономірностей суспільного розвитку*. Статистика є такою наукою, яка специфічними для неї прийомами стосовно якісно-визначеного об’єкта дозволяє розкривати кількісну сторону явищ, а тим самим необхідність, приховану за поверхневою оболонкою випадкових явищ. Можна сказати, що за допомогою свої прийомів статистика допомагає виявляти необхідність, тобто закон досліджуваних явищ. Вона розкриває закономірності і взаємозв’язки, що виражають кількісні зміни масових явищ, але розкриває так, що вони дають можливість пізнати і їхню якісну визначеність, їх сутність. Відмінна риса статистики, що відрізняє її від інших суспільних наук, полягає в тому, що вона досліджує кількісний бік суспільних явищ, застосовуючи різні прийоми кількісного аналізу.

Існування окремої науки, що досліджує кількісну сторону масових суспільних явищ, викликана об'єктивною необхідністю. Як відомо, суспільні явища складні, постійно перебувають у розвитку. На відміну від відносно стійких явищ природи, суспільним явищам притаманні постійні зміни, обумовлені впливом багатьох факторів. Для встановлення, наприклад, вікового складу населення, необхідно обов'язково взяти все наявне на даній території населення, оскільки інакше не можна обчислити такий суттєвий показник, як вік окремих категорій громадян.

Між методами вимірювання природних та суспільних явищ багато схожого. Якщо обсяг і вагу вугілля можна встановити шляхом зважування, то вимірювання продукції галузі та інших узагальнюючих показників потребує вирішення багатьох методологічних та методичних питань. Внаслідок особливої складності та важливості кількісної сторони суспільних явищ та процесів, виникає необхідність у спеціальній науці, яка б розкривала взаємозв'язки та давала характеристику суспільним явищам та процесам.

Дані статистики про розміри і кількісні співвідношення масових суспільних явищ та процесів мають дуже велике значення. Вони широко використовуються для вирішення багатьох практичних питань, а також у наукових дослідженнях. Достовірні дані, отримані за допомогою статистичних досліджень, можуть бути використані для підтвердження або спростування певних теоретичних тверджень. З приводу одного із статистичних оглядів Англії К. Маркс писав: «Як би не виглядали сухо ці збудовані колонками в офіційно надрукованому документі цифри, вони насправді дають більше цінного матеріалу для історії загального розвитку, ніж томи, повні риторичної нісенітниці та політичної балаканини»[1].

Характеризуючи кількісну сторону масових суспільних явищ, статистика сприяє більш глибокому проникненню в їх сутність, дослідженню складу, структури, взаємозв'язків і закономірностей,

---

[1]

внутрішніх особливостей розвитку. Статистика створює фундамент для відкриття законів і закономірностей, служить засобом знаходження нових тенденцій і взаємозв'язків у розвитку суспільного життя. За допомогою статистики підтверджуються або спростовуються раніше відкриті закони. До тих пір, коли певні теоретичні судження не підтвердилися фактами, вони залишаються лише гіпотезами. Тому статистика, підтверджуючи висновки фактами, перетворює гіпотези у доведені теоретичні положення.

Статистика служить одним із інструментів, за допомогою якого використовуються закони ринкової економіки. Це пов'язано із встановленням конкретного стану економіки та певній території у визначений час. Відомо, що будь-який економічний закон у відповідності до своєї суті має певний механізм дії, який реалізується при певних умовах. Встановлюючи відповідний механізм використання законів, здійснюється свідоме їх використання.

В сучасних умовах статистика представляє собою складну та розгалужену систему наукових дисциплін. Основними розділами статистики як науки є:

1) загальна теорія статистики, в якій викладаються загальні принципи та методи дослідження статистики;

2) соціально-економічна статистика, яка розробляє методи кількісного аналізу соціально-економічних явищ та процесів;

3) галузева статистика, яка вивчає окремі галузі суспільних явищ або окремі галузі народного господарства: соціальна статистика, демографічна статистика (статистика населення), статистика промисловості, статистика сільського господарства, статистика будівництва, статистика транспорту, торгівлі, банківська статистика та ін.

Статистика як наука тісно пов'язані з іншими суспільно-економічними науками. Так як статистика досліджує кількісну сторону суспільних явищ та процесів, при вимірюванні величин та встановлення кількісних співвідношень вона виходить із визначених закономірностей та особливостей розвитку цих явищ. Якісний зміст (або якісну сторону) економічних явищ досліджує економічна теорія, соціального життя суспільства та соціальних процесів — соціологія, склад та рух населення та закономірності його розвитку — демографія тощо.

Теоретичною основою статистики є економічна теорія. Економічні поняття, категорії, положення, які нею встановлені, таким чином, є базою, фундаментом для дослідження. Економічна теорія досліджує та формулює фундаментальні закони та закономірності економічного життя розвитку. Спираючись на знання цих законів та закономірностей, статистика дає їм кількісну характеризує у часі і у просторі. Йдеться як про створення загальних теоретичних положень, а так і необхідність кількісного визначення процесів і явлений у процесі суспільного відтворення. З одного боку, багато теоретичних положень, які висуваються науковою думкою, повинні знайти підтвердження в кількісному вираженні явищ і процесів, що відбуваються в економічному та соціальному житті українського суспільства.

Проте й розвиватись суспільні науки по-справжньому можуть лише тоді, коли їх висновки щодо закономірностей та взаємозв'язків розвитку суспільних явищ ґрунтуються на аналізі масового статистичного матеріалу.

Здійснення кількісного аналізу для статистики має велике значення, що визначає її тісний зв'язок з математикою, особливо з математичної статистикою, що з масовими, кількісними відносинами, безпосередньо не пов'язана з якісними особливостями явищ. Особливо велике значення для статистики має використання спеціальних математичних дисциплін — *математичної статистики* та *теорії ймовірностей*. Перша розробляє математичні методи систематизації, обробки та використання статистичних даних для отримання науковий і практичних висновків, друга — досліджує випадкові величини, їх властивості та операції над ними. Застосування математики дозволяє підвищити якість аналізу причинно-наслідкових зв'язків, зробити його точнішим.

Суттєве значення має зв'язок статистики із технологічними дисциплінами. Зокрема, досліджуючи якість продукції, статистика спирається на досягнення інженерної науки. В той же час технологічні дисципліни дуже часто для висновків користуються статистичними методами збору та обробки даних.

## 1.2. Значення та основні завдання статистики

Статистика є основним методом вивчення масових суспільних явищ, основним джерелом інформації про стан та розвиток суспільних явищ та найважливішим методом їх вивчення.

Статистика допомагає виявити сутність закону, бо без неї цієї сутності і зовсім не було видно. Розкриваючи статистично виражену форму прояви закону, статистика *тим самим розкриває та її сутність* — щоправда, у співдружності з іншими науками.

У суспільстві статистика постає як один із засобів, за допомогою якого здійснюється регулювання економічного та соціального розвитку країни. За допомогою статистики збираються та аналізуються дані, необхідні для розробки прогнозів економічного та соціального розвитку країни.

О с н о в н а з а д а ч а статистичної науки полягає у пізнанні за допомогою кількісних характеристик причинно-наслідкового зв'язку між масовими суспільними явищами, в описі та вимірюванні взаємозв'язків та закономірностей суспільного розвитку. Статистика як наука має справу при встановленні причинно-наслідкових зв'язків із масовими явищами, що розрізняються між собою безліччю індивідуальних ознак та величиною останніх. Специфічною рисою статистичної науки є вимір кількісних відносин суспільних явищ у конкретно-історичних умовах, пов'язаних із певним часом та місцем.

Складність залежностей між окремими сторонами явищ, різноманіття взаємовідносин явищ із навколишнім світом роблять статистику необхідним знаряддям пізнання реальних зв'язків.

Специфічні особливості явищ та процесів суспільного життя, що використовуються для пізнання закономірностей та виявлення фактів, що визначають розвиток явищ, полягає у нерозривній єдності якісних та кількісних їх властивостей. Тому при дослідженні суспільних явищ та процесів статистика завжди ґрунтується на відповідній науковій дисципліні. Базою для теоретичного обгрунтування закономірностей розвитку явищ і процесів, які у житті суспільства, їх кількісному вираженні, є загальна теорія статистики.

В сучасних умовах зростання масштабів суспільного виробництва значення і роль статистики суттєво зростає порівняно із попередніми формаціями.

Велика роль статистики в регулюванні економічних та соціальних процесів в країні. Завдяки статистиці регулюючи державні органи можуть отримати всебічну інформацію про об'єкт, будь то народне господарство в цілому або окремі його галузі та підприємства. Статистика подає сигнали, що свідчать про порушення механізму розвитку явищ та процесів, наявність диспропорцій. Регулювати складними соціальними і економічними системами без статистики не можна.

Необхідно підкреслити роль статистики при плануванні та прогнозуванні соціально-економічного розвитку країни. Це проявляється уже на початковому етапі планування. Без статистичних даних неможливо здійснювати будь-яке планування, оскільки планування передбачає наявність інформації про розвиток об'єкта.

Оскільки в теперішній час планування здійснюється за декількома варіантами, то потреба в статистичних даних все більше і більше зростає. Крім того, державному регулюванні піддається все більше і більше економічних та соціальних процесів, що обумовлює необхідність до розширення статистичної інформації.

Значення статистики для планування визначається і тим, що статистика показує, як виконується план по окремим підприємствам, галузям та в цілому по народному господарстві. Зокрема, завдяки статистиці контролюється стан виконання державного та місцевих бюджетів країни. При цьому виявляються можливі диспропорції в народному господарстві, що служить сигналом для державних органів при розробці заходів щодо регулювання соціально-економічних процесів. За допомогою статистики чітко простежується розвиток країни. в процесі контролю за ходом виконання планів завдяки статистичним методам можна виявити резерви збільшення виробництва та підвищення його ефективності.

Статистичні дані необхідні для спостереження за розвитком економіки, впровадженням, у життя економічної політики держави, для аналізу використання трудових, матеріальних та фінансових ресурсів, виявлення невикористаних резервів зростання виробництва та продуктивності праці, для розробки теоретичної проблеми суспільства, для вдосконалення регулювання народним господарством. Усе це визначає виняткову важливість тісного зв'язку статистичної практики із національною економікою.

Українська статистики досягла суттєвих успіхів в розробці аналітичних прийомів, в особливості розробки системи показників, економічних класифікацій. Великі перспективи для аналізу макроекономічних процесів в цілому відкрили роботи щодо впровадження Системи національних рахунків.

Суттєво покращилася робота щодо механізації та автоматизації обробки статистичних даних. Застосування новітньої обчислювальної техніки і сучасних математичних методів дозволяє суттєво підвищити швидкість обробки величезного обсягу інформації, і тим самим забезпечити оперативність і можливість проведення глибокого економічного аналізу.

## 1.3. Стадії статистичного дослідження

Статистичне дослідження кількісної сторони суспільних явищ та процесів проходить в декілька етапів. На *першому етапі* відбувається збирання статистичних даних, під час якого формується первинний статистичний матеріал про соціально-економічні явища, що підлягають дослідженню. Планомірний, науково організований процес збирання даних щодо масових явищ та процесів, які відбуваються в економічній, соціальній та інших сферах життя називається статистичним спостереженням. Статистичне спостереження здійснюється за спеціальною програмою, розробленою на основі статистичної методології. Так, наприклад, при перепису населення реєструють заздалегідь обумовлені ознаки усіх жителів країни по ретельно розробленому плану (вік, стать, сімейний стан, національність, освіта і т. п.).

Статистичне спостереження дозволяє отримати висхідні дані, які характеризують все різноманіття прояву суспільних взаємозв'язків та закономірностей.

На *другому етапі* статистичного дослідження зібрані дані піддаються систематизації, класифікації та групуванню. Ця стадія називається зведенням статистичних даних. Вона включає групування даних статистичного спостереження та розроблення системи показників для характеристики виділених груп і підгруп.

Важливим методом, що застосовується в ній, є метод статистичного групування. Отримані результати спостережень містять якісно різні явища (наприклад, різні за видами економічної діяльності підприємства, різні види продукції). Щоб вивчити взаємозв'язки між двома чи більше ознаками, сукупність одиниць необхідно розчленувати на групи з двох чи більше ознак. Групування, таким чином, є розчленуванням сукупності факторів, що вивчаються, на окремі якісно однорідні види (наприклад, види підприємств відповідно до Господарського кодексу України). Різноманіття явищ та процесів потребують виявлення складу, структури, типів, внутрішній та зовнішніх зв'язків, факторів, що визначають їх розвиток. Так, наприклад, за даними перепису населення необхідно визначити групи населення за статтю, освітою, національністю, рівнем доходу і т. д.

При проведення статистичних досліджень метод групування має принципове значення тому, що він дозволяє виявити однорідні статистичні сукупності, виділити в них типи та групи за окремими ознаками, дати їм узагальнюючу характеристику. На цій стадії відбувається перехід від опису окремих індивідуальних одиниць статистичної сукупності до опису груп та об'єкта дослідження в цілому. Це здійснюється шляхом підрахунків узагальнюючих показників у вигляді середніх, відносних величин.

Третя стадія статистичного дослідження полягає в аналізі та узагальненні статистичних фактів та вияву закономірностей та взаємозв'язків масових соціально-економічних явищ та процесів. На даному етапі застосовується весь арсенал статистичних методів дослідження. Аналіз даних включає комплекс методів обробки багатовимірної системи даних спостережень та характеризується багатьма ознаками. На цій стадії отримують висновки про стан досліджуваного соціально-економічного явища, фактори, що визначають динаміку його розвитку. В ряді випадків аналіз дозволяє встановити такі зв'язки й ознаки, які раніше не розглядалися в процесі статистичного дослідження. Як правило, висновки і сам аналіз здійснюється текстом і супроводжується графічними і табличними ілюстраціями.

Основою аналізу даних статистичного спостереження є з'ясування соціально-економічної сутності досліджуваних явищ та процесів.

Форми і методи аналізу змінюються в залежності від характеру досліджуваних процесів. Так, дослідження щодо визначення резервів використання часу роботи обладнання цеху може бути проведено однією особою із використання окремих статистичних показників, що ґрунтуються на вибірковому методі. Проведення вибіркового дослідження доходів населення потребує ґрунтовної розробки плану, програми спостереження, підготовки великої кількості працівників, застосування великої кількості статистичних методів розрахунку узагальнюючих показників.

Науково-технічний прогрес має суттєве значення для реалізації планів соціально-економічного розвитку країни. Багатофакторність народногосподарських процесів, складність виробничих, міжгалузевих та територіальних зв'язків, посилення інтеграції України у світовий економічний простір при обмеженості статистичної інформації ускладнює вирішення багатьох соціальних та економічних задач. Все це ставить нові задачі перед українською статистикою. Для їх вирішення формуються нові галузі та підгалузі статистики. Зокрема, проходять процес становлення статистика науково-технічного процесу, статистика якості продукції, банківська статистика, екологічна статистика, патентна статистика, митна статистика, статистика науки та інновацій, статистика навколишнього середовища та ін.

Слід відрізняти статистику як суспільну науку від математичної статистики, яка досліджує методи систематизації, обробки та використання статистичних даних для отримання наукових та практичних висновків. Ці науки мають багато спільного. В суспільних науках передбачається дослідження складних об'єктів, які мають різносторонні внутрішні та зовнішні зв'язки. Ці об'єкти перебувають у постійному розвитку, змінюється їх склад, будова, внутрішні зв'язки. При чому зміни відбуваються досить швидко. Звідси витікає спільність прийомів обробки і оцінки даних (статистична перевірка гіпотез, дисперсійний, кореляційний, регресійний аналіз тощо). Різниця між ними полягає в тому, що математична статистика як розділ математики розглядає абстрактно кількісну сторону явищ, тоді як соціально-економічна статистика досліджує їх у нерозривному зв'язку з їх якісною стороною, в конкретних умовах часу та простору.

Таким чином, статистика перетворюється у потужний знаряддя пізнання. Вона допомагає розкрити та виявити соціально-економічні закономірності, оцінити причинно-наслідкові зв'язки явищ, встановити фактори, які визначають їх розвиток.

## 1.4. Особливості статистичної методології. Закон великих чисел та статистична закономірність

Методологічною основою соціально-економічної статистики є діалектичний матеріалізм і політична економія. Загальні принципи, положення наукового пізнання, які розроблені в матеріалістичній діалектиці, служать основою для розробки статистичної методології. Метою наукового дослідження є пред'явлення матеріальної дійсності в її деталях, у її детальному історичному розгортанні від простого до складного. Основними засадами діалектичного матеріалізму є всебічний зв'язок та постійний рух різних систем на основі внутрішніх механізмів руху та розвитку — постійного подолання неминучих протиріч. Особливо важливе значення для розкриття специфіки статистичної методології є трактовка таких філософських категорій, як причина і наслідок, випадковість, закон і закономірність, загальне і окреме, характер руху тощо.

Зупинимося лише на деяких особливостях статистичної методології. Ці особливості можна узагальнити в наступних трьох особливостях: 1) точне вимірювання масових суспільних явищ та процесів; 2) оцінка диференціація явищ; 3) застосування зведених (узагальнюючих) показників для характеристики явищ і закономірностей їх розвитку.

Перша риса, або особливість статистичної методології полягає в тому в статистиці для кількісного вивчення явищ і процесів використовують масове спостереження, яке представляє собою форму зв'язку між чисельністю досліджуваних явищ і ступенем, повнотою прояву загальної закономірності, властивої цим явищам. Але інші науки теж мають справу з масовими явищами. Але особливість статистичної методології полягає в тому, що для встановлення тенденцій і закономірностей розвитку суспільних явищ вона досліджує не окремі об'єкти, окремі одиниці сукупності, а вимірює загальні кількісні співвідношення. Наприклад, статистика досліджує рівень народжуваності не окремої жінки, а сукупності жінок в країні. Так здійснюється і дослідження чисельності населення, собівартість продукції, рух цін і т. п.

Велике значення для статистичної методології має закон великих чисел. *Законом великих чисел* називають загальний принцип, в силу якого кількісні закономірності, властиві масовим суспільним явищам, чітко проявляються лише в досить великій кількості спостережень. Закономірність масового явища може виступити тільки в досить великому числі випадків, коли сукупна дія безлічі випадкових причин призводить до результату, що вже майже не залежить від випадку. Але випадковості можуть взаємно врівноважити один одного тільки тоді, коли число їх *досить велике* для того, щоб могло виявитися дія *закону великих чисел*.

Математичне обґрунтування закону великих чисел, що виражає діалектичну зв'язок між випадковістю і необхідністю, дано теоретично ймовірностей.

Закон великих чисел — якщо його розуміти раціонально — вимагає того, щоб число індивідуальних подій було досить великим, бо тільки в такому випадку може з достатньою переконливістю виявитися об'єктивна закономірність, прихована за окремими випадковостями і пробивається крізь ці випадковості. Коротше кажучи, зміст закону великих чисел полягає в тому, щоб виявити дійсні тенденції та закономірності розвитку суспільних явищ, необхідно, щоб дослідженням було охоплено достатньо велика кількість елементів. І чим більшим буде маса об'єктів, тим із більшою достовірністю може бути зроблений висновок про характер розвиток, що лежить в основі даних випадкових подій.

Величину ознаки в сукупності за певних умов можна розглядати як випадкову величину, розподіл якої не тільки підпорядковується якійсь загальній закономірності розподілу, але і одночасно залежить від дії багатьох незалежних випадкових факторів. Ці випадкові чинники можуть в окремих випадках істотно вплинути на випадкову величину. Тому, щоб зробити певний висновок, потрібно вивчити велику кількість одиниць спостереження. Так, тривалість життя окремої людини визначається загальними умовами, але й багатьма індивідуальними особливостями організму тієї чи іншої людини. Неможливо зробити висновок про середню тривалість життя в країні на підставі обстеження невеликої групи людей, оскільки при невеликій кількості факторів виникає ризик впливу на випадкову величину безлічі факторів, що не залежать від цієї загальної закономірності. У цьому випадку кожна окремо взята величина буде випадковою: «…внутрішній закон, що прокладає собі дорогу через ці випадковості і регулює їх, стає видимим лише тоді, коли вони охоплюються у великих масах» (т. 25, ч. II, с. 396)

В результаті взаємопогашення дії випадкових факторів середні, обчислені для такої сукупності, виражають типові розміри ознак, що коливаються. Закономірності та тенденції мають силу лише у великих масах одиниць спостережень, але не як закони для індивідуального елемента сукупності. Так, наприклад, для суспільства властивий закон зростання продуктивності праці. Однак на окремому підприємстві за певних умов не виключено зниження продуктивності праці. Ця загальна закономірність зростання продуктивності праці виражається як середня закономірність.

Закон великих чисел грає велику роль не тільки як абстрактне положення науки, а й у практичної діяльності людей. Знаменитий швейцарський математик XVII століття Яків Бернуллі, який вперше вдягнув цей закон у математичну форму, писав, що спосіб збільшення спостережень з метою підвищення точності судження «не новий і не незвичайний» і що «навіть найбільш обмежена людина по якомусь природному інстинкту сам собою і без будь-якого попереднього навчання (що дуже дивно) знає, що більше прийнято до уваги таких спостережень, тим менша небезпека не досягти мети».

При збільшенні числа спостережень узагальнюючі показники стають більш точними, вони дозволяють виявити закономірності, властиві розглянутому явищу, які не були помітні при розгляді окремих випадків або при невеликій кількості спостережень. У цьому проявляється дія закону великих чисел. Наприклад, при виготовленні об'єктивів було зроблено 2–3 виміри фокусної відстані і отримана середня була близька до ДСТУ, але це ще не означає, що виробничий процес виготовлення об'єктів на верстаті налагоджений правильно. Але якщо середня з десятків або сотень вимірів мало відрізняється від необхідного розміру, то це вже є доказом правильного ходу виробничого процесу.

При малому числі спостережень досліджуване явище або процес може виявитися випадковим, що не підпорядковується видимим закономірностям. Однак якщо такі явища піддати масовому спостереженню, то в них часто можна виявити певну закономірність. У великій кількості спостережень взаємно погашаються (врівноважуються) випадкові фактори (коливання), що спостерігаються в індивідуальних явищ, в результаті якого чіткіше проявитися внутрішні необхідні зв'язки явищ. Взаємопогашення випадкових факторів і виявлення тенденцій та закономірностей відбувається в межах статистичних сукупностей. При цьому виявлення тенденцій і закономірностей відбувається тим точніше, чим ширше коло явищ охоплено сукупністю. У цьому й проявляється дія закону великих чисел — закону, що діє об'єктивно, і полягає у взаємопогашенні впливу випадкових, другорядних факторів, що не визначають сутність явищ.

Закон великих чисел відображає діалектичний зв'язок між необхідністю і випадковістю. Відображаючи діалектичний зв'язок між необхідністю і випадковістю, закон великих чисел діє у всіх масових явищах. За визначенням академіка А. Н. Колмогорова, закон великих чисел — «загальний принцип, в силу якого сукупна дія великої кількості випадкових факторів призводить, за деяких вельми загальних умов, до результату, що майже не залежить від випадку»[1].

Тому в основі статистичного дослідження завжди лежить велика кількість одиниць спостереження. Але закон великих чисел не є регулятором процесів, що вивчаються статистикою, не пояснює внутрішні закономірності розвитку, властиві тому чи іншому явищу. Він характеризує лише одну з форм прояву закономірностей у масових суспільних явищах та процесів.

У статистичній методології відіграє суттєву роль поняття статистичної закономірності. Статистична закономірність — в загальнонауковому, філософському значенні одна з форм прояву зв'язку між попереднім і наступним станом системи, що утворюється безліччю елементів і перебуває під впливом зовнішніх умов, що постійно змінюються. Під *статистичною закономірністю* розуміється повторюваність, послідовність та порядок у масових процесах чи явищах. Вона представляє собою особливу, статистичну форму вияву існуючих закономірностей. Остання виражає собою необхідний зв'язок не між якими двома явищами або двома сторонами явища, величинами, властивостями і т. д., а між величезним безліччю однотипних процесів або предметів, що знаходяться в особливій взаємодії між собою. Цим виявляється її зв'язок із законом великих чисел.

Статистична закономірність обумовлює малу ймовірність великих відхилень фактичних частот варіантів ознаки від теоретичних. Наприклад, на підприємствах запас матеріалів відповідає середньому використанню в нормальних умовах з резервним запасом, що забезпечує можливі відхилення від нормальних умов. Статистична закономірність з певною ймовірністю обумовлює стійкість середніх величин при збереженні впливу безлічі факторів, що залежать від цієї загальної закономірності.

Зі сказаного випливає, що термін «статистична закономірність» слід розуміти лише в тому сенсі, що він означає ту статистичну форму, в яку вбирається закономірність масового явища. Висловлюючись філософськи, статистична закономірність не особлива форма руху чи існування матерії, а лише зовнішній вираз тих відносин, які лежать у суті даного масового явища. Статистичні дані лише описують явище, сутність якого має проникнути відповідна наука. Статистична закономірність є своєрідним видом наукового узагальнення.

Звідси випливає, що статистичний вираз форми прояву закономірностей так тісно переплетено з сутністю закономірностей явища, що відокремити одну від іншої можна умовно, в абстракції.

Вивчення закономірностей конкретних явищ та процесів має велике теоретичне та практичне значення.

У працях класиків економічної теорії знайшло відображення використання дії закономірностей для вивчення багатьох економічних питань. Так, К. Маркс, розглядаючи складний економічний процес формування вартості товарів, на основі статистичної закономірності вивчав механізм відхилення цін виробництва від вартості, ринкових цін від цін виробництва.

Другою особливістю статистичної методології є розрахунок та оцінка диференціації явищ. Предметом статистичного дослідження є не просто масові суспільні явища. Специфіка методології полягає в тому, що ознаки масових явищ та процесів повинні мати суттєві відмінності між собою, тобто коливатися. Не має сенсу досліджувати явища, які не мають відмінностей між собою. Для цього не потрібно застосовувати весь арсенал статистичної методології. Об'єктом статистичного дослідження виступають лише ті одиниці, значення яких суттєво відрізняється при переході від однієї одиниці до іншої.

Третьою рисою статистичної методології є застосування зведених (узагальнюючих) показників. *Узагальнюючий показник* — показник, що виражає розміри або кількісні співвідношення ознак статистичної сукупності в цілому або її окремих частин (груп). Він характеризує одним числом найбільш типові, найбільш поширені сторони явищ, що вивчаються. Облік сукупності загалом та узагальнення даних масового спостереження дозволяють встановити та виміряти за допомогою узагальнюючих показників взаємозв'язку між явищами та процесами.

Прикладом узагальнюючих показників можуть бути широко застосовувані в статистиці середні величини: середньомісячна заробітна плата в галузях народного господарства; середньорічне число робітників і службовців, урожайність зернових культур, середня ціна реалізації і т. д. Так, закономірність розвитку економіки України за останні роки є зниження народжуваності та зростання смертності населення, а звідси і зменшення природного приросту населення. Використання узагальнюючих показників дає кількість оцінки даному явищу. І справді, якщо за даними ДСС України у 2010 р. кількість народжених на тисячу населення нашій країні становила 10,8, то у 2020 р. вона знизилася на 27,8 %, склавши 7,8. Водночас кількість померлих на тисячу населення за цей час перевищувала рівень народжуваності (наприклад, у 2010 р. вона становила 15,2, у 2020 р. — 15,9). Внаслідок значного перевищення кількості померлих над кількістю народжених природний приріст населення становить негативну величину, що значною мірою зумовило скорочення загальної чисельності населення країни. Цілком зрозуміло, що наведені узагальнюючі показники характеризують зміну кількості народжуваності та смертності всього населення країни в цілому, не є характерними регіонів.

Складність предмета дослідження визначає можливість розрахунку і використання для характеристики об'єкта дослідження різних видів узагальнюючих показників (абсолютних, середніх, різного роду відносних величин і т. п.), що відрізняються один від одного характером розв'язуваних завдань і методикою побудови.

## 1.5. Поняття та категорії статистичної науки

При дослідженні кількісної сторони суспільних процесів та явищ статистика використовує, крім розглянутих вище, низку особливих понять та категорій. До них відносяться насамперед: 1) ознака; 2) варіація; 3) статистична сукупність; 4) показник; 5) система показників.

*Ознакою* (в статистиці) прийнято називати властивість, характерну рису або іншу особливість, яка властива одиниці сукупності (як і одиниці спостереження), що досліджуються статистикою. Так, ознаками промислового підприємства можуть виступати: вид випущеної продукції, обсяг основних засобів, запаси матеріальних цінностей на складі, чисельність робітників і багато інших. Демографічними і соціальними ознаками людини служать стать, вік, освіта, національність, дохід і т. п.

В статистиці прийнято ділити ознаки на кількісні та якісні. Під *якісними* розуміють ознаки, які не мають числового вираження. До якісних ознак, що характеризують якість, властивість досліджуваного об'єкту, відносять стать, освіту, професію, вид економічної діяльності, вид валюти і т. п. Наприклад, професія людини може відрізнятися характером праці: слюсар, бухгалтер, учитель, інженер. Ознака: погодні умови; градації: сильна посуха, нормальна кількість опадів, кількість опадів суттєво більша за норму. В окремих випадках якісні показники можуть мати градації — «так», «ні», зазвичай кодуються як 1, 0 (прийнято рішення про зарахування об'єкта будівництва до першої категорії; постачальник відвантажує продукцію в строк), або більше градацій кодуються зазвичай послідовними цілими числами.

*Кількісними* називають ознаки, окремі значення яких вимірюються в деяких одиницях і задаються виміряними значеннями. Наприклад, вік працівника, розмір заробітної плати, вага деталі, температура тіла, відсоток виконання плану тощо). Кількісна ознака може бути дискретною та неперервною (інтервальною). *Дискретна ознака* — це ознака, що приймає тільки скінченну кількість значень, наприклад, тарифний розряд робітників. *Неперервні ознаки* приймають будь-які значення у визначених межах, виражаються цілими чи дробовими числами, реєструються з визначеним ступенем точності.

Ознаки можуть бути поділені на факторні та результативні. *Факторною* називають ознаку, що впливає на іншу, пов'язану з нею результативну ознаку (залежну ознаку), та зумовлює її зміну — варіацію. Наприклад, продуктивність праці виступає фактором зміни (зниження) собівартості одиниці продукції. Факторні ознаки визначають причини та умови зв'язку між соціально-економічними явищами та процесами. *Результативною* називають ознаку, на яку впливає інша, пов'язана з нею певною залежністю, факторна ознака. Вона характеризує наслідок зв'язку між окремими соціально-економічними явищами та процесами. Наприклад, кваліфікація робітника є фактором зростання продуктивності його праці. Чим вища кваліфікація, тим вища, при інших рівних умовах, продуктивність праці. Факторні та результативні ознаки не є постійними і можуть змінюватися.

Ознаки можуть бути поділені на основні (істотні) та другорядні (несуттєві). *Основні ознаки* визначають головний зміст процесів, явищ. *Другорядні ознаки* не пов'язані безпосередньо із внутрішнім змістом досліджуваних явищ. В статистиці головним чином ведеться спостереження та облік основних ознак, оскільки її головною задачею є визначення внутрішнього змісту явищ та процесів, дослідження взаємозв'язків між залежностей між ними. Другорядні ознаки доповнюють зміст досліджуваних процесів, дозволяють краще зрозуміти механізм функціонування та розвитку того чи іншого масового явища.

Ознаки можуть також бути поділені на *первинні*, які отримуються в результаті статистичного спостереження, і *вторинні* (*похідні*), які отримуються при обробці цих даних. Первинні показники визначаються шляхом зведення та групування даних і подаються у формі абсолютних величин. Похідні показники обчислюються на базі первинних або вторинних показників і мають форму середніх чи відносних величин.

Крім цього, ознаки ділять на варійовані та постійні. *Варійовані ознаки* приймають різні значення у окремих одиниць об'єкта спостереження. Наприклад, вік людини в певній країні коливається від 0 до 115 років, розмір заробітної плати працівників підприємства — від 6 до 25 тис. грн, вага деталі — від 86 до 89 г. *Постійні ознаки* мають незмінні значення ознак у всіх одиниць об'єкта. Наприклад, ціна проїзду в метро, розмір стипендії студентів тощо.

Статистичному дослідження піддаються лише варійовані ознаки. Якщо значення ознак у всіх одиниць сукупності є однаковим, то достатньо дослідити одну одиницю.

*Варіацією* називають зміну значень ознаки у статистичній сукупності, тобто набуття сукупністю або групами одиниць різних значень ознаки. Варіація альтеративної ознаки має у окремих одиниць лише одне із двох значень. Наприклад, стать людини може бути чоловіча або жіноча; верстат працює або не працює; людина здорова або хвора і т. п. Межі, в яких можливі відмінності значень величини кількісно варійованої ознаки у одиниць об'єкту, називають *межі варіації*. Нижня межа — це мінімальне значення ознаки, верхня межа — максимальне. Так, наприклад, якщо за нижню межу величини ваги деталі взяти 485 г, а за верхню — 492 г, то говорять, що вага варіює в межах від 485 до 492 г.

Окреме значення ознаки одиниці сукупності, відмінні від її значень інших одиниць, називається *варіантом* тієї ознаки. Варіанти можуть приймати будь-які значення в даних межах варіації. Деякі одиниці можуть мати однакове значення (той самий варіант) ознаки.

*Статистичною сукупністю* називають безліч одиниць (об'єктів), що володіють деякими загальними властивостями, істотними для їх характеристики і піддаються статистичного дослідження. Наприклад, сукупністю буде населення будь-якої країни, що складається з людей, які різняться за статтю, віком та багатьма іншими ознаками. В іншому випадку (залежно від мети дослідження) статистичною сукупністю будуть промислові підприємства, працівники промисловості або окремого підприємства, вироблена продукція, товарообіг магазину, чисельність безробітних, новонароджених, заробітна плата, шлюб продукції, валовий внутрішній продукт і т.п. д.

Окремі первинні неподільні елементи, або індивідуальні об'єкти, явища, що складають статистичну сукупність, називаються *одиницями сукупності*. Наприклад, статистичною сукупністю буде населення країни у визначений час, а одиницями її будуть окремі громадяни. Одиниці сукупності, володіючи деякими загальними ознаками, можуть різнитися між собою іншими ознаками. Таким чином, масові соціально-економічні явища завжди являють собою сукупності одиниць, які в певному відношенні якісно однорідні, але в інших відносинах відрізняються між собою. Так, із загальної чисельності населення країни можуть бути виділені сукупності громадян спочатку за однією ознакою (за статтю), потім за іншою (віком, освітою, національністю і т. п.).

Розрізняють сукупності однорідні та неоднорідні. *Однорідна сукупність* — статистична сукупність, в якій її складові елементи (одиниці) подібні між собою за істотними для даного дослідження ознаками і відносяться до одного і того ж типу явищ. Однорідна сукупність, будучи однорідною за одними ознаками, може бути різнорідною за іншими. *Неоднорідна сукупність* — статистична сукупність, в якій елементи (одиниці), її складові, відносяться до різних типів досліджуваного явища. Для таких груп обчислення середніх та відносних показників не має значення. За допомогою методу групувань і методу таксономії в неоднорідній сукупності можуть бути утворені однорідні групи.

Склад статистичної сукупності може змінюватися або залишатися незмінним на протязі певного проміжку часу. Якщо вона зберігає склад одиниць сукупності певний час, то така сукупність називається *стабільною*, якщо він змінюється, то така сукупність називається *динамічною*.

*Сукупність* називається *нормальною*, якщо розподіл чисельності її варіантів слідує нормальному закону розподілу.

*Показник* — важлива категорія статистичної науки. Це — *узагальнююча кількісно-якісна характеристика соціально-економічних явищ і процесів*. Якісна характеристика показника відображає сутність явища або процесу в умовах конкретного місця і часу, а кількісна — розмір, абсолютну або відносну величину його. Прикладами показників служать: чисельність населення, обсяг виробленої продукції, рівень продуктивності праці, величина собівартості продукції, площа посівів сільськогосподарських культур, число народжених дітей і т. п. В статистиці саме узагальнюючі статистичні показники виступають у якості основних категорій. Методика розрахунку статистичних показників, що відображають кількісну сторону суспільних явищ, що виступають головними категоріями статистики, складають основний зміст статистичної теорії.

Незважаючи на те, що статистичні показники характеризують кількісну сторону суспільних явищ та процесів, а в кінцевому рахунку кількість у зв’язку з його якісним змістом. Тому можна впевнено сказати, що показники виступають якісно-кількісними характеристиками масових явищ і процесів. У статистичних показниках кількісна і якісна сторони явищ знаходяться у тісному зв’язку. Наприклад, групуючи робітників, статистика виділяє якісно різні його контингенти — робітники, молодший обслуговуючий персонал, спеціалісти, інженерно-технічні робітники; співставлення темпів зростання обсягу національного виробництва приводить до характеристики якісних відмінностей між країнами; відмінності у окремих показниках випущеної продукції підприємства свідчить про якість технологічного процесу на підприємстві і т. п.

Кожен статистичний показник має якісне визначення, визначеність простору, визначеність часу і кількісне визначення. Число, в якому відсутнє хоча б одна із перерахованих ознак, не відноситься до статистичних показників. Наприклад, показник: чисельність населення (населення — якісна визначеність показника чисельності) м. Харкова (визначеність простору) на 1 січня 2022 р. (часова визначеність) — 1542,6 тис. осіб. (кількісна визначеність). Показник — одне із основних понять статистки. На відміну від ознак, що реєструються, показники завжди розраховуються певним чином.

Величина показника визначається в результаті його вимірювання за допомогою системи одиниць вимірювання (див. гл. 6) і відповідної методології. В статистиці часто показники поділяють на *об'ємні* та *якісні*. До перших відносять показники, які пов'язані із вимірюванням загальної величини одиниць статистичної сукупності, наприклад, чисельність населення на певній території або обсяг випущеної продукції і т. п.; до других — показники, які характеризують рівень розвитку явищ, наприклад, середня собівартість одиниці продукції, урожайність сільськогосподарської культури, середня продуктивність працівників цеху або заводу, середня тривалість життя в країні і т. п.

Різноманіття явищ і процесів суспільного життя, їх взаємозв'язків між собою, обумовлює і різноманіття статистичних показників. Показник може характеризувати окрему одиницю статистичної сукупності — окреме явище, об'єкт і т. п., групу одного і того ж виду і всю статистичну сукупність. Відповідно до ступеня охоплення показники поділяють на *індивідуальні*, *групові* та *зведені* (*загальними*). Важливою умовою правильного обчислення групових і загальних показників є якісний аналіз досліджуваних явищ і виділення якісно однорідних сукупностей.

Показники виражаються у формі абсолютних величин, відносних величини, середніх величині т. п. *Абсолютні показники* — кількісні показники, що отримують у результаті проведеного статистичного дослідження та які характеризують розміри (рівні, обсяги) суспільних явищ у певних умовах місця та часу. Абсолютні показники є основою, висхідною базою для розрахунку відносних та середніх показників і дають уявлення про розмір явищ та процесів. Ці показники вимірюються у певних одиницях: натуральних (кг, тонни, штуки, літри і т. п.), вартісних — гривни (євро,ієни тощо), трудових (люд.-год.). *Відносні показники* — показники, що виражають кількісні співвідношення між явищами або процесами суспільного життя, тобто узагальнені показники, що є результатом ділення однієї величини на іншу (виражаються у коефіцієнтах, відсотках, проміле тощо). *Середні показники* — узагальнююча міра варіюючої ознаки у статистичній сукупності. Вони характеризують типовий розмір варіюючої ознаки. Наприклад, середня заробітна плата робітників підприємства, середній вік життя, середній обсяг товарообороту, середня вага деталі і т. п.

Для дослідження сутності, взаємозв'язків та динаміки складних соціально-економічних явищ недостатньо користуватися одним статистичним показником. Необхідно використовувати систему статистичних показників. Сукупність статистичних показників, що всебічно відображають розвиток досліджуваних явищ та процесів, утворюють *систему показників*. Система статистичних показників повинна бути спрямованою на вирішення конкретної задачі. За допомогою системи показників мається можливість дослідити різні сторони явищ і процесів, виявити причини їх зміни, фактори, що впливають.

## 1.6. Сучасна організація статистики в Україні

В Україні державна статистична діяльність організована відповідно до державного устрою та адміністративно-територіальним поділом країни. Органами державної статистики є: центральний орган виконавчої влади, що реалізує державну політику у сфері статистики; функціональні органи державної статистики — підприємства, установи та організації, які знаходяться у сфері управління центрального органу виконавчої влади, що реалізує державну політику у сфері статистики. В Україні правові відносини у сфері державної статистики, права та функції органів державної статистики, організаційні засади провадження державної статистичної діяльності визначаються Конституцією України, Законом України «Про державну статистику», постановою Кабінету Міністрів України ««Положення про Державну службу статистики України», іншими законами та нормативно-правовими актами, а також міжнародними договорами України.

Центральним орган виконавчої влади в галузі статистики є Державна служба статистики України (далі — ДСС України), яка здійснює централізоване керівництво та забезпечує формування та реалізацію державної політики у сфері статистики. В кожній області маються головні статистичні управління, в районах та містах — районні та міські відділи статистики.

Державна служба статистики, головні управління статистики в областях, статистичні органи в містах та районах, наукові заклади, обчислювальні центри та інші організації, підприємства та установи органів державної статистки представляють собою єдину систему статистики в Україні.

ДСС України і органи державної статистики проводять свою роботу на загальних принципах, єдиній методології і організації загальнодержавної статистики виходячи із того, що централізована система статистики в Україні є одним із важливих важелів державного управління народним господарством.

Нинішня система державних органів представляє собою широко розвинену галузь державного управління. Вона має потужну технічну базу — ЕОМ, копіювальну та оргтехніку, науково-технічний персонал, навчальний заклад тощо. Без такого оснащення державна статистики не змогла б справитися з тими великими та складними завданнями, які стоять перед нею в період сучасного соціально-економічного розвитку.

Вкажемо на основні задачі органів Державної служби статистики України.

**ДСС України**. Державна служба статистики України здійснює централізоване керівництво справами обліку і статистики в країні. ДСС України несе відповідальність за організацію і удосконалення обліку і статистики, за відповідність інформації критеріям якості, за забезпечення достовірності наданих і опублікованих нею статистичних даних. Діяльність ДСС України регламентується «Положенням про Державну службу статистики України», затверджене постановою Кабінету Міністрів України від 23.09.2014 р. № 481.

Основними задачами Державної служби статистики України є: а) розробка, вдосконалення та впровадження науково обґрунтованої статистичної методології; б) затвердження звітно-статистичної документації статистичних спостережень; в) організація і проведення статистичних спостережень за соціально-економічними та демографічними явищами і процесами; г) збирання, перевірка, розробка та своєчасне передання статистичної інформації державним органам та органам місцевого самоврядування; д) проведення фундаментальних досліджень та прикладних розробок в галузі статистики.

В задачі ДСС України також входить розробка за участю державних органів та інших заінтересованих юридичних осіб проекти довгострокових програм розвитку державної статистики та щорічних планів державних статистичних спостережень, забезпечення використання в практиці органів державної статистики міжнародних статистичних стандартів та рекомендацій, створення та ведення Єдиного державного реєстру підприємств та організацій України та реєстрів респондентів статистичних спостережень.

Цілковито нова, більш складна задача встала перед ДСС України на сучасному етапі — створення і забезпечення ефективного функціонування автоматизованої системи державної статистики. Рішення цієї задачі передбачає широке впровадження електронно-обчислювальної техніки, новітніх засобів зв'язку, математичних методів.

До головних завдань ДСС України відноситься також розробка, вдосконалення та впровадження звітно-статистичної документації, а також типових форм первинної облікової документації, необхідної для проведення статистичних спостережень; здійснює міжнародні та міжрегіональні статистичні зіставлення; забезпечує підготовку, перепідготовку та підвищення кваліфікації виробників і користувачів статистичної інформації з питань статистики, обліку, звітності та використання сучасних інформаційних технологій.

У відповідності з цими задачами ДСС України розробляє (за участю міністерств і відомств) систему статистичних показників та єдину наукову методологію обліку і звітності в народному господарстві; організує і веде статистику за секторами та видами економічної діяльності, бюджетів сімей, використання природних ресурсів, охорони навколишнього середовища; проводить переписи населення, вибіркові та монографічні обстеження.

ДСС України здійснює загальне керівництво первинним обліком в народному господарстві, проводить перевірки організації і стану обліку та звітності; перевіряє достовірність звітних даних і у випадку виявлення помилок та інших навмисних викривлень звітних даних ставить питання про притягнення осіб, які винні у викривленнях, до відповідальності; про результати перевірок повідомляє відповідним міністерствам і відомствам.

ДСС України веде велику роботу по розвитку механізації та автоматизації обчислювальних робіт у своїй сфері та в народному господарстві в цілому; вона бере участь у розробці та створенні загальнодержавної автоматизованої системи збору та обробки первинної інформації для обліку, планування та регулювання народного господарства.

Очолює ДСС України голова. В ДСС України утворюється колегія у складі голови ДСС України (він же й голова колегії), його замісників, визначеної особи Міністерства економічного розвитку і торгівлі України, а також керівників самостійних структурних підрозділів ДСС України. Колегія на своїх засіданнях розглядає найважливіші питання організації, методології і механізація статистичних робіт; підсумки найважливіших виконаних статистичних і економічних робіт; заслуховує звіти керівних працівників ДСС України, керівників інших органів державної статистики; аналізує стан дотримання в органах державної статистики законодавства з питань державної служби, організаційно-кадрової роботи та виконавчої дисципліни; розробляє пропозиції щодо вдосконалення діяльності Держстату України, територіальних і функціональних органів державної статистики.

В ДСС України маються департаменти, самостійні відділи та деякі інші структурні підрозділи. Так, мається департамент статистики цін, в його складі декілька відділів, серед яких відділ статистики споживчих цін, відділ статистики цін виробників промислової продукції, відділ статистики цін виробників послуг, відділ статистики цін будівництва та житла, відділ міжнародних зіставлень. Серед основних структурних підрозділів в якості прикладу наведемо департаменти: статистики національних рахунків; соціальної статистики; статистики цін; статистики населення; статистики сільського господарства та навколишнього середовища; статистики зовнішньоекономічної діяльності та енергетики; статистичної методології тощо.

**Територіальні (місцеві) органи державної статистики**. Державна служба статистики України здійснює свої повноваження безпосередньо та через утворені в установленому порядку територіальні органи. До територіальних органів відносяться управління статистики в областях, містах та районах. Головні управління статистики в областях є територіальним органом Державної служби статистики України і в межах наданих повноважень здійснює реалізацію державної політики у сфері статистики з метою одержання достовірної інформації про соціально-економічне становище регіону. Статистичні управління статистики в областях, містах і районах, що називають територіальними (місцевими) органами державної статистики, ведуть статистики на належних територіях, збирають, оброблюють та аналізують статистичні дані про розвиток усіх сфер народного господарства,забезпечують місцеві органи необхідними статистичними матеріалами. Вони виконують роботи загальнодержавного плану статистичних робіт, проводять статистичні дослідження за завданнями уряду України, ДСС України, міністерств та відомств; перевіряють стан обліку та звітності, достовірність статистичних даних на підприємствах, організаціях та установах. Вони проводять переписи, одночасні обліки на своїх територіях та інші статистичні роботи, які покладаються на них вищими організаціями.

Територіальні органи статистики складаються з відділів та управлінь, основними серед яких є: відділ планування, координації, моніторингу статистичної діяльності, управління збирання та підготовки даних статистичних спостережень, управління збирання та підготовки даних вибіркових обстежень населення, управління обробки даних статистичних спостережень та ін. Обласні управління статистики, що підпорядковуються ДСС України, є юридичними особами публічного права і складовою частиною єдиної системи органів. Діяльність обласних управлінь статистики здійснюється на основі Положення, затверджених ДСС України.

**Планування статистичних робіт, що виконуються органами державної статистики**. Органи державної статистики проводять свою роботу за планами статистичних робіт, які заздалегідь розробляються. ДСС України щорічно складає план статистичних робіт. Хоча він і називається «План діяльності державної служби статистики України» за відповідний період, він є загальнодержавним планом статистичних робіт органів державної статистики. План діяльності розробляється відповідно до Закону України «Про державну статистику України», Програми розвитку статистики на відповідний період, Плану державних статистичних спостережень на відповідний період. Він містить найменування та коротку характеристику роботи, її періодичність, спосіб представлення, головного виконавця. Крім того, ДСС України розробляються спеціальні плани проведення окремих великих робіт.

План доводиться до статистичних управлінь областей, які на його основі складають свої плани, беручи із нього те, що покладається на них, і добавляють роботи, які вони намічають провести для задоволення місцевих потреб в статистичних даних. Обласні управління статистики складають плани статистичних робіт для районних (міських) органів державної статистики.

**Відомча статистика**. Крім органів державної статистики, статистичну роботу проводять міністерства та відомства. В їх складі маються відповідні структурні підрозділи. Вони називаються по-різному — департаменти, відділи, управління (наприклад, відділ розвитку авіаційної галузі, департамент з питань промисловості та ін.). Вони збирають у межах своєї компетенції і оброблюють статистичні дані, які необхідні для оперативного регулювання галузей та секторів народного господарства і які не збираються органами ДСС України.

Відомча статистика займає важливе місце в статистичній роботі в країні і в багатьох випадках служить джерелом статистичної інформації для аналізу, планування та регулювання соціально-економічних процесів та явищ в країні.

## ЛІТЕРАТУРА


1. *Горкавий В. К.* Статистика: Підручник. Третє вид., переробл. і доповн. Київ: Алерта, 2019. С. 11–36.
2. Статистика: підручник / *С. І. Пирожков, В. В. Рязанцева, Р. М. Моторин* та ін.; Київ. нац. торг.-екон. ун-т. Київ : КНТЕУ, 2020. С. 8–23.
3. *Ткач Є. І., Сторожук В. П.* Загальна теорія статистики: підручник. 3-тє вид. Київ: Центр учбової літератури, 2009. С. 5–16.


## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники, словники*


4. Про державну статистику: Закон України від 17.09.1992 р. № 2614-XII. URL: https://zakon.rada.gov.ua/laws/show/2614-12#Text.
5. *Карпенко Л. М.* Статистика: навч. посіб. Одеса: ОРІДУ НАДУ, 2019.
6. *Козирєва О. В.*, *Федорова В. О.* Статистика: навчальний посібник. Харків: Видавництво Іванченка І.С., 2021.
7. Статистичний словник / [*О. Г. Осауленко*, *О. О. Васєчко*, *М. В. Пугачова* та ін.]; за ред. д-ра держ. упр., проф., член-кор. НАН України О. Г. Осауленка; НТК статистичних досліджень. Київ: ДП «Інформ.-аналіт. агентство», 2012. 498 с.


*Монографії та статті*


8. Нариси з історії статистики України. – 2-ге вид., випр. та доп. Київ: Держкомстат України; НДІ статистики Держкомстату України, 1999. 187 с.

# Г Л А В А 2

# СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ

## 2.1. Поняття про статистичне спостереження. Основні вимоги до статистичного спостереження

Для статистичного дослідження будь-якого масового явища, як про це вказувалося в гл. 1 необхідно зібрати відомості про його стан на певний момент часу (на певну дату) або про результати його розвитку за певний період часу. Це найважливіший специфічний прийом, у якого формується те, що називається статистичним матеріалом чи первинної статистичної інформацією. Наприклад, щодо економіки торгівлі певної території з кожному торговому підприємству збирають дані, що характеризують розвиток торгівлі. У зв'язку з цим вивчення масового явища починається зі статистичної реєстрації фактів як сприйняття дійсності.

Статистичне спостереження не можна ототожнювати з власне спостереженням, тобто безпосереднім сприйняттям явищ з допомогою почуттів. Вище ми підкреслили ту особливість статистичного спостереження, що воно проводиться з метою проведення статистичного дослідження, отримання узагальнюючих характеристик про масові суспільні явища та процеси. Якщо, наприклад, торговий представник вивчив ціни різних виробників вироби, це ще не можна назвати статистичним спостереженням. Поряд з цим фахівці територіальних органів державної статистики для проведення спостереження за цінами відбирають та реєструють ціни на продукти (конкретні товари) з урахуванням певних критеріїв. Ці відомості передають відповідним статистичним органам. Останні на основі цих даних визначають рівень цін у просторі (територіальні відмінності у цінах) і за часом (зміна цін в один період години порівняно з іншим). Так визначаються споживчі ціни, ціни виробників промислової продукції, ціни на будівництво та ін.

*Статистичне спостереження* — планомірний, науково організований збір даних про явища та процеси суспільного життя шляхом реєстрації за заздалегідь розробленою програмою спостереження суттєвих ознак. Статистичне спостереження як перший етап статистичного спостереження спрямоване на формування первинного статистичного матеріалу — даних (висхідної первинної інформації) про спостережувані об’єкти. Від дотримання принципів наукової організації статистичного спостереження значно залежить успіх всієї роботи. Обробка та аналіз зібраних у процесі спостереження відомостей повинні дати статистичну характеристику явища, що вивчається, процесу, обґрунтовані відповіді на питання, поставлені перед дослідником.

*Метою статистичного спостереження* є реєстрація елементів, що становлять досліджуване масове явище, за заздалегідь встановленими найбільш істотними ознаками.

По суті, статистичне спостереження збігається переважно з першим (емпіричним) ступенем пізнання суспільних явищ і процесів. Будь-яке дослідження, зокрема і статистичне, починається зі збору фактичного матеріалу, спостереження; висновків, узагальнення як у науці, і у практиці цінні лише тоді, що вони виведені з урахуванням фактів. Тільки накопичення суттєвих відомостей про досліджуване явище і їх повна достовірність забезпечують цінність зведених матеріалів і аналітичних даних.

Таким чином, статистичне спостереження, як правило, представляє собою початок всього процесу наукового пізнання масового явища і як одна з найважливіших стадій статистичного дослідження.

Отримана в результаті статистичного спостереження інформація має велике наукове та практичне значення. Дані статистичного спостереження використовуються для розрахунку узагальнюючих статистичних показників, необхідних управлінським органам для оперативного керівництва виробництвом та управлінням, для його планування та прогнозування. Та чи інша форма розподілу статистичних даних може бути йому джерелом існування внутрішніх причинних відносин, вказати, у напрямі треба вести пошуки сил, дію яких отримало у цій формі розподілу своє зовнішнє вираження. Без первинного статистичного матеріалу, без фактичних даних не можна отримати обґрунтованих рішень, що враховують особливості місця, часу, впливу різних факторів на результат виробничої діяльності.

Від якості статистичного спостереження великою мірою залежить успіх всього спостереження в цілому. Організація статистичного спостереження має бути побудована таким чином, щоб інформація, отримана в результаті проведення статистичного спостереження, була об'єктивною та відображала дійсний стан цього об'єкта досліджень. Залучення апарату математичної статистики для обробки неповних, неточних або неперевірених первинних статистичних даних призводить до помилкових висновків, порожніх витрат часу, а іноді слугує для отримання бажаних висновком шляхом підгонів.

Маючи на увазі цю обставину, видатний математик і статистики А. А. Чупров писав: «… вправною обробкою одного й того ж матеріалу можна вижати з нього… прямо протилежні висновки».

Неповні, неточні дані недостатньо добре характеризують процес, а тим більше спотворюють його і можуть призвести до помилкових виводів.

Не всякий отриманий масив даних, зібраних у процесі статистичного спостереження, може бути придатним для теоретичного аналізу, узагальнення та обґрунтування висновків. До статистичних даних, зібраних у процесі проведення статистичного спостереження, пред'являються низка вимог. Ці вимоги одночасно пред'являються і до статистичного спостереження.

Перша, основна вимога, що пред'являється до даних, зібраних в результаті статистичного спостереження, полягає в тому, щоб вони були *повними*, а не уривчастими, випадково вихопленими. Такий первинний матеріал виходить в результаті повного статистичного спостереження.

По-перше, повноту статистичного матеріалу слід розуміти щодо повного охоплення одиниць досліджуваної сукупності (повнота просторового охоплення). Наприклад, щодо розвитку сільського господарства необхідно зібрати дані про діяльність всіх без винятку господарюючих суб'єктів, які займаються виробництвом сільськогосподарської продукції, або про групу типових підприємств, що входять у цю галузь.

У більшості випадків повні дані про те чи інше явище суспільного життя є багаточисленними. Однак у деяких випадках можуть охоплювати кілька одиниць. Наприклад, два метрополітени (у м. Києві та м. Харкові) повністю вичерпують усі метрополітени України; одне підприємство (ВАТ «Турбоатом») здійснює виробництво турбін для атомних та теплоелектростанцій; сім нафтопереробних заводів повністю покривають виробництво нафтопродуктів у країні. Дані про основні показники цих підприємств забезпечують повноту охоплення досліджуваного процесу, але за кількістю одиниць сукупності вони нечисленні.

Повнота даних забезпечує отримання вичерпних первинних статистичних даних, всебічне вивчення явища, відтворення процесу загалом. Отримана на основі повних даних первинна інформація є основою встановлення міцного фундаменту фактів, на основі яких можуть бути зроблені правильні висновки та узагальнення. Вони виступають своєрідною гарантією від спотворень, тому що не дозволяють випадковим подіям істотно змінити точність висновків та узагальнень.

Однак у практичній діяльності з ряду явищ і процесів досягти повноти охоплення досліджуваних одиниць сукупності у часі та в просторі дуже складно, а часто і немає можливості. Наприклад, повністю досліджувати доходи та витрати сімей країни, якість масової продукції, структуру використання робочого часу працівниками підприємства тощо. У таких випадках доводиться вдаватися до обстеження частини статистичної сукупності (наприклад, вибіркового обстеження продукції), і на підставі цього обстеження робити висновки про рівень якості усієї випущеної продукції. До дослідження частини статистичної сукупності можуть вдаватися й у деяких інших випадках. Умовою правильності та обґрунтованості висновків та узагальнень, отриманих на підставі обстеження частини сукупності, є правильний відбір цієї частини. Правильність відбору забезпечується тим, що відібрані одиниці сукупності повинні відображати основні зміни у розвитку всієї досліджуваної сукупності. Іншими словами, частина елементів, відібраних для вивчення, повинна містити типові властивості і специфічні особливості досліджуваного явища. Типові дані виступають формою повних даних, необхідні висновків і узагальнень.

По-друге, повноту статистичних відомостей слід розуміти щодо охоплення всіх визначальних ознак масових суспільних явищ, що вивчаються (повнота охоплення сторін явища). Як відомо, будь-яке масове явище, будь-який об’єкт складається з безлічі одиниць, які є носіями нескінченно великої кількості ознак, властивостей, зв’язків, відносин як у собі, так і з усім навколишнім світом. Причому одні ознаки, одні взаємозв'язки є істотними, внутрішньо необхідні, які визнають якість даного об'єкта дослідження, інші — несуттєвими, зовнішніми. Природно, що при науковому підході до статистичного дослідження інтерес становлять виняткова суттєві властивості чи ознаки. Тому при проведенні аналізу доводиться обмежуватися вивченням лише найістотніших із них, що відображають основні, причинно-наслідкові зв'язки та взаємозалежності.

По-третє, при аналізі відомостей за певні періоди часу повнота статистичних відомостей передбачає наявність даних за максимально тривалий безперервний термін (повнота охоплення в часі). Необхідність повноти охоплення явищ, що вивчаються в часі, випливає з того, що явища і процеси постійно змінюються в часі. На них впливають як постійні, так і випадкові фактори. Для усунення чи ослаблення впливу останніх факторів та виявлення закономірностей потрібно вивчити статистичні дані протягом тривалих періодів часу. Наприклад, при вивченні валового збору сільськогосподарських культур для отримання точних висновків потрібно використовувати дані за ряд років, тому що в рослинництві на результати діяльності впливають погодні умови.

Другою важливою вимогою, що пред'являються до даних, зібраних в процесі статистичного спостереження, є достовірність, точність інформації. *Достовірність інформації* — ступінь адекватності відображення інформацією явищ, подій або процесів, що нею описуються. Інформація вважається достовірною, якщо вона повно і правильно відображає події або процеси, що описуються нею. Причому під достовірністю тут розуміється не лише їхня арифметична, вимірювальна точність до гривень, копійок, кілограмів тощо, але й відповідність даних об'єктивної дійсності. Наявність помилок в інформації, а також неповнота відомостей про описане явище, подію або процес знижують достовірність інформації. В українському законодавстві встановлено, що керівники підприємств та організацій відповідають за факти приписок та інших спотворень державної звітності. Винні у цьому особи притягуються до дисциплінарної, матеріальної чи кримінальної відповідальності.

Третьою, особливою вимогою, що висувається до статистичних даних, зібраним у процесі спостереження, виступає вимога їхньої *одноманітності*, *сумісності*. Порівнянність відомостей, збираних у процесі статистичного спостереження, необхідне подальшої обробки, узагальнення і порівняння в просторі і часу.

Такі три основні вимоги до статистичних даних, які піддаються обробці, аналізу та узагальненню. У багатьох випадках статистичні дані використовуються державними органами для прийняття управлінських рішень на всіх рівнях, до них пред'являється ще одна, четверта вимога — *своєчасність*. Дійсно, первинний статистичний матеріал, якщо він цілком достовірний, але надходить невчасно, не може бути дійсно використаний для прийняття оперативних управлінських рішень на всіх рівнях державного управління.

Перед статистикою стоять складні питання, що стосуються організації статистичного спостереження суспільних явищ, а саме: як слід здійснити спостереження і в який час і в які терміни потрібно його проводити, які ознаки необхідно реєструвати в первинній звітності і як контролювати достовірність отриманої інформації і багато які інші питання? Всі ці питання знаходять своє рішення теорії статистичного спостереження.

## 2.2. Програмно-методологічні питання статистичного спостереження

**Програмно-методичні питання спостереження**. Проведення статистичного спостереження починається із складання плану, що поділяється на програмно-методологічну та організаційні частини.

*Програмно-методологічні питання статистичного спостереження* — перелік пунктів плану статистичного спостереження, які відповідають на питання: для чого проводиться статистичне спостереження (мета статистичного спостереження), що спостерігається (об'єкт статистичного спостереження), що є складовими об'єкта статистичного спостереження (одиниці статистичного спостереження), що є джерелом інформації статистичного спостереження (звітна (облікова) одиниця), на які запитання слід отримати відповіді у ході державного спостереження (програма статистичного спостереження), чим характеризується статистичне спостереження (вид, час, період, організаційна форма, спосіб статистичного спостереження й опис системи статистичних показників). Ця частина плану статистичного спостереження містить також перелік методологічної та звітно-статистичної документації, національних та локальних статистичних класифікацій та ін.

Статистичне спостереження проводиться з певною, наперед встановленою метою, організується за планом, у якому передбачається вирішення всіх питань, пов'язаних з організацією спостереження, його проведенням, збиранням та обробкою даних.

Перш ніж розпочати проведення статистичного спостереження, чітко формулюють його ціль, оскільки від цього залежить обсяг роботи, методи та терміни виконання, вартість роботи. *Мета статистичного спостереження* — отримання статистичної інформації (даних), яка є підставою для узагальнення характеристики стану та розвитку явища або процесу з визначенням відповідних закономірностей, взаємозв'язків та тенденцій. Дослідник, який встановлює план збирання статистичних даних, а також програму подальшої їх обробки, повинен чітко визначити те пізнавальне завдання, яке передбачається вирішити за допомогою статистики. Цілі спостереження, як правило, визначаються потребами практичної та наукової діяльності та характером досліджуваного об'єкта статистичного спостереження. Вони визначаються виходячи із завдань статистичного дослідження і в загальному вигляді встановлюються в документі, на підставі якого організується спостереження.

**Об'єкт та одиниця статистичного спостереження**. При проведенні статистичного спостереження необхідно визначити, що підлягає спостереженню, інакше кажучи, необхідно встановити об'єкт спостереження. Безліч явищ і предметів (звана статистикою статистичної сукупністю), про які мають бути зібрані необхідні відомості, утворюють *об'єкт статистичного спостереження*. У практиці статистичних досліджень може бути сукупність підприємств, машин, устаткування, шкіл, жителів країни тощо. Об'єкт статистичного спостереження встановлюється шляхом чіткого визначення поняття те, що, де і коли має вивчатися. На основі аналізу досліджуваного явища потрібно виділити і вказати ознаки та риси, що відрізняють його від інших, подібних до нього об'єктів, визначити межі переходу від одного явища до іншого.

Наприклад, під час перепису встановленого обладнання дається чітке визначення, яке обладнання вважати встановленим (тобто що вивчати), станом на яку конкретну дату проводити спостереження (тобто коли вивчати).

Для успішного проведення статистичного спостереження об'єкт має бути точно визначено. Для цього потрібно вказати ознаки, риси, що відрізняють його від інших, подібних до нього об'єктів. Кожен об'єкт складається, зазвичай, з багатьох елементів, чи одиниць, його складових. Якщо об'єкт визначено неточно, то процесі спостереження деякі частини може бути втрачені, недообліковані, чи, навпаки, в обстеження можуть потрапити частини інших об'єктів, які не належать даної сукупності. В результаті отримані дані не можуть відобразити дійсного положення.

Кожен об'єкт статистичного спостереження складається з окремих елементів — одиниць статистичного спостереження[1]. *Одиницею спостереження* називають первинний елемент об'єкта статистичного спостереження (сукупності), який є носієм ознак, що підлягають реєстрації під час статистичного спостереження. Іншими словами, одиницями спостереження виступають ті первинні ланки, у тому числі складається об'єкт спостереження. Наприклад, під час перепису устаткування одиницею спостереження буде окремий верстат, окремий агрегат, щодо використання робочого дня — витрати праці на виконання певних операцій. Вибір об'єкта спостереження залежить від мети та конкретних умов проведення спостереження. Визначаючи одиницю конкретного статистичного спостереження, потрібно точніше охарактеризувати її, вказавши специфічні риси, які б відрізнити її від близьких до неї за видом одиниць інших об'єктів. Одиницею спостереження, зазвичай, виступає одиниця сукупності. Чітке визначення одиниці спостереження дає змогу правильно визначити об'єкт спостереження.

При організації статистичного спостереження важливо чітко визначити одиницю виміру, оскільки однією з вимог проведення статистичного спостереження є сумісність явищ. Чітке визначення одиниці спостереження дозволяє правильно виділити об'єкт статистичного дослідження.

Залежно від складності об'єкта статистичного спостереження, а також від конкретних завдань дослідження у одному і тому ж самому спостереженні може бути не одна, а кілька одиниць спостережень. Так, якщо збираються відомості про стаж, вік, освіту працівників підприємства, то, природно, одиницею спостереження є людина, бо реєструються ознаки, властиві людині. Але якщо реєструються ознаки, як склад членів сім'ї, середній рівень доходу, отриманого однієї члена сім'ї, особи у сім'ї певного віку і т. п., то цьому випадку одиницею

---

[1] Іноді одиницю спостереження називають одиницею рахунку.

спостереження є сім’я.

Залежно від предмета дослідження одиницями спостереження можуть виступати реально відокремлені фізичні одиниці (населення країни, працівники підприємства, запаси на складах готової продукції, види обладнання, об’єкти незавершеного будівництва тощо), організаційні одиниці (підприємства, організації та установи), окремі події або стану процесу (факти народження або смерті, простою верстата тощо).

Іноді недостатньо загального короткого визначення одиниці спостереження. Зустрічаються такі об’єкти статистичного спостереження, порівнянні елементи яких суттєво відрізняються своїми найменуваннями та деякими іншими ознаками.

Після того, як сформульована мета дослідження, встановлені одиниці статистичного спостереження, що підлягають включенню, об’єкт слід відмежувати у просторі, тобто територіально. Припустимо, що вивчається використання робочого дня працівників підприємства. Виникає питання: взяти всіх основних робітників чи основних робітників якогось одного цеху чи бригади.

Після визначення мети спостереження, об’єкта та одиниці спостереження розробляється програма спостереження.

**Програма статистичного спостереження**. Статистичне спостереження проводиться суворо за визначеним планом, що включає як програмні, і організаційні питання. У плані статистичного спостереження одне з основних питань — програма спостереження. Перелік ознак одиниці спостереження, що підлягає реєстрації у процесі проведення статистичного спостереження, називається *програмою статистичного спостереження*. Зміст програми та кількість поміщених у ній питань залежить від специфіки об’єкта дослідження та від мети статистичного спостереження. Широка програма, що містить значну кількість питань, дозволяє отримати більш різнобічні відомості про об’єкт спостереження і тому глибше досліджувати сутність та особливості взаємозв’язків одиниць сукупності. Якщо здійснюється перепис населення, то про кожну людину необхідні відомості про вік, сімейний стан, національність, освіту і т. п. Вузька програма передбачає збір відомостей з обмеженого кола питань, що часто мають лише практичне значення. Наприклад, при зборі відомостей про виробничий стаж робочих програма спостереження включає лише одне питання. Складаються також спеціалізовані програми, що враховують специфічні особливості об’єкта спостереження, зокрема програми для обліку робочого часу роботи обладнання. З метою вивчення явища за певний період або на певну дату проводиться спеціально організовані спостереження за спеціально розробленою програмою: перепису, одноразові обліки, спеціальні обстеження.

Обсяг показників, які у програмі, значною мірою залежить також від коштів, які мають державні статистичні органи, які проводять спостереження, від терміновості, з якою потрібно отримати статистичні дані, та багатьох інших чинників.

До програми статистичного спостереження *пред'являється ряд вимог*. Розглянемо найважливіші їх.

1. Програма статистичного спостереження повинна містити суттєві ознаки, що характеризують наскільки можна найважливіші властивості досліджуваного явища, його тип, основні риси, специфічні особливості, прямі та непрямі ознаки та інші якісні характеристики сутності об'єкта дослідження. Це дуже важлива наукова вимога до розробки програми статистичного спостереження, яка має вирішальне значення для всього статистичного дослідження. До програми повинні включатися і такі показники, які дозволили б розкрити невикористані резерви виробництва, виявити недоліки та прорахунки в організації та управлінні підприємством, галуззю та народним господарством в цілому. Тобто вона має бути достатньо повною для вирішення поставлених завдань.

2. До програми не слід включати питання, які не можуть бути отримані об'єктивними, достовірними відповідями, зокрема через відсутність у первинному обліку необхідних даних, через незацікавленість обстежуваних у наданні достовірних даних. Слід уникати «зайвих» питань, які не належать до мети спостереження. Крім того, мілкі, незначні властивості об'єкта спостереження лише відволікають дослідника, внаслідок чого поза увагою може виявитися головна ознака чи закономірність у досліджуваному процесі. Розробляючи програму статистичного спостереження, звичайно, необхідно прагнути до повноти збору статистичної інформації, яка отримується в результаті проведення статистичного спостереження.

3. Питання програми мають бути сформульовані чітко, ясно і при цьому не виникала можливість їхнього різного тлумачення. Чим конкретніше питання, тим більше гарантії правильного його розуміння та отримання правильної відповіді. Якщо ймовірна можливість різного тлумачення чи його нерозуміння, необхідні вказівки і пояснення питання. Наприклад, якщо у звітності підприємства ставиться питання

«обсяг продукції», необхідно вказати вид продукції та одиниці її вимірювання з метою порівняння продукції за підприємствами. Детальна розшифровка питань, а також порядок заповнення бланка звітності, переписного листа, анкети наведено в інструкції.

4. Питання програми повинні записуватися в певній послідовності і, по можливості, включати питання контрольного характеру, які служать цілям перевірки та уточнення відомостей, що збираються. Так, у переписному листі перепису населення України 2001 р. питання «Ваша дата народження» кореспондує питанню «Ваш сімейний стан». Два питання ставилися щодо перевірки віку, а саме: про дату народження і кількість років, що виповнилися. Питання у формах статистичної звітності, а також зміст самих форм звітності, що характеризують виробничо-фінансову діяльність підприємств (обсяг виробленої продукції, витрати на виробництво, фонд заробітної плати, собівартість продукції та ін.) взаємопов'язані між собою.

5. Доцільно питання програм формулювати так, щоб отримані відповіді можна було порівняти з даними статистичних спостережень попередніх періодів.

У результаті обробки первинного статистичного матеріалу, зібраного у процесі спостереження, має бути отримана система узагальнюючих статистичних показників (абсолютні, відносні, середні величини), що характеризує як сукупність в цілому, і окремі її групи. Тому попередньо становлять макети статистичних таблиць. Нижче наводимо макет статистичної таблиці, призначеної для характеристики демографічного навантаження на населення України.

Таблиця 2.1

**Демографічне навантаження на населення у віці 16–59 років та 15–64 роки**

<table>
<tr><th rowspan="3">Рік</th><th colspan="3">На 1000 осіб постійного населення у віці 16–59 років</th><th colspan="3">На 1000 осіб постійного населення у віці 15–64 роки</th></tr>
<tr><th rowspan="2">загальне навантаження</th><th colspan="2">у тому числі особами у віці</th><th rowspan="2">загальне навантаження 0–15 років</th><th colspan="2">у тому числі особами у віці</th></tr>
<tr><th>0–15 років</th><th>60 років і старше</th><th>0–15 років</th><th>60 років і старше</th></tr>
<tr><td>2019</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2020</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2021</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2022</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Таким чином, правильно складена програма — це найважливіша умова успішного проведення статистичного спостереження.

Розробка програми — одна з найважливіших теоретичних та практичних проблем статистичного спостереження. Від того, наскільки добре буде розроблена програма статистичного спостереження, значною мірою залежить якість зібраного матеріалу, успіх всього дослідження. Програма статистичного спостереження може бути побудована лише на основі глибокого знання досліджуваного предмета. Правильно розроблена програма спостереження — найважливіша умова успішності будь-якого статистичного дослідження.

Розробка програми спостереження — складний та відповідальний етап статистичної роботи. Тому для успішної розробки програми статистичного спостереження необхідно мати досить повне уявлення про об'єкт дослідження, а також його складових елементів.

Зміст програми багато чому визначається сутністю об'єкта спостереження. Чим складніший об'єкт спостереження, тим глибше і детально має бути розроблена програма спостереження. Тому для успішної розробки програми необхідно мати досить точне уявлення про об'єкт спостереження та його одиниці. Зміст програми статистичного спостереження залежить багато в чому від мети спостереження, потреби різних органів державного управління у певних статистичних даних, необхідних для прийняття важливих державних рішень у сфері економічного, соціального та правового життя суспільства. Так, мету Всеукраїнського перепису населення було задекларовано у Законі України «Про Всеукраїнський перепис населення» від 19 жовтня 2000 р. № 2058-ІІІ, а саме: «Метою проведення Перепису населення є отримання достовірних, об'єктивних та цілісних даних щодо різноманітних характеристик населення країни в цілому та по кожній адміністративно-територіальній одиниці для інформаційного забезпечення управління та прогнозування соціально-економічного розвитку, а також розроблення та реалізації виваженої державної політики з питань народонаселення».

Розробку програми державного спостереження економічних явищ та процесів здійснюють відповідні державні органи (Державна служба статистики України, Кабінет Міністрів України, Міністерство фінансів України та ін.).

Програма статистичного спостереження включає: постановку мети та завдань спостереження; визначення об'єкта спостереження та одиниць сукупності; встановлення переліку ознак, що характеризують одиницю спостереження або, інакше, переліку питань, з яких збираються відомості; формулювання питань програми та підказ ряду можливих відповідей; послідовність постановки питань у програмі.

Одночасно із програмою спостереження складається і програма розробки матеріалів спостереження, яка конкретизує завдання статистичного спостереження, вона ясніше показує, які саме дані потрібно зібрати під час проведення спостереження. Іншими словами, програма розробки дозволяє уточнити програму статистичного спостереження. Якщо ж програма розробки буде складена після збору матеріалів статистичних даних, то може виявитися, що потрібних даних у програмі немає, а багато питань, що включені до програми спостереження, є зайвими.

Програма спостереження отримує своє конкретне втілення у переліку питань, відповіді на які мають бути отримані під час спостереження. Зміст програми визначається метою статистичного дослідження, конкретними завданнями того, що необхідно встановити, виявити. Правильна, науково обґрунтована програма спостереження має вирішальне значення всього статистичного дослідження.

На методологію статистичного спостереження, зокрема, розробку програми спостереження, впливає механізація та автоматизація рахункових робіт та розвиток інформаційних технологій. У пам'ять ЕОМ вносяться та зберігаються статистичні дані. За допомогою певних інформаційних технологій ЕОМ можуть за командою знайти та видати користувачеві дані, здійснювати операції з їхньої обробки. Це дозволяє суттєво підвищити оперативність даних, їхню точність при обробці.

Проведення спостереження передбачає наявність певного інструментарію статистичного спостереження. Під *інструментарієм статистичного спостереження* розуміється перелік бланків та документів, що відносяться до статистичного спостереження. Головні з них — формуляр спостереження та інструкція до нього.

**Формуляр статистичного спостереження**. Збір необхідних даних фіксується у так званому формулярі спостереження або статистичному бланку (бланком зазвичай називається незаповнений формуляр). *Формуляр* (*бланк*) є особливим чином розграфлений аркуш (чи аркуші) паперу, у якому містяться питання програми спостереження і місце відповідей них, і навіть для запису шифрів (кодів) відповіді питання. Обов'язковим елементом будь-якого статистичного формуляра є титульна та адресна частини. У *титульній частині* зазвичай міститься найменування статистичного спостереження, дата, найменування органу, який проводить спостереження, ким і коли затверджено цей формуляр, іноді і номер, присвоєний йому у загальній системі формулярів статистичних спостережень, здійснених цим органом. В *адресній частині* передбачається запис точної адреси одиниці або сукупності одиниць спостереження, здійснених даним органом. У багатьох випадках у формулярах статистичного спостереження, крім того, вказується в які терміни і куди (у які адреси) повинні надсилатися заповнені формуляри, а також передбачаються підписи осіб, відповідальних за правильність утримання в них відомостей.

Прикладом статистичного формуляра є будь-який бланк статистичної звітності, спостережний лист фотографії робочого дня, переписні листи, особиста картка (з обліку кадрів), листок про простій, анкети, опитувальні листи та ін. Використання того чи іншого бланку визначається специфічною особливістю спостереження, що організується.

Дуже важливо, щоб питання були сформульовані ясно і якомога швидше.

Важливою вимогою, що пред’являється до статистичного формуляру (бланку) є коротка, чітка, виразна і вичерпна характеристика питання. Питання бланка нічого не винні допускати різного змісту та різних форм відповідей. Питання має вимагати однієї відповіді, підходити до нього. Якщо кількість можливих звітів невелика, рекомендується підказувати (повністю чи частково) відповіді у самому тексті питання. У бланку може бути передбачена альтернативна форма відповідей: «так», «ні».

У питаннях бланка спостереження слід уникати таких слів, як «інші», «решта», «і т. п.», тому що лічильники мають тенденцію відносити до названих пунктів усі неясні та невизначені випадки.

Сукупність питань у бланку мають забезпечити взаємоперевірку.

У статистичній практиці використовують два основних види бланків – картковий та списковий. *Картковий* називають бланк, що містить дані лише про одну одиницю спостереження, а *списковий* — для запису відомості про кілька одиниць спостереження. У картці загальна кількість карток дорівнює кількості одиниць досліджуваної статистичної сукупності. У списковому формулярі (бланку-списку) кожній одиниці спостереження відводиться горизонтальна або вертикальна смуга, в якій спочатку записується її найменування (наприклад, прізвище, ім’я, по батькові, якщо йдеться про людей), а потім уже відповіді на питання, що містяться в формулярі. Потрібно пам’ятати, що застосування облікової форми бланка можливе лише при відносно невеликій програмі і лише при експедиційному способі спостереження та самореєстрації. Приклад карткового формуляру — звіти підприємств, спискового — переписні листи перепису населення 2001 р.

Названі формуляри мають свої переваги та недоліки. У карткових формулярах витрачається багато часу заповнення титульної і адресної боку, нераціонально використовується папір. Облікові формуляри практично не застосовні за широкої програми.

Програма статистичного спостереження знаходить свій конкретний вираз у переліку питань формуляра спостереження. Скажімо, при організації спостереження робітників заводу поставлено завдання отримати дані про використання робочого часу робітників. Для цього необхідно зафіксувати час роботи робітників. Щоб це зробити, потрібно у формулярі поставити відповідне питання. Але як його сформулювати?

Формулювання питань бланка спостереження має винятково велике значення в організації та успішному проведенні статистичного спостереження. Якість дослідження багато в чому залежить від правильного розуміння усіма спостерігачами (реєстраторами, хронометражистами) питань програми. Питання мають бути сформульовані наскільки можна коротко, ясно і виразно, щоб їх розуміння не викликало труднощів. Так, у переписі населення України 2001 р. питання про вік сформульовано так: Ваша дата народження (число, місяць, рік, скільки виповнилося років). У той же час, якщо питання про вік сформулювати так: «Вік», то неясно, мається на увазі кількість років, що виповнилися, повна дата народження або рік народження.

Формулювання питань бланка спостереження — справа, як правило, дуже складна. Нерідко у статистиці при організації та проведенні спостережень доводиться витратити багато часу, перебираючи багато варіантів формулювань, доки не знайдеться та, яка виявиться найточнішою і буде внесена до формуляру питань. При недостатньо строгому формулюванні питання виникають неясності, що знижує точність отриманих відповідей.

Заключною частиною роботи зі складання програми спостереження є визначення порядку розташування питань статистичного формуляра та розробка його форми.

Якими б ясними були питання формуляра статистичного спостереження, до них зазвичай дається інструкція. *Інструкцією* називають документ, який роз'яснює питання програми статистичного спостереження, порядок заповнення статистичного формуляра та частково організаційні питання його. В інструкції точно вказуються сенс того чи іншого питання, пояснюється порядок проведення спостереження, містяться відповіді на питання про цілі та завдання спостереження, об'єкт та одиниця спостереження, його тривалість тощо.

В інструкції містяться відповіді на питання про завдання спостереження, об'єкт та одиницю спостереження, його тривалість, порядок перевірки даних, основні терміни і т. д. Але головним змістом інструкції є роз'яснення питань формуляра; часом у ній наводяться можливі варіанти відповіді питання. Наприклад, порядок заповнення форми № 2-ОЗ ІНВ «Звіт про наявність та рух необоротних активів, амортизацію та капітальні інвестиції» регламентується інструкцією (роз'ясненнями) щодо заповнення цієї форми, затвердженої наказом Державної служби статистики України від 14.06.2021 р.

Інструкція має бути чіткою і короткою, нерідко вона міститься в примітках, на обороті бланка і т. п. Докладно розробляються лише інструкції щодо складання звітів підприємств та організацій, а також проведення великих переписів.

Проведення спостереження вимагає високої кваліфікації працівників, які здійснюють процес первинного складання та обробки матеріалів. Для підвищення якості спостереження для цих осіб проводять *статистичний інструктаж* — навчання кадрів з програмно-методологічних та організаційних питань статистичного спостереження, що готується. Цей інструктаж найчастіше проводить інструктор-контролер, головним завданням якого є керівництво роботою деякої кількості реєстраторів у період проведення статистичного спостереження, та здійснення контролю за якістю їх роботи та прийом зібраних ними первинних статистичних матеріалів.

Іноді у формулярі після питання відразу ж даються деякі роз'яснення або перераховані можливі відповіді. Перелік можливих відповідей питання статистичного формуляра називається *статистичним підказом*. Якщо дається повний перелік відповідей на поставлене запитання, підказ називається *вичерпним* (*повним*), а якщо перераховуються не всі можливі відповіді, а деякі з них, то підказ називається *неповним*. Наприклад, у переписному листі Всеукраїнського перепису населення 2001 р. на запитання «Ваша стать» надані повні відповіді: 1 — чоловіча, 2 — жіноча. На питання «Ваше громадянство» надані такі можливі відповіді: 1 — Україна, 2 — без громадянства, 3 — інша держава (назва). В першому випадку підказ є повним, у другому — неповним.

Звернемося до питання 13 «Ваші джерела існування». Тут після питання перераховані 15 можливих відповідей, в яких вказано основні джерела існування нашого населення, а в 15 написано «інше джерело» — і в залишене місце для відмітки про це «інше» джерело засобів існування. Тут дана майже вичерпна відповідь, але все ж неповний підказ.

Повний підказ зобов'язує особу, яка заповнює формуляр, вибрати одну з перерахованих відповідей на запитання, неповний же орієнтує, якого характеру повинні даватися відповіді на поставлене питання.

А ось приклад питання, до якого дані пояснення, а не підказ. Питання 15 «Ваше заняття на основній роботі» (вкажіть докладну назву професії, посади або роботи, що виконується Вами). Це не підказ тому, що тут не вказані конкретні відповіді на питання, які могли б бути відміченими у формулярі (крім останнього).

**Час спостереження**. Складовою частиною програми статистичного спостереження є визначення часу проведення. *Час спостереження* — час, станом на який або за який реєструються відомості у процесі статистичного спостереження. Об'єкт статистичного спостереження перебуває у стані постійної зміни. Так, змінюється чисельність персоналу підприємства, його статевий, віковий та інший склад, обсяг випущеної продукції, вартість основних фондів, кількість шлюбів тощо. Тому проектуючи проведення статистичного спостереження, необхідно вирішити низку питань щодо часу його проведення.

Насамперед потрібно вибрати пору року, коли найдоцільніше проводити статистичне спостереження. Пора року залежить від об'єкта, мети та програми спостереження. Наприклад, під час перепису населення необхідно вибрати час, коли найменша міграція населення. Крім цього часу, потрібно ще визначити час, протягом якого слід визначити час, протягом якого слід здійснити спостереження, інакше кажучи, тривалість спостереження. Тривалість спостереження залежить від розмірів спостереження, програми, наявності кадрів, які можна залучити до цієї роботи і т. п. Потім необхідно точно встановити термін спостереження, вказавши дату початку та закінчення його. Наприклад, всеукраїнський перепис населення проводився з 5 по 14 грудня 2001 р.

В залежності від характеру об'єкта, що вивчається, від характеру визначених показників відомості можуть реєструватися або на певну дату (на початок або на кінець місяця, кварталу або року), або за певний проміжок часу (за добу, місяць, квартал, рік тощо), що залежить від характеру досліджуваних явищ. Наприклад, перепис обладнання може бути проведений лише на якийсь момент, а вивчення споживання електроенергії по окремих цехах або обсягу вироблених деталей лише за якийсь період (годину, зміну, день, декаду, місяць).

У першому випадку ми говоримо про так званий критичний момент спостереження, у другому — про період спостереження. Встановлення критичного моменту та періоду спостереження дуже важливо для якісного проведення всього спостереження. Від вибору часу значною мірою залежить повнота та оперативність даних, отриманих при організації та проведенні статистичного спостереження.

При організації статистичного спостереження має бути вирішено про час проведення даного спостереження, включаючи вибір сезону даного спостереження, встановлення терміну (періоду) спостереження.

Всі статистичні показники в залежності від часу спостереження можуть бути поділені на два класи. Один становлять показники, відомості про які можуть реєструватися лише станом на якийсь момент часу, інший — показники, відомості про які можуть реєструватися тільки за якісь проміжки часу. Перші назвемо моментними, другі — інтервальними показниками.

*Моментні показники* — це показники чисельності одиниць досліджуваної сукупності (об'єкта дослідження) та окремих її груп. Наприклад, кількість підприємств та організацій, чисельність працівників підприємства, чисельність робітників окремих професій, кількість одиниць встановленого обладнання, запасів сировини на складі тощо.

*Інтервальні показники* — це показники кількості та вартості виробленої продукції, витрат матеріалів на виготовлення продукції, числа народжених або померлих за певний період часу та ін.

Від вибору часу проведення статистичного спостереження великою мірою залежить точність отримання, успіх проведення дослідження. При цьому потрібно керуватися принаймні такими міркуваннями. Час, протягом якого проводиться спостереження, має відображати характерний стан об'єкта дослідження, тобто об'єкт повинен перебувати в «звичайному» для нього стані. Вибираючи час, необхідно враховувати технічні засоби, за допомогою яких буде проведено спостереження. Якщо об'єкт є дуже рухомим, то час має бути вибраний таким чином, коли цей об'єкт перебуває у найменш рухомому стані.

*Період* (*термін*) *спостереження* — період часу, протягом якого проводиться заповнення формулярів. Найчастіше це невеликий період часу. Однак бувають дуже короткі та надмірно тривалі терміни проведення статистичного спостереження.

Період спостереження визначається багатьма чинниками, найважливішими у тому числі є наступні. Насамперед час спостереження багато в чому залежить від характеру досліджуваного об'єкта, його розміру та стану, в якому він знаходиться. Чим більший об'єкт (за чисельністю одиниць спостереження), тим за інших рівних умов організація та проведення спостереження вимагатиме більшого часу.

Великий вплив на час проведення спостереження надає програма спостереження, її обсяг та складність ознак, що підлягають реєстрації. Чим ширша програма, що більше вона містить поставлених питань, тим більше потрібно часу щодо спостереження. Крім того, порядок встановлення значень ознак може вимагати додаткового часу, протягом якого здійснюється реєстрація одиниць спостереження.

Наявність кадрів, їх кількість і рівень підготовки багато в чому визначають терміни проведення спостереження. Що більше можна залучити кадрів під час проведення спостереження, то швидше можна провести спостереження, і навпаки.

Статистична наука виробила принципи, дозволяють уникнути похибки під час організації та проведення статистичних спостережень. Одним із таких принципів є принцип встановлення критичного моменту спостереження. *Критичним моментом спостереження* (як правило, перепису) називається момент часу, станом на який проводиться реєстрація (облік, запис у статичний формуляр) статистичних відомостей у процесі статистичного спостереження. Він встановлюється з метою отримання одноманітних, порівняннних статистичних даних. Критичний момент вибирають у найбільш зручний час, його визначають таким чином, щоб зібрані відомості давали характеристику об'єкта спостереження у його типових, найстійкіших рисах. Таким критичним моментом Всеукраїнського перепису населення 2001 р. було 12 год. ночі з 14 на 15 грудня 2001 р.

Справді, грудень — зимовий місяць, а взимку рухливість населення значно нижча, ніж навесні та влітку. Із зимових місяців грудень є найбільш вдалим тому, що зібрані у процесі перепису дані найкраще узгоджуються зі статистикою руху населення, яка веде в межах календарного року. З грудня не підходять дні з 21 грудня до 31 грудня (студентські канікули). Із дат 28 грудня не підходить, бо є передсвятковим днем. В результаті висновків зрештою було прийнято рішення як критичний момент при проведенні перепису населення взяти 12 годин ночі з 14 на 15 грудня. Час 12 годин вибрано тому, що рухливість населення у цей час доби мінімальна; цей час ув'язується і з реєстрацією народжень та смертей, що враховує відповідні випадки з точністю до доби.

У практиці статистичного дослідження зазвичай прагнуть до того, щоб реєстрація даних (заповнення формулярів спостереження) відбувалася не надто віддалено від критичного моменту, оскільки чим більше цей розрив, тим більші зміни можуть статися в об'єкті спостереження, що негативно позначиться на його результатах.

Для всіх форм статистичної звітності, які здають підприємства в державні органи статистики, встановлюється термін (число місяця або кварталу), не пізніше якого дані мають бути надані за призначенням. Чим коротший період спостереження, тим швидше мають бути надані відомості після закінчення цього періоду.

## 2.3. Найважливіші організаційні питання статистичного спостереження

**Організаційний план статистичного спостереження**. Крім програмно-методичних питань, підготовка та проведення спостереження включає в себе вирішення проблем організаційного характеру. З метою успішного проведення статистичного спостереження слід скласти докладний організаційний план спостереження. *Організаційний план спостереження* — це спеціальний документ, в якому зафіксовані всі найважливіші організаційні заходи, проведення яких необхідне для успішного здійснення спостереження. Найважливішими організаційними питаннями плану статистичного спостереження є: встановлення інформуючої (звітної) одиниці, місця та часу спостереження (строків спостереження), порядку прийому та здачі матеріалів; порядку отримання та складання попередніх та остаточних підсумків, визначення форм та видів спостереження, підбір та навчання кадрів, розмноження необхідної документації, способів реєстрації фактів.

Важливою умовою успішного складання організаційного плану спостереження є попереднє виконання теоретичного аналізу досліджуваної сукупності — знання сутності явища, що вивчається, особливості його розвитку в даних умовах часу і місця. З метою перевірки дієвості розробленої програми та організаційного плану статистичного спостереження проводять пробні обстеження невеликого обсягу одиниць сукупності.

**Органи спостереження**. В організаційному плані спостереження вказується *орган* (*виконавець*) статистичного спостереження, який виконує функції з організації збору та проведення статистичного спостереження та несе відповідальність за цю роботу. Функції організатора та виконавця статистичного спостереження у загальнодержавному масштабі виконують Державна служба статистики України (далі ДСС — України), а також міністерства та відомства (Державна фіскальна служба України, Національний банк України, Міністерство науки і освіти України тощо).

Міністерствами, відомствами, рідше науковими та іншими установами здійснюється підготовка та проведення статистичних спостережень локального характеру.

Крім постійних органів, які здійснюють статистичне спостереження, іноді, головним чином для проведення великих обстежень, створюються тимчасові органи (комісії, відділи, бюро, сектори тощо) у відповідних установах.

На підприємствах органами спостереження можуть бути різні відділи та служби підприємства: бухгалтерія, технологічний відділ, відділ праці та заробітної плати, відділ головного механіка, відділ технічного контролю, інструментальний відділ, відділ кадрів та ін. Але головним виконавцем статистичного спостереження є планово-економічний відділ та планово-економічне бюро цеху. Саме на ці відділи спільно з іншими займаються організацією та проведенням спостереження за перебігом технологічного процесу, за якістю продукції, станом запасів сировини, матеріалів тощо. На підприємствах інколи виконавцем спостереження може бути одна особа.

**Інформуюча** (**звітна одиниця**). Важливими організаційними питаннями плану статистичного спостереження є встановлення інформаційної (звітної) одиниці[1]. В теорії статистики *інформуюча одиниця* — це той орган (чи окрема особистість), якій адресуються отримання

[1] При спеціально організованому спостереженні застосовується термін інформуюча одиниця, а при звітно(обліковано)-статистичному спостереженні — звітна одиниця.

даних, що характеризують одиниці об'єкта статистичного спостереження. Наприклад, при переписі обладнання заводу інформучою одиницею є окреме підприємство, при фотографії робочого часу — окремий робітник. В окремих випадках одиниця спостереження та інформуюча одиниця можуть збігатися. Так, при обстеженні кваліфікаційного складу робітників на якомусь заводі в одному випадку (залежно від організації обстеження) робітник буде одиницею спостереження та інформаційною одиницею, в іншому випадку робітник буде лише одиницею спостереження, а інформаційною одиницею буде цех.

**Місце спостереження**. З метою одноманітності зібраних відомостей щодо статистичного спостереження відзначаються межі явища. *Місце спостереження* — це місце, де повинна проводитися реєстрація фактів, що спостерігаються, де заповнюються формуляри спостереження.

Місце проведення спостереження вибирають, виходячи зі способу спостереження та особливостей об'єкта спостереження. На практиці складність із визначенням місця спостереження виникає лише при спеціально організованому статистичному спостереженні і притому щодо тих одиниць об'єкта, які можуть змінювати своє місцезнаходження. Від точного визначення місця спостереження у значній мірі залежить повнота охоплення одиниць об'єкта дослідження. Найбільш характерним прикладом спостереження, де виникає складність при визначенні місця спостереження, є перепис населення. Під час проведення перепису люди можуть знаходитися вдома і на роботі, в гостях, на відпочинку, у відрядженні, у дорозі (у поїзді, у літаку, в автомобілі і т. п.). Як визначити місце перебування населення, що підлягає перепису. В цьому випадку необхідне точне визначення місця проведення перепису населення, яке визначається в інструкції до його проведення.

В деяких випадках встановлюються особливі місця для реєстрації об'єктів спостереження. Інформуюча або звітна одиниця повинна з'явитися в ці місця (пункти реєстрації) та надати певні відомості. До таких пунктів реєстрації можна віднести Державну автоінспекцію, відділи реєстрації актів цивільного стану. Хоча Державна автоінспекція не відноситься до статистичних відділів, проте саме від неї органи державної статистики ведуть спостереження щодо забезпечення безпеки дорожнього руху в країні — кількості дорожньо-транспортних пригод, числа постраждалих, кількості та технічного стану транспортних засобів.

## 2.4. Основні організаційні форми, види та способи статистичного спостереження

**Основні організаційні форми статистичного спостереження**. Форми статистичного спостереження виділяють з урахуванням їх найбільш загальних організаційних особливостей. Все різноманіття організаційних форм статистичного спостереження у статистиці зводиться до двох основних форм — звітність та спеціально організоване спостереження.

*Звітність* — це форма статистичного спостереження, при якій відповідні органи отримують від підприємств, установ та організацій необхідні їм статистичні дані у вигляді встановлених у законному порядку документів (статистичних звітів). Звітність характеризується регламентацією всіх програмно-методологічних та організаційних питань та надається у суворо встановлені терміни. Склад, обсяги та методологія розрахунків показників, адреси та рядки подання статистичної інформації, зазначені в звітно-статистичній документації, є *обов'язковими* для всіх респондентів (ст. 18 Закону України «Про державну статистику»). Підприємства та організації зобов'язані надавати статистичні відомості про свою діяльність статистичним органам або державним відомчим установам (або тим та іншим одночасно). Керівництво підприємств, установ та організацій, що надають звітність, несе відповідальність своєчасне надання її та достовірність сполучених відомостей.

В Україні статистична звітність має винятково велике значення. Нею охоплено все народне господарство країни. Надання звітів є обов'язковою функцією підприємств, державних та кооперативних установ та організацій. Статистичні звіти є основним джерелом статистичної інформації, необхідної для прогнозування економічного та соціального розвитку країни, забезпечення потреб регулювання економіки на загальнодержавному, галузевому та регіональному рівнях управління.

У статистичних звітах містяться різні дані, що характеризують результати господарської та фінансової діяльності підприємств, установ та організацій за певний період. Коло ознак та показники статистичних звітів містять інформацію, одержану на основі первинної реєстрації господарських операцій, що здійснюється відповідно до національних стандартів України. Ці дані визначаються на основі безперервної в часі реєстрації фактів, хоча самі форми звітів, в яких даються вже зведені підсумки, можна представляти за укрупненими періодами (місяцях, кварталах тощо). Дані звітів служать для цілей розробки системи заходів щодо регулювання народного господарства та її окремих секторів і галузей, контролю виконання плану, розробки нових планів і оперативного регулювання. Статистичні звіти використовуються не лише у практичних, а й у наукових цілях. Важливо відзначити, що статистична звітність, як правило, базується на документальному обліку господарських операцій підприємств та організацій.

Розрізняється: *звітність загальновідомча* та *звітність внутрішньовідомча*, *звітність типова* та *звітність спеціалізована*, *звітність первинна* та *звітність зведена*; особливо виділяється *звітність тимчасова оперативна*.

Статистичну звітність поділяють на загальнодержавну та внутрішньовідомчу. *Загальнодержавна звітність*, періодичність, способи та терміни її подання є обов'язковою для всіх підприємств, установ та організацій. Статистична інформація, зібрана з урахуванням загальнодержавної звітності, використовується потреб державного управління. *Внутрішньовідомча звітність* збирається для потреб відомствами, міністерствами.

Залежно від суб'єкта подання звітність поділяється на первинну та зведену. *Звітність первинна* — звітність, що надається безпосередньо звітними одиницями — як правило, первинними підрозділами установ статистичної служби країни. *Звітність зведена* — статистичні відомості, що є результатом зведення статистичних даних, що містяться у первинній звітності.

За періодичністю розрізняється *періодична* звітність (надається через однакові проміжки часу або точно в певні дати, наприклад, 3-го числа кожного місяця, не пізніше 1 листопада кожного року) та *одноразова* звітність (надається при необхідності, без певної періодичності або неповторна). У свою чергу, періодична звітність поділяється на *поточну* (період надання якої менше одного року: квартал, місяць, тиждень), *річну* (період надання — календарний рік), а також звітність, що надається один раз на рік поза зв'язком з початком або кінцем календарного року , та звітність з періодом понад рік (двічі на 5 років, один раз на два роки тощо).

В залежності від цілей, які ставляться під час проведення статистичного спостереження, звітність поділяється на оперативну та підсумкову. *Звітність оперативна* призначена для поточного спостереження за соціально-економічними процесами в структурних одиницях, відділеннях, підрозділах, відомствах, дільницях в залежності від потреб управління. На відміну від оперативної звітності, *підсумкова звітність* служить для узагальнення результатів статистичних спостережень за тривалий період, які служать в основі стратегічного управління та регулювання даних процесів або явищ, а також для їх узагальнення, виявлення закономірностей.

В Україні існує і діє організована в державному масштабі суворо регламентована система звітності. Статистична звітність вводиться і затверджується лише відповідними державними органами та в порядку, ними встановленому. Державна служба статистики України затверджує загальнодержавний мінімум показників і форм статистичної звітності для підприємств, установ та організацій. Форми бухгалтерської звітності затверджуються Міністерством фінансів України. Забороняється всім державним органам управління, громадським організаціям вимагати, а підприємствам і організаціям надавати будь-які статистичні звіти, не передбачені державною звітністю.

Кожна форма звітності має свою назву, номер, в ній вказується період, за який вона надається, термін надання, назва підприємства, підпорядкованість його та ін. Обов’язкова сукупність деяких зовнішніх елементів у складі кожної форми звітності називається *реквізитом* даної форми.

Поряд із звітністю найважливішим джерелом статистичних даних виступає спеціально організоване статистичне спостереження (перепису та різні обстеження).

*Спеціально організоване статистичне спостереження* являє собою спостереження, яке проводиться з метою збирання статистичних даних, відсутніх у статистичній звітності або з метою перевірки, уточнення даних звітності, а також для вирішення самостійних науково-практичних завдань. Прикладами спеціально організованого статистичного спостереження можуть бути перепису населення, що проводиться для вивчення чисельності та складу населення (вікового, статевого, національного складу тощо); проведення інвентаризації основних фондів та товарно-матеріальних цінностей на підприємствах та організаціях; облік використання робочого дня працівників окремих професій шляхом проведення хронометражу.

Причина існування спеціально організованого статистичного спостереження полягає в тому, що значна кількість явищ і процесів не знайшла відображення у статистичній звітності. Так, не надаються звіти про зміни у кваліфікації працівників підприємств, про використання робочого часу робочих окремих професій, про норми витрати окремих видів матеріалів на виробництво продукції. Такі дані можна отримати, організувавши спеціальне спостереження. Необхідність проведення спеціально організованого спостереження виникає і в тих випадках, коли за звітами немає можливості встановити результати окремих видів та сфер діяльності підприємств та організацій. Для глибшого пізнання сутності процесів, які відбуваються у суспільному житті країни, часто потрібні більш детальні дані. Наприклад, у звітності не надаються докладні дані про наявність, технічний стан і використання обладнання, транспортних засобів, окремих видів сировини, про витрати праці на виробництво окремих видів продукції тощо. Крім того, нерідко виникає необхідність перевірити чи уточнити відомості, які у звітах підприємств і закупівельних організацій. Інформація, що міститься у звітах, не дозволяє це зробити. Тому інформацію про це збираються шляхом проведення різноманітних переписів та обстежень.

У нашій країні спеціально організоване статистичне спостереження поряд зі звітністю набуло широкого поширення. Для оперативного управління спеціально організоване спостереження є одним з найважливіших джерел статистичних даних по підприємству.

З впровадженням електронно-обчислювальної техніки та математичних методів зароджується нова організаційна форма спостереження. Шляхом введення даних у обчислювальні пристрої створюється великі масиви інформації, які можуть бути швидко знайдені, комбіновані (узагальнені) за певними параметрами та видані споживачеві. Така організаційно-технічна система концентрації, зберігання та видачі статистичних даних має назву *автоматизованих баз даних* (*АБД*). Отримання, накопичення та використання статистичної інформації за допомогою розробки та впровадження АБД є найважливішою формою статистичного спостереження. Найбільш суттєвою її відмінністю від традиційних форм статистичного спостереження є автоматизація функцій зберігання та пошуку даних. Прикладом автоматизованої бази даних є Єдиний державний регістр підприємств і організацій України (ЄДРПОУ) — автоматизована система збирання, накопичення та опрацювання даних про всіх юридичних осіб, що відокремлені підрозділи юридичних осіб, що знаходяться на території

України, а також відокремлені підрозділи юридичних осіб України, що знаходяться за межами України.

**Види статистичного спостереження**. За видами статистичне спостереження класифікується залежно від часу проведення та повноти охоплення спостереженням досліджуваної сукупності.

1. В залежності від ступеня охоплення досліджуваної сукупності статистичне спостереження поділяється на суцільне та несуцільне. *Суцільним* називається спостереження, при якому обстеженню піддаються всі без винятку одиниці досліджуваного об'єкта статистичного спостереження. Типовим прикладом суцільного спостереження є переписи населення, у яких збирають інформацію про кожного жителя цієї країни. До суцільного спостереження відносять реєстрацію новонароджених дітей, облік скоєних злочинів, дорожньо-транспортних пригод, надання звітів державних та приватних підприємств тощо.

*Несуцільним* називають таке спостереження, при якому проводиться обстеження частини об'єкта статистичного спостереження, причому так, щоб на основі відібраної частини можна роботи судження про всю сукупність. Остання обставина особливо важлива: не можна вважати несуцільним спостереження, яке було задумане як суцільне, але частина сукупності чомусь не була піддана спостереженню, і воно тому перетворилося на несуцільне. Точно так само не може бути названо несуцільним спостереження, мета якого — отримати статистичні дані про частину одиниць сукупності для вивчення лише цієї частини, оскільки як стосовно одиниць сукупності, як об'єкта дослідження, воно буде суцільним.

При несуцільному спостереженні відомості збираються не про всі одиниці сукупності, а лише про деяку їх частину, відібрану певним чином. Причому заздалегідь, при організації вибіркового спостереження встановлюється: по-перше, сам факт, що проводиться обстеження частини об'єкта статистичного спостереження; по-друге, яка саме частина одиниць сукупності повинна бути піддана спостереженню і яким чином необхідно відібрати ті одиниці, які повинні бути обстежені.

Найважливішою характерною особливістю несуцільного спостереження є те, що за результатами дослідження частини елементів, що становлять досліджуваний процес, судять про всю сукупність явищ. Наприклад, у масовому та серійному виробництві проводиться обстеження частини випущеної продукції, за результатами якої судять про якість усієї виробленої продукції. Необхідною умовою отримання правильних висновків та узагальнень, отриманих на основі дослідження частини елементів сукупності, як зазначалося вище, є типовою відібраною для обстеження сукупності.

Під час проведення несуцільного спостереження не ставиться завдання повністю суцільного охоплення всіх фактів статистичного спостереження. Його заздалегідь організують як облік частини одиниць сукупності, але досить масовий, щоб мати можливість отримати достовірні та надійні узагальнюючі статистичні характеристики. Так організовано, наприклад, статистику сімейних бюджетів. Іншим прикладом несуцільного спостереження є статистика цін.

Безумовно, суцільне спостереження забезпечує повне охоплення одиниць статистичної сукупності. Висновки, отримані на основі суцільного спостереження, вирізняються високим ступенем точності. Поряд із суцільним спостереженням у статистиці дуже часто користуються несуцільним спостереженням. Це викликане головним чином міркуваннями економії матеріальних ресурсів та праці, що витрачаються на збір, обробку даних, яка досягається при несуцільному спостереженні. У деяких випадках несуцільне спостереження є єдиним можливим, наприклад, при контролі якості промислової продукції або вивченні використання робочого часу працівників окремих професій.

У практиці статистичної роботи застосовуються кілька видів несуцільного спостереження. Основними є: вибіркове спостереження, метод основного масиву, монографічне обстеження.

*Вибірковим* називається таке статистичне спостереження, під час якого реєструються не всі елементи статистичної сукупності, лише деяка їх частина, відібрана з урахуванням проведеного за відомими правилами ненавмисного випадкового відбору. Сутність вибіркового методу полягає в тому, що для вивчення цілого береться частина його, на підставі даних про цю частину припускають скласти судження про склад та властивості всієї сукупності фактів чи явищ. Вибіркове спостереження за правильної організації та проведення дозволяє отримати досить точні дані для характеристики досліджуваної сукупності явищ в цілому.

Вибірковий метод застосовується тоді, коли проведення суцільного обстеження дуже важко і пов'язані з великими витратами праці та часу. Оскільки вибірковому спостереженню присвячена особлива глава, тут немає потреби у його розгляді.

*Метод* (*обстеження*) *основного масиву* являє собою спостереження, при якому обстеженню (суцільному) піддається не вся сукупність, а найбільші одиниці спостереження, питома вага яких переважає в обсязі досліджуваних первинних ознак. Ці одиниці статистичного спостереження відіграють визначальну роль в характеристиці об'єкта статистичного спостереження. Взяті разом вони займають переважну питому вагу в сукупності здебільшого для досліджуваною ознакою (або кількома ознаками). Тому спостереження основного масиву називають ще спостереженням несу цільним недосконалим. У той же час одиниці сукупності, які займають незначну питому вагу, обстеженню не піддаються. Їх чисельність може бути значно більшою, ніж чисельність великих одиниць, проте їх значення для узагальненої характеристики досліджуваного явища або процесу невелике. Наприклад, вирощуванням птиці в Україні у 2020 р. займалося 385 підприємств, у тому числі фермерські господарства, птахофабрики. Однак у 107 підприємств, що становить 27,8 % їх загальної чисельності, містилося 95,8 % від усього поголів'я птиці. У решті 278 підприємств (72,2 %) чисельність поголів'я птиці становило лише 4,2 % від загального поголів'я птиці. Таким чином, обстеження великих одиниць сукупності є достатньою для судження про динаміку всього поголів'я птиці в Україні. Прикладом такого спостереження може бути вивчення використання робочого дня у найбільших цехах підприємства. Результати обстеження цих цехів можуть розглядатися як результати, характерні для всього підприємства.

*Монографічне обстеження* — докладний опис окремих, але типових для певної сукупності одиниць об'єкта статистичного спостереження. Об'єктом статистичного спостереження є представники будь-яких типів явищ, наприклад, передові, середні чи збиткові підприємства; працівники, що допускають великий відсоток браку продукції тощо. Проводиться з метою виявлення наявних чи намічених тенденцій у розвитку явищ або для досконального вивчення окремих типових одиниць статистичної сукупності. Воно може також застосовуватися для виявлення причин, що зумовили високу ефективність того чи іншого підприємства, або навпаки, для вивчення причин, що зумовили недоліки в його роботі тощо. Наприклад, монографічним дослідженням буде детальне вивчення використання робочого часу робітниками, що значно перевиконують норми вироблення (часу), а також систематично не виконують норми виробітку (часу). В обох випадках вивчаються окремі типові об'єкти.

Детальне вивчення одиничних, але типових об'єктів дозволяє зрозуміти закономірності і тенденції розвитку всіх одиниць сукупності. Особливо важливе значення у цьому відношення має вивчення діяльності передових підприємств та організацій.

2. За часом реєстрації спостереження може бути безперервним, або, інакше, поточним та перервним. При *поточному спостереженні* здійснюється безперервна реєстрація фактів у міру їх виникнення. За допомогою безперервного спостереження у принципі забезпечується повне охоплення одиниць статистичної сукупності у часі. При дослідженні явища облік фактів відбувається з їх виникнення. Поточне спостереження дозволяє з вичерпною повнотою врахувати розвиток масового суспільного явища чи процесу у часі, що дуже важливо задля встановлення тенденцій та закономірностей у його розвитку. Прикладом безперервного спостереження можна назвати надання звітності суб'єктами підприємницької діяльності, в основі якої лежить безперервна первинна реєстрація фактів на підприємствах та в установах. Безперервними спостереженнями є облік робочого часу (табельний облік), реєстрація актів цивільного стану ДРАЦСами, облік числа народжень, смертей, дорожніх пригод, пожеж тощо у спеціальних державних органах.

*Перервним* називають таке спостереження, яке проводиться час від часу. Воно відображає, як правило, стан об'єкта та динаміку процесу на певні дати або моменти спостереження. Класичним прикладом безперервного спостереження є перепис населення. До перервного спостереження можна віднести, наприклад, вивчення використання робочого дня на заводі, в цеху чи ділянці, якщо воно проводиться час від часу. Перервними спостереженнями є проведення обліку складських запасів готової продукції, залишків сировини та матеріалів, виконання норм виробітку тощо.

У свою чергу, перервне спостереження поділяється на одноразове та регулярне (періодичне).

*Одноразове спостереження* — таке спостереження, яке організується в одноразовому порядку або проводиться іноді без дотримання суворої періодичності його повторення. Проводиться з метою збирання статистичних даних, які не можна отримати іншим шляхом. До одноразових спостережень можна віднести комплекс хронометражних спостережень, що проводяться з метою розробки технічно обґрунтованих норм виробітку (часу) на будь-яку нову технологічну операцію. Так, наприклад, отримують дані під час перепису населення, при врахуванні норм витрати пального, встановленні використання робочого часу тощо.

*Періодичним* називають таке спостереження, яке повторюється через визначені, рівні проміжки часу. Такі, наприклад, облік основних фондів, устаткування, проведений за станом 1 січня, періодичний облік запасів готової продукції складі, вивчення використання робочого дня проводяться регулярно. Поточне перервне спостереження називають *регулярним*.

Періодичність чи регулярність повторення спостереження — найважливіший принцип проведення науково обґрунтованого статистичного спостереження. Послідовне здійснення даного принципу дозволяє систематично стежити за зміною даного явища або процесу, виявляти закономірності та тенденції у їх розвитку. Найчастіше ці спостереження проводять для характеристики стану будь-яких явищ на визначений момент часу.

Підставою для реєстрації відповіді питання формуляра можуть бути безпосереднє встановлення фактів працівників, які проводять спостереження, або відповідні документи, або показання опитуваних осіб.

3. Залежно від способу збирання даних розрізняють безпосередньо спостереження, документальне спостереження та опитування.

*Безпосереднім* називають таке спостереження, у якому джерелом записів даних у формуляр статистичний служить безпосередній рахунок одиниць спостереження у натурі чи вимір їх ознак (параметрів) з допомогою будь-яких технічних засобів, наприклад, зважування. При безпосередньому спостереженні працівники, які проводять спостереження, фіксують стан елементів досліджуваної сукупності на основі особистого огляду, підрахунку, вимірювання та обліку (наприклад, хронометраж, фотографія робочого дня, моментно-вибіркове спостереження). Такі, наприклад, інвентаризація майна, підрахунок ваги тварин, облік залишків пального складі тощо.

Перевагою безпосереднього спостереження порівняно з іншими способами є те, що він являється досить точним і достовірним способом отримання вихідних статистичних даних. Проте його проведення потребує великих витрат праці та коштів. У багатьох випадках безпосереднє спостереження не застосовується, оскільки статистика часто цікавлять факти, що відбувалися в минулому (час і місце народження мешканців), або факти, що не піддаються такому спостереженню (національність жителів).

При *документальному способі спостереження*, запис відповіді питання формуляра здійснюється виходячи з відповідних (переважно облікових) документів[1]. Так, наприклад, при переписі транспортних засобів різні відомості про автомобілі можуть бути отримані на підставі їх технічного паспорта. Цей спосіб спостереження застосовується головним чином при заповненні підприємствами, організаціями та установами форм статистичної звітності. Статистика тісно пов'язана з первинним, оперативно-технічним та бухгалтерським обліками. Вона користується цими видами обліку як джерелами інформації та черпає з них документів облікових даних для своїх потреб.

Спосіб *опитування* характеризується тим, що відомості про кожну одиницю спостереження записуються за словами опитуваних осіб. На основі опитування заповнювалися, зокрема, формуляри Всеукраїнського перепису населення 2001 р. Підставою для запису відомостей про вік, освіту, сімейний стан, громадянство, джерела доходу тощо служать відповіді опитуваного.

Найбільша точність зібраних відомостей досягається при безпосередньому, а також документальному спостереженні.

**Способи статистичного спостереження**. Статистичне спостереження може здійснюватися різними способами У статистиці застосовуються такі способи спостереження: звітний, експедиційний, самообчислення, анкетний, кореспондентський, явний. Найбільшого поширення набули звітний та експедиційний способи спостереження.

Вище ми вже розглянули звітність як одну з найважливіших форм статистичного спостереження нашій країні. Суть *звітного способу* спостереження, як сказано, полягає в тому, що відповідні органи отримують від підприємств, організацій та установ необхідні їм статистичні дані у вигляді встановлених законом порядку звітних документів (статистичних звітів). Це найбільш поширений спосіб спостереження в українській статистиці. Основні дані, що характеризують соціально-економічний розвиток країни, виконання державного бюджету виходять за допомогою звітності.

Застосування звітного методу має необмежені можливості механізації та автоматизації збору статистичних даних у державних органах статистики. Саме він створює передумови для створення автоматизованої системи збирання, накопичення та опрацювання даних про всіх юридичних осіб, їх філії, відділення, представництва та інші відособлені структурні підрозділи, що знаходяться на території України.

---

[1] Іноді використовується термін «звітний спосіб спостереження».

При звітному методі не створюється інституту спеціальних працівників (лічильників). Заповнення форм статистичної звітності та їх надання здійснюють планові та бухгалтерські підрозділи або спеціальні робітники, які є в штаті підприємств та організацій.

*Експедиційний спосіб спостереження* — спосіб статистичного спостереження, що здійснюється спеціально підготовленими і навченими для цієї мети особами, які з метою реєстрації фактів, що спостерігаються, відвідують кожну одиницю спостереження (наприклад, жителів країни при переписі населення) і самі записують дані в статистичний формуляр. Цей спосіб, як і спосіб звітності, забезпечує отримання достовірної та повної інформації про об'єкт спостереження. Він застосовується лише за спеціально організованому спостереженні. Осіб, які залучають та навчають для проведення такого спостереження, зазвичай називають лічильниками або реєстраторами. При збиранні даних експедиційним методом статистик особисто входить у контакти з обстежуваними одиницями, спостереження здійснюється кваліфікованими кадрами. Статистик особисто оглядає, вимірює, рахує явища (наприклад, оглядає, рахує та записує кількість обладнання на підприємстві при щорічному обліку основних засобів). Тому названий спосіб, як і звітний, забезпечує високу достовірність та точність даних про одиниці спостереження. Крім того, використання кваліфікованих кадрів дозволяє зібрати дані щодо складної програми спостереження. Однак експедиційному способу притаманна й низка недоліків: він громіздкий, дорогий і вимагає використання кваліфікованих працівників.

Незважаючи на великі витрати праці та засобів до цього способу спостереження доводиться вдаватися у тих випадках, коли він є єдиним можливим способом отримання достовірних відомостей про об'єкт спостереження.

При *способі самообчисленні* (*самореєстрації*) обстежуваній особі вручають бланк обстеження і пояснюють питання, бланк заповнює обстежувана особа. Так організується, наприклад, самофотографія робочого дня. Під час самореєстрації також залучають спеціально навчених працівників. Обов'язки лічильників, що залучаються щодо спостереження, перебувають у роздачі формулярів опитуваним, інструктажі їх, зборі заповнених формулярів і перевірці правильності їх заповнення.

Для того, щоб розраховувати на отримання доброякісного матеріалу, необхідно, щоб програма такого спостереження повинна бути коротка і досить проста, щоб не викликати труднощів осіб, які опитуються.

При самореєстрації повнота і точність статистичних даних значно вища, ніж при кореспондентському способі, хоча процес самореєстрації триваліший і обходиться дорожче. У порівнянні з експедиційним способом спосіб самореєстрації дозволяє скоротити витрати лічильників, оскільки вони не несуть витрати часу на заповнення формулярів спостереження. Однак він поступається експедиційному щодо точності отриманих відомостей про об'єкт спостереження. Адже спеціально залучені та навчені працівники, які зазвичай називають лічильниками або реєстраторами, проводять спостереження краще за тих осіб, які отримують короткі роз'яснення щодо порядку проведення спостереження та правил заповнення формуляра спостереження.

*Анкетне обстеження* — спосіб збирання статистичних даних, що полягає у розсилці, роздачі певному колу осіб анкет (запитань), повернення яких у заповненому вигляді є справою добровільною. При анкетному способі особам, від яких хочуть отримати будь-які відомості, розсилають спеціальні запитувачі (анкети). Анкета — структурно організований набір запитань, кожне з яких логічно пов'язане з основною метою дослідження, отримання інформації відбувається шляхом опитування респондентів.

Анкетне обстеження ґрунтується на принципах добровільності та зазвичай на анонімному заповненні анкет (тобто без вказівки заповнюючим своє прізвище, адреси тощо). Зазвичай кількість анкет повертається значно менше, ніж розсилається. З огляду на це анкетне спостереження часто виявляється несу цільним навіть за умови, що анкети розіслані всім одиницям спостереження.

При отриманні анкет неможливо встановити і перевірити точність отриманих у яких відомостей. Крім того, часто неможливо встановити перелік осіб, які не надіслали відповіді. Тому при проведенні спостереження за допомогою анкетного способу неможливо зробити якісну оцінку точності його результатів, визначити помилки спостереження.

У нашій країні анкетний спосіб часто використовується організаціями при контролі роботи свого апарату, до нього іноді вдаються при проведенні соціологічних та маркетингових досліджень. Наприклад, розсилається анкета літературного чи спеціального журналу передплатникам; пропонується анкети учасникам сертифікації для самооцінювання педагогічної майстерності; споживачам надаються анкети з метою оцінки рівня довіри до реклами; видаються анкети визначення рівня мотивації співробітників компанії; жителям мікрорайону надсилаються анкети про своєчасність доставки ним газет, листів тощо.

Теоретично можна уявити таку ситуацію, коли від усіх опитуваних осіб отримано письмові відповіді питання. Отже, анкетне обстеження охоплює всі одиниці об'єкта спостереження, тобто стає суцільним.

Суть *кореспондентського способу* спостереження полягає в тому, що статистичні дані про явища, що вивчаються, повідомляються обізнаними компетентними особами, які не перебувають у штаті статистичного органу, — кореспондентами. Ці особи беруть на себе зобов'язання вести спостереження за будь-якими явищами, процесами й у встановлені терміни повідомляють результати спостережень статистичним органам. Відомості від кореспондентів надходять поштою, телефоном або доставляються особисто.

Статистичні органи забезпечують кореспондентів бланками, інструкціями щодо їх заповнення, а також усім іншим, необхідним для проведення спостереження. Цих кореспондентів зазвичай називають добровільними кореспондентами, оскільки вони не перебувають у штаті статистичних органів. Процес збору статистичних відомостей кореспондентським шляхом вимагає менше витрат часу, праці та коштів на отримання інформації, ніж експедиційним шляхом. Однак при цьому виникають труднощі із забезпеченням повного охоплення одиниць спостереження і збільшуються ризики появи помилок. Якість отриманої інформації залежить від рівня знань та ступеня підготовки кореспондентів. Відомості, що повідомляються кореспондентами, засновані не на документах або особистому встановленні фактів шляхом виміру, зважування, підрахунку тощо, а є суб'єктивною (експертною) оцінкою явища (наприклад, якості виробу).

Кореспондентський метод мав широке поширення при дослідженні попиту населення товари широкого споживання. З його допомогою виробники чи торгові підприємства ведуть спостереження за процесами, що відбуваються на споживчому ринку.

Таким чином, статистичні дані можна отримати різними шляхами та способами. Вибір перерахованих вище організаційних форм, видів та способів спостереження залежить від мети та завдань, поставлених перед спостереженням; від сутності об'єкта, що піддається обстеженню; від програми спостереження; від наявності відповідних кадрів;

від коштів, які мають статистичні органи; від терміновості потреби у тих чи інших статистичних даних.

## 2.5. Точність статистичного спостереження

Точність — найважливіша вимога, що пред'являється до статистичного спостереження, оскільки вихідні матеріали піддаються подальшій обробці та аналізу. Для того, щоб статистичні дані, зібрані в процесі статистичного спостереження, могли бути придатними для обробки та правильних узагальнень, вони мають бути повними, достовірними та порівнянними. Якість вихідного статистичного матеріалу визначає якість узагальнюючих показників, отриманих у результаті статистичної обробки (статистичного зведення). Якщо отримані в результаті статистичного спостереження вихідні дані будуть неякісними, в них будуть мати місце похибки (помилки), то на наступних етапах дослідження виправити ці помилки не вдасться. Тому питанням попередження помилок під час підготовки спостереження завжди необхідно приділяти значну увагу.

*Точністю статистичного спостереження* називають близькість даних про об'єкт статистичного спостереження, отриманих у результаті статистичного спостереження, до дійсних значень. Чим ближче значення показника, отриманого внаслідок організації та проведення статистичного спостереження, до фактичних значень, тим вища точність статистичного спостереження. Розбіжності (невідповідність) між величиною будь-якого показника, знайденої за допомогою статистичного спостереження, і його розмірами називаються *помилками статистичного спостереження*. Вони є наслідком неточностей при реєстрації значень одиниць сукупності, що вивчається.

Точність та достовірність матеріалів статистичного спостереження забезпечується чітким, правильним вирішенням усіх програмно-методологічних та організаційних питань, про які йшлося вище, а також високого рівня постановкою обліку. Проте навіть за досить досконалої організації статистичного спостереження у процесі збирання та обробки масових даних може бути похибки (помилки). Вони призводять до зниження точності результатів, одержуваних під час проведення статистичного спостереження. «У статистиці, — писав акад. В. С. Немчинов, — як і у будь-якій науці, з неправильних передумов не можна отримати істину; не можна шляхом ретельної наукової обробки результатів спостереження отримати правильні висновки, якщо саме спостереження було не науково організованим. Найбільшими та непоправними помилками у статистиці є помилки, що виникають внаслідок неправильного та ненаукового перекладу соціальних, економічних, агрономічних, технологічних та біологічних понять у рахункові категорії. Ніяка наукова обробка неправильно побудованих показників та вимірювачів не може забезпечити отримання правильних результатів». Тому первинний матеріал перед його обробкою піддаються попередньому контролю. Мета контролю – виявлення та виправлення помилок.

**Джерела та типи похибок (помилок) статистичного спостереження**. Помилки статистичного спостереження різноманітні за походженням та характером. Вони можуть виникати внаслідок неповного охоплення підлягають реєстрації одиниць сукупності, пропусків або неясних записів окремих даних, і, нарешті, неправильних записів окремих відповідей. Помилки статистичного спостереження можуть полягати у зв'язку з низькою кваліфікацією або відсутність навичок у реєстраторів, що фіксують ті чи інші події. У деяких випадках трапляються і навмисні помилки, що призводять до викривлення фактів; у таких випадках відповідальних осіб, винних у скоєнні навмисних помилок, притягають до відповідальності осіб.

Помилки статистичного спостереження поділяються на категорії в залежності від причин виникнення та характеру дії. За джерелами виникнення розрізняють помилки ненавмисні і навмисні, за значенням — випадкові і систематичні.

*Помилка реєстрації* — розбіжність між зафіксованим при статистичному спостереженні значенням ознаки одиниці спостереження і дійсним його значенням. Помилки реєстрації виникають або внаслідок неправильного встановлення фактів у процесі спостереження або неправильного їх запису (реєстрації). Ці помилки є результатом низької кваліфікації особи, яка проводить вимірювання, грубих описок або недбалостей. Вони властиві як суцільному, і несуцільному спостереженню і бувають двох видів: випадковими (ненавмисними), чи тенденційними (навмисними).

*Ненавмисна помилка реєстрації* є випадковою, якщо її причини є випадковими (неуважність, обмовка, описка, незосередженість респондента). При цьому ймовірність помилки як у бік перебільшення, і у бік применшення фактів однакова. Тому вони, як правило, не мають істотного впливу на кінцеві результати спостереження, оскільки при великій кількості спостережень випадкові помилки можуть в кінцевому підсумку взаємопогашатися і на результат не впливають. Так, наприклад, щодо використання робочого часу окремі спостерігачі систематично протягом усієї зміни відносять час обслуговування робочого місця до підготовчо-заключного часу. Ненавмисні помилки виникають у результаті нечіткого розуміння спостерігачем тієї чи іншої ознаки.

*Навмисна помилка реєстрації* виникає, якщо особи, які проводять спостереження, з будь-яких причин зацікавлені в прихованні або спотворенні дійсності. Так, існує тенденція до округлення доходів і витрат до цілих чисел (доходів у бік зменшення, витрат — у бік збільшення). Помилки виникають через несправність вимірювальних інструментів, приладів. У цьому випадку реєстратор навмисне спотворює дійсність, тобто надає неправильні відомості. Помилки навмисні носять лише систематичний характер. Їх дія призводить до зсуву результатів у бік збільшення або зменшення.

Помилки реєстрації (навмисні та ненавмисні) виникають завжди з якоїсь певної причини і виражаються у спотворенні результатів спостереження у бік перебільшення чи применшення. Ці помилки мають місце як при суцільному, так і при несуцільному спостереженні. Найчастіше визначити помилку реєстрації неможливо.

*Помилка репрезентативності* — розбіжність між значеннями досліджуваної ознаки вибіркової сукупності та генеральної сукупності. Ця помилка виникає внаслідок того, що вибіркова сукупність недостатньо повно відтворює (репрезентує) сукупність генеральну, хоча встановлення та реєстрація фактів було здійснено точно. Помилки репрезентативності властиві лише несуцільному спостереженню. Вони виникають в результаті того, що склад зібраної з нього частини одиниць сукупності недостатньо повно відображає склад всієї досліджуваної сукупності, хоча реєстрація фактів з кожної відібраної для обстеження одиниці була проведена точно. Ці помилки виявляються при порівнянні результатів несуцільного і суцільного спостережень, проведених по тому самому об'єкту спостережень.

Помилки репрезентативності складаються з помилок власне-репрезентативних та помилок реєстрації (систематичних та випадкових).

*Систематичні помилки* являють собою неточності, які виникають в процесі статистичного спостереження з якоїсь цілком певної причини (однієї або невеликої з числа) і що виражається у спотворенні величини показника у бік перебільшення або применшення. Останні помилки мають місце тільки при вибірковому спостереженні. Часто систематичні помилки характеризуються сталістю знака. Причинами можуть бути неточності чи несправності вимірювальних приладів, перебільшення віку літніми особами, округляють його до чисел 5, 10, і т. п. Наприклад, при не документованому зборі відомостей можливі округлення віку, стажу роботи, зарплати. До систематичних помилок відносяться також помилки суб'єктивних вражень, які також спрямовані в один бік: або в бік збільшення (наприклад, суб'єктивна оцінка врожаю в період дозрівання хлібів), або в бік зменшення. Систематичні помилки накопичуються, тому що вони діють в одному напрямі: залишки сировини применшуються, а обсяги виконаних робіт перебільшуються; сімейні доходи або витрати у зв'язку із забуттям применшуються. Ці помилки можуть бути наслідком свідомого, навмисного та ненавмисного спотворення істини. Причини систематичних помилок, зазвичай, може бути встановлені, отже, самі помилки — виключені з результату розрахунку чи останньому враховані можливі їх межі. Усі систематичні помилки є навмисними помилками і не погашаються у процесі статистичної зведення.

*Випадковими помилками* вважаються неточності, що виникають у процесі статистичного спостереження під час встановлення чи реєстрації фактів з суто випадкових причин (випадкові описки, незнання чи небажання опитуваних відповідати, запам'ятовування і т. п.). Ці помилки мають у кожному окремому випадку різні значення. До помилок такого виду відносяться ненавмисні помилки — як наслідок опису або недостатньо ясного розуміння реєстратором сутності реєстрованих ознак, зокрема, якщо реєстратор не знає, до якої групи належить та чи інша операція, що фіксується в процесі спостереження за роботою бригади складального цеху . Випадкові помилки при статистичному спостереженні не суттєво впливають на кінцеві результати узагальнення; у процесі статистичного зведення зібраних даних в силу дії закону великих чисел вони можуть певною мірою взаємопогашатися.

Таким чином, загальна похибка вибірки складається зі випадкової похибки (внаслідок випадкових розходжень між елементами сукупності, що включені до вибірки, і тими, що не потрапили до неї) і зі зміщення (систематичної похибки), якщо вона існує.

**Контроль матеріалів статистичного спостереження**. Перш ніж приступити до обробки статистичного матеріалу, його піддають контролю. Контроль здійснюється на всіх етапах збору та обробки статистичних даних: при заповненні формулярів статистичного спостереження, складання звітів та статистичних таблиць, розрахунку системи аналітичних показників та підбитті підсумків. Проте особливе значення має контроль на початкових стадіях спостереження, оскільки помилки, допущені цьому етапі, дуже складно буде виправити згодом.

*Контроль* представляє собою перевірку правильності оформлення документів і проведених операцій при їх обробці з метою виявлення та усунення можливих помилок. Контроль достовірності вихідного статистичного матеріалу стосується в основному: 1) повноти зібраних даних; 2) точності відповіді питання її програми; 3) якість обробки матеріалу статистичного спостереження.

*Контроль достовірності* вихідного статистичного матеріалу передбачає передусім перевірку повноти зібраних даних. Вона полягає у з'ясуванні, чи від усіх звітних (облікових) одиниць надійшли дані, чи всі показники заповнені у документах первинного звіту та формах звітності. Потім здійснюється арифметичний та логічний контроль достовірності отриманих даних статистичного спостереження.

*Повнота* статистичних даних перевіряється при прийомі статистичних бланків (формулярів) спостереження як самими реєстраторами, так і особами, відповідальними за організацію та проведення спостереження. Перевірка статистичного матеріалу здійснюється за попередньо складеними списками — підприємств, установ та організацій, робітників цеху, ділянки та інших одиниць спостереження.

Одночасно із перевіркою повноти заповнення бланків спостереження здійснюється контроль за *достовірністю*, правильністю заповнення відповідних бланків. Іноді для перевірки достовірності даних статистичного спостереження вдаються до зіставлення даних, отриманих із різних джерел. Якщо в бланку є порожнє місце, це дуже легко виявити. Значно складніше виявити неточності у відповідях на питання статистичного бланка спостереження.

Для виявлення помилок використовують дві форми контролю: логічний та лічильний. Рахунковий та логічний контроль тісно пов'язані, між ними не завжди можна провести кордон.

*Рахунковий контроль* полягає у перевірці різних арифметичних розрахунків, результати яких наведені у формулярі спостереження, зокрема, підсумків, обчислення відсотків, розрахунків середніх величин, балансових співвідношень тощо. Він зводиться до повторення арифметичних дій, виконаних під час заповнення бланка: до перевірки сум доданків, перевірки розрахунку відсотків, середніх. Головною метою рахункового контролю є перевірка лічильної узгодженості даних, вміщених у формулярах статистичного спостереження, а також правильності підрахунку підсумків. Прикладом можуть бути взаємозв'язки балансових співвідношень: наявність початку періоду *плюс* суми надходжень *мінус* суми вибуття за період повинно *дорівнювати* наявності на кінець періоду.

Наприклад, перевірка результатів різних арифметичних дій:

$$\text{Урожайність} = \frac{\text{Валовий збір}}{\text{Посівна площа}} \text{ і т. п.}$$

Для виявлення помилок, поряд з рахунковим контролем, який дозволяє виявити помилки при обчисленні, також застосовують *логічний контроль*. Логічний контроль ведеться для перевірки правильності змісту відомостей, зібраних з кожної одиниці сукупності. Він полягає також у використанні взаємозв'язків між різними показниками, тільки не лише чисто кількісних, а змістовних. Наприклад, витрати праці та обсяг виготовленої продукції цеху промислового підприємства перебуває у певному взаємозв'язку. У сільському господарстві також взаємопов'язані розміри, наприклад, посівної площі та валовий збір. При логічному контролі перевіряють, чи є між зазначеними показниками якої-небудь неув'язки. Наприклад, якщо зазначений вік 15 років, а освіта вища, то ясно, що відповідь на одне із питань неправильна.

Логічний контроль полягає у зіставленні відповіді питання формуляра спостереження і з'ясування логічної їх сумісності. Логічний контроль здійснюється різними способами:

1) зіставляються дані проведених статистичних спостережень з розробленими нормативами (витратами часу, питомими витратами матеріалів та ін.) або плановими показниками;

2) порівнюються відповіді на різні питання одного і того ж формуляра, наприклад, зіставляються в бланку перепису населення відомості про професію, сімейний стан;

3) зіставляються записи, що належать до звітного періоду, з аналогічними записами попередніх періодів;

4) зіставляються дані проведених статистичних спостережень з результатами вибіркових спостережень, що дозволяють отримати досить достовірні дані про одиниці сукупності, що спостерігаються;

5) порівнюються матеріали статистичного спостереження із середніми для сукупності.

Якщо отримані значні відхилення, отримані дані піддаються всебічної перевірці, звітній одиниці роблять додатковий запис щодо правильності представлених даних. У разі виявлення логічно несумісних відповідей намагаються подальшим зіставленням із відповідями на інші питання встановити, яка з відповідей є неправильною. Таке зіставлення дає змогу знайти неправильну відповідь.

Приклад. Маються такі дані про надходження, продаж та залишки товарів торгівельної організації за квартал (тис. грн):

Таблиця 2.2

| Номер п/п | Найменування товарних груп | Залишок товарів на початок кварталу | Надходження товарів за квартал | Реалізовано та інший документальне витрачання, що не є роздрібним продажем | Продано в роздріб та інше недокументоване витрачання (гр. 1+гр. 2 – гр. 3 – гр. 5) | Залишок товарів на кінець кварталу |
|---|---|---|---|---|---|---|
| А | Б | 1 | 2 | 3 | 4 | 5 |
| 1 | Картопля | 44 | 278 | 72 | 212 | 38 |
| 2 | Капуста | 32 | 103 | 15 | 920 | 28 |
| 3 | Цукор | 47 | 223 | 23 | 213 | 40 |
| 4 | Макаронні вироби | 77 | 218 | 18 | 209 | 68 |
| | Всього | 200 | 822 | 128 | 712 | 174 |

Потрібно перевірити правильність обчислення даних гр. 4 та підсумкових показників. Слід мати на увазі, що рух товарів у торговому підприємстві можна представити у вигляді балансу, що має вигляд наступного рівняння:

$$\begin{matrix}\text{Залишок}\\ \text{на початок}\\ \text{періоду}\end{matrix} + \begin{matrix}\text{Надходження}\\ \text{за період}\end{matrix} = \begin{matrix}\text{Витрати}\\ \text{за період}\end{matrix} + \begin{matrix}\text{Залишок}\\ \text{на кінець}\\ \text{періоду}\end{matrix}$$

При контролі можна застосувати логічну та арифметичну перевірки.

Почнемо перевірку вихідних даних із першої товарної групи — картоплі. Перевіримо взаємозв’язок між показниками:

44+278=72+212+38, тобто 322=322. Перевіряємо (гр. 4) «продано в роздріб»: 44+278 – 70 – 38=212. Відповіді пов'язані логічно та арифметично. За товарною групою — капуста — привертає увагу у гр. 4 цифра 920. Перевіряємо наявність балансової рівності: 32+105≠15+920+28. Балансової рівності між показниками немає.

Знаючи взаємозв'язок між показниками, обчислюємо дані гр. 4. Отримали 90 (32+103–15–28=92). Тепер перевіряємо, чи є балансова ув'язка: 32+103=15+92+28, тобто 135=135. Як видно, воно забезпечується. Отже, допущено помилку. Очевидно, при заповненні звіту механічно замість цифри 92 записано цифру 920, тобто поставлено зайвий нуль. Вносимо виправлення: гр. 4., ряд. 2 замість 920 ставимо 92.

Аналогічно перевіряємо балансову рівність за наступною товарною групою — цукор. Тут балансової ув'язки не отримаємо (47+223≠23+213+40). Перевіряємо розрахунок гр. 4: 47 +223–23–40 = 207. У звіті написано 213, отже, допущена арифметична помилка. Вносимо виправлення: замість 213 записуємо 207.

За макаронними виробами балансова ув'язка виходить 77+218=18+209+68, тобто 295=295. Це означає, що помилок немає. Далі перевіряємо підсумкові показники звіту, підсумовуючи дані щодо кожної групи окремо, а потім ув'язуємо в балансову рівність. Підсумки за гр. 1, 2, 3, 5 правильно підраховані. По гр. 4 з урахуванням внесених виправлень отримуємо 720 (212+92+207+209) та записуємо замість 720. Тепер підсумкові дані відповідають балансовій схемі: 200+822=128+720+174, тобто 1022=1022.

Після внесених виправлень дані можна використовувати для оперативної та аналітичної роботи.

Серед заходів, що сприяють підвищенню точності спостережень, важливе значення має належне роз'яснення відповідним посадовим особам, населенню значення, цілей та порядку проведення спостереження, а також серйозні заходи боротьби зі свідомими спотвореннями статистичних даних.

Для здійснення перевірки раніше проведеного виміру роблять *контрольне вимірювання*. У деяких випадках воно проводиться з метою визначення на підставі кількох контрольних вимірювань будь-якого показника. Результати контрольного вимірювання часто приймають як контрольну величину для оцінки даних наступних спостережень.

Поєднання логічного та рахункового контролю зазвичай призводить до з'ясування та виправлення помилок.

## ЛІТЕРАТУРА


1. *Горкавий В. К.* Статистика: Підручник. Третє вид., переробл. і доповн. Київ: Алерта, 2019. С. 37–54.
2. Статистика: підручник / *С. І. Пирожков*, *В. В. Рязанцева*, *Р. М. Моторин* та ін. Київ: Київ. нац. торг.-екон. ун-т, 2020. С. 47–86.
3. *Ткач Є. І.*, *Сторожук В. П.* Загальна теорія статистики: підручник. 3-тє вид. Київ: Центр учбової літератури, 2009. С. 21–38.


## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники, словники*


4. Глосарій до плану статистичного спостереження: Наказ Держкомстату від 29.12.2009 р. № 498. URL: http://ukrstat.gov.ua/me-tod_polog/glos.htm.
5. *Карпенко Л. М.* Статистика: навч. посіб. Одеса: ОРІДУ НАДУ, 2019. С. 12–16.
6. *Мармоза А. Т.* Теорія статистики: підручник. 2-ге вид., перероб. і доп. Київ: «Центр учбової літератури», 2013. С. 51–71.
7. *Опря А. Т.* Статистика (модульний варіант з програмованою формою контролю знань). Навч. посіб. Київ: Центр учбової літератури, 2012. С. 22–44.
8. *Сєрова І. А.*, *Аксьонова І. В.* Організація статистичних спостережень. Конспект лекцій для студентів спеціальності 8.050 110 усіх форм навчання. Харків: Вид. ХНЕУ, 2008. 237 с.
9. Статистичний словник / [*О. Г. Осауленко*, *О. О. Васєчко*, *М. В. Пугачова* та ін.]; за ред. д-ра держ. упр., проф., член-кор. НАН України О. Г. Осауленка; НТК статистичних досліджень. Київ: ДП «Інформ.-аналіт. агентство», 2012. 498 с.
10. *Шапочка М. К.*, *Маценко О. М.* Теорія статистики: навчальний посібник. Суми: Університетська книга, 2014, С. 27–40.


*Монографії та статті*


11. *Васєчко О. О.* Методологічні основи статистики підприємств. Київ: НТК стат. досліджень, 2005.

## Г Л А В А 3

# ЗВЕДЕННЯ І ГРУПУВАННЯ СТАТИСТИЧНИХ МАТЕРІАЛІВ

### 3.1. Завдання зведення та основний його зміст

Після статистичного спостереження слідує друга стадія статистичного дослідження, що складається із зведення отриманих під час спостереження матеріалів.

В результаті першої стадії статистичного дослідження — статистичного спостереження — отримують відомості про кожну одиницю сукупності. Так, відповідно до Національного положення (стандарту) бухгалтерського обліку 1 «Загальні вимоги до фінансової звітності», затвердженого наказом Міністерства фінансів України від 7 лютого 2013 р. № 73, у річній формі фінансової звітності підприємства «Баланс» надається різнобічна інформація на певну дату про ресурси, контрольовані підприємством, зобов’язання та власний капітал. Вказується вартість основних засобів, незавершених капітальних інвестицій, запасів, грошових коштів та багато інших ознак.

Щоб за науково зібраними фактами глибоко пізнати явище, зробити правильні висновки і розробити конкретні практичні положення, ці факти потрібно узагальнити, піддати зведеній обробці, теоретично осмислити. Без цього не можна зробити теоретичні та практичні висновки. Ці відомості представляють лише вихідний матеріал, який за образним виразом відомого статистика О. А. Кауфмана, становить лише груду каменю, цегли, балок і труб, з яких буде збудовано статистичну будівлю.

Завдання другої стадії статистичного спостереження у тому, щоб упорядкувати і узагальнити первинний матеріал, звести їх у групи і цій основі дати зведену характеристику всієї сукупності фактів з допомогою статистичних показників. Цей етап називається зведенням.

*Статистичне зведення* — це друга стадія статистичного дослідження. Вона полягає у систематизації, обробці та підрахунку групових та загальних підсумків, розрахунку похідних величин. Надалі на основі зведених підсумків розраховуються й інші похідні статистичні показники: середні та відносні величини та різні методи статистичного аналізу. Зведення включає угруповання даних статистичного спостереження, розробку системи показників для характеристики виділених груп чи підгруп. На стадії зведення відбувається перехід від індивідуальних значень окремих ознак явищ до характеристики їх сукупності, від характеристики окремих значень ознак, що варіюються, у окремих одиниць сукупності до характеристики всієї маси об'єктів, явищ і процесів.

Значення зведення як другої наступної після статистичного спостереження стадії статистичного дослідження дуже велике. У підсумку в процесі статистичного спостереження можливо зібраний багатий, чудовий за своєю детальності, ретельності збирання матеріал, який правдиво і всебічно охоплює те чи інше явище життя. Але, якщо будуть допущені помилки під час статистичного зведення, то на основі зібраного матеріалу не можна буде зробити об'єктивних узагальнюючих висновків.

Це означає, що головним змістом процесу зведення є систематизація, обробка та підрахунок групових та загальних підсумків, розрахунок виробничих величин, що розкривають сутність соціально-економічних явищ. В результаті зведення здійснюється перехід від даних, що характеризують окремі явища (підприємства тощо) до даних, що характеризують сукупність явищ у цілому (чисельність населення країни, підприємств галузі промисловості, групи великих, середніх та дрібних підприємств тощо).

Виділяють етапи програми проведення зведення:

1) виділення груп та підгруп;

2) визначення узагальнюючих показників (групових та загальних підсумків, похідних величин);

3) розробка макету таблиць.

На першому етапі програми проведення зведення відбувається поділ статистичної сукупності на окремі групи та підгрупи, що характеризують основні ознаки масових соціально-економічних явищ та процесів. На другому етапі здійснюють розрахунок групових, підсумкових та похідних показників статистичного зведення. Основний зміст третього етапу програми зведення становить система макетів розробних таблиць.

Виділяють кілька підвидів зведення: первинна, вторинна, централізована та ін. *Зведення первинне* — обробка та підрахунок даних безпосередньо в процесі статистичного спостереження. *Зведення вторинне* — обробка та підрахунок даних, отриманих в результаті первинного зведення. *Зведення централізоване* — спосіб виділення зведення, при якому всі первинні матеріали, отримані в результаті статистичного спостереження, зосереджуються в центральному органі, де піддаються зведенню. Зведення централізоване набуває великого значення у зв'язку з широким застосуванням електронно-обчислювальних машин. Це дозволяє комплексно обробляти первинні дані, скоротити витрати на їх обробку, прискорити процес зведення, усунути дублювання в обробці статистичної інформації. *Зведенням децентралізованим* називається комплекс операцій з узагальнення первинних даних, коли результати спостереження обробляються в одному органі, а отримані підсумки надходять в інший орган і там визначаються підсумкові показники.

По глибині та точності обробки матеріалу розрізняють зведення просте та складне. При *простому зведенні* відбувається підрахунок лише загальних підсумків із сукупності одиниць спостереження. *Складне зведення* являє собою комплекс операцій, що включають угруповання одиниць спостереження, підрахунок підсумків за кожною групою та по всьому об'єкту та представлення результатів угруповання та зведення у вигляді статистичних таблиць.

Статистична зведення в більшості випадків є складною операцією з обробки інформації, зібраної в результаті статистичного спостереження, при якому тисячі і багато мільйонів індивідуальних показників статистичної сукупності перетворюються на струнку систему статистичних викладок.

Порядок організації та проведення статистичного зведення визначається за програмою, що складається заздалегідь. У цій програмі визначається перш за все підмет і присудок. *Підмет зведення* становить групи або частини, на які розбивається явища сукупності, що вивчається. *Присудок зведення* становить показники, що характеризують кожну групу і сукупність в цілому. Групи даної сукупності можуть бути отримані за багатьма ознаками та охарактеризовані багатьма показниками.

Результати зведення можуть бути представлені у вигляді статистичних рядів розподілу.

Основний зміст програми зведення становить система розробних таблиць, яка є допоміжною статистичною таблицею, призначену для отримання докладних даних при обробці інформації, зібраної в результаті спостереження. Матеріали розробної таблиці зазвичай застосовуються отримання даних по укрупненим групам явищ досліджуваної сукупності.

Для успішного здійснення статистичного зведення складається його план. У плані зведення містяться вказівки про послідовність та терміни виконання її окремих частин, від осіб, відповідальних за її проведення, про порядок викладу результатів, а також координація установ та організацій, що займаються проведенням зведення.

## 3.2. Метод групувань

**Завдання групувань**. Отримані в результаті статистичного зведення прості підсумкові показники не задовольняють дослідника, оскільки вони дають лише надто загальне уявлення про досліджуваний об'єкт. Наприклад, обмежитися знанням чисельності населення країни не можна. Потрібно знати чисельність чоловічого та жіночого населення, зайнятих в окремих видах економічної діяльності, розміщення населення за окремими областями та районами країни. Детальна характеристика населення за віком, національністю, сімейним станом, джерелом доходу необхідна для державного регулювання національної економіки.

Наприклад, є такі дані про роботу 24 підприємств харчової промисловості (табл. 3.1). Однак прості підсумкові дані дають лише загальне уявлення про об'єкт, що вивчається, і не дозволяють досліднику зробити якісний аналіз. Наприклад, лише за підсумками та окремими показниками важко судити про характер розподілу підприємств, наприклад, за обсягом виробленої продукції, за кількістю працюючих або за вартістю основних виробничих фондів, про те, які значення показників є найбільш характерними для цієї галузі за звітний рік. Іншими словами, від статистики вимагається не тільки кількісна характеристика об'єкта, але й знання його окремих частин, груп. Порівняння окремих груп дозволяє зробити висновок про їхню відмінність, про динаміку та тенденції їх розвитку. Узагальнення даних про напрям розвитку груп дає уявлення про характер розвитку досліджуваного об'єкта в цілому.

Таблиця 3.1

**Основні показники діяльності підприємств харчової промисловості**

| Номер п/п | Середньорічна вартість основних виробничих фондів, млн грн | Середньооблікова чисельність працюючих за звітний період, осіб | Виробництво товарної продукції за звітний період, млн грн | Чистий прибуток (збиток), млн грн |
|---|---|---|---|---|
| 1 | 18,6 | 370 | 37,4 | 1,2 |
| 2 | 28,6 | 380 | 96,5 | 4,2 |
| 3 | 58,3 | 210 | 16,4 | 0,9 |
| 4 | 38,2 | 450 | 43,4 | –1,6 |
| 5 | 29,4 | 392 | 64,7 | 2,6 |
| 6 | 76,0 | 278 | 29,9 | 3,1 |
| 7 | 22,4 | 574 | 96,4 | 2,6 |
| 8 | 45,9 | 212 | 111,7 | 6,2 |
| 9 | 36,2 | 267 | 23,4 | 1,8 |
| 10 | 31,5 | 205 | 37,6 | 2,5 |
| 11 | 56,4 | 200 | 15,4 | –1,2 |
| 12 | 39,6 | 249 | 19,5 | 0,3 |
| 13 | 11,0 | 310 | 37,1 | 2,9 |
| 14 | 67,5 | 415 | 26,6 | 3,1 |
| 15 | 46,7 | 644 | 77,8 | 6,2 |
| 16 | 28,5 | 402 | 39,6 | 2,4 |
| 17 | 61,8 | 307 | 34,7 | –3,8 |
| 18 | 25,9 | 455 | 89,1 | 4,8 |
| 19 | 33,7 | 300 | 22,4 | 1,2 |
| 20 | 67,6 | 340 | 27,9 | 0,6 |
| 21 | 30,8 | 270 | 16,5 | 2,5 |
| 22 | 26,5 | 250 | 127,8 | 12,4 |
| 23 | 70,6 | 445 | 56,7 | 2,9 |
| 24 | 46,2 | 507 | 48,3 | 3,4 |
| | 997,9 | 8 432 | 1 196,8 | 61,2 |

Для виявлення типових характеристик і структури явища, що вивчається, потрібно привести в відомий порядок, тобто згрупувати. Так, наприклад, при переписі обладнання необхідно визначити групи обладнання за технічним станом, часом використання, необхідності капітального ремонту і т. п. угрупованням.

*Процес утворення груп одиниць сукупності однорідних у якомусь суттєвому відношенні, що мають однакові або близькі значення групувальної ознаки* називається *групуванням*. При цьому мається на увазі не тотожність одиниць, що входять до складу групи, а їх однотипність за якоюсь істотною ознакою.

Групування первинного матеріалу статистичного спостереження є однією з найважливіших робіт статистичного дослідження в цілому, від результатів якої значною мірою залежатиме правильність отриманих висновків. Незважаючи на зовнішню простоту та легкість технічного характеру, ця операція є найважчим етапом статистичної роботи.

При статистичному дослідженні складних суспільних явищ великі можливості в науковому та практичному аспекті закладені у використанні методу групувань з іншими методами статистичного дослідження. Насамперед метод групувань пов'язаний з методом кореляції та з показниками рядів розподілу.

Групування дозволяє правильно застосовувати середні величини, оскільки визначення середніх без розчленування даних на якісно однорідні групи може призвести до помилкових висновків.

Приклад. Середній процент виконання норм виробітку в цеху становить 108 %. Цей показник сам собою нічого не говорить про наявні резерви. Для виявлення необхідно провести групування чисельності робочих за рівнем виконання норм, продуктивності праці (табл. 3.2).

Таблиця 3.2

**Групування робітників з виконання норм виробітку**

| Виконання норм виробітку, % | Кількість робітників даної групи, % | Виконання норм виробітку, % | Кількість робітників даної групи, % |
|---|---|---|---|
| 80–90 | 6 | 120–130 | 6 |
| 90–100 | 9 | 130–140 | 5 |
| 100–110 | 57 | 140–150 | 3 |
| 110–120 | 14 | 108 | В середньому по цеху |

Метод групувань дуже часто використовується для вирішення багатьох завдань. Завдання, які вирішуються методом групування, можна звести до трьох основних:

1) вивчення соціально-економічних типів явищ;

2) вивчення структури явища та структурних зрушень, що відбуваються в ньому;

3) вивчення зв'язку та взаємозалежності між явищами.

Кожне з цих завдань вирішується відповідно шляхом побудови та аналізу *типологічних*, *структурних* та *аналітичних* (факторних) групувань.

За допомогою групувань можна вирішити такі завдання, які не можна вирішити за допомогою теоретичних міркувань, ні на підставі окремих прикладів. Велика роль групувань в контролі за ходом виконання плану, оскільки з їх за допомогою можна встановити причини невиконання плану, резерви його перевиконання. За допомогою угрупований виявляються диспропорції у розвитку народного господарства країни.

**Принципи побудови статистичних групувань та класифікацій**. Перш ніж побудувати будь-яке групування, необхідно вирішити два основні завдання: вибрати групувальну ознака (основи групування) і визначити число груп, що виділяються.

Побудова групування починається з визначення групувальних ознак або основи групування. *Основою групування* (або групувальною ознакою) є ознака, за якою одиниці сукупності розподіляють за групами, за числом груп та їх позначення (межі). Якщо, наприклад, проводиться групування підприємств за даними їх звітності, то скласти групи підприємств можна за різними ознаками: за вартістю товарної продукції, чисельністю працівників, вартістю основних виробничих засобів, потужністю електроустановок тощо. В якості основи групування необхідно відбирати, основні, найбільш суттєві ознаки.

Вибрана на підставі попереднього аналізу ознака групування є основою цього групування. Зрозуміло, вибір основи групування залежить від цілей дослідження.

Вибір групувальної ознаки, тобто ознаки, з урахуванням якої розчленовують сукупності на групи, — одне з найважливіших і важливих питань теорії групування і статистичного дослідження соціально-економічних явищ і процесів.

Залежно від ознаки, яка лежить в основі утворення груп, розрізняють групування за атрибутивною (якісною) та кількісною ознаками.

*Атрибутивними* називають ознаки, варіанти якої виражаються словесно. Прикладом побудови групування за атрибутивною ознакою може бути розбивка населення групи за соціальним становищем, статтю, професією, освітою; продукції — за сортами, видами, ступенем новизни, капітальних вкладень — за видами активів, за джерелами фінансування тощо. До атрибутивних ознак можна віднести адміністративно-територіальну та відомчу належність одиниці спостереження. Побудова групування за атрибутивною ознакою вимагає правильного визначення його меж, за допомогою яких встановлюється належність групованих явищ до певного типу. Наприклад, у статистичному щорічнику Державної служби статистики України наводяться такі угруповання: розподіл постійного населення за статтю: обидві статі, чоловіки, жінки; за складом населення: все населення, міське, сільське; розподіл сільськогосподарської продукції за групами виробників: всього, підприємства, господарства населення.

*Кількісні ознаки* виражають властивості досліджуваного явища з допомогою цифр. Прикладом групування за кількісною ознакою є групування працівників за розміром заробітної плати, стажу роботи, віком, групування підприємств за розміром товарної продукції, основних виробничих засобів тощо.

Після того, як визначено основу групування, слід вирішити питання про кількість груп, на які треба розбити досліджувану сукупність.

Число груп залежить від завдань дослідження та групувальної ознаки, одиниць сукупності, ступеня варіації ознаки.

Якщо групування будується за атрибутивною ознакою, кількість груп буде стільки, скільки є градацій, видів стану цієї ознаки. Наприклад, групування електростанцій за потужністю враховує потужності гідроелектростанцій, теплових, атомних, інших (вітрових, сонячних). Відповідно до Господарського кодексу України в редакції від 22.03.2012 року залежно від розмірів виділяють такі суб'єкти підприємництва: суб'єкти мікропідприємництва; суб'єкти малого підприємництва; суб'єкти середнього підприємництва; суб'єкти великого підприємництва.

Разом з тим групування за атрибутивною ознакою найчастіше не знімає питання про основу групувань, тому що при достатності значень атрибутивних ознак створюється надмірна роздробленість явища, що вивчається. Наприклад, при дослідженні роду занять населення можна отримати тисячі варіантів відповідей. При такій величезній кількості атрибутивних ознак виникають труднощі при їхньому віднесенні до певних груп або класів. У зв'язку з цим виникає необхідність складання класифікацій занять, де кожне заняття може бути віднесене до тієї чи іншої групи. В основу групування кладеться групувальна ознака «рід занять», яка, в свою чергу залежить від того, що під нею розуміється.

Дещо складніше справа у випадку, коли групування проводиться за кількісною ознакою. Оскільки кількісні ознаки мають цифровий вираз, віднесення їх в одну групу є співпадінням конкретного значення групувальної ознаки зі значенням цієї ознаки в окремих об'єктів.

В цьому випадку необхідно звернути особливу увагу на кількість одиниць ознаки. При невеликому обсязі сукупності не слід створювати велику кількість груп, оскільки групи будуть включені недостатнє число одиниць явища, що вивчається. В цьому випадку показники, розраховані виходячи з даних груп, не будуть представницькими і дозволять дати адекватну оцінку зв'язків між явищами і процесами.

Систематизація об'єктів за їх природними (внутрішньо притаманними) ознаками називається *класифікацією*. Класифікація є своєрідним стандартом, в якому міститься систематизований перелік об'єктів, кожному з яких надано певний код. Чіткий розділ між групуванням та класифікацією провести складно.

В українській статистиці детально розроблена класифікація за видами економічної діяльності (КВЕД), яка дає уяву про загальну структуру національної економіки на основі віднесення окремих підприємств і виробництв до тієї чи іншої групи і призначена для використання органами державного управління, фінансовими органами та органами статистики.

В основі розроблених статистиками класифікації лежать суттєві ознаки, які мало змінюються та існують тривалий час. Своєрідною класифікацією видів економічної діяльності служить номенклатура (перелік) видів економічної діяльності.

Якщо в основі групування покладена безперервна ознака, то виникає питання по визначення числові межі окремих груп, тобто питання про встановлення інтервалів групування.

## 3.3. Інтервали групувань

Якщо в основі групування покладена кількісна ознака, то крім вибору групувальної ознаки виникає питання про кількісні розміри окремих груп, питання про інтервали угруповання. Інтервал представляє собою значенням варійованої ознаки, яка лежить у певних межах. *Інтервали угруповань* — позначення груп «від» — «до», утворених групуванням за кількісною ознакою. Наприклад, заробітна плата робітників «від 10 000 до 11 000 грн», «від 11 000 до 12 000 грн» тощо.

*Межі інтервалу* — числа, що позначають найменше і найбільше значення ознаки у виділеному інтервалі при групуваннях, що називається відповідно верхньою і нижньою межею інтервалу. Наприклад, групи підприємств за чисельністю працівників можуть позначатися:

до 100
от 100 до 200
от 200 до 300
от 300 до 600 і т. п.

При побудові інтервальних варіаційних рядів у кожен інтервал включаються варіанти, числові значення яких більші за нижню межу і менше або рівні верхньої межі.

*Розмір інтервалу* (*інтервальна різниця*) ($k$) — це різниця між верхньою $x_{i\max}$ та нижньою $x_{i\min}$ межами інтервалу:

$$k_i = x_{i\max} - x_{i\min}. \tag{3.1}$$

Так, величина другого інтервалу становитиме 100 (200 – 100), четвертого — 300 (600 – 300). Знання величини інтервалу дозволяє визначити межі всіх інтервалів низки розподілів. Нижню межу першого інтервалу доцільно приймати рівній мінімального значення ознаки (іноді за початкову межу береться число, найближче до мінімального значення).

Часто буває, що в групуваннях фактичне значення інтервалу невідоме, і тому його замінюють наближеним, тобто середнім (центральним) значенням меж інтервалу. *Центром інтервалу*, або *серединним значенням інтервалу* називається напівсума нижньої та верхньої меж інтервалу кожної групи, утвореної групуванням. Розрахунок центру інтервалу здійснюють за такою формулою:

$$x_{\text{центр}} = \frac{x_{\min} + x_{\max}}{2}, \tag{3.2}$$

де $x_{\min}$ — нижня межа інтервалу;

$x_{\max}$ — верхня межа інтервалу.

Враховуючи, що $x_{\max} - x_{\min} = \Delta$, то отримуємо іншу формулу центру інтервалу[1]:

---

[1] В цьому випадку можна поступити так: мається мінімальне $x_{\min}$ і максимальне $x_{\max}$ значення інтервалу: $x_{\text{центр}} = \frac{x_{\min} + x_{\max}}{2} = \frac{x_{\min} + x_{\max} + x_{\min} - x_{\min}}{2} = \frac{2x_{\min} + x_{\max} - x_{\min}}{2} = $ $= x_{\min} + \frac{x_{\max} - x_{\min}}{2}$; приймаючи $x_{\max} - x_{\min} = \Delta$, отримаємо $x_{\text{центр}} = x_{\min} + \frac{\Delta}{2}$.

$$x_{центр} = x_{min} + \frac{\Delta}{2}. \quad (3.3)$$

Серединне значення другого інтервалу становитиме: $x_{центр} = \frac{100+200}{2} =$ =150 осіб. Використовуючи іншу формулу, отримаємо той самий результат: $x_{центр} = 100 + \frac{100}{2} = 150$ осіб. Аналогічно розраховуються серединні значення та інших інтервалів. Якщо є інтервали відкриті, то для обчислення серединного значення інтервалу для них беруться рівними: для першого — інтервалу другому, для останнього — інтервалу попередньому (передостанньому). Якщо, наприклад, групи були позначені: «до 100», «100–2002, «200–300», «понад 300», то серединне значення першого інтервалу складе 50 $\left(\frac{100-0}{2}\right)$, останнього 350 $\left(300+\frac{300-200}{2}\right)$.

Інтервали можуть бути різними за величиною. В залежності від величини інтервального проміжку між групами розрізняють інтервали рівні та нерівні. Інтервальні варіаційні ряди *з рівними інтервалами* — такі, в яких усі інтервали мають одну й ту саму величину. Рівні інтервали застосовуються у тих випадках, коли зміна кількісної ознаки всередині сукупності відбувається більш менш рівно мірно в порівняно невеликих межах. Цей вид інтервалів є зручним і найчастіше використовується в економічній статистиці.

Варіаційний ряд із рівними інтервалами наведено в табл. 3.3. Якщо ступінь коливання групувальної ознаки велика і його значення варіюється нерівномірно, необхідно використовувати групування *з нерівними інтервалами* (нерівні інтервали застосовуються в основному як прогресивно зростаючи). Нерівні інтервали застосовуються, як правило, у разі виділення якісно різнорідних явищ або у випадку, коли групувальна ознака змінюється нерівномірно.

Таблиця 3.3

**Розподіл робітників за рівнем виконання норм виробітку**

| Групи робітників за ступенем виконання норм виробітку, % | Число робітників |
|---|---|
| 90–100 | 3 |
| 100–110 | 8 |
| 110–120 | 16 |
| 120–130 | 12 |
| 130–140 | 11 |
| Всього | 50 |

В залежності від характеру зміни від нижчих груп до вищих, нерівні інтервали можуть бути прогресивно зростаючими або прогресивно спадаючими. На практиці найбільшого поширення отримали інтервали послідовно зростаючи. Приклад ряду з нерівними зростаючими інтервалами наведено в табл. 3.4.

Таблиця 3.4

**Групування підприємств за розмірами площі, з якої зібрано урожай зернових та зернобобових культур у 2020 р.**

| Групи підприємств за розмірами площі, га | Кількість підприємств | | Валовий збір | |
|---|---|---|---|---|
| | одиниць | у % до загальної кількості | тис. т | у % до загального обсягу виробництва |
| Підприємства з них з площею, га | 32 513 | 100,0 | 51 718,0 | 100,0 |
| до 100 | 19 026 | 58,5 | 1 934,0 | 3,7 |
| 100–200 | 3 559 | 10,9 | 1 899,5 | 3,7 |
| 200–500 | 4 213 | 13,0 | 5 410,0 | 10,5 |
| 500–1 000 | 2 765 | 8,5 | 8 300,4 | 16,1 |
| 1 000–2 000 | 1 880 | 5,8 | 11 694,6 | 22,6 |
| 2 000–3 000 | 566 | 1,7 | 6 694,1 | 12,9 |
| більше 3 000 | 504 | 1,6 | 15 785,4 | 30,5 |

Застосування прогресивно-зростаючих інтервалів обумовлено тим, що з більшості економічних явищ у нижчих групах мала різниця у показниках має значення, тоді як у інших групах ця різниця не суттєва. Так, при виділенні груп підприємств за кількістю робітників для нижніх груп велике значення матиме різниця в 20 і 50 працівників, для вищих груп ця різниця не матиме суттєвого значення, оскільки тут сильно зростає чисельність працівників підприємства. У практиці проведення статистичного спостереження нерівні інтервали застосовуються частіше, ніж рівні. Це пов'язано, насамперед, із необхідністю виділення груп, що мають відмінні якісні характеристики.

Величина нерівних інтервалів ($h_{i+1}$), що зростають в арифметичній прогресії, визначається так:

$$h_{i+1} = h_i + a, \tag{3.4}$$

а в геометричній прогресії:

$$h_{i+1} = h_i \cdot q, \tag{3.5}$$

де $a$ — константа, що має для прогресивно-зростаючих інтервалів знак «+», для прогресивно-спадаючих інтервалів знак «–»;

$q$ — константа (для прогресивно-спадаючих інтервалів >1; в іншому випадку — $q$<1.

Серед прогресивно зростаючих інтервалів особливе місце займають кратні інтервали з різними коефіцієнтами зростання. Якщо коефіцієнт зростання дорівнює 2, кожен наступний інтервал більше попереднього вдвічі. Наприклад, якщо кількість працівників підприємства варіюється від 20 до 1280, то шкалу інтервалів за кількістю робітників можна так:

20–40
40–80
80–160
160–320
320–640
640–1280

Це закриті нерівні кратні інтервали з коефіцієнтом зростання, рівним 2: перший інтервал тут дорівнює 20 (40–20), другий — 40 (80–40), третій 80 (160–80), четвертий 160, п’ятий — 320, шостий — 640 .

При утворенні інтервалів необхідне точне позначення кількісних меж групи. Якщо групувальна ознака є дискретною, межі між групами можуть бути чітко окресленими. Наприклад, угруповання працівників підприємств за стажем роботи (років): до 2 років; 2–4 роки; 5–9 років; 10–14 років; 15–19 років; 20 і більше років.

При групуванні одиниць сукупності за безперервно варіюючою ознакою слід приділяти особливу увагу питанню віднесення окремих одиниць сукупності, значення яких є граничними до тієї чи іншої групи. Часто межі інтервалів позначаються так, що те саме значення ознаки служить верхньою і нижньою межами двох суміжних інтервалів, тобто верхня межа попереднього інтервалу служить нижньою межею наступного інтервалу. Наприклад, групи робочих підприємства за стажем роботи можуть бути позначені:

до 1 року
от 1 до 3
от 3 до 5
от 5 до 10
от 10 до 15 и т. д.

При таких інтервалах виникає питання, в яку групу відносити одиниці спостереження, значення яких збігаються з межами інтервалів; наприклад, куди відносити робітників зі стажем роботи 5 та 10 років в інтервалах: від 3 до 5 років; від 5 до 10 років, від 10 до 15 років? Це питання, як правило, вирішується так: нижня межа формується за принципом «включно», а верхня — за принципом «виключено», тобто в кожен інтервал включаються варіанти, числові значення яких більші від нижньої межі і менші або рівні верхньої межі. Так, робітники зі стажем роботи 5 років входять до 3-ї групи (від 3 до 5 років), а з терміном 10 років — до 4-ї групи (від 5 до 10 років).

Нерідко для крайніх інтервалів вказується лише одна з двох кордонів: для першого інтервалу — верхня, для останнього — нижня, як у першому з наведених прикладів. Наприклад, групуючи робітників за місячною заробітною платою з рівними інтервалами 2 000 грн виділимо такі групи:

До 10 000 грн
10 001–12 000 грн
12 001–14 000 грн
14 001–16 000 грн
16 001–18 000 грн
18 001 грн и вище

При цьому копійки до уваги не беруться. 10 001 грн є нижньою межею другого інтервалу, 12 000 грн — його верхньою межею, 12 001 грн є нижньою межею третього інтервалу, 14 000 Ї його верхньою межею і т. д. Перший та останній інтервали цього групування є відкритими. При знаходженні серединного значення (відкритого) інтервалу з серединного значення другого інтервалу віднімають величину інтервалу. А серединне значення другого інтервалу визначається як напівсума верхніх значень першого інтервалу та другого, тобто (10 000 + 12 000): 2 = 11 000 грн.

Серединне значення нижнього інтервалу визначається 11 000 – 2 000 = 9 000 грн. Серединне значення третього інтервалу утворюється як напівсума верхніх значень другого та третього інтервалів і т. д. Серединне значення верхнього відкритого інтервалу визначається як сума серединного значення передостаннього інтервалу та величини інтервалу.

Недолік нерівних інтервалів полягає в тому, що одиниці спостереження, які мають різний рівень значення ознак, потрапляють в ту ж саму групу. Уникнути це можна шляхом застосування спеціалізованих інтервалів, що відображають економічний зміст груп. Ці інтервали використовують у тому випадку, коли досліджувана сукупність виражена за допомогою кількісних ознак, але різноманітна за групами або типами ознак.

Питання про те, використовувати рівний або нерівний інтервал, залежить і від кількості одиниць, включених до цієї групи, або від ступеня заповнення даного інтервалу. Ступінь заповнення даного інтервалу служить одним з головних показників, на підставі якого можна зробити висновок про доцільність даного інтервалу груп.

За побудовою інтервали бувають закритими (замкнутими) та відкритими. *Закритими* (*замкнутими*) називаються інтервали, у яких позначені конкретними чисельними значеннями обидві межі інтервалів, *відкритими* — такі, у яких зазначено лише одну межу — верхню у першого, нижню — у останнього інтервалу групувань. Як видно з табл. 3.4, ряд має відкриті інтервали у першій «До 100» та останній «Понад 3 000» групах. У табл. 3.3 наведено закриті інтервали.

Питання про число груп при групуванні за кількісним ознакою залежить від мінливості ознаки і числа спостережень. Чим сильніша мінливість значення ознаки, тим більше груп має бути утворено. Однак це загальне правило необхідно застосовувати з певними обмеженнями, керуючись логічною логікою та здоровим глуздом. Виділення занадто великої кількості груп ускладнює проведення аналізу.

Визначення числа груп вирішується на основі виділення однорідних груп, близьких за значенням кількісної ознаки одиниць сукупності. Для визначення ступеня коливання ознаки використовують коефіцієнт варіації. Проте знайомство із цим показником відбудеться дещо пізніше. Тут ми поки що обмежимося лише загальними міркуваннями.

Коли попередньо вирішено питання про кількість рівних інтервалів, можна визначити їх величину. Розрахунок величини інтервалу ($i$) при рівних інтервалах провадиться за формулою:

$$i = \frac{x_{\max} - x_{\min}}{n}, \qquad (3.6)$$

де $x_{\max}$ — максимальне значення досліджуваної ознаки у сукупності;

$x_{\min}$ — мінімальне значення ознаки у сукупності;

$n$ — кількість груп, які намічено виділити під час проведення групувань.

При обчисленні рівних інтервалів часто практично отримуються дробові числа. Однак у багатьох випадках не потрібно встановлювати величину інтервалу дробовим числом. Для приведення дробових чисел у цілі застосовується кілька способів.

Величина рівних інтервалів, вираженого цілим числом, у цьому випадку може бути встановлена за формулою:

$$i = \frac{x_{max} - x_{min} + 1}{n}, \qquad (3.7)$$

де $i$ — величина інтервалу;

$x_{max}$ і $x_{min}$ — максимальне і мінімальне значення ознаки;

$n$ — кількість намічених інтервалів.

Тож, якщо максимальний вік робочого заводу становить 59 року, а мінімальний 20 років, намічено утворити п’ять рівних інтервалів, то величина рівного інтервалу становитиме: $i = \frac{59 - 20 + 1}{5}$=8. Цей результат можна отримати трохи інакше: з максимального значення ознаки (59) відняти не мінімальне його значення (20), а суміжне з ним попереднє значення ознаки (19); у разі величина інтервалу також складе: $i = \frac{59 - 19}{5}$=8.

Якщо при розрахунку величини рівного інтервалу вийде дробове число, рекомендується його округлювати до цілого, відповідно розширивши межі інтервалу, що охоплюється розмах коливання значень ознаки.

Для розрахунку оптимальної величини інтервалу, тобто такої величини, за якої варіаційний ряд не буде дуже громіздким і в ньому не зникнуть особливості явища, рекомендується наближена формула Стерджеса[1]:

$$i = \frac{x_{max} - x_{min}}{1 + 3{,}322 \lg n}, \qquad (3.8)$$

де $x_{max}$ — максимальне значення досліджуваної ознаки;

$x_{min}$ — його мінімальне значення;

$n$ — кількість одиниць сукупності.

Зазначений вираз у знаменнику майже завжди виявляється дробовою величиною, яку округляють до найближчого цілого числа (кількість груп не може бути дробовою величиною).

Ця формула може бути орієнтуванням для визначення числа груп

---

[1] Розглянутий спосіб наближеного визначення інтервалу групування був запропонований в 1926 р. американським ученим Г. А. Стерджесом. Такий прийом рекомендується, наприклад, у книзі Ф. Мілса.

у випадку, коли застосовуються рівні інтервали в групах. Недолік цієї формули полягає в тому, що її застосування дає хороші результати, якщо сукупність складається із великої кількості одиниць і розподіл одиниць за ознакою, покладеною в основу групування, близько до нормального.

Так, якщо в сукупності 100 одиниць найбільший варіант дорівнює 60, а найменший — 0, то

$$i = \frac{60-0}{1+3{,}322\lg 100} = \frac{60-0}{1+3{,}322\cdot 2} = \frac{60}{7{,}644} \approx 8.$$

Отже, у даному випадку оптимальною величиною інтервалу може служити величина 8.

Прирівнявши знаменники наведених формул, отримаємо іншу формулу Стерджеса для визначення числа груп:

$$k \approx 1 + 3{,}322\lg n, \tag{3.9}$$

де $k$ — число груп;

$n$ — число одиниць сукупності.

Ця формула використовується для визначення числа груп у тому випадку, якщо розподіл одиниць сукупності за даною ознакою наближається до нормальної та застосовуються рівні інтервали угруповань.

Використаємо вихідні дані попередньої задачі та визначимо кількість груп:

$$k \approx 1 + 3{,}322\lg 100 = 1 + 6{,}644 = 7{,}644.$$

Округлюючи, отримаємо число груп, що дорівнює 8. Розмір інтервалу складе 7,5 $\left(\frac{60-0}{8}\right)$.

На підставі формула Стерджеса можна скласти таку номограму:

| $N$ | 15–24 | 25–44 | 45–89 | 90–179 | 180–359 | 360–719 | 720–1439 |
|---|---|---|---|---|---|---|---|
| $n$ | 5 | 6 | 7 | 8 | 9 | 10 | 11 |

Величина інтервалу значною мірою залежить від точності даних статистичного спостереження: якщо вихідні значення ознаки представлені цілими числами, то розрахована величина інтервалу округляється до найближчого числа, якщо значення представлені з точністю до 0,1, то величина інтервалу округляється до цілих з десятими і т д.

Число інтервалів не повинно бути надмірно великим, тому що в кожному інтервалі виявиться замало одиниць спостережень для того, щоб закономірність чітко проявилася, і воно не повинно бути надмірно малим, щоб не втратити істотних подробиць розподілу.

## 3.4. Типологічні групування

**Сутність та значення типологічних групувань**. У процесі аналізу суспільних явищ необхідно виділити соціально-економічні типи, оскільки у зародженні, розвитку, боротьбі та відмиранні різних соціально-економічних типів заключна суть історичного процесу розвитку суспільства.

До *типологічних групувань* відносять групування, за допомогою яких у досліджуваній сукупності явищ виділяються одноякісні в істотному відношенні групи, перш за все класи та соціально-економічні типи. Вони характеризують якісні особливості та різницю між типами явищ. В результаті утворюються однорідні групи. Якщо підставою такого угруповання є ознака, що безперервно змінюється, то його граничні значення, що відокремлюють одну групу від іншої, служать точками переходу від однієї якості в іншу (кількості в якість).

*Завдання типологічного групування* полягає у виділенні типів, які становили якісно-однорідні групи. Нижче дається характеристика найважливіших типів у статистиці промисловості:

1) *групування за формами власності*, відповідно до якого всі підприємства поділяються на державні, колективні та приватні. Групування за формами власності має винятково важливе значення. Вона показує питому вагу державної, колективної та приватної власності у виробництві продукції, розподілі засобів виробництва, чисельності зайнятих осіб у виробництві тощо;

2) *групування за ознакою економічного призначення продукції*, згідно з якою підприємства діляться на 2 групи: групу А (1-й під-поділ) — виробництво засобів виробництва та групу Б (2-й під-поділ) — виробництво предметів споживання. Групування підприємств за ознакою економічного призначення продукції дозволяє дослідити процес розширеного відтворення;

3) *групування підприємств за їх розмірами* служить насамперед для дослідження процесів концентрації за окремими видами економі-

чної діяльності і здійснюється за низкою ознак: чисельності зайнятих працівників, вартості активів та обсягу доходу. Згідно з цими критеріями всі підприємства поділяються на великі, середні та малі.

Прикладами типологічного групування є групи населення за родом занять, категорії персоналу підприємства. Наприклад, групування населення за національністю, рівнем освіти, підприємств за формами власності, розмірами, продукцією за економічним призначенням.

Типологічні угруповання мають надзвичайно важливе значення для дослідження багатьох суспільних явищ. Одне із центральних місць займає типові групування за видами власності. Для спостереження за їх розвитком, важливі типологічні групування товарообігу, промислової та сільськогосподарської продукції тощо за формами власності. Великий інтерес представляє типологічне групування населення за класами та суспільними групами. Значну цінність становить класифікація працівників за рівнем освіти, статтю.

У нашій країні депозитні корпорації (банки) можуть видавати кредити у різних видах валюти. У зв'язку з цим значний науковий інтерес становить групування кредитів, виданих депозитними корпораціями, у розрізі валют (табл. 3.5).

Аналіз виданих кредитів за видами валют дозволяє розкрити важливі закономірності розвитку кредитного ринку країни.

Таблиця 3.5

**Групування кредитів, виданих депозитними корпораціями (банками) резидентам, у розрізі валют**
(на кінець року, млрд грн)

| Роки | В національній валюті | В іноземній валюті |
|---|---|---|
| 2010 | 395,5 | 337,3 |
| 2017 | 570,6 | 446,0 |
| 2018 | 614,0 | 459,1 |
| 2019 | 613,7 | 358,2 |
| 2020 | 600,5 | 347,9 |

Не менш важливого значення має визначення оцінка динаміки кредитів, виданих банками у межах окремих секторів економіки (інституційних одиниць). Так, якщо у 2008 р. питома вага домашніх господарств у загальному обсязі виданих кредитів становив 38,2 %, тоді як у 2018 р. він становив 18,7 %, а у 2020 р. — 20,2 %.

Однією з найважливіших характеристик підприємства є його розміри. За цією ознакою підприємства ділять на малі, середні та великі (великі). У зв'язку з цим, істотно, наприклад, виділення основних типів підприємств за їх ролі в обсязі виробленої продукції (табл. 3.6). В основі розподілу підприємств за розмірами лежить три класифікаційні ознаки: вартість активів, обсяг отриманого доходу та чисельність найманих працівників. Визначення групи є важливою умовою правильного ведення обліку та звітності. Розміри підприємств тісно пов'язані з галузевою належністю.

Дослідження галузевих особливостей дозволяє глибше досліджувати процеси концентрації та централізації капіталу, визначити динаміку та основні тенденції їх змін.

Таблиця 3.6

**Обсяг виробленої продукції (товарів, послуг) суб'єктами великого, середнього та малого підприємництва в Україні, %**

| Рік | Суб'єкти великого підприємництва | Суб'єкти середнього підприємництва | Суб'єкти малого підприємництва |
|---|---|---|---|
| 2015 | 41,7 | 36,7 | 21,6 |
| 2016 | 37,5 | 38,5 | 24,0 |
| 2017 | 38,3 | 37,0 | 24,7 |
| 2018 | 37,9 | 36,2 | 25,9 |
| 2019 | 35,9 | 35,9 | 28,2 |

Саме виділення типів не може розглядатися як незмінне, раз і назавжди задане. По-перше, оскільки всім суспільним явищам притаманний розвиток, типи також розвиваються у часі та у просторі (за внутрішнім змістом). По-друге, виділення однорідних типових груп здійснюється відповідно до поставленого завдання дослідження і в залежності від конкретних умов. Тому з однієї й тієї ж безлічі явищ можуть бути виділені різні типи. Наприклад, із сукупності промислових підприємств можуть бути виділені такі типи: за призначенням готової продукції — підприємства, що виробляють засоби виробництва та підприємства, що виробляють предмети споживання; за масштабом виробництва — підприємства масового, серійного та індивідуального виробництва; за розміром — малі, середні та великі підприємства.

Техніка розподілу одиниць статистичного спостереження за окремими типовими групами — справа дуже складна.

**Вибір групувального ознаки та встановлення інтервалу**. Основна проблема складання типологічного групування — вибір основи, або групувальної ознаки, на підставі якої явища будуть згруповані в різні групи. Виділити типове можна за будь-яким ознакою. За основу групування повинні бути взяті найважливіші ознаки, які безпосередньо характеризують сутність явищ. Вирішення питання про основу групування має здійснюватися на основі аналізу сутності досліджуваного явища. При цьому не можна використовувати ту саму групувальну ознаку у всіх випадках. Його вибір значною мірою залежить від цілей аналізу, умов і місця проведення. Групування мають бути обґрунтовані економічно.

В основі типологічного групування лежить якісний ознака — ознака, значення якого виражаються у вигляді понять, найменувань. Наприклад, види підприємств залежно від розмірів (малі, середні, великі), стать працівників (чоловіки, жінки), продукція підприємства (засоби виробництва та предмети споживання), освіта працівників (загальноосвітня, середня, вища) тощо.

Під час проведення типологічного групування важливо правильно встановити інтервал групування, відокремити один клас чи тип від іншого. Питання про інтервал типологічного групування вирішується на підставі визначення таких кількісних меж, які виділяють нову якість.

Саме виділення типів не завжди може обійтися без деякої умовності внаслідок безлічі форм переходу одних типів в інші, безлічі ознак, що пов'язують один з одним. Тому при виділенні типів необхідно брати, як відомо, не окремі ізольовані ознаки, а сукупність ознак, що характеризують багато сторін досліджуваної сукупності.

Число груп у типологічному групуванні залежить від числа дійсно наявних соціально-економічних типів. Візьмемо, наприклад, умови, що визначають розмір підприємства. Чим визначається цей тип? Якісна категорія — тип підприємства може буті чітко відмежована певною кількісною ознакою, тобто ознакою, має числове вираження. Таких ознак виділення типів підприємств досить багато. Це і чисельність працівників, і вартість виробленої продукції, та балансова вартість активів. Залежно чисельних значень цих показників підприємства групують такі типи: малі, середні і великі підприємства. Прикладом типологічного угруповання можуть бути такі дані (табл. 3.7).

Таблиця 3.7

**Основні показники діяльності підприємств за їх розмірами в Україні у 2020 р.**

| Групи підприємств за розміром[1] | Кількість підприємств, одиниць | Кількість зайнятих працівників, тис. осіб | Обсяг реалізованої продукції (робіт, послуг), млрд грн |
|---|---|---|---|
| Великі підприємства | 512 | 1 574,6 | 3 626,4 |
| Середні підприємства | 17 602 | 3 088,4 | 4 359,4 |
| Малі підприємства | 355 708 | 1 703,1 | 2 064,1 |
| Всього | 373 822 | 6 366,1 | 10 049,9 |

[1] Відповідно до Господарського кодексу України у редакції від 22.03.2012 року. URL. https://zakon.rada.gov.ua/laws/show/436-15.

Як свідчать дані табл. 3.7, великі підприємства відіграють провідну роль економіки країни.

Типологічні групування у статистиці застосовуються всюди, де потрібно охарактеризувати якісні особливості окремих груп. До типологічного угруповання в цьому значенні відносять групування підприємств за групами (підприємства, що виробляють засоби виробництва (група А) і предмети споживання (група Б)), роздрібного товарообігу за видами (продовольчі та непродовольчі товари), працівників підприємства за статтю, рівню освіти, місцевості проживання населення та інших. До типологічним групувань відносяться і групування, що визначають нове і передове, виділяють типи підприємств з технічної оснащеності, з економічної ефективності та інших. З розвитком економіки країни значення типологічного групування зростає дедалі більше.

Таким чином, типологічне групування є досить потужним статистичним знаряддям якісно-кількісного аналізу суспільних явищ. Типологічні групування доповнюються визначенням структури окремих соціально-економічних типів структурними зрушеннями, що відбуваються в них, тобто структурними групуваннями.

## 3.5. Структурні групування

Виділені в результаті типологічного угруповання окремі типи явища потребують вивчення їх складу, так як тип взагалі може включати велику кількість одиниць. Виникає необхідність розчленувати типи, статистичні сукупності на групи, охарактеризувати їхню будову, структуру. З цією метою застосовуються структурні групування.

*Структурним групуванням* називається групування, що виявляє склад (будову) однорідної у якісному відношенні статистичної сукупності за певними ознаками. Застосування структурних угруповань дозволяє вирішити друге завдання групувань — виявити внутрішню будову явищ. З допомогою структурного групування насамперед досліджується внутрішня структура типів, статистичної сукупності. Вони дають інформацію про те, з яких частин складається множина явищ, якої будова типів явищ і якими показниками характеризуються окремі частини.

Таким чином, структурні групування розглядаються як етап подальшого, глибшого вивчення однорідних груп з метою з’ясування внутрішньої будови, структури (та структурних зрушень) у сукупності однотипних одиниць. Найчастіше структурне групування виконується після та на основі типологічного.

Велике значення структурних групувань у сфері вивчення концентрації, спеціалізації та територіального розміщення промислових та сільськогосподарських, торгових, будівельних та інших підприємств та організацій.

У цьому виникає питання про основу групувань. Наприклад, для дослідження процесів концентрації виробництва у промисловості можна в якості групування взяти вартість товарної продукції, чисельність працівників, вартість капіталу, основних виробничих засобів, потужність двигунів та ін. Кожна групувальна ознака по-своєму відображає процес концентрації промислового виробництва. Так, для фондомістких галузей промисловості концентрацію найкраще відобразити за допомогою групувань за вартістю основних виробничих засобів. В інших галузях групування за вартістю основних виробничих фондів слід доповнювати групуванням за іншими ознаками.

Структурні угруповання знаходять широке застосування й у аналізі виконання плану підприємства. Зіставлення фактичних та планових випусків за певний період (годину, зміну, день, декаду, місяць та ін.) дозволяє визначити ритмічність виробництва. Групування робочих підприємства за рівнем виконання планових завдань дозволяє розкрити резерви зростання продуктивність праці.

Структурні групування дозволяють розглянути склад сукупності за економічними та адміністративними районами, галузями народного господарства, географічними зонами.

У структурному групуванні число груп має бути мінімальним, але достатнім, щоб дати уявлення про досліджувану структуру і структурні зрушення.

Прикладом структурного групування може бути розподіл зайнятого населення за видами економічної діяльності (табл. 3.8). Дані таблиці дозволяють визначити основні сфери застосування праці, оцінити структурні зміни у зайнятості населення за видами економічної діяльності.

Складання даних структурного групування за часом дає уявлення про *структурні зрушення*, що відбуваються  при аналізі сукупності одиниць явищ. Призначення цього виду групувань зводиться до того щоб охарактеризувати структуру в динаміці, тобто відображати всі зміни, що відбуваються  в  початковій  структурі  статистичної сукуп-

Таблиця 3.8

**Зайнятість населення за видами економічної діяльності, %**

| | 2015 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|---|
| Всього | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 |
| Сільське, лісове та рибне господарство | 17,5 | 17,7 | 18,0 | 18,2 | 17,1 |
| Промисловість | 15,7 | 15,1 | 14,8 | 14,8 | 14,8 |
| Будівництво | 3,9 | 4,0 | 4,1 | 4,2 | 4,2 |
| Оптова та роздрібна торгівля; ремонт автотранспортних засобів і мотоциклів | 21,4 | 21,8 | 22,3 | 22,9 | 22,9 |
| Транспорт, складське господарств, поштова та кур'єрська діяльність | 6,1 | 6,1 | 6,1 | 6,0 | 6,1 |
| Фінансова та страхова діяльність | 1,5 | 1,3 | 1,3 | 1,3 | 1,3 |

*Джерело*: Державна служба статистики України. URL. http://www.ukrstat.gov.ua/

ності.

Наступне групування показує зміну у складі населення України на 1 січня 2000 р. та на 1 січня 2020 р. (табл. 3.9).

Групування показує глибокі зміни у складі населення, що відбулися за 20 років.

Найважливішими структурними групуваннями промислових підприємств є групування підприємств за розмірами продукції, чисельністю робітників, за вартістю основних виробничих засобів, за потужністю силового апарату, ступенем виконання планового завдання та інших ознак.

Таблиця 3.9

**Зміна чисельності міського та сільського населення України за 2000–2020 роки.**

| | Млн осіб | | В процентах до всього населення | |
|---|---|---|---|---|
| | 2000 р. | 2020 р. | 2000 р. | 2020 р. |
| Все населення | 49,4 | 41,9 | 100 | 100 |
| В тому числу: | | | | |
| міське | 33,3 | 29,1 | 67,4 | 69,5 |
| сільське | 16,1 | 12,8 | 32,6 | 30,5 |

Всередині підприємства здійснюється групування робітників за величиною місячної заробітної плати, стажем безперервної роботи, віком, освітою, ступенем виконання планового завдання та іншим ознакам.

## 3.6. Аналітичні групування

*Факторні групування* дають можливість виявити взаємозв'язок між окремими ознаками досліджуваного соціально-економічного явища. За допомогою факторних групувань встановлюються та вивчаються причинно-наслідкові зв'язки між ознаками однорідних явищ, виявляються фактори розвитку сукупності. Такі групування у літературі прийнято називати *аналітичними*[1]. Використовуючи аналітичні групування, визначають факторні та результативні ознаки досліджуваних явищ.

У статистиці ознаки, що впливають на інші, пов'язані з ними ознаки, називають *факторними*, а ознаки, що змінюються під впливом факторних, — результативні. Взаємозв'язок проявляється у систематичній зміні значень результативної ознаки у зв'язку зі зміною факторної ознаки. Значимість роль цих ознак може змінюватися. В одному зв'язку ознака виступає як результативна, в іншому — як факторна. Отже, групувальна ознака є зазвичай факторною, а ознака, що характеризує групування — результативною. Факторні групування ґрунтуються на вивченні того, що в масових суспільних явищах зі зміною однієї або кількох факторних ознак змінюється і результативна ознака.

Метод аналітичних групувань дозволяє як встановити зв'язок між ознаками соціально-економічного явища, а й виявити, наприклад, фактори, що впливають цей зв'язок.

Щоб досліджувати взаємозв'язок між відібраними ознаками за допомогою аналітичних угруповань, необхідно провести групування за факторною ознакою і для кожної виділеної групи розрахувати середнє значення результативної ознаки: якщо вона кількісна, або відносні величини, якщо вона — ознака якісна. Взаємозв'язок проявляється у систематичній зміні значень результативної ознаки у зв'язку із зміною ознаки факторної. Варіація середнього значення результативної ознаки від групи до групи під впливом групувальної ознаки буде вказувати на наявність або відсутність взаємозв'язку.

---

[1] Але ця назва не зовсім коректна, вона хіба що передбачає, що типологічні та структурні групування не несуть аналітичних, тобто пізнавальних функцій, що зовсім не так. Аналітична функція властива і типологічним угрупованням, і структурним.

Якщо кількість значень ознак ряду динаміки невелика, то зв'язок між ними може бути встановлений шляхом безпосереднього огляду рядів цих значень, які називаються *паралельними рядами*. В якості прикладу такого явища представимо паралельні ряди показників по цеху металургійного заводу (табл. 3.10).

Таблиця 3.10

**Залежність між вмістом сірки в чавуні та браком у відливах**

| Номер спостереження | Вміст сірки у чавуні, ‰ ($x$) | Кількість бракованих відливів на 100 ($y$) |
|---|---|---|
| 1 | 20 | 4 |
| 2 | 23 | 5 |
| 3 | 28 | 6 |
| 4 | 30 | 5 |
| 5 | 34 | 8 |
| 6 | 38 | 7 |
| 7 | 42 | 6 |

Зіставляючи два ряди таблиці, можна дійти до висновку, що чим більше вміст сірки в чавуні, то тим більша кількість бракованих виливків. Отже, зниження вмісту сірки є важливим фактором зменшення браку продукції (відливків).

Методологію факторних групувань розглянемо на прикладі встановлення зв'язку між стажем роботи та кількістю випущеної продукції токарями механічного цеху машинобудівного заводу. Перш за все потрібно обґрунтувати, чи існує зв'язок між стажем роботи та виробленням продукції робітниками заводу. Теоретичний аналіз приводить нас до висновку про існування прямого зв'язку між цими показниками, оскільки зі збільшенням стажу роботи підвищується кількість виробів, що виготовляються одним робітником.

Для встановлення фактичного існування зв'язку наведемо конкретні дані про значення ознак окремих робітників. З метою скорочення розміру таблиці, обмежимося даними про 10 робітників (табл. 3.11). Аналіз даних дозволяє зробити висновок, що між стажем і змінним виробітком існує прямий зв'язок, хоча і дуже тісний. Зокрема, виробіток четвертого робітника зі стажем 10 років становить 336 шт., а зі стажем 7 років — відповідно 352 шт.; восьмий робітник, у якого стаж 16 років, виробляє 372 шт. виробів, тоді як робітник зі стажем 14 років — відповідно 404 шт. і т. д.

Таблиця 3.11

**Стаж роботи та випуск продукції робітниками механічного цеху**

| Стаж роботи, років | Випущено виробів за зміну, шт. | Стаж роботи, років | Випущено виробів за зміну, шт. |
|---|---|---|---|
| 2 | 272 | 13 | 380 |
| 4 | 295 | 14 | 404 |
| 7 | 352 | 16 | 372 |
| 10 | 336 | 21 | 408 |
| 11 | 328 | 27 | 396 |

Виникає питання, чому в нашому прикладі зі збільшенням стажу виробіток робітників підвищується, але не у всіх? Очевидно, що на величину виробітку робітників, крім стажу впливають й інші фактори, такі як кваліфікація, проходження спеціального технічного навчання та ін. Дослідника ж цікавить вплив лише одного фактора — стажу роботи. Щоб визначити вплив одного фактора на результативний показник, потрібно усунути або пом’якшити вплив всіх інших факторів. Таким прийомом є факторне групування.

Розглянемо практичне застосування методу групувань за даними табл. 3.1. Застосовуючи метод аналітичного групування, виявимо характер залежності випуску продукції від величини основних виробничих засобів підприємств у даній галузі. При групуванні за факторною ознакою необхідно утворити п’ять груп підприємств з рівними інтервалами.

Перш ніж проводити будь-яке групування, необхідно виділити групувальну ознаку, або основу групування. З економічної теорії відомо, що основні засоби (будівлі, споруди, машини, обладнання тощо) є найважливішим фактором, від якого значною мірою залежить обсяг виробленої продукції. Щоб мати відомості про групи підприємств, цю сукупність підприємств необхідно розчленити за розміром вартості основних виробничих засобів. Отримані групи підприємств охарактеризуємо такими показниками: вартістю основних виробничих засобів, кількістю робітників та товарною продукцією підприємства.

Щоб скласти угруповання, треба спочатку визначити початок відліку інтервалів — нижню межу першого інтервалу. За це береться число, найближче до мінімального значення ознаки.

В якості досліджуваної ознаки візьмемо, наприклад, вартість основних виробничих засобів і побудуємо по ньому ряд розподілу з рівними закритими інтервалами. Під час розрахунку інтервалів утворимо п’ять груп підприємств. Тоді величина інтервалу дорівнюватиме:

$$i = \frac{x_{\max} - x_{\min}}{n} = \frac{76{,}0 - 11{,}0}{5} = 13.$$

Величину інтервалу можна округляти до цілого числа.

Тепер утворюємо групи підприємств, що відрізняються одна від одної за середньорічною вартістю основних виробничих засобів на цю величину (млн грн):

| | | | |
|---|---|---|---|
| 1-а група | 11,0–24,0 | 4-а група | 50,0–63,0 |
| 2-а група | 24,0–37,0 | 5-а група | 63,0–70,0 |
| 3-я група | 37,0–50,0 | | |

Після того, як обрано групувальну ознаку, визначено кількість груп, інтервали групувань, необхідно розподілити підприємства по кожній групі. Розподіляємо сукупність, що вивчається, за виділеними групами з урахуванням того, що інтервал повинен бути однаковим у всіх групах. Показники, що характеризують діяльність підприємства, розносяться по п’яти вищезгаданим групам і підраховуються підсумки.

Техніка тут проста. Попередньо необхідно здійснити ранжування об’єктів спостереження, тобто розмістити значення ознаки у зростаючому чи спадному порядку, і рахунок вести за групами. Можна зробити вибірку необхідних значень із табл. 3.1 та занести їх попередньо до робочої таблиці:

| Групи підприємств за вартістю основних виробничих засобів, млн грн | Число підприємств |
|---|---|
| 11,0–24,0 | \| \| \| |
| 24,0–37,0 | \| \| \| \| \| \| \| \| \| |
| 37,0–50,0 | \| \| \| \| \| |
| 50,0–63,0 | \| \| \| |
| 63,0–76,0 | \| \| \| \| |

Кожна характеристика відповідає одиниці сукупності, тобто одному підприємству.

Складемо макет таблиці із системою показників, куди занесемо результати групування підприємств за середньорічною вартістю основних виробничих засобів (табл. 3.12).

Таблиця 3.12

| Номер п/п | Групи підприємств за середньорічною вартістю основних виробничих засобів, млн грн | Підприємства | | Основні виробничі засоби | | Число робітників | | Товарна продукція | |
|---|---|---|---|---|---|---|---|---|---|
| | | число підприємств | в % до підсумку | млн грн | в % до підсумку | осіб | в % до підсумку | млн грн | в % до підсумку |
| А | Б | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| I | 11,0—24,0 | | | | | | | | |
| II | 24,0—37,0 | | | | | | | | |
| III | 37,0—50,0 | | | | | | | | |
| IV | 50,0—63,0 | | | | | | | | |
| V | 63,0—76,0 | | | | | | | | |
| | Всього | | | | | | | | |

Групування починається з того, що за факторною ознакою окремі одиниці об'єднуються в однорідні групи. В нашому прикладі підприємства ми об'єднаємо в групи за однаковою вартістю основних виробничих засобів, наприклад, в інтервали 11,0–24,0; 24,0–37,0; 37,0–50,0; 50,0–63,0; 63,0–76,0 (мінімальна вартість серед 24 підприємств становить 11,0 млн грн, максимальна — 76,0 млн грн).

При складанні робочої таблиці одна й та ж варіанта має зустрічатися у двох групах, тобто кожна наступна група повинна починатися з нової наступної варіанти.

Для дослідження зібрані дані щодо 24 промислових підприємств. Для заповнення макета таблиці попередньо складемо робочу таблицю (див. табл. 3.13). За даними ранжованого ряду добре видно зміну вартості основних виробничих засобів та легко позначити межі груп. До групи підприємств з розміром вартості основних засобів 11,0–24,0 увійде перші по порядку три підприємства, до другої групи 24,0–37,0 увійдуть наступні по порядку дев'ять підприємств і т. д.

Об'єднуючи окремі підприємства у однорідні групи за вартістю основних виробничих засобів, можна побачити, що у кожній окремій групі виявляться підприємства з різною чисельністю робітників та обсягом товарної продукції. Оскільки інші факторні ознаки в однорідних групах мають випадкові значення, їх вплив у силу дії закону великих чисел може взаємопогашатися, й у групі ці ознаки виявляться якомусь середньому рівні. Тому в результаті зведення всіх інших

Таблиця 3.13

| Номер п/п | Групи підприємств за середньорічною вартістю основних виробничих засобів, млн грн | Номер підпри-ємства | Основні виробничі засоби, млн грн | Число робітників | Товарна продукція, млн грн |
|---|---|---|---|---|---|
| А | Б | 1 | 2 | 3 | 4 |
| I | 11,0—24,0 | 1 | 18,6 | 370 | 37,4 |
| | | 7 | 22,4 | 574 | 96,4 |
| | | 13 | 11,0 | 310 | 37,1 |
| Всього | | 3 | 219,4 | 1 254 | 170,9 |
| II | 24,0—37,0 | 2 | 28,6 | 380 | 96,5 |
| | | 5 | 29,4 | 392 | 64,7 |
| | | 9 | 36,2 | 267 | 23,4 |
| | | 10 | 31,5 | 205 | 37,6 |
| | | 16 | 28,5 | 402 | 39,6 |
| | | 18 | 25,9 | 455 | 89,1 |
| | | 19 | 33,7 | 300 | 22,4 |
| | | 21 | 30,8 | 270 | 16,5 |
| | | 22 | 26,5 | 250 | 127,8 |
| Всього | | 9 | 271,1 | 2 921 | 517,6 |
| III | 37,0—50,0 | 4 | 38,2 | 450 | 43,4 |
| | | 8 | 45,9 | 212 | 111,7 |
| | | 12 | 39,6 | 249 | 19,5 |
| | | 15 | 46,7 | 644 | 77,8 |
| | | 24 | 46,2 | 507 | 48,3 |
| Всього | | 5 | 216,6 | 2 062 | 300,7 |
| IV | 50,0—63,0 | 3 | 58,3 | 210 | 16,4 |
| | | 11 | 56,4 | 200 | 15,4 |
| | | 17 | 61,8 | 307 | 34,7 |
| Всього | | 3 | 176,5 | 717 | 66,5 |
| V | 63,0—76,0 | 6 | 76,0 | 278 | 29,9 |
| | | 14 | 67,5 | 415 | 26,6 |
| | | 20 | 67,6 | 340 | 27,9 |
| | | 23 | 70,6 | 445 | 56,7 |
| Всього | | 4 | 281,7 | 1 478 | 141,1 |
| Разом | | 24 | 997,9 | 8 432 | 1 196,8 |

факторних ознак до середнього рівня на результативну ознаку в кінцевому результаті впливатиме лише одна факторна ознака, за якою проведено факторне угруповання, тобто на обсяг товарної продукції впливатиме лише вартість основних виробничих засобів.

Групові таблиці робочої таблиці занесемо у відповідні рядки та графи макета таблиці та отримаємо остаточну зведену групову таблицю з результатами групування підприємств за середньорічною вартістю основних виробничих засобів (табл. 3.14).

Таблиця 3.14

**Залежність обсягу виробництва від вартості основних засобів**

| Номер п/п | Групи підприємств за середньорічною вартістю основних виробничих засобів, млн грн | Підприємства | | Основні виробничі засоби, млн грн | | Число робітників | | Товарна продукція | |
|---|---|---|---|---|---|---|---|---|---|
| | | число підприємств | в % до підсумку | млн грн | в % до підсумку | осіб | в % до підсумку | млн грн | в % до підсумку |
| А | Б | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| I | 11,0—24,0 | 3 | 12,5 | 52,0 | 5,2 | 1 254 | 14,9 | 170,9 | 14,3 |
| II | 24,0—37,0 | 9 | 37,5 | 271,1 | 27,2 | 2 921 | 34,6 | 517,6 | 43,2 |
| III | 37,0—50,0 | 5 | 20,8 | 216,6 | 21,7 | 2 062 | 24,5 | 300,7 | 25,1 |
| IV | 50,0—63,0 | 3 | 12,5 | 176,5 | 17,7 | 717 | 8,5 | 66,5 | 5,6 |
| V | 63,0—76,0 | 4 | 16,7 | 281,7 | 28,2 | 1 478 | 17,5 | 141,1 | 11,8 |
| | Всього | 24 | 100,0 | 997,9 | 100,0 | 8 432 | 100,0 | 1 196,8 | 100,0 |

Таким чином, на відміну від ряду розподілу (табл. 3.1) групування дозволяє зробити конкретні та змістовні висновки. Дане групування показує, що в невеликих підприємствах зосереджена значні за обсягами ресурси. Близько 50 % підприємств (групи І та ІІ) мають у розпорядженні 32,2 % всіх основних виробничих засобів і виробляють при цьому 57,5 % товарної продукції, маючи при цьому лише 49,5% загальної кількості робітників.

Факторні угруповання є сильним прийомом вивчення взаємозв’язків між явищами. Вони дають можливість зробити висновок про наявність зв’язку, про форму її (пряма, зворотна), приблизно характеризують силу зв’язку. Метод факторних групувань дозволяє не тільки встановити зв’язок між ознаками економічного явища, а й відобразити мінливість і різноманіття зв’язків усередині сукупності, розкрити причину тієї чи іншої явища. Однак факторні групування не дають зведену характеристику, наскільки змінюється результативна ознака при зміні на одиницю факторної ознаки, яка тіснота зв’язку, який із факторів у цьому випадку провідний.

## 4.7. Вторинні групування

При статистичному дослідженні іноді використовують *метод вторинного угруповання* — прийом, що використовується для утворення нових груп на основі раніше зробленого групування, тобто перегрупують статистичний матеріал, вже зведений в групи.

Необхідність у вторинному групуванні виникає у двох основних випадках. По-перше, коли в результаті первинного групування нечітко виявився характер розподілу одиниць сукупності. Це може виникати при розбивці даних у невиправдано дрібні, часто формальні групи. По-друге, необхідність у вторинному угрупованні виникає з потреби у порівняльній та узагальнюючій оцінці явищ за матеріалами, зібраними у різних місцях, що відрізняються специфічними умовами та обробленими по-різному.

Вторинна групування може проводитись шляхом зведення в нові групи за тією самою ознакою, за якою проведено первинне групування статистичних даних. І в цьому випадку проводять або укрупнення інтервалів, або навпаки, розчленовування їх. Вторинне угруповання також використовується для отримання порівняних даних двох угруповань, якщо їх групи утворені з різних підстав під час проведення первинного угруповання. Вторинне групування або перегрупування згрупованих даних застосовується для кращої характеристики явища, що вивчається.

Вторинне групування може вестись як операція з укрупнення або розчленування існуючих груп та утворення нових груп.

При вторинному групуванні застосовується два способи утворення нових груп: 1) зміна величини інтервалів первинного угруповання; 2) дольове перегрупування. Найчастіше здійснюється укрупнення інтервалів, що дає більш яскравішу картину розвитку явища. Застосовуючи метод вторинного групування, зазвичай виходять із припущення, що розміри показників усередині груп змінюються рівномірно.

**Метод укрупнення інтервалів**. На практиці іноді трапляються випадки, коли угруповання даних проведено невдало. Провали, що пов’язані з інтервалами без даних, загальна розтягнутість — усе це не дозволяє уявити з належною чіткістю характерні риси даного розподілу.

Розглянемо для прикладу таке угруповання (табл. 3.15). Залежність не зовсім зрозуміла.

Таблиця 3.15

**Продуктивність ткачів комбінату за годину роботи на кожному верстаті,** *м*

| Групи ткачів за продуктивністю праці, *м* | Число ткачів | Питома вага ткачів групи у відсотках до підсумку, % |
|---|---|---|
| 3,25–3,50 | 4 | 1,0 |
| 3,50–3,75 | 3 | 0,8 |
| 3,75–4,00 | 21 | 5,3 |
| 4,00–4,25 | 74 | 18,5 |
| 4,25–4,50 | 102 | 25,5 |
| 4,50–4,75 | 88 | 22,0 |
| 4,75–5,00 | 62 | 15,5 |
| 5,00–5,25 | 25 | 6,3 |
| 5,25–5,50 | 7 | 1,8 |
| 5,50–5,75 | 6 | 1,5 |
| Понад 5,75 | 8 | 2,0 |
| Всього | 400 | 100,0 |

Наведене групування дозволяє бачити число і структуру сукупності, але не показує чіткої та суворої закономірності зміну груп ткачів за продуктивністю праці. Допустимо, потрібно утворити за даними таблиці замість одинадцяти п’ять груп, що призведе до меншої кількості груп. Виділимо за даними табл. 3.15 п’ять груп ткачів за продуктивністю праці: 3,0–4,0 м за годину, 4,0–4,5 м за годину, 4,5–5,0 м за годину, 5,0–5,5 м за годину та понад 5,5 м за годину. Для розрахунків чисельності та відсотків до підсумку за новими групами відносимо до групи ткачів з продуктивністю праці 3,0–4,0 м на годину 1-у, 2-у та 3-ю групи; групу ткачів з продуктивністю праці 4,0–4,5 м на годину — відповідно 4-у та 5-у групи; у групу ткачів з продуктивністю праці 4,5–5,0 м на годину — відповідно в 6-у та 7-у групи. До групи ткачів з продуктивністю праці 5,0–5,5 м на годину увійдуть повністю 8-а та 9-а групи та понад 5,5 м на годину — відповідно 10-а та 11-а групи.

Збільшення інтервалів призведе до меншої кількості груп. З огляду на це вторинне угруповання призведе до таких даних (табл. 3.16).

Таблиця 3.16

**Продуктивність ткачів комбінату за годину роботи на кожному верстаті**

| Групи ткачів за продуктивністю праці, *м* | Число ткачів | Питома вага ткачів групи у відсотках до підсумку, % |
|---|---|---|
| 3,0—4,0 | 28 | 7,0 |
| 4,0—4,5 | 176 | 44,0 |
| 4,5—5,0 | 150 | 37,5 |
| 5,0—5,5 | 32 | 8,0 |
| Понад 5,5 | 14 | 3,5 |
| Всього | 400 | 100,0 |

У табл. 3.16 нові групи утворені шляхом підсумовування початкових груп. Так, оскільки група ткачів з продуктивністю праці 3,0–4,0 м на годину була утворена шляхом підсумовування 1-ї, 2-ї та 3-ї груп, то чисельність ткачів цієї групи становитиме 28 осіб (4+3+21) , або 7,0 % їх загальної чисельності $\left(\frac{28\cdot 100}{400}\right)$. Аналогічно утворювалися групи і за окремими категоріями персоналу. Групування вийшло компактним і наочним. Якщо в первинному групуванні зміна чисельності працівників залежно від продуктивності праці виступає не надто наглядно, то перегрупування даних показує, що найбільшу частку у кількості ткачів на комбінаті займає група з продуктивністю 4,0–4,5 м за годину (44,0 %). Дещо менша чисельність ткачів припадає на 3-ю групу з продуктивністю праці 4,0–4,5 м за годину (150 осіб, або 37,5%). Найменшою (14 осіб) виявилася група ткачів із продуктивністю праці понад 5,5 м за год.

Таким чином, тут вторинне групування виступає у ролі підсилювача закономірності, що спостерігається в первинному групуванні лише приблизно, не зовсім ясно.

**Спосіб розщеплення інтервалів**. У розглянутому прикладі довелося укрупнювати інтервали. Але на практиці статистик зустрічається і з такими випадками, коли спочатку прийняті інтервали доводиться розкрупнювати. Занадто великі інтервали можуть призвести до того, що особливості розподілу виявляться, так би мовити, «змазаними». Розщеплення інтервалів здійснюють, якщо необхідно перегрупувати варіаційний ряд без звернення до первинного матеріалу. Перегрупування даних за другим способом може здійснюватися кількома способами.

Розглянемо розщеплення інтервалів *за принципом пропорційності*. Спочатку намічаються групи з їхньої питомої ваги за показником числа одиниць сукупності. Потім відповідно до цього розраховуються всі інші показники груп.

Маються такі дані про розподіл підприємств двох галузей за кількістю працівників (табл. 3.17).

Таблиця 3.17

| I галузь | | | II галузь | | |
|---|---|---|---|---|---|
| Номер п/п | Групи підприємств за кількістю працюючих | Число підприємств у відсотках до підсумку | Номер п/п | Групи підприємств за кількістю працюючих | Число підприємств у відсотках до підсумку |
| 1 | До 100 | 0,5 | 1 | До 50 | 1,5 |
| 2 | 100–200 | 3,2 | 2 | 50–75 | 9,0 |
| 3 | 200–300 | 26,9 | 3 | 75–100 | 25,0 |
| 4 | 300–500 | 25,6 | 4 | 100–150 | 20,0 |
| 5 | Понад 500 | 43,8 | 5 | 150–250 | 8,0 |
| | | | 6 | 250–350 | 12,0 |
| | | | 7 | 350–500 | 8,0 |
| | | | 8 | Понад 500 | 16,5 |
| Всього | | 100,0 | Всього | | 100,0 |

Наведені дані не дозволяють зробити порівняння розподілу підприємств двох галузей за чисельністю працівників, оскільки у цих галузях є різна кількість груп підприємств.

Для порівняння підприємств двох галузей за чисельністю робітників необхідно їх привести до порівнянного виду. З метою порівняння показників необхідно провести перегрупування підприємств II галузі за кількістю працюючих, взявши за основу порівняння розподіл підприємств I галузі. Отже, по II галузі треба провести вторинне групування, або перегрупування підприємств, утворивши таку ж кількість груп з тими ж інтервалами, що й у I галузі. В результаті перегрупування отримаємо такі порівнянні дані, що характеризують розподіл підприємств за кількістю працівників (табл. 3.18).

Пояснимо розрахунки. У першу, новостворену групи підприємств II галузі з числом працівників до 100 увійде перші три групи підпри-

Таблиця 3.18

| Номер п/п | Групи підприємств за чисельністю працівників | Питома вага підприємств групи у відсотках до підсумку | | |
|---|---|---|---|---|
| | | I галузь | II галузь | |
| 1 | 2 | 3 | 4 | 5 |
| 1 | До 100 | 0,5 | 35,5 | 1,5+9,0+25,0=35,5; |
| 2 | 100–200 | 3,2 | 24,0 | $20+\frac{50}{100}8=24;$ |
| 3 | 200–300 | 26,9 | 10,0 | $4+\frac{50}{100}12=10;$ |
| 4 | 300–500 | 25,6 | 14,0 | $6+\frac{100}{100}8=14;$ |
| 5 | Понад 500 | 43,8 | 16,5 | 16,5 |
| | Всього | 100,0 | 100,0 | |

ємств, сума частот яких дорівнює (1,5+25,0+9,0). Тепер треба утворити другу групу підприємств із кількістю працівників 100–200. До неї входить четверта група підприємств із числом працівників 100–150, що становить 20,0 % від загальної кількості підприємств, а також частина п’ятої групи. Для обчислення числа підприємств, яке треба взяти з п’ятої групи до новоствореної, умовно приймемо, що це число підприємств має бути пропорційно до питомої ваги відібраних працівників у групі. Питома вага 50 працівників у 5-й групі дорівнює:

$$\frac{50}{250-150}=\frac{50}{100}=0{,}5,$$

тобто становить 50 %. Отже, до нової групи треба взяти половину підприємств із п’ятої групи:

$$\frac{50\cdot 8}{100}=\frac{1}{2}\cdot 8=4.$$

Тоді питома вага підприємств новоствореної групи становить: 20+4=24. Аналогічно здійснюються розрахунки при утворенні інших груп. Якщо поруч із частостями є чисельні значення показників по групам, всі розрахунки показників за новоствореними групами здійснюється у тому ж співвідношеннях, що чисельність одиниць розподілу.

## 3.8. Комбіновані групування

У аналізованих раніше групуваннях одиниці статистичної сукупності групувалися за однією ознакою. В основі групувань лежала одна ознака. У практиці зустрічаються групування, при яких розбиті на групи сукупності піддаються подальшому поділу на групи за одним, а іноді за двома-трьома додатковими ознаками.

Групування, в якій розчленування сукупності одиниць на групи проводиться за двома або більше ознаками, взятими в поєднанні (комбінації), називається *комбінованим*. При комбінованому групуванні групи, утворені за однією ознакою, поділяються на підгрупи за іншою ознакою, а отримані таким чином підгрупи поділяються (кожна окремо) ще на підгрупи і т. д. Наприклад, просте групування робітників підприємства за рівнем кваліфікації — за тарифними розрядам — доповнити групуванням за стажем роботи, виділивши в кожній з цих груп підгрупи за стажем роботи до 3 років, від 3 до 6 років і понад шість років, то отримаємо комбінаційне групування. Результат цього угруповання викладається у комбінаційній таблиці.

Зробимо групування даних заводів за двома ознаками: вартістю товарної продукції та середньообліковою чисельністю промислово-виробничого персоналу, використовуючи групи за вартістю основних засобів, утворених при групуванні за однією ознакою.

Кожну групу і підгрупу охарактеризуємо такими показниками: кількість підприємств, середньорічна вартість основних виробничих засобів, середньооблікова чисельність промислово-виробничого персоналу, товарна продукція. Групування підприємств харчової промисловості за вартістю основних виробничих засобів та чисельності персоналу наведене в табл. 3.19.

У цій таблиці ми бачимо залежність результативного фактора (товарної продукції) від двох факторів — вартості основних виробничих засобів та чисельності персоналу. Залежність товарної продукції від вартості основних виробничих засобів та чисельності персоналу пряма, тобто розмір товарної продукції в більшості груп збільшується зі зростанням розмірів вартості основних засобів та чисельності персоналу.

Правильно побудовані комбіновані групування, взяті у системі, можуть дати багатий аналітичний матеріал для аналізу, наприклад, при аналізі продуктивності праці, собівартості продукції, нормативів питомих витрат сировини, матеріалів тощо.

Таблиця 3.19

| Номер групи | Групи підприємств за вартістю основних виробничих засобів, тис. грн | У тому числі підгрупи за середньообліковою чисельністю персоналу, осіб | Число підприємств | Основні виробничі засоби, тис. грн | Середньооблікова чисельність промислово-виробничого персоналу, осіб | Товарна продукція в оптових цінах, тис. грн |
|---|---|---|---|---|---|---|
| 1 | 11,0–24,0 | 200–350<br>350–500<br>500–650 | 1<br>1<br>1 | 11,0<br>18,6<br>22,4 | 310<br>370<br>574 | 37,1<br>37,4<br>96,4 |
| | Всього | | 3 | 52,0 | 1254 | 170,9 |
| 2 | 24,0–37,0 | 200–350<br>350–500<br>500–650 | 5<br>4<br>– | 158,7<br>112,4<br>– | 1292<br>1629<br>– | 227,7<br>289,9<br>– |
| | Всього | | 9 | 271,1 | 2921 | 517,6 |
| 3 | 37,0–50,0 | 200–350<br>350–500<br>500–650 | 2<br>1<br>– | 85,5<br>38,2<br>– | 461<br>450<br>– | 131,2<br>43,4<br>– |
| | Всього | | 3 | 123,7 | 911 | 174,6 |
| 4 | 50,0–63,0 | 200–350<br>350–500<br>500–650 | 3<br>–<br>2 | 176,5<br>–<br>92,9 | 717<br>–<br>1151 | 66,5<br>–<br>126,1 |
| | Всього | | 5 | 269,4 | 1868 | 192,6 |
| 5 | 63,0–76,0 | 200–350<br>350–500<br>500–650 | 2<br>2<br>– | 143,6<br>138,1<br>– | 618<br>860<br>– | 57,8<br>83,3 |
| | Всього | | 4 | 281,7 | 1478 | 141,1 |
| | Разом по підгрупам | 200–350<br>350–500<br>500–650 | 13<br>8<br>3 | 575,3<br>307,3<br>115,3 | 3398<br>3309<br>1725 | 520,3<br>454<br>222,5 |
| | Разом | | 24 | 997,9 | 8432 | 1196,8 |

Спочатку утворюються групи за однією ознакою, потім виділені групи поділяються на підгрупи за іншою ознакою, у свою чергу виділені підгрупи поділяються на підгрупи за наступною ознакою і т. д. Наприклад, якщо просте групування встановленого обладнання за віком доповнити групуванням за технічним станом, виділивши в кожній з цих груп підгрупи (придатне обладнання; обладнання, що вимагає капітального ремонту; непридатне обладнання, що підлягає списанню), то отримаємо комбіноване угруповання. Результат цього угруповання викладається у комбінованій таблиці.

Між ознаками не може бути задане відношення старшинства ознак, коли значення молодшої ознаки булі визначені на області значень старшої ознаки. Черговість розподілу сукупності за кожною ознакою може бути довільною (на розсуд дослідника).

Групування населення України за місцем проживання та статтю (рис. 3.1).

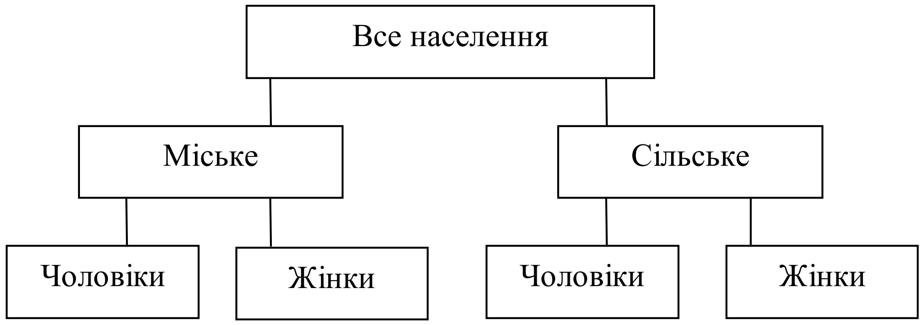



*Рис. 3.1.* Групування населення України за місцем проживання та статтю.

У комбінаційному групуванні, наприклад, групування населення взятим у комбінації ознаками «Місце проживання» та «Стать». Кожною з цих ознак володіє кожен громадянин країни. У цьому прикладі все населення спочатку розбито на дві групи за місцем проживання, а потім кожна з цих груп розбита на дві підгрупи за статтю. Можна зробити інакше. Спочатку виділити дві групи населення за статтю, а потім кожну з них поділити за місцем проживання (рис. 3.2).

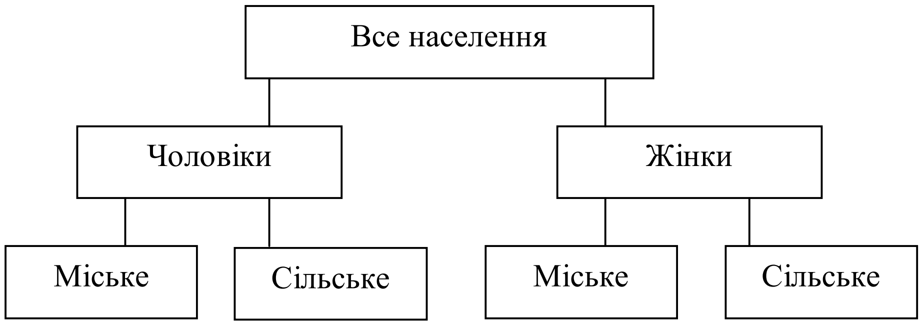



*Рис. 3.2.* Групування населення України за статтю та місцем проживання.

В обох випадках групування спостерігається відмінна ієрархія групувальних ознак. Однак оскільки між зазначеними ознаками відсутня ієрархія старшинства, то виділення груп і підгруп може бути різним залежно від мети дослідження.

Таким чином, комбінаційні групування дозволяють дати більш об'єктивну оцінку впливу окремих факторів на результативний показник. Необхідність застосування комбінованих групувань обумовлюється складністю та багатогранністю економічних процесів. Використання комбінаційних групувань в економіко-статистичному аналізі виробництва, де можуть бути побудовані групування шляхом різного поєднання ознак залежно від поставлених цілей та завдань, відкриває великі можливості в економіко-статистичних дослідженнях.

## ЛІТЕРАТУРА


1. *Горкавий В. К.* Статистика: Підручник. Третє вид., переробл. і доповн. Київ: Алерта, 2019. С. 55–71.
2. *Козирєва О. В.*, *Федорова В. О.* Статистика: навчальний посібник. Харків: Видавництво Іванченка І.С., 2021. С. 34–42.
3. *Ткач Є. І.*, *Сторожук В. П.* Загальна теорія статистики: підручник. 3-тє вид. Київ: Центр учбової літератури, 2009. С. 40–64.


## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники, словники*


4. *Мармоза А. Т.* Теорія статистики: підручник. 2-ге вид.,ketpepeoб. і доп. Київ: «Центр учбової літератури», 2013. С. 78–114.
5. *Педченко Г. П.* Статистика: Навчальний посібник. Мелітополь: Колор Принт, 2018. С. 64–84.
6. Статистичний словник / [*О. Г. Осауленко*, *О. О. Васєчко*, *М. В. Пугачова* та ін.]; за ред. д-ра держ. упр., проф., член-кор. НАН України О. Г. Осауленка; НТК статистичних досліджень. Київ: ДП «Інформ.-аналіт. агентство», 2012. 498 с.

# Г Л А В А 4

# СТАТИСТИЧНІ ТАБЛИЦІ

## 4.1. Сутність табличного викладення статистичних даних

**Поняття та значення статистичних таблиць**. Результати зведення та групування статистичних даних зазвичай викладаються у вигляді статистичних таблиць, що полегшують читання, аналіз та засвоєння даних. Аналітична таблиця — важливий результат лічильної та статистичної роботи. *Статистична таблиця* представляє собою найбільш поширену форму оформлення статистичних даних як систематично розташованих чисел, що характеризують ті чи інші масові явища чи процеси. У статистичних таблицях на відміну будь-яких інших таблиць логарифмів, відсотків, зворотних чисел тощо. завжди даються певні показники, що характеризують ту чи іншу область суспільного життя.

У таблицях дані виступають компактніше. Їх застосовують для кращої наочності та зручності порівняння показників. Таблиці дозволяють стисло і компактною, в статиці або динаміці викласти статистичну інформацію, відобразити кількісні характеристики об’єктів, явищ і процесів, зіставляти і встановлювати взаємозалежність між статистичними показниками, виявляти закономірності зміни характеристик.

У економіці таблиці відіграють значну роль. Працівник управління, економіст повинен уміти правильно їх складати та робити правильні висновки. Таблиці дають можливість встановлювати закономірності розвитку соціально-економічних явищ та процесів, давати їм кількісну оцінку, встановлювати взаємозв’язки між показниками, зробити певні висновки та висновки. Грамотно сконструйована таблиця дає вичерпний аналітичний матеріал, який часто не потребує якихось пояснень. Вона здатна замінити багато сторінок текстового матеріалу. Аналіз зведених у таблиці даних дозволяє точно і швидше виявити тенденції та закономірності розвитку досліджуваного об’єкта, правильно зробити певні висновки та висновки.

**Виникнення таблиць**. Представлення даних у формі таблиць, настільки звичне нашого часу, почали застосовуватися лише у XVIII ст. До цього часу результати статистичних спостережень оформлялися як громіздких текстових описів.

Вперше статистичні таблиці були використані російським географом і статистом Кириловим. У своїй праці «Квітучий стан Всеросійської держави ...», написаної в 1727 р., Кирилов описує окремі області держави і всю державу в цілому за допомогою таблиць, не вдаючись до поіменній строчній побудові. У західній статистиці винахід табличного способу викладу приписують данцю Анхерсену. Однак його робота побачила світ тільки в 1741 р., тобто через 14 років після написання роботи Кирилова. Таблиці Анхерсена були примітивнішими, ніж перші таблиці Кирилова.

В подальшому таблиці стали широко застосовуватися у всіх сферах діяльності, оскільки вони дозволяли значно наочніше і компактніше викласти підсумки статистичної зведення і угруповання.

## 4.2. Елементи таблиці

За змістом таблиці представляє собою думку, виражену у числах, тому подібна до граматичного речення, у якому є підмет і присудок.

*Підмет таблиці* — це одиниці статистичної сукупності або їх групи, що характеризуються цифровими даними. Інакше кажучи — це об’єкт або група об’єктів, про які йдеться в даній таблиці (наприклад, в ній характеризуються промислові підприємства, галузі, продукція, матеріально-технічні цінності, лікарні, школи, населення і т. д.), або досліджуваний об’єкт. Крім того, це можуть бути сукупність, окремі одиниці сукупності в порядку їх переліку або згруповані за однією або декількома ознаками територіальними одиницями, часові періоди і т. д. Наведемо приклад розробки підмету таблиці.

*Присудок* — це сукупність статистичних показників, що виража-

Таблиця 4.1

Перечнева таблиця. Підмет — перелік сільськогосподарських підприємств району.

| Сільськогосподарські підприємства | Озима пшениця | Ячмінь | Овес |
|---|---|---|---|
| АФ «Зоря» | 34,5 | 26,7 | 22,4 |
| ТОВ «Факел» | 39,7 | 25,1 | 23,6 |
| ФГ «Перемога» | 31,5 | 23,7 | 25,9 |

Таблиця 4.2

*Групова таблиця*. Підмет — групи сільськогосподарських підприємств, утворені за провідним напрямком:

| Групи сільськогосподарських підприємств за типом спеціалізації | Озима пшениця | Ячмінь | Овес |
|---|---|---|---|
| Зернові | 32,5 | 27,1 | 22,8 |
| Зерново-свинарські | 29,9 | 24,3 | 20,4 |
| Молочно-зернові | 35,4 | 23,5 | 25,3 |
| Всього | 32,8 | 25,7 | 22,9 |

ються відповідними цифровими даними, якими характеризується підмет (скажімо, показники випуску продукції, чисельності працівників, собівартості продукції тощо).

Розробка присудка може бути простою і складною. *Проста розробка присудка* означає послідовне перерахування показників, що характеризують підлягає. Наприклад, посівні площі зернових та зернобобових культур характеризуються: врожайністю, посівними площами, валовим збором.

*Складна розробка присудка* передбачає наявність додаткової одиниці.

Підмет таблиці зазвичай розташовується зліва у вигляді назви горизонтальних рядків, присудок — справа як найменування вертикальних граф. У випадку, коли присудок дуже громіздкий, а підмет є нескладним, перше рекомендується розміщувати зліва, а друге справа. Походження того й іншого зовсім по-різному.

Розробка присудка статистичної таблиці може бути простою і складною. Проста розробка присудка означає послідовне перерахування показників, що характеризують підмет. Наприклад, робітники промислового підприємства (підмет) характеризуються: числом, розміром заробітної плати, тарифним розрядом, стажем роботи.

Складна розробка присудка передбачає наявність додаткової одиниці вивчення, наприклад, основна заробітна плата (основна) і додаткова заробітна плата (додаткова). Складний присудок при двох і більше ознак цієї одиниці може бути некомбінованим і комбінованим. При складній комбінованій розробці присудка ознаки беруться у поєднанні, у комбінації один з одним, при некомбінованій — ізольовано.

Початком статистичного зведення є визначення підмету таблиці, яке представляє сам об'єкт, перелік його одиниць або їх групи, утворених за однією або декількома ознаками. Визначення підмету передує підрахунку. Присудок таблиці виникає в результаті підрахунку ознак, які характеризують певні значення одиниць спостереження. Ці значення ознак беруться із формуляру (бланка, картки, анкети).

Так, розробляючи статистичну таблицю технічного стану основних засобів різних груп, статистик спочатку сортує картки окремих одиниць основних засобів за групами: окремо відкладе картки, заповнені на будівлі, споруди та передавальні пристрої, окремо на машини та обладнання, окремо на транспортні засоби тощо; в групі машин і обладнання можна виділити верстати, пресове обладнання, верстати, випробувальні стенди та ін. Отримані групи та підгрупи і є підмет таблиці. Тільки після цього статистик починає рахувати залишкову вартість окремих груп основних засобів. З картки, що у цій групі, статистик вибирає величину залишкової вартості основних засобів і складує з величинами залишкової вартості, взятими з інших карток цієї групи основних засобів. В результаті такого розрахунку виходить результат вартості основних засобів за всією підгрупою. Для отримання підсумків по групі вже не потрібно звертатися до карток; можна скласти підсумки з підгруп. Так виникає присудок — результат внутрішньогрупового розрахунку.

При складанні таблиць у них слід включати ті дані, які відображають найважливіші і суттєві боку досліджуваного явища.

Складовим елементом статистичної таблиці є загальний заголовок, в якому дається коротка назва, що вказує на її вміст. Часто загальним заголовком, що міститься над таблицею, праворуч пишеться слово «Таблиця» та порядковий номер, якщо наводяться дві та більше таблиці. Заголовки мають бути точними та короткими.

По виду побудови статистичні таблиці є перетином горизонтальних рядків і вертикальних граф (колонок, стовпців), що утворюють графо-клітини (див. рис. 4.1). Рядки зазвичай служать для запису підмету таблиці, а вертикальні ділення — для ознак, що становлять присудок. Графи і рядки складають як би скелет таблиці. Перетин горизонтальних і вертикальних ліній утворює *клітини таблиці*, у яких розташовуються записи цифрових даних. Дуже часто наприкінці таблиці розташована *підсумковий рядок*, що представляє собою місце для запису результату підрахунку (сумування) чисельності одиниць сукупності та суми значень їх ознак по всій статистичній сукупності.

Проте чи всяка форма, розграфлена в такий спосіб і заповнена цифрами, представляє собою статистичну таблицю. Табличну форму можуть мати опитувальний бланк або якийсь формуляр, або схема розрахунку. Можна назвати дві основні відмінності статистичної таблиці з інших табличних уявлень. Перше, що відрізняє статистичну таблицю з інших табличних уявлень, у тому, що таблиця має бути обов'язково результатом підрахунку фактичних даних. Друга відмінність полягає в тому, що в таблиці воєдино дані щодо багатьох або, принаймні, кількох одиниць або об'єктів спостереження. Статистична таблиця має бути результатом зведення, тобто з'єднання шляхом класифікації, групування та розрахунку багатьох початкових записів, підрахунків, обчислень.

Бланк статистичної таблиці, що має заголовки рядків та граф, які ще не заповнені цифровими даними, називається *макетом таблиці*. Макети таблиць складаються для аналізу статистичних даних, розробки матеріалів статистичного спостереження. Особливо велику роль відіграє раціонально розроблена система макетів статистичної таблиці при проведенні великих статистичних досліджень, наприклад, при проведенні перепису населення в країні.

Принципова схема більшості статистичних таблиць може бути представлена у вигляді, показаному на рис. 4.1.

У більшості випадків підмет розташовується у горизонтальних рядках таблиці, а присудок — у вертикальних графах. Проте можуть бути й винятки, що не змінюють, втім, суті справи.

Упорядкування макетів статистичних таблиць — найважливіша умова планування розробки статистичних матеріалів. Зауважимо, що й самі статистичні таблиці складаються вже узагальнення даних,

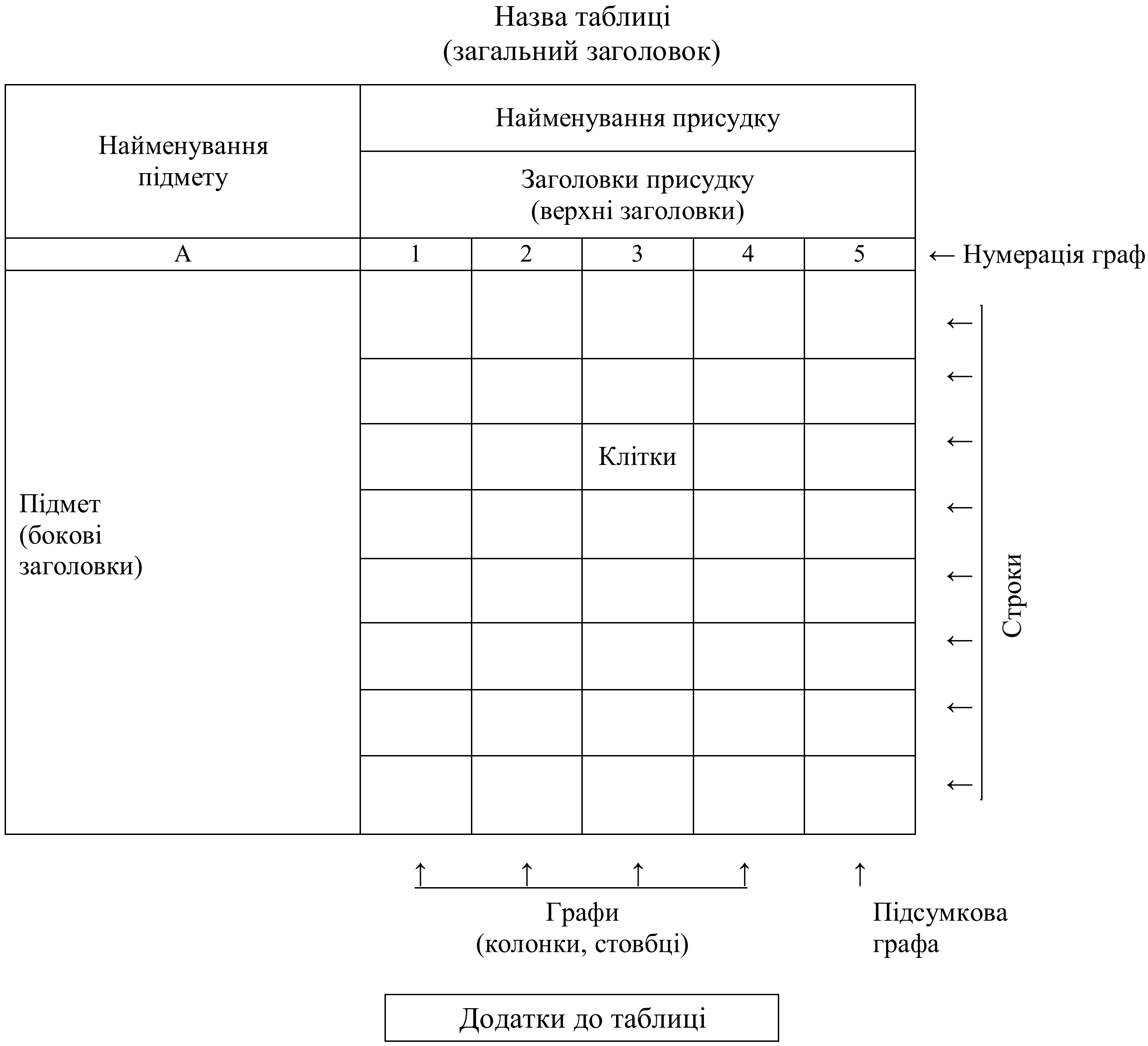



*Рис. 4.1.* Схема статистичної таблиці.

зібраних у процесі статистичного спостереження, то макети таблиць проектуються ще за складанні плану дослідження, тобто перед статистичним спостереженням.

## 4.3. Прості таблиці

*Простими* називаються такі таблиці, в підметі яких немає групувань, а дається лише перелік (найменування) одиниць спостереження (одиниць сукупності), адміністративних районів або періодів часу. Відповідні таблиці можуть бути названі простими переліковими, територіальними чи хронологічними.

Прості таблиці мають широке поширення у статистичній практиці й у економічній статистиці. Статистична характеристика вартості вироблених товарів та послуг за регіонами країни з перерахуванням у підметі всіх областей є простою таблицею. Також простою таблицею є таблиця, що характеризує кількість працівників, зайнятих різними видами економічної діяльності, у цьому місті, районі, області. Прикладом простої таблиці є вартість виробництва, виробленої цьому підприємстві.

Слід зазначити ту обставину, що прості таблиці не дозволяють проаналізувати ні типи досліджуваного явища, ні його структуру, ні зв'язки та взаємозв'язки між окремими ознаками явищ або процесів. Найчастіше прості таблиці мають описовий характер. У простий таблиці, як правило, міститься перелік однорідних ознак, які становлять предмет досліджень.

Прикладом простої перелікової таблиці може бути табл. 4.3, у якій представлені п'ять підприємств однієї галузі.

Таблиця 4.3

**Залежність собівартості продукції від обсягу випущеної продукції за п'ятьма заводами галузі**

| Номер заводу | Обсяг випущеної продукції, млн грн | Собівартість одиниці продукції, грн |
|---|---|---|
| 1 | 13,6 | 547,3 |
| 2 | 21,4 | 526,7 |
| 3 | 23,6 | 517,5 |
| 4 | 36,2 | 486,4 |
| 5 | 41,4 | 472,1 |

В присудку в цій таблиці є графи з показниками обсягу випущеної продукції та собівартістю одиниці продукції. Одиниці сукупності в переліковій таблиці зазвичай розташовуються від кращих до гірших або від дрібних до великих.

Підмет може бути виражено також поєднанням цих ознак. Розглянемо для прикладу просту статистичну таблицю, в якій представлені основні підсумкові зведені показники, що характеризують динаміку економічного розвитку економіки України (табл. 4.4).

Таблиця 4.4 є простою таблицею. Оскільки в її підметі наведені відрізки часу, а в присудку — назва (перелік) об'єктів вивчення без будь-якого групування одиниць сукупності. В якості об'єктів дослідження в ній виступають обсяг реалізованої промислової продукції,

Таблиця 4.4

**Основні показники розвитку економіки України за 2015–2020 рр.**

| Показник | 2015 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|---|
| Валовий внутрішній продукт (у фактичних цінах), млрд грн | 1 988,5 | 2 981,2 | 3 560,3 | 3 977,2 | 4 191,9 |
| Індекс споживчих цін (грудень до грудня попереднього року), відсотків | 143,3 | 113,7 | 109,8 | 104,1 | 105,0 |
| Обсяг реалізованої промислової продукції, млрд грн | 1 776,6 | 2 625,9 | 3 045,2 | 3 019,4 | 3 201,9 |
| Капітальні інвестиції, млрд грн | 273,1 | 488,5 | 578,7 | 624,0 | 508,2 |
| Перевезення вантажів усіма видами транспорту , млрд т | 1,5 | 1,6 | 1,6 | 1,6 | 1,6 |
| Чисельність постійного населення (на кінець року), млн | 42,6 | 42,2 | 42,0 | 41,7 | 41,4 |

капітальні інвестиції, обсяг вантажоперевезень та чисельність населення.

Простими територіальними таблицями є таблиці, у яких наведено перелік різних територій (країн, областей тощо.), кожна з яких має специфічні показники. Як ілюстрацію наведемо таку просту територіальну таблицю (табл. 4.5). В ній представлені дані, що характеризують територію окремих країн світу.

Таблиця 4.5

**Територія окремих країн світу**

| | Територія, тис. км$^2$ |
|---|---|
| Україна | 603,5 |
| Білорусь | 207,6 |
| Туреччина | 774,8 |
| Іспанія | 506,8 |
| Німеччина | 357,0 |

*Джерело:* Державна служба статистики України. URL. http://www.ukrstat.gov.ua/

У практиці статистики часто зустрічаються різні поєднання простих (і не лише таблиць). Прикладом територіально-хронологічної таблиці може бути наступна таблиця, що характеризує динаміку виробництва промислової продукції з окремих країн світу (табл. 4.6).

Таблиця 4.6

**Темпи приросту (зниження) промислової продукції**

(процентів до попереднього року)

| Регіони | 2010 р. | 2015 р. | 2017 р. | 2018 р. | 2019 р. |
|---|---|---|---|---|---|
| Україна | 12,0 | –12,3 | 1,1 | 3,0 | –0,5 |
| Білорусь | 11,7 | –6,6 | 6,1 | 5,7 | 1,0 |
| Німеччина | 11,2 | 1,4 | 2,8 | 1,0 | –4,5 |
| Угорщина | 10,6 | 7,4 | 4,7 | 3,5 | 5,6 |
| Франція | 4,8 | 1,7 | 2,0 | 0,8 | 0,3 |

*Джерело:* Державна служба статистики України. URL. http://www.ukrstat.gov.ua/

Це проста динамічна таблиця. У ній підметом є перелік країн у поєднанні з періодами часу. У присудку дано індекси виробництва промислової продукції (у відсотках до попереднього року).

Прикладом перелікової хронологічної простої таблиці можуть служити дані, що характеризують виробництво продукції тваринництва (у всіх категоріях господарств) в Україні (табл. 4.7) та зміна народжуваності населення по різних країнах (табл. 4.8).

Таблиця 4.7

**Виробництво продукції тваринництва (в усіх категоріях господарств) в Україні за 1990–2020 рр.**

| | 1990 р. | 2000 р. | 2005 р. | 2010 р. | 2020 р. |
|---|---|---|---|---|---|
| М’ясо (у забійній вазі), тис. т | 4 358 | 1 663 | 1 597 | 2 059 | 2 477,5 |
| Молоко, тис. т | 24 508 | 12 658 | 13 714 | 11 249 | 9 263,6 |
| Яйця, млн шт. | 16 287 | 8 809 | 13 046 | 17 052 | 16 167,2 |
| Шерсть, т | 29 804 | 3 400 | 3 195 | 4 192 | 1 573 |

*Джерело:* Державна служба статистики України. URL. http://www.ukrstat.gov.ua/

Наведені у табл. 4.8 дані не дозволяють виділити окремі типи, дати якісну характеристику такому явищу, як народжуваність населення, ні виявити фактори, що надають вирішальний вплив на її зростання чи зниження.

За допомогою простих таблиць можна дати узагальнюючий аналіз розвитку якогось явища або процесу суспільного життя, охарактеризувати його розвиток у часі та у просторі. У свою чергу, прості

Таблиця 4.8

**Загальний коефіцієнт народжуваності населення**

(*кількість народжених на 1000 осіб*)

| Країна | 2010 р. | 2017 р. | 2018 р. | 2019 р. |
|---|---|---|---|---|
| Україна | 10,8 | 9,4 | 8,7 | 8,1 |
| Казахстан | 22,5 | 21,6 | 21,8 | 21,7 |
| Німеччина | 8,3 | 9,5 | 9,5 | 9,4 |
| Японія | 8,4 | 7,5 | 7,3 | 6,9 |
| Китай | 11,9 | 12,4 | 10,9 | 10,5 |

*Джерело:* Державна служба статистики України. URL. http://www.ukrstat.gov.ua/

таблиці можна поділити на ряд груп, серед яких можна виділити: таблиці перерахування, хронологічні та територіальні. Однак, цей поділ можна назвати умовним, оскільки іноді ці види можуть поєднуватися.

Статистичний аналіз простої таблиці обмежений наведенням паралельних рядів. Завдання вивчення взаємної залежності ознак явища, що характеризується, прості таблиці не вирішують. Більш глибокий аналіз простої таблиці неможливий через відсутність причинно-наслідкового зв'язку між показниками підлягає і присудка.

## 4.4. Групові таблиці

*Груповою* називається таблиця, підмет якої складається з груп одиниць сукупності, утворених за однією ознакою. Наприклад, групи підприємств за обсягом виробленої продукції, вартості основних виробничих засобів, чисельності населення і т. п. У підметі групової таблиці окремі елементи даного явища не просто перераховуються, а об'єднуються в групи. Кожен рядок у груповій таблиці поєднує в одне поняття або кілька якісних понять, або кілька характеристик.

Групові таблиці широко застосовуються в пізнавальних і практичних цілях, так як дозволяють охарактеризувати типи господарств, взаємозв'язок показників, структуру сукупності та її зміну у часі та просторі.

Розглянемо табл. 4.9. Витрати на інновації, які приведені в даній таблиці, складаються з окремих видів, які мають різний характер джерел фінансування. До власних коштів підприємств відносяться прибуток, амортизаційні відрахування та інші власні кошти промис-

Таблиця 4.9

**Джерела фінансування інноваційної діяльності промислових підприємств в Україні за 2017 – 2020 рр.**

| | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|
| Витрати на інновації — всього, млн грн | 9117,5 | 12180,1 | 14220,9 | 14406,7 |
| В тому числі за рахунок: | | | | |
| власних коштів підприємств | 7704,1 | 10742,0 | 12474,9 | 12297,7 |
| коштів державного бюджету | 227,3 | 639,1 | 556,5 | 279,5 |
| коштів інвесторів-нерезидентів | 107,8 | 107,0 | 42,5 | 125,3 |
| коштів інших джерел | 1078,3 | 692,0 | 1147,0 | 1704,2 |

*Джерело:* Державна служба статистики України. URL. http://www.ukrstat.gov.ua/

лових підприємств, які вони використовують для фінансування інноваційної діяльності і т. д.

Підмет табл. 4.10 показує структурний поділ населення за рівнем середньодушових еквівалентних доходів.

Таблиця 4.10

**Розподіл населення України за рівнем середньодушових еквівалентних грошових доходів у 2019–2020 рр.,** (процентів)

| | 2019 р. | | | 2020 р. | | |
|---|---|---|---|---|---|---|
| | всі домогосподарства | в том числі, які проживають | | всі домогосподарства | в том числі, які проживають | |
| | | в міських поселеннях | в сільській місцевості | | в міських поселеннях | в сільській місцевості |
| Все населення | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 |
| Із середньодушовими еквівалентними грошовими доходами за місяць, грн | | | | | | |
| До 3 000,0 | 17,6 | 13,0 | 27,2 | 12,5 | 9,7 | 18,8 |
| 3 000,1–4 000,0 | 19,7 | 18,4 | 22,5 | 20,0 | 17,4 | 25,3 |
| 4 000,1–5 000,0 | 18,7 | 18,7 | 18,6 | 20,3 | 20,2 | 20,4 |
| 5 000,1–6 000,0 | 14,0 | 14,7 | 12,6 | 14,4 | 14,5 | 14,2 |
| 6 000,1–7 000,0 | 9,1 | 9,8 | 7,6 | 9,6 | 10,4 | 7,7 |
| 7 000,1–8 000,0 | 6,9 | 8,0 | 4,5 | 7,6 | 9,0 | 4,9 |
| 8 000,1–90 00,0 | 4,0 | 4,4 | 3,1 | 4,9 | 5,6 | 3,2 |
| 9 000,1–10 000,0 | 2,9 | 3,8 | 1,2 | 2,9 | 3,5 | 1,9 |
| 10 000,1–11 000,0 | 2,5 | 3,0 | 1,6 | 2,3 | 2,8 | 1,3 |
| 11 000,1–12 000,0 | 1,5 | 1,9 | 0,6 | 1,8 | 2,2 | 0,9 |
| Понад 12 000,0 | 3,1 | 4,3 | 0,5 | 3,7 | 4,7 | 1,7 |

*Джерело:* Державна служба статистики України. URL. http://www.ukrstat.gov.ua/

Підметом цієї таблиці є групи населення за розміром середньодушового еквівалентного загального грошового доходу. Показники, що характеризують ці групи, становлять присудок таблиці. Таблиця показує, які групи населення та у якій місцевості отримують еквівалентні доходи. Як видно з таблиці, найбільша частина населення країни мала середньодушові загальні доходи у розмірі від 3 000,1 до 4 000,0 грн, від 4 000,1 до 5 000. На ці дві групи в 2019 р. припадало 45,7 % від усієї чисельності населення. Значна частина населення (близько 19%) має середньодушові доходи на місяць менше за прожитковий мінімум.

Таблиця дозволяє провести глибокий аналіз розподілу населення за рівнем середньодушових еквівалентних грошових доходів.

Приклад групової таблиці є табл. 4.11, в якій представлено групування сільськогосподарських підприємств за розміром сільськогосподарських угідь.

Таблиця 4.11

**Розподіл підприємств , які здійснювали сільськогосподарську діяльність, за розміром сільськогосподарських угідь у 2020 р.**

(*на 1 листопада*)

| | Кількість підприємств | | Площа сільськогосподарських угідь | |
|---|---|---|---|---|
| | од. | відсотків до загальної кількості | тис. га | відсотків до загальної площі сільськогосподарських угідь |
| Підприємства, що мали сільськогосподарські угіддя, у тому числі площею, га | 36 277 | 76,3 | 20 254,2 | 100,0 |
| до 5,00 | 1 975 | 4,2 | 6,4 | 0,0 |
| 5,01–10,00 | 1 877 | 3,9 | 14,7 | 0,1 |
| 10,01–20,00 | 3 061 | 6,4 | 47,7 | 0,2 |
| 20,00–50,00 | 9 395 | 19,7 | 353,3 | 1,7 |
| 50,01–100,00 | 4 626 | 9,7 | 333,0 | 1,6 |
| 100,01–500,00 | 7 889 | 16,6 | 1 928,1 | 9,5 |
| 500,01–10 00,00 | 2 716 | 5,7 | 1 957,9 | 9,7 |
| 1 000,01–2 000,00 | 2 409 | 5,1 | 3 458,9 | 17,1 |
| 2 000,01–3 000,00 | 1 030 | 2,2 | 2 500,3 | 12,4 |
| 3 000,01–4 000,00 | 473 | 1,0 | 1 629,9 | 8,1 |
| 4 000,01–5 000,00 | 247 | 0,5 | 1 099,2 | 5,4 |
| 5 000,01–7 000,00 | 263 | 0,6 | 1 535,9 | 7,6 |
| 7 000,01–10 000,00 | 132 | 0,3 | 1 107,0 | 5,5 |
| больше 10 000,00 | 184 | 0,4 | 4 280,1 | 21,1 |
| Підприємства, що не мали сільськогосподарських угідь | 11 246 | 23,7 | | |

*Джерело:* Державна служба статистики України. URL. http://www.ukrstat.gov.ua/

Підметом цій таблиці є групи підприємств за розміром сільськогосподарських угідь. Підметом групування сільськогосподарських підприємств, що наводиться нижче, є групування кількісне. До першої групи можна віднести сільськогосподарські підприємства і з площею 3 га, і з площею 4 га і т. д. Показники, що характеризують ці групи, становлять присудок таблиці. Як видно з таблиці, виробництво продукції сконцентровано на підприємствах площею від 1 000,1 до 2 000,0 га та понад 10 000,0 га. Підприємства цих двох груп, що становлять 5,5 % усіх підприємств, мають у своєму розпорядженні 48,2 % усіх сільськогосподарських угідь.

Групові таблиці широко застосовуються в пізнавальних та практичних цілях, оскільки дозволяють охарактеризувати типи явищ, взаємозв'язок показників, структуру сукупності та її зміну у часі та просторі.

Разом з тим, недоліком групових таблиць є те, що підмет, який входить в кожну групу, не містить інших показників, які характеризує підмет таблиці. Більш складним видом групових таблиць є такі таблиці, у яких підмет містить ряд одиниць, що характеризують об'єкт дослідження.

## 4.5. Комбінаційні таблиці

*Комбінаційними* називаються таблиці, підмет яких складається з груп або підгруп, утворених за двома або більшими ознаками, взятими в комбінації. Побудова комбінаційних таблиць має ряд суттєвих особливостей, що відрізняють її від інших видів. У комбінаційній таблиці кожна з груп, виділена в груповій таблиці, розчленовується на підгрупи за якоюсь новою ознакою (одною або декількома). Наприклад, працівники підприємства за стажем роботи можуть підрозділятися на підгрупи за іншою ознакою — віком, які, у свою чергу, можуть бути розчленовані за третьою ознакою — статтю, освіті і т. д.

Ідеї побудови комбінаційної таблиці була висловлена київським професором Н. І. Зібером і вперше практично застосовані А. Шлікевичем при розробці матеріалів земського перепису Козельського уїзду Чернігівської губернії в 1882 р. Теоретичне обґрунтування цього методу було дано А. Шліхевичем в статті «Що дають і що можуть дати подвірні записи» («Земський збірник Чернігівської губ.», 1890, № 3–4).

Відмінною особливістю комбінаційної таблиці є те, що у ній результат підраховується також у кожній групі, розчленованої на підгрупи. В результаті підсумок за групами і підгрупами називають *груповим*, за сукупністю — *загальним*.

У порівнянні з таблицями простими та груповими, комбінаційна таблиця містить інформацію, що дозволяє виявити та охарактеризувати взаємозв'язок низки показників та закономірності їх зміни як у просторі, так і в часі. Комбінаційні групування використовують із вивчення складних процесів, які відображають взаємозв'язок низки ознак. Правильно побудовані комбіновані групування, взяті в системі, можуть дати багатий аналітичний матеріал для аналізу, наприклад, при аналізі продуктивності праці, собівартості продукції, нормативів питомих витрат сировини, матеріалів тощо.

Якщо прості таблиці головним чином описують явища, дають інформацію, необхідну реалізації статистикою управлінської функції, то групові і комбінаційні служать переважно науково-пізнавальним цілям — виділенню типів і встановленню взаємної залежності ознак явища, що характеризується.

Наведемо групування встановленого обладнання машинобудівельного заводу за двома ознаками: за віком та технічною придатністю. Кожну групу і підгрупу охарактеризуємо такими показниками: кількість устаткування, зокрема застарілого, вартість устаткування, зокрема застарілого. Групування встановленого обладнання за віком та технічною придатністю дане в табл. 4.12.

За цими даними можна встановити вік устаткування для підприємства, і навіть його технічну придатність. Уявлення статистичних даних як комбінаційної таблиці дозволяє якісніше і всебічно охарактеризувати дане соціально-економічне явище життя.

Щоб комбінаційна таблиця була наочною, слід при розробці її підмета обмежуватися двома-трьома ознаками, утворюючи у кожному їх обмежене число груп. Якщо кількість підмету і присудка в комбінаційній таблиці є численним, їх необхідно нумерувати. Графи, у яких наведено перелік об'єктів чи його груп (підмет таблиці) визначають літерами алфавіту, а графи, у яких наведено показники присудка, — арабськими цифрами.

Комбінаційна таблиця, складена за двома ознаками, може бути замінена двома груповими таблицями, які, однак, не дозволяють повно виявити і охарактеризувати взаємозв'язок ряд показників, як комбінаційна таблиця.

Таблиця 4.12

**Угруповання встановленого обладнання за віком та за технічною придатністю**

(умовні дані)

| Групування встановленого обладнання | | Кількість обладнання, шт. | | Вартість обладнання, тис. грн | |
|---|---|---|---|---|---|
| за віком, років | за технічною придатністю | всього | в тому числі застарілого | всього | в тому числі застарілого |
| До 5 | придатне обладнання | 8 | 2 | 8452,4 | 2 198,4 |
| | обладнання, що потребує капітального ремонту | 2 | 1 | 3457,5 | 584,1 |
| | непридатне обладнання, що підлягає списанню | – | – | – | – |
| | По підгрупі в цілому | 10 | 3 | 1 1909,9 | 2 782,5 |
| от 5 до 10 | придатне обладнання | 10 | 2 | 1 1423,5 | 2 754,2 |
| | обладнання, що потребує капітального ремонту | 2 | 1 | 2 104,7 | 687,3 |
| | непридатне обладнання, що підлягає списанню | 3 | 1 | 4 123,2 | 1352,4 |
| | По підгрупі в цілому | 15 | 4 | 17 651,4 | 4 793,9 |
| Понад 10 | придатне обладнання | 13 | 10 | 7 528,3 | 6 009,9 |
| | обладнання, що потребує капітального ремонту | 22 | 20 | 6 851,1 | 6 631,5 |
| | непридатне обладнання, що підлягає списанню | 30 | 27 | 6 808,9 | 5 910,3 |
| | По підгрупі в цілому | 65 | 57 | 21 188,3 | 18 551,7 |
| Всього | | 90 | 64 | 50 749,6 | 26 128,1 |

У деяких випадках окремі моменти в таблицях можна пояснити у примітці. За допомогою приміток уточнюються деякі питання, що стосуються таблиці (джерело даних, час, виключення, відсутність якихось даних та ін.). Наприклад, при порівнянні рівня безробіття в Україні та інших країнах світу, які публікує ДСС України, у примітці вказується: «рівень зареєстрованого безробіття на кінець року».

Для дослідження залежності деяких ознак від трьох факторів будується комбінаційна таблиця з виділенням (у підметі) за трьома групувальними ознаками. Наприклад, щодо впливу фондоозброєності, землезабезпеченості і чисельності працівників витрати і прибуток треба так побудувати таблицю (табл. 4.13). Практична побудова комбінаційних таблиць показує, що групування за трьома і більше озна-

Таблиця 4.13

**Вплив фондоозброєності, землезабезпеченості та витрат на 1 га сільськогосподарських угідь на ефективність сільськогосподарського виробництва (макет)**

| Групи сільськогосподарських підприємств | | | Число сільсько-госпо-дарських підпри-ємств. | Вартість валової продукції на 100 га сіль-ськогоспо-дарських угідь, тис. грн | Прибуток, тис. грн | |
|---|---|---|---|---|---|---|
| за вартістю основних фондів, тис. грн | за розмі-ром ріллі на 1 пра-цівника, га | витрати на 1 га сільсько-господарських угідь, тис. грн | | | на 1 робіт-ника | на 1 га сіль-госп-угідь |
| До 100 | До 7,0<br>7,0 і більше | До 10,0<br>10,0 і більше | | | | |
| В середньому по групі | | | × | | | |
| 100–200 | До 7,0<br>7,0 і більше | До 10,0<br>10,0 і більше | | | | |
| В середньому по групі | | | × | | | |
| 200 і більше | До 7,0<br>7,0 і більше | До 10,0<br>10,0 і більше | | | | |
| В середньому по групі | | | × | | | |
| В середньому за всіма сільськогосподарськими підприємствами | | | × | | | |

ками є дуже громіздким, його важко читати і тому воно застосовуються рідко. Якщо припустити, що кількість груп за кожною ознакою складе 4, то при групуванні за двома ознаками число груп буде 16, при угрупованні за трьома ознаками — 64, при групуванні ж за чотирма ознаками — 256. Така таблиця є дуже громіздкою і важко читається. Тому на практиці найчастіше комбінаційну таблицю будують за двома групами факторів (у підметі) і за трьома групувальними ознаками (присудок).

Відмінною особливістю комбінаційних таблиць є те, що міняти довільне місце цієї ознаки в комбінації ознак не можна. Ознаки в комбінаційній таблиці розміщують або за важливістю, або за послідовністю вивчення.

## 4.6. Розробка присудка таблиць

Розробка присудка статистичних таблиць є найважливішим завданням їх правильної побудови. При великому числі ознак присудку можна по-різному підраховувати чисельні характеристики цих ознак.

Наприклад, дані про виробництво сільськогосподарської продукції країни, областей та районів за категоріями господарств та видами сільськогосподарської продукції можуть бути розміщені в присудку так:

| Продукція сільського господарства за категоріями господарств | | | Продукція сільського господарства за видами | | |
|---|---|---|---|---|---|
| підприємства | господарства населення | всього | продукція рослинництва | продукція тваринництва | всього |
| 1 | 2 | 3 | 4 | 5 | 6 |

При заповненні таблиці вийде достатньо докладна характеристика видів продукції (рослинництва та тваринництва) та ізольована продукція за категоріями господарств.

Але ці дані можуть бути розміщені і по іншому:

| Підприємства | | | Господарства населення | | |
|---|---|---|---|---|---|
| продукція сільського господарства | Із неї | | продукція сільського господарства | Із неї | |
| | продукція рослинництва | продукція тваринництва | | продукція рослинництва | продукція тваринництва |
| 1 | 2 | 3 | 4 | 5 | 6 |

У першому випадку проведена проста розробка присудка, у другому — комбінована. Якщо результаті простої розробки продукція сільського господарства розглядається окремо за видами підприємств та продукції, то комбінаційна розробка дає можливість розглянути продукцію сільського господарства у поєднанні цих двох сторін. Звідси видно, що комбінована розробка присудка цінніше, ніж проста.

Припустимо, що на підставі простої таблиці в підметі лежить територіальний ознака — області та райони, а кожна територіальна одиниця характеризується загальним приростом населення, природним приростом, міграційним приростом і міграційним приростом населення до 18 років і старше 18 років. Тоді заголовки таблиці будуть такими:

| Область, район | Загальний приріст, скорочення (–) | У тому числі | | | |
|---|---|---|---|---|---|
| | | природний приріст, скорочення (–) | міграційний приріст, скорочення (–) | у віці | |
| | | | | до 18 років | 18 років і старше |
| 1 | 2 | 3 | 4 | 5 | 6 |

При заповненні таблиці ми отримали досить детальну характеристику міграційного населення по районам і за двома ізольованими один від одного ознаками: приросту і віку. Така розробка присудку буде простою, оскільки кожна ознака підраховується незалежно одна від одної. Дійсно, за кожною областю чи районом можна отримати із таблиці відомості про весь загальний приріст населення і розподіл його за видами і віком.

Поглиблену статистичну характеристику приросту населення, чисельність його зміни за причинами і віком в комбінації цих ознак може дати комбінаційна розробка обох показників. При цьому заголовки граф таблиці будуть виглядати уже по-іншому:

| Область, район | Загальний приріст, скорочення (–) | У тому числі у віці | | | |
|---|---|---|---|---|---|
| | | до 18 років | | 18 років і старше | |
| | | природний приріст, скорочення (–) | міграційний приріст, скорочення (–) | природний приріст, скорочення (–) | міграційний приріст, скорочення (–) |
| 1 | 2 | 3 | 4 | 5 | 6 |

В цьому макеті таблиці ознаки присудку пов’язані між собою. Мається можливість прослідкувати не лише за приростом населення за причинами і віком окремо, але й визначити чисельність природного та міграційного приросту (скорочення) населення вказаних вікових груп. В простій таблиці ми можемо отримати відомості за кожною групою окремо. При комбінаційній таблиці кожна група може бути охарактеризована різним поєднанням цих ознак. Характер поєднання ознак залежить від мети дослідження. Повернемося до попередньої таблиці і в якості групувальної ознаки присудку візьмемо спочатку вид приросту, а потім вік населення:

<table>
<tr><th rowspan="3">Область, район</th><th rowspan="3">Загальний приріст, скорочення (–)</th><th colspan="4">У тому числі</th></tr>
<tr><th colspan="2">природний приріст, скорочення (–)</th><th colspan="2">міграційний приріст, скорочення (–)</th></tr>
<tr><th>до 18 років</th><th>18 років і старше</th><th>до 18 років</th><th>18 років і старше</th></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr>
</table>

При складанні таблиць необхідно зважити, які ознаки присудку буде простою, а які піддадуться комбінаційній розробці. Проте при складанні макету таблиці слід врахувати, що при комбінаційній розробці присудку таблиця може мати занадто велику кількість граф, які складно буде наочно розглянути. З іншого боку, аналітичне дослідження потребує комбінаційної розробки присудку. Із таблиць з комбінаційної розробки присудку легко можна отримати таблиці з простою розробкою присудку.

## 4.7. Оформлення таблиць

При побудові таблиць важливо дотримуватися суворо певних правил їх побудова та оформлення, основними з яких є наступні.

Таблиця має бути побудована так, щоб усіляко полегшувалося її читання. З цією метою таблицю слід поміщати тільки те, що, безумовно, необхідно для характеристики досліджуваного об’єкта. Не слід завантажувати статистичні таблиці зайвими цифровими матеріалами, так як це ускладнює їх читання та аналіз. Слід пам’ятати, що, чим менше таблиця, тим вона ясніша і дохідливіша.

**Загальний заголовок таблиці**. Назву таблиці (за її наявності) слід поміщати зверху над таблицею. Заголовок таблиці повинен коротко висловлювати основний зміст таблиці, але не перераховувати всього того, що в ній викладено. Наприклад, таблицю з даними про кількість виробленої промислової продукції по Україні в цілому, по областях можна назвати: «Виробництво промислової продукції за видами (в натуральному вираженні)»; таблицю з даними про зміну обсягу виробництва сільськогосподарської продукції за категоріями господарства (підприємства та господарства населення) та за видами (продукція рослинництва та продукція тваринництва) можна назвати так: «Індекси сільськогосподарської продукції». Бажано, щоб заголовок таблиці складався лише з 5–10 слів. При цьому необхідно пам'ятати, що чим лаконічніший текст заголовків, тим доступнішим і дохідливішим стає таблиця для читача.

У назві таблиці необхідно вказувати час, якого належить таблиця. Наприклад, «Посівні площі сорго під урожай 2021 року», або «Обсяг реалізованої промислової продукції за видами діяльності (2015–2022 рр.).

Якщо в таблиці наводяться дані, що охоплюють не всю сукупність, а лише її частину (наприклад, посівні площі озимої пшениці за всіма категоріями господарств та посівні площі озимої пшениці в сільськогосподарських підприємствах), то в назві таблиці слід виділити слова, що визначають, яку частину сукупності охоплює цю таблицю.

**Заголовок підмету і присудка**. Назви граф, що містять підмет і присудок, повинні бути короткими і зрозумілими, а сама таблиця побудована так, щоб вона являла собою закінчене ціле, незалежне від тексту, що її супроводжує. *Текст повинен слугувати висновком* з таблиці, *а не розшифровувати її зміст* і вже принаймні не повторювати те, що вміщено в таблиці.

Всі слова в заголовках підмету і присудку таблиці повинні писатися по можливості повністю. Для скорочення тексту заголовків та підзаголовків граф окремі поняття замінюють літерними позначеннями, встановленими ДСТУ 3582:2013 «Бібліографічний опис скорочення слів і словосполучень українською мовою. Загальні вимоги та правила (ISO 4:1984, NEQ 832:1994, NEQ), або інші позначення, якщо вони пояснені в тексті або наведені в ілюстраціях.

У статистичній таблиці необхідно вказувати *одиниці виміру*, місце та час, до яких належать дані. Якщо всі дані в статистичній таблиці дані мають одну одиницю виміру, то вона вказується в загальному заголовку, якщо різні — то в заголовках рядків або граф, зазвичай через кому. Крім того, у таблиці необхідно мати підсумкові показники. При перенесенні частини таблиці на ту саму або інші сторінки назву поміщають тільки над першою частиною таблиці.

Заголовки граф і рядків таблиці слід писати з великої літери, а підзаголовки граф — з малої літери, якщо вони складають одне речення із заголовком, або з великої літери, якщо вони мають самостійне значення. Наприкінці заголовків та підзаголовків таблиць крапки не ставлять. Заголовки і підзаголовки граф вказуються в однині.

Число показників присудка має бути по можливості обмеженим.

Заголовки присудка виражають короткими фразами. Наприклад, при порівнянні обсягу реалізованої продукції підприємства у 2022 р. порівняно з 2010 р. можна записати: «2022 р. у % 2010 р.».

У випадку, якщо в таблиці наводяться дані за певний період, їх потрібно розташовувати в хронологічному порядку.

Якщо таблиця містить абсолютні та відносні дані, то для зручності користування табличними даними спочатку вказують абсолютні показники, потім — відносні.

**Запис цифр у таблиці**. При округленні треба стежити, щоб після округлення не залишалася хоча б лише одна значна цифра. Наприклад, округлювати посівні площі сільськогосподарських культур в Україні до мільйонів за окремими роками не можна, оскільки за такими даними не можна буде побачити динаміку зміни посівних площ.

Якщо у якійсь клітині таблиці не можна проставити дані через те, що вони нині відсутні, у ній зазвичай ставляться три точки (…), або позначають літерами н. в. (немає відомостей). Якщо заповнити відповідну клітину неможливо через відсутність досліджуваного явища, у ній робиться прочерк (–). Якщо дана клітина не підлягає заповненню, то ній ставиться хрест (×). Якщо явище зустрічається в сукупності, але величина досліджуваного ознаки менше, ніж половина одиниці, в якій, змінюється ознака, то нульові значення статистичних даних в будь-якій клітині в таблицях прийнято позначати знаком тире (–). Позначення «0,0» означає, що число цієї клітини знаходиться за межами точності, прийнятої в таблиці.

Якщо таблиці дані наводяться у відсотках якогось попереднього року, цей рік має бути зазначений у таблиці, попри зазначення їх у заголовку. Наприклад, так (див. табл. 4.14).

Таблиця 4.14

**Темпи зростання необоротних активів за видами активів в Україні за 2019–2021 рр.**

(на кінець року; у відсотках до 2010 р.)

| | 2010 р. | 2019 р. | 2020 р. | 2021 р. |
|---|---|---|---|---|
| Всі необоротні активі | 100 | 394 | 423 | 540 |
| В тому числі: | | | | |
| матеріальні активи | 100 | 461 | 537 | 695 |
| нематеріальні активи | 100 | 163 | 297 | 321 |

У таблицях, що наводяться в аналітичних записках, значущість абсолютних цифр має бути найменшою.

При використанні багатозначних цифр треба кожні три цифри (починаючи праворуч) відокремлювати один від одного. Наприклад, цифру 3542861 потрібно записати 3 542 861. Якщо число має 6–7 знаків, то такі цифри краще округлювати до сотих (3,54).

У випадках, як у тексті робляться посилання рядки і графи таблиці або таблиця друкується на кількох сторінках, її рядки і графи (тільки графи) нумеруються. У присудку нумеруються тільки графи, заповнені показниками. Графи, до яких наводяться найменування підлягає таблиці і одиниці виміру, позначаються великими літерами А, Б і т. д., а графи присудка — арабськими цифрами. Якщо цифра в якійсь графі є результатом арифметичної дії з числами інших граф таблиці, то під заголовком цієї графи в дужках вказується, з цифрами яких граф і які дії треба зробити.

Таблиці з невеликою кількістю граф допускається ділити на частини та поміщати одну частину поряд з іншою на одній сторінці, при цьому розділяють частини таблиці подвійною лінією або лінією завтовшки 2s. При цьому повторюють головку таблиці відповідно до рис. 4.1.

Таблиця …

**Випуск продукції в цеху № 2**

| Місяці | Виробництво продукції, шт. | Місяці | Виробництво продукції, шт. | Місяці | Виробництво продукції, шт. |
|---|---|---|---|---|---|
| Січень | 467 | Травень | 645 | Вересень | 605 |
| Лютий | 482 | Червень | 489 | Жовтень | 669 |
| Березень | 568 | Липень | 447 | Листопад | 608 |
| Квітень | 468 | Серпень | 577 | Грудень | 575 |

Отже, статистична таблиця є не лише формою оптимального і компактного викладу статистичної інформації, а й знаряддям її аналізу, оскільки дозволяє проводити наочні зіставлення даних, дозволяють констатувати факти, встановлювати взаємозалежність між показниками, виявляти закономірності розвитку суспільних явищ.

Залежно від характеру об’єктів та відповідного розташування в їх підмету статистичні таблиці поділяються на прості, групові та комбінаційні. До простих відносяться статистичні таблиці, що не мають у підметі статистичних групувань, у груповій таблиці цифрові дані по окремих одиницях об'єднуються у групи за певними ознаками, у комбінаційній таблиці групи статистичного підмету розбиваються на підгрупи. Проста таблиця ставить собі завдання уявити лише перелік ознак та його кількісне значення. І групові, і комбінаційні таблиці складаються для аналітичних цілей. Вони служать і для вивчення структури явища, і для вивчення взаємної залежності ознак характеризованого явища.

## ЛІТЕРАТУРА


1. *Горкавий В. К.* Статистика : Підручник. Третє вид., переробл. і доповн. Київ: Алерта, 2019. С. 72–74.
2. Статистика: підручник / *С. І. Пирожков*, *В. В. Рязанцева*, *Р. М. Моторин* та ін. Київ: Київ. нац. торг.-екон. ун-т, 2020. С. 76–78.
3. *Ткач Є. І.*, *Сторожук В. П.* Загальна теорія статистики: підручник. 3-тє вид. Київ: Центр учбової літератури, 2009. С. 65–79.


## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники, словники*


4. Мармоза А. Т. Теорія статистики: підручник. 2-ге вид., перероб. і доп. Київ: «Центр учбової літератури», 2013. 115–120.
5. *Педченко Г. П.* Статистика: Навчальний посібник. Мелітополь: Колор Принт, 2018. С. 23–31.
6. Статистичний словник / [*О. Г. Осауленко*, *О. О. Васєчко*, *М. В. Пугачова* та ін.]; за ред. д-ра держ. упр., проф., член-кор. НАН України О. Г. Осауленка; НТК статистичних досліджень. Київ: ДП «Інформ.-аналіт. агентство», 2012. 498 с.

## Г Л А В А 5

# СТАТИСТИЧНІ ГРАФІКИ

## 5.1. Розвиток та застосування графічного методу у статистиці

**Сутність та застосування графічного методу**. Поряд із статистичними таблицями графіки є важливим засобом наочного уявлення та аналізу статистичних даних. *Графіком* у статистиці називаються зображення статистичних величин чи ознак у вигляді графічних (геометричних) образів. З їх допомогою статистичні дані можна надати наочно, лаконічно та доступно для розуміння будь-якої людини, що полегшує сприйняття інформації. Наочно зображуючи статистичні дані, графіки дозволяють за допомогою геометричних образів (крапок, ліній, фігур, їх поєднання), а також спрощених предметних зображень відобразити їх зміст, розміри та динаміку розвитку. Графіки, перетворюючи «сухі» цифри на наочні образи, спонукають людину обмірковувати сказане. Графічні методи зображення набули великого поширення в практичній роботі, особливо в економіці та організації виробництва.

Використання графічного методу дозволяє вирішувати наступні задачі: 1) охопити сукупність явищ; 2) показати розвиток явищ у часі (їх динаміку); 2) з’ясувати структуру досліджуваних явищ і структурні зрушення, які відбуваються у них; 3) порівняти статистичні величини між собою; 4) виявити найбільш характерні співвідношення та зв’язки між явищами; 5) визначити основні тенденції (тренди) розвитку явищ; 6) охарактеризувати виконання плану; 7) дати характеристику розподілу будь-якого показника.

Але, що важливіше і цінніше, графічний метод є потужним знаряддям узагальнення та аналізу статистичних даних, а деяких випадках — єдиним і незамінним способом їх дослідження. У всіх випадках креслення полегшить і прискорить засвоєння та осмислення матеріалу.

Графіки дозволяють наочно уявити закономірності, виявлені у процесі аналізу статистичних даних. Тому його краще використовувати в тих випадках, коли неважливі точні, детальні дані про явище, а необхідно протягом короткого часу пояснити слухачам або читачам суть, основні характеристики явища, що вивчається. Однак графіки поступаються таблицям в точності даних, що відображаються. Якщо показники таблиці важко осмислити, то графік може полегшити і прискорити їх засвоєння.

Тому в усіх тих випадках, коли основними вимогами до результатів статистичної роботи, що представляються споживачеві, є дохідливість і наочність тих характеристик сукупностей або закономірностей, які виявлені в ході статистичного дослідження, замість таблиць бажано використовувати графіки. Графічні образи дозволяють зіставити розміри досліджуваних явищ та процесів, створювати моделі структур, моделі динаміки розміщення та зв'язку явищ. Ці моделі є потужним засобом аналізу та прогнозування.

**З історії графічного методу**. Використання графічного методу у статистиці має двовікову історію. Зародження та становлення графічного методу у статистиці відбувалося у 80-х роках XVIII ст., коли з'явилися перші графічні зображення статистичних даних. Першими значними статистичними роботами із застосуванням графіків були, як відомо, двох «табличних статистиків» і «лінійних арифметиків кінця XVIII — початку XIX ст.: гессенського професора Августа Фрідріха Кроме (Grome, 1753–1833) та англійця Вільяма Плейфера (W. Playfair, 1759–1823). Перший з них видав у 1782 р. «Карту продуктів Європи з книгою, що відноситься до неї: Продукти Європи». У «Карті продуктів» Кроме користується примітивними картограмами, точніше — картами, на які умовними знаками нанесені місця видобутку та виробництва різних продуктів, а також діаграми порівняння. У 1785 р. була опублікована його книга «Про величину і населення усіх європейських держав», в якій було безліч таблиць про площу цих держав, чисельність та щільність населення, а також карта, на якій відносини між державами за цими показниками були зображені за допомогою квадратів , тобто діаграми порівняння.

Інтерес представляють міркування А. Кроме щодо призначення та ролі графічних зображень. «Взаємні відношення між різними величинами, — писав А. Крім, — завжди можуть бути осяжні та з'ясовані набагато легше, якщо вони представлені у вигляді креслень, ніж коли вони виражені числами, тому що в першому випадку прокидається сила уяви. Особливо це спостерігається при числах, що складаються з кількох цифр, наприклад, при порівнянні величин держави». Тому можна вважати, що А. Кроме започаткував застосування картограм і площинних діаграм порівняння для зображення статистичних даних. Незважаючи на те, що графічні зображення А. Кроме відрізнялися примітивністю форми та окремими недоліками, це вважається першим історичним досвідом застосування графічного зображення статистичних даних.

У. Плейфер видав у 1786 р. «Торговий і політичний атлас», який зображує за допомогою гравірованих по міді кольорових графіків процес торгівлі, державних доходів, витрат і боргів Англії протягом усього VIII ст. У роботі статистичні дані зображені як прямолінійних координатних графіків, і навіть кругових, секторальних, стовпчикових діаграм. У. Плейфером було опубліковано ще низку робіт, у яких широко застосовується графічний метод. Серед них найбільш відомі: «Лінійна арифметика», «Кратний виклад статистики», «Дослідження про постійні причини занепаду та загибель могутніх та багатих націй».

У графічних зображеннях У. Плейфер вперше застосував прямокутну систему координат, масштабні шкали. У цьому графіки У. Плейфейра виконані настільки майстерно, що мало чим відрізняються від кращих сучасних зразків. Якщо роботах А. Кроме графічні зображення були представлені в елементарної формі, то у роботах У. Плейфера вони виступають як самостійний метод вивчення статистичних даних. Сам У. Плейфейр так висловлювався від значення і переваги використовуваних ним способів графічного зображення статистичних даних: «Перевага запропонованого методу, — писав він, — полягає не в тому, що він дає більш точне уявлення про предмет, ніж цифри, але у цьому, що у простішому вигляді висловлює думку про поступовий прогрес і порівняльних величинах у різні періоди, представляючи нашому погляду постать, пропорції якої відповідають зображуваним даним».

Із середини XIX ст. графічний метод почав широко обговорюватись серед учених. Вперше питання щодо застосування графічних методів для спеціальних потреб статистики було поставлено у доповіді Організаційної комісії третього конгресу (Відень, 1857)[1].

Початком наукової розробки теорії статистичних графіків вважається серпень 1872 р., коли на восьмому Міжнародному статистичному конгресі в Петербурзі вперше в історії була поставлена і обговорена доповідь про теорію графічних зображень. На конгресі одночасно було представлено та обговорено дві доповіді. Доповідь про теорію графічних зображень була підготовлена Германом Швабе, директором Статистичного бюро (м. Берлін). Одночасно була представлена доповідь Адольфа Фіккера (м. Відень) про методи картографії стосовно завдань статистики. Доповідям на восьмому конгресі передувало обговорення питання про графічний метод на третьому з'їзді конгресу (Відень, 1857) і особливо на сьомому (Гаага, 1869).

Не зайве відзначити, що графічний метод у статистиці не відразу отримав загальне визнання та набув численних прихильників. На початку його цінність і корисність ставилася під сумнів як вченими, так і практиками. «Спочатку цей спосіб зображення статистичних висновків зустрічав численні заперечення як зі сторони окремих представників науки, яким здавався недостатньо «науковим», так і з боку статистичних практиків, для яких він був недостатньо практичним»[2]. Скептичне, а часом і негативне ставлення до графічного методу в статистичній літературі панувало до середини XIX ст.

Широке використання графічного методу в статистиці розпочалося на початку XIX ст., коли майже у всіх країнах почали виникати та розвивати свою діяльність органи державної статистики та наукові статистичні товариства. Яскраве підтвердження цього — поява відомих робіт відомого французького статистика А. Геррі і всесвітньо відомого бельгійського статистика А. Кетле, в основу яких були покладені докладні статистичні дані про кримінальні злочини, що регулярно публікувалися на той час у Франції та Бельгії[3]. Оцінюючи зна-

---

[1] «Emploi de la cartographie et de la methode graphigue en general pour les besoins speclaux de la statictigue». Див.: Compte rendu Венского конгресса, стр. 192–197.

[2] *Майр Г.* Вказ. робота, с. 141.

[3] Мова йде про роботу А. Геррі «Досвід моральної статистики Франції», опублікованої у 1833 р., та знаменитій книзі А. Кетле «Соціальна фізика або досвід розвитку людських здібностей», виданій у 1835 р.

чення книги А. Геррі для розвитку та застосування графічного методу у статистиці, Ю. Янсон писав: «Карти та графічні зображення, прикладені до тексту, становлять найістотнішу частину, в якій текст є поясненням; вони також мають історично-науковий інтерес, як методи викладу результатів статистичного дослідження які передували з того що зроблено у цьому плані Кетле».

Графічні методи у статистиці першої половини XIX ст., за словами відомого фахівця у сфері графічного методу Л. Бізова, більше розповсюджувалися, ніж поглиблювалися, а прийоми, запропоновані У. Плейфейромі А. Кроме, засвоювалися надто повільно і тільки у другій половині XIX ст. настає епоха «ентузіазму» щодо статистичних графіків.

Таким чином, графічний метод давно і міцно увійшов до арсеналу засобів математичної статистики. Якщо говорити про графічний метод у статистиці в сучасному розумінні, то основи були закладені в роботах А. Кроме та У. Пейфейра. Послідовники графічного методу продовжували його розвивати та удосконалювати, що й робилося надалі. В даний час графічний метод отримав широке поширення для вивчення явищ у соціальній, економічній, правовій та інших сферах суспільного життя.

## 5.2. Складові елементи статистичних графіків

В даний час графічне зображення статистичних у статистиці здійснюється за допомогою різних графічних образів — точок, ліній, площин, фігур, їх поєднань та різного розташування. Числові значення статистичних показників перетворюються на графічні образи з допомогою масштабу. Для прочитання графіка необхідно усвідомити його конструкцію, соціально-економічне значення досліджуваного явища чи процесу, спосіб його зображення на графіці.

Як і статистичні таблиці, кожен графік повинен мати з а г о л о -

в о к — назву, що розкриває зміст зображуваного явища, час і місце наведених показників, а також обов'язкове розшифровку умовних позначень (експлікація графіка). Якщо розглядати статистичний графік як площинне зображення, то в ньому можна виділити наступні елементи:1. П о л е г р а ф і к а — простір розміщення знаків, який має певні розміри та пропорції сторін. Розмір поля залежить від призначення графіка. Щоб читач міг зрозуміти графік, даються пояснення застосовуваних знаків, масштаб (наприклад, 1 см деякого знака дорівнює 1 млн грн продукції) та наводяться найменування графіка та його окремих частин.

2. Ш к а л а г р а ф і к а — сукупність ліній, позначок та проставлених у деяких із них чисел або інших символів, що відповідають ряду послідовних значень зображених величин. Якщо рівними відрізками на шкалі відповідають рівні числові інтервали, шкала називається *рівномірною*, якщо нерівні — шкала називається *нерівномірною* (*функціональною*). З нерівномірних часто застосовується логарифмічна шкала. Носіями шкал можуть бути прямі та криві лінії. З криволінійних найпоширенішими є кругові та дугові діаграми. Найчастіше для побудови статистичних графіків у статистиці використовують *подвійну* шкалу, яка являє собою дві системи послідовних числових значень, відповідних явищ або процесів, що зображуються на графіках. Ці шкали, як правило, з різним масштабом, мають у своєму розпорядженні поруч або з двох сторін графіків.

3. Г е о м е т р и ч н і з н а к и — символи понять, що відображаються на графіку; ці символи різноманітні: точки, відрізки прямих, кола, сектори, геометричні постаті, силуети та інших. Графічне вираз статистичних величин характеризує мову графіка. Залежно від геометричних образів графіки поділяються на точкові, коли в якості геометричних образів застосовуються сукупність точок; лінійні — коли застосовуються лінії; стовпчикові, смугові і т.д.

4. П р о с т о р о в і о р і є н т и р и, що визначають розміщення в полі графіка. Ці орієнтири залежить від прийнятої системи координат. У більшості графічних побудов застосовується прямокутна система координат. Іноді використовують полярні координати, сферичні (контурні карти).

5. М а с ш т а б н і о р і є н т и р и — еталони знака, що відображають величину геометричних знаків. Вони зображуються як кіл, прямокутників, квадратів і зазвичай виносяться з поля графіка. Величину явища можна визначити лише порівнянням графічного знака із

зразком. У статистичних графіках, що найчастіше зустрічаються у практичній діяльності правоохоронних органів, застосовуються прямолінійні (рідше криволінійні, наприклад, кругові) масштабні шкали (позначки), які часто розташовуються по осях координат.

6. Експлікація графіка — словесне пояснення змісту графіка та кожного його геометричного знака.

При складанні графіків керуються такими практичними вказівками: по можливості всі показники надаються в лінійних величинах, площі та обсяги, як правило, малонаглядні; вгорі на шкалі ординат поміщають конкретні позначення для кожного значення абсциси; вертикальну шкалу треба вибирати так, щоб нульова лінія була видна на діаграмі; якщо неможливо дати таку шкалу, треба дати у діаграмі горизонтальний розрив[1].

Перед побудовою графіка необхідно зі статистичного матеріалу вибрати статистичні дані, визначити його елементи.

## 5.3. Класифікація графіків

З огляду на наявність великої кількості різноманітних графічних зображень виникла необхідність проведення їх класифікації. Класифікація є систематичне розподіл явищ та об'єктів на певні групи, класи, розряди на основі їх подібності та відмінності. Підставою для класифікації, як правило, є якісна ознака. Однією з основних ознак є наявність полів графіків. Полем графіки може бути або звичайний чистий папір, або географічна карта, або контурна карта. З цього погляду графіки поділяються на *діаграми* та *статистичні карти*.

Графічне зображення статистичних даних, що наочно показує співвідношення між порівнюваними величинами називається *діаграмою*.

За завданням зображення діаграми поділяються на діаграми:

---

[1] Наочніший виклад прийомів складання графіків див. в кн. *Чекотовський Е. В.* Графічний метод у статистиці: історія і теорія. Частина II. Розвиток і застосування графічного методу в статистиці. *Статистика України*. 2009. № 2. С. 83–90. *Чекотовський Е. В.* Графічний метод у статистиці: історія і теорія. Частина II. Розвиток і застосування графічного методу в статистиці. *Статистика України*. 2009. № 2. С. 83–90.

а) порівняння; б) структурні; в) динамічні; г) лінійні; д) кругові балансові; д) потокові; е) виконання плану; ж) графіки зв'язку (див. гл. 9) та ін.

*Статистичні карти* є видом графічних зображень статистичних даних на схематичній географічній карті, тобто показують розміщення явища за територіями. У свою чергу, статистичні карти можна поділити на картограми, картодіаграми та центрограми. Відмінність між картограмами і картодіаграмами полягає у способах зображення в залежності від величини статистичної ознаки. Особливим видом картограм є центрограми, які є зображення міжтериторіальних переміщень населення чи господарських явищ, перетворення галузей економіки за історичними етапами.

*Контрольно-планові графіки* застосовуються на підприємствах з метою оперативного контролю над ходом виконання планових завдань. *Розподільні графіки* зображають зміни варіаційних рядів у формі полігону (прямих ліній, що утворюють багатокутник), гістограми (розміри ознаки виражаються у вигляді стовпчиків), асиметричних, нормальних, ексцесивних двоверхових і біноміальних кривих, кумулянти (послідовного підсумовування даних частот). *Спеціальні графіки* відображають характерну рису взаємозв'язку явищ. До них відносять графік кореляційного поля, криву Лоренца, номограми та ін.

При більш докладному викладі графіки кожна група може бути поділена більш окремі підгрупи. Найбільшого поширення в практиці статистичних досліджень отримали діаграми стовпчикові, стрічкові, лінійні, кругові, квадратні, секторні, виконання плану, діаграми за способом фігур-знаків та радіальні. Тут зупинимося лише на основних, широко застосовуваних видах.

## 5.4. Лінійні діаграми

*Лінійні діаграми* найчастіше використовуються для характеристики зміни явищ у часі. Для побудови статистичної кривої використовують систему прямокутних координат По осі абсцис відкладають годину, а по осі ординат — абсолютний чи відносний розмір явища за цю годину. Геометричними знаками в таких графіках є точки і послідовно з'єднують їх відрізки прямих ліній. Їх перевага полягає в тому, що динаміка розвитку явищ зображена на графіку у вигляді безперервної лінії, яка характеризує хід процесу, розвиток явища у годині. За допомогою лінійної діаграми можна більш наглядно та виразно охарактеризувати динаміку розвитку явища. Проте порівняння динаміки трьох та більш явищ є досить громіздким.

Статистичні криві, що відображають ряди динаміки, розрізняються за побудовою системи координатних сіток, які визначають характер ліній, характер динаміки. Відмінність полягає в масштабах побудови вертикальних і горизонтальних шкал. Це призводить до зміни конфігурації кривих.

У статистичній практиці найчастіше для графічного зображення рядів динаміки застосовується зображення з рівномірними шкалами. По осі часу (абсцис) відрізки беруться пропорційно до числа періодів або дат. По осі рівнів (ординат) шкали пропорційні або самим рівням, або їх логарифмам.

Лінійні графіки використовують також для дослідження взаємозв'язку між двома явищами або двома ознаками одного й того ж явища. В цьому випадку по осі абсцис відкладають розмір ознаки, що впливає, а по осі ординат — абсолютній або відносний розмір ознаки, на який здійснюється вплив. Така, наприклад, діаграма, що показує зниження витрат робочого часу на переробку 1000 ц соняшнику на олійноекстракційних заводах залежно від збільшення розміру добового випуску (рис. 5.1).

В прямокутній системі координат нанесемо на вісь абсцис показники добової продуктивності заводу, а на вісь ординат — дані про затрати праці на переробку 1000 ц соняшнику (в людино-годинах). Масштаб 1 см — 4 люд.-год. Статистична крива має порівно від'ємний нахил, що характеризується зміною затрат праці в залежності від добової потужності переробки соняшнику. Як видно наглядно з графіка, при збільшенні добової потужності олійних заводів затрати праці на переробку соняшнику знижуються.


*Рис. 5.1.* Залежність затрат робочого часу на переробку 1000 ц соняшнику від розмірів фактичного добового випуску на олійних заводах.

До роду лінійних (одномірних) діаграм відносяться статистичні ламані та криві. Власне статистичні криві будуть розглянуті нижче, тут зупинимося на їхньому різновиді — *статистичних ламаних*.

Побудуємо статистичну ламану валового збору соняшнику в Україні за 2014–2020 роки. Дані про виробництво такі:

Таблиця 5.1

| Показник | 2014 р. | 2015 р. | 2016 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|---|---|---|
| Валовий збір соняшнику, тис. ц | 9 995,2 | 11 087,1 | 13 190,6 | 11 937,8 | 13 882,7 | 14 923,4 | 13 135,8 |

На осі абсцис відкладаємо роки, на осі ординат — валовий збір соняшнику. Вісь абсцис ділимо на шість рівних відрізків, що відповідають шести рокам. Всю довжину осі ординат приймемо за 15 000 тис. т; наносимо на шкалу відповідні числа масштабів. З середини відкладених на осі абсцис відрізків відновлюємо перпендикуляри заввишки, пропорційною (за масштабом) валовому збору за кожний рік. Вершини перпендикулярів з'єднуємо відрізками прямої; отримана лінія і є шуканою статистичною ламаною (рис. 5.2).

На рис. 5.2 показані коливання обсягу виробництва соняшнику в Україні за 2014–2020 гг. Статистична крива має порівно великий нахил, характеризується зміною динаміки виробництва. Після зростання у 2015–2016 рр. у 2017 рр. спостерігалася тенденція до скорочення обсягів виробництва. Починаючи з 2018 р. обсяг виробництва цукру зростав, досягнувши у 2019 р. найбільшої величини у 14923,4 тис. т.

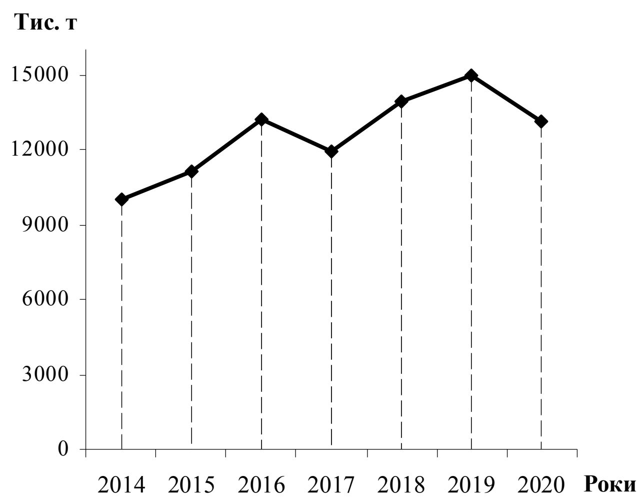



*Рис. 5.2*. Обсяг виробництва (валовий збір) соняшника в Україні за 2014–2020 роки (тис. т).

Істотною перевагою ламаних є те, що вони відкривають можливість нанесення на один графік (координатну сітку) даних про декілька явищ. Так, поряд із діаграмою про валові збори в одній координатній сітці можна побудувати діаграми посівних площ та урожайності соняшнику в Україні за ті ж роки. Дані про посівні площі та урожайність за 2014–2020 рр. наступні:

Таблиця 5.2

| Показник | 2014 р. | 2015 р. | 2016 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|---|---|---|
| Посівні площі соняшнику, тис. га | 4 987,1 | 4 962,0 | 5 756,8 | 5 779,6 | 5 923,4 | 5 759,9 | 6 381,3 |
| Урожайність соняшнику, ц с 1 га | 20,0 | 22,3 | 22,9 | 20,7 | 23,4 | 25,9 | 20,6 |

На основі цих первинних даних побудуємо графік, що характеризує зміну посівних площ, урожайності та валового збору соняшнику в Україні (рис. 5.3). Як бачимо, графік трьох взаємопов'язаних показників дає можливість не лише наочно уявити, а й узагальнити статистичні дані. За даними графіку, зокрема, можна зробити узагальнений висновок про те, у які роки відбувалося зростання досліджуваних показників, а у які — зниження.

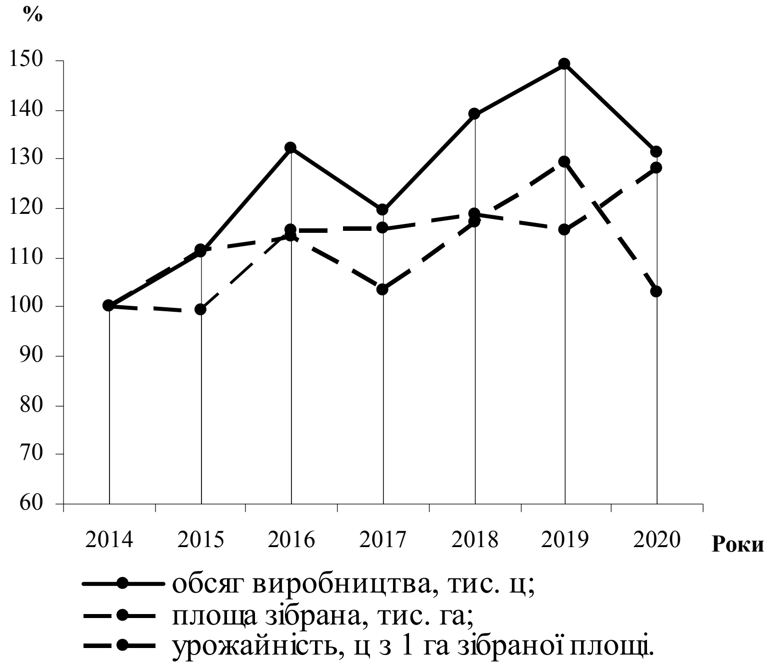



*Рис. 5.3.* Валовий збір, посівні площі та врожайність соняшнику в Україні у 2014–2020 рр.

Крім характеристики рядів динаміки лінійні графіки використовують  і для наочного зображення рядів розподілу і залежності однієї варійованої ознаки від іншого. Їхня перевага полягає в тому, що динаміка зображується у вигляді безперервної лінії, що характеризує безперервний хід процесу, розвиток явища в часі та хід виконання планів.

## 5.5. Діаграми порівняння

*Діаграма* — графічне зображення, що дає уявлення про абсолютні та відносні розміри соціально-економічних явищ та їх статистичні характеристики. За допомогою діаграм можна: 1) показати зміну явищ у часі (їх динаміку); 2) з’ясувати структуру досліджуваних явищ і структурні зрушення, які відбуваються у них; 3) порівняти статистичні величини між собою; 4) вивчити залежність між явищами чи окремими факторами одного явища; 5) охарактеризувати розподіл будь-якого показника. Графіки даного типу бувають стовпчикові та лінійні, розміри величин яких відповідають зображуваним ними цифрам. Найбільшого поширення набули діаграми стовпчикові, лінійні, кругові, квадратні, секторні, діаграми за способом фігур-знаків та радіальні.

*Діаграми порівняння* — групи діаграм, що застосовуються для зіставлення величин. Такого роду діаграми графічно показують співвідношення різних статистичних сукупностей за якоюсь ознакою, що змінюється в динаміці. Особливістю побудови діаграм порівняння є одномірне графічне зображення ознаки, що змінюється, в одному масштабі для всіх значень цього ознаки.

Для порівняння можна використовувати діаграми стовпчикові та смугові. На *стовпчикових діаграмах* статистичні показники зображуються у формі прямокутників — стовпчиків, рівних за величиною основи і розміщених вертикально поруч або на однаковій відстані один від одного. Основа стовпчиків розміщується на горизонтальній лінії, а висота цих стовпчиків відповідно до прийнятого масштабу пропорційна зображуваним величинам.

Побудуємо, наприклад, діаграму території окремих країн світу (табл. 3.3). При побудові стовпчикових діаграм використовують прямокутну систему координат. По осі абсцис розташовуються окремі країни світу, по осі ординат — площа країн, тис. км$^2$. Прийнявши нижній та лівий краї координатної осі за абсцис та ординат, на них, як на шкали, наносимо відповідні масштаби.

З цією метою всю довжину осі ординат прирівнюємо до максимального показника, який необхідно відобразити на графіку (774,8 тис. км$^2$); округляючи, беремо цю довжину за 800 тис. км$^2$, після чого наносимо шкали — 0, 100, 200, 300, 400, 500, 600, 700, 800. На осі абсцис розташовуються п’ять рівних підстав стовпчиків з відповідними повинні перевищувати ширини стовпчиків). Ширина стовпчиків — 1 см. Масштаб на вертикальній осі — 200 тис. км на 1 см. Після нанесення масштабів на кожній із підстав будуємо стовпчики високої, пропорційно відповідній території (рис. 5.4).

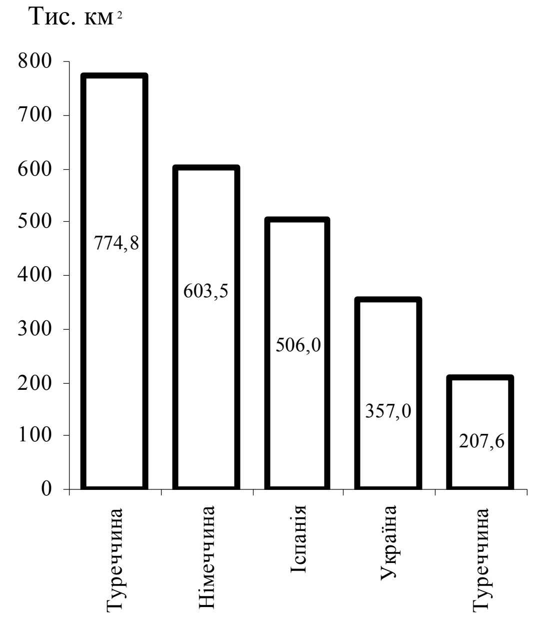



*Рис. 5.4.* Територія окремих країн світу (стовпчикова діаграма з розімкненими) стовпчиками).

На стовпчиковій діаграмі відображається єдина міра — площа. Висота стовпчиків залежить від величини статистичної ознаки. Стовпчики діаграм можуть бути зімкнутими (без інтервалів), їх можна заштрихувати або зафарбувати фарбою одного або кількох кольорів. Вони дають можливість порівняти статистичні величини між собою, а також виявляти структуру та структурні зрушення та показати зміну явищ у часі.

Різновидом стовпчикової діаграми є *стрічкова* (*смугова*) *діаграма*, на якій величини зображуються у вигляді смуг однакової ширини, що розташовуються горизонтально. На стрічковій діаграми координати осі повернені на 90°. При цьому основа смуг (об'єкти) розташовуються на осі $y$, а масштаб — на осі $x$. Внаслідок повороту осей у таких діаграмах стовпчики перетворюються на своєрідні горизонтальні смуги або стрічки, а базова лінія розташовується вертикально (рис. 8.3).

Стрічкові діаграми дають можливість вирішувати ті ж завдання, що й стовпчикові, але в них стовпчики розташовані горизонтально. При зображенні статистичних даних як стовпчикової діаграми часто цифри показників пишуть безпосередньо над стовпчиками. Однак цей спосіб застосовувати не рекомендується, так як

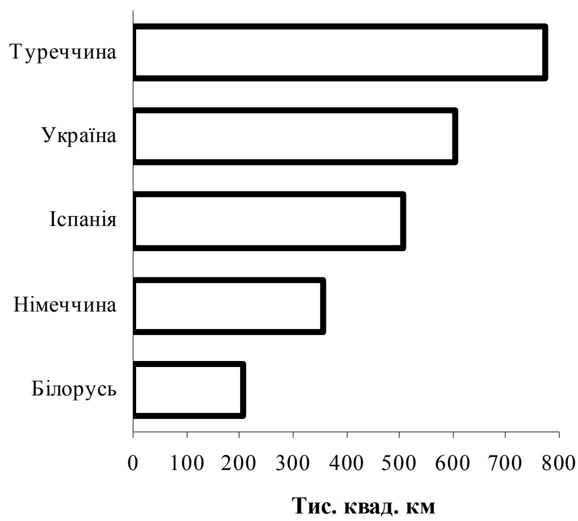



*Рис. 5.5* Територія окремих країн світу (стрічкова діаграма з розімкненими смугами).

цифри на кінці стовпчиків як би подовжують їх, що створює помилкове зорове враження про співвідношення зображуваних показників. Більш доцільно цифри показників можна написати всередині стовпчиків або помістити в один ряд над стовпчиком на рівні закінчення шкали по осі ординат.

При побудові стовпчикових діаграм необхідно виконувати такі умови: 1) шкала, за якою встановлюється висота стовпчика, має починатися з нуля; 2) шкала не повинна перериватися; 3) довжина стовпчиків повинна бути рівною між собою; стовпчики можуть бути розміщені на однаковій відстані один від одного, впритул один до одного або напливом, при якому один стовпчик частково накладається на інший; 4) поряд з розміткою шкали відповідними цифровими написами слід постачати і стовпці.

Початок смуг може бути на одній і тій же вертикальній лінії. Смуги можна розташувати і по обидва боки центральної вертикальної лінії. Так зображують *вікові піраміди* у демографічній статистиці. Вони використовуються для графічного зображення розподілу населення за статтю і за віком.

Кількість населення за віком зображується у формі вікової піраміди, де довжина кожної смуги означає чисельність населення певного віку, причому чисельність чоловіків відхиляється в один бік, а чисельність жінок — в іншу (див. рис. 5.6).

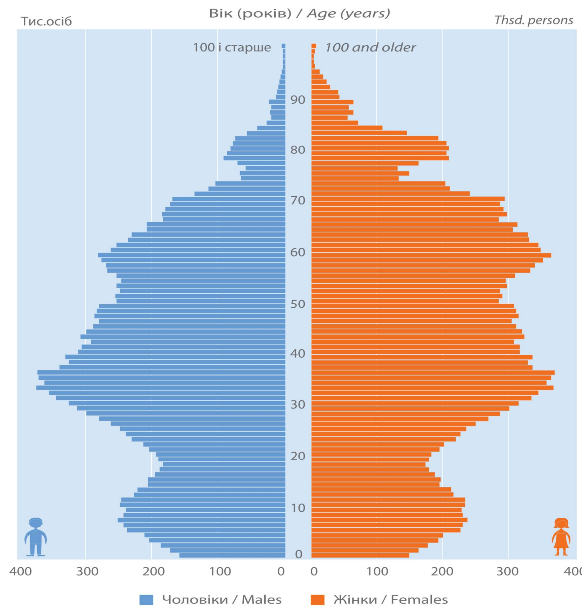



*Рис. 5.6.* Розподіл постійного населення України за статтю та віком на 1 січня 2020 р.

*Джерело*: Державна служба статистики України. URL: http://www.ukrstat.gov.ua/

Стовпчикові та смугові діаграми взаємозамінні, тому що в обох випадках використовується один вимір — висота стовпчика або довжина полоси. При побудові діаграм висоти стовпчиків або довжини полос розташовують у спадному чи зростаючому порядку.

## 5.6. Структурні діаграми

*Структурна діаграма* показує склад (структуру) цілого, розділеного на частини. Відповідно до назви цього виду графіків їх призначення полягає у показі питомої ваги та співвідношень складових частин явищ. До структурних діаграм відносяться насамперед діаграми питомих ваг, що графічно зображають відношення окремих частин сукупності до її загального обсягу. Образотворними засобами для цього можуть служити *діаграма стовпчикова* (висота стовпчика приймається за 100 %), *діаграма секторна*, *діаграма трикутна* та ін. Зображувані частини явищ як у стовпчикових, так і в секторах повинні мати різне забарвлення або штрихування і розташовуватися у певній послідовності: у секторних діаграмах — по руху годинникової стрілки, у стовпчикових — від верху до низу.

*Стовпчикова діаграма* є різновидом діаграми, що зображує статистичні величини у формі прямокутників — стовпчиків, рівних за величиною основи і розміщених вертикально поруч або на однаковій відстані. При цьому загальна висота стовпчика приймається за 100 %, яке частини виражають частки груп чи підрозділів цілого.

Прикладом стовпчикової структурної діаграми може бути рис. 5.7.

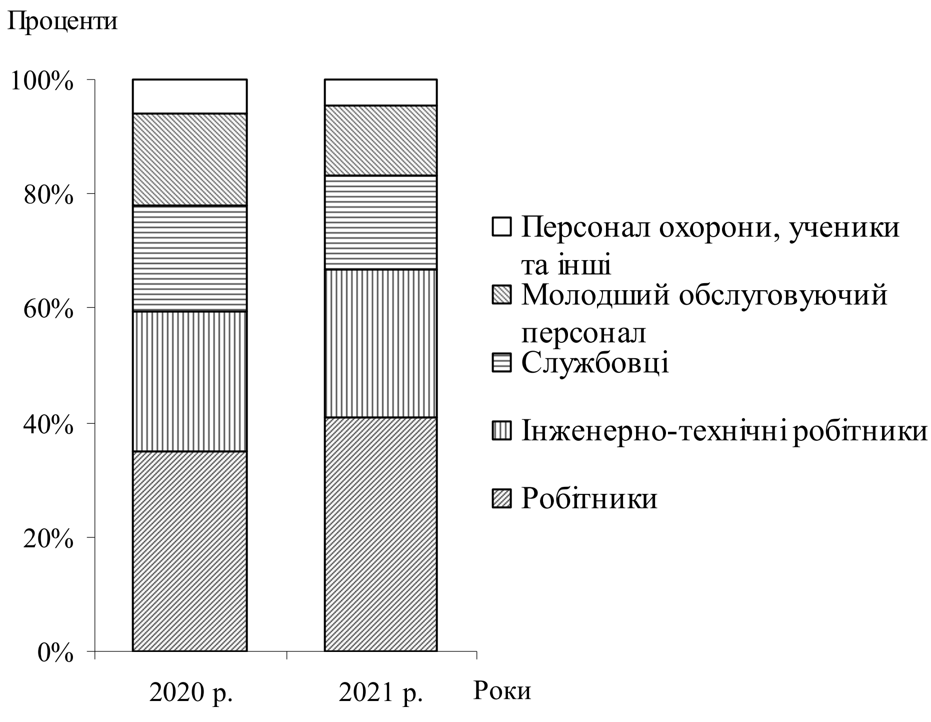



*Рис. 5.7.* Структура персоналу підприємства в 2020 та в 2022 рр. (%).

Стовпчикові діаграми дозволяють проводити наочне порівняння матеріалів статистичного спостереження. В окремих випадках розмір стовпчиків може відповідати абсолютним значенням всього явища.

*Секторна діаграма* є різновид структурної діаграми, зіставлення елементів цілого здійснюється з допомогою площ, утворених секторами кола. Кругова секторальна діаграма будується шляхом поділу кола на сектори пропорційно до питомої ваги частин у цілому. При побудові секторної діаграми необхідно пам'ятати, що 1 % відповідає 3,6°.

Маються дані про склад посівних площ у сільськогосподарських підприємствах та фермерських господарствах (табл. 5.3).

Щодо висхідних умов нашої задачі діаграма буде складатися з двох кіл — посівна площа сільськогосподарських підприємств та фермерських господарств, яка приймається за 100%. Для виділення в кожному колі чотирьох секторів (за кількістю культур) визначаємо за даними про питому вагу посівних площ, зайнятих під окремими культурами, відповідні значення центральних кутів (1%=3,6°):

Таблиця 5.3

**Структура посівних площ України (всі категорії господарств, %)**

| Культура | 2010 р. | 2020 р. |
|---|---|---|
| Зернові та зернобобові | 57,0 | 65,0 |
| Технічні | 11,2 | 13,4 |
| Картопля, овочеві та баштанні продовольчі | 1,7 | 3,7 |
| Кормові | 30,1 | 17,9 |
| Вся посівна площа | 100,0 | 100,0 |

Посівна площа зернових культур:

в сільськогосподарських підприємствах

$$\frac{57,0}{100} \cdot 360 = 205,1\ \%;$$

в фермерських господарствах

$$\frac{65,0}{100} \cdot 360 = 234,1\ \%.$$

Посівна площа технічних культур:

в сільськогосподарських підприємствах

$$\frac{11,2}{100} \cdot 360 = 40,2\ \%;$$

в фермерських господарствах

$$\frac{13,4}{100} \cdot 360 = 48,3\ \%.$$

Посівна площа картоплі та овоче-баштанних культур:

в сільськогосподарських підприємствах

$$\frac{1{,}7}{100}\cdot 360 = 6{,}0\ \%;$$

в фермерських господарствах

$$\frac{3{,}7}{100}\cdot 360 = 13{,}3\ \%.$$

Посівна площа кормових культур:

в сільськогосподарських підприємствах

$$\frac{30{,}2}{100}\cdot 360 = 108{,}7\ \%;$$

в фермерських господарствах

$$\frac{17{,}9}{100}\cdot 360 = 64{,}4\ \%.$$

Використовуючи отримані дані, збудуємо секторні діаграми структури посівних площ (рис. 5.8 та 5.9).

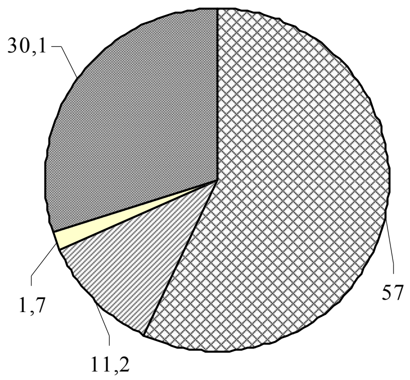



*Рис. 5.8*. Структура посівних площ у сільськогосподарських підприємства на 15 жовтня 2010 р. (в процентах).

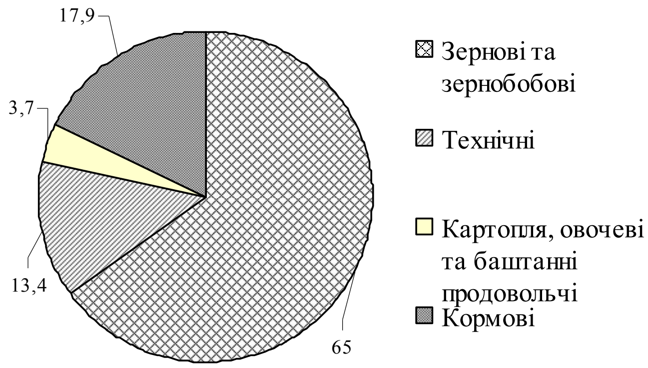
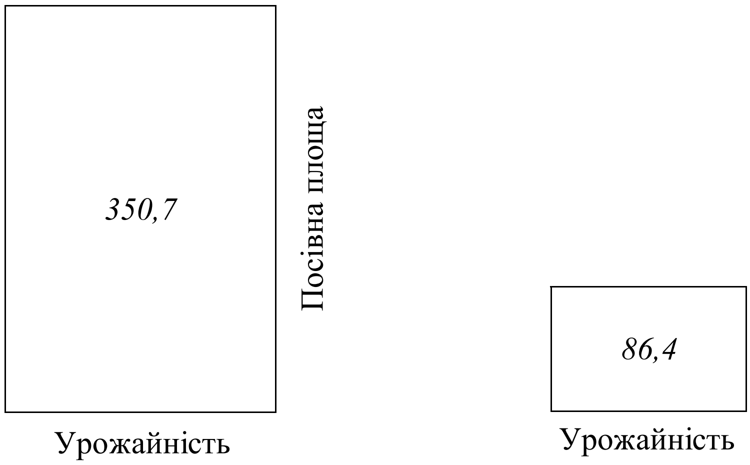



*Рис. 5.9*. Структура посівних площ у фермерських господарствах на 15 жовтня 2010 р. (в процентах).

За допомогою відносних показників виявляється не тільки структура досліджуваної сукупності, а й структурні зрушення.

Особливим різновидом площинної діаграми є *знаки Варзара*, запропоновані видним статистиком В. Є. Варзаром (1851–1940). З її допомогою можна зобразити одночасно три величини: одна зображується основою ($a$) прямокутника, інша ($h$) – його висотою, третя, що дорівнює їх добутку, — розміром отриманої площі ($S$). Графічний знак (або знак Варзара) будується у вигляді прямокутника, основа та висота якого визначаються за масштабом двома факторами-співмножниками, а площа — величиною результативного показника-добутку. Наприклад, $a$ — посівна площа, га; $h$ — урожайність, ц із 1 га; $S = ah$ — валовий збір, ц.

Приклад побудови графіка Варзара показано на рис. 5.10.




а) б)

*Рис. 5.10.* Валовий збір, урожайність і посівні площі в 2020 г.:
а) зернових культур; б) соняшнику.

Діаграма знака Варзара дає можливість порівнювати явища за періодами на території на географічній карті.

## 5.7. Контрольно-планові графіки

*Контрольно-планові графіки* — це графіки, що зображають хід виконання плану. Застосування цих графіків на підприємствах зі складною структурою виробничих процесів (наприклад, у машинобудуванні) суттєво полегшує контроль та планування виробництва. Об'єктами графічного контролю можуть бути: випуск продукції та виконання плану по заводу, цеху, ділянці та бригаді; терміни виконання будь-якої роботи; походження виробу чи процесу; стан заділів у виробництві; хід комплектування збирання тощо.

Виділяють два основні види цих графіків: а) лінійні графіки виконання плану; б) обліково-планові графіки.

Лінійні графіки є зручним способом контролю виконання планових завдань за одним виробом, або за показником. Як за формою, так і за назвою вони можуть бути різними (стовпчиковими, лінійними та ін.). Найпростішим з таких графіків можна представити у вигляді стовпчикової діаграми, де по осі абсцис розташовані періоди (або об'єкти), а по осі ординат — відсотки виконання плану. Приклад графіка виконання плану цехами машинобудівного заводу протягом тижня наведено на рис. 5.12.

Для наочності паралельно осі абсцис може бути проведена лінія стовідсоткового виконання плану. Такі графіки є громіздкими, тому що для кожного об'єкта, взятого за один інтервал часу, доводиться будувати окремий стовпчик. Тому цей графік на практиці не набув широкого поширення.

До контрольно-планових графіків першого типу належать також лінійні діаграми. Координати часу в таких графіках розташовуються здебільшого на осі абсцис, а рівномірність спостережуваного показника (відсоток виконання плану) відкладається в певному масштабі на осі ординат.

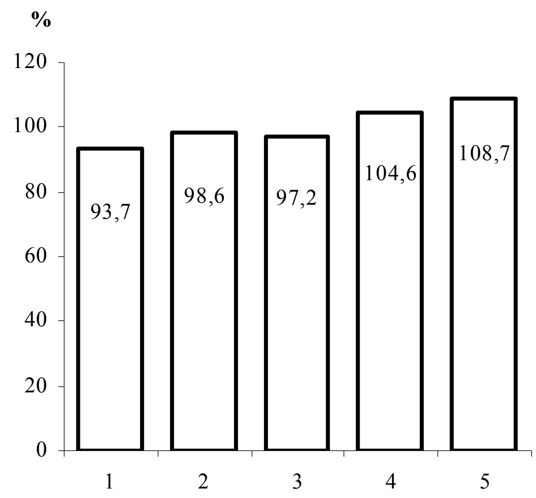



*Рис. 5.11.* Виконання плану цехами заводу: 1 — механічний цех; 2 — збиральний цех; 3 — ремонтний цех; 4 — токарний цех; 5 — термічний цех.

На рис. 5.12 представлений графік виконання плану бригади протягом місяця.

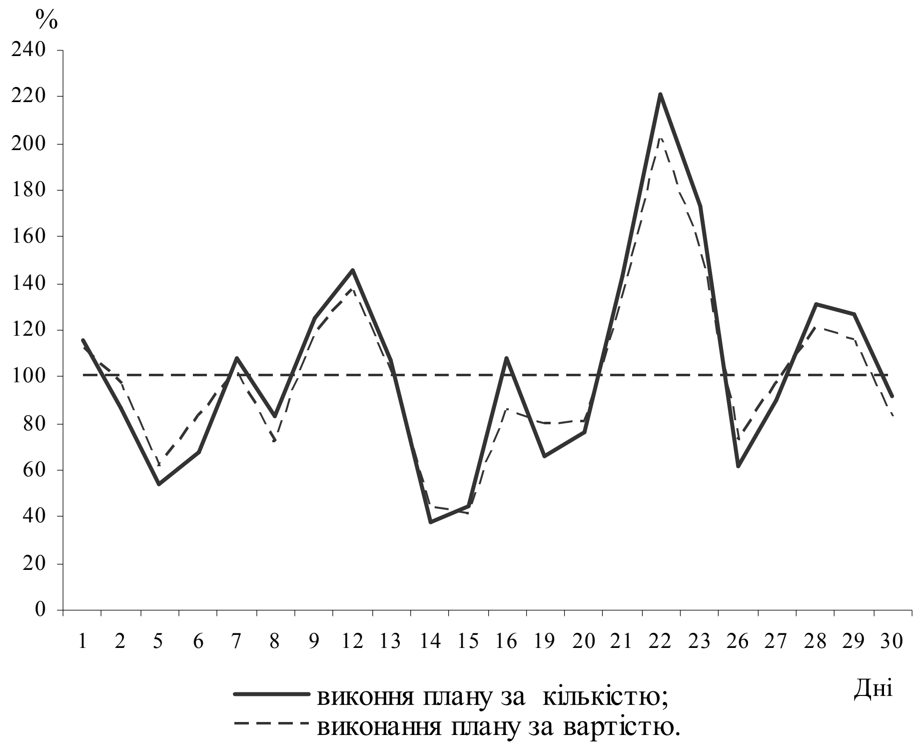



*Рис. 5.12.* Виконання плану бригади випуску продукції за місяць.

Цей графік є дуже поширеним, особливо для контролю за випуском продукції підприємства чи цеху. Він дає змогу оперативно аналізувати стан виконання плану підрозділами підприємства.

Більш зручні *обліково-контрольні графіки*, що застосовуються для поточного спостереження та контролю за ходом виробництва. Ці графіки є інформаційними у сенсі слова й при їх конструюванні необхідно досягнення двох умов: максимальної повноти інформації та максимальної її наочності.

Першим і найпростішим видом обліково-контрольних графіків є графіки лінійного типу. У них кожному об’єкту спостереження відведено один рядок, у якому графічно ведуться позначки. Зміст та характеристики позначок можуть бути різними залежно від об’єкта контролю (робочий, виріб, вузол або процес). У плановій практиці у такий спосіб ведуться багатономенклатурні графіки виконання плану або забезпеченості виробництва сировину, матеріалами, деталями тощо.

Наприклад, мається наступна інформація про виконання плану випуску продукції одним робітником за тиждень:

Таблиця 5.4

| Дні неділі | План на день, шт. | План з початку неділі, шт. | Виробіток за день, шт. | Виконання денного плану, % | Вироблено з початку неділі, шт. |
|---|---|---|---|---|---|
| Понеділок | 20 | 20 | 15 | 75 | 15 |
| Вівторок | 25 | 45 | 20 | 80 | 35 |
| Середа | 30 | 75 | 30 | 100 | 65 |
| Четвер | 30 | 105 | 36 | 120 | 101 |
| П’ятниця | 30 | 135 | 34 | 113 | 135 |

Зобразимо процес виконання плану цим робітникам на графіку (рис. 5.13).


*Рис. 5.13*. Виконання плану випуску продукції робітниками цеху з 15 по 20 квітня 2022 р.

Їх будують на спеціально розграфленій сітці, що має форму таблиці. По горизонталі відкладають одиниці часу (години, зміни, дні, тижні), по вертикалі розміщують об'єкти дослідження. Кожен відрізок по горизонталі відповідає 100 % виконання завдання. На графік нанесена календарна сітка робочих днів зі шкалою відсотків (20, 40, 60 та 80, причому 0 та 100 % не позначені, щоб не захаращувати графік). У верхньому лівому кутку смуги проставлені дані про план на день, у правому — з початку тижня. Ступінь виконання плану зображується двома лініями: тонкою лінією та жирною прямою. Відрізками тонкої прямої зображені відсотки виконання плану за кожен день; при перевиконанні плану тонка лінія продовжується зліва зверху (див. четвер, п'ятницю). Безперервної жирної прямої дається важлива інформація про виконання плану наростаючим підсумком з початку тижня. Ця лінія прокреслюється відразу по всій сітці, але на ній щоденно ставляться позначки про виконання плану. Методика відміток така.

За понеділок жирна лінія дорівнює тонкій, оскільки та й інша відображає однакову величину (15) в одному масштабі. З вироблених у вівторок 20 прим. на покриття невиконання плану за понеділок використано 5 прим. Отже, рахунок виконання плану залишається 15 прим. (20–5). Це відповідає виконанню плану у розмірі 60 % $\left(\frac{15 \cdot 100}{25}\right)$, що і показано позначкою на жирній лінії за вівторок. За середу робітник виробив 30 шт., але з них на покриття невиконання плану за понеділок та вівторок беруться 20 шт., решта 20 шт. становлять до плану середу 67 % $\left(\frac{20 \cdot 100}{30}\right)$, як і відображено на жирної лінії за середу. У четвер робітник виготовив 36 виробів, з яких 10 йде на покриття недоданого за попередні дні; 26 шт. становлять до плану за четвер 87 %. З вироблених у п'ятницю 34 вироби 4 шт. йдуть на покриття заборгованості, а решта 30 точно відповідають денному завданню; отже, сумова лінія за п'ятницю сягає 100 %. Таким чином, жирна лінія виступає своєрідною сумарною лінією, що відображає перевиконання або недовиконання плану за станом на кожен день.

Зображення однією графіку кількох об'єктів (робітників, цехів тощо) дозволяє порівнювати результати їх у цілому оцінювати рівномірність виконання плану.

## 5.8. Картограми

Розташування умовних знаків на географічній карті для характеристики територіального розміщення явищ носить назву *картограми*. На географічній карті штрихуванням різної густоти, точками або забарвленням різного ступеня насиченості показано порівняльну інтенсивність будь-якого показника в межах кожної одиниці нанесеного на карту територіального поділу. Такими показниками можуть бути, наприклад, щільність населення, середня тривалість життя громадян, народжуваність, розмір середнього душового доходу і т. д. Замість розмальовки та штрихування як графічні символи в картограмі можна використовувати точки, умовно визначаючи, яку величину ознаки означає одна крапки. Різновидом картограми є точкові та фонові картограми.

*Картограма точкова* — різновид картограми, де рівень явища показується за допомогою точок, розміщених на контурній карті будь-якої територіальної одиниці. Кожна точка відповідає певній числовій величині. Для наочності зображення на географічній карті щільності або частоти появи певної ознаки крапкою позначають одну одиницю сукупності або деяку кількість.

*Картограма фонова* — різновид картограми, що показує штрихуванням різної густоти або забарвленням різного ступеня інтенсивність будь-якого показника в межах територіальної одиниці. Зазвичай найменше значення показника зображується легкою затушовкою або штрихуванням або зовсім не затушовується.

Картограми застосовуються тоді, коли вирішується завдання графічного відображення географічного розподілу явища, що вивчається. І тут на карту наносяться знаки, що характеризують величину явища у різних містах, районах, областях. Як приклад наведемо фонову картограму захворювання на короновірус у Харківській області (рис. 5.14).

Найбільш неблагополучними щодо кількості захворюваності на короновірус серед районів Харківської області були Харківський і Чугуївський райони. У цих районах 16 серпня 2021 р. було зареєстровано відповідно 348 та 114 випадків захворювання на короновірус. Високий рівень захворюваності на короновірус був у м. Харкові (4 757). У багатьох областях, позначених світло-коричневим кольором, рівень захворюваності не перевищував 100 випадків. Найнижча кількість заражених спостерігалася у Близнюківському (1) та Барвін-

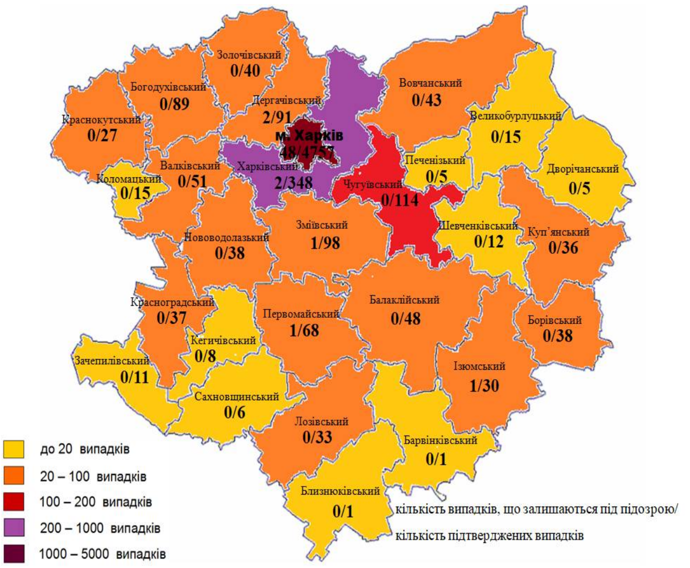



*Рис. 5.14*. Картограма числа випадків захворювання коронавірусом 16 серпня 2021 г. в Харківській області.

*Джерело*: Сайт Медіагрупи «Об'єктив». URL.: https://www.objectiv.tv/o.

ківському (1) районах.

Крім розглянутих графіків у статистиці є низку інших методів графічного зображення статистичного матеріалу, але де вони настільки поширені широкого застосування практичної діяльності не знаходять.

## 5.9. Картодіаграми

*Картодіаграма* — вид картограми, на якій за допомогою статистичного показника в межах кожної одиниці нанесеного на картодіаграму територіального поділу. Прикладом картодіаграми може служити кількість населення та його склад по областях, щільність земельних угідь, виробництво валової продукції промисловості та її зростання за певний період тощо. угідь тощо) всередині окремих одиниць. Приклад картодіаграми наведено на рис. 5.15.

На відміну від картограми, картодіаграми дають можливість наочно відобразити більш складні статистико-географічні побудови, ніж картограми. За допомогою фігур різної форми (стовпчиків, кіл, квадратів) зображують сумарну величину будь-якого явища в межах певної території. Слід враховувати, що картодіаграма не відображає дійсного розміщення явища (валової продукції, чисельності населення, структуру земельних угідь тощо) усередині окремих одиниць. Слід враховувати, що картодіаграма не відображає дійсного розміщення явища (валової продукції, чисельності населення, структуру земельних угідь тощо) усередині окремих одиниць.

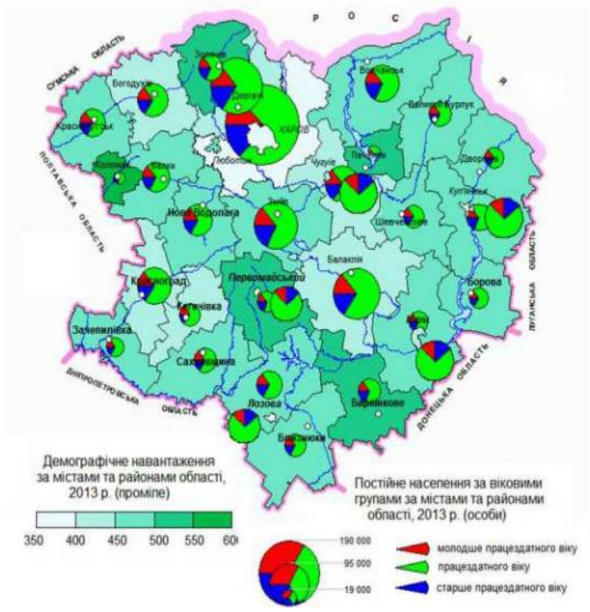



*Рис. 5.15.* Демографічне навантаження адміністративно-територіальних одиниць Харківської області.

*Джерело*: Стратегія розвитку Харківської області на період до 2020 року. Харків, 2021, с. 19.

## 5.10. Радіальні графіки

*Графік радіальний* — вид графіка, побудованого у полярних координатах. Використовується для зображення явищ, що періодично змінюються у часі (переважно сезонних коливань). Час відраховується за годинниковою стрілкою по колу, а значення показника відповідає відстань точки до центру. Радіальний графік у разі зростання величини показника вийде у вигляді спіралі. Якщо до початку наступного циклу показник повертається до вихідної величини, то радіальний графік виявиться замкненим. Так, є вихідні дані про обсяг випущеної продукції підприємством харчової промисловості за місяцями (табл. 5.5).

Таблиця 5.5

**Кількість продукції, випущеної підприємством харчової промисловості за місяцями**

| Місяці | Випуск продукції, тис. т | Місяці | Випуск продукції, тис. т |
|---|---|---|---|
| I | 8,3 | VI | 10 |
| II | 13,8 | VIII | 7,9 |
| III | 14,5 | IX | 7,9 |
| IV | 13,9 | X | 9,7 |
| V | 11,8 | XI | 12 |

За даних табл. 5.5 побудуємо радіальний графік, який представлений на рис. 5.16.

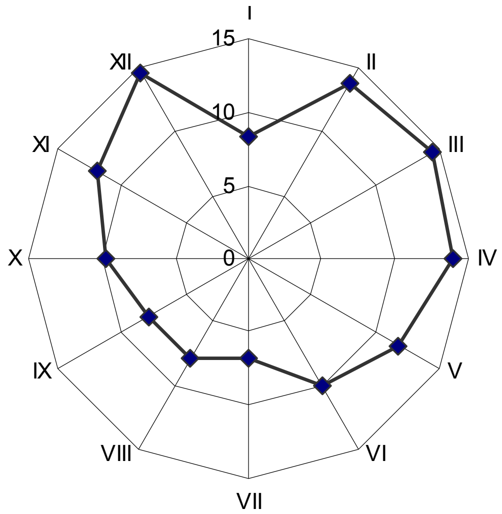



*Рис. 5.16.* Динаміка кількості виробленої продукції підприємством харчової промисловості за місяцями.

## 5.11. Діаграми за методом фігур-знаків

Особливим різновидом стрічкових діаграм виступають *фігурні діаграми*, у яких явища зображуються у вигляді рисунків (фігуризнаки), відповідних матеріального змісту явищ. Як графіки порівняння фігурні діаграми (образотворчі графіки) будуються з використанням спрощених предметних зображень описуваних явищ та процесів. Ці діаграми поєднують переваги стовпчикових та стрічкових діаграм (діаграми порівнюються за одним параметром) та переваги образотворчих символів перед геометричними фігурами. Наприклад, виробництво електроенергії зображується у вигляді рисунків ліній електропередач, вантажообіг залізничного транспорту — вагонів, виробництво молока — бідонів. Ці символи надають фігурним діаграм велику виразність, завдяки чому вони широко використовуються в статистиці.

Як графіки порівняння вони являють собою узагальнені зображення об’єктів вимірювання. Фігурні (образотворчі) діаграми можуть будуватися різними способами: 1) розміри рисунків беруться пропорційними зображуваним величинам, що більше відповідна фігура, то більше вписувалося величина цього об’єкта (рис. 5.17); 2) символи однакового розміру або геометричні фігури представляють певну величину, а кількість цих фігур виражає розмір явищ на різних ділянках або за різні періоди часу (рис. 5.18); 3) графічні елементи (криві, стовпчики і т. п.) можуть супроводжуватися рисунками.

Діаграми за методом фігур-знаків характерні тим, що фігури вказують, яке економічне явище розглядається, а кількість чи величина цих фігур виражає розмір явищ. Наприклад, чисельність корів у сільськогосподарських підприємствах області на кінець року може бути зображено у формі контуру корови різної величини (рис. 5.17).


*Рис. 5.17.* Поголів’я корів за 1990–2020 рр.

Порівнюючи обсяг будівництва високовольтних ліній електропередач за певний період, зображують контур високовольтної лінії. Відповідно до кількості високовольтних ліній, зображення можуть супроводжуватися цифровими значеннями. Прикладом фігурної діаграми може бути рис. 5.18.

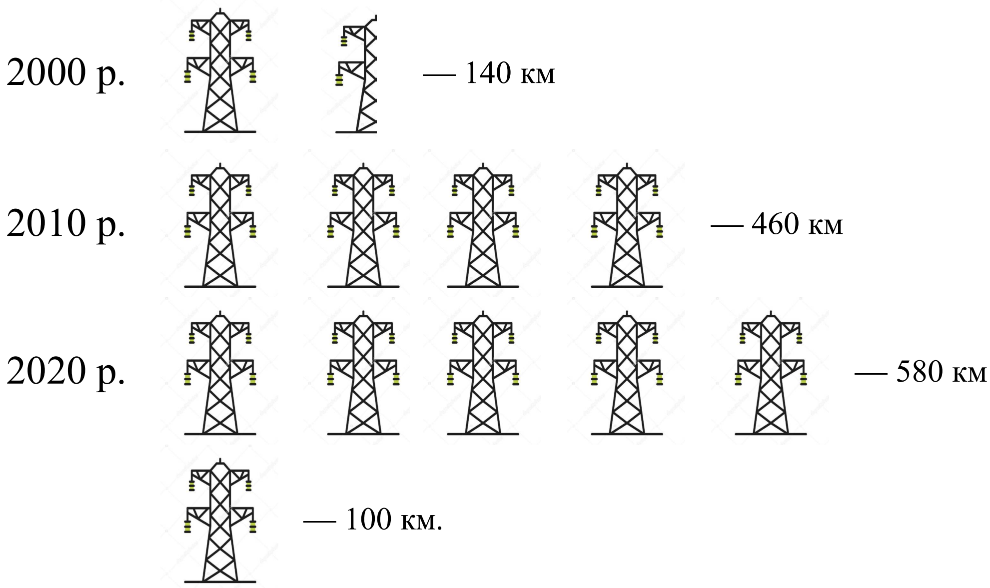



*Рис. 5.18.* Діаграма за методом фігур-знаків.

В окремих випадках обсяг будівництва високовольтних ліній електропередач може бути зображений у вигляді фігурних знаків ліній електропередач різної величини (рис. 5.19).

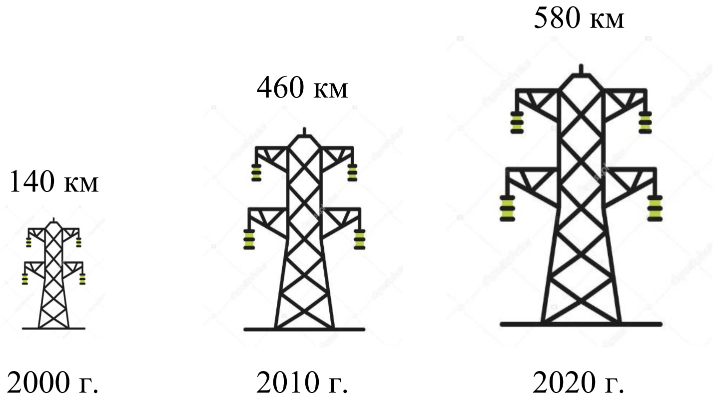



*Рис. 5.19*. Обсяг будівництва ліній електропередач за 2000–2020 рр., км.

Фігурні діаграми можуть характеризувати і діаграму явищ, коли використовуються дані за різні періоди або певні дати. Одним із недоліків фігурних діаграм є труднощі встановлення відмінностей у рівні.

## ЛІТЕРАТУРА


1. *Горкавий В. К*. Статистика: Підручник. Третє вид., переробл. і доповн. Київ: Алерта, 2019. С. 355–371.
2. *Логунова Н. А*., Статистика ІІ: Підручник Київ: Кондор-Видавництво, 2015. 340 с.


## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники*


3. *Клещева С.А.* Графический метод в экономических исследованиях: пособие для самостоят. работы. Пинск: ПолесГУ, 2010. 40 с.
4. *Опря А. Т.* Статистика (модульний варіант з програмованою формою контролю знань). Навч. посіб. Київ: Центр учбової літератури, 2012.
5. *Педченко Г. П.* Статистика: Навчальний посібник. Мелітополь: Колор Принт, 2018. С. 32–57.
6. *Шапочка М. К*., *Маценко О. М*. Теорія статистики: навч. посібник. Суми: Університетська книга, 2014. С. 84–100.


*Монографії та статті*


7. *Чекотовський Е. В.* Графічний метод у статистиці: історія і теорія. Частина І. Зародження і становлення графічного методу в статистиці. *Статистика України*. 2009. № 1. С. 93–99.
8. *Чекотовський Е. В.* Графічний метод у статистиці: історія і теорія. Частина ІІ. Розвиток і застосування графічного методу в статистиці. *Статистика України*. 2009. № 2. С. 83–90.

ГЛАВА 6

# АБСОЛЮТНІ І ВІДНОСНІ СТАТИСТИЧНІ ВЕЛИЧИНИ

## 6.1. Статистичні величини

Явища суспільного життя, досліджувані статистикою, як говорилося в гл. 1, мають кількісну визначеність. Конкретне кількісне вираження будь-яких статистичних показників (ознак) у статистиці отримало назву *статистичних величин*.

Як відомо, статистика вивчає кількісну сторону масових суспільних явищ і процесів у нерозривному зв'язку з їх якісною стороною. Кількісна сторона цих явищ виражається абсолютними величинами, що характеризують статистичну сукупність. *Абсолютний розмір явища* — це його величина, взята сама по собі, безвідносно до інших явищ. Вона безпосередньо визначає розміри суспільних явищ і процесом числом одиниць або в різних конкретних показниках. До таких величин можна зарахувати, наприклад, вартість основних виробничих засобів; чисельність працівників, зокрема і з окремих категорій персоналу; кількість забракованих виробів, стаж роботи, виробництво окремих видів продукції (штук, тонн, кілограм,) тощо.

Проте якісно самі по собі абсолютні величини не пізнаються, не розглядаються. У процесі вивчення їх з чимось порівнюють, зіставляють (бо все пізнається в порівнянні). В результаті отримують відносні величини. *Відносний розмір* — це співвідношення величини даного явища з величиною якогось іншого явища або з величиною того самого явища, але взятої за інший час або по іншій місцевості. Відносні статистичні величини дають зведену, загальну якісну характеристику рівня явищ або виражають зв'язки та співвідношення, а також динаміку явищ. Такі, наприклад, питома вага матеріальних витрат у структурі собівартості продукції; число народжених (померлих) за певний період (зазвичай за рік) у розрахунку на 1000 осіб, які проживають на даній території; співвідношення основних та допоміжних робітників; співвідношення зростання продуктивності праці та середньої заробітної плати та ін.

Відповідно до того, кількісна визначеність явищ характеризується абсолютними та відносними розмірами, статистична наука, досліджую ї, утворює кількісні категорії. Через велику кількість явищ, що підлягають статистичному дослідженню, різноманіття їх типів і форм прояву абсолютних і відносних розмірів, статистика використовує у процесі аналізу широку систему кількісних категорій.

Статистичні величини у найзагальнішому вигляді можуть бути поділені на два великі класи відповідно розмірів вираження суспільних явищ. Один клас становить величини, що відображають абсолютні розміри явищ, інший — величини, що відображають їх відносні розміри. Перші з цих класів величин отримав назву «Абсолютні величини», другий — «Відносні величини».

Абсолютні та відносні величини — це дві основні, найзагальніші форми статистичних величин. Будь-який статистичний показник, який використовується у статистиці для характеристики властивостей явищ та процесів, приймає одну з цих двох форм. Усередині кожного з названих класів виділяють підкласи.

Необхідно мати на увазі, що в деяких випадках важко точно встановити загальну форму статистичної величини. Зустрічаються величини, які володіють, як побачимо нижче, деякими рисами як абсолютних, так і відносних величин. Тому їхнє віднесення до того чи іншого класу дещо умовне.

Іноді виділяють не дві основні, найзагальніші форми статистичних величин, як це зроблено вище, а три, додаючи до двох названих середні величини (див. гл. 7). При цьому не враховують, що самі середні величини виступають у формі або абсолютних або відносних величин.

## 6.2. Абсолютні статистичні величини

**Сутність та значення абсолютних величин**. У процесі статистичного спостереження, а потім зведення та групування статистичних даних отримують абсолютні величини.

*Абсолютні величини* характеризують розміри (обсяги, рівні) суспільних явищ. Відмінною особливістю абсолютних величин як форми статистичних показників є те, що вони безпосередньо пов'язані із соціальною, економічною основою чи речовою формою явищ, до яких вони відносяться. Розміри суспільних явищ можуть бути виражені або в чисельності одиниць сукупності (наприклад, чисельність населення), або у вигляді величини ознаки, що характеризує це суспільне явище (наприклад, обсяг промислової продукції країни, розмір заробітної плати, ціна виробу).

Абсолютні величини завжди виражені у певних вимірниках — у кілограмах, тоннах, штуках, гектарах, літрах, кубічних метрах, метрах, кілометрах, гектарах тощо. Звернемося до статистичного щорічника Державного комітету статистики України «Статистичний щорічник України». У ньому містяться численні дані про соціально-економічні явища, що відбуваються в країні. Відтворимо деякі їх у формі абсолютних величин. Наприклад, у 2020 р. в Україні було видобуто 24,2 млн т кам'яного вугілля, 20,2 млрд $м^3$ газу, вироблено 1549,3 тис. т борошна, 75,0 млн $м^2$ тканини із синтетичних ниток, 180,8 млн дал пива, 4,7 млн пар взуття з гуми або полімерних матеріалів, 20,2 млн чавуну. У 2020 р. в Україні валовий внутрішній продукт (у поточних цінах) становив 4 191,9 млрд грн; було зареєстровано 1 674,2 тис. безробітних. Чисельність осіб, які навчалися у закладах освіти, становила 5 766 тис. осіб. Кількість лікувальних закладів склала 1,7 тис. установ, у яких перебувало 309 тис. лікарняних ліжок. За 2020 р. прибуток до оподаткування підприємств економіки склав 674,0 млрд грн проти 101,9 млрд грн у 2012 р. Ось такого роду статистичні величини, які характеризують розміри (абсолютні) соціально-економічних явищ, їх ознак в одиницях міри протяжності, площі, маси (ваги), і т. п., в одиницях рахунку часу, у грошових одиницях або у вигляді числа елементів (одиниць), що становлять дане масове явище, називаються *абсолютними статистичними величинами.*

Абсолютна величина може відображати обсяг частини сукупності. Так, у 2020 р. капітальні інвестиції становили 419,8 млрд грн, з них в інженерні споруди — 13,3 млрд грн, транспортні засоби — 37,2 млрд грн, у нематеріальні активи — 18,9 млрд грн.

Абсолютні статистичні величини мають велике наукове та практичне значення, вони необхідні для того, щоб мати конкретне уявлення про розмір, обсяг досліджуваного явища. У планах промислових підприємств, торгових організацій, будівництв завжди зазначені конкретні абсолютні дані. Зрозуміло, що для здійснення контролю за виконанням плану треба й у звітних даних мати абсолютні величини. Без абсолютних даних про запаси сировини, чисельності робітників і т. д. не можна здійснити планування та облік на будь-якому підприємстві та в масштабі всієї країни. Абсолютні дані мають бути суворо достовірними та повними.

**Види абсолютних величин**. Розрізняють три види абсолютних величин: індивідуальні, групові та загальні. Групові та загальні величини називають підсумковими, чи сумарними.

*Індивідуальними* називають такі абсолютні статистичні величини, які виражають розміри кількісних ознак в окремих одиниць сукупності, що вивчається. Наприклад, мешканець країни, продукція підприємства, робітники заводу, запаси сталі на складі, число народжених дітей або померлих осіб. Усе це індивідуальні величини певних одиниць сукупностей. Індивідуальні абсолютні величини отримують у процесі статистичного спостереження і реєструються у формулярах спостережень.

Індивідуальні абсолютні величини мають значення у дослідженні соціально-економічних явищ. З них утворюють загальні абсолютні величини, при групуваннях за кількісними ознаками вони є підставою для віднесення кожної одиниці сукупності до тієї чи іншої та груп, що виділяються, вони є засобом аналізу соціально-економічних явищ та процесів, зокрема ефективності виробничо-господарської діяльності підприємств.

*Групові* та *загальні абсолютні величини* утворюються в процесі обробки матеріалів спостереження, узагальнення (зазвичай підсумовуванням) абсолютних розмірів ознаки в окремих одиниць сукупності або в результаті підрахунку числа одиниць сукупності, що входять в окремі групи, або всієї частини сукупності в цілому. З визначення ясно, що підсумкові (загальні та групові) абсолютні величини отримують в результаті підсумовування індивідуальних абсолютних величин, тобто значень ознаки в окремих одиниць сукупності, або шляхом підрахунку числа одиниць, що входять в окремі групи або в сукупність в цілому. Отже, аналізовані величини є результатом зведення даних статистичного спостереження. Якщо, наприклад, обробляються кількість населення за областями, то результаті підрахунку кількості осіб виходить загальна чисельність жителів країни, а підсумовування дає їх чисельність у всіх областях.

**Одиниці виміру абсолютних величин**. Вище говорилося, що абсолютні величини виражаються у певних одиницях. Але які одиниці вибрати в тому чи іншому випадку — питання часом складне і має велике значення в статистиці. Справа в тому, що багато явищ, що вивчаються у статистиці, можуть вимірюватися у різних одиницях. Так, наприклад, для вивчення працівників підприємства можуть використовуватися такі показники, стаж, освіта, стать, наявність підвищення кваліфікації та ін.; продукції — гривні, штуки, комплекти і т.д.

В залежності від якісної особливості досліджуваного явища та завдань дослідження абсолютні величини виражаються у різних одиницях виміру: *натуральних*, *грошових* (*вартісних*) та *трудових*. Вибір одиниць виміру абсолютних величин залежить від сутності явища, що вивчається, його фізичних і соціально-економічних властивостей, а також від цілей дослідження.

У ряді випадків застосовуються *умовно-натуральні одиниці* виміру. Сутність застосування умовних одиниць виміру полягає в тому, що окремі натуральні одиниці досліджуваної сукупності перераховуються в умовно-натуральні одиниці шляхом вираження різновидів явища в одиницях якого-небудь еталона. Тому основне питання застосування умовних одиниць виміру полягає у виборі ознаки, за якою встановлюються відповідні коефіцієнти перерахунку.

Однак у процесі дослідження якого-небудь явища або процесу, а також при плануванні та обліку обмежитися тільки абсолютними величинами не можна. За допомогою лише абсолютних величин неможливо здійснити повний аналіз, зробити чіткі висновки, показати певні закономірності та зв'язки.

## 6.3. Поняття та значення відносних показників

**Сутність та значення відносних величин**. Порівняльну оцінку явищ і процесів життя дають відносні величини. Вони є засобом узагальнення особливостей конкретних суспільних явищ.

*Відносними статистичними величинами* називають величини, що є мірою кількісного співвідношення статистичних показників і відображають відносні розміри соціально-економічних явищ. Ці величини обчислюється шляхом розподілу однієї статистичної величини в іншу. Вихідні показники (об'ємні, середні, відносні) можуть порівнюватися у часі, у просторі або з плановими (порівняння одноймених показників), або з іншими взаємопов'язаними показниками (порівняння різноіменних показників); можуть порівнюватися також частини ознаки або обсягу сукупності зі своїми загальними обсягами (порівняння частини та цілого). Це може бути співвідношення чисельності різних сукупностей явищ, їх окремих якісних показників, розмірів різних ознак однієї й тієї самої сукупності; співвідношення величини показника планової та фактичної, а також величини його за поточний та минулий (базовий) період.

Відносні величини широко застосовуються для вивчення пропорцій у народному господарстві країни, для характеристики рівня виконання планових завдань, виробництва продукції, науково-технічного прогресу, розміщення продуктивних сил по районах країни, ступеня концентрації виробництв та ін. Відносні величини допомагають розкрити якісні особливості та закономірності розвитку суспільних явищ, напрями та темпи їх розвитку, визначити відхилення від намічених планових завдань, виявити резерви, дати оцінку результатів виробничої діяльності.

Відносна величина завжди отримується як частка від ділення двох величин, що особливо важливо для оцінки її ролі в статистичній роботі. Величина, з якою зіставляється якась інша порівняна величина (знаменник дробу), зазвичай називається *основою* або *базисною величиною*, а та, яка порівнюється, називається *поточною*, *порівнянною* або *звітною величиною*.

Відносні величини можуть бути отримані також внаслідок зіставлення різноіменних величин. Для вираження різноіменних величин використовується поєднання найменувань порівнюваної величини та величини, прийнятої за основу порівняння. Наприклад, у 2019 р. споживання паливно-енергетичних ресурсів в Україні становило 89,1 млн т умовного палива (у перерахунку на нафтовий еквівалент). Середньорічна чисельність населення за цей період склала 42,0 млн. осіб. Таким чином, у зазначеному році на одного жителя країни припадало 2,2 т умовного палива.

Відносні величини відіграють важливу роль у проведенні статистичних досліджень у всіх сферах та галузях народного господарства країни. Справа в тому, що в багатьох випадках абсолютна величина, взята сама по собі, не завжди дозволяє дати якісну оцінку досліджуваним процесам та явищам суспільного життя. Крім того, у багатьох випадках безпосереднє зіставлення абсолютних величин є неможливим або можливим, але недоцільним: дані про чисельність населення або про виробництво окремих видів продукції за областями говорять нам дуже мало, тому що області різняться за величиною території, чисельності населення, площі посіву, і т. д. Дані про абсолютне число народжень і смертей за різним роком нічого не характеризують.

Перетворення абсолютних величин у відносні по суті відповідає в арифметиці приведення дробів до загального знаменника. Зіставляючи дані про чисельність населення на 1 км$^2$, ми приводимо їх до однієї площі території. Подібно до того, як у дробах береться спільний знаменник, економіст бере загальну основу: обсягу виробництва в національній економіці — валовий внутрішній продукт на одного жителя країни, для урожайності — площа в 1 га, для народжуваності — 1000 осіб населення, і до цього загального знаменника приводяться співставні величини.

Лише в комплексі, в доповненні з відносною величиною, можливо дати справжнє судження про факти суспільного життя, що відбуваються. Якщо, наприклад, відомо, що на даному підприємстві за рік було отримано прибуток у розмірі 500 тис. грн, то нічого не можна сказати про рівень ефективності виробничої діяльності на даному підприємстві. Тільки зіставивши цю величину із загальною сумою витрат окремих видів ресурсів, ми складемо судження про ефективність господарської діяльності підприємства.

Застосування відносних величин у статистиці забезпечує виявлення ступеня і динаміки розвитку явищ, міри співвідношень між складовими його частинами. За допомогою відносних показників виражаються багато фактів життя: темпи зростання і приросту, процент виконання плану, частку окремих категорій працівників у загальній чисельності персоналу підприємства, вікову структуру робочих, рівень народжуваності і смертності населення та ін.

Однією з важливих властивостей відносних показників і те, що вони абстрагують відмінність абсолютних величин і дозволяють порівнювати такі явища, абсолютні розміри яких не порівнювані. Якщо відомо, що в країні народилося протягом року 1 млн осіб, а в іншій — 2 млн осіб, то зі зіставлення цих абсолютних величин не можна зробити висновок про те, в якій країні рівень народжуваності вищий. Для відповіді на це питання потрібно взяти відносні показники — співвідношення чисельності народжених дітей та чисельності населення в кожній країні

**Форми вираження та види відносних величин**. Відносні величини отримуються у результаті зіставлення одноіменних та різноіменних величин.

Через зіставлення однойменних абсолютних величин отримуються величини, які не мають розмірності. Як правило, вони виражаються у формі кратного відношення, що показує, у скільки разів одна абсолютна величина більша за іншу. Наприклад, чисельність населення України на кінець 2019 р. становила 41,9 млн осіб, а Азербайджану — 10,0 млн осіб, відтак чисельність населення України в 4,2 рази перевищувала чисельність населення Азербайджану. У даному випадку базисна величина приймається за одиницю.

Основа відносної величини прирівнюється до одиниці або якогось числа, кратного 10 (100, 10 000 і т. п.). У першому випадку відносна величина показує, у скільки разів порівнянна величина більша за базисну. В інших випадках — скільки одиниць однієї величини припадає на 100, на 1000, на 10 000 і т. п. іншої, базисної величини.

Найбільш поширеною формою вираження є процентні відносини, за яких базисна величина приймається за 100. Наприклад, встановлена потужність електростанцій у нашій країні становила в 2019 р. 51,4 млн кВт, а в 2020 р. — 55,1 млн кВт; коефіцієнт зростання встановленої потужності електростанцій становив 1,072 $\left(\frac{55,1}{51,4}\right)$, у відсотках — 107,2 % $\left(\frac{55,1 \cdot 100}{51,4}\right)$. Якщо база прирівняна до 1 000, відносний показник виражається у проміле (‰), до 10 000 — у продециміліі (‱). Так, в Україні народжуваність у 2020 р. становила 7,8 особи на 1000 осіб постійного населення.

Приклад показників, розрахованих ДСС України на 100 000 мешканців, наведено у табл. 6.1.

Обчислення відносних показників у проміліі, процедиміліі тощо надає відносним величинам більшої виразності, більш зручнішого вигляду для сприйняття, звільнивши їх від дуже дробових чисел або від великої кількості знаків після коми у десятковому дробі.

Через зіставлення різноіменних абсолютних

Таблиця 6.1

**Смертність населення за окремими причинами в Україні на 100 000 наявного населення**

| Причини | 2005 | 2010 | 2015 | 2019 |
|---|---|---|---|---|
| Хвороб системи кровообігу | 1 037,6 | 1 013,9 | 1 009,5 | 989,2 |
| Новоутворень | 195,0 | 193,5 | 199,8 | 199,3 |
| Хвороб органів дихання | 59,4 | 42,5 | 35,3 | 32,0 |

величин виходять іменовані відносні величини. Їх найменування утворюється шляхом поєднання найменувань порівнюваної та базисної величини. Такими, наприклад, є загальні коефіцієнти народжуваності, смертності та природного приросту (скорочення) населення (на 1000 осіб наявного населення); щільність населення, що вимірюється чисельністю жителів, які припадають на 1 км$^2$; виробництво та споживання продукції на одного жителя та ін.

Залежно від завдань, розв'язуваних з допомогою відносних величин, розрізняють кілька видів. Серед них: *відносна величина динаміки*, *виконання плану* та *планового завдання*, *просторового порівняння*, *порівняння*, *структури*. Розглянемо призначення кожної величини та техніку її обчислення.

## 6.4. Відносні показники динаміки

*Відносною величиною динаміки* називають співвідношення величини показника за цей період (рік, квартал, місяць тощо) і величини його за будь-який аналогічний попередній час, прийнятою за базу порівняння. Вони характеризують інтенсивність і швидкість розвитку досліджуваного явища в часі, або інакше, темп розвитку. Відносна величина динаміки виявляється у формі кратного відношення (коефіцієнтів) або відсотків.

При цьому зазвичай показник вихідного періоду, який називається *базисним*, приймається за 100 або за одиницю, а показники наступних років виражаються у відсотках або коефіцієнтах до базисного. Показники динаміки називають темпами.

Капітальні інвестиції в народне господарство України за рахунок усіх джерел фінансування становили у 2018 р. 578,7 млрд грн, а у 2019 р. — 624,0 млрд грн. Відносна величина динаміки дорівнює: 624,0 : 578,7 = 1,078, або 107,8%. Отже, у 2019 р. капітальні інвестиції у народне господарство країни збільшились порівняно з 2018 р. на 7,8 %.

У випадку, якщо є дані кілька періодів часу, то порівняння кожного цього рівня може здійснюватися або з рівнем попереднього періоду, або з будь-яким іншим, прийнятим за основу. Перші називають відносними величинами динаміки *зі змінною базою порівняння*, або *ланцюговими*, другі — відносними величинами динаміки *з постійною базою порівняння*, або *базисними*.

Приклад. Споживання м'яса країни характеризується такими даними (на душу населення; кг на рік):

| | | | | | |
|---|---|---|---|---|---|
| 1990 | 1995 | 2000 | 2005 | 2010 | 2015 |
| 68,2 | 38,9 | 32,8 | 39,1 | 52,0 | 56,1 |

Для виявлення напрямку та характеру споживання м'яса в країні за період ринкових перетворень порівняно з 1990 визначимо базисні темпи зростання $(T_б)$:

$$T_б = \frac{y_i}{y_0} \cdot 100, \qquad (6.1)$$

де $y_i$ — рівень досліджуваного періоду;

$y_0$ — базисний рівень.

Послідовно порівняємо рівні 1995, 2000, 2005, 2010 и 2015 рр. $y_i$ з рівнем 1990 г. $y_0$:

1995 $\frac{38,9}{68,2} \cdot 100 = 57,0$ %;

2000 $\frac{32,8}{68,2} \cdot 100 = 48,1$ %;

2005 $\frac{39,1}{68,2} \cdot 100 = 57,3$ %;

2010 $\frac{52,0}{68,2} \cdot 100 = 76,2$ %;

2015 $\frac{56,1}{68,2} \cdot 100 = 82,2$ %.

З отриманих базисних відносних величин динаміки (темпів зростання) видно, що за вказані роки споживання м'яса в країні не перевищило рівень 1990 р. Найбільші темпи скорочення споживання м'яса були характерними у 2000 р. У 2015 р. рівень споживання м'яса становив лише 82,2 % від рівня 1990 р.

$$57,0>48,1<57,3<76,2<82,2.$$

Для обчислення ланцюгових відносних величин динаміки ($T_ц$) за базу порівняння кожного рівня здійснюється з рівнем попереднього періоду:

$$T_ц = \frac{y_i}{y_{i-1}} \cdot 100, \qquad (6.2)$$

де $y_i$ — рівень досліджуваного періоду;

$y_{i-1}$ — рівень попереднього періоду.

Визначимо ланцюгові темпи зростання:

1995 $\frac{38,9}{68,2} \cdot 100 = 57,0$ %; 2010 $\frac{52,0}{39,1} \cdot 100 = 133,0$ %;

2000 $\frac{32,8}{38,9} \cdot 100 = 84,3$ %; 2015 $\frac{56,1}{52,0} \cdot 100 = 107,9$ %.

2005 $\frac{39,1}{32,8} \cdot 100 = 119,2$ %;

Запишемо вихідні дані та результати розрахунків в одній таблиці (табл. 6.2).

Таблиця 6.2

| Показник | 1990 р. | 1995 р. | 2000 р. | 2005 р. | 2010 р. | 2015 р. |
|---|---|---|---|---|---|---|
| Споживання м'яса, кг | 68,2 | 38,9 | 32,8 | 39,1 | 52,0 | 56,1 |
| Ланцюгові темпи зростання (у відсотках до попереднього періоду) | – | 57,0 | 84,3 | 119,2 | 133,0 | 107,9 |
| Базисні темпи зростання (в процентах до 1990 р.) | 100 | 57,0 | 48,1 | 57,3 | 76,2 | 82,2 |

З отриманих ланцюгових відносних величин динаміки (темпів зростання) видно, що у окремих етапах економічного розвитку відбувалися коливання споживанні м'яса. У 1995–2000 рр. темпи споживання м'яса країни скорочувалися. Так, у 1995 р. споживання м'яса у країні становило лише 57,0 % від рівня 1990 р. У 2000–2015 рр. темпи зростання споживання м'яса зростали, досягнувши максимальної величини 2010 р. 133,0 %.

Представимо вихідні рівні ряду динаміки у вигляді стовпчикової діаграми (рис. 6.1).

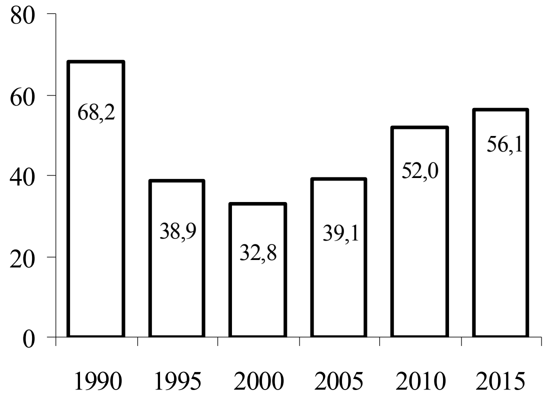



*Рис. 6.1.* Споживання м'яса на душу населення в Україні у 1990–2015 рр. (кг за рік).

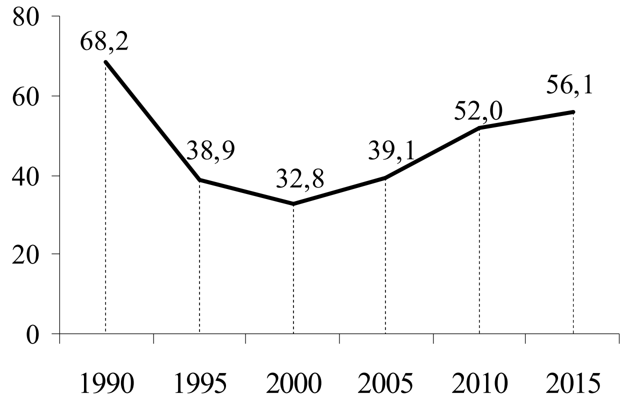



*Рис. 6.2.* Темпи зростання споживання м'яса у 1990–2015 роках. за рік душу населення (у відсотках до 1990 р.).

Представимо графічно отримані розрахунку базисні відносні величини динаміки. Для цієї мети найчастіше використовується лінійна діаграма (рис. 6.2).

У системі координат нанесемо на вісь ординат базисні темпи зростання (у відсотках), але в осі абсцис — показники часу.

З цього графіка видно, що до 2000 р. споживання м'яса стійко скорочувалося. Лише з 2000 р. спостерігається стійке зростання темпів зростання споживання м'яса. Водночас абсолютне скорочення споживання м'яса за 1990–2000 рр. перевищило його зростання за 2000–2015 роки. Як наслідок, рівень споживання м'яса у 2015 р. не перевищив рівень 1990 р.

Ланцюгові та базисні величини пов'язані між собою. Послідовне перемноження ланцюгових відносних величин дає змогу отримати відповідну базову величину. Так, базисні темпи зростання за 1990-2015 роки. складе:

$$\frac{57{,}0}{100}\cdot\frac{84{,}3}{100}\cdot\frac{119{,}2}{100}\cdot\frac{133{,}0}{100}\cdot\frac{107{,}9}{100}=0{,}822,\text{ або }82{,}2\,\%.$$

Частка від ділення даної базисної відносної величини динаміки на попередню дорівнює ланцюговій відносній величині динаміки, наприклад:

$$\frac{48{,}1}{100}:\frac{57{,}0}{100}=0{,}843,\text{ або }84{,}3\,\%,$$

або

$$\frac{0{,}481}{0{,}570}=0{,}843,\text{ або }84{,}3\,\%.$$

Вибір бази для динамічних рядів залежить від цього, з чим ми хочемо порівнювати даний рівень. У деяких випадках від цього вибору залежить наочність графічних уявлень. Припустимо, нам треба показати порівняльний рух двох рівнів (А та Б) на три дати: 2005 р., 2010 р., 2015 р. (тис. грн):

| | | | |
|---|---|---|---|
| А | 230 | 420 | 350 |
| Б | 170 | 260 | 410 |

Якщо порівнювати рух рівнів у процентах, ми отримаємо три можливі варіанти (рис. 6.3).

Завдання порівняння досягнутого рівня з висхідним мабуть, найкраще виконано на 1 рисунку. Дуже часто у зарубіжних публікаціях за 100 приймають і середину інтервалу.

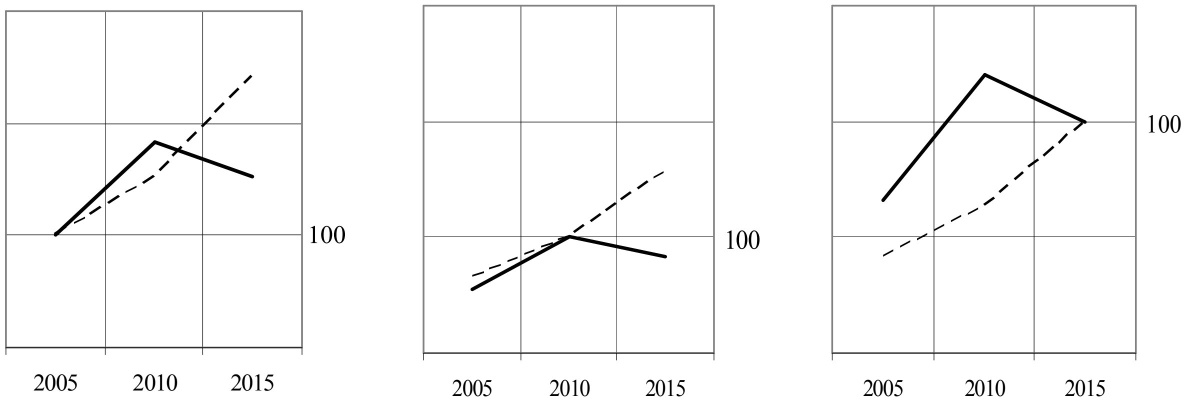



*Рис. 6.3.* Варіанти обчислення процентів.

Відносні величини динаміки мають велике значення і широко поширені при дослідженні масових соціально-економічних явищ, особливо в соціально-економічних дослідженнях. Це зумовлено тим, що вивчення розвитку явища в часі, зокрема, вивчення розвитку народного господарства країни, — найважливіша теоретична та практична задача науки.

## 6.5. Відносні показники виконання плану та планового завдання

**Відносні величини виконання плану**. *Відносна величина виконання плану* представляє собою відношення величини показника, досягнутого за якийсь час (або до якогось моменту часу), і величини його, встановленої за планом на цей же час. Зазвичай відносна величина виконання плану виражається у відсотках і показує, на скільки процентів виконаний план. Різниця між відносною величиною виконання плану і 100 % може дорівнювати нулю, мати позитивне або негативне значення. Різниця, що дорівнює нулю, свідчить про точне виконання планового завдання. Якщо плановий показник такий, що його зростання — явище позитивне (наприклад, кількість виготовленої продукції, середня врожайність сільськогосподарських культур, середня продуктивність тварин тощо), то різниця з позитивним знаком вказує на перевиконання плану, а з негативним — на недовиконання. Якщо ж характер показника такий, що позитивним є зменшення його розмірів (наприклад, скорочення браку продукції, зниження собівартості продукції), то перевищення фактичної величини над його плановою свідчить про невиконання плану, а якщо вона менша від планової — то про переповнення плану.

Відносні величини виконання плану мають велике значення як засіб контролю та аналізу виконання планів. Ці показники широко застосовуються як на рівні окремого підприємства (для визначення ритмічності, для оцінки виконання бізнес-плану), так і на державному рівні (при аналізі виконання державного та місцевих бюджетів, державних цільових програм, планів соціально-економічного розвитку тощо).

Способи обчислення відносних величин виконання плану залежить від того, в якому вигляді дані показники плану. Планове завдання може бути виражене у вигляді *абсолютних величин* (досягти певного обсягу виробництва продукції, чисельності працівників, суми витрат на виробництво тощо), *середніх* (досягти встановленої урожайності, середньої заробітної плати, норм витрати матеріалу тощо) або *відносних величин* (забезпечити зростання випуску продукції, підвищення використання робочого дня, зниження собівартості виробів і т. п.).

При цьому слід розрізняти такі випадки.

1. Якщо планове завдання встановлено у вигляді абсолютних чи середніх величин, відносний показник виконання плану обчислюється як відношення фактично досягнутої у звітному періоді абсолютної (звітної) величини до абсолютної величини планового завдання (планової):

$$\text{Відносна величина виконання плану (\%)} = \frac{\text{фактичний рівень}}{\text{планове завдання}} \cdot 100.$$

В залежності від періоду порівняння названі показники можуть розраховуватися по-різному. Це залежить від того, чи хочемо ми охопити виконання плану за окремими відрізками аналізованого періоду або ставитися питання про виконання плану в наростаючих підсумках з початку періоду.

Припустимо, що виконання плану виробництво продукції в механічному цеху за другий квартал характеризувалося таким чином: у квітні було випущено продукції на 18 млн грн, у травні — 20, у червні — 25 млн грн, разом 63 млн грн. План на квартал було встановлено у 60 млн грн. Виконання квартального плану випуску продукції цехом складе:

за квітень на 30 % $\left(\frac{18}{60}\cdot 100\right)$;

за травень на 33,3 % $\left(\frac{20}{60}\cdot 100\right)$;

за червень на 41,7 % $\left(\frac{25}{60}\cdot 100\right)$;

за квартал на 105 % $\left(\frac{63}{60}\cdot 100\right)$.

Процент виконання плану отримується як сума місячних процентів: 30 + 33,3 + 41,7 = 105.

Але можна представити виконання плану в наростаючих підсумках:

виконання плану за квітень складає 30,0 % $\left(\frac{18}{60}\cdot 100\right)$;

за квітень і травень — 63,3 % $\left(\frac{18+20}{60}\cdot 100\right)$;

за квітень, травень і червень — 105 % $\left(\frac{18+20+25}{60}\cdot 100\right)$.

Для оперативного керівництва виробництвом найчастіше використається другий спосіб.

Найчастіше при розрахунку відносних показників виконання плану порівнюють звітні та планові показники за одні й самі періоди. Це дозволяє визначити рівень виконання плану за кожний період окремо.

Приклад. Маються такі дані про виробництво у звітному періоді продукції окремих цехів промислового підприємства:

Таблиця 6.3

**Дані про виробництво продукції окремих цехів промислового підприємства**

| Цех | За планом, млн грн | Фактично, млн грн | Процент виконання плану, % |
|---|---|---|---|
| Зварювальний | 62,4 | 65,8 | 105,4 |
| Інструментальний | 18,2 | 17,5 | 96,2 |
| Збиральний | 21,6 | 22,3 | 103,2 |
| Всього | 102,2 | 105,6 | 103,3 |

В останній графі визначені відносні величини виконання плану випуску продукції у звітному періоді за окремими цехами та в цілому по підприємству. Так, якщо за планом обсяг випуску у зварювальному цеху становив 62,4 млн грн, а фактично ним було виготовлено продукції на 65,8 млн грн, то відсоток виконання плану буде: $\frac{65,8}{62,4} \cdot 100 = 105,4$ %. Отже, план виконано на 105,4 %, а перевиконання плану складає 5,4%. Аналогічно обчислювалися відносні показники виконання плану з інших цехів.

Для наочного уявлення результатів виконання плану окремими цехами підприємства плану виробництва продукції отримані дані доцільно нанести на полосовий графік (рис. 6.4).

2. Якщо план передбачає не зростання, а зниження рівнів, як наприклад, собівартості виробництва продукції, то відносна величина виконання плану визначається шляхом порівняння фактично досягнутого та запланованого зниження рівня:

$$\text{Відносна величина виконання плана (\%)} = \frac{\text{фактическая величина снижения}}{\text{плановая величина снижения}} \cdot 100.$$

Приклад. Рівень собівартості виробництва 1 т продукції *А* в базисному році становив 826 грн. Планом на 2020 р. намічено зниження витрат на виробництво тонни цієї продукції на 16 грн. Фактична собівартість тонни цієї продукції склала за звітом за 2020 рік 809 грн. В и з н а ч т е відносну величину виконання плану зниження собівартості виробу *А* в 2020 р.

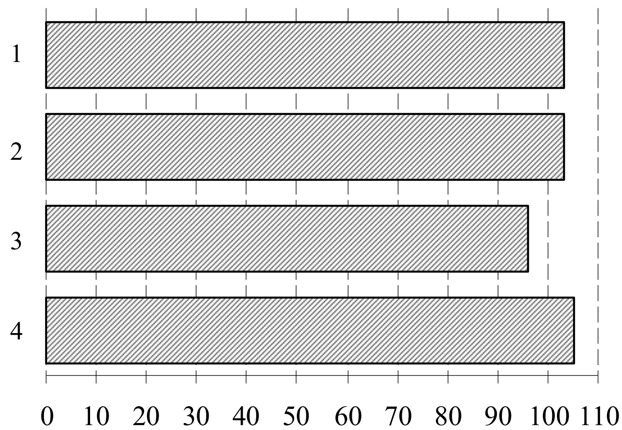



*Рис. 6.4.* Виконання плану виробництва окремих цехів підприємства (у процентах); 1 — зварювальний цех; 2 — інструментальний цех; 3 — складальний цех; 4 — вся продукція підприємства.

Для обчислення процента виконання плану зниження в 2020 р. собівартості виробу *А* знайдемо фактичну величину зниження собівартості: 826–809= 17 (грн).

Розрахуємо відносну величину виконання плану як відношення величини зниження собівартості (17) до величини зниження собівартості за плановим завданням:

$$\text{Відносна величина виконання плана} = \frac{17}{16} \cdot 100 = 106{,}3\ \%.$$

Таким чином, планове завдання щодо зниження собівартості виробу у 2015 р. перевиконано на 6,3 %.

Якщо планове завдання передбачає зниження рівня показника, то результат порівняння фактичного рівня із запланованим, що становив за своєю величиною менше 100%, свідчить про перевиконання плану.

Приклад. Планова трудомісткість виготовлення електродвигуна на машинобудівному заводі становила 285,4 люд.-год., а фактична — 264,5 чол.-год. Перевиконання плану зниження трудомісткості продукції складе:

$$\frac{264{,}5}{285{,}4} \cdot 100 - 100 = 7{,}3\ \%,$$

тобто план зниження трудомісткості виготовлення електродвигуна перевиконано на 7,3 %.

3. Якщо планове завдання виражено у вигляді відносних величин, то в цьому випадку для визначення відносної величини виконання плану потрібно знайти відношення звітної величини до тієї, яка була прийнята як базисна при встановленні планового завдання, і віднести (розділити) отриману відносну величину до планової відносної величини, тобто:

$$\text{Відносна величина виконання плану (\%)} = \frac{\text{коефіцієнт фактичного зростання}}{\text{коефіцієнт планового завдання}} \cdot 100.$$

Приклад. Річний план фабрики на 2021 р. передбачав зростання випуску товарної продукції на 5,5 %.Фактично приріст товарної продукції цей рік становив 8,8 %. В и з н а ч т е відносну величину виконання фабрикою річного плану зростання випуску товарної продукції.

Заданий планом приріст випуску продукції (5,5 %) виявляється у формі коефіцієнта зростання випуску продукції проти 2020 р.: $\frac{100+5{,}5}{100}=1{,}055$.

Фактичний відсоток зростання випуску продукції (108,8%) виражаємо у формі коефіцієнта:

$$\frac{100+8{,}8}{100}=1{,}088.$$

За допомогою отриманих коефіцієнтів розрахуємо відносну величину виконання плану:

$$\frac{1{,}088}{1{,}055}\cdot 100=103{,}1\,\%.$$

тобто досягнутий в 2021 р. план випуску продукції перевиконано фабрикою на 3,1 %.

Порівнювати статистичні дані можна лише спів ставні дані. У зв'язку з цим обов'язковою вимогою при розрахунку відносних величин виконання плану є спів ставність фактичних та планових показників. *Співставність* — можливість порівняння, зіставлення даних з метою виявлення тенденцій, закономірностей розвитку відображуваних ними явищ. Це означає, що:

дані, які використовуються з метою порівняння, повинні відноситися до того самого об'єкта, бо не можна зіставляти планові дані, що відносяться до іншого об'єкта;

дані мають бути порівняні щодо конкретної сутності показника (якщо за планом береться показник рівень якості випуску деталі *A*, то фактичні дані повинні відображати якість випуску деталі *A*, а не обсягу випуску деталей *A*);

розміри показників плану та фактичних показників повинні обчислюватися за єдиною методологією.

Статистичні показники дуже часто виявляються непорівнянні (незрівнянні) в силу відмінностей їх предметного змісту, одиниць вимірювання і т. п. Однак у багатьох випадках непорівнянні дані можна привести до сумісності. Вона досягається перерахунком даних; приведенням даних до одних і тих де одиниць виміру; відмежуванням і виділенням з усієї маси даних однорідної та порівнянної сукупності їх (порівняного кола); заміною абсолютних величин відносними величинами; заміною абсолютних величин середніми величинами; спеціальним групуванням та перегрупування даних.

**Відносні величини планового завдання**. В економічному аналізі та оперативному управлінні підприємства планове завдання може бути виражене і у формі відносної величини, тобто у вигляді коефіцієнтів зростання або приросту рівня в плановому періоді порівняно з рівнем базисного періоду. Якщо планові завдання встановлені у відносних величинах, то розрахунок виконання плану є дещо складнішим. З цією метою розраховують відносні показники виконання планового завдання. Отримуються вони в результаті порівняння плану майбутнього періоду з фактичним рівнем, прийнятим за основу.

*Відносні величини планового завдання* представляють собою відношення величини показника, що встановлюється на планований період до його фактичної величини, досягнутої за попередній період або за який-небудь інший, прийнятий за базу. Отже, відносна величина планового завдання показує, яке планується зміна певного показника в порівнянні з досягнутим на момент складання плану. Розрахунок відносної величини планового завдання здійснюється за формулою:

$$\frac{y_{\text{пл}}}{y_0}, \tag{6.3}$$

де $y_{\text{пл}}$ — величина показника, що встановлюється на планований період;

$y_0$ — фактична величина, досягнута за попередній період або за якийсь інший період.

Наприклад, обсяг товарної продукції машинобудівного заводу у 2021 р. становив 245,7 млн грн, а річною програмою на 2022 р. заплановано у розмірі 256,8 млн грн. Відносний показник планового завдання на 2022 рік складе:

$$\frac{256{,}8}{245{,}7} \cdot 100 = 104{,}5\ \%.$$

Таким чином, за планом на 2022 рік передбачалося збільшити загальний обсяг товарної продукції заводу на 4,5 % порівняно з обсягом, досягнутим у 2021 р.

Відносні показники планового завдання набули широкого поширення у статистиці та плануванні. Зокрема, їх використовують при оцінці ступеня напруженості виробничої програми, планових показників собівартості, витрат окремих видів ресурсів, виконання робітниками норм виробітку тощо.

Відносні величини динаміки, планового завдання та виконання плану пов’язані між собою таким співвідношенням: відносна величина динаміки дорівнює добутку відносних величин планового завдання та виконання плану, тобто

$$\begin{matrix}\text{Відносний} \\ \text{показник} \\ \text{динаміки}\end{matrix} = \begin{matrix}\text{Відносний} \\ \text{показник} \\ \text{виконання плану}\end{matrix} \times \begin{matrix}\text{Відносний} \\ \text{показник} \\ \text{планового завдання}\end{matrix}$$

Позначимо базисні символи $y_0$, планові — $y_{пл}$, звітні — $y_1$. За визначенням

відносний показник динаміки $= \frac{y_1}{y_0}$; (6.4)

відносний показник виконання плану $= \frac{y_1}{y_{пл}}$; (6.5)

відносний показник планового завдання $= \frac{y_{пл}}{y_0}$. (6.6)

Звідси видно, що

$$\frac{y_1}{y_0} = \frac{y_{пл}}{y_0} \cdot \frac{y_1}{y_{пл}}. \tag{6.7}$$

Використовуючи встановлений взаємозв'язок між трьома відносними показниками, визначимо відносний показник виконання плану:

$$\text{Відносний показник виконання плану} = \frac{\text{Відносний показник динаміки}}{\text{Відносний показник планового завдання}} = \frac{y_1}{y_0} : \frac{y_{пл}}{y_0} = \frac{y_1}{y_{пл}}. \tag{6.8}$$

Приклад. Випуск продукції підприємства характеризується наступними даними (тис. грн):

| Базисний період | Звітний період | |
|---|---|---|
| | По плану | Фактично |
| 46 854 | 47 928 | 49 763 |

Визначити відносні величини планового завдання, виконання плану і динаміки.

Відносна величина динаміки:

$$\frac{y_1}{y_0} = \frac{49\,763}{46\,854} = 1{,}062.$$

Відносна величина планового завдання:

$$\frac{y_{пл}}{y_0} = \frac{47\,928}{46\,854} = 1{,}023.$$

Відносна величина виконання плану:

$$\frac{y_1}{y_{\text{пл}}} = \frac{49\ 763}{47\ 928} = 1{,}038.$$

Взаємозв'язок між відносної динаміки, планового завдання та виконання плану:

$$\frac{y_1}{y_0} = \frac{y_{\text{пл}}}{y_0} \cdot \frac{y_1}{y_{\text{пл}}} = 1{,}023 \cdot 1{,}038 = 1{,}062.$$

Таким чином, знаючи будь-які дві з цих трьох величин, завжди можна знайти третю.

## 6.6. Відносні показники координації

*Відносною величиною координації* називають співвідношення розмірів частин деякого цілого між собою. При розрахунку відносних величин координації одну з частин цілого приймають за основу порівняння і знаходять співвідношення до неї всіх інших елементів. Можливо порівнювати дві взаємопов'язані ознаки досліджуваного явища або процесу, і тоді в чисельнику міститься результативна, або залежна ознака, в знаменнику — факторний, або впливаюча ознака. Іноді відносні величини координації нерідко вважають відносними показниками структури, яку вони дають деяке уявлення. Але мають інше завдання, ніж відносні показники структури: відносні величини координації дозволяють виявити невідповідності між окремими частинами єдиного цілого, між розмірами різнорідних, але тісно взаємопов'язаних ознак, диспропорції у розвитку народного господарства. Відносні величини координації виражаються у кратних співвідношеннях.

За допомогою відносних величин координації визначають, у скільки разів порівнювана частина більша або менша частини, прийнятої за основу, або скільки одиниць однієї частини цілого припадає на 1, 10, 100, 1000 і т. д. одиниць іншої його частини, прийнятої за базу порівняння. Прикладом відносних показників координації можуть бути також показники: кількість жінок на 1000 чоловіків (у країні чи будь-якому регіоні); число допоміжних робітників, що припадають на одного основного робітника; число службовців та молодшого обслуговуючого персоналу, що припадає на одного робітника (на підприємстві, у певній галузі народного господарства); кількість студентів, які припадають на одного викладача вузу.

Наведемо приклад. Так, на 1 січня 2021 р. на підприємстві працювало 840 робітників, 60 осіб інженерно-технічного персоналу, 40 осіб молодшого обслуговуючого персоналу. У нашому прикладі можна обчислити, яке співвідношення між чисельністю інженерно-технічних працівників, службовців тощо і чисельністю робітників підприємства. Найчастіше розрахунок відносних показників координації відносно персоналу ведеться на 100 осіб. З прикладу видно, що на 100 робітників припадає приблизно 7 інженерно-технічних працівників $\left(\frac{60}{840}\cdot 100 \cong 7\right)$, близько 5 молодшого обслуговуючого персоналу $\left(\frac{40}{840}\cdot 100 \cong 5\right)$. Розрахунок можна зробити і в іншому порядку, розділивши кількість робітників на число інженерно-технічних працівників, службовців або молодшого обслуговуючого персоналу.

Наведемо інший приклад. Так, на 1 січня 2020 р. у нашій країні чисельність постійного населення за статтю становила 19 343,4 тис. чоловіків та 22 389,3 тис. жінок. Отже, на 1000 чоловіків припадало: (22 389,3: 19 343,4) · 1000 = 1 157 жінки.

## 6.7. Відносні величини інтенсивності

Одним із важливих видів відносних величин, що застосовуються у статистиці, є відносні величини інтенсивності.

*Відносні величини інтенсивності* (*ступеня*) представляють собою співвідношення обсягів двох різних сукупностей, що знаходяться у певному зв'язку одна з одною. Ці показники отримують шляхом порівняння двох різноіменних, але пов'язаних між собою абсолютних величин і обчислюються як відношення величини явища, що вивчається, до обсягу того середовища, в якому відбувається розвиток явища. Одне — середовище (її розмір), у якому відбувається розвиток якого-небудь явища чи які нею породжуються, інше — досліджуваний процес, явище (їх величина). Відносні величини інтенсивності характеризують ступінь розвитку (розповсюдження) того чи іншого процесу, явища у певному середовищі. Цей показник застосовується у разі, коли абсолютна величина виявляється недостатньою для обґрунтування висновків про масштаби, ступінь поширеності явищ, насиченості, щільності. Такі, наприклад, демографічні коефіцієнти смертності, народжуваності, шлюбності; економічні показники продуктивність праці, фондовіддачі, матеріаломісткості, фондовіддачі і енергоозброєності та ін. Цими ж відносними показниками характеризується рівень розвитку, т. е. виробництво найголовніших товарів душу населення.

Так, мається кількість дітей, що народилися на певній території протягом якого-небудь року. Порівнюючи цю чисельність із чисельністю дітей минулого року, можна зробити висновок про збільшення чи зменшення народжуваності. Можна порівняти її із сукупністю дітей, які народилися в іншій області. Але на підставі отриманих даних зробити висновок, де народжуваність вища, не можна, оскільки абсолютна кількість дітей пов'язана з чисельністю населення. Якщо обчислити коефіцієнт народжуваності (відношення числа народжених за рік до середньої чисельності населення, виражене в промілях, тобто на 1 000 жителів), можна порівнювати ці показники по різних регіонах і країнах. В результаті порівняння дізнаємося, що у Харківській області у 2020 р. народжуваність була в 1,6 рази нижчою, ніж у Волинській області, і становила 6,3 осіб на 1000 (тобто 6,3 ‰), у той час як у Полтавській області — 6,3, у Сумській області — 5,7, у Чернігівській — 6,0,, в Херсонській області — 7,5 і т. д.[1]

Аналогічно обчислюється і коефіцієнт смертності як показник кількості померлих протягом року на 1 000 осіб середньорічного населення.

Порівнюючи сукупності народжених і померлих із чисельністю населення, ми цим оцінюємо ступінь, інтенсивність розвитку цього явища. Тому такі показники називають показниками інтенсивності розвитку.

Відносні величини інтенсивності показують, скільки одиниць однієї величини припадає на 1, 100, 1 000 і т. п. одиниць іншої величини, з якою проводиться порівняння. Відносні величини інтенсивності називають також відносними величинами ступеня чи частоти.

Приклад. На кінець 2017 р. чисельність лікарів усіх спеціальностей становила 186 тис. осіб, а чисельність постійного населення на 1 січня 2018 р. — 42 216,8 тис. осіб. Отже, у 2017 р. на кожні 10 000 мешканців припадало 44 лікарі: (186:42216,8) ··1000=44,1.

---

[1] Природний рух населення України за 2020 рік: Статистичний збірник. Київ: Державна служба статистики України, 2021, С. 148.

Значна частина показників, що характеризують якість підприємств, є показниками інтенсивності. Вони дозволяють розкрити важливі процеси розвитку суспільного життя.

Відносні величини інтенсивності є іменованими числами і можуть виражатися в кратних відносинах, відсотках, проміле та інших формах. При розрахунках відносної величини інтенсивності база дорівнює 1; 100; 1 000 і т. п. Відносні величини інтенсивності часто називають коефіцієнтами.

Відносні величини у вигляді часток, відсотків та проміле можна виразити у загальній формі:

$$\frac{a}{b} \times 10^k, \tag{6.9}$$

де $k$ — деякий цілий показник ступеня.

Якщо $k = 0$, то $10^k = 1$, ми отримаємо просто частку (у випадку, коли $\frac{a}{b}$ відношення менше одиниці). Якщо $k = 2$, то відсотки, якщо $k = 3$, то проміле. У демографічній статистиці показники розраховують на 10 000 У деяких випадках у демографії розрахунки ведуть до 10 000 ($k = 4$). Такі різні позначення обумовлені сутністю та особливостями розвитку явища.

Приклад. По одному з міст області отримані наступні дані за 2022 р.:

Таблиця 6.4

| Число народжених | Число померлих | Число браків | Число розлучень | Середньорічна чисельність населення |
|---|---|---|---|---|
| 1 342 | 621 | 720 | 193 | 76 620 |

Визначте відносну величину, що характеризує народжуваність та смертність дітей.

Для вирішення завдання визначимо коефіцієнт народжуваності дітей:

$$\text{Коефіцієнт народжуваності} = \frac{\text{Кількість народжених}}{\text{Середньорічна чисельність населення}} = \frac{1342}{76\,620} = 0{,}0175,$$

або 17,5 ‰ (проміле).

Таким чином, народжуваність дітей у районі для 1 000 осіб населення становила 17,5 дітей.

Чисельник і знаменник відносного показника інтенсивності вимірюються в однакових одиницях, тому показники, хоч і мають свої назви (коефіцієнт народжуваності, коефіцієнт плинності тощо), виражаються в кратних відношеннях, відсотках, проміле, продециміллі та подібних до них формах. Наведемо приклад.

Маються дані про природний рух населення України:

Таблиця 6.5

| Рік | На 1000 осіб | | |
|---|---|---|---|
| | кількість народжених | кількість померлих | природний приріст |
| 2000 | 7,8 | 15,4 | –7,6 |
| 2010 | 10,8 | 15,2 | –4,4 |
| 2020 | 7,8 | 15,9 | –8,1 |

*Джерело:* Природний рух населення України за відповідні роки: Статистичний збірник. Київ: Державна служба статистики України.

Усі демографічні коефіцієнти, розміщені у таблиці, є відносними показниками інтенсивності, вираженими в проміле. Наведені коефіцієнти народжуваності, зокрема, характеризують негативний тренд у зміні природного приросту населення. Справді, народжуваність населення 2020 р. знизилася проти 2010 р. на 27,8 %.

Відносні величини інтенсивності отримали широке поширення при обчисленні диференціації рівня життя, рівня добробуту, демографічних змін і т. п. Наприклад, валовий внутрішніх продукт у порівнянних цінах на душу населення; кількість осіб, які мають середню та вищу освіту, на 1000 населення; ступінь забезпеченості населення окремих міст житлом (житлова площа душу населення); кількість торгових підприємств на 1000 осіб; обслуговування населення медичною допомогою (число лікарів на 10 тис. осіб населення); кількість лікарняних ліжок на 10 тис. населення.

До цієї категорії показників відносяться показники щільності населення (кількість жителів на 1 *кв. км* території), густоти залізничної мережі (довжина залізниць на 1000 *кв. км* площі країни або району), густоту перевезень (млн т перевезених вантажів або мільйон пасажирів на 1 км експлуатаційної довжини колії).

Різновидом відносних величин інтенсивності, що застосовуються для аналізу диференціації доходів, є квінтильний коефіцієнт диференціації доходів та квінтильний коефіцієнт фондів. *Квінтильний коефіцієнт диференціації доходів населення* — співвідношення мінімального рівня доходів серед 20 % найменш забезпеченого населення до максимального рівня доходів серед 20 % найменш забезпеченого населення. *Квінтильний коефіцієнт фондів* — співвідношення сумарних доходів 20 % найбільш і 20 % найменш забезпеченого населення.

Таблиця 6.6

**Диференціація життєвого рівня населення**

| Рік | Квінтильний коефіцієнт диференціації загальних доходів населення, разів | Квінтильний коефіцієнт фондів (по загальних доходах), разів |
|---|---|---|
| 2014 | 1,9 | 3,1 |
| 2015 | 1,9 | 3,2 |
| 2016 | 1,9 | 3,0 |
| 2017 | 1,9 | 3,3 |
| 2018 | 2,0 | 3,3 |
| 2019 | 2,1 | 3,5 |
| 2020 | 2,0 | 3,5 |

*Джерело*: Диференціація життєвого рівня населення. Державна служба статистики України. URL: http://www.ukrstat.gov.ua.

Динаміка показників диференціації рівня доходів населення України за останні роки характеризується даними, наведеними в табл. 6.6. Негативним є збільшення квінтильного коефіцієнта фондів з 3,1 у 2014 р. до 3,5 у 2020 р. Це свідчить про значне збільшення диференціації доходів між 20 % найбільш і 20 % найменш забезпеченого населення. За останні два роки величина цього показника є незмінною, що характеризує стабільні темпи зростання доходів найбільш і найменш забезпеченого населення.

Рівень виробництва продукції тваринництва у сільськогосподарсяких підприємствах оцінюється показниками чисельності поголів’я худоби та продукції тваринництва для 100 га земельної площі. Це також показники інтенсивного розвитку.

Широко застосовуються показники інтенсивності і для характеристики ступеня досконалості виробництва, оснащення його новою технікою (показники фондо-, енерго- та механоозброєності праці, ступеня механізації (автоматизації) виробництва та робіт, коефіцієнт електрифікації технологічних процесів та ін.), характеристики використання обладнання та виробничих площ (скільки продукції припадає на одну гривню устаткування, коефіцієнт завантаження виробничої площі цеху чи ділянки), ефективності капітальних вкладень тощо. На відміну від інших відносних величин показники інтенсивності завжди виражають кількість одиниць сукупності, яка знаходиться в чисельнику (в її одиницях виміру) на одиницю сукупності, яка знаходиться в знаменнику.

## 6.8. Відносні величини порівняння

Оцінка рівня розвитку підприємства, галузей народного господарювання досягається за рахунок відносних величин порівняння чи наочності.

*Відносною величиною порівняння* називають співвідношення величин однойменних показників, що належать до різних об'єктів або різних територій. Порівняння однойменних рівнів відбувається за той самий період або на один і той же момент. Завдання обчислення цього виду відносних величин — визначити величину даного абсолютного показника порівнянням його з аналогічним показником за іншою сукупністю, що розвивається в інших умовах, інакше кажучи, різнохарактерними частинами явища.

Наприклад, порівнюються дані про чисельність жителів, лікарів, рівні доходів населення, обсяг валового внутрішнього продукту в різних країнах на ту саму дату, рівні собівартості однотипних виробів, що виробляються на різних підприємствах і т. п. Так порівнюють ціни (на ринку і в магазині) на картоплю у якомусь місті, приймаючи ціну ринку за базу порівняння. Можна ринкову ціну на молоко в Харкові прийняти за базу і віднести до неї ринкові ціни на молоко на ту ж дату в інших містах.

У статистиці поширене порівняння однойменних показників різних регіонів країни, різних областей, підприємств, міст тощо. Так, на 1 січня чисельність населення м. Харкові становила 1 433,9 тис. осіб, а населення Вінниці — 370,6 тис. людина. Можна сказати, що населення Харкова майже вчетверо більше, ніж у Вінниці.

Серед цих показників дуже важливе значення мають порівняльні дані щодо різних країн, зокрема, дані, що порівнюють рівень виробництва, чисельності населення та ін. Наприклад, відомо, що середньорічна чисельність зайнятих у 2019 р. становила в Україні 16,7 млн. осіб, у Польщі — 16,5 млн, у Казахстані — 8,8 млн, у Білорусії — 4,9 %. Якщо прийняти дані по Україні за 100 %, то чисельність населення у Польщі становила 98,2 %, у Казахстані — 52,4 %, у Білорусії — 29,2 %. Три останні цифри — відносні показники просторового порівняння.

Відносні величини порівняння дають наочне уявлення про співвідношення порівнюваних величин і порівняльну оцінку рівня розвитку регіонів країни за показником, що порівнюється, між окремими країнами. Відносні величини порівняння іноді називають відносними *величинами наглядності*.

Таблиця 6.7

**Середньомісячна заробітна плата в Україні та країнах ЄС у 2017 р.,**
млн дол. США

| Країни | Середньомісячна заробітна плата, дол. США | Відношення середньомісячної заробітної плати в Україні до середньомісячної заробітної плати в інших країнах |
|---|---|---|
| Україна | 267,1 | – |
| Австрія | 4 034,2 | 15,1 |
| Італія | 2 750,2 | 10,3 |
| Литва | 948,3 | 3,6 |

Відносні величини порівняння виражаються у вигляді кратного відношення (у разах, частках одиниці) або у відсотках.

Прикладом відносних показників просторового порівняння можуть бути дані табл. 6.7. Відношення середньомісячної заробітної плати в Україні складе: до середньомісячної заробітної плати в Австрії — $\frac{4034,2}{267,1} = 15,1$; до середньомісячної заробітної плати в Італії — $\frac{2750,2}{267,1} = 10,3$; до середньомісячної заробітної плати в Литві — $\frac{948,3}{267,1} = 3,6$. Таким чином, середньомісячна заробітна плата в Австрії у 2017 р. була вищою щодо середньомісячної заробітної плати в Україні того ж року у 15,1, у Литві — 3,6 раза.

Ці цифри також є відносними показниками просторового порівняння. Вони показують співвідношення рівня середньомісячної заробітної плати в Україні (її рівень прийнято за 1) та деяких країн ЄС.

Другий приклад. Маються дані про відношення чисельності безробітних в Україні та Німеччині у процентах. Назване співвідношення становило:

| | |
|---|---|
| В 2010 р. . . . . . . | 159,3 |
| В 2019 р. . . . . . . | 92,3 |

Ці цифри є також відносними показниками просторового порівняння, вони показують співвідношення рівня безробітних в Україні (її рівень прийнятий за 100 %) та Німеччини. Цифри свідчать, що якщо 2010 р. чисельність безробітних в Україні була на 59,3 % нижчою, ніж у Німеччині, то 2019 р. Україна відставала від Німеччини на 7,7 %.

Особливо часто до показників порівняння вдаються при зіставленні середніх за групами. Наприклад, якщо прийняти середню урожайність зернових та зернобобових культур у сільськогосподарських підприємств із площею до 100 га у 2020 р. за 100 %, то для підприємств із площею 1 000,01–2 000,00 га вона становитиме 153,1 %, а для підприємств площею понад 3 000,00 га — 187,4 %.

Відносні показники порівняння можуть обчислюватися з урахуванням як абсолютних, середніх, а й у основі показників інтенсивності. Так, кількість народжених у 2019 р. в Україні становила 8,1, в Австрії — 9,6, в Угорщині — 9,1, у Японії — 6,9 ‰ (табл. 6.8).

Якщо дані по Україні прийняти за 100 %, то кількість народжених в Австрії складе 118,5 % $\left(\frac{9,6}{8,1} \cdot 100\right)$, в Угорщині — 112,3% $\left(\frac{9,1}{8,1} \cdot 100\right)$, в Японії — 85,2% $\left(\frac{6,9}{8,1} \cdot 100\right)$. Ці показники порівняння застосовуються й у оцінки успіхів передових підприємств, бригад і людей. Якщо, наприклад, у передовий робітник за зміну виготовив 30 деталей, а інший — 20, можна сказати, що продуктивність праці передового працівника у півтора разу більше, чи вище на 50 %.

Таблиця 6.8

**Коефіцієнти народжуваності у 2019 р.**
(на 1000 осіб населення)

| Країни | Число народжених, ‰ | Відношення числа тих, хто народився в інших країнах, до тих, хто народився в Україні, % |
|---|---|---|
| Україна | 8,1 | – |
| Австрія | 9,6 | 118,5 |
| Угорщина | 9,1 | 112,3 |
| Японія | 6,9 | 85,2 |

Наведені проценти — відносні показники просторового порівняння, обчислені на основі інтенсивних показників.

## 6.9. Відносні величини структури

Для характеристики складу тієї або іншій сукупності використовують показники структури.

Відносні величини структури представляють собою співвідношення величини частини будь-якого цілого і величини цього цілого. Вони характеризують склад статистичної сукупності і відповідають питання, яку частку (чи питому вагу) у всій сукупності становлять окремі її частини. Обчислюються відносні величини структури в частках одиниці або у відсотках (відносні величини структури називають також відносними величинами частки, питомою вагою)[1].

За допомогою групувань статистична сукупність поділяється на групи за різними ознаками. Ці групи розрізняються за складом одиниць, що входять до їх складу. Отже, вони мають різну питому вагу у всій сукупності. Ці відносні показники частки й характеризуватимуть склад і структуру сукупності.

Для обчислення відносних величин структури необхідно розділити значення окремих елементів значення загального результату і помножити на 100:

$$d_i = \frac{f_i}{\sum f_i} \cdot 100, \tag{6.10}$$

де $d_i$ — відносний показник структури;

$f_i$ — окремі частини сукупності явища;

$\sum f_i$ — сума окремих частин сукупності.

При визначенні відносних величин структури порівнюваними величинами може бути чисельності окремих груп статистичної сукупності, чи обсяги їх ознак. За основу (базу) порівняння приймається загальний результат статистичної сукупності.

Зазвичай показники структури встановлюються з точністю до 0,1. Якщо ж наводяться питомі ваги всіх груп, їх сума повинна дорівнювати 100 %, навіщо іноді доводиться округляти окремі відносні величини (зазвичай, найбільші). Наприклад, на підприємстві працює 400 основних робітників та 200 допоміжних. У цьому випадку питома вага основних робітників становить 66,66 %, а допоміжних — 33,33 %, а з округленням — 66,7 % та 33,3 % відповідно.

Наведемо приклад розрахунку відносних величин структури. Колектив великого промислового підприємства станом на 15 жовтня 2021 р. складається з 4 000 осіб, з них робітників — 3 200 осіб, інженерно-технічних працівників — 380 осіб, службовців — 80 осіб, молодшого обслуговуючого персоналу — 120 осіб та учнів, персоналу охорони — 220 осіб. Для розрахунку відносних величин структури зіставимо чисельність окремих категорій працівників $f_i$ із загальним

---

[1] Якщо при віднесенні частини до цілого останнє прийнято за 1, відносний показник структури зазвичай називають часткою, якщо ціле прийнято за 100 — питомою вагою. Небезпідставно іноді питомою вагою називають і частки.

підсумком $\sum f_i$, тобто визначимо їх питому вагу у процентах $\left(\frac{f_i}{\sum f_i} \cdot 100\right)$. Структура всього персоналу буде обчислена так (табл. 6.9).

Таблиця 6.9

**Розрахунок структури персоналу підприємства**

| Категорії працівників | Кількість осіб | В % до загальної чисельності |
|---|---|---|
| Робітники | 3 200 | $\frac{3200}{4000} \times 100 = 80$ |
| Інженерно-технічні робітники | 380 | $\frac{380}{4000} \times 100 = 9{,}5$ |
| Службовці | 80 | $\frac{80}{4000} \times 100 = 2{,}0$ |
| Молодший обслуговуючий персонал | 120 | $\frac{120}{4000} \times 100 = 3{,}0$ |
| Учні, персонал охорони | 220 | $\frac{220}{4000} \times 100 = 5{,}5$ |
| Всього | 4 000 | 100,0 |

Обчислені показники (граф. 3) характеризують питому вагу (тобто частку) складових частин цілого у загальній чисельності персоналу. З даних таблиці слідує, що 80 % усієї чисельності персоналу підприємства станом на 15 жовтня 2021 р. складали робітники, 9,5 % — інженерно-технічні працівники, 2,0 % — службовці. Сума відносних величин структури досліджуваної сукупності, завжди дорівнює 100 % або одиниці.

До показників структури відносять відношення чисельності різних категорій промислово-виробничого персоналу до його загальної чисельності; відношення окремих видів доходів та видатків до загальної суми доходів та видатків державного бюджету; відношення окремих статей витрат до загальної суми витрат на виробництво будь-якої продукції; відношення площі сільськогосподарських культур до загальної посівної площі; ставлення викладачів з освіти та стажу науково-педагогічної роботи тощо.

Відмінною рисою показників структури є те, що вони можуть підсумовуватися на відміну показників інтенсивності, які підсумовуванню не підлягають. Ця операція тут можлива тому, що ці показники обчислюються до однієї бази.

З допомогою відносних показників не тільки виявляється структура досліджуваної сукупності, а й структурні зрушення. Обчислені за кілька періодів (моментів) часу, вони дають уявлення про зміни структури, званих *структурним зрушенням*.

У статистичних показниках, що характеризують зміну структури явищ у часі, відображаються найважливіші закономірності розвитку. Процес уповільнення економічного зростання країни супроводжувався зниженням частки валового нагромадження у валовому внутрішньому продукті країни. Якщо порівняти наведені показники 2020 р. з показниками попередніх періодів, то побачимо, що в 2020 р. питома вага валового нагромадження у вартості валового внутрішнього продукту склала 7,5 %, а в 2020 р. — 20,9 %. У той самий час частка споживання домашніх господарств становив відповідно 73,4 % і 63,0 %.

Показники структури характеризують не тільки відношення чисельності одиниць сукупності, а й розподіл ознак за групами. Саме у різному співвідношенні цих ознак часто проявляються характерні риси того чи іншого розподілу.

Як приклад наведемо таку таблицю (табл. 6.10).

Таблиця 6.10

**Групування підприємств за розмірами зібраної площі зернових та зернобобових культур у 2020 р.**[1]

| | Кількість підприємств | | Обсяг виробництва | | Урожайність, ц з 1 га |
|---|---|---|---|---|---|
| | одиниць | у % до загальної кількості | тис. т | у % до загального обсягу виробництва | |
| Підприємства | 32 513 | 100,0 | 51 718,0 | 100,0 | 46,4 |
| з них з площею, га | | | | | |
| до 100,00 | 19 026 | 58,5 | 1 934,0 | 3,7 | 30,2 |
| 100,01–200,00 | 3 559 | 10,9 | 1 899,5 | 3,7 | 36,6 |
| 200,01–500,00 | 4 213 | 13,0 | 5 410,0 | 10,5 | 40,0 |
| 500,01–1000,00 | 2 765 | 8,5 | 8 300,4 | 16,1 | 42,2 |
| 1 000,01–2 000,00 | 1 880 | 5,8 | 11 694,6 | 22,6 | 46,3 |
| 2 000,01–3 000,00 | 566 | 1,7 | 6 694,1 | 12,9 | 48,6 |
| більше 3 000,00 | 504 | 1,6 | 15 785,4 | 30,5 | 56,6 |

[1] Обсяг виробництва наведено у масі після доробки.

*Джерело*: Групування підприємств за розмірами зібраної площі основних сільськогосподарських культур у 2020 році/ Державна служба статистики України. URL: http://www.ukrstat.gov.ua/

Таблиця показує диференціацію урожайності залежно від рівня концентрації посівних площ зернових та зернобобових культур у сільськогосподарських підприємствах країни. Складаючи 1,6 % загальної чисельності, великі сільськогосподарські підприємства забезпечили 30,5 % всього збору зерна, при цьому урожайність у них була значно вищою, ніж у сільськогосподарських підприємствах із малою площею посіву.

За допомогою відносних показників структури характеризується зміна вікового складу населення в Україні, структури промислової продукції, вартості основних засобів, джерел фінансування капітальних інвестицій і т. п. При цьому потрібно мати на увазі, що зниження питомої ваги далеко не завжди означає зменшення абсолютного розміру цієї частини сукупності.

## 6.10. Відносні величини рівня економічного розвитку

*Відносними величинами рівня економічного розвитку* називають показники, що характеризують обсяги виробництва різних видів продукції душу населення. При обчисленні їх необхідно річний обсяги виробництва даного виду продукції розділити на середньорічну чисельність за той же період.

Приклад. Маються дані про виробництво деяких видів промислової продукції в Україні у натуральному вираженні у 2020 р.:

| | |
|---|---|
| Вугілля кам’яне, млн т . . . . . | 24,2 |
| Нафта сира, млн т . . . . . . . | 1,7 |
| Борошно, млн т . . . . . . . . | 1,5 |
| Електроенергія, млрд кВт·год. . . | 147,8 |
| Гас природний, млрд м$^3$ . . . . . | 20,2 |

Необхідно в и з н а ч и т и відносні показники рівня економічного розвитку країни, знаючи, що чисельність населення України на початок 2020 р. становила 41,7 млн осіб та на початок 2021 р. — 41,6 млн осіб. Спочатку визначимо середньорічну чисельність населення за 2020 р.:

$$\frac{41,7+41,6}{2}=41,65 \text{ млн осіб.}$$

Розділивши обсяги виробництва кожного виду продукції на 41,65 млн осіб, отримаємо, що на душу населення припадало: 581 кг кам’яного вугілля, 40,8 кг нафти, 36 кг борошна, 3549 кВт·год. електроенергії, 485 м$^3$ природного газу.

## ЛІТЕРАТУРА


1. *Горкавий В. К.* Статистика: Підручник. Третє вид., переробл. і доповн. Київ: Алерта, 2019. С. 81–87.
2. Статистика: Підручник / *С. С. Герасименко*, *А. В. Головач*, *А.М. Єріна* та ін.; За наук. ред. д-ра екон. наук С. С. Герасименка. 2-ге вид., перероб. і доп. Київ: КНЕУ, 2000. С. 41–46.
3. *Ткач Є. І.*, *Сторожук В. П.* Загальна теорія статистики: підручник. 3-тє вид. Київ: Центр учбової літератури, 2009. С. 113–130.


## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники, словники*


4. Мармоза А. Т. Теорія статистики: підручник. 2-ге вид., перероб. і доп. Київ: «Центр учбової літератури», 2013. 333–396.
5. *Опря А. Т.* Статистика (модульний варіант з програмованою формою контролю знань). Навч. посіб. Київ: Центр учбової літератури, 2012. С. 66–73.
6. *Педченко Г. П.* Статистика: Навчальний посібник. Мелітополь: Колор Принт, 2018. С. 92–99.
7. Статистичний словник / [*О. Г. Осауленко*, *О. О. Васєчко*, *М. В. Пугачова* та ін.]; за ред. д-ра держ. упр., проф., член-кор. НАН України О. Г. Осауленка; НТК статистичних досліджень. Київ: ДП «Інформ.-аналіт. агентство», 2012. 498 с.
8. *Шапочка М. К.*, *Маценко О. М.* Теорія статистики: навч. посібник. Суми: Університетська книга, 2014. С. 109–117.


*Монографії та статті*


9. *Чекотовський Е. В.* Абсолютні статистичні показники: історичні та теоретичні аспекти. Частина І. Зародження теорії абсолютних статистичних показників. *Статистика України*. 2012. № 1. С. 74–78.
10. *Чекотовський Е. В*. Відносні статистичні показники: історія та теорія. Частина ІІ. Еволюція понять відносних статистичних показників, форм їх вираження та принципів побудови і застосування. *Статистика України*. 2013. № 4. С. 42–47.
11. *Чекотовський Е. В*. Відносні статистичні показники: історія та теорія. Частина ІІІ. Еволюція видів відносних статистичних показників та їх класифікації. *Статистика України*. 2014. № 1. С. 32–37.

# Г Л А В А 7

# СЕРЕДНІ ВЕЛИЧИНИ

## 7.1. Середня, її сутність і визначення

**Значення середньої у статистиці**. Для вивчення масових суспільних явищ поряд з абсолютними і відносними показниками необхідні узагальнюючі показники, до яких належать середні величини. Середні величини відіграють важливу роль у статистиці. За допомогою середніх величин можна охарактеризувати сукупність за кількісною ознакою, що варіюється.

Метод середніх, взятий у його загальній формі, як і метод угруповань, є специфічною особливістю статистичної методології. Припустимо, ми хочемо порівняти середній виробіток продукції по двох однотипних підприємствах. Для цього ми не можемо порівнювати виробіток двох робітників, взятих з різних підприємств, оскільки ця ознака суттєво варіюється, та інші робітники можуть мати інший виробіток. Виробіток взятих робітників, наприклад, передовиків чи відстаючих, може бути не типовим для цих підприємств. Ми не можемо порівняти і загальний обсяг виготовлених виробів, оскільки на підприємствах може виявитися різною чисельність робітників. Якщо ж ми розділимо ці суми на чисельність робітників, тобто обчислимо середній виробіток на одного робітника, то ми можемо їх порівнювати і сказати, на якому підприємстві рівень продуктивності вищий.

Виробіток продукції досліджуваної сукупності робітників отримує узагальнену характеристику середньої величини. У середній виражається те загальне, що притаманно всієї сукупності робітників щодо досліджуваної ознаки (виробітку продукції).

За допомогою середніх проводиться багато аналітичних досліджень суспільних явищ та процесів. У багатьох випадках показники розвитку народного господарства країни загалом чи з окремих галузей визначаються у багатьох випадках у вигляді середніх величин.

У статистичній практиці середня найчастіше застосовується не як характеристика розвитку будь-якого процесу, а як характеристика сукупності на певний момент (або період) часу. Наприклад, середня заробітна плата працівників промисловості України у 2022 р., середні ціни продукції сільського господарства, реалізовані підприємствами у 2022 р., середня урожайність зернових культур у 2022 р., середня тривалість життя, середній розмір пенсії тощо. На цих прикладах середня пов'язана з сукупністю явищ, що існують у певний час.

Середні величини широко використовуються у плануванні та аналізі економічних процесів. Так, у «Стратегічному плані діяльності Міністерства економіки України на 2022–2024 роки» у розділі «Створення сприятливого для бізнесу інституційного середовища» визначено, що середньомісячна заробітна плата на 2023 рік зросте до 17159 грн, або на 48,0 % проти 2020 р.[1]

Ряд народногосподарських завдань на перспективу також представлені у вигляді середніх показників, наприклад, довести темпи приросту фізичного обсягу експорту товарів та послуг на 2022–2024 роки у середньому на 5 %, забезпечити зростання номінального ВВП на душу населення у 2025–2030 рр. не менше ніж до 10 тис. доларів США і т.п.

Велике значення середні величини мають при розробці поточних і перспективних планів розвитку підприємств. Часто при плануванні кількісних показників (продукції, запасів сировини, необхідного обладнання тощо) середні величини використовуються як середні норми (витрат матеріалу, палива, використання обладнання тощо), на базі яких ці дані визначаються. Ряд середніх міститься у самих планових показниках (середня тривалість виробничого циклу виготовлення виробу, середня урожайність з 1 га тощо). У цих випадках середні виступають вже як самостійно планований і контрольований вираз тієї чи іншої закономірності у розвитку підприємств.

**Визначення середньої**. *Середньою величиною* в статистиці називається узагальнюючий показник, що виражає типові розміри кількісно варіюючих ознак якісно однорідних суспільних явищ. Вона дає зведену кількісну характеристику ознак масового процесу. Застосування середніх величин дозволяє виявити характерні, типові особли-

[1] Стратегічний план діяльності Міністерства економіки України на 2022–2024 роки. Міністерство економіки України. URL.: www.me.gov.u.

вості досліджуваної сукупності за тими чи її ознаками.

Статистичні середні величини виражають типові розміри та кількісні відносини об'єктивно існуючих властивостей суспільних явищ. К. Маркс зазначив: «... середня величина є завжди середня багатьох різних індивідуальних величин одного й того ж виду» (т. 23, с. 334). Дія різноманітних факторів породжує коливання, варіацію ознаки, що усереднюється. Середня величина є загальною мірою їхньої дії, що рівнодіючою всіх цих факторів. Вона характеризує сукупність за ознакою, що підлягає усередненню, але відноситься до одиниці сукупності.

Середня величина є зведеною, узагальнюючою характеристикою сукупності. Але для того, щоб середня правильно характеризувала цю сукупність, необхідно, щоб її елементи мали однорідні властивості, одними і тими ж ознаками. За всіма іншими ознаками одиниці, що входять до статистичної сукупності, можуть бути якісно різними. Наприклад, при обчисленні середньої урожайності потрібно, щоб вихідні дані спостереження враховувалися або за однією культурою або групою подібних культур. Середня перестає виконувати роль такого узагальнюючого показника, який відображає закономірність явища, якщо ця середня обчислена для сукупності фактів, що не можуть внаслідок своєї різнорідності розглядатись у загальному ряді.

Навіть у тому випадку, якщо сукупність більш менш однорідна, рекомендується доповнювати обчислення загальних середніх величин груповими середніми. Це можна пояснити на такому прикладі. Не можна обмежитися обчисленням середньої заробітної плати в цілому по Україні. В окремих областях або за видами економічної діяльності різниця між середньою заробітною платою може істотно відрізнятися. Так, якщо у січні 2022 р. середньомісячна заробітна плата в Україні становила 14 577 грн., тоді як у м. Києві — 21 347 грн., у Волинській області — 11 735 грн. У даному випадку середні величини необхідно доповнити груповими середніми. Те саме можна сказати про вироблення продукції в різних групах підприємств, професійних груп робітників і т.п.

Середня має ту хорошу властивість, що в ній погашаються індивідуальні особливості масових явищ і усуваються випадкові коливання. Внаслідок цього взаємопогашення у середній виявляються загальне, закономірне, властиве даної сукупності явищ. Наприклад, висока урожайність у середньому району може приховувати низьку величину урожайності окремих господарств. Але це хороша властивість перетворюється на протилежність, коли середня визначається за сукупністю, елементи якої не мають однорідних властивостей. І тут у ній погашатимуть як випадкові відхилення, а й істотні відхилення окремих величин від основного типу.

Але процес погашення випадкових відхилень не можна трактувати як суто математичне наслідок формули простої арифметичної середньої. Середня величина, обчислена виходячи з великої кількості випадків, представляє кількісний результат принципу взаємопогашення випадковостей лише тому випадку, якщо береться велика маса явищ. Обчислення середньої повинно ґрунтуватися лише в масі випадків, пробиваючись, так би мовити, «крізь натовп випадковостей», спирається на «усереднення» різноманітних тенденцій. У середній виявляється основна тенденція, закономірність. Для цього середню необхідно обчислювати на основі великої кількості фактів. Тим самим було середня величина має безпосереднє відношення до закону великих чисел.

Середня величина є загальною мірою їхньої дії, що рівнодіючою всіх цих факторів. Середня величина характеризує сукупність за ознакою, що підлягає усередненню, але відноситься до одиниці сукупності. Прийом середніх дозволяє отримати найбільш узагальнену кількісну характеристику ознаки якісно однорідної сукупності і виміряти їх коливання навколо середнього рівня розвитку. При обчисленні середньої виходять із припущення, що коливання ознаки носить характер випадковості по відношенню до тих загальних умов, що визначають характер розподілу. Це обґрунтування здійснюється на основі аналізу варіаційного ряду.

Будь-яка статистична сукупність має низку властивостей, які характеризуються певними ознаками. За допомогою середньої ми прагнемо дати характеристику якійсь одній властивості досліджуваній сукупності і саме таку властивість, ознаки якого виражені кількісно. Такою кількісною ознакою може бути, наприклад, розмір вироблення продукції на одного робітника, чисельність населення певної території, дохід сім'ї, успішність студентів, розмір урожайності певної культури, удій молока з однієї корови і т. п.

Але у деяких випадках для ознак, виражених кількісно, середню величину не завжди потрібно обчислювати. Немає потреби обчислювати середню, коли величина ознаки кожної одиниці сукупності однакова. Наприклад, відомо, що 200 студентів 1-го курсу одержують стипендію у розмірі 1450 грн на місяць, тобто розмір стипендії для всіх студентів однаковий. І тут немає необхідності обчислювати середню. Середню обчислюють у випадках, коли значення ознак одиниць сукупності відрізняються, або, як кажуть, коливаються.

У *широкому розумінні* терміна під середньою величиною мається на увазі будь-яка узагальнена кількісна характеристика ознаки в статистичній сукупності (середня урожайність, середня заробітна плата, середній вік обладнання, середня собівартість одиниці виробленої продукції, середня ціна реалізації продукції і т. д.).

Необхідно врахувати, що обчислення середньої величини має ґрунтуватися на сутності цього показника. Кожен показник має свій, тільки йому притаманний зміст. Наприклад,

$$\text{Середня урожайність} = \frac{\text{Валовий збір}}{\text{Посівна площа}};$$

$$\begin{matrix}\text{Середній час, витраче-}\\ \text{ний на одну деталь}\end{matrix} = \frac{\text{Весь витрачений час}}{\text{Число деталей}};$$

$$\begin{matrix}\text{Середня заробітна}\\ \text{плата одного працівника}\end{matrix} = \frac{\text{Фонд заробітної плати}}{\text{Чисельність працівників}};$$

$$\begin{matrix}\text{Середня собівартість}\\ \text{одиниці продукції}\end{matrix} = \frac{\text{Загальні витрати на продукцію}}{\text{Кількість продукції}} \text{ і т. д.}$$

Такий підхід дозволяє правильно визначити середню величину ознаки, вибрати форму середньої. На відміну від математичних середніх, які є абстрактними величинами, статистичні середні завжди характеризують конкретні суспільні явища або процеси.

Найважливішим соціально-економічним завданням, що вирішуваної за допомогою середніх, є характеристика типів явищ або загальних умов, загальних закономірностей. Але це загальне завдання має бути конкретизовано, виражено у більш окремих завданнях. У статистиці можна виділити кілька основних задач, які вирішуються за допомогою середніх:

1) визначення рівня розвитку суспільних явищ;
2) порівняння двох або кількох рівнів;
3) аналіз динаміки явищ у часі та у просторі;
4) виявлення та характеристика зв’язків явищ;
5) здійснення розрахунків та оцінок зв’язку у зв’язку з плануван-

ням та прогнозуванням розвитку суспільних явищ і процесів.

Середні величини набули широкого поширення не тільки у статистиці, а й у багатьох інших науках, в планової, облікової, науково-дослідницької діяльності. У статистиці середні мають винятково важливе значення. Теорія середніх величин займають одне з центральних місць у загальній теорії статистики.

Численні прийоми і показники математичної статистики безпосередньо пов'язані з обчисленням середніх величин. Дійсно, ступінь коливання одиниць сукупності характеризується шляхом порівняння кожної індивідуальної ознаки із середньою величиною. Чим ближче значення ознаки до середньої (середньої арифметичної величини), тим менший ступінь коливання одиниць сукупності, і чим різкіша різниця між окремими ознаками, тим сильніше варіює ряд. В теорії ймовірностей і в практичному використанні останньої велике значення має вимір відхилень. Для вимірювання відхилень використовують середнє квадратичне відхилення, яке є не чим іншим, як середньою квадратичною з відхилень окремих варіантів досліджуваної величини від їх середньоарифметичного значення. Формули агрегатних і середніх індексів представляють, власне кажучи, характеристику зміни середнього рівня показника порівняно з рівнем, прийнятим за базу порівняння. Розрахунок параметрів рівняння регресії при виявленні та характеристики загальних тенденцій у розвитку соціально-економічних явищ є не що інше, як знаходження середньої лінії розвитку тренду. Кількісний вимір зв'язків між явищами (показник кореляції) ґрунтується на оцінці взаємовідносин між середніми показниками.

Складність та взаємопов'язаність суспільних явищ зумовлює виняткову важливість застосування методу групувань у їх дослідженні. Середні перетворюються на потужний засіб аналізу, підпорядкований методу групувань. За допомогою методу групувань можна дати характеристику соціально-економічних типів явищ, а сукупність середніх дає кількісну оцінку даних типів. Так, середній обсяг доходу, середня вартість основних засобів, середня чисельність персоналу та інші середні показники дають загальне уявлення про тип підприємства.

## 7.2. Наукові засади обчислення середніх показників

Використовуючи середні величини в економічних дослідженнях, необхідно дотримуватись певних умов їх застосування.

*Перша та найважливіша* умова наукового застосування середніх величин у статистиці полягає в тому, перш ніж обчислювати середню величину, необхідно переконатися в тому, що ознаки, які підлягають аналізу, відносяться до *якісно однорідних*, *однотипних сукупностей*. Тільки за дотримання цієї умови середні величини виражають характерну, типову величину ознаки в одиниць сукупності. Ця умова дуже точно була сформульована К. Марксом у «Капіталі». Враховуючи, що праця, уречевлена у вартості, є праця середньої суспільної якості, тобто прояву *середньої* робочої сили, Маркс пояснює: «Але середня величина є завжди середньою багатьох різних індивідуальних величин *одного і того ж виду.* У кожній галузі промисловості індивідуальний робітник Петро чи Павел більш менш відхиляються від середнього робітника. Такі індивідуальні відхилення погашаються і знищуються, якщо ми беремо значну кількість робітників».

Значить, щоб середня величина правильно надавала узагальнюючу характеристику сукупності однотипних явищ за якоюсь із її ознак, необхідно мати у розпорядженні дані *про значення цієї ознаки у маси одиниць сукупності*. У такому разі погашаються індивідуальні відмінності одиниць сукупності у значеннях ознаки, що підлягає усередненню. Це означає, що взаємне погашення індивідуальних відхилень, що не порушує загальних умов для всієї сукупності, можливе лише за якісної її однорідності. Властивість середніх відображати загальні, типові розміри досліджуваної ознаки маси одиниць реально існуючої сукупності і розкривати їх загальні закономірності за умови взаємопогашення випадкових коливань величини ознаки окремих одиниць сукупності К. Маркс назвав *законом середніх чисел*.

Якщо сукупність різноякісна, то явища підпорядковані різним законам розвитку ознаки, що підлягає усередненню. У цьому випадку середні характеристики отримуються нетиповими та середня втрачає визначеність свого змісту. Використання таких середніх не тільки не збагачує наше пізнання реальної дійсності, а, навпаки, ускладнюють це пізнання, оскільки в таких середніх затушовується дійсність. Середня величина *для маси одиниць будь-якого складу* з різними властивостями по-суті *не має реального змісту*.

Якісна однорідність суспільних явищ, їхня однотипність встановлюється на основі теоретичного дослідження сутності та закономірності цих явищ. Такий теоретичний аналіз має розкрити природу соціально-економічних явищ, закони розвитку. Базуючись на цій теорії, статистики виділяють якісно однорідні сукупності, які розвиваються в загальних для всіх одиниць умовах часу та місця за допомогою методу групувань. Це означає, що обчисленню середніх величин має передувати статистична групування, що поділяє досліджувану сукупність одиниць на якісно однорідні групи. Саме тому науковою основою застосування середніх величин є метод статистичних групувань, тобто розчленування сукупності на якісно однорідні групи. Статистичне групування відкриває можливість реально вивчати типи та форми розвитку досліджуваного суспільного явища, зокрема, процес розвитку народного господарства країни, галузей матеріального виробництва, підприємств та їх окремих ділянок.

Така перша, основна умова наукового застосування середніх величин. Однак цього ще замало.

При глибшому науковому дослідженні не можна обмежуватися обчисленням середніх величин. Середня, будучи узагальнюючою характеристикою сукупності однотипних явищ загалом, приховує відмінність індивідуальних значень ознаки в окремих одиниць, що підлягає усередненню, затушовує особливості у розвитку. Тим часом, ці відмінності навіть у межах якісно однорідної сукупності здебільшого носять не випадковий, а систематичний характер. Крім того, у реальній дійсності безперервно відбувається народження нового та відмирання старого. Але носіями нового та відмирання старого можуть бути нечисленні, а в окремих випадках і поодинокі одиниці сукупності. Тому загальні середні потрібно доповнювати розрахунком середньої кожної групи, виділеної із сукупності, тобто групових середніх. Без знання групових середніх загальні середні можуть привести нас в оману і не розкрити справжніх закономірностей досліджуваних процесів.

Припустимо, вивчається урожайність сільськогосподарської культури у двох районах. Загальна урожайність зернових культур у другому районі виявляє тенденцію до зниження (табл. 7.1).

Але, очевидно, урожайність окремих зернових культур у другому районі вища, ніж у першому районі. Обчислюючи середні за окремими культурами та аналізуючи динаміку не тільки загальної середньої, а й групових середніх, можна виявити, що за всіма зерновим культурами у другому районі урожайність зернових культур зростала, або не змінювалася, а зниження загальної середньої в цілому було обумов-

Таблиця 7.1

| | Перший район | | | Другий район | | |
|---|---|---|---|---|---|---|
| | площа, га | урожайність, ц з га | валовий збір, ц | площа, га | урожайність, ц з га | валовий збір, ц |
| Пшениця | 500 | 30 | 15 000 | 100 | 35 | 3 500 |
| Ячмінь | 100 | 25 | 2 500 | 250 | 25 | 6 250 |
| Горох | 50 | 20 | 1 000 | 400 | 21 | 8 400 |
| Всього | 650 | 28,5 | 18 500 | 750 | 24,5 | 18 150 |

лено зростанням питомої ваги зернових культур із нижчою урожайністю. Зрозуміло, що у цьому прикладі динаміка групових середніх правильніше відобразила зміна урожайності, а динаміка загальної середньої повністю не розкриває цю закономірність.

Як бачимо, і ця вимога наукового методу середніх призводить по суті до того, що середні повинні застосовуватися в органічному зв'язку з групуваннями.

Розглянемо ще приклад. Припустимо, ми порівнюємо два підприємства з однаковою чисельністю робітників за рівнем заробітної плати (табл. 7.2).

Таблиця 7.2

| | Перше підприємство | | | Друге підприємство | | |
|---|---|---|---|---|---|---|
| | кількість робітників | фонд заробітної плати, тис. грн | середня заробітна плата, тис. грн | кількість робітників | фонд заробітної плати, тис. грн | середня заробітна плата, тис. грн |
| 1. Перевиконують норму | 125 | 1 750 | 14 | 50 | 725 | 14,5 |
| 2. Виконують норму | 350 | 3 850 | 11 | 400 | 4 600 | 11,5 |
| 3. Не виконують норму | 25 | 150 | 6 | 50 | 325 | 6,5 |
| Всього | 500 | 5 750 | 11,5 | 500 | 5 650 | 11,3 |

На першому підприємстві чисельність робітників становила 500 осіб, а місячний фонд заробітної плати — 5 750 тис. грн. Отже, середня місячна заробітна плата одного робітника становила 11,5 тис. грн. Друге підприємство має 500 робітників, але місячний фонд заробітної плати тут становить 5 650 тис. грн. Отже, місячна заробітна плата робітників на першому підприємстві трохи вище, ніж на другому (11,5 – 11,3).

Але розглянемо показники заробітної плати, обчислені за окремими групами робітників за ступенем виконання планового завдання. В залежності від ступеня виконання планового завдання виділимо три групи робочих: 1-а група — перевиконують норму, 2-а група — виконують норму і 3-я група — не виконують норму. Виявляється, що за окремими групами робітників місячна заробітна плата у другому підприємстві виявилася дещо вищою, ніж у першому. Це пояснюється тим, що у другому підприємстві питома вага робітників, які перевиконують норму, була дещо нижчою (10 %), ніж у першому підприємстві (25 %). Тоді як частка робітників, які виконують норму (10 %), була вищою, ніж у першому підприємстві (5 %). Зрозуміло, що у цьому прикладі динаміка групових середніх дозволила зробити правильний висновок про зміну типового ознаки, досягнутого у той чи інший період.

Отже, найважливіша вимога наукового методу середніх у тому, що загальні *середні потрібно доповнювати обчисленням групових середніх*. При цьому групи повинні бути утворені за ознаками, залежно від яких систематично змінюється ознака, що підлягає усередненню.

Наступна вимога, що висувається до обчислення середніх, — *правильний вибір явища* (*одиниць сукупності*), *на основі якого розраховується середня*.

У більшості випадків середня обчислюється шляхом ділення обсягу ознаки, взятої за сукупності явищ, на число явищ (одиниць), що мають цю ознаку. Наприклад, середній розмір діаметра виробу розраховується виходячи із загальної кількості виготовлених виробів, середня заробітна плата обчислюється виходячи із загальної кількості працівників і т. д. Керуючись цим положенням, правильніше було б обчислювати середні не за всіма одиницями, а тільки за тими, які на даний момент характеризуються певною ознакою. З іншого боку, це положення не завжди дотримується. Наприклад, у сільськогосподарських підприємствах середньорічний удій молока одну корову обчислюють шляхом поділу загального надою молока не на кількість дійних корів, а на загальну кількість фуражних корів. Середню заробітну плату[1] одного штатного працівника облікового складу в українській

---

[1] Див.: Інструкція зі статистики кількості працівників: Наказ Державного комітету статистики України від 28.09.2005 р. № 286. URL: https://zakon.rada.gov.ua/laws/show/z1442-05#Text.

статистиці визначають діленням суми нарахованого фонду оплати праці штатних працівників не на середньооблікову чисельність працівників, які працювали за відповідний період (місяць, квартал, півріччя, рік), на середньооблікову кількість усіх працівників, включаючи тих, які не працювали з причинами простою. У наведених прикладах число недійних корів і кількість робітників, які фактично з'явилися, але не працювали через простої можна вважати з нульовими значеннями ознак. За такого підходу порушення процедури обчислення середніх не відбувається.

Найважливіша вимога наукового підходу до середніх — *обчислення їх по всьому колу одиниць досліджуваної сукупності або за типовою частиною їх*. Якщо середня розрахована не за всіма одиницями явища, а за деякою частиною, то ця частина має бути репрезентативною. Репрезентативність визначає, наскільки можливо узагальнювати результати дослідження із залученням певної вибірки на всю генеральну сукупність, з якої вона була зібрана. Мінімальна кількість одиниць, якою можуть бути отримані досить представницькі середні, становить 20–30 одиниць. Якщо одиниці явища мають незначні коливання, то середня може бути обчислена на основі невеликого числа одиниць. Чим більше відрізняються один від одного одиниці сукупності, тим більше (щодо обсягу сукупності) має бути обрано одиниць для отримання задовільних характеристик середньої величини.

Обчислюючи середні, необхідно пам'ятати, що вони дають узагальнюючу кількісну характеристику статистичної сукупності лише за однією ознакою. Насправді статистичні сукупності мають різноманітні кількісні ознаки, що характеризують ті чи інші їх сторони розвитку. Тому з метою глибокого аналізу масових суспільних явищ рекомендується обчислювати не одну середню, а *систему середніх*. Використання системи середніх величин дозволить охарактеризувати явища з різних сторін. Крім того, з метою наукової обґрунтованості висновків сама система середніх має застосовуватись у комплексі з іншими статистичними показниками.

## 7.3. Види середньої

Досі ми розглянули загальні методичні питання, що стосуються середніх. І це правильно, оскільки середня дозволяє дати узагальнену характеристику статистичної сукупності. Однак, якщо розглянути ці показники детальніше, то можна побачити, що середні поділяються на ряд видів.

У математичній статистиці розрізняють кілька видів середніх величин: *середня арифметична*, *середня гармонійна*, *середня геометрична*, *середня квадратична*, *середня кубічна* та ін. У першому випадку середню називають простою, у другому — зваженою.

Види середніх, які застосовуються в статистиці, можуть бути алгебраїчно об'єднані загальним поняттям *степеневої середньої*. Якщо маються індивідуальні значення ознаки (варіанти) $x_1^z$; $x_2^z$; $x_3^z$; . . . ; $x_n^z$, то середня з варіантів може бути обчислена за формулою простої незваженої середньої середнього порядку:

$$\bar{x} = \sqrt[k]{\frac{\sum x_i^k}{n}}, \tag{7.1}$$

де $\bar{x}$ — степенева середня (читається «ікс» з рисочкою);

$x_i$ — змінюючи величини ознаки (варіанти);

$n$ — число спостережень;

$k$ — показник ступеня, що визначає тип середньої.

За наявності відповідних частот $f_1$; $f_2$; $f_3$; . . . ; $f_n$ середня обчислюється за формулою зваженої степеневої середньої:

$$\bar{x} = \sqrt[k]{\frac{\sum x_i^k f_i}{\sum f_i}}, \tag{7.2}$$

де $f_i$ — частоти, або статистичні ваги варіантів.

Зміна значення показника рівня середньої ($k$) визначає вид середньої величини. Зокрема, якщо у формулі степеневої середньої $k$=1, отримують середню арифметичну:

$$\bar{x}_a = \frac{\sum x}{n}. \tag{7.3}$$

Якщо $k = -1$, отримаємо середню гармонійну[1]:

$$\bar{x}_h = \frac{n}{\sum \frac{1}{x}}. \tag{7.4}$$

Якщо у формулі степеневої середньої $k$=0, отримаємо середню

[1] Доказ. Якщо $k = -1$, статечна середня перетворюється на середню гармонійну:

геометричну[1]:

$$\bar{x}_g = \sqrt[n]{\prod x}, \tag{7.5}$$

де $\prod$ — знак множення.

Якщо $k$=2, то отримаємо формулу середньої квадратичної:

$$\bar{x}_q = \sqrt{\frac{\sum x^2}{n}}. \tag{7.6}$$

Зі степеневих середніх у статистиці найчастіше застосовується середня арифметична, рідше — середня гармонійна; середня геометрична застосовується обчислення середніх темпів динаміки розвитку явищ, а середня квадратична — лише при обчисленні показників варіації.

Усі середні розрізняються за своєю абсолютною величиною, так, для ряду 1, 2, 3:

Середня арифметична — $\bar{x}_a = \frac{\sum x}{n} = \frac{1+2+3}{3} = 2;$

» гармонійна — $\bar{x}_h = \frac{n}{\sum \frac{1}{x}} = \frac{3}{\frac{1}{1}+\frac{1}{2}+\frac{1}{3}} = 1{,}64;$

» геометрична — $\bar{x}_q = \sqrt[n]{\prod x} = \sqrt[3]{1 \cdot 2 \cdot 3} = 1{,}82;$

» квадратична — $\bar{x}_g = \sqrt{\frac{1^2+2^2+3^2}{3}} = 2{,}08.$

У всіх випадках величини середніх сильно відрізняються. Незважаючи на це, між величинами наведених середніх зберігається певне співвідношення: $\bar{x}_g > \bar{x}_a > \bar{x}_q > \bar{x}_h$. У всіх випадках порівняльна величина наведених середніх збережеться така ж, тобто найменша буде гармо-

---

$$\bar{x} = \left(\frac{\sum x^{-1}}{n}\right)^{-1} = \left(\frac{\sum \frac{1}{x}}{n}\right)^{-1} = \frac{n}{\sum \frac{1}{x}}.$$

[1] Доказ. Якщо $k$=0, степенева функція перетворюється $\bar{x} = \left(\frac{\sum x^0}{n}\right)^{\frac{1}{0}} = \left(\frac{n}{n}\right)^{\frac{1}{0}} = 1^{\infty}$. При розкритті цієї невизначеності за способом Лопіталя отри маємо $\bar{x}_g = \sqrt[n]{x_1 x_2 \ldots x_n}$. Це не що інше, як формула так званої середньої геометричної.

нійна середня, потім геометрична і т. д.

Крім степеневих середніх у статистиці застосовують показники, що характеризують розподіл варійованої ознаки: моду (найчастіше зустрічається варіанта) і медіану (серединна варіанта).

Вибір одного з перерахованих видів середніх для характеристики ознаки проводиться не довільно, а залежно від змісту явища, що вивчається, і від мети, для якої обчислюють середню. Висхідною базою розрахунку і орієнтиром правильності форми середньої є економічне співвідношення, що виражає зміст середніх величин та їх зв'язок.

## 7.4. Середня арифметична та її властивості

Середня арифметична є однією з найбільш поширених форм середньої величини у статистиці. Як і всі інші середні, вона застосовується у формі *простої середньої* та *виваженої середньої*.

Вихідною, визначальною формою служить проста середня. *Середня арифметична проста* (*незважена*) розраховується шляхом ділення суми всіх значень ознаки (варіант), що варіюється, на число одиниць сукупності[1]:

$$\bar{x}_{ap.} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum x}{n}, \qquad (7.7)$$

де $\bar{x}_{ap.}$ — середня арифметична;

$x_1, x_2, x_3 \ldots, x_n$ — окремі значення (варіанти) варіюючої ознаки;

$n$ — число одиниць в сукупності.

Отже, *середня арифметична* є основним видом середніх, що застосовуються у статистиці.

Для обчислення середньої ми повинні мати в розпорядженні ряд величин певної ознаки, що варіюють (відрізняються один від одного).

Для обчислення середньої арифметичної треба підсумувати зна-

---

[1] Правильніше б цю формулу слід записати як

$$\bar{x}_a = \frac{\sum_{i=1}^{n} x_i}{n},$$

але коли підсумовування здійснюється від $i$=1 до $n$, знаки, що вказують, які величини підсумовуються, опускаються, і формула дається у тому вигляді, в якому вона дана у тексті.

До речі, у статистиці майже завжди підсумовування проводиться від $i$=1 до $n$.

чення всіх варіантів і суму розділити на кількість варіантів ($n$). Наведемо приклад розрахунку.

Маються такі дані про виробництво робітниками продукції *A* за зміну, шт. (табл. 7.3).

Таблиця 7.3

| Номер робітника | Випущено виробів за зміну, шт. | Номер робітника | Випущено виробів за зміну, шт. |
|---|---|---|---|
| 1 | 10 | 6 | 13 |
| 2 | 11 | 7 | 8 |
| 3 | 9 | 8 | 8 |
| 4 | 10 | 9 | 11 |
| 5 | 11 | 10 | 9 |

Необхідно о б ч и с л и т и число деталей, що виробляються одним робітником за зміну.

У цьому прикладі ознака, що коливається, — випуск продукції за зміну. Численні значення ознаки (10, 11, 9 і т. д.) у кожної одиниці сукупності називають *варіантами*.

Число деталей, що виробляються всіма робітниками ($\sum x_i$), дорівнює сумі індивідуального вироблення кожного робітника і може бути записано алгебраїчно:

$$x_1 + x_2 + x_3 + ... + x_n = \sum x_i . \tag{7.8}$$

Для визначення середнього виробітку необхідно суму деталей, вироблених усіма робітниками, розділити на кількість робітників:

$$\bar{x} = \frac{\sum x}{n} = \frac{10+11+9+10+11+13+8+8+11+9}{10} = \frac{100}{10} = 10 \text{ шт.}$$

Це і є середня арифметична. Якби всі робітники мали однаковий виробіток, то вони забезпечували б випуск продукції за зміну в кількості: 10 шт.·10 осіб = 100 шт. деталей. Перевіримо правильність використання виду середньої за правилом[1]: сума реальних величин

[1] Визначальній функції відповідає рівняння середніх, де індивідуальне значення ознаки (варіанту) одиниць сукупності замінюється середньою: $\bar{x} + \bar{x} + \bar{x} + ... + \bar{x} = x_1 + x_2 + x_3 + ... + x_n$, або $n\bar{x} = \sum x_i$ , звідси отримуємо середню арифметичну просту: $\bar{x}_{ар} = \frac{\sum x_i}{n}$, де $x_i$ — індивідуальні значення ознаки кожної одиниці сукупності; $n$ — число одиниць сукупності.

повинна дорівнювати сумі середніх, тобто:

$$10+11+9+10+11+13+8+8+11+9=10+10+10+10+10+10+10+10+10+10=100.$$

Для розрахунку середньої арифметичної простої необхідно мати в розпорядженні індивідуальними значеннями варіюючої ознаки, отримані зі спостережень. Зазвичай дані статистичного спостереження записуються у первинний ряд розподілу, коли одиниці з індивідуальним значенням ознаки розташовуються у такому ряду без системи. І тут техніка розрахунку середньої зводиться до підсумовування варіант і діленням отриманої суми на їх число.

*Проста середня арифметична обчислюється у випадках, коли окремі ваги відсутні або їх дуже важко визначити*. Наприклад, потрібно дізнатися про середню ціну реалізації продукції. Щоб правильно обчислити середній рівень цін, ці ціни потрібно зважити. Однак це або дуже важко, або взагалі неможливо зробити. *Просту середню можна обчислити і в тому випадку, якщо ваги, тобто окремі чисельності кожного варіанта, не надто відрізняються*. Для нашого прикладу це означає, що за різними цінами продається приблизно та ж сама кількість товару. Навпаки, у тому випадку, коли ваги сильно відрізняються один від одного, застосування простої середньої призводить до дуже грубих помилок.

Середня арифметична набула широкого поширення у виробничій практиці. Так, за допомогою середньої арифметичної виражається середній рівень трудомісткості виготовлення продукції, продуктивності праці, якість продукції, витрат матеріалів, витрат палива та енергії на одиницю продукції тощо.

Але часто буває так, що окремі варіанти в досліджуваній сукупності зустрічаються неоднакову кількість разів (одних більше, інших менше). У тих випадках, коли кожен варіант ознаки зустрічається не один, а кілька разів, то середня обчислюється інакше. У даному випадку формула простої середньої ускладнюється, тому що необхідно врахувати чисельність кожного варіанта, або, як то кажуть, зважити його значення в сукупності. *Статистичною вагою* називаються числа, які враховують значення величини ознаки (варіанту) в окремої одиниці сукупності.

Якщо вихідні дані статистичного спостереження піддані первинної обробці, то загальну суму значень ознаки всім одиниць можна отримати як суму добутку всіх варіантів $x_i$ на число одиниць, або частот $f_i$.

З цією метою в статистиці та економічному аналізі дуже часто використовують *середню арифметичну зважену*. Формула виваженої середньої арифметичної має такий вигляд:

$$\bar{x}_{ар.} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 \dots + x_n f_n}{f_1 + f_2 + f_3 \dots + f_n} = \frac{\sum xf}{\sum f}, \qquad (7.9)$$

де $f$ — частота окремих значень ознаки (варіанту).

Така середня арифметична за формою є зваженою на відміну від простої арифметичної, яка була отримана як частка від ділення загальної кількості виробів, випущених за зміну, на кількість робітників. Виваженою вона називається тому, що при обчисленні її кожному варіанті було надано «вагу», що відповідає числу випадків, у яких вона зустрічається, тобто її частоті.

**Обчислення середньої арифметичної із варіаційного ряду**. Як відомо, вихідні дані, отримані в результаті спостереження можуть бути представлені у вигляді дискретних або інтервальних рядів. Залежно від вихідних даних техніка розрахунку середньої дещо змінюється.

Для обчислення середньої у дискретних рядах варіанти необхідно помножити на частоти і суму добутку розділити у сумі частот. Покажемо застосування середньої на прикладі.

Маються наступні дані про заробітну плату 50 робітників-відрядників:

Таблиця 7.4

| Місячна заробітна плата, грн | Число робітників |
|---|---|
| 1 800 | 2 |
| 2 000 | 9 |
| 2 300 | 15 |
| 2 700 | 13 |
| 3 000 | 11 |
| Всього | 50 |

На основі цих даних потрібно в и з н а ч и т и середню зарплатню працівників.

За даними наведеного ряду розподілу видно, що одні й самі значення ознаки (варіанти) повторюються кілька разів. Так, якщо варіант $x_1$ зустрічається в сукупності 2 рази, а варіант $x_3$ — 15 разів, то другий варіант «важить» майже у 8 разів більше, ніж перший і це має знайти своє відображення у величині середньої арифметичної.

Число однакових значень ознаки в рядах розподілу називається *частотою* або *вагою* і позначається символом $f$ (табл. 7.5).

Таблиця 7.5

| Місячна заробітна плата (варіанта — $x$), грн | Число робітників $f$ |
|---|---|
| $x_1$=1800 | $f_1 = 2$ |
| $x_2$=2000 | $f_2 = 9$ |
| $x_3$=2300 | $f_3 = 15$ |
| $x_4$=2700 | $f_4 = 13$ |
| $x_5$=3000 | $f_4 = 11$ |
| | 50 |

Неважко переконатися, що результати розрахунку за згрупованими даними досліджуваної сукупності відповідають величині середньої арифметичної обчисленої за даними первинного обліку, не підданого групуванню ряду розподілу.

З'ясовуємо, що представляє собою середня заробітна плата одного робітника, співвідношенням яких величин вона є. Під середньою місячною заробітною платою працівника розуміється сума коштів, яку отримує працівник протягом місяця. Середня місячна заробітна плата, розрахована за формулою середньої зваженої арифметичної, для вищезгаданого прикладу дорівнює:

$$\text{Середня заробітна плата одного робітника} = \frac{\text{Заробітна плата всіх робітників (фонд заробітної плати)}}{\text{Число робітників}}.$$

Отже, нам відомі значення знаменника, але не відомі значення чисельника. Їх можна розрахувати. Для цього потрібно місячну заробітну плату помножити на кількість працівників.

По варіаційному ряду для отримання загальної суми заробітної плати слід врахувати повторюваність кожної варіанти, тобто помножити її на частоту та отримати загальну суму таких добутків $\sum xf$; для отримання загальної чисельності працівників ($f$) треба підрахувати суму частот, тобто $\sum f$.

Беремо дані із вихідної таблиці. Підставляємо у вихідне співвідношення обидві необхідні нам величини і отримуємо середню:

$$\bar{x} = \frac{(1\,800 \cdot 2) + (2\,000 \cdot 9) + (2\,300 \cdot 15) + (2\,700 \cdot 13) + (3\,000 \cdot 11)}{2 + 9 + 15 + 13 + 11} =$$

$$\frac{124\,200}{50} = 2\,484 \text{ грн.}$$

Отже, середня місячна заробітна плата робітників-відрядників дорівнює 2 484,0 грн.

Фонд заробітної плати по кожній групі робітників дорівнює добутку варіанти на частоту, а сума цих добутків дає загальний фонд заробітної плати всіх робітників. Позначивши чисельність працівників через $x$, місячну заробітну плату через $f$ і середню як звичайно через $\bar{x}$, напишемо формулу в алгебраїчному вигляді: $\bar{x} = \frac{\sum xf}{\sum f}$. Це — *середня арифметична*.

З формули видно, що середня залежить не тільки від значень ознаки, а й їх частот, тобто від складу сукупності, від її структури. Змінивши за умови завдання склад робітників і обчислимо середню у зміненій структурі (табл. 7.6).

Таблиця 7.6

| Місячна заробітна плата (варіанта — $x$), грн | Число робітників $f$ | $x \cdot f$ |
|---|---|---|
| 1 800 | 4 | 7 200 |
| 2 000 | 10 | 20 000 |
| 2 300 | 16 | 36 800 |
| 2 700 | 12 | 32 400 |
| 3 000 | 8 | 24 000 |
| | 50 | 120 400 |

Середня заробітна плата одного робітника складе:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{120\,400}{50} = 2\,408 \text{ грн.}$$

Обчислення середньої арифметичної, зваженої частотами, може розглядатися, по суті, лише як скорочення розрахунків. Якби всі 50 працівників не були б зведені до варіаційного ряду, то підсумовування їх для обчислення середньої потребувало б більших зусиль. Зрозуміло, обчислення середньої арифметичної зваженої за даними групувань може призвести до певної помилки. Передбачається, що окремі значення ознак, що потрапляють у межі інтервалу, розподіляються більш менш рівномірно, не накопичуючись біля меж. Таким чином, середні значення інтервалів досить близькі до середніх значень даних, що потрапили в той чи інший інтервал. В результаті можуть виникнути помилки при обчисленні середньої. Але, як показує практика, ці помилки зазвичай невеликі, і тому задля їх усунення не слід відмовлятися від вигод, пов'язаних із можливостями скорочення рахункових операцій.

Слід застерегти від помилки, що зустрічається при обчисленні середньої арифметичної. Часто при розрахунках виникає питання, а чому не можна у формулі виваженої середньої скоротити частоти *f*? Так робити не можна, бо сума добутку не дорівнює добутку сум, тобто $\sum xf \neq \sum x \sum f$. Не входячи в теоретичні міркування, досить точно звернути увагу на підсумки нашого прикладу:

$$\sum xf = 120400;$$

$$\sum x = 1800 + 2000 + \ldots + 300 = 11800;$$

$$\sum f = 50;$$

$$120400 \neq 11800 \cdot 50.$$

Отже, при обчисленні середньої арифметичної зваженої слід використовувати статистичні ваги, тобто частоти.

Нерідко статистичний матеріал у результаті обробки може бути представлений не тільки у вигляді дискретних рядів розподілу, а й у вигляді згрупованих в інтервали даних.

У інтервальних рядах варіанти ознаки, що підлягають усередненню, представлені у вигляді певного інтервалу «від … до …», тобто величина варіанта коливається у визначених межах. При цьому інтервальні варіаційні ряди можуть бути як із закритими, так і з відкритими інтервалами. Як обчислити у цьому випадку середню арифметичну?

Для того, щоб обчислити середню арифметичну *інтервального ряду*, треба для кожного інтервалу спочатку визначити центр інтервалу (тобто середнє значення), а після цього проводити обчислення за відомим нам способом. Середні значення (центр) інтервалів $x'$ визначаються як напівсуми крайніх їх значень: $x' = \frac{x_{\min} + x_{\max}}{2}$. У цьому випадку середня арифметична зважена обчислюється за формулою: $\bar{x}_a = \frac{\sum x'f}{\sum f}$.

Розглянемо розрахунок середньої арифметичної для таких рядів.

У варіаційному ряду, що представляє розподіл робітників за кількістю виробленої продукції, інтервали виявилися закритими.

Обчислимо середнє вироблення виробів одним робітником за зміну.

| Групи робітників за кількістю виробленої продукції за зміну, шт. $x$ | Число основних робітників $f$ |
|---|---|
| 8–10 | 28 |
| 10–12 | 58 |
| 12–14 | 72 |
| 14–16 | 29 |
| 16–18 | 13 |
| | 200 |

В якості ознаки, що підлягає усередненню ($x$), тут виступає кількість продукції, виробленої робітниками за зміну. Однак у цьому ряду варіанти ознаки, що усереднюється, (продукція за зміну) представлені не одним числом, а вигляді інтервалу «від … до …». Робочі першої групи виробляють продукції від 8 до 10 шт., робітники другої групи від 10 до 12 шт. і т. д.

Щоб застосувати цю формулу середньої зваженої арифметичної, необхідно перейти до дискретного ряду, тобто варіанти ознаки виразити одним числом (дискретним). За таке дискретне число приймається середня арифметична проста з верхнього та нижнього значення інтервалу. Так, для першої групи дискретна величина буде рівною:

$$\frac{8+10}{2}=9.$$

Подальший розрахунок провадиться звичайним методом арифметичної середньої. Результати розрахунків дискретних чисел, що характеризують серединну величину інтервалів груп робітників за кількістю виробленої продукції за зміну, наведено у табл. 7.7. Якщо в нашому прикладі перемножити всі варіанти (середній виробіток) на число основних робітників, то отримаємо виробництво продукції всіма робітниками. Сума цих величин дорівнюватиме 2 482 шт. Якщо всі виготовлені деталі розділити на число робітників, тобто на 200, отримаємо середній виробіток на одного робітника.

Подальший розрахунок проводиться звичайним методом визначення середньої арифметичної зваженої:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{2\,482}{200} = 12{,}41 \approx 12 \text{ штук}.$$

Таблиця 7.7

| Групи робітників за кількістю виробленої продукції за зміну, шт. | Число основних робітників ($f$) | Середина інтервалу ($x$) | Добуток ($x \cdot f$) |
|---|---|---|---|
| 8–10 | 28 | $\frac{8+10}{2} = 9$ | 252 |
| 10–12 | 58 | 11 | 638 |
| 12–14 | 72 | 13 | 936 |
| 14–16 | 29 | 15 | 435 |
| 16–18 | 13 | 17 | 221 |
| Всього | 200 | | 2 482 |

Отже, усі працівники зробили 2 482 шт. виробів за зміну, а кожен у середньому виробив 12 шт.

Перетворимо розглянутий вище варіаційний ряд розподілу на ряд із *відкритими інтервалами*. В окремих випадках частота повторення відповідних варіантів може бути виражена у відсотках. У таких випадках розрахунок середньої арифметичної дещо змінюється. Розглянемо порядок розрахунку середньої.

Наприклад, поставлене завдання — знайти середній стаж основних робітників. Маються такі дані про розподіл робітників механічного цеху за стажем роботи:

Таблиця 7.8

| Групи робітників за стажем, років | Питома вага основних робітників, % |
|---|---|
| До 3 років | 4 |
| Від 3 до 6 років | 10 |
| » 6 » 10 | 16 |
| » 10 » 15 | 18 |
| » 15 » 25 | 38 |
| Понад 25 років | 14 |
| Всього | 100 |

Щодо стосується відкритих інтервалів, то за центральний варіант першого інтервалу береться половина другого інтервалу, за центральний варіант останнього інтервалу — сума нижньої межі та половини останнього інтервалу. У прикладі центральні варіанти будуть такі: 1,5 (3:2); 4,5 [(3+6):2]; 8,0 [(6+10):2]; 12,5 [(10+12):2]; 20,0 [(15+25):2]; 30,0 [(25+(10:2)].

Для визначення середнього стажу основних робітників залишається помножити центральні варіанти на відповідні частості, виражені у відсотках, підсумувати ці добутки та отриману суму поділити на число робітників. Усі вихідні та розраховані дані представлені в наступній таблиці:

Таблиця 7.9

| Групи робітників за кількістю виробленої продукції за зміну, шт. | Число основних робітників в процентах ($w$) | Центральний варіант ($x$) | Добуток ($x \cdot w$) |
|---|---|---|---|
| До 3 років | 4 | 1,5 | 6 |
| Від 3 до 6 років | 10 | 4,5 | 45 |
| » 6 » 10 | 16 | 8,0 | 128 |
| » 10 » 15 | 18 | 12,5 | 225 |
| » 15 » 25 | 38 | 20,0 | 760 |
| Понад 25 років | 14 | 30,0 | 420 |
| Всього | 100 | | 1 584 |

Звідси

$$\bar{x} = \frac{\sum xf}{100} = \frac{1584}{100} \approx 15{,}8 \text{ років.}$$

Приймаючи центральні варіанти в якості $x_1, x_2, x_3, ..., x_n,$ умовно припускаємо, що варіанти у кожному інтервалі розташовуються рівномірно, хоча насправді цього може виявитися.

**Обчислення середньої арифметичної у випадку, коли частоти представлені відносними величинами структури**. Для розрахунку середньої можна використовувати не частоти, а відносні показники структури $(d_i)$, які представляють собою питому вагу окремої частоти в загальній сумі всіх частот, виражених у процентах: $d_i = \frac{f_i}{\sum f_i} \cdot 100.$ Якщо частоти підраховуються у процентах, такий показник питомої ваги може служити для розрахунку середньої арифметичної. Замінюючи частоти відносними величинами структури, отримаємо формулу середньої арифметичної зваженої: $\bar{x}_a = \frac{\sum x_i d_i}{\sum d_i}$. Метод розрахунку середньої та кінцевий результат від цього не зміняться.

Приклад. Представимо дані про чисельність робітників за умови наведеної вище типової задачі у відносних величинах:

Таблиця 7.10

| Місячна заробітна плата ($x$), грн | Показники структури | | $x \cdot d$ | $x \cdot w$ |
|---|---|---|---|---|
| | питома вага, у відсотках до підсумку ($d_i$) | Частість ($w_i$) | | |
| 1 800 | 8,0 | 0,08 | 14 400 | 144 |
| 2 000 | 20,0 | 0,20 | 40 000 | 400 |
| 2 300 | 32,0 | 0,32 | 73 600 | 736 |
| 2 700 | 24,0 | 0,24 | 64 800 | 648 |
| 3 000 | 16,0 | 0,16 | 48 000 | 480 |
| | 100,0 | 1,00 | 240 800 | 2 408 |

Середня заробітна плата робітника (середня арифметична), зважена за процентними співвідношеннями, буде дорівнюватиме:

$$\bar{x} = \frac{(1800 \cdot 8) + (2000 \cdot 20) + (2300 \cdot 32) + (2700 \cdot 24) + (3000 \cdot 16)}{100} =$$

$$= \frac{240\,800}{100} = 2\,408 \text{ грн.}$$

При розрахунку середньої можна використовувати не частоти, а частості $(w_i)$, які є відношенням частот до загальної суми всіх частот, що визначається для кожної варіанти:

$$w_i = \frac{f_i}{\sum f_i}. \tag{7.10}$$

Якщо вагами є частоти, виражені в коефіцієнтах, то обчислення спрощуються. Оскільки сума коефіцієнтів завжди дорівнює одиницю ($\sum d_i = 1$), то розрахунок середньої зводиться до визначення суми добутку варіант на частоти (у разі коефіцієнти): $\bar{x}_{ap} = \sum x_i w_i$ . Середня заробітна плата в даному випадку становитиме:

$$\bar{x} = 1\,800 \cdot 0{,}08 + 2\,000 \cdot 0{,}20 + 2\,300 \cdot 0{,}32 + 2\,700 \cdot 0{,}24 +$$

$$+ 3\,000 \cdot 0{,}16 = 2\,408 \text{ грн.}$$

Незалежно від характеру ваги величина середньої арифметичної заробітної плати робітників має однакове значення і в даному випадку дорівнює 2 408 грн.

Застосування середньої арифметичної залежить від вихідного співвідношення, що виражає сутність (зміст) показника, що підлягає усередненню, і наявних даних.

**Найважливіші властивості середньої арифметичної**. Тепер перейдемо до характеристики властивостей середньої арифметичної. Середня арифметична має ряд властивостей, що мають велике застосування як у галузі теоретичної статистики, так і для різноманітних практичних обчислень. Деякі з цих властивостей можуть бути використані для скорочення обчислень або перевірки їх точності.

Найважливіші з цих властивостей наступні.

1. **Середня величина, помножена чисельність всієї сукупності, дорівнює сумі добутку кожного варіанта на його чисельність**. Алгебраїчно це можна записати так:

$$\bar{x}\sum f = \sum xf. \tag{7.11}$$

Ця властивість витікає із самої формули середньої арифметичної. Оскільки

$$\bar{x} = \frac{\sum xf}{\sum f}, \quad \text{тоді} \quad \bar{x}\sum f = \sum xf. \tag{7.12}$$

Ця властивість дуже широко використовується у практиці розрахунків, при плануванні та аналізі економічних явищ та процесів. Так, наприклад, середня заробітна плата, помножена на чисельність працівників, дозволяє обчислити фонд заробітної плати; середня трудомісткість виготовлення виробу, помножена на кількість виробів, дає трудомісткість продукції і т.д.

У прикладі це означає: 2 408·50=120 400 грн. З зазначеної властивості випливає, що, обчислюючи середню, ми зрівнюємо варіанти, замінюємо їх одним середнім числом, яке як постійний множник виносимо з-під знака суми.

**2. Якщо від кожного варіанта (*x*) відняти довільне постійне число і з результатів віднімання, тобто відхилень варіантів від цієї постійної величини ($x - A = x'$) знайти середню ($\overline{x'}$), то нова середня зменшиться на те ж число**. Позначимо постійне число $A$, тоді алгебраїчно це можна записати так:

$$\frac{\sum (x - A)f}{\sum f} = \bar{x} - A, \tag{7.13}$$

звідки

$$\bar{x} = \frac{\sum (x - A)}{\sum f} + A. \qquad (7.14)$$

Тому, щоб отримати середню варіантів ($\bar{x}$), потрібно до знайденої середньої ($\overline{x'}$) додати ту ж постійну величину ($A$):

$$\bar{x} = \overline{x'} + A, \qquad (7.15)$$

якщо $x' = x - A$.

На практиці в якості постійного числа $A$ вибирають одне з значень усередненеої ознаки, розташованого в середні ряду.

Застосуємо цю властивість на вирішення вже наведеного прикладу. Зменшимо всі варіанти на 1 800 ($A$ = 1 800). У цьому випадку згідно з правилом нова середня має становити: 2 408 – 1 800=608.

Таблиця 7.11

**Обчислення середньої із зменшених варіант**

| Середнє значення інтервалу ($x$) | $x - A = x'$ | Число робітників ($f$) | $(x - A)f$ |
|---|---|---|---|
| 1 800 | 0 | 4 | 0 |
| 2 000 | 200 | 10 | 2 000 |
| 2 300 | 500 | 16 | 8 000 |
| 2 700 | 900 | 12 | 10 800 |
| 3 000 | 1 200 | 8 | 9 600 |
| Всього | — | 50 | 30 400 |

Середня з цих зменшених варіантів дорівнюватиме $\bar{x} - A = \frac{30400}{50} = 608$; а з незменшених варіантів була рівною $\bar{x} = 608+$ +1 800=2 408. Як бачимо, при зменшенні всіх варіантів на 1 800 і середня зменшилася на цю величину.

За допомогою такого прийому суттєво полегшується техніка розрахунку середньої арифметичної, оскільки відбувається перехід від низки складних чисел до простих. Використана при розрахунках величина $\overline{x'}$ називається умовним моментом першого порядку. Метод обчислення середньої, заснований на розібраній властивості, називається методом відліку від умовного нуля, методом моментів або методом редукції.

**3. Якщо до кожного варіанту додати якесь довільне постійне число, то нова середня збільшиться на те ж число**. Алгебраїчно це позначається так:

$$\frac{\sum (x+A)f}{\sum f} = \bar{x} + A, \qquad (7.16)$$

звідки

$$\bar{x} = \frac{\sum (x+A)}{\sum f} - A. \qquad (7.17)$$

Розглянемо застосування цієї властивості на практиці. Збільшимо всі варіанти на 1 800 ($A$=1 800). У цьому випадку згідно з правилом нова середня має становити: 2 408 + 1 800 = 4 208.

Таблиця 7.12

**Обчислення середньої зі збільшених варіант**

| Середнє значення інтервалу ($x$) | $x + A$ | Число робітників ($f$) | $(x + A)f$ |
|---|---|---|---|
| 1 800 | 3 600 | 4 | 14 400 |
| 2 000 | 3 800 | 10 | 38 000 |
| 2 300 | 4 100 | 16 | 65 600 |
| 2 700 | 4 500 | 12 | 54 000 |
| 3 000 | 4 800 | 8 | 38 400 |
| Всього | — | 50 | 210 400 |

$$\bar{x} - A = \frac{210400}{50} = 4\,208; \qquad \bar{x} = 4\,208 - 1\,800 = 2\,408.$$

Друга і третя властивість означають, що якщо до кожної варіанти відібрати (або додати) довільне, постійне число $A$, то нова середня зменшиться (відповідно збільшиться) на те саме число $A$.

**4. Якщо кожен варіант розділити на довільне постійне число, і з групових обчислити середню, то нова середня арифметична зменшиться в стільки ж разів**. Зменшимо всі ваги варіантів в $A$ раз і запишемо формулу для середньої з новими вагами:

$$\frac{\sum \frac{x}{A} f}{\sum f} = \frac{\bar{x}}{A}, \qquad (7.18)$$

звідки

$$\bar{x} = \frac{\sum \frac{x}{A} f}{\sum f} \cdot A. \qquad (7.19)$$

Тому, щоб отримати середню з варіантів ($\bar{x}$), потрібно знайдену середню ($\overline{x'}$) помножити на ту ж постійну величину ($A$), тобто

$$\bar{x} = \overline{x'} \cdot A. \qquad (7.20)$$

Перевіримо цю властивість на наведеному прикладі. Розділимо всі варіанти на 100 ($A$ = 100). У цьому випадку нова середня арифметична, отримана в результаті розрахунків, повинна бути в 100 разів менше фактичної середньої арифметичної. Її величина має становити $\frac{2408}{100} = 24{,}08$ (див. табл. 7.13).

Таблиця 7.13

**Обчислення середньої із скорочених варіант**

| Середнє значення інтервалу ($x$) | $\frac{x}{A}$ | Число робітників ($f$) | $\frac{x}{A}f$ |
|---|---|---|---|
| 1 800 | 18 | 4 | 72 |
| 2 000 | 20 | 10 | 200 |
| 2 300 | 23 | 16 | 368 |
| 2 700 | 27 | 12 | 324 |
| 3 000 | 30 | 8 | 240 |
| Всього | — | 50 | 1 204 |

$$\bar{x} = \frac{\sum \frac{x}{A} f}{\sum f} \cdot A = \frac{1204}{50} \cdot 100 = 2\,408.$$

Завдяки зменшенню ваг техніка обчислення середньої арифметичної значно спрощується.

**5. Якщо кожну варіанту помножити на довільне постійне число, то нова середня арифметична збільшиться в стільки ж разів**. Збільшимо всі ваги варіантів в $A$ раз і запишемо формулу для обчислення середньої:

$$\overline{x'} = \frac{\sum (x \cdot A) f}{\sum f} = \bar{x} \cdot A, \qquad (7.21)$$

звідки

$$\bar{x} = \frac{\sum (x \cdot A) f}{\sum f} : A. \qquad (7.22)$$

Перевіримо цю властивість на наступному прикладі. Помножимо всі варіанти на 5 ($A = 5$). Отримана нова середня має бути в 5 разів більше фактичної середньої, вона дорівнюватиме 2 408·5=12 040 (табл. 7.14).

Таблиця 7.14

**Обчислення середньої зі збільшених варіант**

| Середнє значення інтервалу ($x$) | $x \cdot A$ | Число робітників ($f$) | $x \cdot A \cdot f$ |
|---|---|---|---|
| 1 800 | 9 000 | 4 | 36 000 |
| 2 000 | 10 000 | 10 | 100 000 |
| 2 300 | 11 500 | 16 | 184 000 |
| 2 700 | 13 500 | 12 | 162 000 |
| 3 000 | 15 000 | 8 | 120 000 |
| Всього | — | 50 | 602 000 |

$$\bar{x} = \frac{\sum (x \cdot A) f}{\sum f} = \frac{602\ 000}{50} = 12\ 040; \quad \bar{x} = 12\ 040:5 = 2\ 408.$$

Як бачимо, результат розрахунків відповідає фактичній величині середньої. Ці дві властивості мають виключно велике практичне значення, дозволяючи скорочувати іноді дуже стомлюючі обчислення середньої.

**6. Якщо всі частоти (ваги) розділити або помножити на якесь довільне число, то середня арифметична не зміниться**. Розглянемо доказ цієї властивості. Нехай загальне кратне $k$:

$$\frac{\sum x \frac{f}{k}}{\sum \frac{f}{k}} = \frac{\sum xf}{\sum f} = \bar{x}. \tag{7.23}$$

Таким чином, числове значення середньої від заміни частот ($f$) на скорочені частоти ($f'$) не зміниться.

Перевіримо цю властивість на тому ж прикладі. Розділимо всі частоти на 5 і помножимо на 100 (тобто виразимо у відсотках). Величина отриманої нової середньої арифметичної має змінитися.

Результат вийшов той самий, тільки техніка обчислення середньої значно спростилася (табл. 7.15).

$$\bar{x} = \frac{120400}{5} = 2\ 408.$$

У силу цієї властивості заміна в варіаційному ряду частот частос-

Таблиця 7.15

**Обчислення середньої зі скороченими частотами**

| Середнє значення інтервалу ($x$) | Частоти ($f$) | $\frac{f}{50}100$ | $x \cdot f \frac{100}{50}$ |
|---|---|---|---|
| 1 800 | 4 | 8 | 7 200 |
| 2 000 | 10 | 20 | 20 000 |
| 2 300 | 16 | 32 | 36 800 |
| 2 700 | 12 | 24 | 32 400 |
| 3 000 | 8 | 16 | 24 000 |
| Всього | 50 | 100 | 120 400 |

тями не веде за собою зміни величини середньої арифметичної. Справді, перетворення частот на відносні частоти — частості представляє собою операцію ділення всіх частот однією й ту саму величину — загальну чисельність низки, тобто воно рівносильне зменшенню або збільшенню частот в одне і те ж число разів.

**7. Сума відхилень окремих значень ознаки (варіантів) від середньої арифметичної, помноженої на ваги (частоти), дорівнює нулю:**

$$\sum (x - \bar{x}) = 0 \qquad (7.24)$$

(якщо всі варіанти дорівнюють одиниці),
або

$$\sum (x - \bar{x}) f = 0. \qquad (7.25)$$

Практично ця властивість середньої арифметичної може бути використана для перевірки обчислень: якщо сума негативних відхилень від середньої не дорівнює сумі позитивних відхилень, це означає, що при обчисленні середньої була допущена помилка.

Доказ. Запишемо символами цю властивість середньої арифметичної (див. робота Т. В. Рябушкіна)

$\sum (x_i - \bar{x}) = \sum d_i = 0$ для первинного ряду;

$\sum (x_i - \bar{x}) f = \sum d_i f_i = 0$ для згрупованих даних.

Значення ознаки (варіанту) для окремих одиниць:

$$x_1, x_2, x_3 \ldots, x_n.$$

Визначимо відхилення $d$ значень ознаки у кожної одиниці від середньої арифметичної та знайдемо суму відхилень для всіх одиниць:

$$\begin{aligned} & x_1 - \bar{x} = d_1 \\ + \ & x_2 - \bar{x} = d_2 \\ & x_3 - \bar{x} = d_3 \\ & . \ . \ . \ . \ . \\ & x_n - \bar{x} = d_n \end{aligned}$$

$$x_1 + x_2 + x_3 + \ldots + x_n - n\bar{x} = \sum d_i.$$

Але $\bar{x}$ обчислений у нас як $\frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$. Підставимо замість $\bar{x}$ його значення. Тоді отримаємо:

$$x_1 + x_2 + x_3 + \ldots + x_n - n \cdot \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \sum d_i. \tag{7.26}$$

Після скорочення в лівій рості рівності ми отримаємо $\sum d_i = 0$, а з урахуванням частот $\sum d_i f_i = 0$, що й необхідно було довести.

Таким чином, $\sum x = \sum \bar{x}$, тобто рівність, що використовується у виведенні формули середньої величини. Якщо кожен варіант повторюється $f$ разів, то, очевидно, запис розглянутої властивості середньої буде наступною:

$$\sum (x - \bar{x}) f = 0; \quad \sum xf = \sum \bar{x} f. \tag{7.27}$$

Це означає, що якщо взяти відхилення від кожного варіанта від середньої величини і зважити за чисельністю, а потім скласти їх, то отримаємо нуль.

Розглянута властивість середньої арифметичної дозволяє перевіряти її розрахунок, застосовувати способи спрощення її розрахунку. Останнє можливо через те, що сума відхилень варіантів від будь-якої іншої величини не виявиться рівною нулю; та сума, що технічно набагато легше підраховується, може бути використана для розрахунку середньої величини з відповідними коригуваннями.

Перевіримо цю властивість тому ж прикладі (табл. 7.16.). У наведеному вище прикладі середня заробітна плата становила 2 408 грн. Якщо цю величину відняти з кожної варіанти, алгебраїчна сума дорів-

Таблиця 7.16

**Перевірочний розрахунок, що $\sum (x-\bar{x})f = 0$ при $\bar{x}$=2408**

| Середнє значення інтервалу ($x$) | Число робітників ($f$) | Відхилення від середньої $x_i - \bar{x} = d_i$ ($x_i$–2408) | $d_i f_i$ | Сума відхилень |
|---|---|---|---|---|
| 1 800 | 4 | –608 | –2432 | –8240 |
| 2 000 | 10 | –408 | –4080 | |
| 2 300 | 16 | –108 | –1728 | |
| 2 700 | 12 | 292 | 3504 | +8240 |
| 3 000 | 8 | 592 | 4736 | |
| Всього | 50 | | | 0 |

нюватиме нулю.

У разі нулю дорівнює сума творів частот і відхилень (–8 240 + 8 240). Нічого не змінюється, якщо цей розрахунок зроблено для розгрупованих даних.

На підставі цієї властивості заснований технічний прийом обчислення середньої арифметичної, відомий під назвою способу моментів.

**8. Сума квадратів відхилень варіантів від їх середньої арифметичної величини менша, ніж сума квадратів їх відхилень індивідуальних значень будь-якої іншої величини**.

$$\sum (x_i - \bar{x})^2 = \sum d_i^{\,2} = \min . \tag{7.28}$$

Це означає, що вона менша за суму квадратів відхилень цих варіантів від будь-якого довільного числа:

$$\sum (x_i - \bar{x})^2 f < \sum (x_i - A)^2 f . \tag{7.29}$$

Розглянемо доказ цієї властивості (див. робота П. П. Маслова). Візьмемо суму квадратів відхилень членів низки від будь-якої величини $A$. Очевидно, що ми можемо представити цю суму як функцію $A$:

$$\sum (x - A)^2 = \varphi(A) . \tag{7.30}$$

Візьмемо першу похідну від цієї функції, звівши попередньо ліву частину рівності в квадрат:

$$\left[\sum (x - A)^2\right]' = \varphi'(A) \tag{7.31}$$

$$\left(\sum x^2 - 2\sum Ax + \sum A^2\right)' = 2\sum x - 2\sum A = \varphi'(A) . \tag{7.32}$$

Скорочуючи на 2, отримаємо

$$\sum x - \sum A = \varphi'(A)\,. \tag{7.33}$$

З математики нам відомо, що якщо перша похідна дорівнює нулю, то функція набуває мінімального (або максимального) значення.

Прирівняємо отриману похідну нулю $\sum x - \sum A = 0;$ звідси $\sum x = \sum A.$ Але сума $A$ може дорівнювати сумі членів ряду тільки за умови, якщо $A = \bar{x}$, бо у всіх інших випадках сума відхилень членів ряду від величини $A$ не буде дорівнює нулю. Отже, сума квадратів відхилень може приймати мінімальні розміри лише за умови, якщо відхилення обчислюються від середньої арифметичної.

Вивчення цих властивостей має велике практичне значення для обчислення середньої за даними варіаційного ряду.

**9. Зважена середня з групових (часткових) середніх дорівнює загальній середній:**

$$\frac{\sum \bar{x}_i \cdot n_i}{\sum n_i} = \bar{x}_{\text{общ}}\,, \tag{7.34}$$

де $\bar{x}_i$ — групові (часткові) середні, тобто середні окремих груп сукупності;

$n_i$ — чисельності одиниць у них.

**Обчислення середньої арифметичної з варіаційного ряду способом моментом або методом умовного нуля**. На підставі властивостей середньої (2-го, 3-го, 4-го та 5-го) побудований спосіб відліку від умовного нуля[1]. Цим методом зручно користуватися у випадках, коли для розрахунку представлені великі числові значення варіантів. Якщо величину вибрати з числа варіант, що мають найбільшу частоту і лежать близько до центру розподілу, застосування цього способу забезпечує великі вигоди для обчислювача. У результаті скорочення всіх відхилень варіант від «довільного початку» на величину інтервалу Ці обчислення можна зробити лише у випадку з рівними інтервалами.

*Метод моментів* є технічним прийомом для обчислення середньої арифметичної в інтервальному ряду. Зміст обчислення середньої арифметичної методом відліку від умовного нуля у тому, що всі варіанти в досліджуваному ряду зменшують на одну яку-небудь величину

---

[1] Цей спосіб іноді називають також способом «умовної середньої» або «довільного початку».

(зазвичай значення одного з варіантів). Після такого зменшення проводять обчислення середньої. Але обчислена середня буде меншою за дійсну на ту величину, на яку зменшені всі варіанти, отже, для визначення істинної середньої треба буде до зменшеної середньої додати величину, на яку зменшували всі варіанти.

Алгебраїчно розглянуту властивість середньої арифметичної можна представити так:

$$\frac{\sum (x-A)f}{\sum f} = \frac{\sum xf}{\sum f} - \frac{A\sum f}{\sum f} = \bar{x} - A. \tag{7.35}$$

З рівності слідує, що:

$$\bar{x} = A + \frac{\sum (x-A)f}{\sum f}. \tag{7.36}$$

Розрахунки можна ще більш спростити, якщо отримані варіанти зменшити (збільшити) на величину інтервалу і обчислити середню з них, перетворюючи ряд відхилень варіант від «довільного початку» в ряд натуральних чисел. Розрахунки можна проводити лише у випадку з рівними інтервалами. Для обчислення середньої арифметичної на отриманий кінцевий результат завдають поправки, але у зворотному порядку. Використання властивостей середньої арифметичної дозволяє значно спростити її обчислення.

Обчислення середньої арифметичної здійснюється в такій послідовності:

1) умовне прийняття центру інтервалів значення ознак всіх одиниць у кожному інтервалі;

2) віднімання з усіх варіантів початку відліку або «хибного нуля» ($x_0$) (властивість 2)[1];

3) розподіл всіх варіантів або відхилень варіантів від початку відліку на загальний множник (*k*) (властивість 4) та отримання умовних варіантів;

4) використання як ваги скорочених частот ($f'$) або відносних величин структури або координації (властивість 6).

Формула обчислення середньої методом відліку від умовного

---

[1] Використання цієї властивості дає практично відчутний ефект у разі прийняття за початок відліку варіанта (або центру інтервалу) з найбільшою частотою (або частістю).

нуля або методом моментів має вигляд: $\bar{x} = \overline{x'} \cdot k + x_0$, де $\overline{x'} = \frac{x - x_0}{k}$, тобто відхилення варіантів від початку відліку ділиться на загальний множник, а обчислення середньої $x'$ залежно від того, які ваги ми маємо, проводиться за однією з наступних формул:

$$\overline{x'} = \frac{\sum x'f'}{\sum f'}; \qquad \overline{x'} = \frac{\sum x'w'}{\sum w'}; \qquad \overline{x'} = \frac{\sum x'w'}{100}, \tag{7.37}$$

де $w$ — відносні величини координації.

Розглянемо спрощений спосіб розрахунку середньої арифметичної методом моментів на прикладі. За даними табл. 7.17 (колонка 1 і 2) обчислити середню трудомісткість виготовлення деталей, використовуючи метод відліку від умовного нуля і переконатися в тотожності результату з безпосереднім розрахунком середньої за формулою:

$$\bar{x} = \frac{\sum xf}{\sum f}. \tag{7.38}$$

Заповнюємо табл. 7.17.

Таблиця 7.17

**Розрахунок середньої арифметичної методом відліку від умовного нуля або методом моментів**

| Групи деталей щодо трудомісткості обробки, *хв.* | Число деталей $f_i$ | Середина інтервалу $x$ | Добуток $x \cdot f$, необхідний для розрахунку середньої | Відхилення від початку відліку $x' - x_0$ ($x_0$=40,5) | Умовні варіанти чи розрахункові одиниці $x' = \frac{x - x_0}{k}$ ($k$=3) | $x'_i \cdot f_i$ |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 30–33 | 5 | 31,5 | 157,5 | –9 | –3 | –15 |
| 33–36 | 10 | 34,5 | 345 | –6 | –2 | –20 |
| 36–39 | 23 | 37,5 | 862,5 | –3 | –1 | –23 |
| 39–42 | 18 | $x_0$=40,5 | 729 | 0 | 0 | 0 |
| 42–45 | 12 | 43,5 | 522 | +3 | +1 | +12 |
| 45–48 | 8 | 46,5 | 372 | +6 | +2 | +16 |
| 48–51 | 4 | 49,5 | 198 | +9 | +3 | +12 |
| Всього | 80 | – | 3 186 | – | – | –18 |

Дії в таблиці зроблено в наступній послідовності:

1) обчислюємо середні значення інтервалів; для першого інтервалу середнє значення становитиме 31,5 мм $\left(\frac{30+33}{2}\right)$, для другого — 34,5 мм $\left(\frac{33+36}{2}\right)$ і т. д. (колонка 3);

2) за початок відліку $x_0$ приймаємо значення інтервалу, що стоїть у середині (у разі $x_4'$=40,5 хв.) і знаходимо відхилення від початку відліку (колонка 5). Тоді для групи деталей зменшений варіант становитиме — 9 (31,5 – 40,5), другої групи — 6 (34,5 – 40,5) тощо;

3) розділимо відхилення $x - x_0$ на загальний множник $k$=3, що міститься в них, і величини $\frac{x-x_0}{k}$ позначимо через $x'$ (колонка 6);

4) в якості ваги використовуємо частоти ($f$), що містяться в колонці 2, і здійснимо зважування умовних варіантів $x_i' f_i$ (колонка 7).

Тепер розрахуємо умовну середню арифметичну з умовних відхилень (цю величину позначимо $\overline{x'}$):

$$\overline{x'} = \frac{\sum x_i' f_i}{\sum f_i} = \frac{-18}{80} = -0{,}225.$$

Тепер, щоб отримати дійсну середню, потрібно на отриману величину накласти поправки. Це означає, що з визначення середньої трудомісткості обробки деталі необхідно умовну середню збільшити в $k$ на раз і додати $x_0$, тобто.

$$\bar{x} = \overline{x'}k + x_0. \tag{7.39}$$

Для прикладу, що розглядається:

$$\bar{x} = -0{,}225 \cdot 3 + 40{,}5 = 39{,}825 \text{ хв.},$$
$$\bar{x} \approx 39{,}8 \text{ хв.}$$

тобто середня трудомісткість обробки деталі дорівнює 39,8 хв.

Поєднавши весь процес розрахунку, напишемо його формулу

$$\bar{x} = \frac{\sum\left(\frac{x-x_0}{k}\right)f_i}{\sum f_i}k + x_0. \tag{7.40}$$

Середня арифметична може обчислюватися за формулами зваженої та незваженої простої. У нашому прикладі розрахунок здійснювався за формулою середньої зваженої. Незважені середні можна застосовувати в тому випадку, коли всі варіанти ознаки, що усереднюється, мають однакові ваги або зустрічаються тільки по одному разу.

Порівнюючи знайдену середню із середньою, обчисленою безпосередньо за формулою $\bar{x} = \frac{\sum xf}{\sum f} = \frac{3186}{80} = 39{,}825$, бачимо їх тотожність. Перевага ж розрахунків методом відліку від умовного нуля очевидна.

**Обчислення середньою методом сум**. При рівноважних значеннях ознаки обчислення середньої арифметичної можна проводити методом відліку від умовного нуля з використанням накопичених частот або частостей. В цьому випадку нагромадження та підсумовування частостей проводиться таким чином (див. табл. 7.18). $w_1$

Таблиця 7.18

**Нагромадження та сумування частостей**

| Середина інтервалу $x$ | Число деталей (в процентах до підсумку) $w_i$ | $x'$ | Сумування частостей |  |  |  |
|---|---|---|---|---|---|---|
|  |  |  | в символах |  | фактично |  |
| А | 1 | 2 | 3 |  | 4 |  |
| 31,5 | $w_1$ 6,25 | –3 | $w_1$ | зі знаком мінус | 6,25=6,25 | – 72,5 |
| 34,5 | $w_2$ 12,50 | –2 | $w_1 + w_2$ |  | 12,5+6,25=18,75 |  |
| 37,5 | $w_3$ 28,75 | –1 | $w_1 + w_2 + w_3$ |  | 28,75+18,75=47,50 |  |
| 40,5 | $w_4$ 22,50 | 0 |  |  |  |  |
| 43,5 | $w_5$ 15,00 | 1 | $w_7 + w_6 + w_5$ | зі знаком плюс | 15,0+15,0=30,0 | +50 |
| 46,5 | $w_6$ 10,00 | 2 | $w_7 + w_6$ |  | 5,0+10,0=15,0 |  |
| 49,5 | $w_7$ 5,00 | 3 | $w_7$ |  | 5,0=5,0 |  |
| Всього | 100 |  | $-(3w_1 + 2w_2 + w_3) + (w_5 + 2w_6 + 3w_7)$ |  | – 72,5<br>+ 50,0<br>–22,5 |  |

Використовуючи результат колонки 4 і застосовуючи спосіб відліку від умовного нуля, отримаємо $\overline{x'} = \frac{\sum x'w}{\sum w} = -0{,}225$, а далі, як і попередньому розрахунку $\bar{x} = \overline{x'} \cdot k + A = -0{,}225 \cdot 3 + 40{,}5 = 39{,}825$.

## 7.5. Середня гармонійна

Середня арифметична, з обчисленням якої ми познайомимо, є найбільш елементарною, але в той же час часто використовуваною середньою, але далеко не єдиним видом середніх величин. Практика показує, що застосування такої середньої дуже часто призводить до помилкових результатів, тобто вона в певних випадках не застосовується.

Розглянемо такий приклад. У цеху працює три токарі і кожен токар обробляє однакові деталі, причому перший токар на обробку деталі витратив 6 хв., другий — 10 хв. та третій — 20 хв. Потрібно в и з н а ч и т и, скільки часу витрачає в середньому на обробку однієї деталі робітник із середньою продуктивністю, якщо кожен із них відпрацював однаковий час, наприклад, зміну.

На перший погляд, здається, це завдання легко можна вирішити за формулою середньої арифметичної простої. Середні витрати часу на обробку однієї деталі обчислимо за формулою середньої арифметичної складуть: $\bar{x} = \frac{\sum x}{n} = \frac{6+10+20}{3} = \frac{36}{3} = 12$ хв. Отже, протягом години один токар виробляє 5 деталей (60:12), а протягом 8-годинного робочого дня — 40 деталей. Три токарі за 8 год. вироблять 120 деталей. Цей розрахунок зроблено за допомогою середньої арифметичної величини. Отримана середня була б правильною, якби кожен робітник зробив би по одній деталі. Перевіримо правильність розрахунку з урахуванням витрат часу кожним робочим. Перший робітник за годину виробить 10 деталей (60:6), а за 8 годин — 80 деталей; другий робітник — відповідно 6 (60:10) та 48 деталей та третій робітник — відповідно 3 (60:20) та 24 деталі. Троє токарів вироблять 152 деталі (80+48+24). У разі формула середньої арифметичної до розрахунку середніх витрат часу не відповідає особливостям явища. Неправильність отриманого результату свідчить, що формула середньої була обрана не правильно.

Якби кожен робітник обробляв даний вид продукції не безперервно, а мав би певне завдання в одиницях, то середній час визначалося б як середня арифметична зважена за кількістю одиниць продукції. Але протягом дня окремими робітниками було виготовлено різну кількість деталей. Для визначення числа деталей, виготовлених кожним робітником, скористаємося таким співвідношенням:

$$\text{Середній час, витрачений на одну деталь} = \frac{\text{Весь витрачений час}}{\text{Кількість деталей}};$$

Цей розрахунок та формула будуть вірними незалежно від того, що прийняти за одиницю часу. Звідси середня норма часу як відношення витрат часу на кількість деталей становитиме:

$$\text{Середній час, витрачений на одну деталь} = \frac{8\cdot 60+8\cdot 60+8\cdot 60}{152} = 9{,}47 \text{ хв.}$$

Число деталей, виготовлених кожним робітником, визначається відношенням всього часу роботи до середнього часу, витраченого на одну деталь. Тоді середній час, необхідний виготовлення однієї деталі, дорівнює:

$$\bar{x} = \frac{8\cdot 60 + 8\cdot 60 + 8\cdot 60}{\frac{8\cdot 60}{6} + \frac{8\cdot 60}{10} + \frac{8\cdot 60}{20}} =$$

$$= \frac{1440}{80+48+24} = \frac{1440}{152} = 9{,}47 \text{ хв.}$$

Це рішення можна представити інакше:

$$\bar{x} = \frac{3\cdot 8\cdot 60}{8\cdot 60\left(\frac{1}{6} + \frac{1}{10} + \frac{1}{20}\right)} = \frac{1\,440}{152} = 9{,}47 \text{ хв.}$$

Результати обчислення за допомогою середньої та за фактичними даними збігаються, отже ця середня є правильною. Вона називається *середньою гармонійною незваженою*. Таким чином, середня гармонійна — зворотне значення середньої зі зворотних значень ознаки, що варіює, тобто коли підсумовуванню підлягають не самі варіанти, а зворотні їм величини: $\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_n}$. Як і середня арифметична вона буває простою та зваженою. Обчислюється середня гармонічна за наступним формулам:

$$\bar{x}_{\text{гарм}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x}} \text{ — незважена;} \quad (7.41)$$

$$\bar{x}_{\text{гарм}} = \frac{f}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots + \frac{f_n}{x_n}} = \frac{\sum f}{\sum \frac{1}{x} f} \text{ — зважена,} \quad (7.42)$$

де $\bar{x}$ — середня величина;

$\frac{1}{x}$ — зворотне значення варіантів ознаки;

$n$ — кількість варіантів ознаки, з яких розраховується середня;

$f$ — вага.

У практиці української статистики середня гармонійна застосовується при обчисленні середньої трудомісткості одиниці виробленої продукції, середньозваженого індексу цін, середньої тривалості будівництва об’єктів та ін.

## 7.6. Середня геометрична

*Середня геометрична* обчислюється шляхом обчислення кореня ступеня, що дорівнює кількості варіантів, мінус одиниця із співвідношення кінцевого та початкового показників розвитку явища.

Середня геометрична розраховується за формулою:

$$\bar{x}_{\text{геом}} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n}. \tag{7.43}$$

Для того, щоб надати формулі більш компактного вигляду, можна записати

$$\bar{x}_{\text{геом}} = \sqrt[n]{\prod_{j=1}^{n} x_j}, \tag{7.44}$$

де $\bar{x}_{\text{геом}}$ — середня геометрична;

$x_j$ — відносні величини динаміки, виражені кратним відношенням $j$-го значення показника до попереднього $(j-1)$-го;

$\prod$ — символ добутку;

$n$ — кількість значень ознаки (число варіантів).

$x_1, x_2, x_3, \ldots, x_n$ — коефіцієнти зростання за період.

Ми взяли формулу середньої геометричної незваженої. Вона найчастіше використовується практично. Для її застосування інтервали часу мають бути однаковими. У випадку, якщо інтервали часу неоднакові, то для розрахунку середніх темпів зростання використовують середню геометричну зважену. Якщо використовувати частоти ($f$), то середня геометрична зважена набуде наступного вигляду:

$$\bar{x}_{\text{геом}} = \sqrt[\Sigma f]{x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_n^{f_n}}. \tag{7.45}$$

Для надання формули компактний вигляд, формулу середньої зваженої геометричної можна записати наступним чином:

$$\overline{x}_{\text{геом}} = \sqrt[\Sigma x]{\prod (x^m)}\ . \qquad (7.46)$$

Обчислення середньої геометричної значною мірою спрощуються застосуванням логарифмування. Для незваженої середньої геометричної $\overline{x}_{гм} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n}$ отримуємо:

$$\lg \overline{x}_{\text{геом}} = \frac{\sum \lg x}{n}. \qquad (7.48)$$

Для зваженої середньої геометричної:

$$\overline{x}_{\text{геом}} = \sqrt[\Sigma f]{\prod (x)^f}\ ; \qquad (7.49)$$

$$\lg \overline{x}_{\text{геом}} = \frac{\sum (\lg x) f}{\sum f}. \qquad (7.50)$$

Таким чином, логарифм середньої геометричної є середня арифметична з варіантів логарифмів (див. формулу середньої арифметичної).

Середня геометрична використовується при обчисленні середніх темпів зміни явища за кілька років. Краще обчислювати її вилученням кореня ступеня, що дорівнює кількості варіантів, мінус одиниця із співвідношення кінцевого та початкового показника розвитку явища.

Розглянемо приклад, коли потрібно обчислити особливу середню, яка називається *середньою геометричною*.

Випуск промислової продукції проводився підприємством у таких розмірах:

Таблиця 7.19

| Показник | Січень I | Лютий II | Березень III | Квітень IV |
|---|---|---|---|---|
| Випуск промислової продукції, млн грн | 10,3 | 11,2 | 11,4 | 12,1 |

Це так званий хронологічний ряд чи ряд динаміки. Потрібно дати характеристику обсягу виробництва. Для характеристики зміни будь-якого явища у часі використовують різноманітні показники. Щоб визначити середній місячний коефіцієнт і темпи зростання промислової продукції, визначає помісячні коефіцієнти зростання $k_p$, які у даному разі і є варіантами:

$$k_p\ (\text{II:I}) = \frac{11{,}2}{10{,}3} = 1{,}08738;$$

$$k_p\ (\text{III:II}) = \frac{11{,}4}{11{,}2} = 1{,}01786;$$

$$k_p\ (\text{IV:III}) = \frac{12{,}1}{11{,}4} = 1{,}06140.$$

Якщо взяти великий період *n* років, то таких індивідуальних коефіцієнтів буде *n* – 1. Тепер ставиться завдання в и з н а ч и т и, як у середньому зростали дана величина (темпи зростання).

Розрахуємо середню з індивідуальних коефіцієнтів зростання. Вона дорівнюватиме 1,056 $\left(\frac{1{,}08738 + 1{,}01786 + 1{,}06140}{3}\right)$. Тепер перевіримо, наскільки отримана середня величина темпи зростання відповідає вихідним показникам.

Для цього виходитимемо з того, що, застосовуючи середній коефіцієнт від року до року, ми повинні отримати зрештою загальну суму виробленої продукції в розмірі 12,13 млн грн. Результати обчислень запишемо у такому вигляді (табл. 7.20).

Таблиця 7.20

**Розрахункова таблиця**

| Період | Сума виробленої продукції, млн грн | Коефіцієнти зростання ($K$) | Сума умовно виробленої продукції, млн грн |
|---|---|---|---|
| I | 10,3 | – | 10,3 |
| II | 21,18 | 1,08738 | 10,3·1,056=10,88 |
| III | 32,67 | 1,01786 | 10,88·1,056=11,49 |
| IV | 44,80 | 1,06140 | 11,49·1,056=12,13 |

Фактична величина виробленої продукції для останнього року та обчислена не збігаюся між собою. Значить, необхідно застосувати іншу формулу середньої — середню геометричну:

$$\bar{x}_{\text{геом}} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n}. \qquad (7.51)$$

Зі знайдених трьох помісячних коефіцієнтів зростання (варіантів) визначаємо середній місячний коефіцієнт зростання ($\bar{k}_p$) за формулою середньої геометричної. Для цього знайдені коефіцієнти зростання перемножуються і з добутку обчислюється корінь третього ступеня:

$$\bar{k}_p = \sqrt[3]{1{,}08738 \cdot 1{,}01786 \cdot 1{,}06140} = \sqrt[3]{1{,}17476} \approx 1{,}055.$$

Це означає, що середньомісячний темп зростання дорівнює 1,055, або 105,5 %. Подивимося тепер, чи придатна ця середня для вираженя відносин, що існують між показниками. Якби щомісячно випуск промислової продукції став зростати на 5,5 %, то у квітні рівень її виробництва досяг би теж 12,1 млн грн, у тому, що це так, переконаємось, здійснивши перевірочний розрахунок (табл. 7.21).

Таблиця 7.21

| Показник | Січень I | Лютий II | Березень III | Квітень IV |
|---|---|---|---|---|
| Випуск промислової продукції, млн грн | 10,3 | $10{,}3 \times 1{,}055=$ $=10{,}9$ | $10{,}9 \times 1{,}055=$ $11{,}5$ | $11{,}5 \times 1{,}055=$ $12{,}1$ |

Примітка. При вилучення коренів здійснюють логарифмування, так $\overline{x}_{\text{геом}} = \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$; $\lg \overline{k}_p = \frac{1}{n} \lg\,(x_1 \times x_2 \times x_3, ..., x_n)$; в нашому прикладі $\lg \overline{k}_p = \frac{1}{n} \lg 1{,}17476$; (значення $\bar{k}_p$ знаходять за таблицею логарифмів).

Обчислена величина дорівнює фактичної. При розрахунку середньої геометричної ми зв'язали значення показника на початку і на кінець періоду, тобто знаючи початковий рівень показника і середню геометричну, можна обчислити кінцевий показник низки динаміки. Нас не повинно бентежити, що середня геометрична, обчислена за кожний період, відрізняється від фактичного коефіцієнта зростання. Середня величина як у цьому, так і у інших випадках не може і повинна збігатися з індивідуальним значенням кожної ознаки, оскільки вона є середня, тобто узагальнююча характеристика розвитку явища чи процесу.

## 7.7. Середня квадратична

*Середня квадратична* отримується, якщо формулу степеневої середньої підставити $k$=2:

$$\overline{x_{\kappa\varepsilon}} = \sqrt{\frac{\sum x^2}{n}} = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + ... + x_n^2}{n}}, \tag{7.52}$$

середня квадратична незважена і

$$\overline{x_{\kappa\varepsilon}} = \sqrt{\frac{\sum x^2 \cdot f}{\sum f}} = \sqrt{\frac{x_1^2 \cdot f_1 + x_2^2 \cdot f_2 + x_3^2 \cdot f_3 + ... + x_n^2 \cdot f_n}{f_1 + f_2 + f_3 + ... + f_n}}, \tag{7.53}$$

середня квадратична зважена.

Середня квадратична застосовується тільки в тих випадках, коли варіанти представляють собою відхилення фактичних величин від їх середньої арифметичної або від заданої норми.

Маються результати виміру відхилень фактичної довжини виробів від заданої норми. (табл. 7.22).

Таблиця 7.22

| Відхилення фактичної довжини виробів від заданої норми (в *мм*) ($x$) | Кількість виробів ($f$) | $x^2$ | $x^2 \cdot f$ |
|---|---|---|---|
| –3,6 | 2 | 12,96 | 25,92 |
| –1,6 | 6 | 2,56 | 15,36 |
| +0,4 | 8 | 0,16 | 1,28 |
| +2,4 | 2 | 5,76 | 11,52 |
| +4,4 | 2 | 19,36 | 38,72 |
| Всього | 20 | – | 92,80 |

О б ч и с л и м о середню величину відхилень.

Знаходимо середню квадратичну зважену; при цьому обчислює в табл. 13 колонки 3 та 4:

$$\overline{x_{кв}} = \sqrt{\frac{\sum x^2 \cdot f}{\sum f}} = \sqrt{\frac{92,80}{20}} = \sqrt{4,64} \approx 2,15 \ мм.$$

Отже, середня величина відхилень фактичної довжини виробів заданої норми становить 2,15 *мм*. У цьому випадку середня арифметична була б непридатна, тому що в результаті ми отримали б нуль:

$$\overline{x_{ар}} = \frac{\sum xf}{\sum f} = \frac{(-3,6)\cdot 2 + (-1,6)\cdot 6 + (+0,4)\cdot 8 + (2,4)\cdot 2 + (4,4)\cdot 2}{20} =$$

$$= \frac{-16,8 + 16,8}{20} = 0.$$

Середня квадратична має обмежену сферу використання. Головною сферою застосування середньої квадратичної є вимір коливання (варіації) ознак, про що йтиметься в гл. 8. На практиці за допомогою середньої квадратичної визначається середні діаметри коліс, труб, середні боки квадратів.

## 7.8. Структурні середні

Крім розглянутих вище середніх величин є інші показники, які дають деяке уявлення про структуру сукупності, наприклад, мода і медіана. Медіану та моду відносять до середніх величин, але між ними та середньою арифметичною є суттєва відмінність.

Величина середньої арифметичної визначається сукупністю значень ознаки у даному ряду розподілу. Для наочності представимо деяку статистичну сукупність, в якій всі спостережені значення ознаки замінимо середньою арифметичною. У цій сукупності $\sum xf = \sum \bar{x} f$, тобто сукупність значень ознаки не зміниться. Такою властивістю ні медіана, ні мода не мають, тому що їх величини не визначаються всіма значеннями ознаки, що зустрічаються в даному розподілі з різними частотами. Ці показники визначаються лише структурою розподілу. Тому їх іноді називають *структурними позиційними середніми*. Моду і медіану використовують у тих випадках, коли обчислення середньої степеневої є неможливим або недоцільним.

**Мода**. *Мода* (*Мо*) є величиною ознаки (варіанту), яка найчастіше зустрічається в сукупності або в варіаційному ряду (модальна ціна одиниці товару, модальний розмір взуття і т. д.). В ряду розподілу це буде величина, що має найбільшу вагу — частоту. Таким чином, мода представляє собою найбільш типове значення випадкової величини. Якщо ряд розподілу подати графічно, то модою буде абсцис вершини його кривої.

Мода по-різному визначається для дискретного та інтервального рядів.

**Знаходження моди у дискретному варіаційному ряду**. У дискретному варіаційному ряду, де значення ознаки задані числовими значеннями, мода визначається візуально, тобто відшукується переглядом чисельностей, які мають варіанти (значення) ознаки. Значення ознаки, що має чисельність (в абсолютному або відносному виразі), більшу, ніж будь-яке інше його значення, і є мода.

Для дискретного ряду мода визначається за частотами варіантів та відповідає варіанту з найбільшою частотою.

Приклад. При визначенні якості насіння жита було отримано наступний розподіл насіння за відсотком схожості:

| Процент схожості | 70 | 85 | 80 | 83 | 85 | 90 | 92 | 93 | Понад 93 |
|---|---|---|---|---|---|---|---|---|---|
| Число проб в процентах до підсумку | 0,5 | 0,5 | 6,0 | 12 | 30 | 40 | 7 | 2 | 2 |

Найбільше проб у процентах до підсумку (40 %) має варіант з 90 % схожістю насіння жита. Це і буде мода — варіант, що має найбільшу частоту, або, інакше варіант, що має найбільшу чисельність у розподілі. До моди вдаються виявлення величини ознаки, має найбільше поширення (наприклад, вік осіб, розмір взуття, стаж роботи тощо).

У моді не погашаються індивідуальні відхилення варіантів, тому вона має описовий характер. Вона завжди відповідає певному варіанті. Тому мода не вимагає свого знаходження розрахунків, якщо відомі значення ознаки.

**Розрахунок моди у інтервальному варіаційному ряду**. В інтервальному варіаційному ряду *модою* вважають таке значення ознаки, на яке припадає найбільша щільність розподілу (тобто навколо якого концентрується найбільша частота).

Для інтервального ряду визначення моди складніше. У разі інтервального розподілу з рівними інтервалами модальний інтервал (тобто мода, що містить) визначається за найбільшою частотою, а при нерівних інтервалах — по найбільшій щільності.

Якщо варіанти в ряду розподілу задані у вигляді інтервалів, то спочатку знаходиться модальний інтервал, тобто інтервал, що має найбільшу частоту, а потім — наближене значення модальної величини ознаки за формулою:

$$Mo = x_0 + i\frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})}, \qquad (7.54)$$

де $Mo$ — мода;
$x_0$ — нижня межа модального інтервалу;
$i$ — величина модального інтервалу;
$f_{m-1}$ — частота інтервалу, що передує модальному;
$f_m$ — частота модального інтервалу;
$f_{m+1}$ — частота інтервалу наступного за модальним.

Розглянемо приклад розрахунку моди. За даними вибіркового обстеження собівартість одиниці однойменної продукції на підприємствах галузі характеризується такими показниками (табл. 7.23).

Щоб знайти моду, спочатку визначають модальний інтервал. Знаходимо найбільшу чисельність. З прикладу видно, що найбільша кількість підприємств (10) має собівартість одиниці продукції в інтервалі 3,2–3,6. Отже, цей інтервал є модальним інтервалом ряду розподілу. Введемо такі позначення:

Таблиця 7.23

| Групи підприємств за собівартістю одиниці продукції, грн | Число підприємств | Групи підприємств за собівартістю одиниці продукції, грн | Число підприємств |
|---|---|---|---|
| 1,6—2,0 | 2 | 2,8—3,2 | 7 |
| 2,0—2,4 | 3 | 3,2—3,6 | 10 |
| 2,4—2,8 | 5 | 3,6—4,0 | 3 |
| | | | 30 |

$$x_0 = 3{,}2; \qquad f_{m-1} = 7;$$
$$i = 0{,}4; \qquad f_{m+1} = 3.$$
$$f_m = 10;$$

Підставимо числові значення з нашого прикладу у формулу моди і зробимо обчислення:

$$Mo = x_0 + i\frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} =$$
$$= 3{,}2 + 0{,}4\frac{10-7}{(10-7)+(10-3)} = 3{,}2 + 0{,}4\frac{3}{3+7} = 3{,}32 \text{ грн.}$$

Таким чином, можна сказати, що мода, або значення показника собівартості продукції, що найчастіше зустрічається, для 30 підприємств складе 3,32 грн.

Зміст цієї формули полягає в наступному: величину тієї частини модального інтервалу, яку потрібно додати до її мінімальної межі, визначають залежно від величини частот попереднього та наступного інтервалів. В даному випадку до 3,2 додаємо 0,12, тобто більше половини інтервалу (0,4), тому що частота наступного інтервалу 3 більша за частоту попереднього інтервалу (7).

Формула (7.54) придатна для розподільних рядів з рівними інтервалами. Зауважимо, що з обчисленні моди для дискретних рядів останні часто призводять до інтервальної форми і користуються формулою (7.54). Цим шляхом досягається більша точність у визначенні моди.

Можуть бути розподіли, де варіанти зустрічаються однаково часто, у цьому випадку моди немає. У варіаційному ряду може бути кілька мод. Слід зазначити, що може бути так, що не одна, а два варіанти можуть мати однакову, найбільшу в порівнянні з іншими чисельність. В цьому випадку у варіаційному ряду буде дві моди. Ряд із двома модами називають *бімодальним* (*двохмодальним*). Бімодальний розподіл може сигналізувати про наявність якісно неоднорідних елементів сукупності за досліджуваною ознакою. Тоді слід перевірити однорідність об'єднуваних у такому ряду явищ, оскільки багатомодальність може бути сигналом про наявність неоднорідних елементів.

Модою користуються у випадках, коли обчислення середньої арифметичної просто беззмістовне. Так наприклад, не можна обчислити середній обсяг випуску одягу чи взуття. Тут слід обчислювати модальну величину. Статистика цін також дуже часто вдається до модальних величин.

Мода знаходить широке застосування для вирішення практичних завдань. Так, при плануванні масового випуску одягу та взуття встановлюється розмір продукції, що користується найбільшим способом (модальний розмір). При вивченні змін цін (тарифів) на споживчі товари береться модельна ціна, тобто ціна, за якою продається максимальна кількість товарів того чи іншого виду.

Мода має важливе значення при її зіставленні із середньою величиною. Чим ближче значення обчисленої середньої до моди, тобто до величини, що найбільше зустрічається в дійсності, тим, отже, середня більш типова і рівновіддалена від крайніх значень ознаки.

**Медіана**. В якості характеристики варіаційного ряду застосовується медіана. *Медіаною* (*Me*) називається величина ознаки у одиниці, що знаходиться в середині ранжованого (упорядкованого) ряду. Тобто це варіанта, що кількісно ділить навпіл ряд, розташований у зростаючому або спадному порядку за значеннями ознаки.

Медіана цікава тим, що показує якісну межу значення ознаки, що варіюється, яку досягла половина членів сукупності.

Як і мода, медіана по-різному визначається для дискретного та інтервального рядів розподілу. Розглянемо кілька прикладів визначення медіани з різних рядів. Почнемо дуже елементарних прикладів, які допоможуть більш ясному уявленню про медіану і способи її визначення.

**Розрахунок медіани у дискретному варіаційному ряду**. Якщо ряд розподілу дискретний і має непарне число членів, то медіаною буде варіанта, що знаходиться в середині впорядкованого ряду (упорядкований ряд — це розташування одиниць сукупності у зростаючому або спадаючому порядку). Так, якщо непарний варіаційний ряд складається з $2m+1$ випадку, значення ознаки у випадку $m + 1$ буде медіанним.

Формули для обчислення медіани при непарному та парному числі варіантів:

$$Me = x_{m+1}; \tag{7.55}$$

$$Me = \frac{x_m + x_{m+1}}{2}. \tag{7.56}$$

Приклад. Вік дев'яти одиниць обладнання інструментального цеху, розташованих у порядку зростання, становив (років):

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_{9=2m+1}$ |
|---|---|---|---|---|---|---|---|---|
| 3 | 4 | 4 | 6 | 7 | 9 | 11 | 12 | 15 |

О б ч и с л и т и медіану.

Маємо непарну кількість варіантів:

$$2m+1=9;$$

$$2m=8;$$

$$m=4.$$

Знаходимо медіану:

$$Me = x_{4+1} = x_5 = 7.$$

У такому впорядкованому ряді медіана — $x_5$ зі значенням ознаки 7 років. По обидва боки від неї є однакова кількість обладнання.

Можна знайти медіану іншим способом. Для цього потрібно суму частот розділити навпіл і до отриманого результату додати ½. Так, у розподілі 9 одиниць обладнання інструментального цеху медіаною буде: $\frac{9}{2}+\frac{1}{2}=5$, тобто 5-а варіанта ($x_5$), яка ділить впорядкований ряд навпіл.

Якщо впорядкований ряд складається з парного числа членів, то медіаною буде середня арифметична з двох варіантів, розташованих у середині ряду. Розрахунок здійснюється за формулою:

$$Me = \frac{\sum f}{2}. \tag{7.57}$$

Нехай тепер буде не дев'ять одиниць обладнання, а дванадцять:

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | $x_{12=2m}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 4 | 4 | 6 | 7 | 9 | 11 | 12 | 15 | 17 | 20 | 23 |

Маємо парну кількість варіантів:

$$2m=12;$$

$$m=6.$$

У цьому ряді є два варіанти, що стоять у центрі ряду. Це варіанти 9 і 11. Середня арифметична з цих значень буде медіаною ряду.

Знаходимо медіану:

$$Me = \frac{x_6 + x_7}{2} = \frac{9+11}{2} = 10 \text{ років.}$$

Щоб знайти медіану невпорядкованого ряду (такого ряду, де порядок запису членів не залежить від їх величини), його значення мають попередньо за зростанням або спаданням.

Наприклад, середня кількість працівників цеху за місяцями на рік виражалася в таких цифрах: січень — 12, лютий — 13, березень — 17, квітень — 14, травень — 16, червень —15, липень — 18, вересень — 19, жовтень — 20, листопад — 22, грудень — 21 (передбачається, що в серпні цех не працював).

Розташуємо чисельність працівників цеху по місяцях у вигляді ранжованого ряду за зростанням кількості працівників (табл. 7.24).

Таблиця 7.24

| № п/п | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Назва місяця | січень | лютий | квітень | червень | травень | березень | липень | вересень | жовтень | грудень | листопад |
| Чисельність працівників | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |

Чисельність працівників у 6-му місяці (середина з 11 місяців), тобто в березні, дорівнює 17, і буде медіаною.

Якщо варіаційний ряд має парне число членів (наприклад, якщо розглядати чисельність працівників не за 11, а за 12 місяців), то медіана дорівнює напівсумі двох середніх варіантів. У нашому прикладі — це півсума за 6-й та 7-й місяці (тобто $\frac{17+18}{2}$=17,5 осіб).

На відміну від середньої арифметичної, мода та медіана можуть бути визначені майже без будь-яких обчислень, що робить їх зручним для застосування у дослідженні економічних процесів.

Розглянемо приклад розрахунку медіани в дискретному ряду, що містить частоти (частості) значень варіантів.

Маються такі дані про заробітну плату робітників-відрядників:

Таблиця 7.25

| Місячна заробітна плата, грн | Число працівників | Сума нагромаджених частот |
|---|---|---|
| 1 | 2 | 3 |
| 1 800 | 2 | 2 |
| 2 000 | 9 | 11 (2+9) |
| 2 300 | 15 | 26 (11+15) |
| 2 700 | 13 | |
| 3 000 | 12 | |
| Всього | 51 | |

Для визначення медіани слід підрахувати суму накопичених частот ряду (гр. 3 табл. 7.25). Нарощування підсумку триває до отримання нагромадження суми частот, що перевищує половину. Для того, щоб це з'ясувати, потрібно нагромадити частоти, починаючи з найменшої варіанти. Сума частот 1-ї та 2-ї варіант дорівнює 11. Якщо додати до 11 частоту 3-ї варіанти, отримаємо суму, рівну 26 (11+15). Отже, 26-а варіанта відповідає третьому значенню ознаки, що варіюється, і медіаною буде працівники, які мають заробітну плату 2 300 грн.

Знайти медіану можна й іншим способом. У прикладі сума частот становила 51. Так як у розподілі 51 працівників за розміром заробітної плати медіаною буде: $\frac{51}{2}+\frac{1}{2}=26$, тобто 26 варіанти ділить впорядкований ряд навпіл. Варіанта, що відповідає цій сумі, тобто 2 300 грн, є медіана ряду.

Якби ми мали парну суму частот (скажімо 50), то, застосовуючи зазначену формулу, ми отримаємо номер медіанної варіанти 25,5 $\left(\frac{50}{2}+\frac{1}{2}\right)$. Оскільки варіанти з дробовим номером не буває, отриманий варіант вказує, що медіана знаходиться посередині між 25-м та 26-м варіантами.

Змінивши значення частот за умови попередньої типової задачі, розрахуємо медіану:

Медіана дорівнюватиме:

$$Me=\frac{2\,300+2\,700}{2}=2\,500\,\text{грн.}$$

Особливістю медіани і моди є те, що в них не погашаються

Таблиця 7.26

| Місячна заробітна плата, грн | Число працівників | Сума нагромаджених частот |
|---|---|---|
| 1 800 | 2 | 2 |
| 2 000 | 8 | 10 (2+8) |
| 2 300 | 15 | 25 (10+15) |
| 2 700 | 13 | — |
| 3 000 | 12 | — |
| Всього | 50 | |

індивідуальні відхилення. Тому мода та медіана у дискретних рядах не вимагають для свого знаходження розрахунків. Проте в інтервальному варіаційному ряду для знаходження медіани вдаються до розрахунків.

**Розрахунок медіани у інтервальному варіаційному ряду**. Набагато складніше визначити медіану в інтервальному варіаційному ряді. При обчисленні медіани інтервального варіаційного ряду спочатку знаходять інтервал, що містить медіану, шляхом використання нагромаджених частот або частостей. Медіанному інтервалу відповідає перша з нагромаджених частот або частостей, що перевищує половину всього обсягу сукупностей.

Візьмемо ряд розподілу (табл. 7.27) та пояснимо спосіб розрахунку медіани в інтервальному ряду.

Таблиця 7.27

| Групи підприємств за собівартістю одиниці продукції, грн | Число підприємств | Сума накопичених частот (кумулятивні частоти) |
|---|---|---|
| 1 | 2 | 3 |
| 1,6–2,0 | 2 | 2 |
| 2,0–2,4 | 3 | 5 (2+3) |
| 2,4–2,8 | 5 | 10 (5+5) |
| 2,8–3,2 | 7 | 17 (10+7) |
| 3,2–3,6 | 10 | – |
| 3,6–4,0 | 3 | – |
| Всього | 30 | |

У ряді розподілу, що розглядається, парне число членів — 30. Отже, значення ознаки, яка є медіаною, має 15-й член, тобто (30:2), якщо рахувати з початку ряду. Для того, щоб визначити, в якому інтервалі знаходиться цей член ряду, необхідно скласти ряд так званих нагромаджених частот, підсумовуючи частоти з початку ряду. У перших двох інтервалах міститься 5 членів ряду, у перших трьох — 10 і т. д. Таке підсумовування частот зроблено лише для чотирьох інтервалів, оскільки сума частот у них вже перевищує 15.

Таким чином, встановлено, що 15-й член ряду приходиться на четвертий інтервал, тобто на інтервал «2,8–3,2». Але в цьому інтервалі знаходиться всього 7 членів ряду, серед якого і 15-й з початку ряду член носить, якщо рахувати вже з початку інтервалу, порядковий номер 5, тобто 15 – 10. Якщо допустити далі, що всі 7 членів ряду розподіляються в інтервалі рівномірно, кожен з них «зсуває» інтервал на 0,057, тобто 0,4:7. А 5 членів інтервал, останній з яких є власником медіани, «зсувають» інтервал на 0,29, тобто 0,057×5. Таким чином, медіана дорівнює: 2,8 (початок інтервалу) + 0,29 = 3,09.

Медіана інтервального варіаційного ряду розподілу визначається за такою формулою:

$$Me = x_{Me} + i_{Me}\frac{\frac{\sum f}{2} - S_{Me-1}}{f_{Me}}, \qquad (7.58)$$

де $x_{Me}$ — початкове значення інтервалу, що містить медіану;

$i_{Me}$ — величина медіанного інтервалу;

$\sum f$ — сума частот ряду;

$f_{Me}$ — частота медіанного інтервалу;

$S_{Me-1}$ — сума нагромаджених частот, що передують медіанному інтервалу.

З цієї формули слідує, що до нижньої межі медіанного інтервалу ($x_0$) додається та частина медіанного інтервалу, яка пропорційна питомій вазі в частоті медіанного інтервалу частини її, розташованої від нижньої межі інтервалу до медіани (*Me*). Слід зазначити, що сума абсолютних величин лінійних відхилень від *Me* мінімальна, що дуже важливо за практичного застосування такого показника, як медіана.

Визначимо її значення за наведеною вище формулою. Відомо що:

$x_{Me}$=2,8; $i_{Me}$=0,4; $\sum f$=30; $S_{Me-1}$=10; $f_{Me}$=7.

Підставивши в цю формулу необхідні величини, отримаємо обчислене значення медіани:

$$Me = x_{Me} + i_{Me}\frac{\frac{\sum f}{2} - S_{Me-1}}{f_{Me}} =$$

$$= 2{,}8 + 0{,}4\frac{\frac{30}{2} - 10}{7} = 2{,}8 + 0{,}4\frac{15-10}{7} = 2{,}8 + 0{,}29 = 3{,}09.$$

Що означає цей показник? Він говорить про те, що у 50 % підприємств собівартість продукції менша за 3,09 грн, а у 50 % — більш ніж 3,09 грн. Таким чином, медіана може доповнити ту узагальнену характеристику сутності фактів, яка дається за допомогою середньої арифметичної.

Порядок розрахунку медіани за формулою (7.58): спочатку визначаємо $\frac{\sum f}{2}$, потім $S_{Me-1}$, і нарешті, всі інші величини — $x_{Me}$, $f_{Me}$ і $i_{Me}$.

Формула (7.58) придатна і у разі рядів розподілу з нерівними інтервалами, $i_{Me}$ у ній тоді позначає довжину медіанного інтервалу.

За якою б формулою ми не розраховували медіану, її суть від цього не змінюється. Медіана — значення ознаки, що знаходиться в середині ряду і, отже, розділяє варіаційний ряд на дві рівні частини за сумою частот.

Медіана має ту властивість, що сума абсолютних величин відхилень варіантів від медіани менша, ніж від будь-якої іншої величини (у тому числі і від середньої арифметичної):

$$\sum |x - Me| = \min . \qquad (7.59)$$

Відмінність між модою та медіаною полягає в наступному. Мода є значенням ознаки, що зустрічається частіше, ніж всі інші. Медіанна теж характеризує значення ознаки, але значення ознаки, що знаходиться в середині ряду і, отже, ділить ряд на дві рівні частини. Медіанна, як і мода, по-різному визначається для дискретного та інтервального рядів розподілу.

Велике значення має медіана для оцінки середньої величини, тому що, чим ближче значення обчисленої середньої до медіани, тим, отже, реальніше і типовіше отримана середня. Однак близькість до моди більшою мірою характеризує реальність та типовість середньої величини, ніж її наближення до медіани.

## 7.9. Застосування середніх в економіко-статистичних дослідженнях

Середні величини служать важливим інструментом вивчення соціально-економічних процесів і явищ. Величезне пізнавальне значення середніх полягає в об'єктивно-точному вимірюванні досліджуваних економічних явищ і в практичній цінності як одного з могутніх знарядь розрахунків. Класики політичної економії дуже часто використовували середні величини для економічних досліджень. Так, К. Маркс в економічних дослідженнях часто користувався поняттям середніх величин для аналізу процесу розвитку капіталізму, коли писав про середню норму прибутку, середній рівень заробітної плати, середню тривалість робочого дня. «Ясно у всякому випадку, що сукупний робочий день великої кількості одночасно зайнятих робітників, будучи розділений на число робітників, є вже сам по собі днем суспільної середньої праці» (т. 23, с. 334).

У практичній діяльності середні величини використовуються для узагальнюючої характеристики однорідних явищ і процесів, оскільки середня відображає те загальне, що характеризує індивідуальні особливості кожної одиниці сукупності, наприклад, середня заробітна плата, середня витрата матеріалів, середня ціна, середній тарифний розряд і т. п.

Вивченню процесів розвитку галузей народного господарства, підприємств, і навіть дослідженню виробництва всередині підприємств передує, як відомо, поділ маси одиниць на якісно однорідні групи. Групування відкривають широкі можливості для використання середніх величин при характеристиці різних форм і типів суспільних явищ, дозволяють уникнути застосування розрахунку «загальних» середніх, що не розкривають якісних особливостей явищ та процесів. Наприклад, при розподілі доходів населення виявляють соціальні групи, що відрізняються за величиною доходу. Тому поряд із середньою необхідно враховувати особливості формування доходів кожної соціальної групи.

Серед середніх величин на практиці найбільшого поширення набула середня арифметична. Популярність середньої арифметичної складається під впливом кількох причин. Середня арифметична має простий зміст щодо двох сумарних величин, які разом із середньою утворюють систему взаємозалежних показників. Середні витрати матеріалу на виробництво виробу ми обчислюємо за формулою середньої арифметичної з витрат матеріалів на виготовлення окремих виробів. У чисельнику середньої — витрати матеріалів, у знаменнику — кількість виробів.

У розрахунках при складанні перспективних і поточних планів розвитку підприємства часто користуються фактичними та нормативними середніми величинами, наприклад, середніх витрат на одиницю вироблюваного виробу, середнього тарифного розряду робітників, середньої заробітної плати, середніх (питомих) витрат матеріалів на одиницю продукції. Ці величини пов'язані у формулі середньої арифметичної простим співвідношенням, з якого знаючи два показники, можна визначити третій. Так, у практиці планування фонд заробітної плати визначається як добуток числа робітників на середню заробітну плату. Ні мода, ні медіана таких властивостей не має. В окремих випадках середні виступають вже як самостійно планований і контрольований показник, що виражає загальну характеристику того чи іншого соціально-економічного явища (наприклад, середня урожайність, середня ціна тощо).

Контроль і аналіз виконання планових завдань ґрунтується на використанні статистичних показників, до яких належать середні величини. На практиці для контролю виконання планових завдань часто використовуються такі середні величини, як середній час роботи обладнання, середня норма виробітку, норма витрати матеріалу на виготовлення виробу, середній відсоток виконання планового завдання та ін. Використання середніх дозволяє визначити ступінь досягнення планових показників, здійснювати контроль за виконанням управлінських рішень, виявити резерви зростання виробництва і підвищення його ефективності.

На основі середніх величин, наприклад, середньої заробітної плати, середньої врожайності, середньої собівартості та ін., обчислених за ряд періодів часу, визначають основні тенденції (тренди) розвитку суспільних явищ, зокрема, процесів виробництва продукції. За формулою середньої геометричної обчислюється інтенсивність змін рівня динаміки. Вимірювання сезонних коливань рівня рядів динаміки також ґрунтується на розрахунку відхилень від середньої арифметичної або від інших характеристик центру розподілу (медіани).

Середні величини широко залучаються для прогнозування розвитку виробництва, тенденцій у зміні попиту, зокрема, при вирівнюванні емпіричних рівнів за способом найменших квадратів, безпосередньо пов'язаного із середніми величинами.

У практиці статистики широко застосовується індексний метод, який представляє собою подальший розвиток методу середніх величин (див. гл. 11). Кількісне визначення впливу окремих ознак-факторів на результативні ознаки, наприклад, вплив заробітної плати на обсяг продукції, вивчається зокрема, за допомогою методів дисперсного, кореляційного та регресійного аналізу, що базуються на середніх величинах (див. р. 9).

Середні величини часто використовуються при проведенні економічного аналізу виробничо-господарської діяльності підприємства. Зокрема, розподіл робітників підприємства (одиниць сукупності) на однорідні групи за ступенем виконання норм виробітку, та обчислення для кожної групи середньої величини (групової середньої) дозволяє виявити як відстаючі, так і передові ділянки. Це дозволяє визначити резерви підвищення продуктивності праці.

Середні величини необхідні при управлінні поточною діяльністю підприємств, зокрема, при прийнятті оперативних рішень та контролю за їх виконанням. До таких середніх відносяться середня норма запасу матеріалів, середня тривалість виробничого циклу, середній час ремонту обладнання та ін. Для контролю управлінських рішень також залучаються середні величини.

Значно менше, але досить значне застосування в економічних дослідженнях має медіана. Медіана отримала найбільше застосування у практиці статистичного контролю якості продукції. Контрольованими показниками служать різні параметри — дотримання рецептури або хімічного складу, геометричних розмірів і форм деталей, припустимо по 5 шт. у кожній пробі. На карту у вигляді точок наносяться контрольовані показники, наприклад, діаметр виробів. Так як результати вимірів виявляються на шкалі впорядкованими за величиною параметра, то знаходження медіани для кожної проби проводиться на око, без жодних розрахунків. Спостерігаючи за становищем медіан на контрольній карті, робітник отримує можливість систематично стежити за станом показників якості продукції, не допускаючи раптового і, що особливо важливо, непоміченого розладу обладнання чи іншого відхилення від нормального ходу виробничого процесу.

Нерідко трапляються випадки, коли у ряді розподілу зустрічаються окремі варіанти, значення яких істотно відрізняється від основної маси значень ознаки. Таке явище спостерігається при встановленні норм виробітку або норм витрати матеріалів на виготовлення продукції. У цьому випадку середня арифметична не буде типовою для даного ряду, а її використання може призвести до суттєвих помилок. Якщо виключити такі «сумнівні» варіанти з подальших розрахунків, як це іноді застосовується на практиці, це призводить до значних спотворень. Медіана, як відомо, не залежить від того, як далеко відстоять від основної маси випадків окремі варіанти. Тому в таких випадках кращим є обчислення медіани

Наведені дані не вичерпують всіх можливостей застосування середніх величин, але достатньо показують їх значення для аналізу розвитку народного господарства.

## ЛІТЕРАТУРА


1. *Горкавий В. К. Статистика*: Підручник. Третє вид., переробл. і доповн. Київ: Алерта, 2019. С. 88–103.
2. Статистика: підручник / *С. І. Пирожков, В. В. Рязанцева, Р. М. Моторин* та ін. Київ: Київ. нац. торг.-екон. ун-т, 2020. С. 105–123.
3. *Ткач Є. І.*, *Сторожук В. П.* Загальна теорія статистики: підручник. 3-тє вид. Київ: Центр учбової літератури, 2009. С. 131–155.


## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники, словники*


4. Загальна теорія статистики: підруч. [для студентів ВНЗ] / *Є. І. Ткач, В. П. Сторожук*. 3-тє вид. Київ : Центр учбової літератури, 2018. 440 с.
5. *Козирєва О. В.*, *Федорова В. О.* Статистика: навчальний посібник. Харків: Видавництво Іванченка І.С., 2021. С. 54–66.
6. *Мармоза А. Т.* Теорія статистики: підручник. 2-ге вид., перероб. і доп. Київ: «Центр учбової літератури», 2013. С. 141–170.
7. *Опря А. Т.* Статистика (модульний варіант з програмованою формою контролю знань). Навч. посіб. Київ: Центр учбової літератури, 2012. С. 73–82.
8. *Педченко Г. П.* Статистика: Навчальний посібник. Мелітополь: Колор Принт, 2018. С. 100–108.
9. Статистика - Statistics: підручник / С. В. Заєць, В. М. Томіленко ; Нац. ун-т держ. податк. служби України. Ірпінь : Нац. ун-т ДПС України, 2015. 510 с.
10. Статистичний словник / [*О. Г. Осауленко*, *О. О. Васєчко*, *М. В. Пугачова* та ін.]; за ред. д-ра держ. упр., проф., член-кор. НАН України О. Г. Осауленка; НТК статистичних досліджень. Київ: ДП «Інформ.-аналіт. агентство», 2012. 498 с.

11. *Шапочка М. К.*, *Маценко О. М.* Теорія статистики: навч. посібник. Суми: Університетська книга, 2014. С. 125–144.

Г Л А В А 8

# ХАРАКТЕРИСТИКА РЯДІВ РОЗПОДІЛУ. ПОКАЗНИКИ ВАРІАЦІЇ

## 8.1. Поняття про варіацію ознак

Середні величини у статистиці мають важливе теоретичне та практичне значення. Вони дають узагальнюючу характеристику сукупності за ознаками, що варіюють, показують типовий для даних умов рівень цих ознак. Однак для повної та всебічної характеристики досліджуваних явищ та процесів однієї середньої виявляється недостатньо. Статистичним величин характерно зміна (коливання) значень ознаки в межах сукупності, які дуже часто істотно відрізняються від середніх величин. У цьому цікаві не тільки крайні відхилення, а й сукупність всіх відхилень. Вивчення цих відхилень дозволяє більш повно і глибоко досліджувати явища та процеси, що відбуваються у суспільному житті.

*Варіацією* називається коливання ознаки, мінливість величини ознаки в одиниць, що входять до складу сукупності. У ній виражається та дія множини випадковостей, яка характерна для масових явищ. Наприклад, коливання норм виробітку в окремих робітників на заводі, коливання урожайності на різних ділянках в районі і т. п.

Причини варіації лежать у природі масових суспільно-економічних явищ. Всім суспільним явищам притаманні специфічні закономірності розвитку. Кожне явище в будь-який момент перебуває в певній стадії, що у значною мірою визначається законом його розвитку. Крім внутрішніх причин у варіації виражається результат дії багатьох випадкових подій, характерних для масового суспільного явища. До цієї групи відносять різноманітні зовнішні умови та обставини, що діють незалежно один від одного в різних напрямках та з різною інтенсивністю. До таких факторів можна віднести аварії на виробництві, зрив поставок сировини, простої через хвороби робітників і т. п.

Таким чином, наявність відмінностей у значеннях ознаки викликається сукупністю взаємодіючих необхідних та випадкових, внутрішніх та зовнішніх причин. При цьому характер та напрямок їх дії дуже часто відбувається у різних напрямках. Зокрема, дія випадкових причин позначається по-різному на внутрішню тенденцію розвитку явища. Так, досвід роботи машинобудівних заводів показує, що аварія обладнання несе велику шкоду слабким підприємствам, які не мають матеріальних та фінансових резервів зростання виробництва. Отже, необхідність, закон не тільки визначає середній рівень явища, що йшлося на початку глави, а й істотно регулює вплив дії різноманітних випадкових чинників, у тому мірою які призводять до варіації ознак.

Отже, поряд з обчисленням середньої величини при аналізі статистичного розподілу важливу роль відіграє оцінка варіації (коливання) ознаки в цьому розподілі. Розміри коливань або варіація ознаки, у відношенні якої розраховується середня, можуть бути більшими або меншими, що часто є важливою характеристикою особливо якості групи і характеристикою надійності самої середньої.

Візьмемо два абстрактних приклади, у яких варіанти однакові, а розподіл частот по-різному (табл. 8.1).

Таблиця 8.1

| Приклад 1 | | | Приклад 2 | | |
|---|---|---|---|---|---|
| *x* | *f* | *xf* | *x* | *f* | *xf* |
| 1 | 1 | 1 | 1 | 35 | 35 |
| 2 | 4 | 8 | 2 | 15 | 30 |
| 3 | 25 | 75 | 3 | 5 | 15 |
| 4 | 62 | 248 | 4 | 60 | 240 |
| 5 | 25 | 125 | 5 | 5 | 25 |
| 6 | 4 | 24 | 6 | 15 | 90 |
| 7 | 1 | 7 | 7 | 35 | 245 |
| – | 122 | 488 | – | 170 | 680 |

Середня в обох випадках однакова:

$$\bar{x}_1 = \frac{488}{122} = 4; \qquad \bar{x}_2 = \frac{680}{170} = 4.$$

Значення середньої величини обох прикладах однаково (4), але відхилення від середніх мають різний характер. У першому прикладі окремі показники близько примикають до середнього рівня, показники другого прикладу відрізняються сильною мінливістю або коливанням. Так, у першому прикладі 112 (25 + 62 + 25) випадків із 122 (тобто 93,3 %) відхиляються від середньої лише на одиницю, у другому прикладі лише 70 (5 + 60 + 5) випадків з 170 (тобто 41 %) відхиляються не більше одиниці. Однак у першому варіанті варіація ознак значно менше, ніж у другому, і можна сказати, що перший варіант за своїм складом щодо досліджуваної ознаки більш однорідний, ніж другий. Це явище — більші або менші коливання значень варіантів ознаки не вловлюється за допомогою використання тільки середніх величин, так як в обох сукупностях величина середньої виявилася однаковою. Якщо значення ознаки сильніше відхиляються від середньої (як у другому прикладі), то узагальнююча варіація перебуває під впливом різноманітних умов і сукупність менш однорідна. Отже, і середня величина, що характеризує цю менш однорідну сукупність, менш надійна.

Наведемо розподіл робітників за тарифними розрядами, підрахувавши, скільки елементів сукупності володіє тим чи іншим значенням ознаки (табл. 8.2). У таблиці нижня межа варіації становить 2 розряд, а верхня — 6 розряд.

Таблиця 8.2

**Розподіл робітників механічного цеху за тарифними розрядами за 14 грудня 2021 р.**

| Тарифний розряд | Число робітників |
|---|---|
| 2 | 2 |
| 3 | 5 |
| 4 | 8 |
| 5 | 3 |
| 6 | 2 |
| Всього | 20 |

Зміни величин статистичної сукупності мають межі. Які ж причини коливань ознаки, що варіюється? Так, наприклад, змінний виробіток токаря механічного цеху визначається рівнів кваліфікації робітника, технічним станом обладнання, забезпеченістю сировину, основними та допоміжними матеріалами, мотивацією та іншими причинами.

Тривалість життя людини залежить від способу і стилю життя, генетики, екології та стану навколишнього середовища, системи охорони здоров'я тощо. Саме варіація і визначає необхідність статистики. Якби тривалість життя залишалася б однаковою для всіх людей, а робітники виробляли щодня одну і ту ж кількість виробів, то необхідність у статистиці відпала б. Очевидно, позбавлене сенсу статистичне вивчення ціни трамвайного квитка в Харкові, бо це явище слід віднести до класу таких, що «не варіюється».

Отже, поруч із обчисленням середньої величини під час аналізу статистичного розподілу постає питання про оцінку варіації значень ознаки в цьому розподілі.

Вивчення варіації поряд із застосуванням середніх та відносних показників має важливе наукове та практичне значення. «Показники варіації, — зазначав М. К. Дружинін, — разом із середніми величинами представляють той фундамент, на який спираються конструкції математико-статистичної методології». Так, дані про варіацію тарифних розрядів токарів заводу допомагають охарактеризувати особливості кваліфікації робітників. Зокрема, на обліку варіації ґрунтується статистичний аналіз зв'язків між явищами.

Варіація характеризує ступінь однорідності сукупності, стійкості значень з цієї ознакою. Порівняємо результати лабораторних випробувань на міцність 200 зразків пряжі, виготовленої двома фабриками (табл. 8.3). З таблиці видно, що фабрика 1 випускає більш однорідну продукцію, ніж фабрика 2. На фабриці 1 міцність пряжі знаходиться в межах від 160 до 259 г, на другий від 120 до 269 г. За допомогою показників варіації є можливість встановити ступінь впливу на цю ознаку інших варійованих ознак, встановити, наприклад, які фактори і в якій мірі впливають на продуктивність корів, на смертність населення і т. п.

Таблиця 8.3

**Розподіл контрольних партій пряжі по міцності**

| Міцність пряжі в грамах | Кількість зразків пряжі | |
|---|---|---|
| | 1 фабрика | 2 фабрика |
| 120–139 | 0 | 1 |
| 140–159 | 0 | 13 |
| 160–179 | 37 | 52 |
| 180–199 | 59 | 59 |
| 200–219 | 64 | 46 |
| 220–239 | 29 | 22 |
| 240–259 | 11 | 6 |
| 260–269 | 0 | 2 |
| Всього | 200 | 200 |

Вимір варіації необхідно при організації вибіркового спостереження, щодо взаємозв'язку між будь-якими ознаками та інших випадках.

Варіація властива всім явищам суспільства і природи. Між варіацією природних та суспільних явищ є суттєва відмінність. Для природних явищ є характерним є значно повільна зміна, ніж для супільних. Так, середня річна кількість опадів у межах якогось географічного району, хоч і відрізняється в різні роки, проте коливається у відносно вузьких межах навколо відносно постійного рівня.

Вимірювання коливань ознак має велике значення в економіці. За допомогою показників варіації вимірюють ступінь зв'язку та взаємозалежності між ознаками, оцінюють ступінь однорідності сукупності, типовості середньої, визначають величину можливої похибки вибіркового спостереження. Наведемо кілька прикладів.

1. Важливою умовою раціональної організації виробництва на підприємствах є досягнення рівномірного, ритмічного ходу виконання плану випуску продукції. Контроль за його здійсненням вимагає розрахунку показників, здатних вимірювати ступінь порушення ходу виконання плану — показників коливання.

2. Продукція, що випускається підприємствами, повинна відповідати вимогам та показникам якості, встановленим державними стандартами, технічними умовами та іншими нормативними документами. Дотримуватись стандартів — це означає не лише забезпечувати середній рівень якості, а й стежити за тим, щоб показники окремих одиниць продукції не виходили за межі встановлених меж. Контроль за виконанням цього завдання потребує розрахунку спеціальних показників якості продукції.

3. Для аналізу та оцінки динаміки продуктивності праці важливе значення має не тільки визначення середнього рівня продуктивності праці, а й оцінка ступеня коливання продуктивності праці окремих працівників. Чим вище чисельність передовиків, тим сильніша коливання норм виробітку. Підтягування основної маси до передовиків знижує коливання норм виробітку.

Таким чином, поряд з обчисленням середньої величини при аналізі статистичного розподілу важливу роль відіграє оцінка варіації (коливання) ознаки в цьому розподілі. У варіації виражається результат дії безлічі випадкових подій, характерних для масового суспільного явища. Отже, щоб розширити вивчення варіаційного ряду, дослідник повинен обчислити та проаналізувати показники, що характеризують варіацію ознаки.

## 8.2. Варіаційні ряди та їх характеристики

**Поняття про варіаційні ряди**. Першим кроком, який робить дослідник на шляху статистичного вирішення завдання, що стоїть перед ним, є побудова рядів розподілу. Ряди розподілу характеризують явища у певний період у статиці. *Ряд розподілу* представляє собою саму елементарну таблицю, що містить у підметі значення ознаки, яка коливається, а в присудку частоти варіантів або інтервалів. Головною метою побудови рядів розподілу є виявлення основних властивостей та закономірностей досліджуваної статистичної сукупності.

Окремі конкретні значення варіюючої ознаки представлені в цьому ряду, називаються *варіантами*. Умовимося кожне значення позначати $x_1$; $x_2$; $x_3$; …; $x_n$. Варіанти зображуються дискретним числом або інтервалом (від — до). Наприклад, кількість членів сім'ї: 2, 3, 4, 5 і т. д., або вік обладнання: до 3 років, 3–5, 5–10, 10–15 і понад 15 років тощо. У ряді розподілу деякі значення варіюючої ознаки можуть мати однакове значення (один і той самий варіант). Іноді замість варіант говорять варіанта.

Припустимо, що досліджувана ознака — якість деталей, а більш конкретно — вага деталей. Позначимо її через *x*. Нехай зміна ваги, наприклад, 50 деталей дала наступний невпорядкований ряд результатів окремих спостережень, розміщених за списком (*г*):

84, 86, 82, 83, 85, 83, 80, 85, 81, 82, 83, 83, 81, 83, 85, 80, 86, 86, 82, 83, 84, 84, 81, 82, 83, 84, 82, 80, 86, 83, 81, 86, 85, 83, 84, 82, 80, 84, 85, 82, 83, 84, 83, 85, 85, 82, 83, 84, 81, 82.

Щоб розібратися в отриманому матеріалі і зробити необхідні висновки, дані повинні бути упорядковані, систематизовані. Але яким чином систематизувати отримані статистичні дані? Першим кроком у справі впорядкування первинного ряду є його ранжування, тобто розташування окремих одиниць спостереження в порядку зростання (або зменшення).

Розташуємо значення ознаки первинного ряду у зростаючому порядку та отримаємо *ранжований ряд*:

80, 80, 80, 80, 81, 81, 81, 81, 81, 82, 82, 82, 82, 82, 82, 82, 82, 82, 83, 83, 83, 83, 83, 83, 83, 83, 83, 83, 83, 83, 84, 84, 84, 84, 84, 84, 84, 84, 85, 85, 85, 85, 85, 85, 85, 86, 86, 86, 86, 86.

Недостатньо ранжувати статистичні дані в порядку зростання або зменшення. Дослідник не може виявити закономірності, які закладені у розподілі даних. Близько якої величини групуються більшість показників варіаційного ряду, як відхиляються показники ряду динаміки від середньої величини, яка спрямованість розподілу одиниць сукупності? Крім того, у переважній більшості випадків доводиться досліджувати статистичні сукупності, що складаються з великої кількості одиниць, які дуже складно уявити у вигляді ранжованого ряду. На ці та інші питання він не може відповісти, ранжувавши показники варіаційного ряду за зростанням або зменшенням.

Безпосередньо за списком, а ще легше по ранжированному ряду можна підрахувати, як часто кожен варіант зустрічається в даній сукупності. В результаті отримаємо розподіл ознаки або *варіаційний ряд*. Найчастіше складання варіаційного ряду є першим кроком, який робить дослідник на шляху статистичного розв'язання завдання, що стоїть перед ним. У наведеному прикладі варіаційний ряд містить розподіл деталей за вагою. Інакше кажучи, варіаційний ряд характеризує варіацію значень даного ознаки в одиниць досліджуваної сукупності, у даному випадку — варіацію показників ваги деталей.

Таким чином, для подальшого узагальнення отриманих статистичних даних дослідник може об'єднати їх у групи, що збігаються або не дуже відрізняються один від одного, тобто групування даних. В результаті замість довгого ряду, що складається з сотень і тисяч окремих одиниць будь-якого масового явища, буде отримано порівняно короткий ряд даних, що складається з невеликої кількості груп. Сама техніка зведення такого масиву даних у групи не складна. Спочатку визначають число груп та інтервали, в яких повинні знаходитися значення досліджуваного ознаки в кожній групі. Інтервали у варіаційному ряду утворюються, як правило, рівні та закриті. Після того, як визначено кількість груп та межі інтервалів, підраховують загальну кількість одиниць, які володіють такими значеннями.

Але якщо рознесення показників об'єкта спостереження та підрахунок підсумків не становить великої складності, то визначення числа та розміру інтервалу є складнішим завданням. Інтервали мають бути вибрані так, щоб уникнути провалів, пов'язаних з інтервалами без даних, обмежити загальну розтягнутість групування.

Побудуємо варіаційний ряд для наведеного вище прикладу. Для цього знаходимо варіант із найменшим значенням ознаки. У нашому прикладі його величина складе 80 *г*, і підраховуємо його частоту. Так, варіант 80 *г* зустрічається 4 рази, варіант 81 *г* — 5 разів і т. д. Розташуємо отримані варіанти наступним чином (табл. 8.4).

Такий ряд називається варіаційним рядом; він характеризує зміну (варіювання) якої-небудь кількісного ознаки (у нашому прикладі варіювання ваги деталей). Отже, варіаційний ряд представляє собою

Таблиця 8.4

**Варіаційний ряд**

| Окремі значення ознаки — варіанти (вага однієї деталі в *г*) $x$ | $x_1$ 80 | $x_2$ 81 | $x_3$ 82 | $x_4$ 83 | $x_5$ 84 | $x_6$ 85 | $x_7$ 86 | Обсяг сукупності (всього деталей) |
|---|---|---|---|---|---|---|---|---|
| Частоти або чисельності (число деталей) $f$ | $f_1$ 4 | $f_2$ 5 | $f_3$ 9 | $f_4$ 12 | $f_5$ 8 | $f_6$ 7 | $f_7$ 5 | $f_n$ 50 |

групову таблицю, що містить два рядки (або колонки). В одній із них наводяться варіанти, в іншій — частоти.

У варіаційних рядах чисельність одиниць сукупності може бути як абсолютним числом, і відносним показником (табл. 8.5).

Таблиця 8.5

**Розподіл 50 деталей за вагою**

| Варіанти (в *г*) | Абсолютні частоти (число деталей) | Відносні частоти (в процентах) |
|---|---|---|
| 80 | 4 | 8 |
| 81 | 5 | 10 |
| 82 | 9 | 18 |
| 83 | 12 | 24 |
| 84 | 8 | 16 |
| 85 | 7 | 14 |
| 86 | 5 | 10 |
| Всього | 50 | 100 |

Сума всіх частот дорівнює, очевидно, обсягу сукупності. Відношення абсолютної частоти до обсягу сукупності називають відносною частотою (див. стовпець 3 табл. 8.5).

У мінливості ознаки спостерігається наступна закономірність. Один із варіантів ознаки, що знаходиться в центрі, зустрічається з найбільшою частотою. У міру віддалення від цього варіанта у бік більшої чи меншої ваги деталі частоти варіантів поступово зменшуються. Характер наростання чи зменшення частот висловлює загальну закономірність, властиву змінам ваги деталі.

За характером своїх варіантів ознаки діляться на *кількісні*, виражені числами, і *атрибутивні*, які числами не виражаються. Так, наприклад, вага, міцність, трудомісткість, енергоємність виготовлення деталі є кількісними ознаками, пристосованість до транспортування, раціональність форми та зручність експлуатації — атрибутивними.

**Види рядів розподілу**. Залежно від того, чи є ознака, взята за основу групування, атрибутивною або кількісною, ряди розподілу поділяються на два види: атрибутивні та варіаційні.

Ряди розподілу, побудовані за атрибутивними ознаками, називають *атрибутивними*. Прикладом атрибутивних рядів може служити розподіл населення за статтю, професіями, характером трудової діяльності, національністю і т. п. При групуванні за атрибутивною ознакою в ряду розподілу складають окремі групи, що вказуються в підметі, і чисельність, або питома вага кожної групи, що вказуються в присудку. Наприклад, наведений нижче ряд розподілу є розподілом за атрибутивною ознакою (табл. 8.6).

Таблиця 8.6

**Кількість працівників, задіяних у виконанні наукових досліджень і розробок, за категоріями персоналу у 2020 р.**

| | Осіб |
|---|---|
| Кількість працівників, задіяних у виконанні наукових досліджень і розробок — усього, осіб | 78 860 |
| У тому числі: | |
| дослідники | 51 427 |
| техніки | 7 117 |
| допоміжний персонал | 20 316 |

*Варіаційний ряд* — ряд, в якому дається розподіл одиниць сукупності, що вивчається, за значеннями якої-небудь ознаки. Іншими словами, варіаційний ряд характеризує варіацію значень даної ознаки в одиниць досліджуваної сукупності. Звідси і походить сама назва ряду — варіаційний ряд. Варіаційний ряд являє собою групову таблицю, що містить дві графи — значення ознаки, що варіює, і чисельність одиниць у кожній групі. Інтервали в варіаційному ряду утворюються, як правило, рівні і закриті.

За характером варіації розрізняють дискретні та безперервні варіаційні ряди.

*Дискретні варіаційні ряди* об'єднують варіанти, окремі значення ознаки (варіанти) яких відрізняються один від одного на деяку кінцеву величину (звичайне ціле число), тобто дані у вигляді безперервних чисел. В основу кожного рядка дискретного варіаційного ряду розподілу покладено один певний варіант, який може приймати тільки цілі значення. Наприклад, тарифний розряд, кількість дітей у сім'ї, кількість робітників на підприємстві, кількість верстатів, що обслуговуються одним робітником, кількість насіння в 1 кг тощо. Наведене вище групування робочих машинобудівного заводу за розрядами (перші дві графи табл. 8.7) може бути прикладом дискретного варіаційного ряду.

У практиці статистичних досліджень дискретні варіаційні ряди застосовуються не часто. Складання їх не становить великої складності, оскільки склад груп визначається конкретними варіантами, які мають досліджувані групувальні ознаки. Для побудови *дискретного ряду* з невеликим числом варіантів виписуються всі варіанти значень ознаки, що позначаються через $x_i$, а потім підраховується частота повторення кожного варіанту $f_i$. Наприклад, розподіл робітників токарного цеху за розрядами характеризується такими даними (табл. 8.7).

Таблиця 8.7

| Тарифний розряд робітника ($x_i$) | Число робітників, які мають цей розряд ($f_i$) | Частості ($w_i$) | Нагромаджені частоти ($S_i$) |
|---|---|---|---|
| 2 | 6 | 0,10 | 6 |
| 3 | 12 | 0,20 | 18 |
| 4 | 22 | 0,37 | 40 |
| 5 | 13 | 0,22 | 53 |
| 6 | 7 | 0,11 | 70 |
| Всього | 60 | 1,00 | |

Цей варіаційний ряд охоплює 60 робітників із різним тарифним розрядом. Число робочих може бути тільки цілим, дискретним числом, отже, і ряд розподілу — дискретним варіаційним рядом. Таким чином, первинний ряд, що характеризує розряд 60 робітників цеху, замінений коротким рядом, що складається з п'яти груп. Замість абсолютного числа робітників, які мають певний розряд (частоти), можна використовувати питому вагу робітників цього розряду.

*Безперервною* називається варіація, при якій значення ознаки можуть відрізнятися одне від одного на скільки завгодно малу величину і в певних межах приймати будь-які значення. В якості прикладу безперервної варіації можна навести розмір заробітної плати робітників, процент перевиконання робітниками норм виробітку, міцність зразків пряжі в грамах, відсоток шлюбу, вага одного насіння і т. д.

Для побудови ряду розподілу при безперервній варіації вказуються інтервали «від—до», оскільки значення ознаки окремих одиниць можуть взагалі не повторюватися.

*Інтервальні варіаційні ряди* поєднують варіанти ознак або дискретних ознак, що змінюються в широких межах. Тобто частоти відносяться вже не до окремого значення ознаки, як це буває при дискретній варіації, а до всього інтервалу. Інтервальні ряди будуються, якщо число варіант велике і розподіл за кожним окремим значенням варіант стає мало наочним, або якщо має безперервне варіювання. Інтервальні варіаційні ряди можуть бути з рівними та нерівними інтервалами.

*Варіаційні ряди з рівними інтервалами* — це ряди, в яких інтервали мають одну й ту ж саму величину. Як приклад варіаційного ряду з рівними інтервалами наведемо розподіл робочих цеху машинобудівного заводу за процентом виконання норм виробітку (табл. 8.8). *Варіаційні ряди з нерівними інтервалами* — це ряди, де інтервали мають різну величину. Наприклад, при групуванні міст за чисельністю населення краще використовувати нерівні інтервали, тому що для невеликих міст має значення і мала різниця в чисельності жителів, а для великих міст різниця в десятки та сотні тисяч не має суттєвого значення. До нерівних, послідовно збільшується інтервалів вдаються в тому випадку, коли необхідно виділити якісно відмінні типи явищ. Наведемо варіаційний ряд із нерівними інтервалами (табл. 8.9).

Таблиця 8.8

**Розподіл робітників за рівнем виконання норм виробітку**

| Групи робітників за виконанням норм виробітку, % (інтервали) | Число робітників (частоти) |
|---|---|
| 80–90 | 4 |
| 90–100 | 18 |
| 100–110 | 27 |
| 110–120 | 11 |
| 120–130 | 5 |
| Всього | 65 |

Таблиця 8.9

**Групування сільськогосподарських підприємств України за чисельністю корів у 2020 р.**

| Групи підприємств із чисельності корів (інтервали) | Число підприємств (частоти) |
|---|---|
| до 50 | 594 |
| 50–99 | 211 |
| 100–499 | 742 |
| 500–999 | 175 |
| більше 1000 | 69 |
| Всього | 1791 |

Крім того, зустрічаються іноді ряди з відкритими інтервалами, першими чи останніми.

В інтервальних варіаційних рядах частоти відносяться не до окремого значення ознаки, як це буває при дискретній варіації, а до всього інтервалу. Часто значення інтервалу приймають його середину, тобто центральне значення. Тому при побудові рядів розподілу інтервальних ознак необхідно встановити число груп (інтервалів), на які слід розбити всі одиниці досліджуваної сукупності.

Інтервальні варіаційні ряди у статистиці набули більш широкого поширення. При їх складанні виникає низка складних питань, у тому числі у виборі кількості груп та величини інтервалів. Принципи вирішення цих питань викладено у розділі угруповання статистичних даних (глава третя).

В інтервальних варіаційних рядах у кожному закритому інтервалі розрізняють нижню та верхню межі інтервалу:

нижня межа інтервалу $x_{\min}$;

верхня межа інтервалу $x_{\max}$.

Визначення величини інтервалу $k$ для побудови варіаційного ряду з рівними інтервалами проводиться у наступній послідовності:

1) обчислюють різницю між максимальним та мінімальним значеннями варіанта ознаки ряду розподілу (так званий розмах варіації, $R$):

$$R = x_{\max} - x_{\min}; \qquad (8.1)$$

2) розділити розмах варіації на кількість груп $K$, тобто $k = \frac{R}{K}$. Для вибору оптимальної величини інтервалу, тобто такої величини інтервалу, при якій варіаційний ряд не буде занадто громіздким, тому що в кожному інтервалі виявиться замало одиниць ознаки, і не надмірно малим, щоб не втратити характерні особливості розподілу ознак явищ можна рекомендувати формулу:

$$k \approx \frac{x_{\max} - x_{\min}}{1 + 3{,}2\lg n}, \qquad (8.2)$$

де $n$ — число одиниць у сукупності;

$x_{\max} - x_{\min}$ — розмах варіації.

Так, якщо в сукупності 100 одиниць найбільший варіант дорівнює 80, а найменший 0, то

$$k = \frac{80 - 0}{1 + 3{,}2 \cdot \lg 100} = \frac{80}{1 + 3{,}2 \cdot 2} = \frac{80}{7{,}4} \approx 11.$$

Отже, у даному випадку оптимальною величиною інтервалу може бути величина 11.

Для визначення числа груп можна скористатися й іншою формулою $K \approx \log_2 n + 1$, де $n$ — число одиниць сукупності, що вивчаються. Число груп можна приблизно визначити і за формулою

$K \approx 1 + 3{,}322 \lg n.$

Величина інтервалу в значній мірі залежить від необхідної точності спостереження. Якщо вихідні дані представлені цілими числами, то розрахована величина інтервалу округляється до найближчого цілого числа, якщо дані представлені з точністю до 0,1, то величини інтервалу округляється до цілих з десятими і т. д.

Визначення величини інтервалу дозволяє побудувати ряд розподілу ознаки у вигляді таблиці. Нижню межу першого інтервалу зазвичай приймають рівною мінімальному значенню ознаки (іноді береться число, найближче до мінімального значення). У кожен інтервал включають варіанти, числові значення які більші за мінімальне значення і менше або дорівнює верхній межі інтервалу. Наведемо такий приклад. Маються наступні дані про розмір партії валиків у кількості 20 шт., виготовлених бригадою робітників механічного цеху заводу протягом зміни (діаметр, *мм*): 49,945; 49,946; 49,961; 49,957; 49,950; 49,946; 49,934; 49,945; 49,944; 49,951; 49,942; 49,941; 49,946; 49,921; 49,936; 49,939; 49,943; 49,949; 49,934; 94,927. Найменше та найбільше із значень, що зустрічаються, — 49,921 і 49,961. Цими двома числами визначається варіація ознаки. Визначимо кількість груп інтервального варіаційного ряду

$$K \approx 1 + 3{,}322 \cdot \lg 20 = 1 + 3{,}322 \cdot 1{,}301 = 5{,}32.$$

Округляючи, отримаємо число груп, що дорівнює 5. Розмір інтервалу для даного ряду розподілу складе 0,008 *мм* $\left(\frac{49{,}961 - 49{,}921}{5}\right)$. В результаті отримаємо ряд розподілу валиків за величиною діаметра (див. графи 1, 2 табл. 8.10). Значення окремих варіантів співпадали з межами інтервалів (наприклад, значення 49,945). Так як максимальне значення ознаки збігається з верхньою межею інтервалу, то ці значення слід включати в інтервал, верхня межа якого збігається із зазначеним значенням. Наприклад, значення ознаки 49,945 включається до третього інтервалу.

Таблиця 8.10

| Діаметр (в *мм*) $x_i$ | Число деталей $f_i$ | Нагромаджені частоти $S_i$ |
|---|---|---|
| 1 | 2 | 3 |
| 49,921–49,929 | 2 | 2 |
| 49,929–49,937 | 3 | 5 |
| 49,937–49,945 | 7 | 12 |
| 49,945–49,953 | 6 | 18 |
| 49,953–49,961 | 2 | 20 |
| Всього | 20 | |

**Частота**. Елементи варіаційного ряду групуються за варіантами ознак, і кожної класифікаційної групи підраховується частота. *Частота* (*чисельність*) — абсолютне число, що показує, скільки разів (як часто) зустрічається в сукупності те чи інше значення ознаки або, що те ж саме, скільки одиниць у сукупності володіють тим або іншим значенням ознаки. Частота позначається $f_1$; $f_2$; $f_3$; …; $f_n$. Наприклад, якщо в сукупності робітників 4 особи одержують заробітну плату в місяць 10 тис. грн, то число 4 є частотою варіанта 10 тис. грн. Іноді частоти називають статистичними вагами. За допомогою частот розраховуються різні узагальнюючі показники масового явища — відносні, середні та інші показники (середня тривалість життя, середня заробітна плата, середня вага деталі, питома вага браку тощо). Ці показники виражають певні пропорції і співвідношення, які складаються в мінливості ознак, що варіюються, і характеризують досліджувану сукупність.

**Частість**. Часто замість абсолютних значень частот використовують відносні величини. Частоти, представлені у відносному вираженні, називаються частостями. Частість ($w_i$) — відносна величина, що визначає частоти окремих варіантів або інтервалів ($f_i$) у загальній сумі всіх частот ($\sum f_i$):

$$w_i = \frac{f_i}{\sum f_i}. \tag{8.3}$$

Ми маємо частоти:

$$f_1;\ f_2;\ f_3;\ \dots;\ f_n.$$

Для отримання суми всіх частот їх потрібно скласти:

$$f_1 + f_2 + f_3 + \dots + f_n.$$

У математиці застосовується знак $\sum$ (грецька буква сигма), що означає підсумовування.

Отже, можна записати:

$$f_1 + f_2 + f_3 + \dots + f_n = \sum_{i=1}^{n} f_i. \tag{8.4}$$

де над знаком ($n$) $\sum i$, під $\sum$ $(i=1)$ показують, що сумуванню підлягають всі частоти ($f_i$) за умови, що $i$ приймає всі цілі значення від 1 до $n$.

Надалі в подібних випадках, тобто при підсумовуванні по підрядковому значку $i$ від $i = 1$ до $i = n$, ми будемо цей запис або її частини опускати, якщо зміст формули і без того зрозумілий.

Таким чином, частість першого варіанта або інтервалу ($w_1$)

$$w_1 = \frac{f_1}{\sum f_i}, \tag{8.5}$$

частість другого варіанта або інтервалу ($w_2$)

$$w_2 = \frac{f_2}{\sum f_i}. \tag{8.6}$$

Обчислимо частості, використовуючи дані табл. 8.7:

$$w_1 = \frac{6}{60} = 0{,}10; \quad w_2 = \frac{12}{60} = 0{,}20; \quad w_3 = \frac{22}{60} = 0{,}37;$$

$$w_4 = \frac{13}{60} = 0{,}22; \quad w_5 = \frac{7}{60} = 0{,}11.$$

Результати розрахунку частостей наведено у гр. 3 табл. 8.7.

Сума усіх частостей рівна одиниці:

$$\sum w_i = 1.$$

В нашому прикладі: 0,10+0,20+0,37+0,22+0,11=1,00.

Частості можуть бути виражені і у процентах. В цьому випадку сума частостей дорівнює 100 %. Заміна частот частостями дозволяє зіставити варіаційні ряди з різним числом спостережень і з більшою чіткістю виразити типові риси розподілу. Частості в статистичному дослідженні використовуються також для визначення ймовірності події, яка вимагає для свого точного математичного визначення повної вказівки всіх можливих випадків.

**Нагромаджені частоти (або частості)**. Нагромаджені частоти (або частості) — частоти варіантів від нижньої межі до заданого значення. Їх обчислюють шляхом підсумовування частот всіх варіантів, не більших за цей варіант. Нагромаджені частоти можна отримати у висхідному та низхідному порядках. Накопичення (підсумовування) часто або часто називають також кумуляцією. Кумулювати можна частоти і варіантів, і інтервалів.

Обчислимо за даними прикладу (графа 1 і 2 табл. 8.11) нагромаджені частості у висхідному та в низхідному порядку.

**Щільність розподілу**. Для правильного уявлення про характер розподілу застосовують щільність розподілу, яку обчислюють як відношення частот або частостей до величини інтервалів. Розрізняють

Таблиця 8.11

| Групи населення за віком (інтервали) | Чисельність постійного населення України у відсотках (частості) | Накопичені частості, отримані підсумовуванням | |
|---|---|---|---|
| | | висхідному порядку | низхідному порядку |
| Від 0 до 4 | 4,1 | 4,1 | 100 |
| 5–9 | 5,5 | 9,6 | 95,9 |
| 10–14 | 5,5 | 15,1 | 90,4 |
| 15–19 | 4,6 | 19,7 | 84,9 |
| 20–24 | 4,8 | 24,5 | 80,3 |
| 25–29 | 6,2 | 30,7 | 75,5 |
| 30–34 | 8,1 | 38,8 | 69,3 |
| 35–39 | 8,4 | 47,2 | 61,2 |
| 40–44 | 7,5 | 54,7 | 52,8 |
| 45–49 | 7,2 | 61,9 | 45,3 |
| 50–54 | 6,6 | 68,5 | 38,1 |
| 55–59 | 7,0 | 75,5 | 31,5 |
| 60 і старше | 24,5 | 100 | 24,5 |
| Всього | 100 | – | – |

абсолютну та відносну щільність розподілу.

*Абсолютна щільність* розподілу:

$$m_i^{\text{а}} = \frac{f_i}{k}; \tag{8.7}$$

та *відносна щільність* розподілу:

$$m_i^{\text{о}} = \frac{w_i}{k}, \tag{8.8}$$

де $m_i^{\text{а}}$ и $m_i^{\text{о}}$ — щільності розподілу, абсолютна (зі значком а) та відносна (зі значком о).

Ці показники використовуються для створення нових груп на основі раніше проведеного групування, що буває необхідним при порівняльній оцінці даних, зібраних за різними сукупностями та порізному оброблених. Наприклад, по двох підприємствах відомий розподіл робітників за розміром заробітної плати (табл. 8.12).

Наведені дані не дозволяють порівняти розподіл робітників двох підприємств за розміром заробітної плати, тому що в цих підприємствах є різна кількість груп робітників. Для перегрупування даних з метою отримання порівнянних груп та їх аналізу можна скористатися простим укрупненням інтервалів. У цьому випадку для двох фабрик

Таблиця 8.12

| Фабрика 1 | | Фабрика 2 | |
|---|---|---|---|
| Групи робітників за розміром заробітної плати, грн | Кількість робітників, у % до підсумку | Групи робітників за розміром заробітної плати, грн | Кількість робітників, у % до підсумку |
| До 8 000 | 2 | До 9 000 | 7 |
| 8 000–9 000 | 4 | 9 000–11 000 | 41 |
| 9 000–10 000 | 48 | 11 000–14 000 | 19 |
| 10 000–11 000 | 30 | 14 000–17 000 | 16 |
| 11 000–13 000 | 9 | Понад 17 000 | 17 |
| 13 000–14 000 | 5 | | |
| 14 000–15 000 | 2 | | |
| Всього | 100 | Всього | 100 |

буде використано однакові варіанти (табл. 8.13).

Можна скористатися й іншим групуванням за розміром заробітної плати (грн): 1 група — до 9 000; 2 група — 9 000–10 000; 3 група —10 000–11 000; 4 група — 11 000–13 000; 5 група — 13 000–15 000. Для проведення перегрупування необхідно розщеплення інтервалів ряду розподіл заводу 2.

Таблиця 8.13

| Групи робітників за розміром заробітної плати, грн | Кількість робітників, у % до підсумку | |
|---|---|---|
| | завод 1 | завод 2 |
| До 9 000 | 6 | 7 |
| 9 000–11 000 | 78 | 41 |
| 11 000–14 000 | 14 | 19 |
| 14 000 і вище | 2 | 33 |
| Всього | 100 | 100 |

Якщо відома відносна щільність розподілу, то частості відповідного інтервалу можна визначити множенням щільності на величину інтервалу, тобто $w_i = m_i^{\text{o}} \cdot k$.

За даними табл. 8.12 визначаємо відносну щільність розподілу робітників за розміром заробітної плати для інтервалів:

другого 9 000–11 000 $m_i^{\text{o}} = 0{,}02050\left(\frac{41}{2000}\right)$;

третього 11 000–14 000 $m_i^{\text{o}} = 0{,}00633\left(\frac{19}{3000}\right)$;

четвертого 14 000–17 000 $m_i^{\text{o}} = 0{,}00533\left(\frac{16}{3000}\right)$.

Тоді кількість робітників (в % до підсумку) заводу 2, розмір заробітної плати яких становить 11000–13000 грн., визначається так:

$$0{,}00633 \cdot 2000 = 12{,}7\ \%;$$

для групи робітників із розміром заробітної плати 13 000–15 000 грн:

$$0{,}00633 \cdot 1000 + 0{,}00533 \cdot 1000 = 11{,}7\ \%.$$

Таблиця 8.14

| Групи робітників за розміром заробітної плати, грн | Кількість робітників, у % до підсумку | |
|---|---|---|
| | завод 1 | завод 2 |
| До 9 000 | 6 | 7 |
| 9 000–10 000 | 48 | 20,5 |
| 10 000–11 000 | 30 | 20,5 |
| 11 000–13 000 | 9 | 12,7 |
| 13 000–15 000 | 7 | 11,7 |
| 15 000 і вище | – | 27,6 |
| Всього | 100 | 100 |

Результати перегрупування представлені у табл. 8.14. Використовуючи відносні щільності, можна здійснити перегрупування варіаційних рядів розподілу. Значення щільності розподілу полягає в тому, що на їх основі можна дати характеристику коливання групувальної ознаки.

## 8.3. Графічне зображення рядів розподілу

Поряд з табличною формою вираження рядів розподілу в математичній статистиці широко використовують графічні зображення. Графічне зображення рядів розподілу полегшує їх аналіз і дозволяє судити про характер розподілу одиниць об'єкта спостереження. Зображення рядів розподілу може бути представлене у вигляді полігону, гістограми, кумуляти, кривої концентрації (Лоренца), огива та ін. У цьому параграфі ми розглядаємо графічні способи зображення рядів розподілу.

**Полігон розподілу**. Дискретний варіаційний ряд зображується як так званого полігону або багатокутника розподілу частот. *Полігон розподілу* (дослівно — багатокутник розподілу частот) будується в прямокутній системі координат, в якій величини ознаки $x$ відкладаються на осі абсцис, відповідні їм частоти $m$ або частості $w$ (точніше щільність розподілу) — по осі ординат. На осі абсцис відзначаються точки, відповідні величині варіантів, їх відновлюються ординати (перпендикулятори), довжина яких відповідає чисельності цих варіантів. Вершини перпендикулярів у послідовному порядку з'єднуємо відрізками прямої. Для замикання полігону крайні вершини з'єднуються з точками на осі абсцис, віддаленими один поділ у прийнятому масштабі від $x_{max}$ і $x_{min}$. Замінивши координати точок (на $x_i; p_i$), аналогічно можна побудувати полігон розподілу ймовірностей для випадкової дискретної величини, де $x_i$ — можливе значення дискретної випадкової величини; $p_i$ — відповідні ймовірності їх появи.

Побудуємо полігон частот для розподілу робочих машинобудівного заводу за розрядами (табл. 8.15). На осі абсцис відкладаємо шість рівних відрізків (за кількістю розрядів), на осі ординат — чисельність робітників. З середини кожного відрізка відновлені перпендикуляри, висота яких відповідає числу робітників (за прийнятим масштабом) у цьому інтервалі. З точок на осі абсцис, що відповідає розрядам, відновлюємо перпендикуляри заввишки, пропорційної числам робітникам. Вершини ординат з'єднуються прямими лініями. Для замикання крайні точки отриманої ламаної лінії з'єднуються з серединами попереднього (перед начальним) та наступного (за останнім) інтервалів, у яких частоти або частості дорівнюють нулю. Замкнена з віссю абсцис ламана лінія представляє полігон розподілу частот (рис. 8.1).

Таблиця 8.15

| Розряд ($x$) | Число робітників ($f$) |
|---|---|
| 1 | 10 |
| 2 | 20 |
| 3 | 84 |
| 4 | 190 |
| 5 | 136 |
| 6 | 60 |
| Всього | 500 |

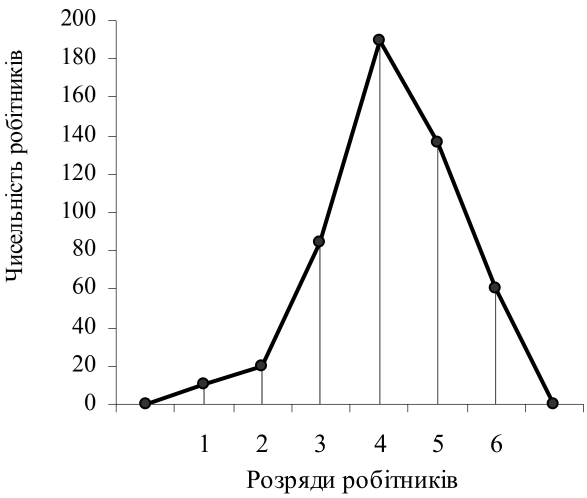



*Рис. 8.1.* Полігон частот розподілу робітників машинобудівного заводу за розрядами.

Найчастіше полігони використовуються для зображення дискретного варіаційного ряду. Сума довжини всіх ординат полігону дорівнює сумі частот варіаційного ряду $n$ або $\sum m$. Але полігони можуть бути використані і для зображення інтервальних рядів. У цьому випадку ординати, пропорційні частоті або частості інтервалу, відновлюють перпендикулярно осі абсцис у точці, що відповідає середині цього інтервалу. Крайні ординати з'єднують із серединою інтервалів, у яких частоти або частості дорівнюють нулю.

**Гістограма**. Інтервальний варіаційний ряд як правило зображується у вигляді особливих ступінчастих графіків — гістограм. *Гістограма* — графічне зображення інтервального варіаційного ряду у вигляді прямокутників рівної висоти, основи яких — відрізки осі абсцис, що відповідають інтервалам зміни ознаки. Гістограма будується аналогічно до полігону в прямокутній системі координат. Але на відміну від полігону при побудові гістограми на осі абсцис беруться не крапки, а відрізки, що у прийнятому масштабі відповідають величині інтервалів варіаційного ряду. Замість ординат, що відповідають частотам або частостям окремих варіантів, будують прямокутники з висотою, пропорційною частотам або частостям інтервалу, а при нерівності — щільності розподілу (абсолютним або відносним). При цьому щільність кожного «ступеня» гістограми відповідає частоті «частості» інтервалу, а загальна площа гістограми дорівнює чисельності сукупності, якщо гістограма побудована за абсолютною щільністю, або одиниці (або 100%), якщо гістограма побудована за відносною щільністю.

Мається варіаційний ряд розподілу робітників цеху машинобудівного заводу за процентом виконання норм виробітку (табл. 8.16). По отриманому варіаційному ряду побудуємо гістограму розподілу. Будуємо систему координат. На осі абсцис відкладаємо процент виконання норм виробітку робітників, починаючи з 80 до 130, на осі ординат проставляємо число робітників (в %). Частоти зображуються у вигляді прямокутників, побудованих на кожному інтервалі (див. рис. 8.2). Для цього на осі ординат відкладаємо частоти; так, щоб

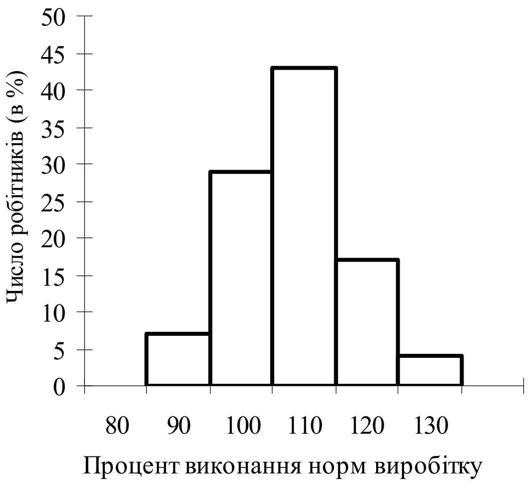



*Рис. 8.2*. Гістограма розподілу робітників цеху машинобудівного заводу за процентом виконання норм виробітку.

Таблиця 8.16

**Розподіл робітників за процентом виконання норм виробітку**

| Групи робітників за процентом виконання норм виробітку | Число робітників (в %) |
|---|---|
| 80–90 | 7 |
| 90–100 | 29 |
| 100–110 | 43 |
| 110–120 | 17 |
| 120–130 | 4 |
| Всього | 100 |

основи прямокутників були рівні між собою, а висота була пропорційна числу робітників у кожному інтервалі. Отримана фігура називається гістограмою розподілу.

Зображення ряду розподілу у вигляді прямокутників ґрунтується на припущенні, що щільність частоти залишається постійною всередині кожного інтервалу і змінюється стрибкоподібно на краях інтервалу. Це умовне припущення певною мірою спотворює закономірність ряду розподілу. Чим більша довжина інтервалу, тим сильніше відбувається спотворення ряду розподілу.

Більш близьким до дійсності є припущення, що в межах інтервалів частоти розташовані рівномірно, а тому можуть бути віднесені до конкретного значення, яке знаходиться в центрі інтервалу. Значить, середина — центр інтервалу як би «навантажується» всіма частотами. Такий спосіб називається способом «навантажених ординат». Це припущення призводить до зображення інтервальних рядів як полігонів розподілу, які будуємо наступним чином (рис. 8.3).


*Рис. 8.3.* Полігон розподілу для інтервального ряду.

З середини кожного інтервалу встановлюємо ординату, пропорційно до щільності частотам. Вершини перпендикулярів з'єднуємо прямими лініями. Дві крайні точки прямокутників замикаються по осі абсцис на середини інтервалів, в яких частоти (частості) рівні нулю. Полігони, що зображують інтервальні ряди розподілу, називають полігонами з навантаженими ординатами.

У випадку, коли гістограма будується для варіаційного ряду з нерівними інтервалами, то по осі ординат слід наносити показники щільності інтервалів (абсолютні або відносні). Висоти прямокутників гістограми будуть зображати величини щільності ряду розподілу.

Якщо збільшується інтервал числа спостережень із однієї і тієї ж сукупності, то це призводить до збільшення груп інтервального ряду. А це у свою чергу призводить до зменшення величини інтервалу. При цьому число сторін полігону буде рости, і ломана крива буде перетворюватися в плавну криву, яка отримала назву крива розподілу.

Крива розподілу характеризує в узагальненому вигляді закономірності розподілу частот всередині досліджуваної сукупності.

**Кумулятивна крива**. У ряді випадків для зображення варіаційних рядів використовується крива кумулятивна (кумулята). *Кумулятивна крива* (крива сум — кумулята) — графічне зображення варіаційного ряду, складене за послідовно підсумовуваними, тобто накопиченими, частотами або частостями. Для її побудови треба розрахувати накопичені частоти або частості. Нагромаджені частоти показують, скільки одиниць сукупності мають значення ознаки не більше, ніж значення, що розглядається, і визначаються послідовним підсумовуванням частот інтервалів. При побудові кумуляти дискретної ознаки на вісь абсцис наносяться значення ознаки (варіанти). По осі ординат відкладають накопичені підсумки частот або частостей, що показують, скільки одиниць сукупності мають значення ознаки, що не перевищують дане значення. З’єднанням вершин ординат прямими лініями отримуємо ламану лінію — кумуляту.

При побудові кумуляти інтервального ряду розподілу нижньої межі першого інтервалу відповідає частота, що дорівнює нулю, а верхній межі — вся частота цього інтервалу. Так, верхньою межею другого інтервалу відповідає частота, що дорівнює сумі частот перших двох інтервалів, і т. д. Верхньою межею останнього інтервалу відповідає накопичена частота, що дорівнює сумі всіх частот, тобто чисельності сукупності.

Побудуємо кумулятивну криву за наведеними даними (див. табл. 8.17) про розподіл робітників за розміром місячної заробітної плати. Накопичені частоти обчислені у графі 3 табл. 8.17 (див. рис. 8.4).

Таблиця 8.17

**Розподіл робітників розміром місячної заробітної плати**

| Групи робітників за розміром заробітної плати, тис. грн $x_i$ | Число робітників $f_i$ | Нагромаджені частоти $S_i$ |
|---|---|---|
| 7,4–8,2 | 7 | 7 |
| 8,2–9,0 | 8 | 15 |
| 9,0–9,8 | 12 | 27 |
| 9,8–10,6 | 8 | 35 |
| 10,6–11,4 | 5 | 40 |
| Всього | 40 | |

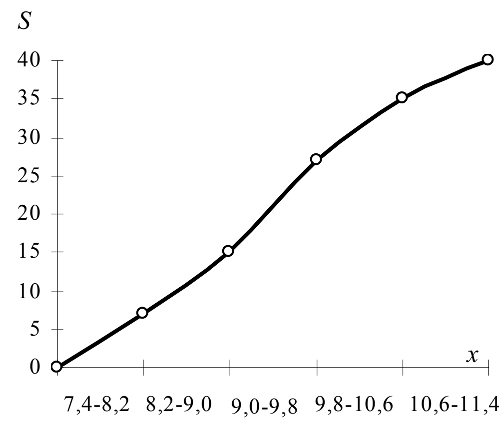



*Рис. 8.4*. Кумулята інтервального варіаційного ряду: $S$ — нагромаджені частоти, $x$ — середня денна заробітна плата, тис. грн.

Зображення варіаційного ряду у вигляді кумуляти є особливо зручним при порівнянні варіаційних рядів, а також в економічних дослідженнях, зокрема для аналізу концентрації виробництва.

**Огіва**. *Огіва* представляє собою графічне зображення рядів розподілів за накопиченими частотами. Огіва будується аналогічно кумуляті з тією лише різницею, що з осі абсцис відкладають нагромаджені частоти (або частості), а, по осі ординат — значення ознаки по зростаючій величині. З'єднуючи точки, отримують огіву розподілу. Якщо аркуш паперу, на якому зображено кумулята, повернути на 90° і подивитися на нього зі зворотного боку на світло, то можна побачити огіву.

У табл. 8.18 наведено дані про чисельність мігруючого населення країни.

Таблиця 8.18

| Групи населення за віком (інтервали) | Чисельність населення у процентах (частості) | Нагромаджені частості, отримані підсумовуванням |
|---|---|---|
| До 5 | 4,6 | 4,6 |
| 5–10 | 2,4 | 7,0 |
| 10–15 | 2,4 | 9,4 |
| 15–20 | 17,7 | 27,1 |
| 20–25 | 24,0 | 51,1 |
| 25–30 | 19,5 | 70,6 |
| 30–35 | 8,1 | 78,7 |
| 35–40 | 5,7 | 84,4 |
| 40–45 | 4,6 | 89,0 |
| 45–50 | 3,4 | 92,4 |
| 50–55 | 2,3 | 94,7 |
| 55–60 | 5,3 | 100,0 |
| Всього | 100,0 | – |

У стовпці 3 обчислені накопичені частості у зростаючому порядку. Так, для другої групи населення (5–10 років) нагромаджені частості становитимуть 7,0 % (4,6 + 2,4), для третьої групи (10–15 років) — відповідно 9,4 % (7,0 + 2,4) тощо.

За даними табл. 8.18, використовуючи накопичені частоти, у висхідному порядку побудуємо огіву (рис. 8.5)

**Крива концентрації (крива Лоренца)**. *Крива концентрації* (*крива Лоренца*) — кумулятивне зростання величин двох взаємозалежних ознак (у процентах до підсумку), нанесене на один графік, пока-


*Рис. 8.5*. Огіва інтервального варіаційного ряду.

зує ступінь концентрації окремих елементів сукупності по групам. Так, наприклад, знаючи суму обсягу випуску продукції окремих груп підприємств, доходу окремих груп сімей, за допомогою кривої Лоренца можна графічно зобразити ступінь концентрації значень ознаки. Для її побудови на осі абсцис відкладатиме нагромаджені частості, а на осі ординат — відповідні суми значень ознаки (у одиниць сукупності зі значенням ознаки до даного).

Наприклад, вивчення концентрації доходів населення координатами кривої Лоренца служать чисельність населення та загальний обсяг доходів, отриманий цим населенням. При рівномірному розподілі доходів крива Лоренца має вигляд прямої, що починається на початку координат. Так, якщо взяти в якості варійованої ознаки розмір доходів населення, то при рівномірному розподілі всієї суми доходу між членами суспільства попарні нагромаджені частоти і відповідно суми доходу будуть рівні (10 % населення мають 10 % доходу, 20 % населення — 20 % доходу і т. д.). При нерівномірному розподілі доходу лінія буде опуклою чи увігнутою. Відхилення кривої від лінії рівномірного розподілу характеризуватиме різний ступінь концентрації.

Побудова кривої концентрації здійснюється у наступній послідовності: 1) побудова варіаційних інтервальних рядів; 2) подальшій обробці даних, обчисленні добутків величин варійованої ознаки в кожній групі ($x$) на частоту ($m$) з наступним знаходженням частостей (у процентах до підсумку) як по колонці $m$, так і по колонці $xm$; 3) обчислюють відносні числа (у процентах до підсумку), і потім і нагромаджені (кумулятивні) відносні числа (те у процентах до підсумку).

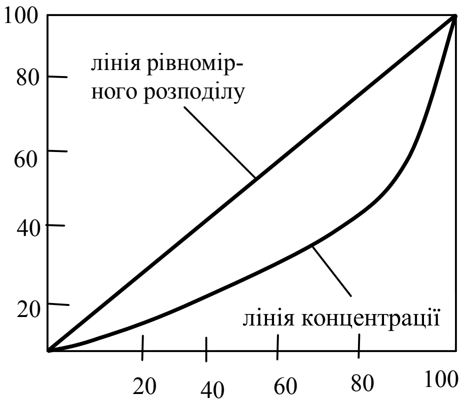



*Рис. 8.6.* Крива Лоренца.

Наносячи їх на графік, зображуємо криву концентрації (криву Лоренца). Відхилення від лінії рівномірного розподілу характеризуватиме різний ступінь концентрації. Крива Лоренца наведена на рис. 8.6. За допомогою кривої Лоренца можна дослідити ступінь концентрації виробництва продукції, числа робітників, доходів населення, земельних ділянок у власників тощо.

За наступними умовними даними (див. колонка А, 1 і 2 табл. 8.19) побудуємо криву Лоренца (див. рис. 8.7).

Таблиця 8.19

**Рівень концентрації виробництва**

| Групи підприємств за кількістю робітників (інтервали) $x$ | Номер групи $i$ | Абсолютна чисельність підприємств (у тис.) (частоти) $m$ | Абсолютна чисельність робітників (у тис.) $x_i m_i$ | Відносне число (у процентах до підсумку) | | Кумулятивні відносні числа (у відсотках до підсумку) | |
|---|---|---|---|---|---|---|---|
| | | | | підприємств $w_i = \frac{m_i}{\sum m_i} \cdot 100$ | робітників $R_i = \frac{x_i m_i}{\sum x_i m_i} \times 100$ | підприємств $\sum_1^i w_i$ | робітників $\sum_1^i R_i$ |
| А | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| До 5 | 1 | 85,3 | 212,3 | 45,8 | 2,6 | 45,8 | 2,6 |
| 5–20 | 2 | 49,4 | 567,1 | 26,4 | 6,9 | 72,2 | 9,5 |
| 2–50 | 3 | 24,1 | 824,7 | 12,9 | 10,1 | 85,1 | 19,6 |
| 51–100 | 4 | 11,8 | 836,5 | 6,3 | 10,2 | 91,4 | 29,8 |
| 101–250 | 5 | 10,5 | 1 542,2 | 5,6 | 18,9 | 97,0 | 48,7 |
| 251–500 | 6 | 3,6 | 1 324,9 | 1,9 | 16,2 | 98,9 | 64,9 |
| 500–1000 | 7 | 1,4 | 1 126,3 | 0,7 | 13,8 | 99,6 | 78,7 |
| 1001–2500 | 8 | 0,5 | 915,4 | 0,3 | 11,2 | 99,9 | 89,9 |
| 2501 і більше | 9 | 0,2 | 824,3 | 0,1 | 10,1 | 100,0 | 100,0 |
| Всього | – | 186,8 | 8 173,7 | 100,0 | 100,0 | – | – |

На рис. 8.7. виразно видно крайня нерівномірність концентрації робітників на підприємствах. Чим більший розрив між лінією абсолютної рівності доходів і кривою Лоренца, тим вищий ступінь нерівності у розподілі доходів в суспільстві.

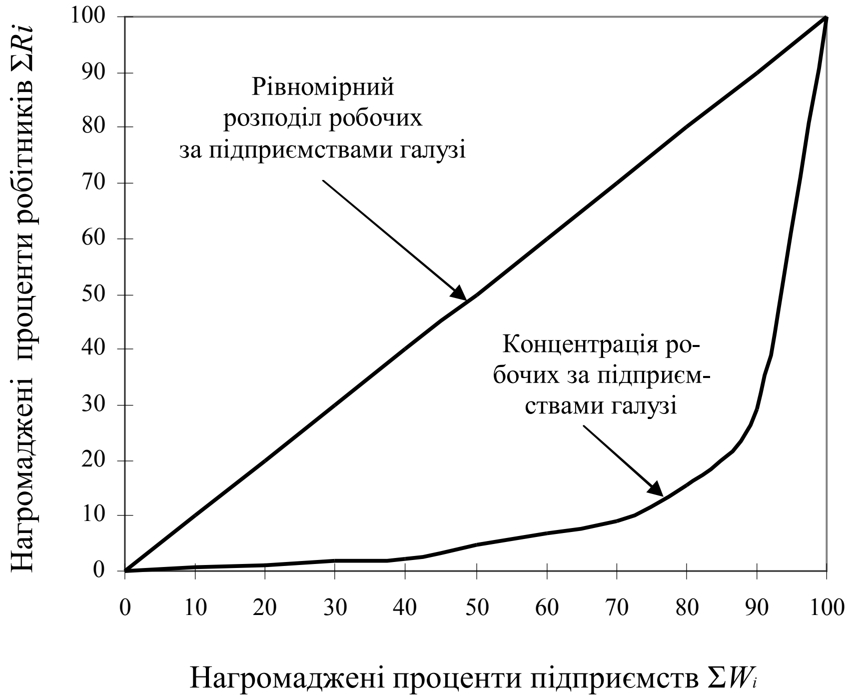



*Рис. 8.7.* Концентрація виробництва.

## 8.4. Основні характеристики варіаційного ряду

Для характеристики розподілу варіантів ознак недостатньо графічного зображення рядів динаміки. Для виявлення закономірностей і властивостей рядів розподілу, що вивчаються, обчислюють узагальнюючі показники варіаційного ряду. Система таких показників може бути представлена при порівнянні особливостей декількох розподілів. Припустимо, по п'яти виробничим дільницям відомі дані про розподіл 100 робітників за кваліфікацією (табл. 8.20).

Розподіл робітників першої і другої ділянки мають однаковий розмах варіації і характер розподілу частот, але відрізняються величиною ознаки, що варіюється, яка є центром розподілу. Таким чином, одну із груп узагальнюючих показників становлять *характеристики центру групування*. В якості таких показників використовують середню арифметичну, медіану і моду.

Розподіл робітників другої та третьої ділянок мають один і той же центр групування та симетричне розташування частот навколо нього, але відрізняються межами варіації.

Отже, крім показників центру групування, для цілей аналізу та порівняльної характеристики різних рядів розподілу необхідно мати уявлення про ступінь коливання ознаки або мати *показники ступеня* *варіації*. Розподіл робітників третьої та четвертої ділянок мають один і той же центр групування, однакові межі варіювання, симетричний характер розташування частот, але різний ступінь витягнутості вздовж осі ординат. Ця вигнутість характеризується *показниками ексцесу*. І нарешті, розподіл робочих четвертої і п'ятої ділянок показує, що вони відрізняються характером розподілу частот щодо центру групування. У першому випадку частоти варіантів розподілені на однаковій відстані від центру групування. Такий розподіл має симетричний характер. У другому випадку розташування частот навколо центру неоднаково, тобто частоти з обох боків від центру мають неоднаковий характер зміни. Тут спостерігається несиметричність (скошеність) ряду розподіл. Ступінь відхилення розподілу частот від симетричної форми характеризується *показниками асиметрії*. Показники ексцесу та асиметрії характеризують *форму розподілу*.

Таблиця 8.20

| Розряд робітників | Число робітників ділянки | | | | |
|---|---|---|---|---|---|
| | першої | другої | третьої | четвертої | п'ятої |
| II | 15 | – | 10 | 2 | 2 |
| III | 70 | 15 | 20 | 10 | 10 |
| V | 15 | 70 | 40 | 76 | 60 |
| V | – | 15 | 20 | 10 | 20 |
| VI | – | – | 10 | 2 | 8 |
| Всього | 100 | 100 | 100 | 100 | 100 |

Таким чином, для цілей аналізу та порівняльної характеристики різних рядів розподілу виділяють три групи узагальнюючих показників: 1) показники центру розподілу (центру групування); 2) показники ступеня варіації; 3) показники форми розподілу.

## 8.5. Показники центру розподілу

Для характеристики центру розподілу значень ознаки у варіаційному ряду використовуються середня арифметична, мода і медіана. Біля цих величин, як правило, концентрується основна маса спостережень. Середні величини характеризують положення графіка розподілу по осі абсцис. Такі величини зазвичай розташовані в центральній частині графіка і тому вони називаються середніми точками. Загальні поняття про середню та порядок її обчислення дано в гл. 7. У цьому параграфі розглянемо розрахунок показників центру розподілу для варіаційних рядів.

Найбільш відомою серед середніх величин є середня арифметична, яка розраховується за формулою:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\sum x}{n}. \tag{8.9}$$

У варіаційному ряду, де окремі варіанти ознаки повторюються кілька разів, частота може бути врахована при обчисленні середньої арифметичної, і розрахунок проводиться за формулою:

$$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i}, \tag{8.10}$$

де $x_i$ — варіанти значень ознаки;

$f_i$ — частота повторення цього варіанта.

Формула (8.9) дає просту середню арифметичну, формула (8.10) — зважену середню арифметичну.

Числа $f_i$ у формулі (8.10) називають вагами. Чим більша вага, з якою даний варіант входить у варіаційний ряд, тим сильніше вплив цього варіанта на величину середньої арифметичної.

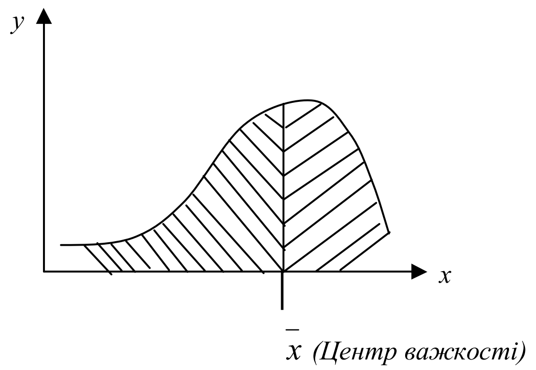


*Рис. 8.8.* Середня арифметична.

В інтервальному варіаційному ряду середня арифметична визначається за зваженою формулою $\bar{x} = \frac{\sum x_i' \cdot f_i}{\sum f_i}$, де $x_i'$ — середина відповідного інтервалу.

Приклад розрахунку середнього заробітку робітників заводу для такого ряду (табл. 8.21).

$$\bar{x} = \frac{760\,000}{100} = 7\,600 \text{ грн.}$$

У табл. 8.21 стовпці 1 і 2 містять вихідні дані, стовпці 3 і 4 складені в процесі обчислення середньої арифметичної. Уявімо собі вісь абсцис у вигляді балки, що несе навантаження у формі графіка розподілу (рис. 8.8). На графіку точка $x$ є середньою арифметичною. Таким чином, середня арифметична представляє собою абсцис центру тяжкості графіка розподілу.

На відміну від середньої арифметичної, яка розраховується на основі всіх варіантів значень ознаки, *мода* і *медіана* характеризують

Таблиця 8.21

| Заробітна плата, грн | Число робітників ($f$) | Центр інтервалу ($x$) | $xf$ |
|---|---|---|---|
| 6 000–7 000 | 20 | 6 500 | 130 000 |
| 7 000–8 000 | 50 | 7 500 | 375 000 |
| 8 000–9 000 | 30 | 8 500 | 255 000 |
| Всього | 100 | – | 760 000 |

величину варіанта, що займає певне становище в ранжированому ряду. *Медіаною* (*Me*) називається величина ознаки, яка володіє наступною властивістю: число елементів сукупності зі значеннями даної ознаки, більшими за цю величину, дорівнює числу зі значеннями ознаки, меншими її. Положення медіани визначається її номером $N_{Me} = \frac{n+1}{2}$, де $n$ — число досліджуваних одиниць.

Наводимо результати вимірювання діаметрів дев'яти деталей (відхилення від номіналу в мікронах): 23, 21, 18, 22, 17, 25, 24, 21, 26. Знайдемо медіану. Перш за все розташуємо ці дані в порядку зростання: 17, 18, 21, 21, 22, 23, 24, 25, 26. Діаметр деталі, що займає в цьому ряду центральне, 5-е місце, виявився рівним 22 мм, отже, медіана — 22.

Якщо кількість даних парна, то на центральних місцях виявляється два елементи; як медіана приймаю напівсуму значень ознаки у цих елементів.

Наприклад, є десять вимірів деталей, розташованих у порядку зростання: 17, 18, 21, 21, 22, 23, 24, 25, 26, 27. У цьому випадку медіана складе $\frac{22+23}{2} = 22{,}5$.

У такий спосіб обчислюють медіану при невеликій кількості даних. У разі великої кількості даних слід застосовувати групування та проводити розрахунок медіани для ряду розподілу.

Використовуючи дані прикладу, наведеного в табл. 8.10, для визначення медіани та моди. $N_{Me} = \frac{60+1}{2} = 30{,}5$, тобто медіана середньої арифметичної 30 і 31 значень ознаки. За нагромадженими частотами визначаємо, що 30-й та 31-й члени ряду мають величину ознаки, що дорівнює 4-му розряду, тобто медіана дорівнює четвертому розряду. *Мода* (*Mo*) — значення ознаки, яке найбільш часто зустрічається в сукупності. Для дискретних рядів мода визначається без обчислень, безпосередньо за таблицею або за графіком розподілу. Для цього ряду сукупності розподілу робітників за розрядами мода також дорівнює четвертому розряду. В інтервальному ряду розподілу відразу можна визначити інтервал, в якому будуть мода і медіана. Для їх розрахунку використовуються такі формули:

$$Me = x_{Me} + i_{Me}\frac{\frac{\sum f}{2} - S_{Me-1}}{f_{Me}}, \qquad (8.11)$$

де $x_{Me}$ — нижня межа медіанного інтервалу;

$i_{Me}$ — довжина інтервалу;

$\sum f$ — сума частот ряду;

$f_{Me}$ — частота медіанного інтервалу;

$S_{Me-1}$ — сума нагромаджених частот, що передують медіанному інтервалу.

Використовуючи дані прикладу, наведеного в табл. 8.9, розрахуємо медіану. У нашому прикладі

$$x_{Me}=49{,}937; \quad i_{Me}=0{,}008; \quad \sum f=20; \quad S_{Me-1}=5; \quad f_{Me}=7.$$

За нагромадженими частотами визначаємо, що медіана перебуває в інтервалі 49,937–49,945 і тоді

$$Me = 49{,}937 + 0{,}008\frac{\frac{20}{2} - 5}{7} = 49{,}943\ \text{мм}.$$

Найбільша частота відповідає інтервалу 49,937–49,945 мм, тобто мода повинна бути в цьому інтервалі і її величину визначаємо за формулою:

$$Mo = x_0 + i\frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})}, \qquad (8.12)$$

де $x_0$ — нижня межа модального інтервалу;

$i$ — величина модального інтервалу;

$f_{m-1}$ — частота інтервалу, що передує модальному;

$f_m$ — частота модального інтервалу;

$f_{m+1}$ — частота інтервалу наступного за модальним.

Наведена формула може бути використана у варіаційних рядах з рівними інтервалами

$$Mo = 49{,}937 + 0{,}008\frac{7-3}{(7-3)+(7-6)} = 49{,}943\ \text{мм}.$$

Медіану та моду можна визначити з використанням графічного методу. Медіана визначається по кумуляті (рис. 8.9).

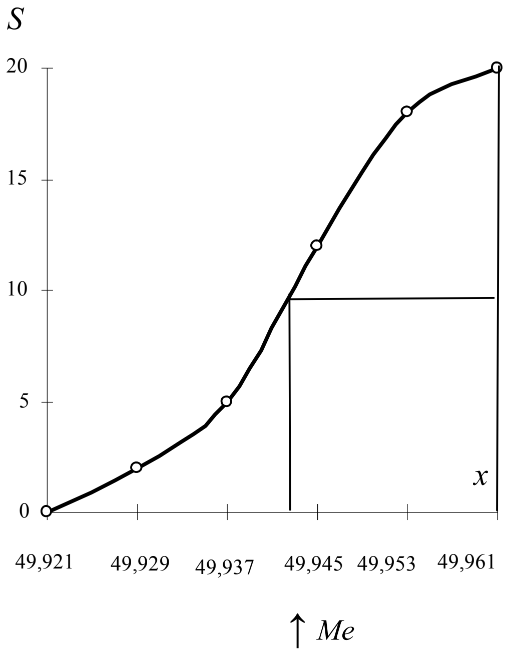



*Рис. 8.9.* Кумулята інтервального варіаційного ряду: *S* — нагромаджені частоти; *x* — діаметр деталей, мм.

Для її визначення висоту найбільшої ординати, що відповідає загальній ординаті, яка відповідає загальній чисельності сукупності, ділять навпіл. Через отриману точку проводять пряму, паралельну осі абсцис до перетину її з кумулятою. Абсцис точки перетину є медіаною. Моду можна також визначити графічно.

Основним показником центру розподілу для варіаційних рядів є середня арифметична. В окремих випадках середня арифметична може бути доповнена обчисленням моди та медіани. Наприклад, при статистичному контролі якості продукції замість середньої найчастіше використовують медіану, оскільки визначення медіани не потрібно спеціального розрахунку; крім того, вона не чутлива до крайніх значень.

Структурні середні набули широкого поширення при дослідженні попиту населення продукції народного споживання (наприклад, взуття, одягу тощо). Тут інтерес представляє визначення модального розміру, тобто розміру, що має найбільший попит. В цьому випадку розрахунок середньої тут немає економічного сенсу.

## 8.6. Показники коливання (варіації) ознак

Середні величини, що характеризують варіаційний ряд одним числом, не враховують варіацію ознаки, тим часом ця варіація існує. Щоб розширити дослідження варіаційного ряду, дослідник повинен обчислити та проаналізувати показники, що характеризують варіацію ознаки.

Варіаційний ряд хоч і показує наявність відмінностей у значеннях ознак у одиниць сукупності явищ, але він не дає єдиного узагальнююючого показника варіації. Як же можна виміряти та оцінити варіацію ознаки?

Варіація може бути виражена узагальнюючими величинами. Для вимірювання ступеня варіювання ознаки в статистиці найчастіше застосовують такі показники варіації: розмах варіації, середнє лінійне відхилення, дисперсія, середнє квадратичне відхилення і коефіцієнт варіації. Кожен із наведених показників використовується для вирішення певних завдань статистичного аналізу.

*Значення показників варіації полягає в наступному*[1]:

1. Показники варіації доповнюють середні величини, за якими приховуються індивідуальні відмінності.
2. Показники варіації характеризують ступінь однорідності статистичної сукупності за цією ознакою.
3. Показники варіації характеризують межі варіації ознаки.
4. Співвідношення показників варіації характеризує взаємозв'язок між ознаками.

Розглянемо конкретні показники варіації.

**Розмах варіації**. Найбільш простим показником, яким може охарактеризувати варіацію ознаки у цьому ряду розподілу, є розмах варіації. *Варіаційний розмах* (*розкид*) — амплітуда коливання, або широта розсіювання, є різницею між максимальним і мінімальним значеннями варіаційного ряду:

$$R = x_{\max} - x_{\min}, \tag{8.13}$$

де $R$ — розмах варіації;

$x_{\max}$ — найбільше значення ознаки варіаційного ряду;

$x_{\min}$ — найменше значення ознаки варіаційного ряду.

*Розмах варіації характеризує межі коливання (варіації) індивідуальних значень (або варіантів) ознаки у статистичній сукупності (або варіаційному ряду).* Величина $R$ нестійка і залежить від випадкових обставин.

Для ряду, що становить розподіл деталей за вагою (табл. 8.4), цей показник має таку величину:

$$R = 86 - 80 = 6.$$

Це найпростіший і найменш точний показник варіації, який, влас-

---

[1] *Суслов И. П.* Общая теория статистики. Учеб. пособие. Изд. 2-е, перераб. и доп. Москва, «Статистика», 1978, с. 189.

не, характеризує межі варіації в сукупності. Застосовується як приблизна оцінка варіації. Зрозуміло, що величина цього показника цілком залежить від випадковостей розташування крайніх значень ознаки ряду розподілу.

Розмах варіації показує, в яких межах відбувається коливання ознаки, що утворює ряд розподілу, і є мірою його варіації. Обчислюється у тих самих одиницях виміру, як і ознака, варіацію якого він характеризує. Розрахунок показника розмаху варіації стажу роботи робочих окремих професій наведено в табл. 8.22.

Розмах має самостійне значення. Наприклад, в промисловості, де допуски точності виробів, що виготовляються, встановлюються в певних межах, відповідних іноді величині розмаху варіації їх ознак.

Обчислення розмаху варіації відрізняється простотою, що дозволяє використовувати його при дослідженні суспільних явищ і процесів. У той самий час розмах варіації неспроможний повною мірою охарактеризувати коливання ряду, оскільки: 1) визначається лише двома крайніми значеннями, що робить певною мірою його величину нестійкою, яка надзвичайно залежить від випадкових обставин; 2) не враховує варіювання значень ознаки в основній масі членів ряду, а тому не відображає коливання ряду в цілому.

Таблиця 8.22

**Розрахунок розмаху варіації**

| Професії | Стаж роботи, років | | $R$ |
|---|---|---|---|
| | максимальний | мінімальний | |
| Токар | 7 | 1 | 6 |
| Слюсар | 16 | 2 | 14 |
| Формувальник | 22 | 3 | 19 |
| Штампувальник | 24 | 4 | 20 |

Ця обставина робить його дуже не надійним показником варіації. Тому математична статистика віддає перевагу використовувати такі показники варіації, які враховують варіювання ознак у всіх членів ряду даної сукупності. Більш надійним показником для вимірювання варіації ознак є середній розмах.

**Середнє абсолютне відхилення**. Індивідуальні значення ознак відхиляються від середньої в той чи інший бік. Щоб визначити варіацію ознаки одиниць сукупності, треба знайти середню міру відхилень кожного значення ознаки від середньої.

Позначимо ознаки, що варіюються, в окремих одиниць сукупності: $x_1, x_2, x_3, \dots, x_n$, де $n$ — число одиниць сукупності.

Віднімаючи з кожного значення ознаки його середню величину, отримаємо:

$$x_1 - \bar{x} = d_1, \quad x_2 - \bar{x} = d_2, \quad x_3 - \bar{x} = d_3 \text{ і т. д.}$$

При цьому завжди передбачається, що середню віднімають із індивідуальних значень ознаки, а не навпаки. Отже, відхилення значень ознаки від середньої можуть бути позитивними або негативними значеннями. Позитивне значення відхилення завжди вказує, що дана варіанта більше середньої, а негативне відхилення показує, що варіанта менше середньої.

З властивостей середньої арифметичної нам відомо, що алгебраїчна сума цих відхилень ($\sum d_i$) дорівнює нулю, так як індивідуальні відхилення від середньої в силу властивості середньої взаємно погашаються. Отже, щоб обчислити середню арифметичну з відхилень, потрібно умовно припустити, що всі відхилення, що мають позитивні та негативні, мають однаковий знак. Тоді для визначення *середнього абсолютного відхилення*, яке часто називають лінійним відхиленням, необхідно взяти значення відхилень за абсолютною величиною без урахування знака.

*Середнє абсолютне відхилення* або *лінійне відхилення* є узагальнюючим показником і представляє собою середню арифметичну з абсолютних (тобто без урахування знака) значень відхилень варіантів від їх середнього рівня.

В залежності від відсутності чи наявності частот обчислюють середнє лінійне відхилення незважене чи зважене.

Середнє абсолютне відхилення для первинних даних, де часто варіант дорівнює одиниці, $\bar{d}$ визначається за наступною формулою:

$$\bar{d} = \frac{\sum d}{n} = \frac{\sum |x - \bar{x}|}{n} =$$

$$= \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + \ldots + |x_n - \bar{x}|}{n}. \qquad (8.14)$$

де $\bar{d}$ — середнє абсолютне відхилення;
$d$ — абсолютне відхилення;
$x$ — значення варійованої ознаки;
$n$ — число одиниць сукупності.

Якщо дані спостереження представлені у вигляді дискретного ряду розподілу з частотами, середнє лінійне відхилення обчислюється за наступною формулою:

$$\bar{d} = \frac{\sum \left| x - \bar{x} \right| f}{\sum f} =$$

$$= \frac{\left| x_1 - \bar{x} \right| f_1 + \left| x_2 - \bar{x} \right| f_2 + \left| x_3 - \bar{\bar{x}} \right| f_3 + \ldots + \left| x_n - \bar{x} \right| f_n}{f_1 + f_2 + f_3 + \ldots + f_n}. \qquad (8.15)$$

де $f$ — частота повторень окремих значень (варіантів) ознаки.

Цей показник представляє собою зважену частотами середню з абсолютних (тобто без урахування знака) відхилень варіант від середньої арифметичної. Прямі дужки, в яких заключні різниці між варіантами і середньою, показують, що безпосереднє підсумовування і підсумовування після зважування проводиться без урахування знаків. Це дає змогу не врахувати знаки відхилень варійованої ознаки від середньої.

Розмірність середнього абсолютного відхилення відповідає розмірності ознаки, що варіює.

Розглянемо приклад розрахунку середнього лінійного відхилення зваженого за вище наведеною формулою, використовуючи умовні дані про розподіл робітників за стажем роботи (табл. 8.23).

Таблиця 8.23

**Розподіл робітників за стажем роботи на підприємстві**

| Стаж роботи робітників, років ($x$) | Кількість робітників ($f$) | $xf$ | $\lvert x-\bar{x}\rvert$ | $\lvert x-\bar{x}\rvert f$ |
|---|---|---|---|---|
| 2 | 2 | 4 | 3 | 6 |
| 3 | 5 | 15 | 2 | 10 |
| 5 | 8 | 40 | 0 | 0 |
| 7 | 3 | 21 | 2 | 6 |
| 10 | 2 | 20 | 5 | 10 |
| Всього | 20 | 100 | — | 32 |

Визначимо середній стаж роботи:

$$\bar{x} = \frac{\sum xf}{f} = \frac{100}{20} = 5 \text{ років.}$$

Відхилення кожного значення ознаки від середньої і зважені відхилення представлені в таблиці.

Розрахуємо середнє лінійне відхилення:

$$\bar{d} = \frac{\sum \left| x - \bar{x} \right| f}{\sum f} = \frac{32}{20} = 1{,}6 \text{ роки.}$$

У даному випадку це означає, що відхилення стажу роботи окремих робітників відхиляються у той чи інший бік в середньому на 1,6 року.

Якщо статистичні дані представлені у вигляді інтервального ряду розподілу, то попередньо визначається дискретна величина ознаки в кожній групі, а потім проводиться розрахунок за середньою арифметичною зваженою, як зазначено вище.

Зважаючи на те, що середнє лінійне відхилення не враховує знака відхилення, воно є недостатньо повною характеристикою варіації. Завдяки штучному характеру цієї операції математичні властивості показника $\bar{d}$ є дуже складними, що робить його використання незручним при детальному вивченні статистичних даних. В результаті абстрагування від знака відхилення середнє абсолютне відхилення як міра варіації застосовують у статистиці дуже рідко. Цей показник використовується, наприклад, для характеристики однорідності ниток та пряжі у текстильній промисловості. Тому в математичній статистиці віддають перевагу іншому способу усунення знака, а саме зведенню кожного значення відхилення від середньої в квадрат.

**Середнє квадратичне відхилення**. Для повнішої характеристики варіації застосовуються показники середнього квадратичного відхилення та дисперсії. Середнє квадратичне відхилення ($\sigma$) є також узагальнюючим показником коливання ознаки і характеризує середній показник відхилення варіант ряду від їх загальної середньої. Виражається середнє квадратичне відхилення у тих самих іменованих числах, у яких виражені варіанти і середня величина (гривні, кілограми, тонни, відсотки і т. п.). Середнє квадратичне відхилення, так само як і лінійне відхилення, показує, на скільки в середньому відхиляються індивідуальні ознаки середньої.

*Середнє квадратичне відхилення* дорівнює квадратному кореню із середнього квадрата відхилень окремих значень ознаки від середньої арифметичної.

Залежно від відсутності чи наявності частот обчислюють середнє квадратичне відхилення незважене та зважене.

*Середнє квадратичне відхилення незважене* обчислюють за такою формулою:

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}, \qquad (8.16)$$

або $\sigma$

$$\sigma = \sqrt{\frac{\sum d^2}{n}}, \qquad (8.17)$$

де $\sigma$ — середньоквадратичне відхилення;

$x$ — величина варіантного ряду;

$\bar{x}$ — середньоарифметична ряду;

$\sum$ — знак, що означає суму;

$n$ — кількість варіантів ряду;

$d$ — абсолютне відхилення ($d = x - \bar{x}$).

Цей показник широко використовується в якості загальноприйнятого показника варіації не тільки у статистиці, а й у техніці, біології та інших галузях знань.

Примітка. При обчисленні середнього квадратичного відхилення для малих вибірок (наприклад, 15–20) сума квадратів відхилень ділиться на варіант мінус одиниця, тобто

$$\sigma = \sqrt{\frac{\sum d^2}{n-1}}.$$

Дільник тут називається число ступенів свободи. Обґрунтування цього прийому ґрунується на абстрактних математичних викладках.

Слід зазначити, що розрахунок середнього квадратичного відхилення, як і середнього лінійного відхилення, має логічне значення в тому випадку, якщо фактичні значення досліджуваної ознаки дуже близькі до нормального. Для явно асиметричних розподілів його розрахунок немає значення.

Наведемо приклад розрахунку середнього квадратичного відхилення. В табл. 8.24 наведені індивідуальні значення ознаки (обсягу відпрацьованих станко-годин).

Таблиця 8.24

**До розрахунку середнього квадратичного відхилення**

| Цех | Відпрацьовано станко-годин | $x-\bar{x}$ | $(x-\bar{x})^2$ |
|---|---|---|---|
| 1 | 70 | – 10 | 100 |
| 2 | 60 | –20 | 400 |
| 3 | 80 | 0 | – |
| 4 | 80 | 0 | – |
| 5 | 90 | + 10 | 100 |
| 6 | 100 | + 20 | 400 |
| Всього | 480 | 0 | 1 000 |

Розрахуємо середню:

$$\bar{x} = \frac{70+60+80+80+90+100}{6} = \frac{480}{6} = 80 \text{ (станко-годин)},$$

тоді

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} = \sqrt{\frac{1000}{6}} = 12{,}9 \text{ (год.)}.$$

За формулою (8.16) ми обчислюємо $\sigma$ для не згрупованих даних. Вона називається *простою*.

Для варіаційних рядів *середнє квадратичне відхилення* визначається з урахуванням частот за формулою:

$$\sigma = \sqrt{\frac{\sum (x-\overline{x})^2 f}{\sum f}} = \sqrt{\frac{\sum d^2 f}{\sum f}}, \tag{8.18}$$

де $f$ — вага (частоти).

Формули середнього квадратичного відхилення визначаються із формул дисперсії (8.22) та (8.23).

Чим більше $\sigma$, тим більше мінливість ознаки в досліджуваній сукупності; що менше $\sigma$, то більш однорідна сукупність. Разом з тим оцінка величини $\sigma$ значною мірою визначається сутністю досліджуваних явищ. Наприклад, коливання заробітної плати робітників у розмірі 100 грн не характеризують суттєвих відмінностей в оплаті праці, але водночас коливання ціни реалізації продукції в кілька гривень характеризуватимуть різке коливання цього показника.

Перевагою цього показника в порівнянні з середнім лінійним відхиленням є те, що при його обчисленні жодного умовного припущення про необхідність підсумовування відхилень варіантів від середньої без урахування знаків ми не робимо, а використовуємо формулу середньої квадратичної, за якою при зведенні відхилень у квадрат їх знак байдужий.

Ми розглянули розрахунок середнього квадратичного відхилення у найпростіших випадках при обмеженій кількості варіант, які можна все розташувати підряд. Для рядів, де варіанти розгруповані так, що кожної групи дана своя чисельність, розрахунки середнього квадратичного відхилення дещо ускладнюються.

Приклад обчислення зваженої $\sigma$ для дискретних рядів розподілу наведено в табл. 8.25.

Обчислимо середній виробіток на одного робітника:

$$\overline{x} = \frac{\sum xf}{\sum f} = \frac{300}{50} = 6 \text{ деталей}.$$

Значить, середнє квадратичне відхилення становитиме:

$$\sigma = \sqrt{\frac{\sum (x-\overline{x})^2 \cdot f}{\sum f}} = \sqrt{\frac{74}{50}} = \pm 1{,}22 \text{ деталей}.$$

Середнє квадратичне відхилення показує, наскільки в середньому коливається величина ознаки від їхнього середнього значення.

Таблиця 8.25

**Розподіл робітників за обсягом виробленої продукції**

| Вироблено продукції одним робітником за зміну, шт. ($x$) | Кількість робітників ($f$) | $xf$ | $x-\bar{x}$ | $(x-\bar{x})^2$ | $(x-\bar{x})^2 f$ |
|---|---|---|---|---|---|
| 4 | 7 | 28 | – 2 | 4 | 28 |
| 5 | 10 | 50 | – 1 | 1 | 10 |
| 6 | 15 | 90 | 0 | 0 | 0 |
| 7 | 12 | 84 | + 1 | 1 | 12 |
| 8 | 6 | 48 | + 2 | 4 | 24 |
| Всього | 50 | 300 | | | 74 |

Так, у цьому прикладі середня величина коливання кількості продукції, виробленої одним робітником, становить 1,22 деталі. Використовуючи частості $p'$ можна розрахувати середнє квадратичне відхилення $\sigma$ за такою формулою:

$$\sigma = \sqrt{\sum (x-\bar{x})^2 p'}. \quad (8.19)$$

Розглянемо приклад розрахунку середнього квадратичного відхилення за формулою (8.19) за даними табл. 8.26.

Таблиця 8.26

**Розподіл робітників за обсягом виробленої продукції**

| Вироблено продукції одним робітником за зміну, шт. ($x$) | Кількість робітників ($f$) | Частості $\frac{f}{\sum f} = p'$ | $(x-\bar{x})^2$ | $(x-\bar{x})^2 p'$ |
|---|---|---|---|---|
| 4 | 7 | 0,14 | 4 | 0,56 |
| 5 | 10 | 0,20 | 1 | 0,20 |
| 6 | 15 | 0,30 | 0 | 0 |
| 7 | 12 | 0,24 | 1 | 0,24 |
| 8 | 6 | 0,12 | 4 | 0,48 |
| Всього | 50 | 1,00 | | 1,48 |

Використовуючи у розрахунках зокрема, наведені у табл. 8.26 визначаємо середнє квадратичне відхилення за формулою (8.19). Результат буде таким самим:

$$\sigma = \sqrt{\sum (x-\bar{x})^2 p'} = \sqrt{1,48} = 1,22.$$

Таким чином, при середньому виробітку в 6 деталей склад робітників за рівнем виробітку не дуже відрізняється, так як $\sigma = 1{,}22$.

У разі інтервальної форми рядів розподілу у обчисленнях беруть участь центри інтервалів (тобто ряд перетворюється на дискретний). Обчислимо середнє відхилення для інтервального ряду (табл. 8.27).

Таблиця 8.27

| Денна заробітна плата, тис. грн ($x$) | Число робітників ($f$) | Центр інтервалу $x'$ | $x'f$ | $x' - \bar{x}$ | $(x' - \bar{x})f$ | $(x' - \bar{x})^2 f$ |
|---|---|---|---|---|---|---|
| 1,7–2,3 | 3 | 2,0 | 6 | –1,23 | –3,69 | 4,5387 |
| 2,3–2,9 | 4 | 2,6 | 10,4 | –0,63 | –2,52 | 1,5876 |
| 2,9–3,5 | 6 | 3,2 | 19,2 | –0,03 | –0,18 | 0,0054 |
| 3,5–4,1 | 3 | 3,8 | 11,4 | 0,57 | 1,71 | 0,9747 |
| 4,1–4,7 | 4 | 4,4 | 17,6 | 1,17 | 4,68 | 5,4756 |
| Всього | 20 | | 64,6 | | –12,78 | 12,5820 |

$$\bar{x} = \frac{64{,}6}{20} = 3{,}23 \text{ тис. грн;}$$

$$\sigma^2 = \frac{12{,}5820}{20} = 0{,}6291 \text{ грн;}$$

$$\sigma = \sqrt{0{,}6291} = 0{,}7932.$$

Середнє квадратичне відхилення завжди виражається у тих самих іменованих числах (зріст — у сантиметрах, вага — у тоннах, стаж — у роках, урожайність — у центнерах і т. п.), як і середня арифметична.

Математична статистика віддає перевагу середньому квадратичному відхиленню середнього абсолютного відхилення. Воно використовується у статистичних дослідженнях дуже широко (у вимірі зв’язків, при організації вибіркового спостереження тощо). Середнє квадратичне відхилення є дуже важливою характеристикою ще й тому, що воно засноване на математичній властивості середньої (див. § 7, властивість 8).

Середнє квадратичне відхилення від середньої так само, як і середнє лінійне відхилення, показує, у скільки разів у середньому відхиляються конкретні варіанти від середнього їх значення. Розрахунок $\sigma$ трохи складніше, ніж розрахунок $\bar{d}$, але $\sigma$ більш чутливо реагує на варіацію і органічно вписується в апарат математичної статистики.

**Зв’язок середнього квадратичного відхилення ($\sigma$) із середнім**

**лінійним відхиленням ($\bar{d}$)**. Між середнім квадратичним відхиленням ($\sigma$) і середнім лінійним відхиленням ($\bar{d}$) існує певне співвідношення. За якістю міжорантності середнє квадратичне відхилення $\sigma$ завжди більше середнього лінійного відхилення $\bar{d}$. Якщо обсяг сукупності досить великий і розподіл ознаки близький до нормального або симетричного розподілу, то між $\sigma$ і $\bar{d}$ є певне співвідношення:

$$\sigma = \bar{d}\sqrt{\frac{\pi}{2}} \approx d \cdot 1{,}25 \quad \text{або} \quad \bar{d} \approx 0{,}8\sigma. \tag{8.20}$$

Знаючи це співвідношення, можна за відомим показником визначати невідомий, скажімо, по $\bar{d}$ визначати $\sigma$. Відхилення $\frac{\sigma}{\bar{d}}$ від 1,25 залежить від близькості розподілу до нормального.

Приклад. За даними табл. 8.8 знайти відношення між $\sigma$ і $\bar{d}$.

Маємо:

$$\sigma = 0{,}7932;$$

$$\bar{d} = \frac{|-3{,}69| + |-2{,}52| + |-0{,}18| + |1{,}71| + |4{,}68|}{20} = 0{,}6390.$$

Звідси

$$\frac{\sigma}{\bar{d}} = \frac{0{,}7932}{0{,}6390} = 1{,}2418 \approx 1{,}24.$$

Це відношення не набагато відрізняється від теоретичного значення (1,25), що опосередковано свідчить про близькість взятого розподілу до нормального.

**Поправка на варіацію всередині інтервалів**. При обчисленні узагальнюючих статистичних показників для інтервальних рядів розподілу дійсні значення ознаки замінюються центральними значеннями інтервалів, які у більшій або меншій мірі відрізняються від середньої арифметичної, включеної в інтервал. Це призводить до появи систематичної похибки під час розрахунку дисперсії. В. Ф. Шеппард встановив, що помилка дисперсії, викликана застосуванням при розрахунку згрупованих даних, становить $1/12$ квадрата величини інтервалу, тобто скоригована дисперсія дорівнює $\sigma^2 - 1/12\, i^2$.

Таким чином, з урахуванням цієї поправки величина дисперсії дорівнює

$$\sigma_c{}^2 = \sigma^2 - \frac{1}{12}k^2, \tag{8.21}$$

де $\sigma^2$ — величина дисперсії, отримана виходячи з прирівнювання всіх значень всередині інтервалів їх центрам;

$k$ — величина інтервалів (при рівних інтервалах);

$\sigma_c^2$ — дисперсія з поправкою Шеппарда.

Поправка Шеппарда повинна застосовуватися за таких умов: 1) розподіл відноситься до ознаки з безперервним характером варіації; 2) розподіл характеризується тісною близькістю з віссю абсцис на кінцях кривої; 3) розподіл побудовано за великою кількістю вихідних даних ($n$>500). При зменшенні величини інтервалу також зменшується величина поправки. При $k$<0,5 поправки Шеппарда нехтують.

Зміст середнього квадратичного відхилення той же самий, що і середнього абсолютного відхилення. Воно представляє середню величину відхилень варіант ряду від середньої арифметичної в ту чи іншу сторону.

Середнє квадратичне відхилення використовується в статистичних дослідженнях досить широко (при оцінці зв'язку між факторами, у вибіркових спостереженнях і т. п.). Середнє квадратичне відхилення є і відомим «мірилом» середньої.

**Дисперсія**. У статистиці найчастіше в якості міри коливання ознаки застосовується середній квадрат відхилень — дисперсія. *Дисперсією* називається середня із квадратів відхилень варіантів від їх середньої (середній квадрат відхилень). Якщо дисперсію обчислюють для всієї сукупності, то її позначають $\sigma^2$ і називають загальною дисперсією.

Дисперсія знаходиться за формулами:

для незваженого ряду

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}, \tag{8.22}$$

для зваженого ряду

$$\sigma^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f}. \tag{8.23}$$

Таким чином, загальна дисперсія є середня арифметична із квадратів відхилень від їх середньої арифметичної.

Обчислення дисперсії зазвичай передує підрахунку середнього квадратичного відхилення, проте дисперсія, як видно з подальшого, маємо самостійне значення.

Дисперсію за формулою зваженого ряду обчислюють у такій послідовності. Спочатку обчислюють середню. Далі необхідно визначити різниці варіанта ознаки і середньої, звести їх у другу ступінь і зважити.

Для обчислення дисперсії зручно користуватися такими формулами:

$$\sigma^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \overline{x^2} - \overline{x}^2. \tag{8.24}$$

Таким чином, дисперсія дорівнює різниці середньої із квадратів мінус квадрат середньої.

Якщо дисперсія зважується, вона виражається формулою:

$$\sigma^2 = \frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2. \tag{8.25}$$

Формулу для розрахунку дисперсії можна перетворити наступним чином:

$$\sigma^2 = \frac{\sum (x_i - \overline{x})^2}{n} = \frac{\sum (x_i^2 - 2\overline{x}x_i + \overline{x}^2)}{n} = \frac{\sum x_i^2 - 2\overline{x}\sum x_i + n\overline{x}^2}{n} =$$

$$= \frac{\sum x_i^2}{n} - 2\overline{x}\overline{x} + \overline{x}^2 = \frac{\sum x_i^2}{n} - \overline{x}^2 = \overline{x^2} - \overline{x}^2. \tag{8.26}$$

Таблиця 8.28

| Розряд робітників механічного цеху | Число робітників ($f$) | $xf$ | $x^2 f$ |
|---|---|---|---|
| 2 | 4 | 8 | 16 |
| 3 | 16 | 48 | 144 |
| 4 | 23 | 92 | 368 |
| 5 | 12 | 60 | 300 |
| 6 | 5 | 30 | 180 |
| Всього | 60 | 238 | 1 008 |

Цю формулою використовують і при машинній обробці вихідних даних. Знайдемо дисперсію за даними табл. 8.28.

Скористаємося формулою 8.25:

$$\sigma^2 = \frac{1\,008}{60} - \left(\frac{238}{60}\right)^2 = 1{,}07.$$

Більші значення дисперсії свідчать про більші відхилення показників ряду розподілу від центру розподілу.

**Коефіцієнт варіації**. Усі перелічені вище міри варіації (коливання) представляють собою абсолютні величини, виражені у тих самих одиницях виміру, як і варіанти. Так, наприклад, середнє квадратичне відхилення кількості оброблених деталей вимірюється в штуках, тривалість горіння лампочок у хвилинах і т. д. Порівняння коливання і надійності середніх в різних рядах за допомогою $\sigma$ неможливо, якщо, наприклад, дані цих рядів виражені в різноіменних величинах. Так, наприклад, обчисливши середнє квадратичне відхилення відпрацьованих верстато-годин і виробітки токарними верстатами, ми не може визначити, варіація якої ознаки більше, тому що в першому випадку вона вимірюється верстато-годинах, у другому випадку — в штуках.

Враховуючи, що середнє лінійне відхилення і середнє квадратичне відхилення являють собою абсолютні величини, виражені в тих же одиницях, що і варіанти, для характеристики коливання різних ознак використовується відносний показник міри варіації — коефіцієнт варіації.

*Коефіцієнт варіації* ($V$) представляє собою відношення середньоквадратичного відхилення до середньої величини варійованої. Він характеризує статистичну сукупність з погляду самого ступеня розкиданості окремих значень ознаки $x_i$ навколо своєї середньої величини $\bar{x}$. Зазвичай виявляється у процентах. Представляє собою відношення абсолютних величин: або варіаційного розмаху ($R$), або середнього лінійного відхилення ($\bar{d}$), або середнього квадратичного відхилення ($\sigma$) до середньої арифметичної ($\bar{x}$) (при цьому $\bar{x} \neq 0$).

Принцип побудови коефіцієнта варіації такий:

$$\text{Коефіцієнт варіації} = \frac{\text{Іменований показник варіації}}{\text{Середня арифметична або величина, що її замінює}^{1}}$$

Коефіцієнт варіації за середнім квадратичним відхиленням обчислюється за формулою:

$$V_\sigma = \frac{\sigma}{\bar{x}}; \qquad V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100; \tag{8.27}$$

де $V_\sigma$ — коефіцієнт варіації;

$\sigma$ — середньоквадратичне відхилення;

$\bar{x}$ — середня величина ознаки.

Таким чином, коефіцієнт варіації будь-якого ряду дорівнює його середньому квадратичному відхиленню, поділеному на середню ари-

---

[1] Чаще всього мода і медіана.

фметичну.

Якщо в якості іменованого показника варіації використовується середнє лінійне відхилення, то коефіцієнт варіації ($V_{\bar{d}}$) може бути обчислений за наступною формулою:

$$V_{\bar{d}} = \frac{\bar{d}}{\bar{x}}; \quad V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100. \tag{8.28}$$

Видозмінений показник коефіцієнта варіації по середньому лінійному відхиленні ($V_{\bar{d}}$) є показником нерівності. Він знайшов широке застосування в текстильній промисловості як міра коливання при вивченні нерівності пряжі (за товщиною, вагою та ін.).

Якщо як іменований показник варіації використовується розмах варіації ($R$), то коефіцієнт варіації ($V_R$) може бути обчислений за наступною формулою:

$$V_R = \frac{R}{\bar{x}}; \quad V_R = \frac{R}{\bar{x}} \cdot 100. \tag{8.29}$$

Іноді його називають коефіцієнтом *осциляції*.

В окремих випадках для розрахунку коефіцієнта варіації використовується наступна формула:

$$V_{\bar{d}} = \frac{Mo}{\bar{x}} \cdot 100. \tag{8.30}$$

де $Mo$ — мода.

Перший коефіцієнт найбільш уживаний, і якщо немає інших вказівок, то під коефіцієнтом варіації мається на увазі саме його.

Оскільки середнє квадратичне відхилення вимірюється у тих самих одиницях, як і сама ознака, і служить одним із найважливіших показників, що характеризують силу коливання значень ознаки, то самі величини середніх квадратичних відхилень не можуть бути безпосередньо порівняні. Це і зумовлює необхідність обчислення відносного показника, в якому б погашалися відмінності у вимірі абсолютних показників варіації. Ділення іменованого показника варіації на середню величину ознаки робить коефіцієнт варіації абстрактним числом. У цьому вигляді його можна використовувати для порівняння коливання явищ у різних сукупностях (наприклад, заробітної плати працівників різних цехів, спеціальностей; урожайності різних культур, вироблення продукції тощо).

Чим більший коефіцієнт варіації, тим більше значення ознаки відрізняються один від одного за величиною, і, отже, менш однорідна статистична сукупність. І навпаки, чим коефіцієнт варіації менше, тим однорідніша сукупність. Велике значення має коефіцієнт варіації для оцінки надійності середньої величини. Чим менший коефіцієнт варіації, тим надійніше середня величина, що для її практичного застосування дуже важливо. Коефіцієнт варіації, будучи відносною величиною, абстрагує відмінності абсолютних величин варіації різних ознак і дає можливість порівняння її.

Коефіцієнт варіації дозволяє порівнювати силу коливань ознак у різних сукупностей явищ, різних і в значенні обсягів самих сукупностей, і в значенні різнойменних або різномаштабних величин варійованої ознаки.

На відміну від середнього відхилення коефіцієнт варіації є відносною величиною, що дозволяє його використовувати для оцінки варіації ознак різних сукупностей. Його також використовують не тільки для порівняльної оцінки, але і як характеристику однорідної сукупності. Сукупність вважається однорідною, якщо коефіцієнт варіації не перевищує 33 % (для розподілів, близьких до нормального).

Коефіцієнти варіації дають уявлення про ступінь коливання ознак сукупності і дозволяють порівнювати ступінь варіації ознак у варіаційних рядах з різним рівнем середніх. Наприклад, якщо для урожайності ячменю в одному районі $\sigma$ = 12 ц і $\bar{x}$= 48 ц, а в іншому районі $\sigma$ =10 ц і $\bar{x}$= 25 ц, то по абсолютній величині варіація в першому районі більше, так як 12> 10, а відносна варіація менша, оскільки

$$V_1 = \frac{12}{48} \cdot 100 = 25{,}0\%; \qquad V_2 = \frac{10}{25} \cdot 100 = 40\%.$$

Коефіцієнт варіації зручний для порівняння варіації різних явищ.

Наприклад, що більше варіює: денна заробітна плата робітника або його виробіток? Застосуємо коефіцієнт варіації для порівняльної оцінки коливання різних ознак. Порівняємо варіацію двох різних ознак (заробітної плати та обсягу виробництва продукції) в одній сукупності (табл. 8.29).

Таблиця 8.29

**Показники цеху №2 за вересень 2021 р.**

| Найменування показника | Середня арифметична ($\bar{x}$) | Середнє квадратичне відхилення ($\sigma$) | Коефіцієнт варіації, % ($V_\sigma$) |
|---|---|---|---|
| Місячна заробітна плата робітників-відрядників грн | 10 800 | 3 780 | 35,0 |
| Вироблено продукції, шт. | 150 | 12 | 8,0 |

Щоб відповісти на питання про те, за яким із показників (за місячною заробітною платою або за кількістю виробленої продукції) вищеколивання ознаки, недостатньо знати тільки середнє квадратичне відхилення, так як $\sigma$ розраховується за відхиленнями від середніх, а середні рівні ознак різні за величиною.

У цьому випадку необхідно зіставити середнє квадратичне відхилення з величиною відповідної йому середньої.

Коефіцієнт варіації за розміром місячної заробітної плати працівників становитиме

$$V = \frac{3\,780}{10\,800} \cdot 100 = 35{,}0\ \%,$$

а за кількістю виробленої продукції — відповідно

$$V = \frac{12}{150} \cdot 100 = 8{,}0\ \%.$$

Отже, варіація місячної заробітної плати значно вища, ніж виробленої продукції на одного робітника.

Ґрунтуючись на коефіцієнті варіації, можна зробити висновок, що у кількості виробленої продукції сукупність є однорідною.

Коефіцієнт варіації певною мірою дозволяє оцінити типовість середньої. Так, якщо значення коефіцієнта варіації є досить великим (перевищує, скажімо 40 %), це означає, що середня характеризує сукупність за ознакою, яка істотно змінюється в окремих її одиниць. Типовість такої середньої не велика.

У коефіцієнті варіації усувається не тільки непорівнянність, пов’язана з різними одиницями вимірювання досліджуваного ознаки, але і непорівнянність, що виникає внаслідок відмінностей у величині середніх арифметичних.

Розглянемо коефіцієнти варіації глибини свердловин на трьох нафтопереробних підприємствах (у табл. 8.30). У таблиці порівнюється варіація однієї ознаки (глибини свердловин) в трьох сукупностях. Найнижчий коефіцієнт варіації на заводі № 2, що свідчить про меншеколивання глибини свердловин.

Таблиця 8.30

**Глубина скважин в районе бурения**

| Номер підприємства | Середня глибина в районі буріння, *м* ($\bar{x}$) | Коефіцієнт варіації, % ($V_\sigma$) |
|---|---|---|
| 1 | 400 | 21 |
| 2 | 620 | 12 |
| 3 | 720 | 28 |

Як очевидно, коефіцієнт варіації дозволяє порівнювати силу коливання (варіацію) ознак у різних сукупностей, і в значенні різноіменних або різномаштабних величин варійованої ознаки.

**Квартилі та децилі**. У системі порядкових характеристик варіаційного ряду назва квартилів отримали варіанти, що займають певне місце: десяте, дванадцяте, п'ятдесяте і т. д.

Як відомо, медіана — це варіант, який ділить впорядкований варіаційний ряд на дві рівні за обсягом групи. У свою чергу, кожну групу можна поділити на дві підгрупи, що відповідає певному варіанту. Такі варіанти називають квартилями. Розрізняють верхній та нижній квартилі. Іноді обчислюють і децилі, тобто такі варіанти, які ділять варіаційний ряд на 10 рівних за обсягом груп, перцентілі, ділять варіаційний ряд на 100 рівних частин, і т. д.

*Квартилі* — значення ознаки в ранжованому ряду розподілу, що поділяють його на чотири рівні за обсягом частини. Виділяють верхній і нижній квартилі.

*Квартилі нижній і верхній* — варіанти, що відокремлюють знизу і зверху в порядку зростання ознаки по одній четвертій частині сукупності. При відношенні обсягу підгруп як $\frac{1}{4}$ до $\frac{3}{4}$ отримуємо нижній квартиль ($Q_1$), а при відношенні обсягів підгруп до — верхній квартиль ($Q_3$). Квартилі визначаються за формулами, аналогічними наведеною вище формулою для розрахунку медіани.

Для розрахунків квартилів інтервального варіаційного ряду використовуються формули:

$$Q_1 = x_{Q_{1\min}} + k \cdot \frac{\frac{1}{4}\sum f - V_{Q_1 - 1}}{f_{Q_1}}; \quad (8.31)$$

$$Q_3 = x_{Q_{3\min}} + k \cdot \frac{\frac{3}{4}\sum f - V_{Q_3 - 1}}{f_{Q_3}}, \quad (8.32)$$

де $x_{Q_{1\min}}$ — мінімальна межа інтервалу, що містить нижній квартиль (визначається за нагромадженими частотами);

$x_{Q_{3\min}}$ — то же, для верхнього квартиля;

$k$ — інтервальна різниця або величина інтервалу;

$V_{Q_1 - 1}$ — нагромаджена частота інтервалу, що передує нижньому квартилю;

$V_{Q_3-1}$ — то же, для верхнього квартиля;

$f_{Q_1}$ — частота інтервалу, що містить нижній квартиль;

$f_{Q_3}$ — то же, для верхнього квартиля частка одиниць у генеральній сукупності, які не мають досліджуваної ознаки.

Медіану можна як другий квартиль, тобто $Q_2$ .

З підсумку колонки 2 табл. 7.27 видно, що чисельність сукупності ряду дорівнює 30. Отже, нижній квартиль відповідає 25. З колонки нагромаджених частостей (колонка 3) видно, що нижній квартиль міститься в інтервалі 2,4–2,8, так як перша з нагромаджених частостей , що перевищують 25, — це нагромадження частостей даного інтервалу, що дорівнює 26,7%. Отже, $x_{Q_{1\min}}$= 2,4; $V_{Q_1-1}$=5; $f_{Q_1}$ =5; $k$=0,4.

Знаходимо нижній квартиль:

$$Q_1 = 2{,}4 + 0{,}4 \cdot \frac{\frac{1}{4} \cdot 30 - 5}{5} = 2{,}60 \text{ грн.}$$

Знаходимо другий квартиль (медіану). Вихідні дані: $x_{Q_{2\min}}$ 2,8; $V_{Q_2-1}$=10; $f_{Q_2}$ =7; $k$=0,4. Звідси

$$Q_2 = 2{,}8 + 0{,}4 \cdot \frac{\frac{30}{2} - 10}{7} = 3{,}09 \text{ грн;}$$

Верхній квартиль відповідає 75 і міститься в інтервалі 3,2–3,6, так як перша з накопичених частостей, що перевищує 75, дорівнює 90 і відповідає цьому інтервалу. Вихідні дані: $x_{Q_{3\min}}$ =3,2; $V_{Q_3-1}$=17; $f_{Q_3}$=10; $k$=0,4.

Знаходимо верхній квартиль:

$$Q_3 = 3{,}2 + 0{,}4 \cdot \frac{\frac{3}{4} \cdot 30 - 17}{10} = 3{,}42 \text{ грн.}$$

Квартильне відхилення дорівнює:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{3{,}42 - 2{,}60}{2} = 0{,}41 \text{ грн.}$$

Так як на квартильне відхилення не впливають відхилення всіх значень ознаки, то його не можна вважати точною мірою варіації.

Обчислення верхніх та нижніх децилів нічим не відрізняється від обчислень медіани та квартилів. При відношеннях обсягів груп $\frac{1}{10}$ до $\frac{9}{10}$, $\frac{2}{10}$ до $\frac{8}{10}$ і т. д. отримуємо децилі. Так, перший і другий децилі можуть бути обчислені за формулою:

$$Д_1 = x_{Д_{1\min}} + k \cdot \frac{\frac{1}{10}\sum f - V_{Д_1 - 1}}{f_{Д_1}}; \tag{8.33}$$

$$Д_2 = x_{Д_{2\min}} + k \cdot \frac{\frac{2}{10}\sum f - V_{Д_2 - 1}}{f_{Д_2}} \tag{8.34}$$

і т. д.

Замість частот при обчисленні квартилів і децилей можна використовувати частості.

**Найважливіші математичні властивості дисперсії**. Середній квадрат відхилень $\sigma^2$ (дисперсія) володіє рядом математичних властивостей, деякі з них дозволяють спростити її обчислення.

Властивість 1. *Дисперсія постійної величини дорівнює нулю*:

$$\sigma_A^2 = 0, \tag{8.35}$$

де $A$ — постійна величина;

$\sigma_A^2$ — дисперсія постійної величини.

Властивість 2. *Якщо значення варіантів ознаки x зменшити на постійну величину* ($x_0$), *то середній квадрат відхилень (дисперсія) не зміниться*.

$$\sigma_{(x-x_0)}^2 = \sigma^2. \tag{8.36}$$

Доведення для незваженої дисперсії.

Маємо: $x_1; x_2; x_3; \ldots; x_n$ зі середньою $\bar{x}$ і $x_1 - x_0 = x_1'$; $x_2 - x_0 = x_2'$; … $x_n - x_0 = x_n'$ із середньою

$$\overline{x'} = \frac{\sum x'}{n} = \frac{\sum (x - x_0)}{n} = \frac{\sum x}{n} - \frac{\sum x_0}{n} = \bar{x} - x_0. \tag{8.37}$$

Тоді

$$\sigma_{(x-x_0)}^2 = \frac{\sum (x' - \overline{x'})^2}{n} = \frac{\sum \left[(x - x_0) - (\bar{x} - x_0)\right]^2}{n} =$$

$$\frac{\sum(x-x_0-\overline{x}+x_0)^2}{n}=\frac{\sum(x-\overline{x})^2}{n}=\sigma_x^2. \qquad (8.38)$$

Фактично це означає, що дисперсію можна обчислювати за наявними варіантами, а, по відхиленням їх від деякого постійного числа. Це відхилення будуть значно меншими порівняно із самими варіантами, що дозволяє значно спростити розрахунки.

Властивість 3. *Якщо всі варіанти значень ознаки зменшити в одне й теж саме число разів* ($k$ разів), *то дисперсія зменшиться в* $k^2$ *раз*.

$$\sigma^2_{\left(\frac{x}{x_0}\right)}=\sigma^2:x_0^2. \qquad (8.39)$$

Отже, всі варіанти можна розділити на будь-яке постійне число (скажімо, на інтервал ряду), обчислити середнє квадратичне відхилення, а потім помножити його на це постійне число:

$$\sigma=\sigma_{\left(\frac{x}{x_0}\right)}\cdot x_0. \qquad (8.40)$$

Доведення (див. робота Г. С. Кільдишева, Т. В. Рябушкіна).

Позначимо $\frac{x}{k}=x'$. Тоді

$$\sigma_k^2=\frac{1}{n}\sum\left[\frac{1}{k}x-\left(\overline{\frac{1}{k}\quad x}\right)\right]^2. \qquad (8.41)$$

Відповідно до властивості середньої

$$\left(\overline{\frac{1}{k}x}\right)=\frac{1}{k}\overline{x}. \qquad (8.42)$$

Отже

$$\sigma_k^2=\frac{1}{n}\sum\left(\frac{1}{k}x-\frac{1}{k}\overline{x}\right)^2=\frac{1}{k^2n}\sum(x-\overline{x})^2=$$

$$=\frac{\frac{1}{k^2}\sum(x-\overline{x})^2}{n}=\frac{1}{k^2}\sigma_x^2. \qquad (8.43)$$

Звідси:

$$\sigma^2 = k^2 \sigma_k^2. \tag{8.44}$$

Властивість 4. *Якщо віднімати середній квадрат відхилень від будь-якої величини* ($A$), *що тією чи іншою мірою відрізняється від середньої арифметичної* ($\bar{x}$), *то він завжди буде більшим за середній квадрат відхилень, обчисленого від середньої арифметичної*:

$$\sigma_A^2 < \sigma^2 . \tag{8.45}$$

При цьому більше на певну величину — на квадрат різниці між середньою і цією умовно взятою величиною, тобто на $(\bar{x} - A)^2$.

Все це можна записати так:

$$\sigma_A^2 = \sigma^2 + (\bar{x} - A)^2 \quad \text{або} \quad \sigma^2 = \frac{\sum (x - A)^2 f}{\sum f} - (\bar{x} - A)^2, \tag{8.46}$$

де $\sigma^2$ — середній квадрат відхилень від середньої арифметичної ($\bar{x}$);

$\sigma_A^2$ — середній квадрат відхилень від довільної величини ($A$).

Значить, дисперсія від середньої завжди буде меншою дисперсією, обчисленою від будь-яких інших величин. Ця властивість показників дисперсії називається *властивістю мінімальності*. Перевіримо викладене правило на прикладі.

Обчислимо відхилення не від середньої арифметичної (3,23), як від довільної (3,8). Тоді, згідно з записаною формулою, середній квадрат відхилень від 3,8 дорівнюватиме $0{,}6291+(3{,}23–3{,}8)^2= =0{,}6291+ +0{,}3249=0{,}954$.

Той самий результат дає безпосередній розрахунок (див. табл. 8.31).

Таблиця 8.31

**Обчислення дисперсії ($\sigma_A^2$) при $A$=3,8**

| Денна заробітна плата, тис. грн ($x$) | Число робітників ($f$) | Центр інтервалу $x'$ | $x' - 3{,}8$ | $(x' - 3{,}8)^2$ | $(x' - 3{,}8)^2 f$ |
|---|---|---|---|---|---|
| 1,7–2,3 | 3 | 2,0 | +1,8 | 3,24 | 9,72 |
| 2,3–2,9 | 4 | 2,6 | –1,2 | 1,44 | 5,76 |
| 2,9–3,5 | 6 | 3,2 | –0,6 | 0,36 | 2,16 |
| 3,5–4,1 | 3 | 3,8 | 0 | – | – |
| 4,1–4,7 | 4 | 4,4 | +0,6 | 0,36 | 1,44 |
| Всього | 20 | – | – | – | 19,08 |

$$\sigma_A^2 = \frac{19{,}08}{20} = 0{,}954.$$

Можна використати цей розрахунок для обчислення показника дисперсії ($\sigma^2$). Для цього треба відняти квадрат різниці між $x_0$ і $\bar{x}$, тобто $(3{,}8 - 3{,}23)^2 = 0{,}57^2 = 0{,}3249$.

Отже, $\sigma^2 = 0{,}954 - (0{,}57)^2 = 0{,}954 - 0{,}3249 = 0{,}6291$.

У разі коли $x_0$ дорівнює нулю і, отже, не обчислюються відхилення, формула набуває такого вигляду:

$$\sigma^2 = \frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2, \tag{8.47}$$

або $\sigma^2 = \overline{x^2} - \bar{x}^2$.

Отже, середній квадрат відхилень дорівнює середньому квадрату значень ознаки мінус квадрат середнього значення ознаки.

Приведемо приклад розрахунку. Висхідні дані наведені в табл. 8.32.

Таблиця 8.32

**Обчислення дисперсії за формулою $\sigma^2 = \overline{x^2} - \bar{x}^2$**

| Центр інтервалу $x$ | Число робітників $(f)$ | $x^2$ | $x^2 f$ |
|---|---|---|---|
| 2,0 | 3 | 4 | 12 |
| 2,6 | 4 | 6,76 | 27,04 |
| 3,2 | 6 | 10,24 | 61,44 |
| 3,8 | 3 | 14,44 | 43,32 |
| 4,4 | 4 | 19,36 | 77,44 |
| | 20 | – | 221,24 |

$$\overline{x^2} = \frac{221{,}24}{20} = 11{,}062;$$

$$\bar{x}^2 = (3{,}23)^2 = 10{,}4329;$$

$$\sigma^2 = \overline{x^2} - \bar{x}^2 = 11{,}062 - 10{,}4329 = 0{,}6291.$$

Цим способом розрахунку дисперсії широко користуються на практиці.

На наведених вище властивостях дисперсії ґрунтуються способи розрахунки, що спрощують обчислення середнього квадратичного відхилення.

**Обчислення дисперсії та середнього квадратичного відхилення способом відліку від умовного нуля**. У більшості випадків розрахунок середнього квадратичного відхилення та дисперсії представляє собою дуже трудомістку операцію. Цей розрахунок можна значно спростити, якщо застосувати той самий спосіб відліку від умовного нуля, інакше кажучи, спосіб моментів, який застосовувався при обчисленні середньої арифметичної. Він зазвичай застосовується для рядів розподілу з рівними інтервалами. Обчислення дисперсії відліку від умовного нуля проводять за наступною формулою:

$$\sigma^2 = \frac{\sum \left( \frac{x - x_0}{k} \right)^2 f}{\sum f} k^2 - (\bar{x} - x_0)^2, \qquad (8.48)$$

де $k$ — інтервальна різниця.

Із останньої формули при $x_0$=0 і $k$=1 виводиться ще одна форма дисперсії:

$$\sigma^2 = \overline{x^2} - (\bar{x})^2, \qquad (8.49)$$

де $\overline{x^2}$ — середня з квадратів варіантів;

$(\bar{x})^2$ — квадрат середньої.

Для розрахунку дисперсії методом відліку від умовного нуля скористаємося вихідними даними нашого прикладу та обчислення будемо вести у табличній формі (табл. 8.33).

Таблиця 8.33

**Обчислення дисперсії ($\sigma^2$) способом від умовного нуля**

| Центр інтервалі $x$ | Число робітників $f$ | $x - x_0$ | $\frac{x - x_0}{k} = x'$ | $x'f$ | $x'^2 f$ |
|---|---|---|---|---|---|
| 2,0 | 3 | –1,2 | –2 | –6 | 12 |
| 2,6 | 4 | –0,6 | –1 | –4 | 4 |
| 3,2 | 6 | 0 | 0 | 0 | – |
| 3,8 | 3 | 0,6 | 1 | 3 | 3 |
| 4,4 | 4 | 1,2 | 2 | 8 | 16 |
| | 20 | | $\left.\begin{matrix}-10\\+11\end{matrix}\right\}1$ | 1 | 35 |

Приймаємо $x_0$=3,2; $k$=0,6, тоді $\overline{x'} = \frac{1}{20} = 0{,}05;$

$$\bar{x} = \bar{x}'k + x_0 = 0{,}05 \cdot 0{,}6 + 3{,}2 = 3{,}23;$$

$$\sigma^2 = \frac{\sum \left( \frac{x - x_0}{k} \right)^2 f}{\sum f} k^2 - (\bar{x} - x_0)^2 = \frac{35}{20} \cdot 0{,}6^2 - (3{,}23 - 3{,}2)^2 = 0{,}6291.$$

Результат збігається з дисперсією, одержаною за цими даними у прикладі (табл. 8.27).

**Обчислення дисперсії та середнього квадратичного відхилення способом моментів**. Розрахувати дисперсію можна і за способом моментів. Для цього визначаються перший ($m_1$) та другий ($m_2$) моменти, а потім дисперсію розраховують за формулою:

$$\sigma^2 = k^2 (m_2 - m_1^2). \qquad (8.50)$$

$$m_2 = \frac{\sum \left( \frac{x - A}{k} \right)^2 f}{\sum f}; \quad m_1 = \frac{\sum \left( \frac{x - A}{k} \right) f}{\sum f}; \qquad (8.51)$$

Покажемо це на прикладі.

$A$=3,2; $k$=0,6 (інтервал).

На основі вихідних даних обчислимо момент першого порядку:

Таблиця 8.34

**Розрахунок дисперсії ($\sigma^2$) способом моментів**

| Середина інтервалів $x$ | Число робітників ($f$) | $\frac{x - 3{,}2}{0{,}6} = x'$ | $x'f$ | $x'^2 f$ |
|---|---|---|---|---|
| 2,0 | 3 | –2 | –6 | 12 |
| 2,6 | 4 | –1 | –4 | 4 |
| 3,2 | 6 | 0 | 0 | 0 |
| 3,8 | 3 | +1 | +3 | 3 |
| 4,4 | 4 | +2 | +8 | 16 |
| | 20 | | $\left.\begin{matrix}-10\\+11\end{matrix}\right\} 1$ | 35 |

$$m_1 = \frac{\sum x'f}{\sum f} = \frac{1}{20} = 0{,}05.$$

Для розрахунку $\sigma^2$ необхідно розрахувати момент другого порядку ($m_2$). Для цього потрібно $x'$ звести до квадрата, а потім помножити на частоту ($f$). Розділивши цю суму на число частот, отримаємо момент другого порядку:

$$m_2 = \frac{\sum x'^2 f}{\sum f} = \frac{35}{20} = 1,75.$$

Дисперсія, обчислена за способом моментів, дорівнює квадрату величини інтервалу, помноженому на різницю моменту другого порядку і квадрата моменту першого порядку:

$$\sigma^2 = k^2(m_2 - m_1^2). \tag{8.52}$$

У нашому прикладі квадрат першого моменту дорівнює $m_1^2 = 0,05^2$ =0,0025. Отже, $\sigma^2 = 0,6^2$ (1,75–0,00250)=0,36·1,7475=0,6291.

Середнє квадратичне відхилення, обчислене способом моментів, дорівнює величині інтервалу, помноженої на корінь квадратний з різниці моменту другого порядку і квадрата моменту першого порядку:

$$\sigma = i\sqrt{(m_2 - m_1^2)}. \tag{8.53}$$

У нашому прикладі воно буде рівним:

$$\sigma = 0,6\sqrt{1,75 - (0,05)^2} = 0,6\sqrt{1,75 - 0,0025} =$$
$$= 0,6\sqrt{1,7475} = 0,6 \cdot 1,13293 = 0,7932.$$

Покажемо розрахунок $\overline{x'}$ та $\sigma$ способом моментів ще на одному прикладі. Припустимо, що розподіл річного надою молока на молочній фермі сільськогосподарського підприємства характеризується такими даними (табл. 8.35).

Таблиця 8.35

| Річний удій молока, л | Кількість корів |
|---|---|
| До 1 000 | 4 |
| 1 000–2 000 | 16 |
| 2 000–3 000 | 41 |
| 3 000–4 000 | 26 |
| 4 000–5 000 | 13 |
| Всього | 100 |

Обчислюємо середній удій, дисперсію, середнє квадратичне відхилення та коефіцієнт варіації.

Виберемо в якості умовної середню центральну варіанту ($A$=2 500). Всі відхилення варіант ряду від цієї умовної середньої скоротити на величину інтервалу ($k$ =1 000). Сума таких відхилень, взятих у першій ступені і помножена на частоти, дорівнюватиме 28, а

Таблиця 8.36

**Розрахунок дисперсії ($\sigma$) способом моментів**

| Річний удій молока, л | Число корів ($f$) | Середина інтервалу ($x$) | $\frac{x-A}{k}$ | $\left(\frac{x-A}{k}\right)f$ | $\left(\frac{x-A}{k}\right)^2 f$ |
|---|---|---|---|---|---|
| До 1 000 | 4 | 500 | –2 | –8 | 16 |
| 1 000–2 000 | 16 | 1 500 | –1 | –16 | 16 |
| 2 000–3 000 | 41 | 2 500 | 0 | – | – |
| 3 000–4 000 | 26 | 3 500 | 1 | 26 | 26 |
| 4 000–5 000 | 13 | 4 500 | 2 | 26 | 52 |
| | 100 | – | – | 28 | 110 |

сума відхилень, взятих у квадраті і теж помножена на частоти, складе 110.

Середнє квадратичне відхилення, отже, дорівнюватиме:

$$\sigma = i\sqrt{(m_2 - m_1^2)} = 1000\sqrt{\frac{110}{100} - \left(\frac{28}{100}\right)^2} = 1000\sqrt{1{,}1 - 0{,}0784} =$$

$$= 1\,000\sqrt{1{,}0216} = 1\,000 \cdot 1{,}0174 = 1\,010{,}7.$$

Визначимо далі величину середньої арифметичної:

$$\bar{x} = A + im_1 = 2\,500 + 1\,000\left(\frac{28}{100}\right) = 2\,780 \text{ л.}$$

Коефіцієнт варіації складе:

$$V_\sigma = \frac{\sigma \cdot 100}{\bar{x}} = \frac{1010{,}7 \cdot 100}{2780} = 36{,}4\ \%.$$

Показники варіації набули широкого поширення як в теорії статистики, так і в практичній діяльності.

## 8.7. Види дисперсій та правило їх складання

Внутрішньогрупова та міжгрупова дисперсія. Якщо сукупність розбита на групи (або частини) за ознакою, що вивчається, то для такої сукупності можуть бути обчислені чотири дисперсій: 1) загальна дисперсія; 2) групова дисперсія; 3) середня з групових дисперсій; 4) міжгрупова дисперсія.

*Загальна дисперсія* — дисперсія ($\sigma^2$), обчислена для всієї статистичної сукупності в цілому і дорівнює середньому квадрату відхилень окремих значень ознаки $x$ від загальної середньої $\bar{x}$. Залежно від вихідних даних дисперсія може бути обчислена за середньою арифметичною простою або зваженою:

$$\sigma^2 = \frac{\sum (x-\bar{x})^2}{n}$$ — дисперсія незважена (проста); (8.54)

$$\sigma^2 = \frac{\sum (x-\bar{x})^2 f}{\sum f}$$ — дисперсія зважена; (8.55)

де $x$ — значення ознаки;
$\bar{x}$ — їх загальна середня;
$f$ — вага.

Загальна дисперсія вимірює ступінь коливання залежної ознаки (її називають результативною ознакою), її варіацію, що породжується всією сукупністю діючих на нього факторів.

Якщо сукупність розбита на групи, то кожної групи може бути обчислена своя дисперсія, так звана групова. Оскільки дисперсія є середній квадрат відхилень варіантів ознаки від їх середньої величини, значить і групова дисперсія це також середній квадрат відхилень.

*Групова дисперсія* дорівнює середньому квадрату відхилень значень ознаки одиниць сукупності в групі, що є складовою частиною статистичної сукупності, від їх середньої величини (середньої арифметичної). Вона може бути обчислена за такими основними формулами:

середньої арифметичної незваженої —

$$\sigma_j^2 = \frac{\sum (x_{ij} - \bar{x}_j)^2}{n_j}; \tag{8.56}$$

середньої арифметичної зваженої —

$$\sigma_j^2 = \frac{\sum (x_{ij} - \bar{x}_j)^2 f_j}{\sum f_j}, \tag{8.57}$$

де $\sigma_j^2$ — групова дисперсія;
$x_{ij}$ — значення (варіант) ознаки $i$-одиниці j-ї групи;
$\bar{x}_j$ — групова середня величина ознаки в $j$-й сукупності;
$f_j$ — вага;
$n_j$ — чисельність одиниць $j$-й сукупності.

Ця дисперсія характеризує варіацію ознаки в межах групи, обумовлену дією на нього всіх інших факторів, крім покладеного в підставі групування (групувальної ознаки). Щоб виміряти таку варіацію для сукупності в цілому, необхідно знайти середню з групових дисперсій. З окремих, тобто внутрішньогрупових дисперсій може бути знайдена середня.

*Дисперсія внутрішньо групова* (*дисперсія, середня з групових дисперсій*) — це дисперсія, обчислена як середня арифметична, зважена з дисперсій, розрахованих за кожною групою, на які розбита статистична сукупність. Визначається за формулою:

$$\overline{\sigma_j^2} = \frac{\sum \sigma_j^2 f_j}{\sum f_j}, \tag{8.58}$$

де $\overline{\sigma_j^2}$ — групова середня.

Середня з групових дисперсій вимірює ступінь коливання (варіацію) ознаки у всій сукупності в цілому за рахунок дії на неї всіх інших факторів (ознак), крім покладеного в основу групування, тобто групувальної ознаки.

Групові середні, як правило, відрізняються одна від одної і від загальної середньої, тобто коливаються. Їх варіацію називають міжгруповою варіацією. Для її характеристики обчислюють міжгрупову дисперсію.

*Міжгрупова дисперсія* дорівнює середньому квадрату відхилення середніх величин ознаки в кожній групі (середніх групових), від середньої загальної для всієї статистичної сукупності в цілому. Міжгрупова дисперсія ($\delta^2$) обчислюється за такою формулою:

$$\delta^2 = \frac{\sum (x_j - \overline{x})^2 f_j}{\sum f_j}. \tag{8.59}$$

Міжгрупова дисперсія, або, інакше, дисперсія групових середніх, вимірює ступінь коливання (варіацію) ознаки в сукупності за рахунок фактора, покладеного в основу групування (групувальної ознаки).

Розглянемо приклад розрахунку дисперсій. Припустимо, на сортовипробувальній сільськогосподарській станції були проведені посіви нового сорту пшениці на 100 ділянках (в м$^2$ кожна). Дані про розподіл числа ділянок озимої пшениці за урожайністю наведені в табл. 8.37.

Обчислимо середню урожайність, дисперсію, середнє квадратичне відхилення (табл. 8.38). Для обчислення групових і міжгрупової дисперсії можна застосовувати будь-який з розглянутих вище спосо-

Таблиця 8.37

| Урожайність (в г з $м^2$) | 100–110 | 110–120 | 120–130 | 130–140 | 140–150 | 150–160 | 160–170 | 170–180 | 180–190 |
|---|---|---|---|---|---|---|---|---|---|
| Число ділянок (в $м^2$) | 2 | 4 | 10 | 13 | 15 | 29 | 16 | 8 | 3 |

бів обчислення середнього квадрата відхилень (відповідно до наявних даних).

Таблиця 8.38

**Розрахунок $\bar{x}$ і $\sigma$ способом моментів**

| Вага урожаю (в г з 1 $м^2$) | Варіанти ($x$) | Число ділянок (частини) ($f$) | Умовне відхилення (відхилення варіантів від 145, поділене на величину інтервалу) ($x'$) | $x'f$ | $x'^2 f$ |
|---|---|---|---|---|---|
| 100–110 | 105 | 2 | –4 | –8 | 32 |
| 110–120 | 115 | 4 | –3 | –12 | 36 |
| 120–130 | 125 | 10 | –2 | –20 | 40 |
| 130–140 | 135 | 13 | –1 | –13 | 13 |
| 140–150 | 145 | 15 | 0 | – | – |
| 150–160 | 155 | 29 | 1 | 29 | 29 |
| 160–170 | 165 | 16 | 2 | 32 | 64 |
| 170–180 | 175 | 8 | 3 | 24 | 72 |
| 180–190 | 185 | 3 | 4 | 12 | 48 |
| Всього | – | 100 | – | 44 | 334 |

Виберемо як умовну середню центральну варіанту ($A$=145). Всі відхилення варіант ряду від цієї умовної середньої поділимо на 10 (величина інтервалу). Суму таких відхилень, взятих у першому ступені і помножених на відповідні частоти, дорівнюватиме 44, а сума відхилень, взятих у квадраті і теж помножених на частоти, складе 334.

Середнє квадратичне відхилення становитиме:

$$\sigma = i\sqrt{m_2 - m_1^2} = 10\sqrt{\frac{334}{100} - \left(\frac{44}{100}\right)^2} = 10\sqrt{0{,}34 - (0{,}44)^2} =$$

$$= 10\sqrt{3{,}1464} = 10 \times 1{,}774 = 17{,}7 \text{ г}.$$

Далі знаходимо середню арифметичну. Приймаємо $x_0$= 145; $k$=10, тоді $x' = \frac{44}{100}$ =0,44. Звідси: $\bar{x} = \bar{x}'k + x_0$ = 0,44·10+145=149,4 *г*.

Це загальна середня. Загальну дисперсію знаходимо за формулою:

$$\sigma^2 = \frac{\sum\left(\frac{x - x_0}{k}\right)^2 f}{\sum f}k^2 - (\bar{x} - x_0)^2 = \frac{334}{100}\cdot 10^2 - (149{,}4 - 145)^2 =$$

$$= 334 - 19{,}36 = 314{,}64.$$

Отже, середнє квадратичне відхилення

$$\sigma = \sqrt{314{,}64} = 17{,}7 \text{ г с } 1 \text{ м}^2.$$

І коефіцієнт варіації

$$V = \frac{\sigma \cdot 100}{\bar{x}} = \frac{17{,}7 \cdot 100}{149{,}4} = 11{,}8\ \%.$$

**Внутрішньогрупова та міжгрупова варіація**. Варіація ознаки визначається різними факторами. В даному випадку урожайність озимої пшениці залежить від багатьох факторів (якості зерна, родючості ґрунту, догляд за посівами тощо). Загальна дисперсія у разі вимірює коливання урожайності з допомогою всіх цих чинників.

Припустимо, що 100 ділянок знаходилося на двох різних масивах: I група — посівні площі, на яких не проводився полив 42 ділянок), II — площі, що піддавалися поливу (58 ділянок).

Ці два масиви ми розглядали разом і обчислювали для них загальну середню та загальну дисперсію. Обчислимо тепер ці показники окремо за масивами (за групами), складемо таблицю (табл. 8.39).

Розрахунок середньої та дисперсії для першого масиву (*без поливу*):

$$m_1 = \frac{-36}{42} = -0{,}857; \qquad m_1^2 = 0{,}734;$$

Приймаючи $A$=145; $k$=10, отримаємо:

$$\bar{x} = A + im_1 = 145 + 10(-0{,}857) = 145 - 8{,}57 = 136{,}43;$$

$$m_2 = \frac{138}{42} = 3{,}286;$$

$$\sigma_1^2 = i^2(m_2 - m_1^2) = 10^2(3{,}286 - 0{,}734) = 100 \times 2{,}552 = 255{,}2.$$

Таблиця 8.39

**Розрахунок групових середніх урожайності та їх дисперсій**

| Середня вага урожаю (в г з 1 $м^2$) | Число ділянок (частоти) (*f*) | В тому числі | | $\frac{x-145}{10}$ ($x'$) | Розрахунок для 1-го масиву | | Розрахунок для 2-го масиву | |
|---|---|---|---|---|---|---|---|---|
| | | на 1-му масиві | на 2-му масиві | | $x'f_1$ | $x'^2 f_1$ | $x'f_2$ | $x'^2 f_2$ |
| 105 | 2 | 2 | – | –4 | –8 | 32 | – | – |
| 115 | 4 | 4 | – | –3 | –12 | 36 | – | – |
| 125 | 10 | 9 | 1 | –2 | –18 | 36 | –2 | 4 |
| 135 | 13 | 11 | 2 | –1 | –11 | 11 | –2 | 2 |
| 145 | 15 | 7 | 8 | 0 | 0 | 0 | 0 | 0 |
| 155 | 29 | 6 | 23 | 1 | 6 | 6 | 23 | 23 |
| 165 | 16 | 2 | 14 | 2 | 4 | 8 | 28 | 56 |
| 175 | 8 | 1 | 7 | 3 | 3 | 9 | 21 | 63 |
| 185 | 3 | – | 3 | 4 | – | – | 12 | 48 |
| Всього | 100 | 42 | 58 | | –36 | 138 | 80 | 196 |

$$\sigma = i\sqrt{m_2 - m_1^2} = 10\sqrt{3{,}286 - 0{,}734} = 16{,}0.$$

$$V_\sigma = \frac{16{,}0 \cdot 100}{136{,}43} = 11{,}7\ \%.$$

Розрахунок середньої та дисперсії для другого масиву (*що піддавалися поливу*):

$$m_1 = \frac{80}{58} = 1{,}379; \qquad m_1^2 = 1{,}902;$$

Приймаючи $A$=145; $k$=10, отримаємо:

$$\bar{x} = A + im_1 = 145 + 10(1{,}379) = 145 + 13{,}79 = 158{,}79;$$

$$m_2 = \frac{196}{58} = 3{,}379;$$

$$\sigma_2^2 = i^2(m_2 - m_1^2) = 10^2(3{,}379 - 1{,}902) = 100 \times 1{,}477 = 147{,}7.$$

$$\sigma = i\sqrt{m_2 - m_1^2} = 10\sqrt{3{,}379 - 1{,}902} = 12{,}2.$$

$$V_\sigma = \frac{12{,}2 \cdot 100}{158{,}79} = 7{,}7\ \%.$$

Розглянемо результати розрахунків разом із даними за всіма ділянками (табл. 8.40).

Таблиця 8.40

**Розрахунок $\bar{x}$ і $\sigma$ способом моментів**

| | $f$ | $\bar{x}$ | $\sigma^2$ | $\sigma$ | $V$ |
|---|---|---|---|---|---|
| Усі ділянки | 100 | 149,4 | 314,64 | 17,7 | 11,8 |
| Ділянки, на яких не здійснювався полив | 42 | 136,43 | 255,2 | 16,0 | 11,7 |
| Ділянки, на яких здійснювався полив | 58 | 158,79 | 147,7 | 12,2 | 7,7 |

Оскільки урожайність значною мірою залежить від поливу, то, як і слід було очікувати, групові середні суттєво відрізнялися від загальної: на ділянках, на яких не здійснювався полив, урожайність була нижчою за середню, а на ділянках з поливом — вищою за середню.

Водночас на варіацію урожайності впливає не лише фактор поливу, а й інші фактори. Мірою впливу цих факторів на загальну варіацію будуть показники внутрішньогрупової варіації.

Узагальнюючим показником внутрішньогрупової варіації буде середня їх внутрішньогрупових дисперсій $\left(\overline{\sigma_j^2}\right)$.

Обчислимо її:

$$\overline{\sigma_j^2} = \frac{\sum \sigma_j^2 f}{\sum f} = \frac{255{,}2 \cdot 42 + 147{,}7 \cdot 58}{42 + 58} = \frac{10718{,}4 + 8566{,}6}{100} = \frac{19285}{100} = 192{,}85.$$

Як бачимо, внутрішньогрупова дисперсія визначає більше половини (точніше 61,3%) загальної варіації (314,64). Таким чином, у нашому прикладі полив відіграє істотну роль у варіації урожайності: дисперсія групових середніх, яка відображає вплив цього фактора, становить 61,3% загальної дисперсії. Решта загальної дисперсії, очевидно, обумовлена впливом групувальної ознаки. Можна виміряти і цю частину загальної варіації (міжгрупову варіацію), якщо розглядати групові середні як варіанти і виміряти їх коливання близько середньої (табл. 8.41).

Таблиця 8.41

| $\bar{x}$ | $f$ | $(\bar{x}-149{,}4)$ | $(\bar{x}-149{,}4)^2$ | $(\bar{x}-149{,}4)^2 f$ |
|---|---|---|---|---|
| 136,43 | 42 | –12,97 | 168,2209 | 7 065,28 |
| 158,79 | 28 | +9,39 | 88,1721 | 5 113,98 |
| | 100 | – | – | 12 179,26 |

Дисперсія групових середніх дасть нам узагальнену характеристику міжгрупової варіації:

$$\delta^2 = \frac{\sum(\bar{x} - 149{,}4)^2 f}{\sum f} = \frac{12179{,}26}{100} = 121{,}79.$$

**Правило складання дисперсій**. Між загальною дисперсією $\sigma^2$, середньою з внутрішньогрупової $\overline{\sigma_j^2}$ та міжгрупової дисперсією $\delta^2$ існує певне співвідношення, яке носить назву *правило складання дисперсій* (*варіацій*). Це правило у загальному вигляді свідчить, що *загальна дисперсія ознаки завжди дорівнює сумі середньої внутрішньогрупової дисперсії та дисперсії групових середніх*. Це правило можна записати так

$$\sigma^2 = \overline{\sigma_j^2} + \delta^2 . \qquad (8.60)$$

Логіка цього співвідношення проста: загальна дисперсія, що виникає під впливом всіх факторів, повинна дорівнювати сумі дисперсії, що виникає під впливом всіх інших факторів, і дисперсії, що виникає за рахунок фактора групування.

Якщо виявиться, що $\overline{\sigma_j^2}$=0, а відношення міжгрупової дисперсії до загальної дорівнює одиниці. Це означає, що ніякі інші фактори, крім ознаки, покладеної в основу групування, не впливають на результативну ознаку, що вся варіація його обумовлена дією на нього саме цієї факторної ознаки. Отже, між фактором-ознакою і результативним ознакою існує тісний зв’язок (можна припустити, що існує функціональний зв’язок).

Якщо $\delta^2$=0, то можна стверджувати, що зв’язок між досліджуваними ознаками майже або навіть повністю відсутній.

За допомогою цього правила, знаючи два види дисперсій, можна визначити третю величину:

$$\delta^2 = \sigma^2 - \overline{\sigma_j^2}; \qquad \overline{\sigma_j^2} = \sigma^2 - \delta^2. \qquad (8.61)$$

На нашому прикладі ми можемо переконатися, що склавши середню внутрішньогрупову дисперсію з дисперсією групових середніх, отримаємо загальну дисперсію:

$$\sigma^2 = \overline{\sigma_j^2} + \delta^2 = 192{,}85 + 121{,}79 = 314{,}64.$$

Результат збігається з дисперсією, обчисленою у прикладі (табл. 8.38) без розчленування варіаційного ряду на групи.

Теоретичний і практичний інтерес цього правила полягає в тому, воно дає можливість судити про силу впливу фактора, покладеного в основу угруповання, в освіті загальної варіації ознаки.

## 8.8. Дисперсія альтернативної ознаки

Поруч із кількісної варіацією ознаки часом виникає необхідність виміряти варіацію альтернативного ознаки. Варіація альтернативної ознаки виявляється у тому, що одні одиниці мають дану ознаку, інші — не мають. Наприклад, одні вироби є придатними, інші — бракованими, один верстат — працює, інший — простоює, одні робітники — пройшли підвищення кваліфікації, інші — ні, кожна доросла людина одружена чи ні і т. д. Така варіація носить назву альтернативної.

У разі такої варіації середнє відхилення можна визначити наступним чином.

Так, наприклад, розгляд випущеної продукції з точки зору її якості, тобто придатності до подальшого використання, дає альтернативну ознаку. Позначаючи наявності ознаки через 1, відсутність через 0, частку[1] одиниць, що мають ознаку, через $p$, а частку одиниць, що не володіють ними, через $q$, знаходимо спочатку середню альтернативну ознаку:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{1 \cdot p + 0 \cdot q}{p + q} = \frac{p}{p + q} = p, \tag{8.62}$$

так як $p + q = 1$.

Потім знаходимо дисперсію альтернативної ознаки[2]:

$$\sigma^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f} = \frac{(1 - p)^2 p + (0 - p)^2 q}{p + q}, \tag{8.63}$$

а так как $1 - p = q$, то

$$\sigma^2 = \frac{q^2 p + p^2 q}{p + q} = \frac{pq(p + q)}{p + q} = pq. \tag{8.64}$$

---

[1] Часткою називається відношення відібраної частини одиниць до всієї кількості одиниць у сукупності. Так, якщо $m$ — число одиниць у відібраній частині, а $N$ — число одиниць у всій сукупності, то $p$ — частка дорівнюватиме $p = \frac{m}{N}$.

[2] Частка, виражена у відсотках, називається питомою вагою.

Отже, $\sigma^2 = pq = p(1-q)$.

Значить, дисперсія альтернативної ознаки дорівнює добутку частки на доповнення її до одиниці. Якщо, наприклад, частка одиниць, що мають ознаку, дорівнює 80 %, а які не володіють ним — 20 %, то дисперсія дорівнює 0,16 (0,8·0,2). Максимальне значення твору *pq* дорівнює 0,25 (коли половина володіє даним ознакою, а інша половина — ні: 0,5·0,5).

З дисперсії альтернативної ознаки вилученням кореня знаходимо *середнє квадратичне відхилення*:

$$\sigma = \sqrt{pq} = \sqrt{p(1-p)}. \tag{8.65}$$

Приклад. Сукупність складається із 10 000 електричних лампочок. Наприклад, в результаті приймального контролю якості 40 лампочок виявилися бракованими. Знайти дисперсію ознаки та середнє квадратичне відхилення.

Таблиця 8.42

| $x$ | $w$ | $xw$ | $(x-\bar{x})$ | $(x-\bar{x})^2$ | $(x-\bar{x})^2 w$ |
|---|---|---|---|---|---|
| 1 | $p$=0,004 | 0,004 | 0,996 | 0,992016 | 0,003968064 |
| 0 | $q$=0,996 | 0,00 | –0,004 | 0,000016 | 0,000015936 |
| | $\sum w = 1{,}00$ | $\sum xw = 0{,}004$ | | | $\sum (x-\bar{x})^2 w = 0{,}003984$ |

$$\bar{x} = \frac{\sum xw}{\sum w} = \frac{0{,}004}{1{,}000} = 0{,}004;$$

$$\sigma^2 = \frac{\sum (x-\bar{x})^2 w}{\sum w} = \frac{0{,}003984}{1{,}00} = 0{,}003984\ ; \qquad \sigma = \sqrt{0{,}003984} = 0{,}0631.$$

Легко показати, що тут $\bar{x} = p$, а $\sigma^2 = pq$. Справді:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{1 \cdot p + 0 \cdot q}{p+q} = \frac{p}{p+q} = p, \tag{8.66}$$

так як $p + q = 1$.

$$\sigma^2 = \frac{\sum (x-\bar{x})^2 f}{\sum f} = \frac{(1-p)^2 p + (0-p)^2 q}{p+q}, \tag{8.67}$$

а так як $1 - p = q$, то

$$\sigma^2 = \frac{q^2 p + p^2 q}{p+q} = \frac{pq(p+q)}{p+q} = pq. \qquad (8.68)$$

Отже, $\sigma^2 = pq$.

З дисперсії альтернативної ознаки вилученням кореня знаходимо *середнє квадратичне відхилення*:

$$\sigma = \sqrt{pq}. \qquad (8.69)$$

Отже, можна було б не здійснювати всіх цих обчислень і набагато простіше отримати величини $\bar{x}$, $\sigma^2$ і $\sigma$:

$$\bar{x}=p=0{,}004; \quad \sigma^2 = pq = 0{,}004 \times 0{,}996 = 0{,}003984;$$

$$\sigma = \sqrt{0{,}003984} = 0{,}0631.$$

## 8.9. Моменти розподілу

Узагальнюючими характеристиками варіаційних рядів, що визначають характеристики розподілу, являються моменти. *Моментом розподілу* називається середня арифметична тих чи інших ступенів відхилень індивідуальних значень ознаки від певної вихідної величини. Моменти є числові характеристики розподілу ймовірностей. Спосіб моментів був розроблений математиком П. Л. Чебишевим. А. А. Марковим спосіб моментів було застосований до розгляду можливостей використання закону нормального розподілу щодо сум великого, але кінцевого числа незалежних випадкових величин.

При обчисленні середньої в якості моментів можуть бути використані частоти, частості або ймовірності.

Моментом $k$-го порядку називається середня з $k$-х ступенів відхилень ознаки $x$ від деякої постійної величини $A$. Момент розподілу в загальному вигляді дорівнює:

$$M_k = \overline{(x-A)^k} = \frac{\sum (x-A)^k f_i}{\sum f_i}, \qquad (8.70)$$

де $A$ — величина, від якої визначаються відхилення;

$k$ — ступінь відхилення (порядок моменту).

Порядок моменту визначається величиною $k$ Моменти називаються: $k$=0 — нульового порядку, при $k$=1 — першого порядку, при $k$=2 другого порядку і т. д. У статистичній практиці користуються моментами першого, другого, третього та четвертого порядку.

Залежно від цього, що приймають за постійну величину ($A$), розрізняють три виду моменту: початкові, центральні і умовні (початкові відносно $x_0$).

1. *Початкові моменти* $M_k$ отримують, коли постійна величина $A$ дорівнює нулю ($A$ = 0). Наведемо формулу розрахунку всіх початкових моментів:

$$M_k = \frac{\sum x^k \cdot f_i}{\sum f_i}. \tag{8.71}$$

Представимо початкові моменти перших чотирьох порядків. Тоді при $k$=0 отримаємо

$$M_0 = \frac{\sum x^0 \cdot f_i}{\sum f_i}$$ — начальний момент нульового порядку;

при $k$ =1

$$M_1 = \frac{\sum x \cdot f_i}{\sum f_i}$$ — начальний момент першого порядку;

при $k$ =2

$$M_2 = \frac{\sum x^2 \cdot f_i}{\sum f_i}$$ — начальний момент другого порядку;

при $k$ =3

$$M_3 = \frac{\sum x^3 \cdot f_i}{\sum f_i}$$ — начальний момент третього порядку;

при $k$ =4

$$M_4 = \frac{\sum x^4 \cdot f_i}{\sum f_i}$$ — начальний момент четвертого порядку.

Обчислимо початкові моменти перших чотирьох порядків, якщо варіанти мають як позитивні, і негативні значення.

Занесемо всі розрахунки в таблицю.

Обчислюємо моменти:

$$M_0 = 1;$$

$$M_1 = -\frac{13}{30} = -0{,}433;$$

Таблиця 8.43

| $x'$ | Частоти ($f$) | $x'f$ | $(x')^2 f$ | $(x')^3 f$ | $(x')^4 f$ |
|---|---|---|---|---|---|
| –3 | 3 | –9 | 27 | –81 | 243 |
| –2 | 5 | –10 | 20 | –40 | 80 |
| –1 | 8 | –8 | 8 | -8 | 8 |
| 0 | 6 | 0 | 0 | 0 | 0 |
| 1 | 4 | 4 | 4 | 4 | 4 |
| 2 | 2 | 4 | 8 | 16 | 32 |
| 3 | 2 | 6 | 18 | 54 | 162 |
| Всього | 30 | –13 | 85 | –55 | 529 |

$$M_2 = \frac{85}{30} = 2,8333;$$

$$M_3 = -\frac{55}{30} = -1,83333;$$

$$M_4 = \frac{529}{30} = 17,63333.$$

2. *Центральні моменти* $\mu_k$ отримують, коли за постійну величину $A$ приймаю середню $\bar{x}$ ($A=\bar{x}$):

$$\mu_k = \overline{(x-\bar{x})^k} = \frac{\sum (x_i - \bar{x})^k \cdot f_i}{\sum f_i}. \tag{8.72}$$

При підстановці різних значень $k$ отримуємо центральні моменти відносно $\bar{x}$. Тоді:

при $k=0$ отримаємо

$$\mu_0 = \frac{\sum (x-\bar{x})^0 \cdot f_i}{\sum f_i}$$ — центральний момент нульового порядку;

при $k=1$

$$\mu_1 = \frac{\sum (x-\bar{x}) \cdot f_i}{\sum f_i}$$ — центральний момент першого порядку;

при $k=2$

$$\mu_2 = \frac{\sum (x-\bar{x})^2 \cdot f_i}{\sum f_i}$$ — центральний момент другого порядку;

при $k=3$

$$\mu_3 = \frac{\sum (x-\bar{x})^3 \cdot f_i}{\sum f_i}$$ — центральний момент третього порядку;

при $k = 4$

$$\mu_4 = \frac{\sum (x-\bar{x})^4 \cdot f_i}{\sum f_i}$$ — центральний момент четвертого порядку.

Обчислимо центральні моменти перших чотирьох порядків за даними табл. 8.44 (колонки 1, 2).

Розрахунки наведені у колонках 3–8.

Таблиця 8.44

| $x$ | Частоти ($f$) | $x \cdot f$ | $x-\bar{x}$ | $(x-\bar{x})f$ | $(x-\bar{x})^2 f$ | $(x-\bar{x})^3 f$ | $(x-\bar{x})^4 f$ |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 14 | 2 | 28 | –3,6 | –7,2 | 25,92 | –93,312 | 335,9232 |
| 16 | 6 | 96 | –1,6 | –9,6 | 15,36 | –24,576 | 39,3216 |
| 18 | 8 | 144 | +0,4 | +3,2 | 1,28 | +0,512 | 0,2048 |
| 20 | 2 | 40 | +2,4 | +4,8 | 11,52 | +27,648 | 66,3552 |
| 22 | 2 | 44 | +4,4 | +8,8 | 38,72 | +170,368 | 749,6192 |
| Всього | 20 | 352 | – | –16,8<br>+16,8<br>0 | 92,80 | +80,640 | 1 191,4240 |

Обчислимо середню арифметичну:

$$\bar{x} = \frac{\sum x \cdot f}{\sum f} = \frac{352}{20} = 17{,}6.$$

$$\mu_0 = \frac{\sum (x-\bar{x})^0 \cdot f_i}{\sum f_i} = 1;$$

$$\mu_1 = \frac{\sum (x-\bar{x}) \cdot f_i}{\sum f_i} = 0;$$

$$\mu_2 = \frac{\sum (x-\bar{x})^2 \cdot f_i}{\sum f_i} = \frac{92{,}80}{20} \approx 4{,}64;$$

$$\mu_3 = \frac{\sum (x-\bar{x})^3 \cdot f_i}{\sum f_i} = \frac{80{,}64}{20} \approx 4{,}03;$$

$$\mu_4 = \frac{\sum (x-\bar{x})^4 \cdot f_i}{\sum f_i} = \frac{1191{,}4240}{20} \approx 59{,}57.$$

Розглядаючи формули моментів, можна побачити, що центральний момент першого порядку завжди дорівнює нулю. Центральний момент другого порядку відповідає дисперсії, яка використовується при обчисленні варіації ознаки ряду розподілу. Центральний момент третього порядку характеризує симетричність ряду розподілу та використовується для розрахунку показників асиметрії. Центральний момент четвертого порядку використовується для обчислення показників ексцесу, що характеризують островершинність ряду розподілу.

З центральних моментів порядку $k > 2$ практично використовують лише моменти третього і четвертого порядку. Моменти $k > 4$ не використовуються, оскільки для їх надійного визначення необхідно мати вибірку дуже великого обсягу.

3. *Умовні моменти* $M'_k$ одержують, якщо $A$ не дорівнює середній арифметичній і відмінна від нуля, а відповідає певній довільно обраній величині ($A = x_0$):

$$M'_k = \overline{(x - x_0)^k} = \frac{\sum (x_i - x_0)^k \cdot f_i}{\sum f_i}. \tag{8.73}$$

Підставляючи різні значення $k$ отримаємо формули розрахунку умовних моментів. Так:

при $k = 0$ отримаємо

$$M'_0 = \frac{\sum (x - x_0)^0 \cdot f_i}{\sum f_i}$$ — умовний момент нульового порядку;

при $k = 1$

$$M'_1 = \frac{\sum (x - x_0) \cdot f_i}{\sum f_i}$$ — умовний момент першого порядку;

при $k = 2$

$$M'_2 = \frac{\sum (x - x_0)^2 \cdot f_i}{\sum f_i}$$ — умовний момент другого порядку;

при $k = 3$

$$M'_3 = \frac{\sum (x - x_0)^3 \cdot f_i}{\sum f_i}$$ — умовний момент третього порядку;

при $k = 4$

$$M'_0 = \frac{\sum (x - x_0)^4 \cdot f_i}{\sum f_i}$$ — умовний момент четвертого порядку.

Обчислимо умовні моменти перших чотирьох порядків за даними табл. 8.43 (колонки 1 та 2).

Таблиця 8.45

| $x$ | $f$ | $x-x_0$ | $(x-x_0)f$ | $(x-x_0)^2 f$ | $(x-x_0)^3 f$ | $(x-x_0)^4 f$ |
|---|---|---|---|---|---|---|
| 85 | 3 | –15 | –45 | 675 | –10 125 | 151 875 |
| 90 | 5 | –10 | –50 | 500 | –5 000 | 50 000 |
| 95 | 8 | –5 | –40 | 200 | –1 000 | 5 000 |
| 100 | 6 | 0 | 0 | 0 | 0 | 0 |
| 105 | 4 | 5 | 20 | 100 | 500 | 2 500 |
| 110 | 2 | 10 | 20 | 200 | 2 000 | 20 000 |
| 115 | 2 | 15 | 30 | 450 | 6 750 | 101 250 |
| Всього | 30 | – | –65 | 2 125 | –6 875 | 330 625 |

$$M'_0 = \frac{\sum (x-x_0)^0 \cdot f_i}{\sum f_i} = 1;$$

$$M'_1 = \frac{\sum (x-x_0) \cdot f_i}{\sum f_i} = -\frac{65}{30} = -2{,}17;$$

$$M'_2 = \frac{\sum (x-x_0)^2 \cdot f_i}{\sum f_i} = \frac{2\,125}{30} = 70{,}83;$$

$$M'_3 = \frac{\sum (x-x_0)^3 \cdot f_i}{\sum f_i} = -\frac{6\,875}{30} = -229{,}17;$$

$$M'_4 = \frac{\sum (x-x_0)^4 \cdot f_i}{\sum f_i} = \frac{330\,625}{30} = 11\,020{,}83.$$

Умовними моментами широко користуються для спрощення техніки розрахунків деяких статистичних показників.

*Найважливіші властивості моменту* $M'_k$:

1) $\bar{x} = M'_1 + x_0$, тобто середня арифметична дорівнює початковому моменту першого порядку щодо початку відліку плюс початок відліку. З формули умовного моменту першого порядку можна записати:
$M'_1 = \frac{\sum (x-x_0) \cdot f_i}{\sum f_i} = \frac{\sum xf - x_0 \sum f}{\sum f} = \bar{x} - x_0$. Звідси $\bar{x} = M'_1 + x_0$;

2) якщо при розрахунку $M'_k$ значення $x$ і $x_0$ мають загальний множник $C$, то на нього можна розділити всі відхилення $(x-x_0)$ на загальний множник $C$, а потім помножити на цей множник у відповід-

ному ступені. Звідси умовний момент може бути обчислений за такою формулою:

$$M'_k = \frac{\sum\left(\frac{x - x_0}{C}\right)^k \cdot f_i}{\sum f_i} \cdot C^k. \tag{8.74}$$

Зокрема, для середньої арифметичної

$$\overline{x} = x_0 + M'_1 \cdot C. \tag{8.75}$$

При порівнянні з обчисленням середньої методом відліку від нуля видно, що $\overline{x'}$ і $M'_1$ тотожні. Тому обчислення середньою методом відліку від умовного нуля іноді називають методом моментів.

Якщо статистичні дані представлені у вигляді інтервального ряду розподілу, то розрахунок моментів для такого ряду має особливості. Для інтервального ряду всі відхилення $x - x_0$ мають загальний множник, що дорівнює величині інтервалу (*C*=*i*). Тому для подальшого спрощення обчислень можна розділити відхилення на величину інтервалу, тобто визначити величину $\frac{x - x_0}{i}$.

Позначимо моменти цих величин через $x'_k$, тобто

$$x'_1 = \frac{\sum\left(\frac{x - x_0}{i}\right) f}{\sum f}; \qquad x'_2 = \frac{\sum\left(\frac{x - x_0}{i}\right)^2 f}{\sum f}; \tag{8.76}$$

$$x'_3 = \frac{\sum\left(\frac{x - x_0}{i}\right)^3 f}{\sum f}; \qquad x'_4 = \frac{\sum\left(\frac{x - x_0}{i}\right)^4 f}{\sum f}. \tag{8.77}$$

Для того, щоб отримати умовні моменти через величини $x'_k$, потрібно помножити на величину інтервалу у відповідній ступені ($i^k$), тобто

$$M'_1 = x'_1 \cdot i;$$

$$M'_2 = x'_2 \cdot i^2;$$

$$M'_3 = x'_3 \cdot i^3;$$

$$M'_4 = x'_4 \cdot i^4.$$

Наведемо приклад. Обчислимо умовні моменти відносно $x_0$=100 перших чотирьох порядків за даними колонок 1 та 2 табл. 8.46.

Візьмемо в якості довільної величини $x_0$ (початок відліку) відхи-

Таблиця 8.46

| Центр інтервалу $x$ | Частота $f$ | $x-x_0$ | $\frac{x-x_0}{i}=x'$ |
|---|---|---|---|
| 85 | 3 | –15 | –3 |
| 90 | 5 | –10 | –2 |
| 95 | 8 | –5 | –1 |
| 100 | 6 | 0 | 0 |
| 105 | 4 | 5 | 1 |
| 110 | 2 | 10 | 2 |
| 115 | 2 | 15 | 3 |
| Всього | 30 | – | – |

лень індивідуальних значень ознаки приймемо величину 100 ($x_0$=100). Обчислимо всі відхилення значень $x$ від довільної величини (колонка 3). Розділимо отримані в колонці 3 відхилення на загальний множник $C$ (величину інтервалу), який для $x$ і $x_0$ складе 5 і занесемо в колонку 4.

Для розрахунку умовних моментом використовуємо знайдені в попередньому прикладі початкові моменти, помножимо їх на $C$ ($C$=5) у відповідній ступені:

$$M'_0=1\cdot5^0=1;$$

$$M'_1=-0{,}433\cdot5^1=-2{,}17;$$

$$M'_2=2{,}8333\cdot5^2=70{,}83;$$

$$M'_3=-1{,}83333\cdot5^3=-229{,}17;$$

$$M'_4=17{,}63333\cdot5^4=11\,020{,}83.$$

Як бачимо, показники умовних моментів відповідають умовним моментам, обчисленим за формулою $M'_k=\frac{\sum(x_i-x_0)^k\cdot f_i}{\sum f_i}$.

Розрахунок умовних моментів здійснюють у наступній послідовності:

зі всіх значень варіант $x$ обчислюють деяку довільну величину $x_0$ (початок відліку) і знаходять відхилення ($x-x_0$);

ділять всі відхилення від початку відліку на загальний множник $C$: $\frac{x-x_0}{C}=x'$;

обчислюють початкові моменти для значень $x'$ ($M_k$);

початкові моменти множать на $C$ у відповідному ступені.

У статистиці обчислення моментів дозволяє спростити розрахунок окремих узагальнюючих показників. Так, застосування умовних моментів першого і другого порядку ($M_1'$ і $M_2'$) дозволяють спростити техніку розрахунку $\sigma$. Для використання умовних моментів при розрахунку $\sigma$ необхідно застосувати формулу переходу від моментів центральних до моментів умовних, яка здійснюється за наступною рівністю:

$$\sigma^2 = \mu_2 = M_2' - (M_1')^2. \tag{8.78}$$

При розрахунку середнього квадратичного відхилення ($\sigma$) під коренем знаходиться центральний момент другого порядку $\sigma = \sqrt{\mu_2}$. Таким чином, формула середнього квадратичного відхилення, виражена в умовних моментах, буде

$$\sigma = \sqrt{M_2' - (M_1')^2}. \tag{8.79}$$

**Розрахунок центральних моментів**, що використовуються для характеристики варіаційного ряду розподілу, за формулою $\mu_k = \frac{\sum (x_i - \bar{x})^k \cdot f_i}{\sum f_i}$, є дуже громіздким. Для спрощення обчислення спочатку визначають умовні моменти відносно величини $x_0$, а для знаходження центральних моментів використовують формулу переходу від умовних моментів до центральних:

$$\mu_k = M_k' - C_k^1 \cdot M_{k-1}' \cdot M_1' + C_k^2 \cdot M_{k-2}' \cdot (M_1')^2 -$$
$$- C_k^3 M_{k-3}' \cdot (M_1')^3 + \ldots + (-M_1')^k. \tag{8.80}$$

де $C_k^1, C_k^2, C_k^3 \ldots$ — числа поєднань із $k$ по 1, із $k$ по 2, із $k$ по 3 і т. д.

Користуючись цією формулою, визначимо центральні моменти першого, другого, третього та четвертого порядків:

$$\mu_0 = M_0' = 1;$$

$$\mu_1 = M_1' - M_1' = 0;$$

$$\mu_2 = M_2' - 2M_1' \cdot M_1' + (M_1')^2 = M_2' - M_1';$$

$$\mu_3 = M_3' - 3M_2' \cdot M_1' + 3(M_1')^3 - (M_1')^3 = M_3' - 3M_2'M_1' + (2M_1')^3;$$

$$\mu_4 = M_4' - 4M_3' \cdot M_1' + 6M_2'(M_1')^2 - 3(M_1')^4.$$

Використовуючи дані табл. 8.45, де розраховані умовні моменти відносно $x_0$= 100, обчислимо центральні моменти перших чотирьох порядків за відповідними формулами і звіримо отримані результати.

З прикладу маємо:

$M'_0 = 1;$ $M'_3 = -229,17;$

$M'_1 = -2,17;$ $M'_4 = 11\,020,83.$

$M'_2 = 70,83;$

За формулами центральних моментів отримуємо, використовуючи початкові моменти:

$$\mu_0 = 1;$$

$$\mu_1 = 0$$

$$\mu_2 = 70,83 - 2,17^3 = 70,83 - 4,7089 = 66,1211;$$

$$\mu_3 = -229,17 - 3 \cdot 70,83 \cdot (-2,17) + 2 \cdot (-2,17)^3 = 211,5;$$

$$\mu_4 = 11020,83 - 4(-229,17) \cdot (-2,17) + 6 \cdot 70,83 \cdot (-2,17)^2 -$$
$$-3(-2,17)^4 = 10966,3.$$

Порівняємо центральні моменти перших чотирьох порядків, обчислених за вказаними формулами, з центральними моментами, обчисленими за наведеними вище формулами.

Таблиця 8.47

| $x$ | $f$ | $x \cdot f$ | $x - \bar{x}$ | $(x - \bar{x}) f$ | $(x - \bar{x})^2 f$ | $(x - \bar{x})^3 f$ | $(x - \bar{x})^4 f$ |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 85 | 3 | 255 | –12,833 | –38,50 | 494,06 | –6 340,24 | 81 364,33 |
| 90 | 5 | 450 | –7,833 | –39,17 | 306,78 | –2 403,00 | 18 822,73 |
| 95 | 8 | 760 | –2,833 | –22,66 | 64,21 | –181,90 | 515,32 |
| 100 | 6 | 600 | 2,167 | +13,00 | 28,18 | +61,06 | 132,31 |
| 105 | 4 | 420 | 7,167 | +28,67 | 205,46 | +1 472,56 | 10 553,82 |
| 110 | 2 | 220 | 12,167 | +24,33 | 296,07 | +3 602,31 | 43 829,25 |
| 115 | 2 | 230 | 17,167 | +34,33 | 589,41 | +10 118,43 | 173 703,12 |
| Всього | 30 | 2 935 | – | –110,33<br>+110,33<br>0 | 1 984,17 | 6 329,21 | 328 920,87 |

$$\mu_0 = \frac{\sum (x - \bar{x})^0 \cdot f_i}{\sum f_i} = 1;$$

$$\mu_1 = \frac{\sum (x - \bar{x}) \cdot f_i}{\sum f_i} = 0;$$

$$\mu_2 = \frac{\sum (x - \bar{x})^2 \cdot f_i}{\sum f_i} = \frac{1984{,}17}{30} \approx 66{,}1;$$

$$\mu_3 = \frac{\sum (x - \bar{x})^3 \cdot f_i}{\sum f_i} = \frac{6329{,}21}{30} \approx 211{,}0;$$

$$\mu_4 = \frac{\sum (x - \bar{x})^4 \cdot f_i}{\sum f_i} = \frac{328920{,}87}{30} \approx 10964{,}0.$$

Порівнюючи центральні моменти перших чотирьох порядків, обчислені за вказаними формулами, з центральними моментами, обчисленими безпосередньо за формулою $\mu_k = \frac{\sum (x_i - \bar{x})^k \cdot f_i}{\sum f_i}$, переконуємося в порівняльній простоті обчислення центральних моментів за наведеними в цьому параграфі формулами. Незначна різниця вийшла в результаті округлень.

## 8.10. Аналіз варіаційних рядів. Криві розподілу

Найважливішою умовою обчислення узагальнюючих показників варіації є однорідність узагальнюючих у них сукупностей, що встановлюється на основі глибокого теоретичного аналізу. Сукупності, що досліджуються в математичній статистиці, мають суттєві відмінності від сукупностей соціально-економічних. У математичній статистиці в статистичну сукупність включаються, зазвичай, однорідні явища по типу (закономірності) розподілу одиниць сукупності (частот або частостей). При дослідженні соціально-економічних явищ у сукупність поєднуються однорідні явища за типом соціальної якості. Тип не менше в більшості випадків характер розподілу явищ у математичній статистиці збігається з розподілом явищ у соціально-економічних сукупностях. Будь-яка соціально-економічна сукупність належить до одного типу розподілу частот, хоча кожен однорідний тип розподілу частот об’єднує одну соціальну сукупність. Так, тип розподілу сімей за чисельністю дітей практично не залежить від того, яка їхня національність. Це означає, що національність громадян країни дозволяє їх віднести до певного типу (закономірності) розподілу частот, а один тип розподілу сімей за чисельністю дітей включає групи населення різної національності.

Закономірності варіаційних рядів виражаються в типі розподілу їх частот. Як відомо, варіація ознак ряду визначається взаємодією випадкових подій випадкових подій і внутрішніх тенденцій розвитку. У першому випадку у варіації ознаки відображаються відмінності, спричинені зовнішніми, випадковими подіями. У другому випадку зміна значень ознак варіаційного ряду відбувається внаслідок внутрішніх тенденцій розвитку, органічно властивих всіх суспільних явищ. Вплив випадкових подій як би «затемняє» основну закономірність зміни величини ознаки. Саме при характеристиці розподілу варіантів виявляються повторювані риси явища, які і визначають закономірності рядів розподілу.

Закономірності варіаційних рядів найнаочніше простежуються на графіках — гістограмі і полігоні розподілу частот. Їх розгляд (рис. 8.2 і 8.1) показує, що при переході від однієї групи до іншої розподіл одиниць сукупності (частот або частостей) у полігоні відбувається поступово, а в гістограмі виявляється велика стрибкоподібність.

При збільшенні чисельності інтервалів варіаційного ряду і відповідно зменшенні величини інтервалу кількість стовпців полігону розподілу буде зростати, ламана лінія полігону починає згладжуватися і перетворюється на якусь плавну криву. Така крива носить назву *кривої розподілу*.

Крива розподілу характеризує теоретичний розподіл, тобто в найбільш узагальненому вигляді показує тип варіації, закономірність розподілу частот всередині одноякісної сукупності явищ при повному погашенні випадкових величин, що «затемняють основну тенденцію». Дослідження закономірності або форми розподілу включає рішення трьох послідовних завдань: 1) з'ясування загального характеру розподілу; 2) вирівнювання емпіричного розподілу, яке полягає в тому, що на підставі емпіричного розподілу будується крива $f(x)$ із заданою формою; 3) перевірку відповідності знайденого теоретичного розподілу емпіричному.

Під час проведення статистичних досліджень доводиться зустрічатися з різними розподілами. У практиці статистичних досліджень часто застосовують насамперед *криву нормального розподілу*. Вона має форму фігури дзвона та одну вершину, права та ліва гілки якої рівномірно та симетрично спадають, асимптотично наближаючись до осі абсцис (рис. 8.10).

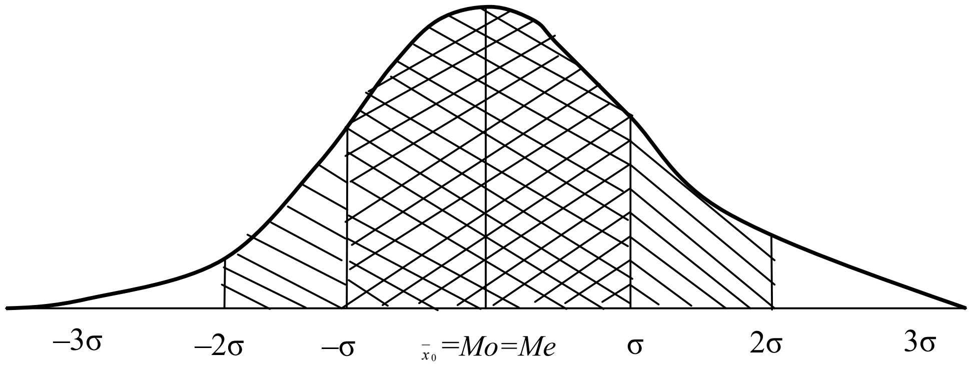



*Рис. 8.10.* Крива нормального розподілу.

Якщо всю площу між кривою та віссю абсцис прийняти за 100 %, то в межах $\bar{x} \pm \sigma$ укладено 68,3 % частот, у межах $\bar{x} \pm 2\sigma$ — 95,4 %, у межах $\bar{x} \pm 3\sigma$ — 99,7 %. Отже, при стандартному нормальному розподілі майже всі значення перебувають у межах ±3σ (правило трьох сигм).

Формула його має такий вигляд:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\bar{x})^2}{2\sigma^2}}, \tag{8.81}$$

де $y$ — ордината кривої розподілу (частість);

$x$ — випадкова величина ознаки;

$\bar{x}$ — середня арифметична чи математичне очікування;

$\sigma$ — середнє квадратичне відхилення;

$\pi$ — відома в математиці величина, що визначає відношення кола до діаметра;

$e$ — основа натуральних логарифмів.

Нормальний розподіл можливий у тому випадку, коли на величину ознаки впливає велика кількість випадкових причин. Дія цих причин незалежна, і жодна з причин не має переважаючого впливу над іншими.

Істотними факторами, що визначають центр групування та форма кривої розподілу, є $\bar{x}$ и $\sigma$. Якщо позначити вираз $\frac{(x-\bar{x})}{\sigma}$ через $t$, і при $\sigma = 1$ отримується найпростіше, канонічне рівняння нормальної кривої $y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2}}$, яку назвемо нормованою функцією.

Визначений інтеграл виду

$$\Phi(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^t e^{\frac{-t^2}{2}} du \qquad (8.82)$$

носить назву нормованої функцією Лапласа і характеризує площу під кривою в проміжку від нуля до $t$.

Нормальний розподіл, як і біномінальний розподіл, конкретно виражає закономірність масового явища, яка формується як сума багатьох незалежних доданків. При цьому кількість цих доданків має бути досить великою. Якщо варіація явища викликана дією багатьох випадкових, слабо взаємозалежних величин, то розподіл такої величини за досить великої кількості доданків має відповідати нормальному розподілу; якщо розподіл не відповідає нормальному, це означає, що у варіацію ознаки вплинули якісь невипадкові чинники. Нормальний розподіл властиво багатьом явищам, наприклад, деякі характеристики живих організмів у популяції, вимірювання параметрів виробів при контролі їх якості і т. д. Нормальний розподіл служить однією з основоположних передумов застосування методів математичної статистики, так як доведено, що результати багатьох масових емпіричних вимірювань та обчислень розподілені нормально.

**Побудова кривої нормальної розподілу**. Для побудови кривої нормального розподілу використовуються кілька методів. Розглянемо розрахунок показників кривої нормального розподілу з використанням *першого методу*. Для того, щоб побудувати криву нормального розподілу, користуються наступною формулою:

$$f(t) = \frac{N \cdot k}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \qquad (8.83)$$

де $N$ — число проведених випробувань, що дорівнює сумі частот емпіричного ряду $\sum f$;

$k$ — величина інтервалу дроблення емпіричного ряду;

$\sigma$ — середнє квадратичне відхилення ряду;

$t$ — нормоване відхилення, тобто $t = \frac{x - \overline{x}}{\sigma}$.

Величина $\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ може бути обчислена за спеціальними таблицями (див. додаток 1). Розглянемо розрахунок кривої нормального розподілу з прикладу. У табл. 8.48 наведено емпіричний розподіл ваги деталей та розрахунок частот нормального розподілу.

Таблиця 8.48

| $x$ | $f$ | $xf$ | $x^2$ | $x^2 f$ | $x-\bar{x}$ | $\frac{x-\bar{x}}{\sigma}$ | $f(t)$ | $f'$ | $f'$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 4,0 | 3 | 12 | 16 | 48 | –2,6 | –2,1 | 0,044 | 3,841 | 4 |
| 4,5 | 15 | 67,5 | 20,25 | 303,75 | –2,1 | –1,7 | 0,094 | 8,206 | 8 |
| 5,0 | 12 | 60 | 25 | 300 | –1,6 | –1,3 | 0,171 | 14,928 | 15 |
| 5,5 | 27 | 148,5 | 30,25 | 816,75 | –1,1 | –0,9 | 0,266 | 23,222 | 23 |
| 6,0 | 31 | 186 | 36 | 1 116 | –0,6 | –0,5 | 0,352 | 30,730 | 31 |
| 6,5 | 32 | 208 | 42,25 | 1 352 | –0,1 | –0,1 | 0,397 | 34,658 | 35 |
| 7,0 | 32 | 224 | 49 | 1 568 | +0,4 | +0,3 | 0,381 | 33,261 | 33 |
| 7,5 | 28 | 210 | 56,25 | 1 575 | +0,9 | +0,7 | 0,312 | 27,238 | 27 |
| 8,0 | 22 | 176 | 64 | 1 408 | +1,4 | +1,1 | 0,218 | 19,031 | 19 |
| 8,5 | 8 | 68 | 72,25 | 578 | +1,9 | +1,5 | 0,13 | 11,349 | 11 |
| 9,0 | 6 | 54 | 81 | 486 | +2,4 | +1,9 | 0,066 | 5,762 | 6 |
| 9,5 | 2 | 19 | 90,25 | 180,5 | +2,9 | +2,3 | 0,028 | 2,444 | 3 |
| 10 | 2 | 20 | 100 | 200 | +3,4 | +2,7 | 0,01 | 0,873 | 1 |
| Всього | 220 | 1 453 | 682,5 | 9 932,00 | | | | 215,544 | 216 |

Для знаходження значень теоретичних частот спочатку використовуємо середню арифметичну:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{1453}{220} = 6{,}6.$$

Потім знаходимо дисперсію, скориставшись формулою $\sigma^2 = \frac{\sum x^2 f}{\sum f} - \left(\bar{x}\right)^2$:

$$\sigma^2 = \frac{9932}{220} - (6{,}6)^2 = 1{,}59; \ \sigma = \pm 1{,}26.$$

Потім визначаємо величину $t$. З цією метою розраховуємо різницю $x - \bar{x}$ (графи 6) і $\frac{x-\bar{x}}{\sigma}$ (графи 7). З таблиці значень (додаток 1) для даних у графі 7 знаходимо відповідні величини (графа 8). Далі знаходимо множник $\frac{N \cdot k}{\sigma}$. За умовами нашого прикладу:

$$\frac{N \cdot k}{\sigma} = \frac{220 \cdot 0{,}5}{1{,}26} = 87{,}3.$$

Помножуючи на множник відповідні значення $f(t)$, отримуємо теоретичну частоту кожного варіанту (графа 9). Зважаючи на те, що частоти можуть бути тільки цілими числами, округляємо їх до цілих і отримаємо теоретичні частоти, які будемо позначати $f'$ (графа 10).

Побудуємо графік емпіричних та теоретичних значень. На рис. 8.11 суцільною лінією дано зображення емпіричного розподілу, а пунктирної — побудова його на основі теоретичного розподілу. Обидва емпіричні розподіли добре відтворюються теоретичним нормальним розподілом.

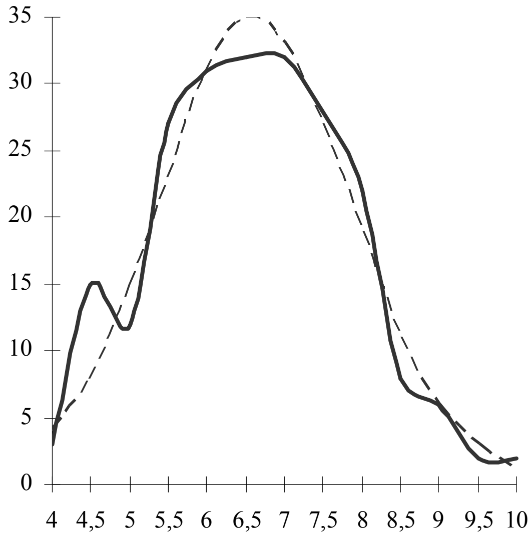



*Рис. 8.11.* Розподіл 220 замірів деталей.

Для нормального розподілу характерний нормальний або симетричний розподіл частот усередині одноякісної сукупності. У ньому відображаються тільки відмінності, викликані лише коливання сукупності одиниць. Разом з тим для багатьох суспільних явищ емпіричні частоти можуть сильно відрізнятися від нормального розподілу, оскільки зовнішні фактори часто призводять до істотних зрушень у розподілі коливання ознак одиниць сукупності. Тому ряди і криві розподілу частот суспільних явищ, як правило, істотно відрізняються від кривих нормального розподілу, в них частоти зростають до максимуму і зменшуються від нього нерівномірно.

Для однорідних сукупностей характерним є одновершинний розподіл. Багатовершинність свідчить про неоднорідність досліджуваної сукупності. Поява двох або більше вершин говорить про необхідність перегрупування даних з метою виділення більш однорідніших груп. З’ясування загального характеру розподілу передбачає оцінку ступеня його однорідності, а також обчислення асиметрії та ексцесу. *Симетричним* є розподіл, у якому частоти варіантів, рівновіддалених по обидві сторони від центру розподілу, рівні між собою. Для симетричних розподілів середня арифметична, мода та медіана рівні між собою (див. гл. 7). Якщо ж розташування варіантів навколо середньої не однаково, тобто частоти по обидва боки від середньої змінюються порізному, то варіаційний ряд називають асиметричним або скошеним.

**Коефіцієнт асиметрії**. В якості показник міри скошеності (несиметричності) варіаційного ряду застосовується простий емпіричний *коефіцієнт асиметрії* ($K_A$), що представляє собою відношення різниці між середньою арифметичною і модою до середньоквадратичного відхилення:

$$K_A = \frac{\bar{x} - Mo}{\sigma}. \quad (8.84)$$

Коефіцієнт асиметрії характеризує ступінь відхилення варіаційного ряду від симетрії, тобто міру скошеності графіка розподілу вліво або вправо від максимальної ординати. Величина показника $K_A$ може бути позитивною і негативною величиною. Чим більша абсолютна величина коефіцієнта, тим більший ступінь скошеності.

В залежності від значення коефіцієнта асиметрії, розрізняють кілька видів, що характеризують асиметрію (скошеність) рядів розподілу:

якщо $K_A$>0, то скошеність правостороння; «довга» частина розподілу лежить праворуч, і ряд має вершину зліва від центру; при правосторонній асиметрії між показниками центру розподілу існує певне співвідношення $Mo<Me<\bar{x}$;

якщо $K_A$<0, то скошеність лівостороння; у цьому випадку «довга» частина розподілу лежить ліворуч від центру і ряд має вершину праворуч; між показниками центру розподілу при лівосторонній асиметрії є певне співвідношення $Mo>Me>\bar{x}$;

якщо $K_A$=0, то варіаційний ряд симетричний (симетрично нормальний розподіл); необхідною, але недостатньою умовою симетричності є рівність трьох характеристик: середньої арифметичної, моди та медіани, тобто $\bar{x}=Mo=Me$.

Приклад. За даними табл. 8.27 обчислимо коефіцієнт асиметрії.

Маємо: $\bar{x}$=3,23; $\sigma$=0,7932.

Розрахуємо моду за даними табл. 8.9.

Для даних табл. 8.27 $f_m$=6; $i$=0,6; $f_{m+1}$=3; $f_{m-1}$=4; $x_0$=2,9. Значення моди складе:

$$Mo = 2{,}9 + 0{,}6 \cdot \frac{6-4}{(6-4)+(6-3)} = 2{,}9 + 0{,}6 \cdot \frac{2}{5} = 3{,}14;$$

$$K_A = \frac{3{,}23 - 3{,}14}{0{,}7932} = \frac{0{,}09}{0{,}7932} \approx 0{,}113.$$

Величина показника асиметрії свідчить про наявність незначної правосторонньої асиметрії у розподілі відібраних одиниць.

Для виявлення асиметрії можна також рекомендувати наступний показник:

$$Па = \frac{Q_3 - Q_1 - 2Me}{Q_0} =$$

$$= \frac{2\left[(Q_3 - Me) - (Me - Q_1)\right]}{Q_3 - Q_1}, \qquad (8.85)$$

де $\left.\begin{matrix} Q_3 \\ Q_1 \end{matrix}\right\}$ — верхній и нижній квартилі;

$Q_0$ — квартиль;

$Me$ — медіана.

У разі повної симетрії показник асиметрії дорівнюватиме нулю ($Па$=0).

Особливості розподілу варіаційних рядів при різних значеннях розглянутих характеристик представлено на рис. 8.12–8.15.

Розглянемо графік розподілу симетричної форми (рис. 8.12). Очевидно, що на такому графіку точка максимуму, лінія, що ділить площу і центр ваги лежать на осі симетрії. Отже, для симетричного розподілу мода, медіана та середня арифметична співпадають.

На рис. 8.13 графік має правосторонню асиметрію. У такому разі мода розташована лівіше, а середня арифметична — правіше медіани. Медіана ділить маси ряду розподілу на дві однакові частини, але ліва частина розташована на короткій, а права на довгій основі. Тому ліва частина кривої вище за праву. Перетин по осі абсцис медіани з кривою ряду розподілу не може служити точкою рівноваги, оскільки площа, розташована на довгому плечі, перетягне площу, розташовану

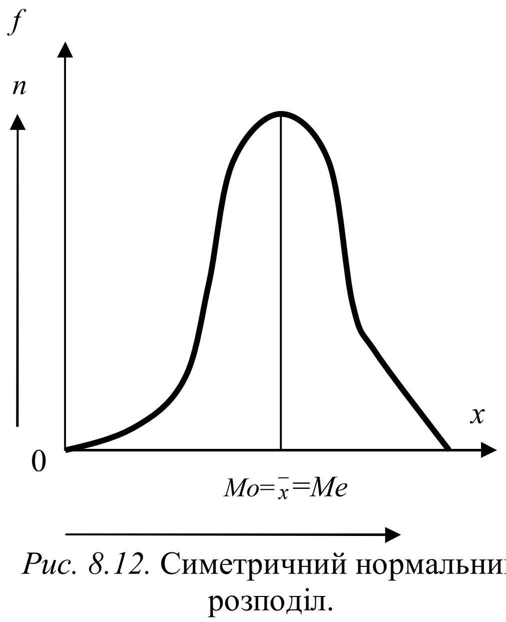



*Рис. 8.12.* Симетричний нормальний розподіл.
$\bar{x} = Mo = Me; K_A = 0.$

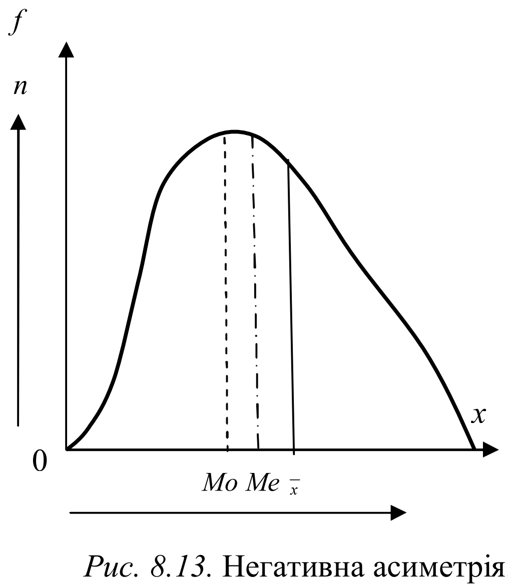



*Рис. 8.13.* Негативна асиметрія.
$\bar{x} < Me$; $\bar{x} < Mo$;
$K_A$ — від'ємний.

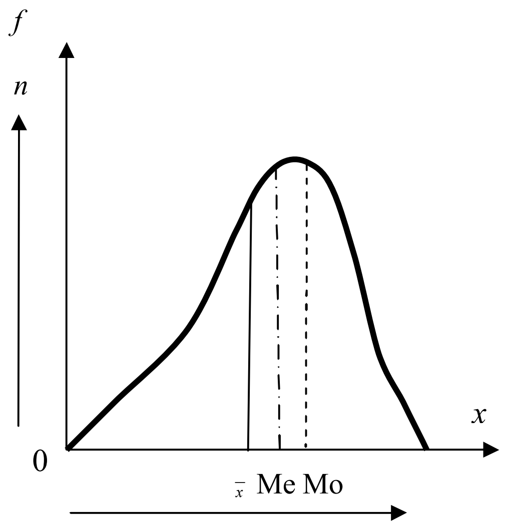



*Рис. 8.14.* Позитивна асиметрія.

$\bar{x}>Mo$; $\bar{x}>Me$; $K_A$ — позитивний.

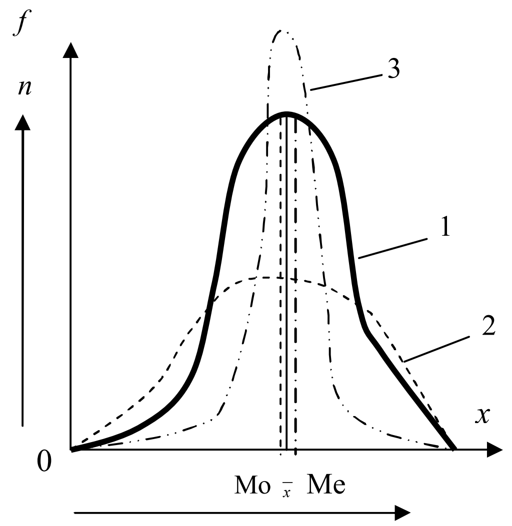



*Рис. 8.15.* Різновершинні симетричні розподіли.

$\bar{x}<Mo$; $\bar{x}<Me$;

1 – нормальний розподіл $\gamma=3$

2 – плосковершинний розподіл $\gamma<3$;

3 – крутовершинний розподіл $\gamma>3$.

на короткому плечі. Отже, середня арифметична знаходиться правіше за медіану.

Таким чином, при правосторонній асиметрії лівіше розташована мода, далі медіана і правіше — середня арифметична. При лівосторонній асиметрії має місце зворотне розташування показників, що характеризують центр розподілу. Таким чином, розбіжність між трьома показниками, що характеризують центр розподілу, залежить від характеру асиметрії варіаційного ряду: чим більш асиметричний графік, тим більша відстань між його середніми точками.

Ліндбергом був запропонований інший показник асиметрії, який розраховується за формулою:

$$K_A = П - 50, \tag{8.86}$$

де $П$ — процент тих значень ознаки, які перевищують за величиною середню арифметичну.

Найбільш точним і поширеним показником міри асиметрії або скошеності варіаційного ряду є показник косості, заснований на визначені центрального моменту третього порядку.

$$A_s = \frac{\mu_3}{\sigma^3}. \tag{8.87}$$

де $A_s$ — коефіцієнт асиметрії (показник косості);

$\sigma^3$ — куб стандартного вибіркового відхилення;

$\mu_3$ — центральний емпіричний момент третього порядку.

Для несгрупованого варіаційного ряду центральний емпіричний момент третього порядку розраховується за формулою:

$$\mu_3 = \frac{\sum_{i=1}^{n}(x_i - \widetilde{x})^3}{n}; \tag{8.88}$$

для згрупованого варіаційного ряду:

$$\mu_3 = \frac{\sum_{i=1}^{n}(x_i - \widetilde{x})^3 f_i}{\sum f_i}, \tag{8.89}$$

де $\widetilde{x}$ — вибіркова середня;

$x_i$ — варіанти дискретного ряду або середини інтервалів інтервального ряду;

$n$ — число ознак варіаційного ряду;

$f_i$ — відповідні частоти.

Якщо коефіцієнт асиметрії $A_s$>0, це вказує на наявність правосторонньої асиметрії (права гілка щодо максимальної ординати витягнута більше, ніж ліва), якщо $A_s$<0, це свідчить про наявність лівосторонньої асиметрії. Чим більший коефіцієнт асиметрії, тим більша скошеність варіаційного ряду. У статистиці прийнята наступна умовна градація: якщо $|A_s|$<0,25, то асиметрія незначна, якщо 0,25<$|A_s|$<0,5, то помірна і якщо $|A_s|$>0,5, то суттєва.

Оцінка ступеня суттєвості цього показника дається за допомогою середньої квадратичної помилки, яка залежить від обсягу спостережень *n* і розраховується за формулою:

$$\sigma_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}. \tag{8.90}$$

Якщо відношення $\frac{|A_s|}{\sigma_{A_s}}$<3, то асиметрія незначна і розподіл ознаки близько до нормального розподілу. Якщо $\frac{|A_s|}{\sigma_{A_s}}$>3, то асиметрія істотна і розподіл ознаки в генеральній сукупності не є симетричним.

Обчислимо показники асиметрії для емпіричного ряду розподілу, представленого в табл. 8.9.

Таблиця 8.49

| Денна заробітна плата, тис. грн ($x$) | Число робітників ($f$) | Центр інтервалу $x'$ | $x'f$ | $x'-\bar{x}$ | $\left(x'-\bar{x}\right)f$ | $\left(x'-\bar{x}\right)^2 f$ | $\left(x'-\bar{x}\right)^2 f$ |
|---|---|---|---|---|---|---|---|
| 1,7–2,3 | 3 | 2,0 | 6 | –1,23 | –3,69 | 4,5387 | –5,5826 |
| 2,3–2,9 | 4 | 2,6 | 10,4 | –0,63 | –2,52 | 1,5876 | –1,0002 |
| 2,9–3,5 | 6 | 3,2 | 19,2 | –0,03 | –0,18 | 0,0054 | –0,0002 |
| 3,5–4,1 | 3 | 3,8 | 11,4 | 0,57 | 1,71 | 0,9747 | 0,5556 |
| 4,1–4,7 | 4 | 4,4 | 17,6 | 1,17 | 4,68 | 5,4756 | 6,4065 |
| Всього | 20 | | 64,6 | | –12,78 | 12,5820 | 0,3791 |

Середня денна заробітна плата становитиме:

$$\tilde{x} = \frac{64{,}6}{20} = 3{,}23 \text{ тис. грн;}$$

Центральний момент третього порядку

$$\mu_3 = \frac{\sum_{i=1}^{n}(x_i - \tilde{x})^3 f_i}{\sum f_i} = \frac{0{,}3791}{20} = 0{,}01895.$$

Дисперсія

$$\sigma^2 = \frac{\sum (x_i - \tilde{x})^2 f}{\sum f} = \frac{12{,}5820}{20} = 0{,}6291,$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{0{,}6291} = 0{,}7932.$$

Коефіцієнт асиметрії складе:

$$A_s = \frac{\mu_3}{\sigma^3} = \frac{0{,}01895}{0{,}7932} \approx 0{,}024 > 0,$$

тобто розподіл має несуттєву правосторонню асиметрію.

Характер асиметрії іноді вказує на напрямок розвитку, що має важливе практичне значення при вивченні багатьох економічних явищ і процесів. При дослідженні варіації ознак, що характеризують розподіл робочих цеху за величиною їх виробітки, лівостороння асиметрія може сигналізувати про відсталість середніх норм і необхідність їх перегляду у бік підвищення. При дослідженні ознак, що характеризують розподіл виробів за кількістю витраченої на них сировини, спостерігається правостороння асиметрія, то в даному випадку мова може піти про необхідність перегляду витрат сировини в сторону зниження.

**Ексцес**. Для симетричних розподілів ознаки варіаційного ряду розраховується показник ексцесу. *Ексцес* — числова характеристика гостроверхості статистичного розподілу. Ексцес представляє собою випад вершини емпіричного розподілу вгору або вниз від вершини нормального розподілу. Для його обчислення використовують центральний момент четвертого порядку $r_4 = \frac{\mu_4}{\sigma^4}$, який називається мірою крутості.

Показник *ексцесу* ($\varepsilon$) може бути виражений формулою[1]:

$$\varepsilon = \beta - 3 = \frac{\mu_4}{(\sigma^2)^2} - 3. \tag{8.91}$$

де $\mu_4$ — центральний емпіричний момент четвертого порядку:

$$\mu_4 = \frac{\sum_{i=1}^{n}(x_i - \widetilde{x})^4}{n}$$ — для несгрупованого варіаційного ряду;

$$\mu_4 = \frac{\sum_{i=1}^{n}(x_i - \widetilde{x})^4 f_i}{\sum f_i}$$ — для згрупованого варіаційного ряду.

Якщо $\varepsilon>0$, то розподіл високовершинний («островершинний») щодо еталонного нормального розподілу, якщо $\varepsilon<0$, то розподіл низьковершинний. Для нормального розподілу $\varepsilon= 0$, оскільки у разі відношення $\frac{\mu_4}{\sigma^4}$=3. Чим більша величина ексцесу за модулем ($|\varepsilon|$), тим суттєвіша висота в розподілі ознаки.

Для вимірювання ексцесу користуються тією обставиною, що у нормальній кривою розподілу відношення $\frac{\mu_4}{(\sigma^2)^2}$ дорівнює трьом. У разі крутості, що відрізняється від нормальної, ця рівність порушується. Таким чином, різниця між $\beta$ і трьома безпосередньо вказує на величину і знак міри крутості — ексцесу: чим $\beta$ більше трьох, тим ряд гостріший, і навпаки.

На рис. 8.15 представлено два розподіли: один гостровершинний (величина — позитивна), другий — плосковершинний (величина екс-

---

[1] «Мінус три» в кінці формули введено для того, щоб коефіцієнт ексцесу стандартного нормального розподілу дорівнював нулю.

цесу від'ємна).

Середня квадратична помилка ексцесу розраховується за такою формулою:

$$\sigma_{Ex} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}, \qquad (8.92)$$

де $n$ — число спостережень.

Покажемо приклад розрахунку показника ексцесу за даними табл. 8.50.

Таблиця 8.50

| Денна заробітна плата, тис. грн ($x$) | Число робітників ($f$) | Центр інтервалу $x'$ | $x' - \bar{x}$ | $(x' - \bar{x})^4 f$ |
|---|---|---|---|---|
| 1,7–2,3 | 3 | 2,0 | –1,23 | 6,86660 |
| 2,3–2,9 | 4 | 2,6 | –0,63 | 0,63012 |
| 2,9–3,5 | 6 | 3,2 | –0,03 | 0,00000 |
| 3,5–4,1 | 3 | 3,8 | +0,57 | 0,31668 |
| 4,1–4,7 | 4 | 4,4 | +1,17 | 7,49555 |
| Всього | 20 | | | 15,30895 |

Розрахуємо центральний момент четвертого порядку

$$\mu_4 = \frac{\sum (x_i - \tilde{x})^4 f_i}{\sum f_i} = \frac{15{,}30895}{20} = 0{,}76845.$$

Середнє квадратичне відхилення склало $\sigma$ =0,7932, то для розглянутої задачі $(\sigma^2)$=(0,7932$^2$)=0,6291;

$$\beta = \frac{0{,}76845}{(0{,}6291)^2} \approx 1{,}94; \quad \varepsilon = 1{,}94 - 3 = -1{,}06.$$

За результатами обчислення ексцесу видно, що $\varepsilon$ від'ємний (–1,06). Від'ємна величина показника ексцесу свідчить про значну плосковершинність емпіричного розподілу. В цілому зазначений розподіл значно нижчий, ніж нормальний розподіл

Розрахуємо помилку ексцесу:

$$\sigma_{Ex} = \sqrt{\frac{24 \cdot 20 \cdot (20-2)(20-3)}{(20-1)^2(20+3)(20+5)}} = 0{,}84,$$

$\frac{|Ex|}{\sigma_{Ex}} = 1{,}095$, оскільки відношення $\frac{|Ex|}{\sigma_{Ex}} < 3$, слід припустити, що ексцес не властивий розподілу ознаки у генеральній сукупності.

Показник ексцесу може бути характеристикою більшої однорідності сукупності (при крутовершинному ряді і $\varepsilon > 3$), або навпаки, її меншої однорідності і більш менш зрівняному розподілі одиниць спостереження за варіантами ряду (при плоско-вершинному ряді і $\varepsilon < 3$). Таким чином і ексцес є однією з якісних характеристик їх сукупності.

## 8.11. Перевірка гіпотези про відповідність емпіричного розподілу типу кривих нормального розподілу

**Перевірка близькості теоретичного та емпіричного розподілу**. Між частотами емпіричного і теоретичного розподілу завжди маються певні розбіжності. У деяких випадках причини розбіжностей можуть бути пояснені випадковостями при організації вибірки. В інших випадках причиною невідповідності може бути та обставина, що розподіл ознаки в генеральній сукупності не відповідає нормальному розподілу.

При аналізі рядів розподілу дуже важливо встановити, до якого типу вони належать, чи є однорідними. Попередньо встановити форму розподілу емпіричних даних у деяких випадках на основі теоретичного аналізу. Для отримання приблизного уявлення про форму розподілу використовують графічний метод, коли варіаційний ряд зображений у вигляді гістограм і полігонів. Число спостережень, за якими будується емпіричний розподіл, зазвичай невеликим і представляє собою вибірку з генеральної сукупності. Щоб відповісти на питання про тип розподілу, встановлюють близькість розподілу емпіричних даних до теоретичного розподілу, що встановлюється за допомогою побудови відповідної кривої розподілу. Однак у більшості випадків ні теорія, ні безпосередній розгляд емпіричного ряду даних не дозволяє визначити форму розподілу. Тоді зазвичай емпіричні дані досліджують на їхню близькість до нормального розподілу.

У статистиці для перевірки значущості розбіжності емпіричних (спостережних) і теоретичних (очікуваних) частот і однорідності досліджуваної сукупності вироблені різні критерії відповідності або згоди. Ці критерії було запропоновано різними вченими, які займалися дослідженнями цих питань. У статистиці обчислюються критерії згоди Пірсона, Романовського, Колмогорова та Ястремського. Найбільш поширеним є критерій згоди Пірсона $\chi^2$ («хі-квадрат»).

*Критерій згоди Пірсона* ґрунтується на визначенні величини $\chi^2$, яка обчислюється як сума квадратів різниць емпіричних і теоретичних частот, віднесених до теоретичних частот, тобто

$$\chi^2 = \sum \frac{(f_i - f_i')^2}{f_i'}, \tag{8.93}$$

де $f_i$ і $f_i'$ — відповідно частоти емпіричного та теоретичного розподілу у визначеному інтервалі.

Чим більше буде різниця між спостережуваними і теоретичними частотами, тим більше буде величина критерію К. Пірсона.

Для оцінки того, наскільки емпіричний розподіл відтворюється нормальним розподілом, обчислюють за розподілом Пірсона ймовірності досягнення $\chi^2$ даного значення $P(\chi^2)$.

Значення $P(\chi^2)$ обчислені для різних $\chi^2$ наводяться в додатку2, де по рядках вказуються значення $\chi^2$, а по стовпцях — значення ступенів свободи варіювання емпіричного ряду розподілу ($k$). Число ступенів свободи варіації визначається для даного ряду розподілу і дорівнює числу груп у ньому мінус чисельність обчислених характеристик (середня, дисперсія, моменти розподілу і т. д.), використаних при обчисленні теоретичного розподілу.

Щоб відрізнити суттєві значення $\chi^2$ від значень, які могли виникнути в результаті випадковостей вибірки, розраховані критерії порівнюються з табличним значенням $\chi^2$ при відповідному числі ступеня свободи і заданому рівні значимості. Рівень значущості вибираємо в такий спосіб, що ($\chi^2_{\text{розр}} > \chi^2_{\text{табл}}$)= $a$ (величина a приймається 0,05 чи 0,01).

При перевірці значущості розбіжностей між емпіричними (спостережуваними) та теоретичними (очікуваними) частотами ми можемо зустрітися з такими варіантами:

1) $\chi^2_{\text{розр}} > \chi^2_{\text{табл}}$, тобто розбіжність між теоретичними та емпіричними розподілами істотна і її не можна пояснити випадковими коливаннями вибіркових даних; таким чином перше мало відображає друге;

2) $\chi^2_{\text{розр}} \le \chi^2_{\text{табл}}$, тобто розрахований критерій не перевищує максимально можливу величину розбіжностей емпіричних і теоретичних частот, яка може виникнути внаслідок випадкових обставин.

Як було зазначено вище, табличне значення критерію Пірсона визначається при фіксованому рівні значущості і відповідному числі ступенів свободи.

За даними прикладу з табл. 8.48 розглянемо розрахунок критерію (хі-квадрат) (табл. 8.51).

Таблиця 8.51

| Номер інтервалу | Емпіричні частоти $f_i$ | Теоретичні частоти $f_i'$ | $(f_i - f_i')^2$ | $\frac{(f_i - f_i')^2}{f_i'}$ |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 1 | 3 | 4 | 1 | 0,25 |
| 2 | 15 | 8 | 49 | 6,13 |
| 3 | 12 | 15 | 9 | 0,60 |
| 4 | 27 | 23 | 16 | 0,70 |
| 5 | 31 | 31 | 0 | 0,00 |
| 6 | 32 | 35 | 9 | 0,26 |
| 7 | 32 | 33 | 1 | 0,03 |
| 8 | 28 | 27 | 1 | 0,04 |
| 9 | 22 | 19 | 9 | 0,47 |
| 10 | 8 | 11 | 9 | 0,82 |
| 11 | 6 | 6 | 0 | 0,00 |
| 12 | 2 | 3 | 1 | 0,33 |
| 13 | 2 | 1 | 1 | 1,00 |
| Всього | – | – | – | 10,62 |

$$\chi^2=10,62;\ k=12.$$

З таблиці $P(\chi^2)$ (додаток 2) для $\chi^2=11$ знаходимо імовірність $P(\chi^2)=0,5289$; вона досить велика. Отже, розходження теоретичних і емпіричних частот у нашому прикладі випадкові, емпіричний ряд однорідний, розподіл частот дуже близько до нормального.

Значення $\chi^2$ наводяться у спеціальних таблицях з математичної статистики. За рівня значимості 0,05 і числі ступенів свободи 12 $\chi^2=19,675$ (додаток 3). Таким чином, розраховане значення критерію Пірсона не перевищує табличного значення при $a=0,05$, тобто наведений розрахунок дає право не відкидати гіпотезу про нормальний характер емпіричного розподілу.

При розрахунку критерію Пірсона слід дотримуватися таких умов: 1) кількість спостережень має бути досить великою, принаймні $n \geq 50$; 2) якщо теоретичні частоти в деяких інтервалах менше 5, то такі інтервали об’єднуються так, щоб частоти були більшими за 5.

Використовуючи величину $\chi^2$, *І. В. Романовський* запропонував оцінку близькості емпіричного розподілу кривої нормального розподілу проводити по відношенню:

$$\frac{\chi^2 - k}{\sqrt{2k}}, \tag{8.94}$$

де $k$ — число ступенів свободи.

Якщо відношення $\frac{\chi^2 - k}{\sqrt{2k}}$>3, то розбіжність частот емпіричного розподілу і розрахованих частот нормального розподілу не можна визнати випадковим, і гіпотезу про нормальний розподіл слід відкинути. Якщо $\frac{\chi^2 - k}{\sqrt{2k}}$≤3, то пропонується розбіжність між теоретичним та емпіричним розподілами вважати несуттєвою. Несуттєвість розбіжності (коли величина відносини Романовського менше трьох) дає можливість прийняти гіпотезу про нормальний характер емпіричного розподілу. Продовжуючи наведений вище приклад, розрахуємо критерій В. І. Романовського:

$$\frac{\left|\chi^2 - k\right|}{\sqrt{2k}} = \frac{\left|10,62 - 12\right|}{\sqrt{2 \cdot 12}} = \frac{1,38}{4,9} = 0,28.$$

Так як розраховане відношення значно менше трьох, тому розбіжності між теоретичними і емпіричними частотами можна вважати несуттєвими, і, таким чином, слід прийняти гіпотезу про нормальність емпіричного розподілу.

Для визначення відповідності між теоретичним та емпіричним розподілом академік А. М. Колмогоров запропонував використовувати різницю між значеннями інтегральної функції теоретичного розподілу ($\lambda$).

*Критерій $\lambda$ Колмогорова* встановлює близькість теоретичних і емпіричних розподілів шляхом порівняння різниці між значеннями інтегральної функції теоретичного розподілу ($F_{(x)}$) і значеннями інтегральної функції емпіричного розподілу ($F'_{(x)}$) і обчислюється виходячи з $D$ — максимальної верхньої межі значення різниці їх накопичених частот до квадратного кореня з числа спостережень $N$:

$$\lambda = \frac{D}{\sqrt{N}}, \tag{8.95}$$

де $D$ — максимальна різниця різниці: нагромаджених теоретичних частот ($M'$), нагромаджених емпіричних частот.

Наведемо таблицю значень $p(\lambda)$, ймовірності того, що $\lambda$ досягне величини (див. табл. 8.52).

Таблиця 8.52

| $\lambda$ | $p(\lambda)$ | $\lambda$ | $p(\lambda)$ |
|---|---|---|---|
| 0,30 | 1,0000 | 1,10 | 0,1777 |
| 0,35 | 0,9997 | 1,20 | 1122 |
| 0,40 | 9972 | 1,30 | 0681 |
| 0,45 | 9874 | 1,40 | 0397 |
| 0,50 | 9639 | 1,50 | 0222 |
| 0,55 | 9228 | 1,60 | 0120 |
| 0,60 | 8643 | 1,70 | 0062 |
| 0,70 | 7112 | 1,90 | 0015 |
| 0,75 | 6272 | 2,00 | 0007 |
| 0,80 | 5441 | 2,10 | 0003 |
| 0,85 | 4653 | 2,20 | 0001 |
| 0,90 | 3927 | 2,30 | 0001 |
| 0,95 | 3275 | 2,40 | 0000 |
| 1,00 | 2700 | 2,50 | 0000 |

Розглянемо застосування цього критерію на наведеному раніше прикладі (табл. 8.53).

Таблиця 8.53

| $x$ | Емпіричні частоти $f_i$ | Теоретичні частоти $f_i'$ | $M$ | $M'$ | $M - M'$ |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 3 | 4 | 3 | 4 | –1 |
| 2 | 15 | 8 | 18 | 12 | 6 |
| 3 | 12 | 15 | 30 | 27 | 3 |
| 4 | 27 | 23 | 57 | 50 | 7 |
| 5 | 31 | 31 | 88 | 81 | 7 |
| 6 | 32 | 35 | 120 | 116 | 4 |
| 7 | 32 | 33 | 152 | 149 | 3 |
| 8 | 28 | 27 | 180 | 176 | 4 |
| 9 | 22 | 19 | 202 | 195 | 7 |
| 10 | 8 | 11 | 210 | 206 | 4 |
| 11 | 6 | 6 | 216 | 212 | 4 |
| 12 | 2 | 3 | 218 | 215 | 3 |
| 13 | 2 | 1 | 220 | 216 | 4 |

Максимальна величина різниці абсолютних нагромаджених частот дорівнює 7 (див. графу 6 табл. 8.53).

$$\lambda = \frac{D}{\sqrt{N}} = \frac{7}{\sqrt{220}} = \frac{7}{14{,}8} \cong 0{,}5.$$

У таблиці ймовірностей $p(\lambda)$ знаходимо для $\lambda$=0,5; $p(\lambda)$=0,9639. Це велика ймовірність вказує на те, що розбіжність між теоретичним і емпіричним розподілами є несуттєвою і може бути випадковим.

## ЛІТЕРАТУРА


1. *Горкавий В. К.* Статистика: Підручник. Третє вид., переробл. і доповн. Київ: Алерта, 2019. С. 113–139.
2. Статистика: підручник / *С. І. Пирожков*, *В. В. Рязанцева*, *Р. М. Моторин* та ін. Київ: Київ. нац. торг.-екон. ун-т, 2020. С. 130–152.
3. Статистика: Підручник / *С. С. Герасименко*, *А. В. Головач*, *А.М. Єріна* та ін.; За наук. ред. д-ра екон. наук С. С. Герасименка. 2-ге вид., перероб. і доп. Київ: КНЕУ, 2000. С. 60–80.
4. *Ткач Є. І.*, *Сторожук В. П.* Загальна теорія статистики: підручник. 3-тє вид. Київ: Центр учбової літератури, 2009. С. 156–170.


## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники, словники*


5. *Козирєва О. В.*, *Федорова В. О.* Статистика: навчальний посібник. Харків: Видавництво Іванченка І.С., 2021. С. 70–79.
6. *Мармоза А. Т.* Теорія статистики: підручник. 2-ге вид., перероб. і доп. Київ: «Центр учбової літератури», 2013. С. 177–203.
7. *Опря А. Т.* Статистика (модульний варіант з програмованою формою контролю знань). Навч. посіб. Київ: Центр учбової літератури, 2012. С. 87–116.
8. *Педченко Г. П.* Статистика: Навчальний посібник. Мелітополь: Колор Принт, 2018. С. 119–133.
9. Статистичний словник / [*О. Г. Осауленко*, *О. О. Васєчко*, *М. В. Пугачова* та ін.]; за ред. д-ра держ. упр., проф., член-кор. НАН України О. Г. Осауленка; НТК статистичних досліджень. Київ: ДП «Інформ.-аналіт. агентство», 2012. 498 с.
10. *Шапочка М. К.*, *Маценко О. М.* Теорія статистики: навч. посібник. Суми: Університетська книга, 2014. С. 155–170.

Г Л А В А 9

# СТАТИСТИЧНІ МЕТОДИ ВИВЧЕННЯ ВЗАЄМОЗВ'ЯЗКІВ

## 9.1. Взаємозв'язки суспільних явищ та необхідність їх статистичного вивчення

**Поняття про кореляцію**. Вивчення дійсності показує, явища життя формуються під впливом численних, різноманітних і взаємозалежних факторів. Це значною мірою визначає складність вивчення суспільних явищ. Щоб правильно зрозуміти сутність явищ суспільного життя — найскладніших явищ, треба керуватися положеннями діалектичного матеріалізму про загальний зв'язок і взаємообумовленість в їх розвитку. Відповідно до цих принципів суспільні явища органічно взаємопов'язані між собою, залежать один від одного і знаходяться в постійному розвитку і русі. Для того, щоб пізнати якесь явище, необхідно вивчити його у всіх взаєминах з явищами-факторами, досліджувати зв'язки з іншими явищами. Їхнє пізнання вимагає вивчення як впливу, що здійснюється одним окремо взятим явищем на інше, так і спільного їхнього впливу. Так, наприклад, обсяг виробництва продукції підприємства залежатиме від числа робітників, рівня їх кваліфікації, рівня продуктивності праці, технічного стану основних фондів, ступеня вдосконалення технології та організації виробництва та безлічі інших факторів.

Зв'язки між суспільними явищами та процесами дуже складні та різноманітні, а іноді носять суперечливий характер. Зокрема, в економіці — це зв'язки між виробництвом, розподілом, обміном і споживанням суспільного продукту, окремими галузями народного господарства, окремими підприємствами, окремими показниками їх роботи. Наприклад, впровадження нового обладнання на підприємстві призводить до зростання продуктивності праці, що веде до зниження собівартості продукції та зростання прибутку. І це, в свою чергу, дозволяє спрямувати частину коштів у збільшення оплати праці, що створює передумови подальшого зростання продуктивність праці.

Вивчення взаємозв'язків та взаємозалежностей явищ — одне з найважливіших завдань, що стоять перед статистикою. Використання методів і прийомів для пізнання суспільних явищ дозволяє не тільки виявити наявність та напрямок зв'язків, але також виміряти їх тісноту та оцінити достовірність одержуваних результатів. У процесі статистичного дослідження залежностей розкриваються причинно-наслідкові зв'язки між явищами і процесами, що дозволяє виявити фактори (ознаки), що надають вирішальний вплив на варіацію результативної ознаки. Особливості статистичного вивчення зв'язку між явищами полягають у тому, що вони дають можливість не тільки оцінити тісноту зв'язку, але і виразити зв'язок аналітично.

У математичній статистиці взаємозв'язок між явищами вивчається методами кореляції. *Кореляція* — залежність між випадковими величинами, що не має строго функціонального характеру, за якої зміна однієї з випадкових величин приводить до зміни математичного сподівання іншої. Цей зв'язок (залежність) між явищами та ознаками складається з елементів, одна частина яких підпорядкована суворої причинної залежності, а інша індивідуально, властива кожному з явищ, що зіставляються (вивчаються). Особливість вивчення взаємозв'язку явищ чи ознак методами кореляції полягає в тому, що не можна ізолювати вплив сторонніх факторів або тому, щоб ці фактори невідомі, або тому, що їхня ізоляція неможлива. Цей по сторонній вплив перешкоджає повному і чистому прояву функціональних зв'язків, приховуючи їх однозначність і нерозривність, тобто створюючи множинність причин і наслідків. Тому метод кореляції застосовується для того, щоб при складному впливі між явищами встановити, яка була б залежність між ознаками за умови, якби побічні, особливі для кожного з порівнюваних явищ впливу, не змінювалися і своєю зміною не спотворювали б основну залежність.

Для вивчення та кількісної оцінки взаємозв'язків і взаємозалежностей між ознаками в статистиці використовується ряд методів. Вибір методу залежить від характеру вихідного матеріалу та цілей дослідження. Методи вивчення зв'язків, розроблені в статистиці, застосовують не тільки для вивчення простих зв'язків між двома явищами, але і множинними, між багатьма явищами.

Дослідження об'єктивно існуючих зв'язків між явищами має велике значення в соціально-економічній статистиці. Все зростаючі масштаби суспільного виробництва, вдосконалення та ускладнення його структури та господарських зв'язків пред'являють у сучасних умовах підвищені вимоги до регулювання соціально-економічними процесами. Ефективне регулювання соціально-економічних процесів має спиратися на глибокий аналіз взаємозв'язків і взаємозалежностей у народному господарстві. У міру розвитку економіки та суспільства роль та значення статистичних методів в економічному аналізі підвищуються, розширюється сфера їх застосування, удосконалюються методи та методологія обчислення. Розроблені в даний час статистичні методи дозволяють вивчити, виміряти і дати кількісне вираження взаємозв'язків між явищами суспільного життя, встановленими на основі якісного аналізу.

Застосування методів кореляційного аналізу має важливе значення для виявлення наявності, для характеристики форми (спрямованості), для визначення ступеня (заходи) і для пізнання самої причинності взаємозв'язків між явищами економічного життя, а також при вирішенні широкого кола інженерно-технічних питань. сов розвитку та вдосконалення виробництва, у різних сферах народного господарства.

Використання методів статистики для виявлення та вимірювання зв'язків між витратами на виробництво і факторами, що визначають їх формування, має не тільки суто теоретичне, а й велике практичне значення, оскільки дозволяє перейти від констатації фактів до активного впливу на них.

Тому статистичним методам вивчення зв'язку у статистиці приділяється велика увага.

**Результативна та факторна ознаки**. При вивченні взаємозв'язку і взаємодії між явищами одні ознаки виступають як фактори, що зумовлюють зміни інших ознак. Ознаки цієї першої групи ми називатимемо *ознаками-факторами* (*факторними ознаками*); а ознаки, які є результатом впливу всіх цих факторів, будемо називати *результативними ознаками*. В процесі дослідження необхідно встановити, яка з ознак є факторною, а якою є результативною. Це питання вирішується переважно з урахуванням якісного аналізу. Для вирішення цього завдання використовується, насамперед, логічний аналіз.

Приклад 1. Обсяг виробництва промислової продукції окремого підприємства залежить багатьох чинників, зокрема вартості основних промислово-виробничих засобів. Обсяг виробництва виступає у разі результативним ознакою, а вартість основних промислово-виробничих засобів — як факторна.

Приклад 2. Щоб судити про переваги великих підприємств перед дрібними, розглянемо, як збільшувалася врожайність зернових та зернобобових культур залежно від збільшення розмірів підприємства.

Таблиця 9.1

**Групування підприємств за розмірами зібраної площі зернових та зернобобових культур у 2020 р.***

| Групи підприємств за розміром збираної площі, га | Кількість підприємств | Урожайність, ц с 1 га |
|---|---|---|
| Всього | 3 2513 | 46,4 |
| із них з площею га: | | |
| до 100 | 1 9026 | 30,2 |
| 100,01–200,00 | 3559 | 36,6 |
| 200,01–500,00 | 4213 | 40,0 |
| 500,01–1000,00 | 2765 | 42,2 |
| 1000,01–2000,00 | 1880 | 46,3 |
| 2000,01–3000,00 | 566 | 48,6 |
| понад 3000,00 | 504 | 56,6 |

* *Джерело*: Групування підприємств за розмірами зібраної площі основних сільськогосподарських культур у 2020 році/ Державна служба статистики України. URL: http://www.ukrstat.gov.ua/

У цій таблиці бачимо залежність урожайності зернових і зернобобових культур від розмірів сільськогосподарських підприємств. Цей зв'язок прямий: чим більший розмір прибраної площі зернових та зернобобових культур, тим вища їхня урожайність. Ознака групування — площа зернових та зернобобових культур — є факторною, урожайність результативною ознакою.

Від обсягу виробництва також залежить врожайність цукрових буряків, про що свідчить табл. 9.2. З таблиці ясно видно прямий зв'язок між розміром виробництва і врожайністю соняшнику. Зі збільшенням убраної площі соняшнику зростає і врожайність. Однак залежність результативної ознаки (урожайності) від факторної носить не обов'язковий характер. Якщо у загальній масі ми спостерігаємо цей зв'язок, то в окремих групах бувають і відступи від загальної закономірності. Такі відступи — характерна особливість статистичного зв'язку, про яку буде сказано нижче.

Одна і та ж ознака в одному випадку виступає як факторна, а в іншому — результативна. Так, підвищення продуктивності праці

Таблиця 9.2

**Групування підприємств за розмірами зібраної площі соняшнику у 2020 р.***

| Групи підприємств за розміром збираної площі, га | Кількість підприємств | Урожайність, ц с 1 га |
|---|---|---|
| Всього | 21856 | 21,4 |
| із них з площею га: | | |
| до 100 | 12692 | 16,4 |
| 100,01–200,00 | 2894 | 19,6 |
| 200,01–500,00 | 3424 | 20,7 |
| 500,01–1000,00 | 1704 | 20,8 |
| 1000,01–2000,00 | 824 | 22,1 |
| 2000,01–3000,00 | 181 | 24,7 |
| понад 1000,00 | 137 | 24,6 |

* *Джерело*: Групування підприємств за розмірами зібраної площі основних сільськогосподарських культур у 2020 році/ Державна служба статистики України. URL: http://www.ukrstat.gov.ua/

супроводжується зменшенням витрат праці на виробництво продукції, що за інших рівних умов призводить до зниження собівартості продукції. В даному випадку продуктивність праці — факторна ознака, собівартість продукції — результативна ознака. В свою чергу, собівартість впливає на розмір прибутку, чим вища собівартість продукції, тим менший обсяг прибутку. Тут собівартість прибутку — факторна ознака, прибуток — результативна ознака. У кожному конкретному випадку для встановлення факторного і результативного ознаки при вивченні взаємозв’язку різних ознак необхідно провести аналіз сутності та природи зв’язку.

## 9.2. Види зв’язків ознак між явищами

Зв’язки між явищами та ознаками через велику їх різноманіття класифікують у статистиці за рядом підстав. Існують різні види і форми зв’язків, що розрізняються по суті, характеру, за напрямом, тісноті, за аналітичним виразом і т.д.

**Функціональні та статистичні (кореляційні) зв’язки**. Залежно від ступеня тісноти зв'язку розрізняють два види зв'язку: функціональну (повну) та кореляційну (неповну).

При *функціональному зв'язку* певному значенню однієї величини (незалежної змінної, факторної ознаки) відповідає строго певне значення іншої (залежної змінної, результативної ознаки). Функціональні зв'язки характеризуються тим, що при таких зв'язках факторіальна ознака повністю визначає величину результативної ознаки. Зазвичай функціональні зв'язки виражаються як формул. У більшості випадків функціональні зв'язки переважають у точних науках, головним чином у фізиці та математиці. Знаючи, наприклад, радіус кола *x*, знайдемо площу цього кола у за формулою: $S = \pi R^2$. Тут *y* функціонально пов'язаний з *x*. Функціональний зв'язок може показувати залежність між результативною ознакою та кількома факторами. Наприклад, площа прямокутника *y* залежить від його довжини *x* та висоти *z*, тобто $y=xz$. Прикладом функціональної залежності в галузі економіки може служити залежність між відрядною розцінкою, кількістю виготовленої продукції та відрядною заробітною платою робітника.

При цьому важливо відзначити, що функціональна залежність *з однаковою силою проявляється у всіх одиниць сукупності*, незалежно від зміни інших ознак даного явища. Так, для прямокутника будь-якої величини буде діяти математична формула, що характеризує залежність його площі від довжини та ширини. Тому достатньо встановити форму функціональної залежності для одного випадку, щоб можна було її застосувати у всіх інших випадках.

Функціональний зв'язок схематично можна уявити рівнянням:

$$y_i = f(x_i), \tag{9.1}$$

де $y_i$ — результативна ознака;

$x_i$ — факторна ознака;

$f(x_i)$ — відомий функціональний зв'язок цих ознак.

В якості факторної ознаки може бути й не одна, а безліч ознак. Однак залежність від однієї ознаки-фактора або від двох і більше факторів повинна повністю визначити величину результативної ознаки. Тому функціональні зв'язки називають зв'язками повними, «жорсткими», точними.

Функціональні зв'язки майже не зустрічаються в явищах суспільного життя, що відрізняються складністю і різноманіттям існуючих взаємозв'язків і взаємозалежностей. Вони характерні лише для найпростіших явищ, які не вимагають спеціального дослідження. Наприклад, валовий збір сільськогосподарських культур дорівнює добутку врожайності на посівну (збиральну) площу.

У практичній діяльності превалюють складніші і менш «жорсткі» залежності, які виявляються над кожному одиничному випадку, а в загалі, за наявності великої кількості спостережень. Це свідчить про те, що між ознаками-факторами і результативною ознакою існує не повний, а зв'язок, що проявляється лише в середньому. Такі зв'язки називаються кореляційними.

При *кореляційної зв'язку* одному й тому самому значенню ознаки-фактора можуть відповідати різні значення результативної ознаки (при тому самому рівні кваліфікації робочих продуктивність їх праці може мати різні значення). Кореляційні залежності, які вивчає статистика, характеризуються тим, що зміна того чи іншого показника, тобто його варіація, є результатом впливу цілого комплексу факторів, частина яких має основне значення і властива всім одиницям, а інша частина має другорядне значення і властива лише окремим одиницям. Вивчення залежності варіації ознаки від чинників і становить зміст теорії кореляції[1].

Поняття кореляційної залежності є окремим випадком ширшого поняття — залежності стохастичної. Зв'язок між випадковими величинами проявляється в тому, що одна з них впливає на зміну інший зміною свого закону розподілу. Такий зв'язок між випадковими величинами називають *ймовірнісним* або *стохастичним зв'язком.*

Так, при зміні обсягу виробництва виникає зміна собівартості продукції, але ця зміна не пропорційна зміни обсягу випуску і коливається близько середньої величини. Прикладом використання кореляційних зв'язків може бути залежність між кількістю внесених добрив і урожайністю сільськогосподарських культур. При збільшенні кількості внесених добрив зростає врожайність, але зростає на різну величину. Причому може статися так, що на тих ділянках, де було внесено менше добрив, урожайність вища. Це пов'язано з тим, що на урожайність впливають ще й інші, не враховані фактори: якість насіння, терміни посіву і збирання, культура землеробства, попередники, рельєф грунту та інші.

[1] Основоположниками теорії кореляції вважаються англійські учені *Ф. Гальтон* (1822–1911) і *К. Пірсон* (1857–1936).

Або другий приклад: рівень продуктивності праці робітника на машинобудівних заводах однакової спеціалізації залежить від рівня кваліфікації робітника (розряду). Разом з тим на практиці часто трапляються випадки, коли на заводі навіть в робочих однакової кваліфікації виробіток різний. Це пов'язано з тим, що, крім кваліфікації, на продуктивність праці впливають інші, не враховані чинники. До таких факторів можна віднести виробничий стаж, технічний стан обладнання, забезпечення матеріалами, інструментами і т. п. Тому, порівнюючи продуктивність праці та кваліфікацію робітників на двох окремих заводах, можна і не роздивитися між ними зв'язки. Може трапитися так, що на заводі, в якому вища кваліфікація робітників, виявиться нижчою рівень продуктивності праці. Це означає, що на рівень продуктивності праці вплинули якісь інші фактори, що її знижують. Але якщо ми візьмемо досить велику кількість робітників, то можна виявити певний взаємозв'язок між рівнем кваліфікації робітника і її продуктивністю. Значить, важлива особливість кореляційних зв'язків полягає в тому, що *вони проявляються не в поодиноких випадках, а лише при великій кількості спостережень* у вигляді певної залежності між середнім значенням результативної ознаки і ознаками-факторами. Оскільки тільки при великому числі спостережень індивідуальні особливості та другорядні фактори згладжуються і залежність, якщо вона має істотну силу, виявиться досить чітко.

Знаходження взаємозв'язку між ознаками-факторами і результативною ознакою при кореляційних зв'язках засноване на дії закону великих чисел: при великій масі спостережень погашаються індивідуальні особливості та усуваються випадкові коливання.

Друга особливість кореляційних зв'язків у тому, що вони є *неповними*. Кореляційний зв'язок характерний тим, що на результативну ознаку, крім факторного, впливають безліч інших ознак, що діють у різних напрямках, одночасно і послідовно. Тому кореляційні зв'язки — це зв'язки неповні, несуворі, наближені.

Кореляційний зв'язок в загальному вигляді можна виразити рівнянням:

$$y_i = \varphi(x_i) + \varepsilon_i, \tag{9.2}$$

де $\varphi(x_i)$ — частина результативної ознаки, що сформувалася під впливом врахованих відомих факторних ознак (одного або множини), що перебувають у кореляційному зв'язку ($\varphi$) з ознакою ($y_i$);

$\varepsilon_i$ — частина результативної ознаки, що виникла внаслідок другорядних та випадкових факторів.

Відмінною особливістю кореляційних зв'язків є те, що в них результативна ознака не повністю визначається дією факторних ознак. Цей вплив проявляється лише в середньому, а в окремих випадках виходять результати, що навіть суперечать установленому зв'язку. Тому, навіть за досить велику кількість спостережень, де погашаються індивідуальні особливості та усуваються випадкові фактори, виявлені залежності неноситимуть повного, функціонального характеру. Вони певною мірою наближатимуться до функціональних зв'язків, але дія інших «неврахованих» факторів виявиться в тому, що кореляційний зв'язок виявиться неповним, він не досягне за силою зв'язку функціонального.

**Прямі та зворотні зв'язки**. Залежно від напряму дії виділяють зв'язок прямий і зворотний. При *прямому зв'язку* обидва види ознак (факторна і результативна) змінюються в тому самому напрямку, тобто зі збільшенням значення факторної ознаки збільшується і результативна і, навпаки, зі зменшенням факторної ознаки зменшується і результативний. У разі *зворотного зв'язку* значення результативної ознаки змінюються під впливом факторної, але у протилежному напрямі по порівнянню зі зміною факторного ознаки. Так, зі збільшенням фактора $x$ результативна ознака зменшується. У явищах суспільного життя зустрічаються і ті та інші зв'язки. Наприклад, чим вища фондоозброєність праці робітників на заводі, то вище рівень продуктивність праці — прямий зв'язок. А чим вищий рівень продуктивності праці, тим нижча собівартість продукції — зворотний зв'язок.

**Прямолінійні та криволінійні зв'язки**. За аналітичним виразом (формою) виділяють зв'язки прямолінійні (просто лінійні) та нелінійні. При *прямолінійному зв'язку* результативна ознака зі збільшенням факторної ознаки поступово зростає або зменшується. Математично такий зв'язок виражається рівнянням прямою $y = a_0 + a_1 x,$ а графічно — прямою лінією. Звідси її більш коротка назва — лінійний зв'язок. При *криволінійному зв'язку* із зростанням величини факторної ознаки зростання (або спадання) результативної ознаки відбувається нерівномірно або напрям його зміни змінюється на зворотній. Геометрично криволінійний зв'язок може бути виражений рівнянням будь-якої кривої лінії (параболи, гіперболи тощо). При цьому необхідно пам'ятати, що тільки функціональні зв'язки можуть бути виражені точно за допомогою аналітичного рівняння. Кореляційні зв'язки виражаються лише приблизно, за умови абстрагування впливу всіх інших чинників. Тому на графіку і спостерігатиметься розкид точок ($y$; $x$) поблизу теоретичної лінії.

**Однофакторні та багатофакторні зв'язки**. При вивченні як функціональних, так і кореляційних взаємозв'язків велике значення має кількість ознак-факторів. За кількістю взаємопов'язаних факторів виділяють однофакторні та багатофакторні. При *однофакторному* (*парному*) *зв'язку* досліджується зв'язок між однією ознакою-фактором і результативною ознакою (при абстрагуванні впливу інших). При *багатофакторному зв'язку* зміна результативної ознаки відбувається під впливом двох або більшого числа найрізноманітніших факторних ознак.

Математично багатофакторний зв'язок у загальному вигляді можна описати наступною формулою:

$$y = f(x_1, x_2, ..., x_n), \tag{9.3}$$

де $x_1, x_2, ..., x_n$ — власне фактори перший, другий, $n$-й.

У випадку багатофакторного зв'язку мається на увазі, що фактори діють комплексно, тобто водночас і у взаємозв'язку.

**Інші види зв'язків**. Крім того, за ступенем впливу існують також зв'язки *безпосередні* та *непрямі*. Фактор $x$ може безпосередньо впливати на зміну результативної ознаки у або опосередковано, через інший фактор $v$. Залежно від сили впливу на результативну ознаку фактори можуть надавати *слабкий* і *сильний* вплив, який у статистиці вимірюється за допомогою спеціальних показників.

Тому, перш ніж приступати до вивчення зв'язку між явищами, необхідно встановити характер цих зв'язків.

**Основні статистичні методи вивчення взаємозв'язків**. Взаємозв'язки між явищами та процесами можуть бути вивчені, виміряні та кількісно виражені за допомогою різних статистичних методів. Ці методи, будучи основним засобом вивчення масових, якісно однорідних, повторюваних явищ, відіграють важливу роль у прогнозуванні поведінки узагальнюючих показників. Коли зв'язок між аналізованими характеристиками не функціональний, а стохастичний, тобто піддається впливу випадку і впливу різних факторів, статистичні та імовірнісні методи будуть практично єдиними інструментами дослідження.

Для дослідження функціональних зв'язків у статистиці застосовується балансовий та індексний методи. Індексний метод буде розглянутий в гл. 11, балансовий розглядається в наступному параграфі.

Для дослідження кореляційних зв'язків використовуються різні методи: для зв'язків між атрибутивними ознаками — метод взаємної сполученості, для кількісно-варіюючих ознак — метод паралельних рядів, балансовий метод, графічний метод, метод аналітичних групувань, кореляційно-регресійний аналіз. Вони будуть розглянуті в наступних параграфах. Для вивчення багатовимірних статистичних сукупностей застосовують кореляцію, регресію, дисперсійний та факторний аналіз. Вирішення цього питання про форму зв'язку спирається на знання питання, логічний аналіз, попередні дослідження та на статистичні дані. Допомога у встановленні зв'язків надають кореляційні таблиці, графіки.

Багаторічна практика застосування цих методів для вивчення взаємозв'язків між суспільними явищами свідчить про їхню високу ефективність. Найбільшого поширення в аналізі набули методи парного та множинного кореляційного аналізу.

## 9.3. Метод порівняння паралельних рядів

Перейдемо тепер до розгляду методів вивчення кореляційних зв'язків між кількісними ознаками. Для встановлення кореляційного зв'язку між ознаками використовуються так звані елементарні прийоми (паралельне зіставлення рядів значень результативної і факторної ознак, побудова групової, кореляційної та комбінаційної таблиць), а також метод дисперсійного аналізу.

Найпростішим прийомом встановлення зв'язку є *порівняння* двох *паралельних рядів* — рядів значень аргументу і відповідних йому значень функції. Метод зіставлення паралельних рядів полягає у порівнянні двох чи кількох показників. Таке порівняння проводиться після того, як теоретично доведено можливість зв'язку між досліджуваними показниками. Для цього значення аргументу розташовують у зростаючому або спадному порядку і паралельно записуються значення результативної ознаки. Потім простежують, як змінюються цікаві для нас ознаки: зростають вони або спадають і в якій мірі, настільки швидко відбуваються ці процеси і т. п. За характером змін значень функції роблять висновок про наявність (якщо зміна величин одного ряду слідує за змінами величин іншого ряду) або про відсутність (якщо жодної твердої, стійкої відповідності в їх змінах немає) зв'язку. Розташувавши таким чином значення ознаки-фактора і результативної ознаки, виявляються існування зв'язку та її напрямок. Результативну ознаку (функцію) надалі позначатимемо через $y$, факторну ознаку — через $x$.

Таке порівняння, здійснене після того, як теоретично доведена можливість зв'язку між досліджуваними показниками, дозволяє встановити наявність зв'язку і отримати уявлення про її характер.

Наприклад, візьмемо дані по 20 заводах однієї галузі промисловості про рівень машиноозброєності праці ($x$ — тис. грн на одного працюючого) і про рівень продуктивності праці ($y$ — виробіток деталей на одного робочого). Дані ці зведені в табл. 9.3, де заводи розташовані в порядку зростання машиноозброєності.

Таблиця 9.3

**Залежність виробітку на 1 робітника від зміни машиноозброєності**

| № п/п | Машино-озброєність, тис. грн/осіб $x$ | Виробіток деталей на 1 робітника за місяць, шт. $y$ | № п/п | Машино-озброєність, тис. грн/осіб $x$ | Виробіток деталей на 1 робітника за місяць, шт. $y$ |
|---|---|---|---|---|---|
| 1 | 5,4 | 700 | 13 | 10,8 | 1 280 |
| 2 | 5,8 | 750 | 14 | 11,5 | 1 300 |
| 3 | 6,0 | 830 | 15 | 12,9 | 1 220 |
| 4 | 6,4 | 910 | 16 | 13,0 | 1 350 |
| 5 | 7,1 | 980 | 17 | 14,7 | 1 290 |
| 6 | 8,2 | 1 010 | 18 | 15,4 | 1 380 |
| 7 | 8,8 | 1 160 | 19 | 16,2 | 1 350 |
| 8 | 9,0 | 1 080 | 20 | 17,4 | 1 460 |
| 9 | 9,3 | 1 150 | 21 | 17,8 | 1 440 |
| 10 | 9,6 | 1 070 | 22 | 18,1 | 1 520 |
| 11 | 9,9 | 1 200 | 23 | 18,2 | 1 560 |
| 12 | 10,5 | 1 150 | 24 | 19,4 | 1 680 |

Просте зіставлення показників цієї таблиці дає можливість встановити, що між машиноозброєністю праці і продуктивністю праці є певний зв'язок. Зі збільшенням ознаки $x$ — машиноозброєності праці зростає (щоправда, не завжди) результативна ознака $y$ — продуктивність праці. Отже, між $x$ і $y$ є пряма залежність, хоч і не повна, але виражена досить ясно. У кожному окремому випадку значення середньої продуктивності праці залежатиме не тільки від рівня машиноозброєності праці, а й від того, як вплинуть інші фактори, що визначають рівень середньої продуктивності праці. У тих випадках, коли зростання величини факторної ознаки веде за собою і зростання результативної ознаки, говорять про *прямий кореляційний зв'язок*. Якщо зі зростанням значень факторної ознаки відбувається зменшення значень результативного ознаки, то зв'язок між ознаками вважається *зворотним*.

Потрібно особливо відзначити, що характер зміни ознак у паралельних рядах не завжди може виражати наявність дійсного зв'язку між досліджуваними ознаками. Звичайно, залежність між явищами, як правило, завжди має виражатися кількісно у відомій узгодженості у змінах зіставних ознак. Але вона може виникнути і при простому співіснуванні цих явищ і не відображати певною мірою причинно-наслідкових зв'язків між ними. Встановлення зв'язку між ознаками використовують методи теоретичного аналізу. Завдання математико-статичних методів полягає в кількісному вимірі зв'язків між ознаками.

Цей метод набув найбільшого поширення під час аналізу невеликої кількості одиниць, що складають досліджувану сукупність. Однак наявність великої кількості значень ознаки-фактора ускладнює сприйняття методу порівняння паралельних рядів.

### 9.4. Балансовий метод дослідження взаємозв'язків

У статистиці набув широкого поширення балансовий метод як інструмент вивчення взаємозв'язків між суспільними явищами і процесами. В даний час балансовий метод отримав найбільше поширення як засіб забезпечення необхідної пропорційності в народному господарстві.

Суть *балансового методу* полягає у побудові балансів — таблиць, у яких результат однієї частини дорівнює підсумку іншої. При застосуванні балансового методу зв'язок між окремими показниками виражається у вигляді рівності цілої відповідної сумі частин або рівності збільшення відповідної різниці «приходу» та «витрат». Балансова рівність у цій таблиці здійснюється рівністю суми підсумків горизонтальних рядків і суми підсумків вертикальних граф. За допомогою балансового методу можна виявити не тільки економічні зв'язки та пропорції у народному господарстві, а й розкрити наявні диспропорції. Цей метод аналізу отримав назву балансового тому, що історично першим прикладом ув'язки великої кількості показників господарської діяльності зазначеним способом був бухгалтерський баланс.

Відмінною особливістю балансового методу і те, що балансові зв'язки кількісно відбивають суворо встановлену взаємозалежність окремих елементів. Зміна одного елемента в балансі неминуче викликає зміну іншого.

У найбільш загальній формі такого роду баланси можуть бути представлені у вигляді рівності:

$$O_{\text{н}} + П = P + O_{\text{к}}, \tag{9.4}$$

Стосовно матеріальних балансів окремі члени наведеної рівності означають:

де $O_{\text{н}}$ — залишки ресурсів на початок року;
$П$ — надходження ресурсів у досліджуваному періоді;
$P$ — використання (розподіл, витрата) ресурсів у досліджуваному періоді;
$O_{\text{к}}$ — кінцевий залишок ресурсів, що переходять на наступний рік.

В основі всіх балансових побудов лежать балансові рівняння, які виражаються у формі підсумків, отриманих в результаті різних математичних дій над цими показниками. Балансові рівняння характеризують внутрішню єдність і взаємозалежність елементів суспільного виробництва та розподілу. Встановлювана таким шляхом рівність (баланс) підсумків показує, що враховані всі взаємодіючі чинники і відповідні їм узагальнюючі показники і, отже, зв'язок між ними подано правильно.

Розробка балансу провадиться в наступному порядку. Спочатку виявляється потреба у окремих видах продукції до різних призначень. Потім визначаються матеріальні ресурси, що надходять із виробництва та інших джерел; при цьому розробляються заходи щодо збільшення ресурсів як за рахунок виробництва, так і за рахунок залишків готової продукції на початок та кінець запланованого періоду. Надалі виробляється взаємна ув'язка потреби з ресурсами.

Баланси сировини та основних матеріалів складаються за прикладною схемою, показаною в табл. 9.4.

У першій частині матеріального балансу «*Ресурси*» включаються: залишки продукції початку планованого періоду; надходження із закупівель у плановому році; виробництво продукції розмірах, передбачених планом (фактично); імпорт, виходячи з існуючих договірних угод; інші надходження, що утворюють, наприклад, за рахунок дострокової доставки.

Таблиця 9.4

**Приблизна схема балансу сировини та матеріалів**

| Ресурси | Кіль-кість | Розподіл (потреби) | Кіль-кість |
|---|---|---|---|
| Ресурси — всього<br>В тому числі:<br>Наявність сировини (напівфабрикатів) на початок року<br>Надходження по закупкам у запланованому році<br>Виробництво<br>Імпорт<br>Інші надходження | | Розподіл — всього<br>В тому числі:<br>Ремонтно-експлуатаційні потреби<br>Витрата сировини на вироблення продукції<br>Втрати сировини при транспортуванні та зберіганні<br>Капітальне будівництво<br>Інші витрати<br>Залишок на кінець року | |

Другим розділом баланси сировини та матеріалів є *витратна частина* (*розподіл*). У ній передбачаються статті, що відображають основні напрямки розподілу матеріально-технічних ресурсів. Наприклад, виділяється продукція на ремонтно-експлуатаційні потреби, виробництво, будівництво, інші витрати і т. п. Крім того, у витратній частині враховуються втрати продукції при транспортуванні та зберіганні.

Баланси поділяються на три підгрупи: матеріальні баланси; баланси праці; фінансові баланси. Принципової різниці між цими підгрупами немає. Загальна схема побудови балансів однакова. Відмінності викликаються переліком статей збільшення ресурсів та їх використання та прийнятими у балансових зіставленнях вимірниками досліджуваних показників.

Широкі можливості виявлення пропорцій і взаємозв'язків, що складуються у процесі відтворення, відкриваються під час використання балансового методу на рівні сукупності підприємств, галузі, економічного району, нарешті, для держави загалом. В цьому випадку за допомогою балансового методу можна дослідити взаємозв'язки та основні пропорції в утворенні та розподілі ресурсів між окремими підприємствами, економічними районами, галузями народного господарства. Такі баланси дозволяють аналізувати пропорції та зрушення в процесі розширеного відтворення. Для аналізу міжгалузевих зв'язків і пропорцій у надходженні та використанні ресурсів між галузями будують *міжгалузеві баланси*.

Балансові таблиці, що аналізують відтворювальні процеси на рівні народного господарства країни, зазвичай являють собою дуже складні побудови, що складаються з цілої системи статистичних таблиць. Як було зазначено, балансовим шляхом вивчають рух (обіг) як матеріальних ресурсів, а й робочої сили в, фінансів тощо.

*Баланс трудових ресурсів* характеризує наявність і зайнятість робочої сили за видами економічної діяльності, джерела забезпечення додаткової потреби в робочій силі та необхідні територіальні переміщення трудових ресурсів. Баланс складають у масштабі народного господарства країни та в межах областей. Призначення балансу трудових ресурсів полягає в тому, щоб визначити найважливіші темпи і пропорції відтворення трудових ресурсів, їх розподіл та використання за сферами діяльності та галузям народного господарства. Склад показників балансу дає можливість визначити пропорції розподілу трудових ресурсів між виробничою та невиробничою сферами, між видами економічної діяльності.

*Баланс грошових доходів і витрат населення* включає систему показників, що відображають рух тієї частини національного доходу, яка розподіляється між окремими членами суспільства в грошовій формі. За допомогою цього балансу визначають співвідношення між платоспроможним попитом населення та його товарним забезпеченням. Дані балансу застосовують для планування товарообігу, платежів і заощаджень населення, для планування загалом грошового звернення. Схема балансу і двох частин: доходів і витрат. Утворення коштів у населення показано в доходній частині за основними напрямами: оплата праці; доходи від підприємницької діяльності та самозайнятості; пенсії, стипендії, соціальна допомога, інші надходження (відсотки за вкладами, виплати за позиками та лотереями). Витрати грошей показано за основними напрямами їх використання: на продукти харчування; на предмети домашнього споживання, побутову техніку та поточний утримання житла; на охорону здоров'я; транспорт; зв'язок; на оплату товарів та послуг у порядку обміну між групами населення. Крім цього, у показниках знаходить відображення величина зміни залишків коштів у населення, облік якої має важливе значення для планування грошового обігу.

Найбільш повний розвиток балансовий метод отримує в *балансі народного господарства* — системи науково обґрунтованих і взаємопов'язаних показників, що відображають масштаби і темпи розширеного відтворення, що показують основні пропорції розвитку народного господарства країни.

Міжгалузевий баланс дає можливість в цілому дослідити, як із руху продукції окремих галузей утворюється складний процес розширеного відтворення сукупного суспільного продукту. З його допомогою можна визначити коефіцієнти прямих і повних витрат продукції, співвідношення між двома підрозділами суспільного виробництва, між виробничим і невиробничим споживанням, споживанням і накопиченням, усіма елементами вартості сукупного суспільного продукту. Докладно баланс народного господарства вивчається в курсі економічної статистики.

В даний час українська статистика розробила цілу систему балансів, що дають докладну характеристику відтворення суспільного продукту, систему кількісних та якісних зв'язків у народному господарстві.

## 9.5. Графічний метод виявлення кореляційних залежностей

Для попереднього виявлення наявності зв'язку можуть допомогти графіки. Графічне зображення досліджуваних явищ дозволяє не тільки встановити наявність або відсутність зв'язку між явищами, а й розкрити характер цього зв'язку, інакше кажучи, вивчити форму зв'язку та його тісноту. Графічний метод вивчення зв'язку зазвичай застосовують разом із іншими методами. Найчастіше графічно зображують результати аналітичних групувань. Зображення форми зв'язків може здійснюватися кореляційним полем або емпіричної лінії зв'язку (регресії).

*Кореляційне поле* — точковий графік у прямокутній системі координат; на осі абсцис відкладається масштаб однієї ознаки ($x$), на осі ординат — іншої ($y$). На полі наносяться точки з координатами ($x$, $y$), які відповідають певним значенням факторної та результативної ознак у окремих одиниць, що піддалися спостереженню.

Відсутність зв'язку між факторною і результативною ознаками буде відображено безладним розташуванням точок на графіку. Чим сильніший зв'язок між цими ознаками, тим точніше точки будуть групуватися навколо певної лінії, що виражає форму зв'язку (прямолінійна, параболічна та ін.). На рис. 9.1 наведено три види залежності факторної та результативної ознак.

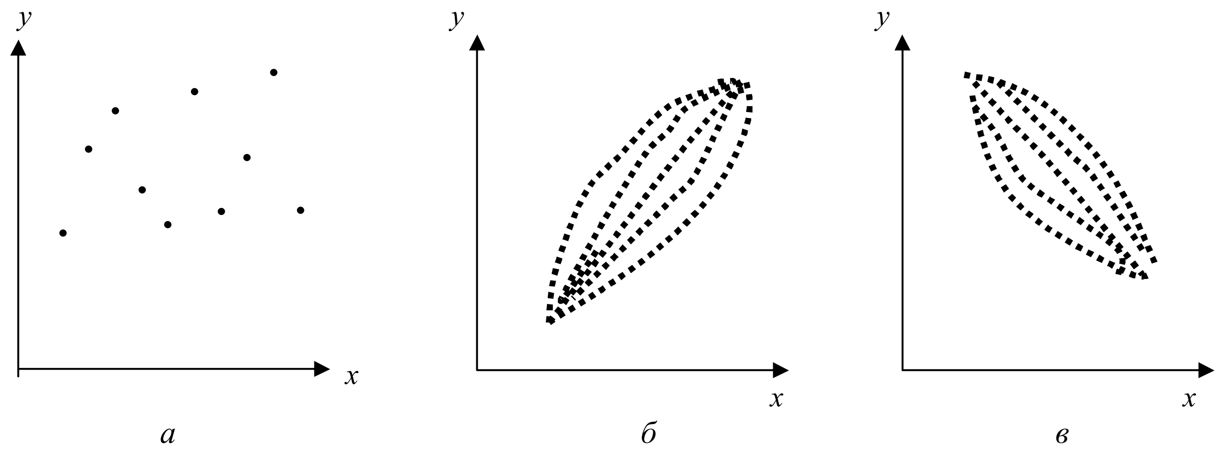



*Рис. 9.1.* Розподіл кореляційного поля за різного виду залежностей:
а — залежності між *y* і *x* немає; б — залежності між *y* і *x* пряма;
в — залежність між *y* і *x* зворотна.

Маючи перед собою числові характеристики факторіального та результативного ознак того самого явища, можна кожну пару чисел зобразити у вигляді точки на площині.

Звернемося до прикладу залежності вироблення продукції від виробничого стажу за групою 10 робочих (див. табл. 3.11). Побудуємо кореляційне поле в прямокутній системі координат, де по осі абсцис у певному масштабі нанесемо фактичні значення факторної ознаки (стажу), по осі ординат — результативної ознаки (змінного виробітку продукції). На перетині абсцис та ординат, що відповідають конкретним значенням ознаки, відзначимо точки. Для нашого прикладу поле кореляції має такий вигляд (рис. 9.2).

Положення кожної точки на графіку визначається величиною двох ознак — стажем роботи та відповідним йому середнім виробітком робітника. Точки кореляційного поля не лежать на одній лінії, але витягнуті певною смугою зліва направо. З’єднуючи послідовно відрізками прямих нанесені на графік точки, отримаємо так звану *емпіричну лінію зв’язку*. Якщо емпірична лінія зв’язку за своїм виглядом наближається до прямої, то можна припустити про наявність прямолінійної кореляції. Якщо ж є тенденція нерівномірної зміни значень результативної ознаки, то має місце криволінійний кореляційний зв’язок.

На рисунку розташування точок наочно підтверджує прямий зв’язок між стажем і змінним виробленням робітників, тобто, з під-

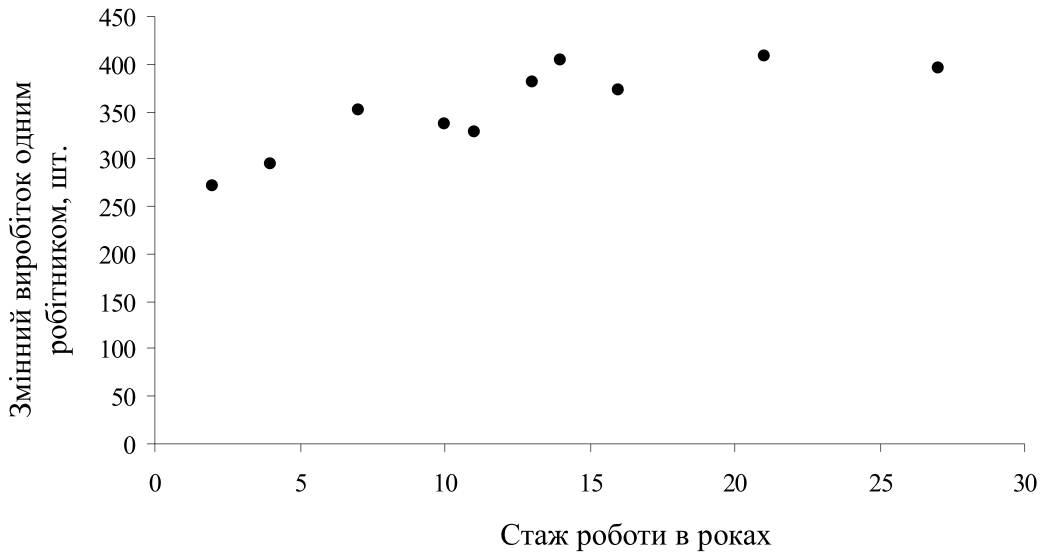



*Рис. 9.2.* Кореляційне поле залежності змінного виробітку від стажу по 10 робітникам.

вищенням стажу збільшується змінний виробіток робітників.

Кореляційне поле використовується для зображення зв'язків усередині невеликого обсягу.

*Емпірична лінія зв'язку* (*регресії*) будується зображення зв'язків за сукупностями, розчленованим на групи. Тут факторна ознака зазвичай береться в інтервалах, результативна у вигляді середніх або відносних показників, обчислених за групами всередині сукупностей. Про відмінність у трудомісткості продукції ливарного виробництва залежно від потужності ливарних цехів свідчать дані табл. 9.5.

Таблиця 9.5

**Вплив концентрації ливарного виробництва на трудомісткість продукції**

| Групи ливарних цехів за потужністю, *т* | Обсяг лиття, *т* | Загальні витрати праці на виробництво продукції, тис. люд.-год. | Середня трудомісткість тонни лиття (у людино-годинах) |
|---|---|---|---|
| А | 1 | 2 | 3 (2:1) |
| До 200 | 167 | 50,7 | 304 |
| 201–300 | 254 | 57,6 | 227 |
| 301–400 | 376 | 47,4 | 126 |
| 401–500 | 468 | 39,3 | 84 |
| 501–600 | 546 | 26,7 | 49 |
| Всього | 1 811 | 221,7 | 122 |

Зіставлення граф А і 3 чітко показує зворотний зв'язок між потужністю ливарних цехів та трудомісткістю тонни лиття.

Побудуємо графік зв'язку між потужністю ливарних цехів і трудомісткістю лиття. Якщо, наприклад, по осі $x$ у системі координат відкласти значення обсягу лиття (факторної ознаки), а по осі $y$ — середньої трудомісткості лиття (значення результативної ознаки), то зв'язок між факторним і результативним ознакою може бути виражений ламаною лінії. Відмінність полягає в тому, що по осі ординат відкладаються середні витрати праці на тонну лиття за групами ливарних цехів, по осі абсцис — середини інтервалів. Будемо мати ряд точок, з'єднавши які отримуємо ламану лінію. Отримана ламана крива є емпіричною *лінією* (*ламаною*) *регресії* (рис. 9.3). Як видно з графіка, зі збільшенням обсягу лиття на заводах знижується середня трудомісткість виготовлення тонни лиття, що говорить про зв'язок між цими ознаками, причому зв'язок зворотній.

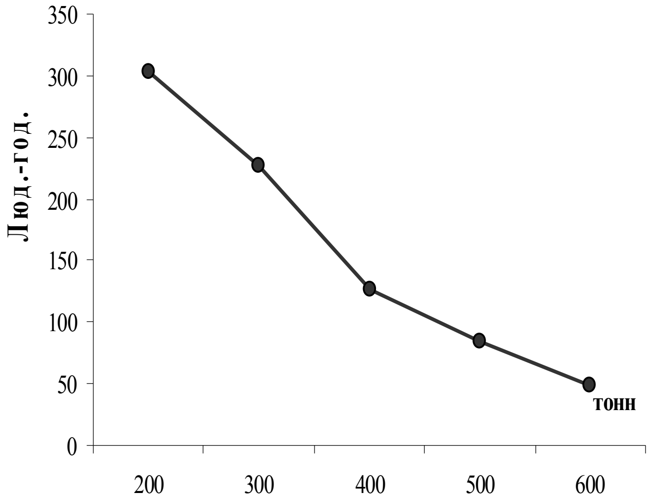



*Рис. 9.3.* Залежність середньої трудомісткості лиття від потужності ливарних цехів.

Графік підкреслює ту залежність напрямом ломаної кривої з верхнього лівого кута в нижній правий кут. Такого ж роду залежність будемо спостерігати на рис. 9.4, вивчаючи зв'язок між товарообігом і обсягом товарних запасів. Як і в попередньому прикладі, факторну ознаку — величину товарообігу — відкладатимемо на осі абсцис, а результативну — товарні запаси на осі ординат. Тут також ясно виражена зворотна залежність між факторною і результативною ознаками.

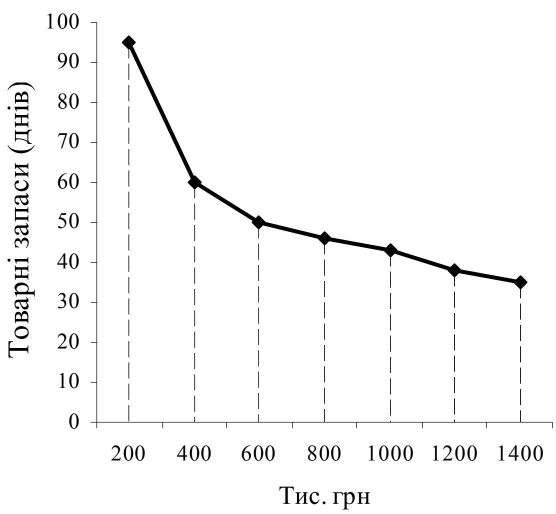



*Рис. 9.4.* Залежність товарних запасів магазинів від їх товарообороту.

По-іншому виглядатиме графік залежності товарообігу від числа робочих місць (рис. 9.5). Тут ми спостерігаємо яскраво виражений прямий зв'язок між двома ознаками: збільшення кількості робочих місць супроводжується зростанням розміру товарообігу магазину.

Графічний метод, як і метод зіставлення паралельних рядів, дозволяє судити про наявність зв'язку, про форму її (пряма або зворотна), приблизно характеризує силу зв'язку. Він наочно ілюструє залежність, виявлену групуванням.

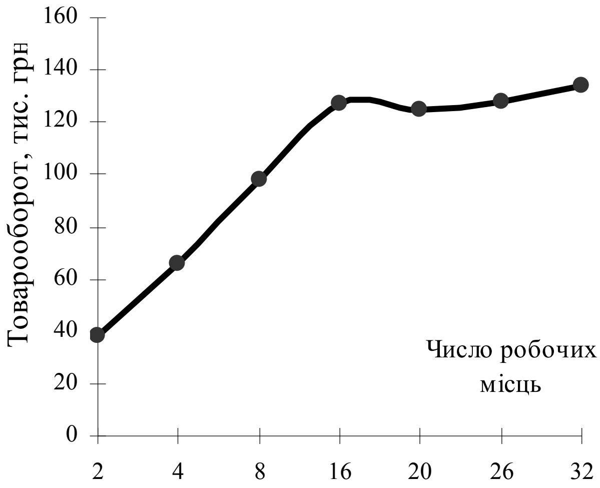



*Рис. 9.5*. Залежність товарообігу для 1 працівниказа квартал від числа робочих місць.

Однак графічний метод не дозволяє дати відповідь на дуже важливе питання — наскільки тісно пов'язані між собою факторна і результативна ознаки, яка кількісна характеристика цього зв'язку. Недолік графічного методу вивчення зв'язку полягає ще в тому, що він дозволяє виявити зв'язок лише між двома ознаками.

Для того, щоб встановити характер зв'язку між явищами, виміряти його тісноту і дати кількісне вираз цього зв'язку, у статистиці використовують кореляційний метод.

## 9.6. Вимірювання тісноти зв'язку між атрибутивними (якісними) ознаками

При дослідженні соціально-економічних явищ і процесів велике значення має вивчення якісних показників та ознак, що не мають кількісної оцінки. Для вимірювання кореляційних зв'язків між атрибутивними (якісними) ознаками застосовується ряд показників. У принципі цей вимір грунтується на спостереженні частоти спільної появи атрибутивних ознак: зв'язок між ними вважається тим сильнішим, чим вище відносна частота такої появи.

**Коефіцієнти взаємного узгодження Пірсона та Чупрова**. Для вимірювання тісноти зв'язку між атрибутивними ознаками, кожна з яких складається більш ніж з двох груп, застосовується коефіцієнти взаємного узгодження.

Для розрахунку коефіцієнтів взаємного узгодження найчастіше застосовуються формули, запропоновані відомим англійським статистиком К. Пірсоном і видним російським статистиком А. А. Чупровим.

*Коефіцієнт взаємного узгодження А. А. Чупрова* ($K_{\text{сопр. Чупр}}$) обчислюється за такою формулою:

$$K_{\text{сопр.Чупр}} = \sqrt{\frac{\varphi^2}{(m_1 - 1)(m_2 - 1)}}, \tag{9.5}$$

де $m_1$ — позначає число можливих значень першої статистичної величини (кількість груп);

$m_2$ — число можливих значень другої статистичної величини;

$\varphi^2$ — показник взаємної узгодженості.

Величина $\varphi^2$, що входить у цю формулу, визначається як сума відносин квадратів частот кожної клітини таблиці розподілу до добутку підсумкових частот відповідного стовпця і рядка. Віднімаючи з цієї суми одиницю, отримаємо величину $\varphi^2$.

*Коефіцієнт взаємної узгодженості К. Пірсона* ($K_{\text{сопр. Пирс}}$) розраховується за такою формулою:

$$K_{\text{сопр. Пирс}} = \sqrt{\frac{\varphi^2}{1 + \varphi^2}}, \tag{9.6}$$

де $\varphi^2$ — має однакове значення з $\varphi^2$ А. Чупрова і є показником узгодженості.

Розглянемо розрахунок коефіцієнта взаємної узгодженості за даними завдання (табл. 9.6).

Таблиця 9.6

| Урожай | Полив | | | |
|---|---|---|---|---|
| | Слабий | Середній | Добрий | Всього |
| Низький | 28<br>(784)<br>8,000 | 41<br>(1681)<br>6,975 | 4<br>(16)<br>0,158 | 73<br>–<br>15,133; 0,207 |
| Середній | 52<br>(2704)<br>27,592 | 148<br>(21904)<br>90,888 | 46<br>(2116)<br>20,950 | 442<br>–<br>139,430; 0,567 |
| Високий | 18<br>(324)<br>3,306 | 52<br>(2704)<br>11,220 | 51<br>(2601)<br>25,752 | 161<br>–<br>40,278; 0,333 |
| Всього | 98 | 241 | 101 | 440; 1,107 |

У кожній клітині таблиці записані частоти, їх квадрати, квадрати частот, поділені на суму частот по стовпцю. У підсумкових стовпцях записані суми частот, сума результатів ділення, а також ділення нижнього числа на верхнє.

Техніка розрахунку $\varphi^2$ зводиться до такого. Спочатку у верхній частині кожної клітини цієї схеми записується частоти таблиці розподілу задачі (табл. 9.6). Частоти зводяться в квадрат, результат записується під ними в дужках. Потім зведені у квадрат частоти ділять на підсумки частот відповідного стовпця поділяються (784:98); частка від ділення проставляється в нижньому рядку кожної клітини.

Після цього числа, записані внизу кожної клітини, сумуються для кожного рядка, суми записуються в нижній частині клітин підсумкового стовпця (8,000 +6,975 +0,158 = 15,133). Зазначені в підсумковому стовпці суми діляться на відповідні підсумки частот, записаних у тих же клітинах (суми 15,133; 139,430; 40,278 діляться на відповідні підсумки рядків 73, 442, 161 та отримані частки 0,207, 0,567 і 0,333 записують в останньому стовпчику). Сума цих чисел за підсумковим стовпцем (1,107) без одиниці і являє собою $\varphi^2$:

$$\varphi^2 = 1{,}107 - 1{,}00 = 0{,}107. \qquad (9.7)$$

Число груп по стовпцях у нашій комбінаційній таблиці ($m_1$) складає 3, по рядках ($m_2$) — теж 3.

Підставляємо ці значення у наведені вище формули коефіцієнтів сполученості:

коефіцієнт Пірсона

$$K_{\text{сопр. Пирс}} = \sqrt{\frac{\varphi^2}{1+\varphi^2}} = \sqrt{\frac{0{,}107}{1+0{,}107}} = \sqrt{0{,}09666} \approx 0{,}311;$$

коефіцієнт Чупрова

$$K_{\text{сопр. Чупр}} = \sqrt{\frac{\varphi^2}{(m_1-1)(m_2-1)}} = \sqrt{\frac{0{,}107}{(3-1)(3-1)}} =$$

$$\sqrt{0{,}02675} \approx 0{,}163.$$

Результати обчислення коефіцієнта Пірсона свідчать про значний зв’язок між урожайністю та поливом посівів. Коефіцієнт взаємної узгодженості Чупрова є більш гнучким, оскільки він враховує число утворених за кожною ознакою груп $m_1$ і $m_2$. Тому результат 0,163 є точнішим порівняно з обчисленим коефіцієнтом взаємної узгодженості Пірсона.

**Коефіцієнт контингенції**. Для дослідження тісноти зв'язку двох атрибутивних ознак, кожна з яких складається тільки з двох груп, можливе використання так званих «тетрахорічних показників». Для їх обчислення дані зводять у комбінаційну чотириклітинну таблицю (таблиця чотирьох полів). Розрахункова таблиця складається з чотирьох ячеєк (що позначаються латинськими $a$, $b$, $c$ і $d$) і показує зв'язок між двома явищами, кожне з яких має бути альтернативним, тобто що складається з двох якісно відмінних один від одного ознак (наприклад, одружений або не одружений).

Таблиця 9.7

| Значення другої ознаки | Значення першої ознаки | | |
|---|---|---|---|
| | 1-е | 2-е | Всього |
| 1-е | $a$ | $b$ | $a+b$ |
| 2-е | $c$ | $d$ | $c+d$ |
| Всього | $a+c$ | $b+d$ | $a+b+c+d$ |

Тут $a$, $b$, $c$, $d$ — частоти співставлених ознак.

Можливості використання тетрахоричних показників дуже різноманітні. Наприклад, масова продукція, що підлягає експлуатації, може бути поділена на дві групи: надійна та ненадійна. Група робітників підприємства, що обстежується за станом здоров'я, може бути розділена на дві підгрупи: практично здорові і потребують лікування.

Для вимірювання тісноти зв'язку між альтернативними ознаками проводиться розрахунок *коефіцієнта контингенції*. Розрахунок здійснюється за наступною формулою:

$$K_к = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}. \tag{9.8}$$

де $a$, $b$, $c$ і $d$ — числа одиниць у чотириклітинній таблиці.

За своїм значенням він дорівнює коефіцієнту взаємної узгодженості. Коефіцієнт контингенції змінюється від –1 до +1, але завжди менше коефіцієнта асоціації. Зв'язок вважається суттєвим, якщо $K_к \geq 0{,}3$. За відсутності зв'язку $K_к$ дорівнює нулю. Від'ємна величина вказує на існування зворотного зв'язку між зображеними у чотириклітинній таблиці ознаками. Розглянемо приклад розрахунку коефіцієнта контингенції.

Наприклад, кількість населення, що прибуло і вибуло, можна розділити на дві групи в залежності від типу місцевості: міська місцевість і сільська місцевість. Результати обстеження представлені у табл. 9.8.

Таблиця 9.8

**Міграційний рух населення України за типом місцевості у 2002-2020 рр., осіб**

| Міграційний рух | 2002 р. | | | 2010 р. | | | 2020 р. | | |
|---|---|---|---|---|---|---|---|---|---|
| | міська місцевість | сільська місцевість | сума | міська місцевість | сільська місцевість | сума | міська місцевість | сільська місцевість | сума |
| Кількість прибулих | 69,8 | 30,2 | 100,0 | 70,1 | 29,9 | 100,0 | 70,6 | 29,4 | 100,0 |
| Кількість вибулих | 68,0 | 32,0 | 100,0 | 68,6 | 31,4 | 100,0 | 67,3 | 32,7 | 100,0 |
| Сума | 137,8 | 62,2 | – | 138,7 | 61,3 | – | 137,8 | 62,2 | – |

Обчислимо коефіцієнти контингенції:

для 2002 р. $$K_к = \frac{(69{,}8 \cdot 32{,}0) - (68{,}0 \cdot 30{,}2)}{\sqrt{100 \cdot 100 \cdot 137{,}8 \cdot 62{,}2}} =$$

$$= \frac{2233{,}6 - 2053{,}6}{\sqrt{85711600}} = \frac{180}{9258{,}1} = 0{,}019;$$

для 2010 р. $$K_к = \frac{(70{,}1 \cdot 31{,}4) - (68{,}6 \cdot 29{,}9)}{\sqrt{100 \cdot 100 \cdot 138{,}7 \cdot 61{,}3}} =$$

$$= \frac{2201{,}1 - 2051{,}1}{\sqrt{85023100}} = \frac{150}{9220{,}8} = 0{,}016;$$

для 2020 р. $$K_к = \frac{(70{,}6 \cdot 32{,}7) - (67{,}3 \cdot 29{,}4)}{\sqrt{100 \cdot 100 \cdot 137{,}8 \cdot 62{,}2}} =$$

$$= \frac{2308{,}6 - 1978{,}6}{\sqrt{85711600}} = \frac{330}{9258{,}1} = 0{,}036.$$

Як бачимо, у нашому прикладі коефіцієнт контингенції є досить незначним, хоча у 2020 р. він зріс більш ніж у 2 рази. Це свідчить про те, що зв'язку між міграційним рухом і типом населення не має.

**Коефіцієнт асоціації**. Якщо варіація обох атрибутивних ознак обмежена двома групами (тобто має альтернативний характер), то коефіцієнт взаємної узгодженості може бути обчислений значно простіше. Як метод для встановлення зв'язку між атрибутивними ознаками Юл запропонував обчислення коефіцієнта асоціації.

*Коефіцієнт асоціації* — показник оцінки тісноти зв'язку між двома альтернативними ознаками. Коефіцієнт асоціації обчислюють за формулою:

$$K_a = \frac{ad - bc}{ad + bc}, \tag{9.9}$$

де $a$, $b$, $c$ і $d$ — число одиниць одноразової появи альтернативних ознак.

Якщо коефіцієнт асоціації $K_a \geq 0{,}5$, можна говорити про наявність суттєвого зв'язку між ознаками. Близькість коефіцієнта асоціації до одиниці свідчить про тісний позитивний зв'язок. У порівнянні з іншими показниками виміру асоціативного зв'язку цей коефіцієнт обчислюється більш простим способом.

Для обчислення коефіцієнта асоціації будується чотириклітинна таблиця кореляції, яка виражає зв'язок між двома явищами, кожне з них у свою чергу має бути лише *альтернативним*, тобто що складається тільки з двох видів, якісно відмінних один від одного.

Розглянемо розрахунок коефіцієнта асоціації на наступному прикладі. Так, досліджувався зв'язок між виконанням норм виробітки робітниками та проходженням ними підвищення кваліфікації. При вивченні залежності норм виробітку від проходженням робітниками підвищення кваліфікації виділяємо з проходження підвищення кваліфікації та виконання норм лише по дві групи. При цій умові можна побудувати наступну чотириклітинну таблицю (табл. 9.9).

Таблиця 9.9

| Групи робітників | Виконують норми | Не виконують норми | Всього |
|---|---|---|---|
| Пройшли підвищення кваліфікації | 82 | 18 | 100 |
| Не пройшли підвищення кваліфікації | 36 | 64 | 100 |
| Всього | 118 | 82 | 200 |

Числа, які стоять на перетині рядків і граф, показують, скільки робітників зустрічається з тим чи іншим рівнем виконання норм, що пройшли або не пройшли підвищення кваліфікації.

Розрахуємо коефіцієнт асоціації:

$$K_a = \frac{82 \cdot 64 - 36 \cdot 18}{82 \cdot 64 + 36 \cdot 18} = \frac{4600}{5896} = 0{,}78.$$

Таким чином, між рівнем продуктивності праці робітників і проходженням ними підвищення кваліфікації існує тісний прямий взаємозв’язок.

Розглянемо ще один приклад. Нехай, наприклад, потрібно виміряти зв’язок між щепленнями проти кору і зниженою захворюваністю на кір у щеплених дітей в одному з районів (табл. 9.10). У нашому прикладі $a$ означає число хворих на кір із числа щеплених, тобто 61; $b$ — число незахворілих щеплених, що дорівнює 6759; (6809–61); $c$ — число захворілих нещеплених, тобто 344–61=282; $d$ — число незахворілих нещеплених, рівне (18517 – 6809) – 283 = 11425.

Таблиця 9.10

| | Заболіло | Не заболіло |
|---|---|---|
| Привиті | 61 | 6 748 |
| Не привиті | 283 | 11 425 |
| Сума | 344 | 18 173 |

Отже, коефіцієнт асоціації становитиме:

$$K_a = \frac{61 \cdot 11425 - 283 \cdot 6748}{61 \cdot 11425 + 283 \cdot 6748} = \frac{-1\,212\,759}{2\,606\,609} = -\,0{,}47.$$

Отримане значення коефіцієнта асоціації досить високе, що доводить вельми тісний зв’язок між щепленнями ($x$) і числом хворих ($y$). Знак мінус вказує на зворотну залежність, тобто зі збільшенням $x$ (числа щеплень) $y$ (кількість хворих) зменшується, і, навпаки, зі зменшенням $x$ відбувається збільшення $y$. Величина коефіцієнта нашому прикладі відповідає середньому розміру зв’язку. Отже, в даному випадку між захворюваністю на кір і щепленнями проти цієї хвороби мала місце зворотна кореляція середнього розміру.

## 9.7. Метод аналітичних групувань

Найбільш простим методом вивчення та кількісної оцінки взаємозв’язків між ознаками є метод групувань, що дозволяє виявити наявність зв’язків шляхом порівняння не індивідуальних, а групових середніх. З таким способом ми вже познайомилися в гл. 3.

Групування дозволяють встановити наявність або відсутність залежності досліджуваного явища від факторів, що здійснюють на нього різний за силою вплив. Однак за допомогою цього методу можна лише встановити загальний напрямок взаємозв’язків між факторами, її тенденцію. Наприклад, групування машинобудівних підприємств дає можливість встановити, що зі збільшенням обсягу виробництва знижується собівартість одиниці виробу. Але як відбувається це зниження, в яких розмірах, наскільки і до якої величини відповісти на ці питання за допомогою групувань складно.

За допомогою статистичних групувань можна отримати охарактеризувати загальні зв'язки між типами явищ (типологічні групування), визначити ступінь впливу одного або більше факторів на результативну ознаку (факторні групування). За допомогою комбінаційних групувань можна виявити зв'язки між результативною ознакою і двома і більше факторами. На основі групувань можна встановити загальний напрямок зв'язку та отримати наближене уявлення про його форму. Велика роль належить групуванням щодо впливу якісного ознаки на показники варіації кількісних ознак.

Основна складність при проведенні групування полягає в тому, щоб утворити таку кількість груп, при якій у варіації середніх в максимальній мірі позначився б вплив групувальної ознаки. Якщо при групуванні візьмемо невелику кількіст груп, в велика ймовірність того, що в групових середніх буде погашена поряд із варіацією, обумовленою іншими факторами, і частина варіації, що обумовлена групувальною ознакою. В той же час не можна значно збільшувати кількість груп, особливо при невеликій кількості спостережень; в цьому випадку висока ймовірність того, що групові середні не будуть характеризувати типові риси виділеної групи, а у значній мірі будуть визначатися випадковими обставинами. Тому при визначенні кількості груп необхідно дотримуватися загального правила: *чим більше груп ми створюємо, тим більше збільшується міжгрупова варіація*.

Основний принцип виявлення зв'язків за допомогою методу групування полягає в тому, що в якості групувальної ознаки зазвичай вибирають факторну. У присудку таблиці розміщують середні (абсолютні або відносні) значення одного або кількох результативних ознак. Зміна факторної ознаки спричиняє і зміни результативної ознаки. З цих змін можна будувати висновки про ступеня взаємодії чинників на результативний показник.

При проведенні групування виходять із того, що фактор-ознака, прийнятий в основі групування, береться ізольовано і досліджується вплив на варіацію іншої ознаки. Варіація результативної ознаки проявляється у групових середніх. Ця варіація (міжгрупова варіація) відображає лише вплив фактора, покладеного в основу групування. Інша частина варіації (внутрігрупова) відображає вплив інших факторів. Отже, потрібно вибрати золоту середину: оптимальне число груп для даного конкретного випадку, коли групові середні перестануть носити випадковий характер у той же час групувальна ознака проявить себе повною мірою. Корисно при цьому керуватися вказівкою А. А. Чупрова: «Чим більше груп ми в змозі нарізати, не наштовхуючись на жодний виняток, тим міцніше висновок, що знайдений зв'язок або помічена відсутність зв'язку не випадкові і свідчать про дійсні взаємовідносини між досліджуваними ознаками» (див. А. Чупров, 1960, с. 11).

Використовуючи дані табл. 9.3, проведемо групування заводів за ознакою машиноозброєності та обчислимо в кожній групі середні. При дослідженні впливу озброєності праці на його продуктивність як факторна ознака фігурує рівень машиноозброєності праці, як результативний — вироблення продукції з розрахунку на 1 робітника. У результаті розрахунків буде наочно видно, як змінюється продуктивність робітників у залежності від рівня машиноозброєності праці (табл. 9.11).

Таблиця 9.11

**Порівняння трьох групувань 24 заводів за рівнем машиноозброєності праці (*x*)**

| Перше групування (4 групи) | | | | Друге групування (5 груп) | | | | Третє групування (6 груп) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| групи по *x* | *n* | *y* | $y_{гр}$ | групи по *x* | *n* | *y* | $y_{гр}$ | групи по *x* | *n* | *y* | $y_{гр}$ |
| 5,4–8,89 | 7 | 6340 | 905,7 | 5,4–8,19 | 5 | 4170 | 834,0 | 5,4–7,72 | 5 | 4170 | 834,0 |
| 8,9–12,39 | 7 | 8230 | 1175,7 | 8,2–10,99 | 8 | 9100 | 1137,5 | 7,73–10,05 | 6 | 6670 | 1111,7 |
| 12,4–15,89 | 4 | 5290 | 1322,5 | 11,0–13,79 | 3 | 3870 | 1290,0 | 10,06–12,38 | 3 | 3730 | 1243,3 |
| 15,9–19,4 | 6 | 8830 | 1471,7 | 13,8–16,59 | 3 | 3890 | 1296,7 | 12,39–14,71 | 3 | 3910 | 1303,3 |
| | | | | 16,6–19,4 | 5 | 7660 | 1532,0 | 14,71–17,03 | 2 | 2550 | 1275,0 |
| | | | | | | | | 17,03–19,4 | 5 | 7660 | 1532,0 |
| Всього | 24 | 28690 | 1195,4 | Всього | 24 | 28690 | 1195,4 | Всього | 24 | 28690 | 1195,4 |

Але якщо потрібно було б виявити залежність між явищами не на прикладі 24 заводів, а на прикладі сотень і тисяч одиниць спостере ження, як це зазвичай буває при проведенні статистичного спостереження. У таких випадках для виявлення залежностей між явищами застосування аналітичних групувань є єдино можливим.

Для цього проводиться розрахунок всіх підприємств за рівнем машиноозброєності і в кожній отриманій групі підприємств розраховується середня продуктивність праці (вироблення деталей на 1 робітника в одиницю часу). У результаті виходить поєднання двох ознак: машиноозброєності праці та продуктивності праці. Якщо поряд зі збільшенням машиноозброєності праці при переході від однієї групової середньої до іншої виявляється якась тенденція, то ми говоримо про зв'язок машиновооруженості праці з продуктивністю праці. Якщо ж тенденція у зміні групових середніх немає, або зв'язок між ознаками, що вивчаються, відсутній, або групування цей зв'язок не вловила.

У першому групуванні прийнято 4 інтервали, інтервал при цьому дорівнює 3,4. У другому угрупованні — 5 груп, інтервал дорівнює 2,8; у третьому угрупованні — 6 груп, інтервал дорівнює 2,33. Розглянемо, як змінюються групові середні ($y_{гр}$). У першій і в другій групах умова А. А. Чупрова витримується: у групах немає жодного виключення у тенденціях зміни групової середньої. У третій групі ця умова порушується: у п'ятій групі середня (1275,0) проти четвертої групою (1303,3) не зростає, а зменшується. Отже, тут порушується умова А. А. Чупрова. Як бачимо, оптимальними є 1-а і 2-а групи.

Аналітичні групування характеризують лише загальні риси досліджуваної сукупності, напрямок її розвитку і не дозволяють кількісно виміряти залежність результативної ознаки від одного або декелькох факторів. Але з урахуванням аналітичних угруповань це завдання можна вирішити з урахуванням розрахунку емпіричного кореляційного отношения.

**Емпіричне кореляційне відношення**. Зробимо розрахунок групових середніх ($\delta^{l}$), приймаючи за основу групування на п'ять груп (табл. 9.12).

Таблиця 9.12

**Розрахунок дисперсій групових середніх рівнів машиноозброєності праці**

| Групи підприємств за рівнем машиозброєності праці (тис. грн за місяць на одного робітника) | Число підприємств ($n$) | Виробіток деталей на 1 робітника для всіх підприємств, шт. | В середньому на одне підприємство ($y_{гр}$) | $y_{гр} - \bar{y}$ | $(y_{гр} - \bar{y})^2$ | $(y_{гр} - \bar{y})^2 n$ |
|---|---|---|---|---|---|---|
| 5,4–8,19 | 5 | 4 170 | 834,0 | –361,4 | 130609,96 | 653049,80 |
| 8,2–0,99 | 8 | 9 100 | 1 137,5 | –57,9 | 3352,41 | 26819,28 |
| 11,0–13,79 | 3 | 3 870 | 1 290,0 | +94,6 | 8949,16 | 26847,48 |
| 13,8–16,59 | 3 | 3 890 | 1 296,7 | +101,3 | 10261,69 | 30785,07 |
| 16,6–19,40 | 5 | 7 660 | 1 532,0 | +336,6 | 113299,56 | 566497,80 |
| Всього | 24 | 28 690 | 1 195,4 | – | – | 1 303 999,43 |

Дисперсія групових рівнів продуктивності праці дорівнює:

$$\delta^2 = \frac{1\,303\,999{,}43}{24} = 54333{,}31.$$

Розрахуємо тепер загальну дисперсію рівнів продуктивності праці за індивідуальними даними 24 заводів. Для цього складемо розрахункову таблицю (табл. 9.13).

Таблиця 9.13

**Розрахунок загальної дисперсії продуктивності праці (*y*)**

| $y$ | $y^2$ | $y$ | $y^2$ | $y$ | $y^2$ |
|---|---|---|---|---|---|
| 700 | 490 000 | 1 150 | 1 322 500 | 1 290 | 1 664 100 |
| 750 | 562 500 | 1 070 | 1 144 900 | 1 380 | 1 904 400 |
| 830 | 688 900 | 1 200 | 1 440 000 | 1 350 | 1 822 500 |
| 910 | 828 100 | 1 150 | 1 322 500 | 1 460 | 2 131 600 |
| 980 | 960 400 | 1 280 | 1 638 400 | 1 440 | 2 073 600 |
| 1 010 | 1 020 100 | 1 300 | 1 690 000 | 1 520 | 2 310 400 |
| 1 160 | 1 345 600 | 1 220 | 1 488 400 | 1 560 | 2 433 600 |
| 1 080 | 1 166 400 | 1 350 | 1 822 500 | 1 680 | 2 822 400 |

$$\sum y = 2\,8820; \quad \sum y^2 = 36\,093\,800; \quad \bar{y} = \frac{28820}{24} = 1\,200{,}8.$$

Розрахуємо загальну дисперсію:

$$\sigma^2 = \frac{\sum y}{n} - (\bar{y})^2;$$

$$\sigma^2 = \frac{36\,093\,800}{24} - (1\,2008)^2 =$$

1 503 908,33 – 1 441 920,64=61 987,69.

Розрахуємо тепер коефіцієнт детермінації ($\eta^2$) та емпіричне кореляційне відношення ($\eta$):

$$\eta^2 = \frac{54\,333{,}31}{61\,987{,}69} = 0{,}877, \text{ або } 87{,}7\ \%; \quad \eta = \sqrt{\frac{54\,333{,}31}{61\,987{,}69}} = 0{,}936.$$

Коефіцієнт детермінації показує, що машиноозброєність праці на 87,7 % визначає варіацію продуктивності праці робітників машинобудівних заводів в нашому прикладі. А кореляційне відношення показує, що зв’язок між машиноозброєністю та продуктивністю праці високий.

## 9.8. Кореляційний та регресійний аналіз взаємозв'язків

Для того, щоб встановити характер зв'язку між явищами, виміряти її тісноту і дати кількісне вираз цього зв'язку, в статистиці використовують кореляційно-регресійний аналіз, заснований на дослідженні супутньої варіації.

*Кореляційний аналіз* або *кореляційний метод* представляє собою метод дослідження взаємозалежності ознак у генеральній сукупності, які є випадковими величинами, мають розподіл нормальне багатовимірне. *Регресійний аналіз* — метод аналізу, в якому досліджується характер і форма залежності результативної ознаки від факторних ознак. Він включає в себе методи обчислення оцінок достовірності отриманих результатів.

Сутність кореляційно-регресійного аналізу полягає в побудові та аналізі статистичної моделі у вигляді рівняння регресії (рівняння кореляційного зв'язку), що наближено виражає залежність результативної ознаки від одного або декількох ознак-факторів. Метод кореляції є ефективним інструментом для досліджень, так як він дає можливість виміряти ступінь впливу кожного окремого фактора при локалізації впливу інших і всієї сукупності факторів за їх одночасної дії.

Кореляційно-регресійний аналіз є продовженням та розвитком факторних угруповань. Як зазначав акад. В. С. Нємчинов, «метод кореляційних рівнянь дозволяє як би довести метод групування до кожного окремо об'єкта спостереження. У порівнянні зі звичайним методом групування метод кореляційних рівнянь дозволяє, образно висловлюючись, визначити групові середні для випадку, коли в групах є лише по одному спостереженню і, навіть тоді, коли в деякі з цих груп не потрапило жодного спостереження ... ».

За допомогою кореляційно-регресійного аналізу вирішуються дві основні задачі:

1) визначається за допомогою рівнянь регресії аналітичної форми ступінь (сила) впливу факторних ознак у загальній зміні результативної ознаки;

2) встановлюється єдина міра тісноти кореляційного зв'язку між ознаками (якою мірою варіація $x$ обумовлює варіацію $y$).

Шляхом побудови та аналізу рівняння регресії можна відповісти на питання, яка роль досліджуваного фактора (факторів) у загальній зміні результативної ознаки. Кореляційний і регресійний аналіз дає можливість виміряти ступінь (силу) впливу факторних ознак на результативні, дозволяє відокремити уявні зв’язки від дійсних.

Основне завдання вивчення кореляційних зв’язків полягає у пошуку причин, які визначають зміну досліджуваного явища, події, факту. Факторна ознака постає як ознака-причина, а результативна — як ознака-слідство.

Вимірювання зв’язку за допомогою кореляційно-регресійного аналізу має свої особливості. В основі вимірювання зв’язку лежать дані статистичного спостереження, де дія випадкових, неврахованих факторів не дозволяє однозначно судити про ступінь впливу залежності, яка цікавить дослідника. Розвиток суспільних явищ у більшості випадків здійснюється під впливом не одного, а багатьох факторів. Якби була можливість закріпити багато другорядних, несуттєвих факторів на постійному рівні, то між основним фактором і результативною ознакою був б функціональний зв’язок. Але практично закріпити дію багатьох ознак-причин на незмінному рівні не представляється можливим. Звідси перше завдання статистичного виміру зв’язку полягає в тому, щоб з’ясувати на основі великої кількості вихідних даних, як змінився б результативний показник при зміні цікавого для нас фактора, якби інші значення ознак-факторів не змінювалися. Друге завдання — визначити ступінь впливу інших аргументів на зміну результативної ознаки.

Якщо суворо виходити із сутності, то кореляційно-регресійний аналіз слід розділити на дві частини: регресійний та кореляційний аналіз. *Регресійний аналіз* вирішує питання побудови та оцінки різних рівнянь регресії, що характеризують залежність результативного показника від інших величин (змінних, факторів, аргументів). За допомогою *кореляційного аналізу* вирішуються ще одне коло питань, що належать до дослідження тісноти зв’язку між ознаками та ролі факторів у загальному результаті. Основними завданнями кореляційного аналізу є оцінка параметрів багатовимірної нормально розподіленої генеральної сукупності (генеральних середніх, дисперсій і парних коефіцієнтів кореляції), множинних та групових коефіцієнтів кореляції; перевірка значимості оцінюваних параметрів взаємозв’язку, отримання інтервальних оцінок для значимих із них, виявлення структури взаємозалежності ознак. У подальшому викладі, якщо не буде обумовлено інше, кореляційно-регресійний аналіз ми називатимемо просто кореляційним аналізом.

Використання кореляційного аналізу передбачає наявність ряду вимог (точніше, емпірично встановлених правил). До цих вимог належать:

ознаки в генеральній сукупності є випадковими величинами та незалежні одна від одної;

дії окремих факторів носить стійкий та незалежний характер;

сталість дисперсії результативної ознаки при зміні факторних ознак;

нормальність розподілу досліджуваних змінних в генеральній сукупності;

однорідність сукупностей (зокрема, відсутність так званих аномальних спостережень), достатній обсяг спостережень.

Насправді щодо кореляційної залежності результативної ознаки від ознак-факторів перелічені умови (обмеження) нерідко порушуються. Зокрема, нерідко виявляються порушеними найважливіші вимоги щодо незалежності окремих одиниць сукупності (спостережень) одна від одної та нормальності розподілу значень ознак. Наприклад, у процесі дослідження вибирають окремі одиниці сукупності, залежні один від одного. Незважаючи на це, результати розрахунків з використанням кореляційного аналізу часто виявляються задовільними.

Основними етапами проведення кореляційного аналізу є такі: 1) попередній теоретичний аналіз; 2) визначення об’єкта дослідження, в тому числі і вибір факторів; 3) збір та обробка відомостей про об’єкт; 4) побудова статистичної моделі у вигляді рівняння регресії та перевірка її адекватності; 5) розрахунок показників тісноти зв’язку.

Розглянемо ці етапи (крім п’ятого, який розглянуто нижче) на прикладі залежності результативної ознаки від одного фактора.

**Роль попереднього теоретичного аналізу**. Раніше ми зазначали, що особливістю кореляційних зв’язків є те, що вони виявляються при великій кількості спостережень, і те, що вони є неповними. У процесі попереднього аналізу необхідно встановити наявність або відсутність взаємозв’язку між досліджуваними явищами. Іноді початкове вивчення вихідних даних може намітити певну взаємозв’язок між явищами, яка у процесі глибокого теоретичного аналізу не знайде свого підтвердження.

Наприклад, є такі дані про тарифний розряд і кількість бракованих деталей у чотирьох токарів механічного цеху:

| | Тарифний розряд | Кількість бракованих виробів за день |
|---|---|---|
| 1 токар | II | 4 |
| 2 » | III | 4 |
| 3 » | IV | 6 |
| 4 » | V | 7 |
| 5 » | IV | 9 |

Зіставлення тарифного розряду та кількості бракованих виробів за день може створити видимість, що між цими ознаками існує тісний прямий зв'язок: зі збільшенням тарифного розряду збільшується і кількість браку. Однак це початковий висновок, отриманий на основі попереднього аналізу, є помилковим і в процесі подальшого теоретичного аналізу може не знайти підтвердження. У разі причиною браку може бути наявність дефектів у вихідній сировині. Крім того, цей висновок не враховує кількість виготовлених виробів. Як правило, у робітників вищого тарифного розряду продуктивність праці вища. Це означає, що за зміну вони можуть зробити більшу кількість виробів. В результаті абсолютна кількість бракованих виробів у них виявиться вищою. Таким чином, наявність зв'язку між тарифним розрядом і кількістю бракованих виробів у процесі глибокого теоретичного аналізу не знайшла підтвердження, виявилася хибною, не обумовленою дією даного фактора.

Помилковий зв'язок може виявитися і в інших випадках. Якщо ми говоримо, що рівень продуктивності праці залежить від ступеня механізації виробничих процесів, оскільки заміна ручних засобів праці машинами і механізмами істотно знижує витрати праці на виробництво продукції, а, отже, підвищує продуктивність праці. Але це значить наявність зворотної залежності, саме, що рівень механізації виробничих процесів залежить від продуктивність праці. Тим часом аналіз варіації зазначених ознак може підтвердити наявність стійкої прямої зв'язку між ними. Таким чином, за узгодженості у варіюванні ознаки причинного зв'язку може і не бути. Отже, не будь-яка узгоджена зміна ознаки може бути предметом кореляційного аналізу. Крім того, в окремих випадках близькість розташування рівень динамічного ряду може визначатися одними і тими ж причинами (явище так званої автокореляції).

Таким чином, лише після того, як з'ясована сутність економічного явища, має сенс дослідити його розвиток і прояв властивих йому закономірностей за допомогою статистичних методів. Це положення повністю відноситься і до застосування кореляційно-регресійного аналізу.

Попередній аналіз передбачає: перевірку однорідності сукупності; виключення аномальних спостережень; визначення необхідного обсягу вибірки; встановлення законів розподілу досліджуваних змінних.

У процесі попереднього теоретичного аналізу необхідно довести, що між досліджуваним ознакою, який ми обираємо як фактор, і ознакою, який є результатом дії цього фактора (факторів), існує певний зв'язок (прямий або зворотний, лінійний або криволінійний).

**Визначення об'єкта дослідження. Відбір факторів**. Керуючись залежностями, виявленими в ході теоретичного аналізу, визначають об'єкт дослідження. Поряд з об'єктом встановлюють одиницю спостереження.

Як і будь-які статистичні показники, показники кореляції можуть бути обчислені лише на основі даних, отриманих з дотриманням принципу однорідності статистичної сукупності. Тому в процесі аналізу з безлічі сукупностей необхідно виділити однорідні сукупності. З зазначеної точки зору поняття однорідності необхідно розглядати завжди тільки у відносному значенні, маючи на увазі питання: щодо яких саме ознак однорідна дана сукупність. Наприклад, у промислових підприємств — однотипність виробничих процесів, однорідність готової продукції. Розглядаючи стаж і ступінь кваліфікації працівників, слід намагатися виключити неоднорідність досліджуваної сукупності за всіма іншими ознаками (сімейний стан, вік та ін.).

Найважливішим питанням кореляційного аналізу є встановлення результативної та факторної (факторних) ознак. Більше це завдання вирішується з допомогою парної кореляції. Скажемо, поставлене завдання — дослідити взаємозв'язок продуктивності праці та енергоозброєності праці. Очевидно, тут результативною ознакою буде продуктивність праці, а факторна — енергоозброєність праці.

**Збір та відомості про об'єкт**. Джерелом даних для проведення кореляційного аналізу служить обліково-статистична, планова і нормативна інформація, а також інформація, одержувана безпосередньо шляхом проведення спеціально організованого спостереження.

Зібрана інформація перевіряється та оцінюється з різних точок зору. Насамперед, перевіряється повнота зібраних даних. У процесі перевірки з'ясовують, чи від усіх звітних (облікових) одиниць надійшли відомості, чи заповнені документи первинного обліку та звітності, а також бланки статистичного спостереження. Далі перевіряється достовірність, правильність відповідей на поставлені питання. У процесі контролю визначається відповідність інформації тим вимогам та обмеженням, які ставляться перед кореляційно-регресійним аналізом (незалежність спостережень та фактів, нормальність розподілу ознак та ін.). І, нарешті, перевіряється порівнянність одержаного статистичного матеріалу. Недостовірна, незрівнянна інформація виключається з подальшої обробки. Отримана інформація має бути не поодинокою, а досить великого обсягу. При цьому достатність обсягу інформації має бути характерною не тільки для загального масиву даних (числа спостережень), але і для кожної з виділених груп. Зазначена умова дозволить забезпечити достатньою мірою погашення випадкових факторів у розвитку. Лише інформація, що задовольняє зазначеним вимогам, може бути використана надалі для проведення кореляційного аналізу.

**Побудова статистичної моделі у вигляді рівняння регресії**. Спостерігаючи статистичний зв'язок між ознаками, математична статистика прагне виразити цей зв'язок за допомогою математичної функції (формули). *Емпірична формула* дозволяє зафіксувати тенденції в динаміці явищ або залежність між результатами спостережень кількох ознак шляхом згладжування або вирівнювання фактичних даних. Розрахунок емпіричної формули, за допомогою якої здійснюється вирівнювання, складається з двох етапів: 1) вибір типу функції; 2) знаходження параметрів обраної функції.

Перш ніж приступити до побудови статистичної моделі, слід встановити форму кореляційного зв'язку, тобто вибрати певний вид математичної функції, що дає найкраще наближення залежності між досліджуваними ознаками. Вибір форми зв'язку має вирішальне значення у кореляційному аналізі. Всі подальші ретельні розрахунки можуть бути знецінені, якщо форма зв'язку обрана неправильно. Пошуку теоретичної форми зв'язку передує якісний аналіз досліджуваного зв'язку. Після теоретичного аналізу взаємозв'язків між ознаками явищ приступають до вибору форми зв'язку.

Вирішення цього питання про тип аналітичної функції спирається на знання природи досліджуваного явища, логічний аналіз, попередні дослідження та на статистичні дані, простоти аналітичної функції, покладеної в основу зв'язку. У визначенні типу функції можуть допомогти кореляційні таблиці, графіки.

При складанні кореляційного рівняння вид функції, тобто характер зв'язку між функцією та аргументами, визначається на основі аналізу сутності досліджуваних явищ. Теоретичний аналіз поряд з використанням загальноекономічних методів включає розгляд досвіду попередніх досліджень, використання експертних оцінок фахівців і т. п. Емпіричний шлях полягає у вивченні вихідних даних за допомогою кореляційних полів, аналізу паралельних рядів. Вивчення фактичного матеріалу дозволяє зробити попередній висновок про наявність або відсутність залежностей між явищами. Так, якщо аналіз показує, що величина явища змінюється приблизно рівномірно, у відповідності зі зміною величини впливу фактора — зв'язок прямолінійний, з однаковими темпами — експоненціальний і т. п.

*Рівняння кореляційного зв'язку* (*рівняння регресії*) — аналітичне рівняння, за допомогою якого виражається зв'язок між ознаками. При цьому потрібно прагнути знайти таку математичну функцію, що характеризує основну залежність між ознаками, яка давала б найменше відхилення від отриманих при спостереженні значень їх ознак. Рівняння цієї функції буде *рівнянням* зв'язку між результативним та факторними ознаками.

Залежно від характеру зміни у зі зміною *x* розмежовують лінійний зв'язок (пряма лінія) і криволінійний (параболічну, гіперболічну та ін). Розглянемо рівняння зв'язку для залежностей від однієї ознаки при різних формах зв'язку (лінійної, криволінійної параболічної, гіперболічної) і для множинного зв'язку:

1) у тих випадках, коли результативна ознака зі збільшенням факторної ознаки рівномірно зростає або зменшується, використовують рівняння прямої лінії (рис. 9.6):

$$\overline{y}_x = a_1 x; \quad \text{и} \quad \overline{y}_x = a_0 + a_1 x; \qquad (9.10)$$

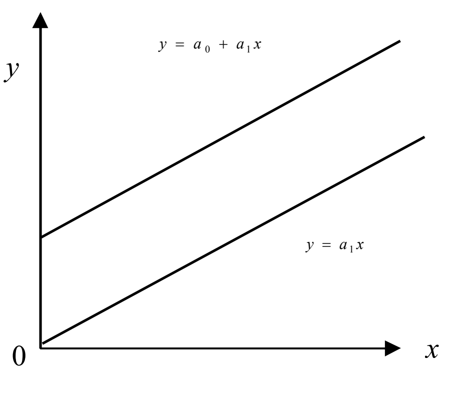



*Рис. 9.6.* Прямі лінії.

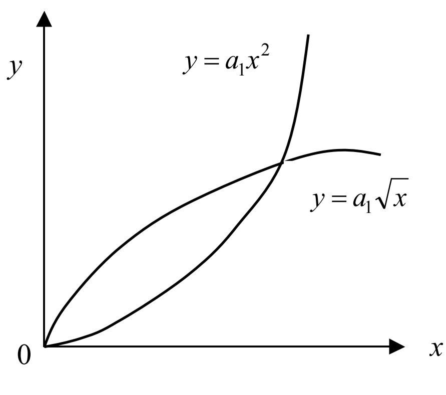



*Рис. 9.7.* Параболічна функція.

2) у тих випадках, коли криві дугоподібні і мають один вигин або коли є рівномірна зміна швидкості, використовують параболи другого

порядку (рис. 9.7–9.8);

$$\overline{y}_x = a_2 x^2; \quad \overline{y}_x = a_2 \sqrt{x}; \quad \overline{y}_x = a_0 + a_1 x + a_2 x^2; \tag{9.11}$$

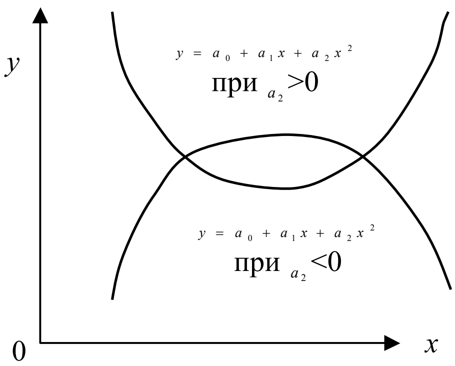



*Рис. 9.8.* Параболічні функції.

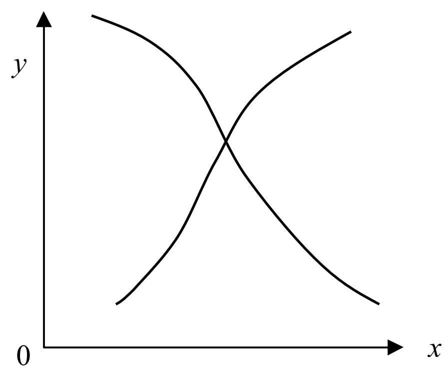



*Рис. 9.9.* Параболи третього порядку.

3) у тих випадках, коли точки ознаки утворюють криву з двома екстремальними значеннями (максимум і мінімум), що має перегин; криві мають *S*-подібну форму (два вигини), використовують параболу третього порядку (рис. 9.9):

$$\overline{y}_x = a_0 + a_1 x + a_2 x^2 + a_3 x^3; \tag{9.12}$$

4) іноді використовують параболу *n*-го порядку:

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n; \tag{9.13}$$

5) у разі, коли величина *y* спочатку зростає прискореними темпами при рівномірній зміні *x*, потім темп зростання уповільнюється і, нарешті, повністю припиняється, застосовується рівняння логістичної функції:

$$\overline{y}_x = \frac{a_0}{1 + a_1 e_{bx}}; \tag{9.14}$$

6) у разі, якщо зі зміною ознаки *x* в арифметичній прогресії відбувається:

а) зростання результативної ознаки *y* геометричній прогресії або рівні змінюються з більш менш відносним приростом (за правилом складних відсотків), то застосовується рівняння показової кривої — експоненти:

$$\overline{y}_x = a_0 \cdot a_1^{\,x}; \tag{9.15}$$

б) уповільнене зростання результативної ознаки *y*, то використовують логарифмічну криву:

$$y = a_0 + a_1 + a_2 \ln x; \tag{9.16}$$

7) іноді використовують рівняння степеневої функції (рис. 9.10):

$$\bar{y}_x = a_0 x^{a_1}, \tag{9.17}$$

або

$$\ln \bar{y}_x = \ln a_0 + a_1 \ln x; \tag{9.18}$$

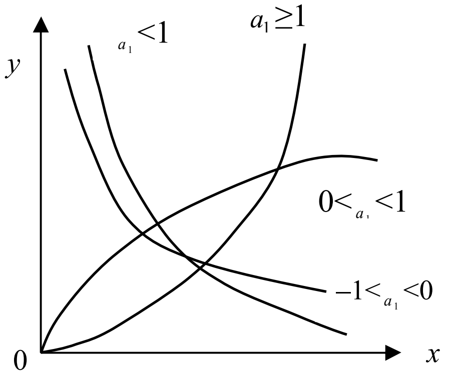



*Рис. 9.10.* Степеневі функції.

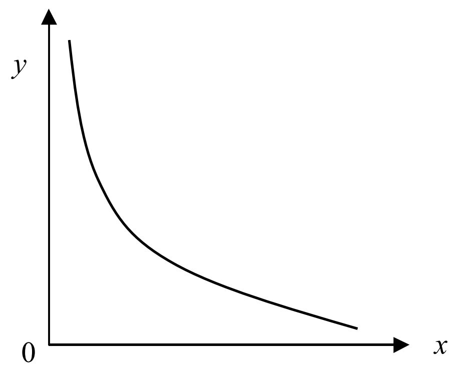



*Рис. 9.11.* Гіпербола другого порядку.

8) в окремих випадках застосовується рівняння гіперболи 2-го порядку (рис. 9.11 та рис. 9.12):

$$\bar{y}_x = a_0 + a_1 \frac{1}{x}; \qquad \bar{y}_x = a_1 \frac{1}{x}; \tag{9.19}$$

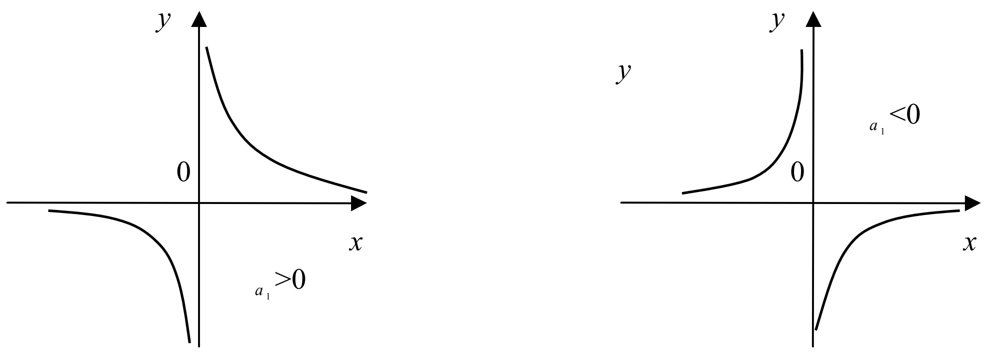



*Рис. 9.12.* Рівностороння гіпербола
(зворотньо-пропорційна функція $\bar{y}_x = a_1 \frac{1}{x}$).

.

9) іноді застосовують рівняння логарифмічного зв'язку:

$$\bar{y}_x = a_0 + a_1 \ln x. \tag{9.20}$$

У цих формулах $a_0$, $a_1$ і $a_2$ — постійні величини, що називаються *параметрами*; $\bar{y}_x$ — значення результативної ознаки, що залежить тільки від факторного ознаки. Параметр $a_0$ показує усреднений вплив на результативну ознаку неврахованих (не виділених для дослiдження) факторівів. Параметр $a_1$ — коефіцієнт регресії показує, на скільки змінюється в середньому значення результативної ознаки при збільшенні факторного на одиницю.

На практиці найбільшого поширення набули лінійні форми зв'язку, оскільки такі моделі простіші, легко інтерпретуються і вимагають меншого обсягу обчислень. Але навіть у тому випадку, коли аналіз взаємозв'язків вимагає застосування криволінійної функції, остання у ряді випадків без великої похибки може бути приблизно виражена прямолінійною. Прямолінійний зв'язок передбачається і в тому випадку, коли форма зв'язку ще не відома.

Однак насправді лінійна форма зв'язку між факторною і результативною ознаками існує відносно рідко. Вибір лінійного зв'язку часто сприймається як спрощений приклад складних взаємозв'язків між явищами і процесами. Лінійна форма не завжди чітко відображає залежність між явищами. З метою точнішого виявлення залежностей між явищами використовуються рівняння криволінійної кореляції.

Вибір теоретичної форми зв'язку завжди пов'язаний з деякою умовністю. Якщо отримане рівняння зв'язку досить точно описує зміна досліджуваного явища, то теоретичне рівняння та його параметри набувають великого практичного значення, перетворюючи теорію кореляції на потужний засіб у планових та економічних розрахунках.

Як найбільш простий метод знаходження параметрів обраної функції використовують метод обраних точок (метод натягнутої лінії), метод середніх значень.

Найбільш загальним методом емпіричної формули є *метод найменших квадратів*[1], за допомогою якого знаходять параметри рівняння математичної функції. Сутність цього способу полягає в тому, щоб сума квадратів відхилень емпіричних значень від значень, отриманих на підставі рівняння зв'язку, була мінімальною. Спосіб найменших квадратів застосовують для знаходження параметрів рівнянь, різних

---

[1] Метод найменших квадратів, який був створений Гаусом і Лапласом, спочатку мав досить вузьку сферу застосування, головним чином при обробці результатів спостережень в астрономічних та геодезичних розрахунках. Тільки після створення теорії кореляції та регресії метод найменших квадратів набув широкого поширення в теорії та практики статистичних досліджень. Зазначимо лише кілька спеціальних робіт: *Барковська Н. В.*, *Барковський В. В.*, *Лопатін О. К.* Теорія ймовірностей та математична статистика. 5-те вид. Київ: Центр учбової літератури, 2010;

для зв'язків кожного виду. Щоб відзначити, що залежність між ознаками виражається в середньому, значення результативної ознаки, знайденої за рівнянням зв'язку, позначають $\overline{y}_x$.

Вирівнювання за способом найменших квадратів ґрунтується на припущенні, що зміни досліджуваного ряду можуть бути приблизно виражені певним математичним співвідношенням (рівнянням плавного рівня), яке знаходиться на основі теоретичного аналізу. При цьому методі необхідно, щоб сума квадратів відхилень фактичних даних від вирівняних (див. рис. 9.13) була найменшою:

$$(y_1-\overline{y}_1)^2+(y_2-\overline{y})^2+(y_3-\overline{y})^2+\ldots(y_n-\overline{y}_n)^2=S=\min. \qquad (9.21)$$

Критерій методу найменших квадратів можна записати у вигляді наступної формули:

$$S=\sum_{i=n}^{n}(y_i-\overline{y}_x)^2=\min. \qquad (9.22)$$

В залежності від вихідних даних в якості плавних рівнів можуть бути обрані різні типи кривих (включаючи пряму як окремий випадок). Покажемо на прикладі вирівнювання по прямій лінії $\overline{y}_x=a+bx$, де $a$ — початковий рівень (за умови, що перше значення $x$ є нуль); $b$ — швидкість ряду.

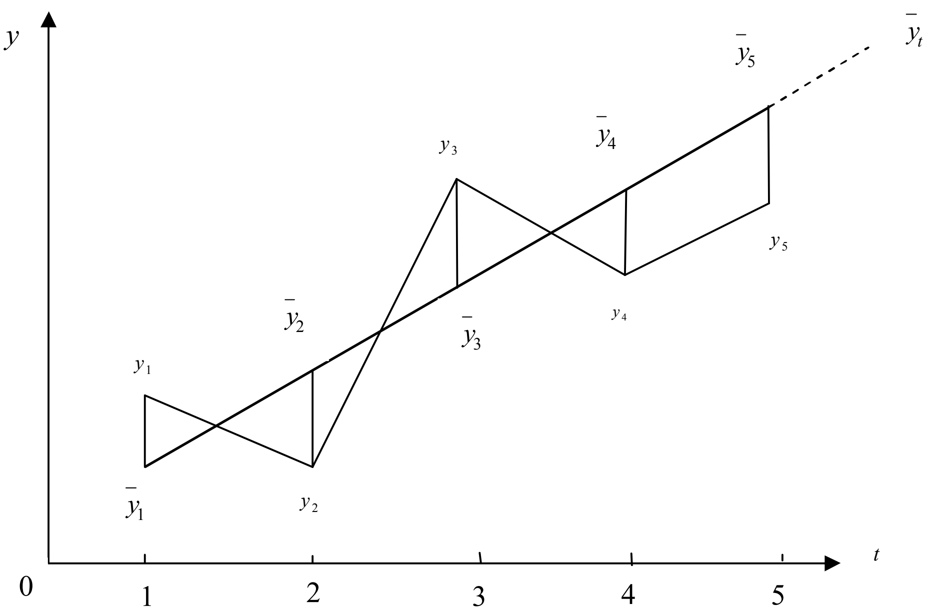



*Рис. 9.13*. Відхилення фактичних рівнів від вирівняних.

Підставивши функцію прямої $\overline{y}_x=a+bx$ в критерій методу найменших квадратів, отримаємо

$$S = \sum_{i=n}^{n} [y_i - (a + bx)]^2 = \min . \qquad (9.23)$$

Отже, для вирішення задачі методом найменших квадратів для визначення $a$ і $b$ необхідно вирішити задачу на екстремум. Функція двох змінних $S(a, b)$ може досягти екстремуму в тому випадку, якщо окремі похідні цієї функції дорівнюють нулю, тобто

$$\frac{ds}{da} = 0; \quad \frac{ds}{db} = 0 . \qquad (9.24)$$

Використовуючи умови екстремуму, обчислимо окремі похідні:

$$\frac{ds}{da} = -2\sum_{i=1}^{n}(y_i - a - bx) = 0; \qquad (9.25)$$

$$\frac{ds}{db} = -2\sum_{i=1}^{n}(y_i - a - bx)x = -2\sum_{i=1}^{n} y_i x + ax - bx^2 = 0. \qquad (9.26)$$

Після перетворень отримаємо систему нормальних рівнянь способу найменших квадратів

Щоб $\sum_{i=n}^{n}[y_i - (a+bx)]^2$ була мінімальною, параметри повинні задовольняти наступній системі нормальних рівнянь:

$$\begin{cases} an + b\sum x = \sum y; \\ a\sum x + b\sum x^2 = \sum yx. \end{cases} \qquad (9.27)$$

де $n$ — число членів ряду.

Якщо дані згруповані (наприклад, представлені у вигляді кореляційної таблиці), то система нормальних рівнянь має вигляд:

$$\begin{cases} a\sum f_{xy} + b\sum xf_x = \sum yf_y; \\ a\sum xf_x + b\sum x^2 f_x = \sum yxf_{xy}. \end{cases} \qquad (9.28)$$

де $f_y$ — частота повторення даного варіанту значень $y$;

$f_x$ — частота повторення даного варіанту $x$;

$f_{xy}$ — частота повторення даного поєднання значень $x$ і $y$.

Якщо сукупність представлена одиницями з безперервною ознакою, то повинні бути використані центральні значення інтервалів, тобто відповідно величини $x'$ і $y'$.

Підставляючи у знайдене рівняння зв’язку значення факторіального ознаки, можна обчислити наперед середнє значення результативної ознаки. Таким чином, рівняння зв'язку є методом узагальнення статистичних зв'язків, що спостерігаються, методом їх вивчення.

Після обчислення параметрів рівняння приступають інтерпретації, яка полягає в тлумаченні отриманих параметрів регресії. На основі отриманих даних регресійної моделі формулюються висновки щодо отриманих рівнянь та коефіцієнтів регресії. Аналіз регресійної моделі дозволяє охарактеризувати зв'язок між змінними, кількісно її оцінити, визначити фактори, що надають вирішальний вплив на зміну результативного показника. Отримані параметри рівняння дозволяють досліднику за заданими значеннями факторів мати теоретичні значення результативної ознаки, підставляючи в нього фактичні значення факторів. Обчислюючи параметри теоретичної лінії зв'язку, ми отримуємо однозначну (за формою) зміну результативного показника зі зміною фактора (факторів). Отримана регресійна модель може бути використана на практиці для впливу на фактори з метою отримання певних результатів. Використання регресійних моделей дає можливість вирішувати різні завдання в галузі прогнозування. Прогнозні значення розраховуються шляхом підстановки в рівняння регресії параметрів значень змінних.

**Перевірка адекватності моделі** включає: перевірку статистичної суттєвості рівняння загалом та її окремих параметрів; перевірку гіпотез, що у основі обраного методу оцінювання; перевірку відповідності формальних властивостей оцінок завданням дослідження.

## 9.9. Однофакторний кореляційно-регресійний аналіз

Спочатку розглянемо однофакторне рівняння регресії. Однією з основних проблем, яку необхідно вирішити перед тим, як обчислювати параметри рівняння регресії, є вибір типу функції.

**Вирівнювання по прямій**. Рівняння зв'язку як рівняння прямої застосовується у разі рівномірного наростання (зменшення) результативної ознаки зі збільшенням (або зменшенням) ознака факторіальної. Така залежність буде залежністю лінійною (прямолінійною).

Розглянемо знаходження теоретичного рівняння зв'язку на прикладі залежності між середньорічною вартістю промислово-виробничих основних засобів і виробленням цукру, що має лінійний характер. При лінійній кореляції формула кореляційної залежності $y$ від $x$ у загальному вигляді має вигляд рівняння прямої лінії, тобто

$$\overline{y}_x = a_0 + a_1 x, \qquad (9.29)$$

де $\overline{y}_x$ — рівень, знайдений за рівнянням;

$a_0$ і $a_1$ — параметри рівняння;

$x$ — час або інший аргумент.

Завдання вирівнювання зводиться до визначення параметрів цього рівняння: $a_0$ — початкового рівня, тобто відрізка на осі ординат від нуля до точки перетину з нею прямою, і $a_1$ — середньої швидкості приросту (зниження) рівня ряду динаміки, тобто тангенсу кута нахилу прямої лінії на осі абсцис.

Для знаходження параметрів $a_0$ і $a_1$ необхідно вирішити систему *нормальних рівнянь*. Нормальні рівняння визначаються застосуванням методу найменших квадратів. Він дає можливість знайти показники рядів динаміки, які мають як прямолінійний, так і криволінійний характер. Метод найменших квадратів базується на вимозі, щоб сума квадратів відхилень даних ($y$) від теоретичних ($\overline{y}_x$) була найменшою.

Для вирівнювання по прямій система нормальних рівнянь має вигляд:

$$\left.\begin{aligned} a_0 n + a_1 \sum x &= \sum y; \\ a_0 \sum x + a_1 \sum x^2 &= \sum xy. \end{aligned}\right\} \qquad (9.30)$$

де $n$ — кількість отриманих при спостереженні пар взаємопов'язаних величин;

$\sum x$ — сума значень факторної ознаки;

$\sum x^2$ — сума квадратів значень факторної ознаки;

$\sum y,$ — сума значень результативної ознаки;

$\sum xy$ — сума добутку значень факторіальної ознаки на значення результативної ознаки.

Знайшовши по рівнянню значення $a_0$ і $a_1$ і підставивши в рівняння зв'язку $\overline{y}_x = a_0 + a_1 x,$, ми зможемо знайти значення $\overline{y}_x$, що залежать від $x$.

У цій системі відомо кількість спостережень $n$, значення $\sum x$, $\sum y$, $\sum x^2$ і $\sum xy$ розраховується на основі фактичних вихідних даних. Невідомо лише два параметри $a_0$ і $a_1$.

Після вирішення рівняння прямолінійної залежності ми можемо перевірити попередній висновок про напрямок залежності між явищами, тобто зв’язок є прямим або зворотнім. Якщо параметр $a_1$ — величина від’ємна (тобто зі знаком мінус), залежність буде від’ємною, бо зі збільшенням значень $x$ значення $\overline{y}_x$ будуть зменшуватися. Якщо $a_1$ буде зі знаком плюс — зв’язок прямий.

Розглянемо знаходження лінійного рівняння, що характеризує зв’язок між випуском цукру і середньорічною вартістю промислово-виробничих засобів на прикладі 10 цукрових заводів і оцінимо значимість параметрів.

Вихідні дані, необхідні визначення параметрів $a_0$ і $a_1$, представлені в табл. 9.14.

Таблиця 9.14

| Номер підприємства | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Середньорічна вартість промислово-виробничих засобів, млн грн | 2,0 | 2,8 | 4,0 | 4,5 | 5,0 | 5,7 | 6,5 | 7,0 | 7,8 | 8,7 |
| Виробництво цукру тис. ц | 210 | 135 | 240 | 264 | 300 | 310 | 365 | 380 | 440 | 500 |

Зробимо аналітичне вирівнювання за прямою на основі даних про середньорічну вартість промислово-виробничих засобів та вироблення цукру 10 підприємств.

Відомо, що зі збільшенням вартості основних засобів до певних меж в середньому буде збільшуватися виробництво цукру. Тому зв’язок між величиною промислово-виробничих засобів та виробленням цукру в певних межах можна вважати прямолінійним.

Позначивши через $x$ середньорічну вартість основних промислово-виробничих засобів, а через $y$ — вироблення цукру в тис. ц, ми можемо визначити зв’язок між цими змінними за допомогою рівняння прямої $\overline{y}_x = a_0 + a_1 x$.

Розрахунки зазначених у системі рівнянь сум зробимо в розрахунковій таблиці (табл. 9.15).

Із таблиці знаходимо:

$$n=10;\ \sum x=54;\ \sum y=3144;\ \sum x^2=333{,}36;$$
$$\sum xy=19027{,}5.$$

Для знаходження параметрів прямий будуємо систему двох рівнянь із двома невідомими:

Таблиця 9.15

**Вирівнювання по прямій**

| Номер підприємства | Середньорічна вартість промислово-виробничих засобів, млн грн $x$ | Виробництво цукру, тис. ц $y$ | $xy$ | $x^2$ | $\bar{y}_x$ |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 2,0 | 210 | 420 | 4,00 | 147,50 |
| 2 | 2,8 | 135 | 378 | 7,84 | 186,77 |
| 3 | 4,0 | 240 | 960 | 16,00 | 245,68 |
| 4 | 4,5 | 264 | 1 188 | 20,25 | 270,22 |
| 5 | 5,0 | 300 | 1 500 | 25,00 | 294,76 |
| 6 | 5,7 | 310 | 1 767 | 32,49 | 329,13 |
| 7 | 6,5 | 365 | 2 372,5 | 42,25 | 368,40 |
| 8 | 7,0 | 380 | 2 660 | 49,00 | 392,94 |
| 9 | 7,8 | 440 | 3 432 | 60,84 | 432,21 |
| 10 | 8,7 | 500 | 4 350 | 75,69 | 476,39 |
| Всього | 54,0 | 3 144 | 19 027,5 | 333,36 | 3 144 |
| В середньому | 5,4 | 314,4 | 1 902,75 | 33,336 | |

$$10a_0 + 54a_1 = 3144;$$
$$54a_0 + 333{,}36a_1 = 19027{,}5.$$

Вирішуємо систему нормальних рівнянь у такій послідовності:

помноживши всі члени першого рівняння на 5,4 (54:10), отримаємо наступну систему рівнянь:

$$54a_0 + 291{,}6a_1 = 16977{,}6;$$
$$54a_0 + 333{,}36a_1 = 19027{,}5.$$

Віднімемо з другого рівняння перше і отримаємо:

$$41{,}76a_1 = 2049{,}9.$$

Звідки $a_1 = \frac{2049{,}9}{41{,}76} = 49{,}088$. Підставляючи значення $a_1$ у перше рівняння, отримаємо:

$$10a_0 + 54 \cdot 49{,}088 = 3144; \qquad 10a_0 + 2650{,}752 = 3144,$$

звідки

$$10a_0 = 493{,}248 \quad \text{і} \quad a_0 = \frac{493{,}248}{10} \approx 49{,}324.$$

Отже, теоретичне рівняння зв’язку між середньою вартістю основних засобів і виробництвом цукру набуває вигляду:

$$\overline{y}_x = 49{,}324 + 49{,}088x.$$

*Логічний та економічний зміст параметрів кореляційного рівняння.* Про що говорять коефіцієнти рівняння зв’язку, визначені методом найменших квадратів?

Перший із показників $a_0$ — ордината при $x$=0. У прикладі $a_0$=49,32 і характеризує вироблення продукції, яка залежить від основних виробничих засобів. Показник $a_1$ говорить про те, як у середньому для всіх спостережень зміниться $y$ при зміні $x$ на одиницю в межах встановлених спостереженням варіації. Згідно з цим рівнянням кожен мільйон вартості основних промислово-виробничих засобів призводить в середньому до збільшення виробітку цукру на 49,088 тис. ц, за умови, що вплив на неї всіх інших факторів еліміновано.

Підставляючи це рівняння конкретні значення $x$ з табл. 9.15, знаходимо для всіх десяти підприємств $\overline{y}_x$, тобто теоретичні значення виробництва цукру. Показники прямої і будуть виражати загальну тенденцію розвитку, отриману шляхом послідовного додавання до кожного члена ряду, починаючи з першого, середнього збільшення обсягу виробництва цукру протягом періоду, що вивчається ($a_1$=49,088). Так наприклад,

при $x$=2,0:

$$\overline{y}_x = 49{,}324 + 49{,}088 \cdot 2{,}0 = 147{,}50;$$

при $x$=2,8:

$$\overline{y}_x = 49{,}324 + 49{,}088 \cdot 2{,}8 = 186{,}77 \text{ і т. п.}$$

Фактично вироблення цукру на підприємстві № 1 (при $x$=2,0 млн грн) було вищим за теоретичне значення на 62,5 тис. ц. Це говорить про те, що в даному випадку інші, невраховані тут фактори, здійснювали сприятливий вплив на виробництво цукру. А ось на підприємстві № 2 співвідношення вийшло інше. Тут знайдене за рівнянням розрахункове виробництво склало 186,77 тис. ц, а фактично вона дорівнює 135 тис. ц. У цьому випадку інші фактори в цілому впливали на вироблення у бік його зниження.

У графі 6 табл. 9.15 дано вирівняні значення вироблення цукру $\overline{y}_x$ для всіх значень середньорічної вартості основних промислово-виробничих засобів. Знайдене вище теоретичне рівняння зв’язку (рівняння регресії) виражає залежність $y$ від $x$.

Графічне зображення фактичних і вирівняних даних наведено на рис. 9.14. На графіці значення $\bar{y}_x$ утворюють пряму лінію, а фактичні значення зображені ламаною лінією, яка дає графічне зображення невирівняних значень $y$. Такий графік сам по собі дає деяке уявлення про тісноту зв'язку між $x$ і $y$. Чим ближче примикає ламана лінія до прямої $\bar{y}_x$, тим тісніше зв'язок.

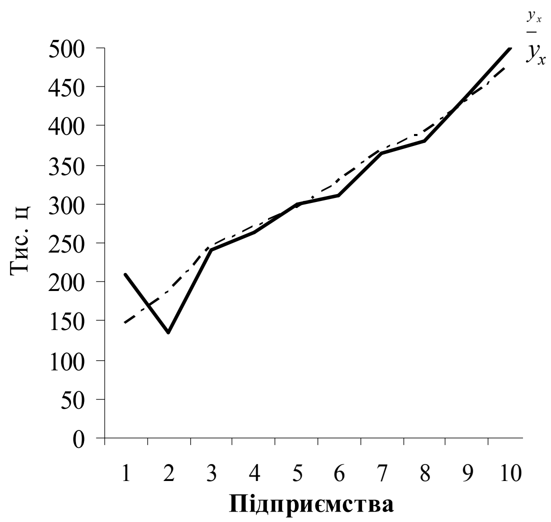


*Рис. 9.14.* Залежність вироблення цукру від вартості виробничих засобів.

Зображені фактичні та обчислені значення на рис. 9.14 показує, що рівняння зв'язку відображає залежність, що спостерігається в середньому.

Проведемо перевірку знайдених параметрів:

$$\text{I.}\quad 10\cdot 49{,}32+54\cdot 49{,}088=493{,}2+2650{,}752=$$
$$=3143{,}952\approx 3144;$$
$$\text{II.}\quad 54\cdot 49{,}32+333{,}36\cdot 49{,}088=2663{,}28+16363{,}97568=$$
$$=19507{,}92768\approx 19508.$$

Правильність розрахунку параметрів рівняння кореляційної залежності може бути перевірено порівнянням двох сум: $\sum y$ і $\sum \bar{y}_x$. У прикладі вони рівні ($\sum y$=3144=$\sum \bar{y}_x$), що підтверджує правильність розрахунків параметрів рівняння.

Параметри рівняння прямої можна визначати не тільки шляхом підстановки, а безпосередньо за формулами, що дещо спрощує їх обчислення:

$$a_0=\frac{\sum x^2\cdot\sum y-\sum xy\cdot\sum x}{n\sum x^2-(\sum x)^2},\qquad a_1=\frac{n\sum xy-\sum x\cdot\sum y}{n\sum x^2-(\sum x)^2}. \tag{9.31}$$

Звідки

$$a_0=\frac{\sum x^2\cdot\sum y-\sum xy\cdot\sum x}{n\sum x^2-(\sum x)^2}=\frac{333{,}36\cdot 3144-19027{,}5\cdot 54}{10\cdot 333{,}36-54^2}\approx 49{,}32;$$

$$a_1=\frac{n\sum xy-\sum x\cdot\sum y}{n\sum x^2-(\sum x)^2}=\frac{10\cdot 19027{,}5-54\cdot 3144}{10\cdot 333{,}36-54^2}=\frac{20499}{417{,}6}=49{,}088.$$

Як видно, параметри рівняння прямої, обчислені за формулами, збігаються з параметрами, обчисленими методом підстановок.

Параметри рівняння можна визначити також за такими формулами:

$$a_1 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - (\overline{x})^2}; \qquad a_0 = \overline{y} - a_1 \overline{x}. \tag{9.32}$$

$$a_1 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - (\overline{x})^2} = \frac{1902{,}75 - 5{,}4 \cdot 314{,}4}{33{,}336 - 5{,}4^2} \approx 49{,}088;$$

$$a_0 = \overline{y} - a_1 \overline{x} = 314{,}4 - 49{,}088 \cdot 5{,}4 \approx 49{,}32.$$

Отже, зі збільшенням вартості основних виробничо-промислових засобів на 1 млн грн виробництво цукру збільшується в середньому на 49,088 тис. ц.

Іноді для знаходження параметрів прямий для спрощення розрахунків застосовується і інший спосіб, заснований на заміні в рівняннях значень часу та відповідних їм даних рівнів ряду на відхилення від їх середніх величин. Це геометричному сенсі відповідає перенесення початку координат в точку $(\overline{x}, \overline{y})$.

Саме рівняння прямої матиме такий вигляд:

$$(y - \overline{y}) = a_0 + a_1(x - \overline{x}). \tag{9.33}$$

Система нормальних рівнянь визначення параметрів прямий має такий вид:

$$\left.\begin{aligned} a_0 n + a_1 \sum x &= \sum y; \\ a_0 \sum x + a_1 \sum x^2 &= \sum xy; \end{aligned}\right\} \tag{9.34}$$

Замінимо в цій системі рівнянь значення аргументу на відхилення від їхнього середнього значення та значень рівня ряду динаміки на відхилення від їхнього середнього значення та отримаємо наступну систему рівнянь:

$$\left.\begin{aligned} a_0 n + a_1 \sum (x - \overline{x}) &= \sum (y - \overline{y}); \\ a_0 \sum (x - \overline{x}) + a_1 \sum (x - \overline{x})^2 &= \sum (x - \overline{x})(y - \overline{y}). \end{aligned}\right\} \tag{9.35}$$

Параметри $a_0$ і $a_1$ (візьмемо їх для простоти без частот) виглядатимуть так:

$$a_0 = \frac{\sum (y - \overline{y}) \sum (x - \overline{x})^2 - \sum (x - \overline{x})(y - \overline{y}) \sum (x - \overline{x})}{n \sum (x - \overline{x})^2 - \sum (x - \overline{x}) \sum (x - \overline{x})}; \tag{9.36}$$

$$a_1 = \frac{n\sum(x-\bar{x})\sum(y-\bar{y}) - \sum(y-\bar{y})\sum(x-\bar{x})}{n\sum(x-\bar{x})^2 - \sum(x-\bar{x})\sum(x-\bar{x})}. \tag{9.37}$$

Відповідно до 7 властивості середньої арифметичної (сума відхилень окремих значень ознаки (варіантів) від середньої арифметичної, помноженої на ваги (частоти), дорівнює нулю):

$$\sum(y-\bar{y}) = 0, \tag{9.38}$$

$$\sum(x-\bar{x}) = 0, \tag{9.39}$$

звідси підставивши їх значення (0) у формули $a_0$ і $a_1$ отримаємо:

$$a_0 = 0;$$

$$a_1 = \frac{\sum(y-\bar{y})(x-\bar{x})}{\sum(x-\bar{x})^2}. \tag{9.40}$$

Параметр $a_0$ можна обчислити із системи нормальних рівнянь, отриманих способом найменших квадратів. Якщо перше рівняння із системи нормальних рівнянь, за якою перебували параметри прямої ($a_0 n + a_1\sum x = \sum y$), розділити на *n*, то формула для розрахунку $a_0$ набуде наступного вигляду:

$$a_0 = \bar{y} - a_0\bar{x}. \tag{9.41}$$

Розрахунки параметрів рівняння прямий цим спрощеним способом зробимо за тими ж показниками, що і в табл. 9.15.

Використовуючи дані таблиці, визначаємо параметри рівняння прямої:

$$a_1 = \frac{\sum(y-\bar{y})(x-\bar{x})}{\sum(x-\bar{x})^2} = \frac{2049{,}9}{41{,}76} = 49{,}088,$$

$$a_0 = \bar{y} - a_0\bar{x} = 314{,}4 - 49{,}088 \cdot 5{,}4 = 49{,}32.$$

Таким чином, параметри $a_0$ і $a_1$, розраховані спрощеним способом, рівні параметрам, обчисленим звичайним способом. Для перевірки правильності визначення загальної тенденції розвитку методом аналітичного вирівнювання необхідно дотриманням деяких умов. Перша — сума відхилень емпіричного ряду ($\sum y$) повинна дорівнювати сумі рівнів вирівняного ряду динаміки ($\sum y_x$) і друга — сума відхилень рівнів емпіричного ряду від рівнів теоретичного (вирівняного) ряду динаміки повинна дорівнювати нулю ($y - y_x$). Як очевидно з

Таблиця 9.16

**Спрощене обчислення параметрів прямої**

| Номер підприємства | Середньорічна вартість промислово-виробничих засобів, млн грн ($x$) | Виробництво цукру, тис. ц ($y$) | $(x-\bar{x})$ | $(x-\bar{x})^2$ | $(y-\bar{y})$ | $(x-\bar{x})(y-\bar{y})$ |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 2,0 | 210 | –3,4 | 11,56 | –104,4 | +354,96 |
| 2 | 2,8 | 135 | –2,6 | 6,76 | –179,4 | +466,44 |
| 3 | 4,0 | 240 | –1,4 | 1,96 | –74,4 | +104,16 |
| 4 | 4,5 | 264 | –0,9 | 0,81 | –50,4 | +45,36 |
| 5 | 5,0 | 300 | –0,4 | 0,16 | –14,4 | +5,76 |
| 6 | 5,7 | 310 | +0,3 | 0,09 | –4,4 | –1,32 |
| 7 | 6,5 | 365 | +1,1 | 1,21 | +50,6 | +55,66 |
| 8 | 7,0 | 380 | +1,6 | 2,56 | +65,6 | +104,96 |
| 9 | 7,8 | 440 | +2,4 | 5,76 | +125,6 | +301,44 |
| 10 | 8,7 | 500 | +3,3 | 10,89 | +185,6 | +612,48 |
| Всього | 54,0 | 3144 | 0,0 | 41,76 | 0,0 | 2 049,90 |
| В середньому | $\bar{x}=5,4$ | $\bar{y}=314,4$ | | | | |

табл. 9.15 (графи 3, 6) та табл. 9.16 (графа 6) у нашому прикладі ці два правила дотримані та підтвердили правильність визначення загальної тенденції.

Ми розглянули розрахунок параметрів рівняння регресії безпосередньо за індивідуальними значеннями, що характеризують вартість основних виробничих засобів і обсяг виробництва цукру. Якщо обсяг сукупності великий, то розрахунок параметрів рівняння вручну є дуже трудомістким і проблематичним. При великих обсягах сукупності обчислення ведуться за згрупованими даними. Наприклад, велику сукупність підприємств за вартістю основних виробничих засобів можна розділити на групи підприємств, що мають обсяг виробничих фондів від 2 до 4 млн грн, від 4 до 6 млн грн і т. п. Розрахунок параметрів рівняння зв’язку в даному випадку буде дещо відрізнятися від викладеного вище тільки тим, що у формули $a_0$ і $a_1$ будуть введені частоти $f$ (за аналогією з формулами середньої арифметичної без частот і з частотами — $\frac{\sum x}{n}$ і $\frac{\sum xf}{n}$).

У такому разі формули розрахунку рівняння регресії мають такий вигляд:

$$a_0 = \frac{\sum yf \sum x^2 f - \sum xyf \sum xf}{\sum f \sum x^2 f - \sum xf \sum xf}; \qquad (9.42)$$

$$a_1 = \frac{\sum f \sum xyf - \sum yf \sum xf}{\sum f \sum x^2 f - \sum xf \sum xf}. \qquad (9.43)$$

Як видно з наведених формул, для обчислення параметрів рівняння необхідно визначити $\sum f$, $\sum xf$, $\sum yf$, $\sum x^2 f$, $\sum xyf$.

Рівняння прямої для визначення загальної тенденції динамічного ряду застосовується у разі, коли збільшення (зменшення) рівня відбувається щодо рівномірно і прямолінійно, що найчастіше буває при внутрішньорічних коливаннях чи коротких рядах. У соціально-економічних явищах, що охоплюють тривалий період, загальну тенденцію досить рідко можна виразити рівнянням прямої, оскільки в рядах динаміки розвиток явища відбувається не рівномірно: після підйому в розвитку починається спад, або навпаки, після спаду починається зростання. У таких випадках статистична модель у вигляді рівняння прямої не може вловити ці зміни. Для точного відображення існуючих зв'язків між ознаками використовують різні математичні функції нелінійного характеру.

Так вивчаються прямолінійні зв'язки.

**Коефіцієнт еластичності**. Для економічної інтерпретації лінійних та нелінійних зв'язків між двома явищами застосовують коефіцієнт еластичності.

*Коефіцієнт еластичності* (*E*) показує зміну результативного ознаки $y$ ($\bar{y}_x$) у процентах в залежності від зміни факторної ознаки $x$ на 1 %. Для інтерпретації коефіцієнта $a_1$ у разі лінійної залежності коефіцієнт еластичності розраховується за формулою:

$$E = a_1 \frac{\bar{x}}{\bar{y}}. \qquad (9.44)$$

де $\bar{x}$ і $\bar{y}$ — середні арифметичні, отримані за вибіркою.

Коефіцієнти еластичності відображають відносний вплив окремих факторів на досліджуваний показник. Щоб показати, якою мірою кожен фактор впливає на функцію в абсолютному значенні, розраховується так званий «додатковий продукт»:

Використовуючи дані табл. 9.16, розрахуємо середні значення факторної та результативної ознак:

$$\bar{x} = \frac{54}{10} = 5{,}4 \text{ млн грн}; \quad \bar{y} = \frac{3144}{10} = 314{,}4 \text{ тыс. ц.}$$

Тоді для нашого прикладу коефіцієнт еластичності складе:

$$E = 49{,}088 \frac{5{,}4}{314{,}4} = 0{,}84.$$

Це означає, що при зростанні вартості основних виробничих фондів на 1 % в порівнянні з $\bar{x}$ виробництво цукру зростає в середньому на 0,84 %.

Коефіцієнти еластичності можна обчислити за кожним підприємством. У нашому прикладі коефіцієнт еластичності на першому підприємстві при $x$=2,0 дорівнюватиме: $E = 49{,}088 \frac{2{,}0}{210} = 0{,}47$. Отже, при зростанні на 1 % приросту середньорічної вартості основних виробничих засобів виробництво цукру зростає на 0,47 %. На четвертому підприємстві при $x$=4,5 $E$=0,84$\left(49{,}088 \cdot \frac{4{,}5}{264}\right)$, на восьмому $x$=7,0 $E$=0,90, на десятому при $x$=8,7 $E$=0,854. Таким чином, коефіцієнти еластичності змінюють своє значення зі зміною величини $x$.

Якщо залежність представлена у вигляді параболи, то коефіцієнт еластичності має вигляд:

$$E = (a_1 + a_2 x)\frac{\bar{x}}{\bar{y}}. \tag{9.45}$$

## 9.10. Нелінійні залежності

В попередньому розділі ми розглянули приклад вирівнювання за прямою. Лінійна залежність — це така, де $y_x$ змінюється пропорційно зміні $x$, тобто тангенс кута нахилу теоретичної лінії зв’язку залишається незмінним. Це означає, що при збільшенні вартості основних виробничих фондів на 1 млн грн випуск цукру зростає на 49088 т. Отже, зберігається однаковий приріст результативного показника в залежності від зміни на одиницю ознаки-фактора. В аналізі масових суспільних явищ найбільш часто застосовуються наступні нелінійні залежності: гіперболічна, напівлогарифмічна, парабола другого порядку. У практиці кореляційно-регресійного аналізу застосовуються й інші функції, але рідше, і ми їх у даному курсі не розглядатимемо.

**Вирівнювання по гіперболі**. З нелінійних залежностей, що зустрічаються в суспільних явищах, найбільш поширені такі, які описуються рівнянням гіперболи:

$$\bar{y}_x = a_0 + a_1 \frac{1}{x}. \qquad (9.46)$$

Для визначення параметрів цього рівняння спосіб найменших квадратів дає таку систему нормальних рівнянь:

$$\begin{cases} na_0 + a_1 \sum \frac{1}{x} = \sum y; \\ a_0 \sum \frac{1}{x} + a_1 \sum \left(\frac{1}{x}\right)^2 = \sum y \frac{1}{x}. \end{cases} \qquad (9.47)$$

де $\sum \frac{1}{x}$ — сума величин, обернених значенням факторіальної ознаки;

$\sum \frac{1}{x^2}$ — сума їх квадратів.

Щоб визначити параметри рівняння гіперболи методом найменших квадратів, необхідно привести його до лінійного вигляду. Для цього зробимо заміну змінних $\frac{1}{x} = x_1$, отримаємо наступну систему нормальних рівнянь:

$$\begin{cases} na_0 + a_1 \sum x = \sum y; \\ a_0 \sum x + a_1 \sum x^2 = \sum xy. \end{cases} \qquad (9.48)$$

Вирішуючи систему нормальних рівнянь, визначаємо параметри рівняння гіперболи.

Покажемо специфіку розрахунків за даними про сполучені значення факторної і результативної ознак: знайдемо рівняння регресії. Дослідимо, наприклад, зв’язок між вартістю основних засобів та собівартістю одиниці виробу по 75 однорідним підприємствам. Теоретичний аналіз приводи до висновку, що зв’язок між ними має гіперболічний вигляд.

Для визначення параметрів рівняння гіперболи складемо розрахункову таблицю (табл. 9.17).

Підставивши до наведеної вище системи нормальних рівнянь суми з табл. 9.17, отримаємо наступні два рівняння з двома невідомим:

$$\begin{cases} 6a_0 + 3{,}75a_1 = 75 \\ 3{,}75a_0 + 4{,}7681a_1 = 52{,}86 \end{cases} \qquad (9.49)$$

Розділимо кожен член обох рівнянь на коефіцієнти при $a_0$, отри-

Таблиця 9.17

**Залежність між розміром підприємств за вартістю основних засобів та собівартістю одиниці виробу**

| Групи підприємств за вартістю основних засобів, млн грн ($x$) | Собівартість одиниці виробу, грн ($y$) | Середина інтервалу $x$ | $\frac{1}{x} = x_1$ | $x_1^2$ | $yx_1$ | $\bar{y}_x$ |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| До 1 млн грн | 16 | 0,5 | 2,00 | 4,0000 | 32,00 | 15,9 |
| 1–2 | 12 | 1,5 | 0,67 | 0,4489 | 8,04 | 12,6 |
| 2–3 | 13 | 2,5 | 0,40 | 0,1600 | 5,20 | 11,9 |
| 3–4 | 13 | 3,5 | 0,28 | 0,0784 | 3,64 | 11,7 |
| 4–5 | 10 | 4,5 | 0,22 | 0,0484 | 2,00 | 11,5 |
| 5 і більше | 11 | 5,5 | 0,18 | 0,0324 | 1,98 | 11,4 |
| Всього | 75 | 18,0 | 3,75 | 4,7681 | 52,86 | 75,0 |

маємо:

$$a_0 + 0{,}625\,a_1 = 12{,}5$$
$$a_0 + 1{,}2715\,a_1 = 14{,}096.$$

Знаходимо відніманням і другого рівняння першу величину $a_1$:

$$0{,}6465\,a_1 = 1{,}596,$$

звідки

$$a_1 = \frac{1{,}596}{0{,}6465} \approx 2{,}47.$$

Підставивши замість $a_1$ його значення, отримаємо $a_0$:

$$a_0 + 0{,}625 \cdot 2{,}47 = 12{,}5;$$
$$a_0 = 12{,}5 - 0{,}625 \cdot 2{,}47 = 10{,}96.$$

Рівняння зв'язку між вартістю основних засобів і собівартістю одиниці виробу матиме такий вигляд:

$$\bar{y}_x = 10{,}96 + \frac{2{,}47}{x}.$$

Підставимо кожне значення $x$ до рівняння, знаходимо $\bar{y}_x$ за будь-яким рядком таблиці:

$$\bar{y}_1 = 10{,}96 + \frac{2{,}47}{0{,}5} \approx 15{,}9;$$

$$\bar{y}_2 = 10{,}96 + \frac{2{,}47}{1{,}5} \approx 12{,}6,$$

і т. д. Результати розрахунків наведено у стовпці 7 табл. 9.17. Видно, що показники собівартості одиниці виробу, обчислені за допомогою функції, досить близькі до фактичних значень.

Зауважимо, що суми, обчислені за рівнянням $\sum \bar{y}_x$, повинні збігатися з фактичними значеннями результативної ознаки $\sum y_i$. Це дозволяє контролювати результати обчислень. У нашому прикладі $\sum \bar{y}_x$ =75=$\sum y_i$, що підтверджує правильність отриманих параметрів рівняння регресії. Будуємо ламану по парам $x$ і $y$ і криву по $x$ і $\bar{y}_x$. Лінії фактичної та розрахункової собівартості, обчислені за знайденим рівнянням гіперболи, наводяться на рис. 9.15. Ломана і крива дуже близько знаходяться одна до одної.

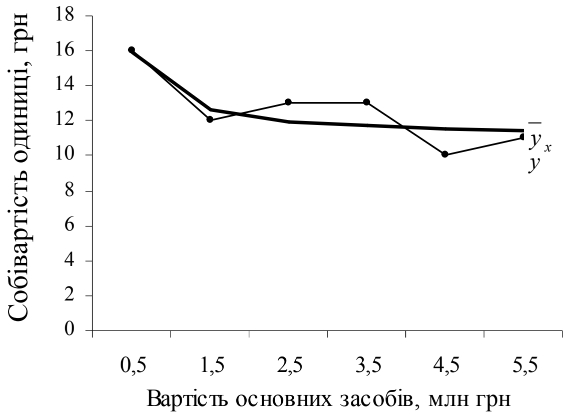



*Рис. 9.15.* Залежність собівартості одиниці виробу від вартості основних засобів.

Дещо змінюється порядок розрахунків, коли вихідні дані зважені за частотами. Вирішуючи способом визначників наведену вище систему нормальних рівнянь для гіперболи, знаходимо (записуємо з частотами):

$$a_0 = \frac{\sum yx \sum \frac{1}{x^2} f - \sum \frac{y}{x} f \sum \frac{1}{x} f}{\sum f \sum \frac{1}{x^2} f - \sum \frac{1}{x} f \sum \frac{1}{x} f}; \tag{9.50}$$

$$a_1 = \frac{\sum f \sum \frac{y}{x} f - \sum \frac{1}{x} f \sum yf}{\sum f \sum \frac{1}{x^2} f - \sum \frac{1}{x} f \sum \frac{1}{x} f}. \tag{9.51}$$

Розрахунок необхідних даних для розрахунку наводимо в табл. 9.18.

Звідки

$$a_0 = \frac{19565{,}9 \cdot 0{,}178 - 826{,}15 \cdot 4{,}192}{100 \cdot 0{,}178 - (4{,}192)^2} \approx 85{,}89;$$

$$a_1 = \frac{100 \cdot 826{,}15 - 4{,}192 \cdot 19565{,}9}{100 \cdot 0{,}178 - (4{,}192)^2} \approx 2618{,}46.$$

Отже, рівняння зв’язку між собівартістю та урожайністю зернових набуває вигляду:

Таблиця 9.18

| Урожайність зернових, ц/га $x$ | Посівна площа, % $f$ | Собівартість 1 ц, грн $y$ | $yf$ | $\frac{y}{x}f$ | $\frac{1}{x}f$ | $\frac{1}{x^2}f$ | $\bar{y}_x$ | $\bar{y}_x f$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 20 | 13 | 221,6 | 2 880,8 | 144,04 | 0,650 | 0,033 | 216,81 | 2 818,57 |
| 22 | 17 | 210,4 | 3 576,8 | 162,58 | 0,773 | 0,035 | 204,91 | 3 483,49 |
| 24 | 35 | 192,3 | 6 730,5 | 280,44 | 1,458 | 0,061 | 194,99 | 6 824,74 |
| 26 | 22 | 187,2 | 4 118,4 | 158,40 | 0,846 | 0,033 | 186,60 | 4 105,20 |
| 28 | 13 | 173,8 | 2 259,4 | 80,69 | 0,464 | 0,017 | 179,41 | 2 332,28 |
| | 100 | × | 19 565,9 | 826,15 | 4,192 | 0,178 | 982,72 | 19 564,28 |

$$y_x = 85{,}89 + \frac{2618{,}46}{x}.$$

Підставивши в рівняння гіперболи конкретні значення $x$ по кожній із п'яти груп, знаходимо теоретичні $\bar{y}_x$ (див. гр. 8 та 9 табл. 9.18). Відхилення $\bar{y}_x$ від $y$ незначні, що підтверджує гіпотезу про наявність тут зв'язку, що описується гіперболою. Теоретичні та емпіричні значення розрізняються незначно; майже збігаються і суми $yf$ і $\bar{y}_x f$ цих рівнів (19565,9 і 19564,28); розбіжність сталася за рахунок округлень.

**Параболічна залежність. Вирівнювання за степеневою функцією**. Степенева функція виду $\bar{y}_t = a_0 x^{a_1}$ застосовується в економічних дослідженнях для характеристики слабо лінійного зв'язку між результативними та факторними ознаками. Параметр $a_1$ має економічний сенс. Він показує, що зі збільшенням ознаки $a_1$ на 1 % результативна ознака збільшується на проценти. Параметр $a_1$ є коефіцієнтом еластичності.

Для визначення параметрів рівняння регресії, вираженої статечною функцією $\bar{y}_x = a_0 x^{a_1}$, приводять функцію до лінійного вигляду шляхом логарифмування. Прологарифмувавши степеневу криву, отримаємо рівняння прямої, в якому рівні та параметри замінені їх логарифмами:

$$\lg y = \lg a_0 + a_1 \lg x. \tag{9.52}$$

Звідси система рівнянь визначення параметрів запишеться:

$$\begin{aligned} n \lg a_0 + a_1 \sum \lg x &= \sum \lg y; \\ \lg a_0 \sum \lg x + a_1 \sum (\lg x)^2 &= \sum \lg y \lg x. \end{aligned} \tag{9.53}$$

Замінемо

$$\lg y = y_1; \quad \lg a_0 = b; \quad \lg x = x_1. \qquad (9.54)$$

Запишемо рівняння (9.52) в нових позначеннях:

$$y = b + a_1 x. \qquad (9.55)$$

Будуємо систему нормальних рівнянь для рівняння (9.55):

$$\sum y = nb + a_1 \sum x;$$
$$\sum xy = b \sum x + a_1 \sum x^2. \qquad (9.56)$$

Вирішуючи систему нормальних рівнянь, визначаємо параметри $a_1$ і $b$ та рівняння (9.55).

Переходячи до початкових позначень $\lg a_0 = b,$ визначаємо параметри $a_0$.

Маються такі дані щодо 13 підприємств:

| Номер підприємства | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Випуск продукції ($x$), млн грн | 152 | 116 | 100 | 108 | 129 | 141 | 147 | 156 | 156 | 163 | 170 | 178 | 187 |
| Собівартість одиниці виробу ($y$), грн | 47,6 | 34,8 | 31,6 | 32,6 | 38,2 | 42,1 | 45,0 | 47,3 | 47,4 | 49,0 | 51,5 | 53,2 | 55,6 |

Залежність між випуском продукції та собівартістю одиниці виробу виражається ступеневою функцією виду $y = a_0 x^{a_1}$.

Для визначення параметрів степеневої функції методом найменших квадратів шляхом логарифмування наведемо її до лінійного вигляду:

$$\lg y = \lg a_0 + a_1 \lg x. \qquad (9.57)$$

Звідси система рівнянь для визначення параметрів запишеться:

$$\sum \lg y = n \lg a_0 + a_1 \sum \lg x;$$
$$\sum \lg y \lg x = \lg a_0 \sum \lg x + a_1 \sum (\lg x)^2. \qquad (9.58)$$

Замінимо $\lg y = y_1; \quad \lg a_0 = b; \quad \lg x = x_1.$

Запишемо систему нормальних рівнянь у нових позначеннях:

$$\sum y_1 = nb + a_1 \sum x_1;$$
$$\sum x_1 y_1 = b \sum x_1 + a_1 \sum x_1^2. \qquad (9.59)$$

Для визначення параметрів рівняння (9.55) побудуємо розрахункову таблицю.

Таблиця 9.19

| Номер п/п | Випуск продукції, тис. шт. ($x$) | Собівартість одиниці виробу ($y$), грн | $\lg x = x_1$ | $\lg y = y_1$ | $x_1 y_1 = \lg y \lg x$ | $(\lg x)^2 = x_1^2$ |
|---|---|---|---|---|---|---|
| 1 | 152 | 47,6 | 2,18184 | 1,67761 | 3,66028 | 4,76044 |
| 2 | 116 | 34,8 | 2,06446 | 1,54158 | 3,18253 | 4,26199 |
| 3 | 100 | 31,6 | 2,00000 | 1,49969 | 2,99937 | 4,00000 |
| 4 | 108 | 32,6 | 2,03342 | 1,51322 | 3,07701 | 4,13481 |
| 5 | 129 | 38,2 | 2,11059 | 1,58206 | 3,33909 | 4,45459 |
| 6 | 141 | 42,1 | 2,14922 | 1,62428 | 3,49094 | 4,61914 |
| 7 | 147 | 45,0 | 2,16732 | 1,65321 | 3,58304 | 4,69726 |
| 8 | 156 | 47,3 | 2,19312 | 1,67486 | 3,67318 | 4,80980 |
| 9 | 156 | 47,4 | 2,19312 | 1,67578 | 3,67519 | 4,80980 |
| 10 | 163 | 49,0 | 2,21219 | 1,69020 | 3,73903 | 4,89377 |
| 11 | 170 | 51,5 | 2,23045 | 1,71181 | 3,81810 | 4,97490 |
| 12 | 178 | 53,2 | 2,25042 | 1,72591 | 3,88403 | 5,06439 |
| 13 | 187 | 55,6 | 2,27184 | 1,74507 | 3,96453 | 5,16126 |
| Всього | | | 28,05799 | 21,31528 | 46,08632 | 60,64215 |
| В середньому | | | 2,15831 | 1,63964 | 3,54510 | 4,66478 |

Підставимо значення фактичних даних у систему нормальних рівнянь:

$$21{,}31528 = 13b + 28{,}05799a_1;$$
$$46{,}08632 = 28{,}05799b + 60{,}64215a_1.$$

Параметри рівняння (9.56) визначимо за формулами:

$$a_1 = \frac{\overline{x_1 y_1} - \overline{x_1}\,\overline{y_1}}{\overline{x^2} - (\overline{x_1})^2} = \frac{3{,}54510 - 1{,}63964 \cdot 2{,}15831}{4{,}66478 - (2{,}15831)^2} = 0{,}96459;$$

$$b = \overline{y}_1 - a_1 \overline{x}_1 = 1{,}63964 - 0{,}96459 \cdot 2{,}15831 = -0{,}44224;$$

$$b = \lg a, \quad \lg a_0 = -0{,}44224, \quad \text{звідси} \quad a_0 = 0{,}3612.$$

Підставимо значення параметрів у шукане рівняння (9.56), тоді рівняння 2 набуде вигляду:

$$\lg y = \lg 0{,}3612 + 0{,}96459 \lg x,$$

або

$$\lg y = 0{,}3612 + 0{,}96459 \lg x.$$

Рівняння кореляційного зв'язку набуде вигляду:

$$y = 0{,}3612 x^{0{,}96459}.$$

Параметр $a_1$ показує збільшення собівартості продукції на 0,96 % зі збільшенням випуску продукції на 1 %.

**Параболічна залежність. Вирівнювання по параболі другого порядку**. Якщо зв'язок між ознаками нелінійний і при рівномірному зростанні факторної ознаки має місце прискорене зростання або спадання результативної ознаки, то кореляційна залежність може бути виражена *параболою другого порядку*: $\overline{y}_t = a_0 + a_1 x + a_2 x^2$.

Розрахунок параметрів параболи другого порядку провадиться за методом найменших квадратів. Параметри $a_0$, $a_1$ і $a_2$ обчислюються шляхом вирішення системи 3 нормальних рівнянь:

$$\begin{cases} a_0 n + a_1 \sum x + a_2 \sum x^2 = \sum y; \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 = \sum yx; \\ a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 = \sum yx^2. \end{cases} \tag{9.61}$$

Слід зазначити, що в залежності від знаків параметрів $a_1$ і $a_2$ рівняння параболи другого порядку можуть відображати різні типи розвитку явищ.

При позитивних знаках параметрів $a_1$ і $a_2$ у ряді динаміки рівень явища збільшується з прискоренням. Якщо ж знак параметра $a_2$ буде від'ємним, а знак параметра $a_1$ залишиться позитивним, то це означає, що відбувається збільшення розвитку явища з тенденцією до уповільнення. Від'ємні знаки $a_1$ параметра говорять про зменшення розвитку явища. Але зменшення явища може відбуватися з подальшим напрямом у бік збільшення при позитивних значеннях параметра $a_2$. Параметр $a_2$ характеризує прискорення у розвитку явища, а також величину вигину параболи.

Методику розрахунку параметрів рівняння параболи другого порядку за індивідуальними значеннями розглянемо на прикладі зміни собівартості виробу.

Маються такі дані по 15 однотипних заводах:

| Номер робітника | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Випуск продукції ($x$), тис. шт. | 2 | 3 | 4 | 4 | 5 | 6 | 6 | 6 | 7 | 8 | 9 | 10 | 12 | 13 | 14 |
| Собівартість одного виробу *А* ($y$), грн | 8 | 10 | 7 | 6 | 5 | 5 | 4 | 3 | 4 | 5 | 3 | 2 | 1 | 1 | 2 |

Теоретично зв'язок між випуском продукції і собівартістю одиниці виробу полягає в наступному: при збільшенні випуску продукції собівартість одиниці виробу знижується, оскільки не всі витрати зростають пропорційно, частина з них не збільшується (залишається постійною) або знижується. Це зрештою призводить до зниження собівартості одиниці виробу. Однак за різних обсягів виробництва швидкість цього зниження різна.

Дані про собівартість продукції, складені в ряд і особливо графічне зображення (рис. 9.16) виявляють зниження собівартості одиниці виробу з тенденцією до уповільнення. Тому загальна тенденція розвитку цього ряду динаміки відповідає рівнянню параболи другого порядку $\overline{y}_t = a_0 + a_1 x + a_2 x^2$. Таким чином, форма зміни собівартості одного виробу криволінійна і виражається параболою другого порядку.

Вважаючи, що залежність описується рівнянням параболи другого порядку, визначимо за способом найменших квадратів коефіцієнти $a_0, a_1, a_2$.

Для визначення параметрів рівняння параболи другого порядку будується розрахункова таблиця.

Таблиця 9.20

| Номер п/п | Випуск продукції, тис. шт. $x$ | Собівартість одного виробу, грн $y$ | $x^2$ | $x^3$ | $x^4$ | $xy$ | $x^2y$ | $\overline{y}_x$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 8 | 4 | 8 | 16 | 16 | 32 | 9,0 |
| 2 | 3 | 10 | 9 | 27 | 81 | 30 | 90 | 7,8 |
| 3 | 4 | 7 | 16 | 64 | 256 | 28 | 112 | 6,6 |
| 4 | 4 | 6 | 16 | 64 | 256 | 24 | 96 | 6,6 |
| 5 | 5 | 5 | 25 | 125 | 625 | 25 | 125 | 5,6 |
| 6 | 6 | 5 | 36 | 216 | 1 296 | 30 | 180 | 4,7 |
| 7 | 6 | 4 | 36 | 216 | 1 296 | 24 | 144 | 4,7 |
| 8 | 6 | 3 | 36 | 216 | 1 296 | 18 | 108 | 4,7 |
| 9 | 7 | 4 | 49 | 343 | 2 401 | 28 | 196 | 3,9 |
| 10 | 8 | 5 | 64 | 512 | 4 096 | 40 | 320 | 3,2 |
| 11 | 9 | 3 | 81 | 729 | 6 561 | 27 | 243 | 2,6 |
| 12 | 10 | 2 | 100 | 1 000 | 10 000 | 20 | 200 | 2,2 |
| 13 | 12 | 1 | 144 | 1 728 | 20 736 | 12 | 144 | 1,6 |
| 14 | 13 | 1 | 169 | 2 197 | 28 561 | 13 | 169 | 1,4 |
| 15 | 14 | 2 | 196 | 2 744 | 38 416 | 28 | 392 | 1,4 |
| Всього | 109 | 66 | 981 | 10 189 | 115 893 | 363 | 2 551 | 66 |

Підставимо до системи нормальних рівнянь дані з табл. 9.20:

$$\begin{cases} \text{I.} \quad 15a_0 + 109a_1 + 981a_2 = 66; \\ \text{II.} \ 109a_0 + 981a_1 + 10\,189a_2 = 363; \\ \text{III.} \ 981a_0 + 10\,189a_1 + 115\,893a_2 = 2\,551. \end{cases}$$

Домножимо друге рівняння на 9 і віднімемо друге рівняння з третього рівняння:

$$\text{II.} \quad 981a_0 + 8829a_1 + 91701a_2 = 3267$$
$$-$$
$$\text{II.} \quad 981a_0 + 10189a_1 + 115893a_2 = 2551$$
$$A \qquad -1360\,a_1 - 24192\,a_2 = 716$$

Домножимо перше рівняння на 65,4 і віднімемо перше рівняння з третього рівняння:

$$\text{I.} \quad 981a_0 + 7128{,}6a_1 + 64157{,}4a_2 = 4316{,}4$$
$$-$$
$$\text{III.} \quad 981a_0 + 10189a_1 + 115893a_2 = 2551$$
$$B \qquad -3060{,}4a_1 - 51735a_2 = 1765{,}4\,.$$

Рівняння *A* домножимо на 2,25029 $\left(\frac{3\,060{,}4a_1}{1\,360a_1}\right)$ і віднімемо з рівняння *B* рівняння *A*:

$$B \qquad -\ 3\,060{,}4a_1 - 51735{,}6a_2 = 1765{,}4$$
$$-$$
$$A \qquad -\ 3\,060{,}4a_1 - 54439{,}016a_2 = 1611{,}208$$
$$2703{,}416a_2 = 154{,}192$$

звідки

$$a_2 = 0{,}0571\,.$$

Підставимо значення параметра $a_2$ в рівняння *A* і обчислимо параметр $a_1$:

$$-\ 1\,360\,a_1 - 24\,192 \cdot (0{,}0571) = 716;$$
$$-\ 1\,360a_1 = 2097{,}3632;$$
$$a_1 = -1{,}542.$$

Підставимо значення $a_1$ і $a_2$ рівняння I і обчислимо параметр $a_0$:

$$15a_0 + 109 \cdot (-1{,}542) + 981 \cdot (0{,}0571) = 66;$$
$$15a_0 - 168{,}078 + 56{,}0151 = 66;$$
$$15\,a_0 = 178{,}0629\,;$$
$$a_0 = 11{,}87.$$

Рівняння зв'язку, отримане способом найменших квадратів для визначення залежності між $x$ (випуском продукції) і $y$ (собівартістю одиниці виробу) набуде наступного вигляду:

$$y_x = 11{,}87 - 1{,}542x + 0{,}0571x^2.$$

Розрахункові показники собівартості одиниці виробу були визначені з рівняння, побудованого за способом найменших квадратів. Підставимо в це рівняння відповідні значення ознаки $x$, розрахуємо теоретичну лінію регресії:

$$y_1 = 11{,}87 - 1{,}542 \cdot 2 + 0{,}0571 \cdot 2^2 = 9{,}0;$$
$$y_2 = 11{,}87 - 1{,}542 \cdot 3 + 0{,}0571 \cdot 3^2 = 7{,}8;$$
$$y_3 = 11{,}87 - 1{,}542 \cdot 4 + 0{,}0571 \cdot 4^2 = 6{,}6 \text{ і т. д.}$$

Отримані значення результативної ознаки відображають середню залежність $y$ від $x$ у вигляді кореляційної залежності. Рівність сум $\sum \bar{y}_x = \sum y$ свідчить про правильність проведених розрахунків.

Лінії фактичної та розрахункової собівартості одиниці виробу, обчислені за рівнянням $y_x = 11{,}87 - 1{,}542x + 0{,}0571x^2$, наводяться на рис. 9.16. Графічний збіг фактичних і розрахункових даних показує майже повний збіг ходу обох ліній, що говорить про хороше відтворення емпіричних даних розрахунковими середніми значеннями результативної ознаки.

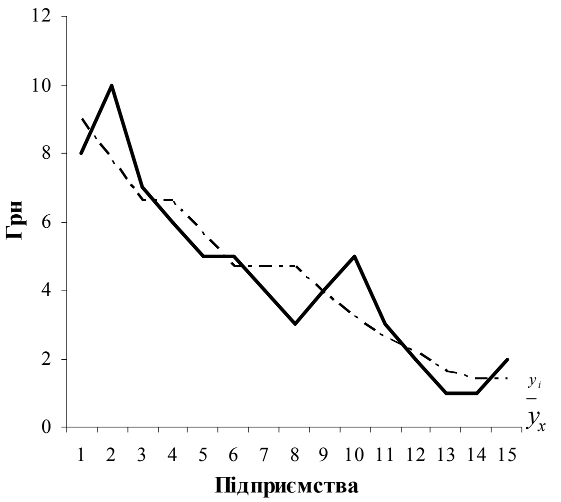



*Рис. 9.16*. Фактичні дані про виробництво виробу *А* та вирівняні за рівнянням параболи другого порядку.

Що стосується системи нормальних рівнянь, то підбором способу відліку можна спростити знаходження параметрів, оскільки

$$\sum t = 0 \text{ т } \sum t^3 = 0.$$

Якщо $\sum t$ і $\sum t^3$ (як і будь-яка сума непарних ступенів) дорівнюють нулю, то система рівнянь набуде вигляду:

$$\left.\begin{aligned} a_0 n + a_2 \sum t^2 &= \sum y; \\ a_1 \sum t^2 &= \sum yt; \\ a_0 \sum t^2 + a_2 \sum t^4 &= \sum yt^2. \end{aligned}\right\} \qquad (9.62)$$

Знайти значення $a_0$, $a_1$ і $a_2$ можна способом підстановки. Зокрема, $a_0$ знаходять з другого рівняння, а $a_1$ і $a_2$ легко можна отримати за допомогою визначників.

У табл. 9.21 наведено дані про випуск продукції по підприємствах. Вирівняємо ці дані по параболі другого порядку і зобразимо графічно.

Таблиця 9.21

| Показник | Роки | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
| Виробництво виробу *А* (тис. шт.) | 4 | 13 | 20 | 28 | 46 | 63 | 87 | 147 | 200 |

Для отримання сум, що входять до системи нормальних рівнянь, збудуємо розрахункову таблицю:

Таблиця 9.21

| Роки | Виробництво (в тис. шт.) виробу *А* ($y$) | $t$ | $t^2$ | $yt$ | $t^4$ * | $yt^2$ | $\bar{y}_t$ |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2014 | 4 | –4 | 16 | –16 | 256 | 64 | 12,7 |
| 2015 | 13 | –3 | 9 | –39 | 81 | 117 | 8,7 |
| 2016 | 20 | –2 | 4 | –40 | 16 | 80 | 12,3 |
| 2017 | 28 | –1 | 1 | –28 | 1 | 28 | 23,4 |
| 2018 | 46 | 0 | 0 | 0 | 0 | 0 | 42,2 |
| 2019 | 63 | 1 | 1 | 63 | 1 | 63 | 68,6 |
| 2020 | 87 | 2 | 4 | 174 | 16 | 348 | 102,6 |
| 2021 | 147 | 3 | 9 | 441 | 81 | 1 323 | 144,2 |
| 2022 | 200 | 4 | 16 | 800 | 256 | 3 200 | 193,3 |
| | 608 | 0 | 60 | –48<br>+1 478<br>1 355 | 708 | 5 223 | 608,0 |

* Підсумок колонки 6 може бути отриманий без зведення в 4 ступінь використанням формули $\sum t^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}$.

Маємо, при $n$=4, $\sum t^4 = \frac{4 \cdot 5 \cdot 9 \cdot 59}{30} = 354$. Подвоїмо результат і отримаємо 708.

Складаємо систему нормальних рівнянь:

$$
\begin{array}{rl}
\mathrm{I} & \left\{ \begin{array}{l} 9a_0+60a_2=608 \\ 60a_0+708a_2=5223 \end{array} \right. \\
\mathrm{II} & \\
\mathrm{III} & 60a_1=1355,
\end{array}
$$

звідки

$$a_1 = \frac{1355}{60} = 22{,}58333.$$

Ділимо перше рівняння на 9, друге на 60:

$$
-\left\{ \begin{array}{l} a_0+6{,}66667a_2=67{,}55556 \\ a_0+11{,}8a_2=87{,}05 \end{array} \right.
$$

$$-5{,}1333a_2=-19{,}49444$$

$$a_2 = \frac{-19{,}49444}{5{,}13333} = 3{,}79762.$$

Підставляємо значення $a_1$ і $a_2$ в перше рівняння і обчислимо $a_0$:

$$a_0=42{,}23808;$$
$$a_1=22{,}58333;$$
$$a_2=3{,}79762.$$

Проведемо перевірку знайдених параметрів[1]:

$$\text{I.}\quad 9\cdot 42{,}23808+60\cdot 3{,}79762=380{,}14272+227{,}85720= =607{,}99992\approx 608;$$

$$\text{II.}\quad 60\cdot 42{,}23808+708\cdot 3{,}79762=2534{,}28480+5222{,}99976= =5222{,}99976\approx 5223.$$

Аналітичне рівняння параболи другого порядку набуде наступного вигляду:

$$\overline{y}_t = 42{,}23808 + 22{,}58333 \cdot t + 3{,}79762 \cdot t^2.$$

Підставляючи значення $t$ (від – 4 до +4), отримаємо вирівняні значення виробництва виробу А (див. колонку 8 табл. 9.21).

$$\overline{y}_{t=-4} = 42{,}23808 + 22{,}58333 \cdot (-4) + 3{,}79762 \cdot (-4)^2 \approx 12{,}7$$
$$\overline{y}_{t=-3} = 42{,}23808 + 22{,}58333 \cdot (-3) + 3{,}79762 \cdot (-3)^2 \approx 8{,}7$$

---

[1] Після вирішення системи нормальних рівнянь завжди слід перевіряти знайдені параметри. Однак надалі проводити перевірку на сторінках підручника ми не будемо, надамо можливість читачеві зробити це самостійно, шляхом підставки в нормальні рівняння.

$$\bar{y}_{t=-2} = 42{,}23808 + 22{,}58333 \cdot (-2) + 3{,}79762 \cdot (-2)^2 \approx 12{,}3$$
$$\bar{y}_{t=-1} = 42{,}23808 + 22{,}58333 \cdot (-1) + 3{,}79762 \cdot (-1)^2 \approx 23{,}4$$
$$\bar{y}_{t=-0} = 42{,}23808 + 22{,}58333 \cdot (0) + 3{,}79762 \cdot (0)^2 \approx 42{,}2$$
$$\bar{y}_{t=1} = 42{,}23808 + 22{,}58333 \cdot (1) + 3{,}79762 \cdot (1)^2 \approx 68{,}6$$
$$\bar{y}_{t=2} = 42{,}23808 + 22{,}58333 \cdot (2) + 3{,}79762 \cdot (2)^2 \approx 102{,}6$$
$$\bar{y}_{t=3} = 42{,}23808 + 22{,}58333 \cdot (3) + 3{,}79762 \cdot (3)^2 \approx 144{,}2$$
$$\bar{y}_{t=4} = 42{,}23808 + 22{,}58333 \cdot (4) + 3{,}79762 \cdot (4)^2 \approx 193{,}3$$
$$608{,}0$$

Зобразимо графічно фактичні дані та вирівняні за параболою другого порядку (рис. 9.17). Будуємо по парах *x* і *y* ламану криву. Потім кожну пару чисел нанесемо на графік. Ламана крива і крива, розрахована на основі рівняння регресії, знаходяться дуже близько одна до одної, що говорить про гарне відтворення фактичних даних розрахунковими значеннями результативної ознаки.

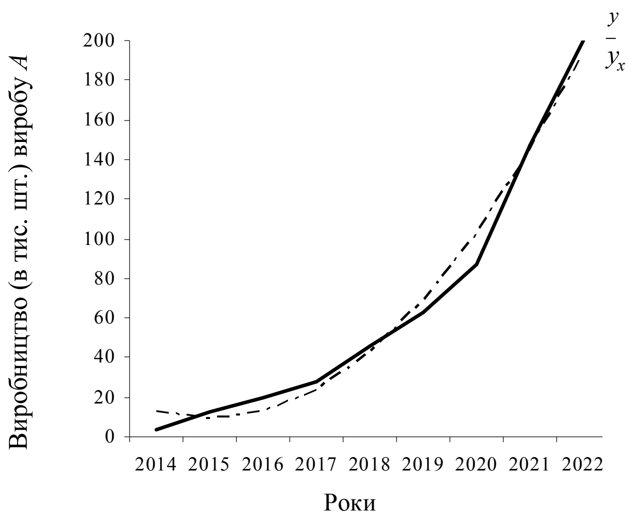



*Рис. 9.17.* Фактичні та вирівняні дані виробництва виробів *A* за 2014–2022 рр.

У практиці вивчення кореляційних взаємозалежностей, крім параболи 2-го порядку, застосовуються параболи і вищого порядку. Чим вищий порядок параболи, тим точніше вона відтво рює фактичні дані.

Якщо рівняння зв’язку є параболу 3-го порядку , то система нормальних рівнянь набуде вигляду:

$$\begin{cases} na_0 + a_1\sum x + a_2\sum x^2 + a_3\sum x^3 = \sum y; \\ a_0\sum x + a_1\sum x^2 + a_2\sum x^3 + a_3\sum x^4 = \sum yx; \\ a_0\sum x^2 + a_1\sum x^3 + a_2\sum x^4 + a_3\sum x^5 = \sum yx^2; \\ a_0\sum x^3 + a_1\sum x^4 + a_2\sum x^5 + a_3\sum x^6 = \sum yx^3. \end{cases} \qquad (9.63)$$

Для знаходження потрібних сум рівняння параболи 3-го порядку складають розрахункову таблицю за такою схемою:

| $x$ | $y$ | $x^2$ | $x^3$ | $x^4$ | $x^6$ | $x^6$ | $yx$ | $yx^2$ | $yx^3$ |
|---|---|---|---|---|---|---|---|---|---|

Ці дані таблиці застосовуються для знаходження необхідних сум. Розв'язавши систему 4 рівнянь, знайдемо параметри $a_0, a_1, a_2, a_3$, отже, рівняння зв'язку. Підставляючи значення факторної ознаки ($x$) у рівняння регресії, отримуємо вирівняні (розрахункові) значення результативної ознаки ($y$).

Якщо розрахунок здійснюється не за індивідуальними даними, а на основі аналітичного групування, то система рівнянь методом найменших квадратів набуває такого вигляду:

$$a_0\sum_{j=1}^{k} f_j + a_1\sum_{j=1}^{k} x_j f_j + a_2\sum_{j=1}^{k} x_j^2 f_j = \sum_{j=1}^{k} y_j f_j;$$

$$a_0\sum_{j=1}^{k} x_j f_j + a_1\sum_{j=1}^{k} x_j^2 f_j + a_2\sum_{j=1}^{k} x_j^3 f_j = \sum_{j=1}^{k} y_j x_j f_j; \qquad (9.64)$$

$$a_0\sum_{j=1}^{k} x_o^2 f_j + a_1\sum_{j=1}^{k} x_j^3 f_j + a_2\sum_{j=1}^{k} x_j^4 f_j = \sum_{j=1}^{k} y_j x_j^2 f_j.$$

Вирішуючи систему рівнянь, знаходять параметри $a_0, a_1, a_2$. Методику розрахунку параметрів рівняння покажемо на прикладі.

Зв'язок між випуском виробів та витратами на упаковку характеризуються такими даними по 50 однотипних заводах:

Таблиця 9.22

| Вартість упаковки, тис. грн $y$ | Випуск виробів, тис. шт. $x$ | | | | Кількість заводів |
|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | |
| 2 | 1 | | | | 1 |
| 4 | 5 | 4 | 2 | | 11 |
| 6 | 4 | 10 | 8 | 1 | 23 |
| 8 | | 1 | 5 | 6 | 12 |
| 10 | | | | 3 | 3 |
| Кількість заводів | 10 | 15 | 15 | 10 | 50 |

Вважаючи, що залежність між випуском виробів і вартістю упаковки описується рівнянням параболи другого порядку, визначимо коефіцієнти $a_0$, $a_1$ і $a_2$.

Для розрахунку параметрів рівняння параболи другого порядку складемо розрахункову таблицю:

Таблиця 9.23

**Кореляційна таблиця**

| $y$ \ $x$ | 2 | 3 | 4 | 5 | $f_y$ | $yf_y$ | $xyf_y$ | $x^2yf_{xy}$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 1 | | | | 1 | 2 | 4 | 8 |
| 4 | 5 | 4 | 2 | | 11 | 44 | 120 | 352 |
| 6 | 4 | 10 | 8 | 1 | 23 | 138 | 450 | 1554 |
| 8 | | 1 | 5 | 6 | 12 | 96 | 424 | 1912 |
| 10 | | | | 3 | 3 | 30 | 150 | 750 |
| $f_x$ | 10 | 15 | 15 | 10 | 50 | 310 | 1148 | 4576 |
| $xf_x$ | 20 | 45 | 60 | 50 | 175 | | | |
| $x^2f_x$ | 40 | 135 | 240 | 250 | 665 | | | |
| $x^3f_x$ | 80 | 405 | 960 | 1250 | 2695 | | | |
| $x^4f_x$ | 160 | 1215 | 3840 | 6250 | 11465 | | | |
| $\bar{y}_i$ | 4,60 | 5,27 | 6,60 | 8,60 | | | | |

Система рівнянь має такий вигляд:

$$50a_0 + 175a_1 + 665a_2 = 310;$$
$$175a_0 + 665a_1 + 2695a_2 = 1148;$$
$$665a_0 + 2695a_1 + 11465a_2 = 4576.$$

Перетворимо систему шляхом розподілу відповідних значень на коефіцієнти при перших членах:

$$a_0 + 3,5a_1 + 13,3a_2 = 6,2;$$
$$a_0 + 3,8a_1 + 15,4a_2 = 6,56;$$
$$a_0 + 4,0526a_1 + 17,2406a_2 = 6,8857.$$

Віднімемо від другого рівняння перше та від третього рівняння друге, отримаємо систему рівнянь із двома невідомими:

$$0,3\,a_1 + 2,1\,a_2 = 0,36;$$
$$0,2526\,a_1 + 1,8406\,a_2 = 0,3257.$$

Виразимо з першого рівняння показник $a_1$:

$$a_1 = \frac{0,36 - 2,1a_2}{0,3},$$

і підставимо його в друге рівняння:

$$0,2526 \cdot \frac{0,36 - 2,1a_2}{0,3} + 1,8406\,a_2 = 0,3257.$$

Вирішуючи систему рівнянь, отримаємо:

$$a_0=5{,}49305; \quad a_1=-0{,}98316; \quad a_2=0{,}31188.$$

Отже, рівняння взаємозв'язку між випуском виробів та вартістю упаковки має такий вигляд:

$$\overline{y}_x = 5{,}49305 - 0{,}98316x + 0{,}31188x^2.$$

Підставляючи значення факторної ознаки ($x$) в рівняння, отримаємо значення результативної ознаки ($y$).

**Вирівнювання по напівлогарифмічній кривій**. Для прикладу вирівнювання по напівлогарифмічній кривій охарактеризуємо залежність між виторгом продавця ($y$) і зростанням загального товарообігу магазину ($x$). Між цими показниками існує наступний зв'язок: із зростанням товарообороту магазину розмір товарообігу, що припадає на одного продавця (тобто виторг продавця, зростає). Однак зростання це не постійне. Середній виторг на одного продавця в дрібних магазинах росте досить швидко, а потім з укрупненням підприємства темпи її зростання уповільнюються. Найкраще відображає таку залежність напівлогарифмічна функція: $\overline{y}_x = a_0 + a_1 \lg x.$

Покажемо застосування цієї функції на конкретному прикладі (табл. 9.24).

Для знаходження параметрів напівлогарифмічної функції необхідно вирішити систему двох рівнянь:

$$a_0 n + a_1 \sum \lg x = \sum y;$$
$$a_0 \sum \lg x + a_1 \sum (\lg x)^2 = \sum y \lg x. \qquad (9.65)$$

Згідно табл. 9.24

$$n=24; \quad \sum \lg x = 38{,}4431; \quad \sum (\lg x)^2 = 63{,}58;$$

$$\sum y=137{,}4; \quad \sum y \lg x = 225{,}74.$$

Отже, рівняння приймуть вид:

$$24a_0 + 38{,}4431a_1 = 137{,}4;$$
$$38{,}4431a_0 + 63{,}58a_1 = 225{,}74.$$

Прирівняємо коефіцієнти при $a_0$, помножуючи перше рівняння на 1,602 $\left(\frac{38{,}4431}{24}\right)$. Отримуємо: $38{,}4431\,a_0+61{,}586\,a_1=220{,}115$. Віднімаємо з другого рівняння знову отримане: $1{,}994\,a_1=5{,}625$, звідки $a_1=2{,}821$. Підставляючи значення перше рівняння, маємо: $24\,a_0+38{,}4431\cdot 2{,}821=137{,}4$; звідки $a_0=1{,}206$.

Таблиця 9.24

**Вирівнювання по напівлогарифмічній кривій**

| $x$ | $y$ | $\log x$ | $(\log x)^2$ | $y \log x$ | $\overline{y}_x$ |
|---|---|---|---|---|---|
| 11,2 | 4,6 | 1,0492 | 1,101 | 4,826 | 4,2 |
| 11,4 | 4,7 | 1,0569 | 1,117 | 4,967 | 4,2 |
| 16,7 | 4,6 | 1,2227 | 1,495 | 5,624 | 4,7 |
| 17,8 | 4,9 | 1,2504 | 1,564 | 6,127 | 4,7 |
| 22,1 | 4,5 | 1,3444 | 1,807 | 6,050 | 5,0 |
| 22,9 | 4,7 | 1,3598 | 1,849 | 6,391 | 5,0 |
| 24,1 | 5,0 | 1,3820 | 1,910 | 6,910 | 5,1 |
| 24,8 | 4,7 | 1,3945 | 1,944 | 6,554 | 5,1 |
| 28,6 | 5,9 | 1,4564 | 2,121 | 8,593 | 5,3 |
| 36,5 | 5,1 | 1,5623 | 2,441 | 7,968 | 5,6 |
| 37,4 | 6,1 | 1,5729 | 2,474 | 9,595 | 5,6 |
| 38,2 | 5,2 | 1,5821 | 2,503 | 8,227 | 5,7 |
| 43,1 | 5,0 | 1,6345 | 2,672 | 8,172 | 5,8 |
| 56,2 | 6,3 | 1,7497 | 3,062 | 11,023 | 6,1 |
| 50,1 | 5,9 | 1,6998 | 2,889 | 10,029 | 6,0 |
| 61,6 | 6,3 | 1,7896 | 3,203 | 11,274 | 6,3 |
| 70,4 | 7,2 | 1,8476 | 3,414 | 13,303 | 6,4 |
| 74,3 | 6,8 | 1,8710 | 3,501 | 12,723 | 6,5 |
| 74,7 | 6,1 | 1,8733 | 3,509 | 11,427 | 6,5 |
| 75,9 | 6,2 | 1,8802 | 3,535 | 11,657 | 6,5 |
| 79,2 | 6,5 | 1,8987 | 3,605 | 12,342 | 6,6 |
| 85,8 | 7,0 | 1,9335 | 3,738 | 13,534 | 6,7 |
| 103,1 | 7,0 | 2,0133 | 4,053 | 14,093 | 6,9 |
| 104,3 | 7,1 | 2,0183 | 4,073 | 14,330 | 6,9 |
| $\sum$ | 137,4 | 38,4431 | 63,580 | 225,740 | 137,4 |

Отже, рівняння зв’язку отримуємо таке:

$$\overline{y}_x = 1{,}206 + 2{,}821 \lg x.$$

На основі цього рівняння розрахуємо значення (див. останню стовпчик табл. 9.24). Порівняємо розрахункові та фактичні дані на рис. 9.18.

У разі, якщо складно визначити форму зв’язку між ознаками, то можуть бути побудовані різні рівняння регресії (наприклад, рівняння прямої та гіперболи), які надалі піддаються порівнянню. Більш адекватним вважається те рівняння регресії, у якому середні відхилення теоретичних рівнів від емпіричних (по модулю або залишкова дисперсія). Для оцінки адекватності рівняння може бути застосований осо-

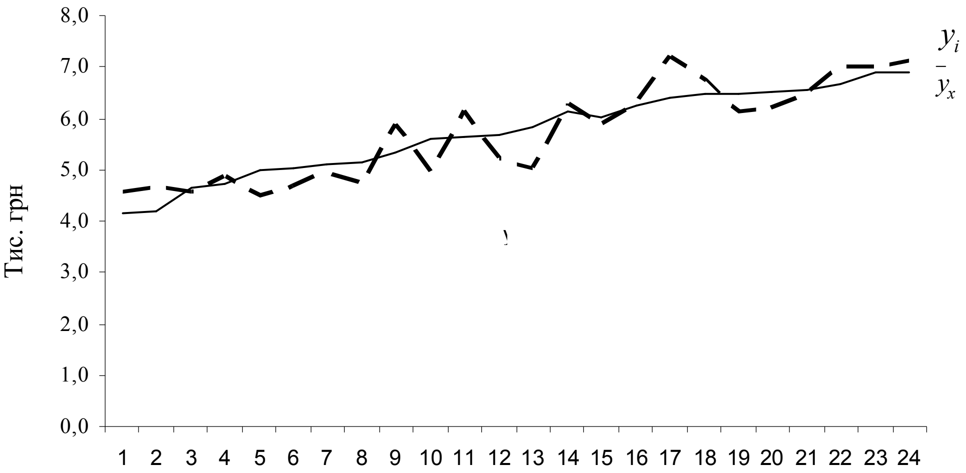



*Рис. 9.18.* Фактичні дані та вирівняні за напівлогарифмічною кривою.

бливий показник — середня помилка апроксимації[1]:

$$\bar{\varepsilon} = \frac{1}{n}\sum \frac{\left|y - \bar{y}_x\right|}{y} \cdot 100. \qquad (9.66)$$

Критичний рівень цієї помилки встановлюється статистиком залежно від конкретного матеріалу і задач, які стоять перед дослідженням (у прикладі з годинною заробітною платою і виробничим стажем $\bar{\varepsilon}$= 6,5%). При порівнянні $\bar{\varepsilon}$, обчислених за різними моделями, більш адекватною вважається модель із меншою середньою помилкою.

## 9.11. Парний коефіцієнт кореляції та кореляційне відношення

**Індекс кореляції (кореляційне ставлення)**. Рівняння зв'язку вирішує перше завдання кореляції — математично виражає характер зв'язку, тобто показує, як змінюється результативна ознака під факторний вплив. Але воно не дає єдиної міри тісноти зв'язку. Тому для вимірювання тісноти зв'язку потрібен відносний показник. При лінійній і криволінійній залежності між ознаками тіснота зв'язку між результативним і факторним ознаками визначається за допомогою показника — *теоретичного кореляційного відношення*, або *індексу*

[1] Апроксимація — наближений вираз одних величин чи геометричних образів через інші, простіші. Наприклад, при обчисленні довжин кривих ліній вдаються до апроксимації кривих ліній ламаними. За наведеною формулою розраховується ступінь наближення ординат обчисленої теоретичної лінії ($\bar{y}_x$) над фактичним ординатам ($y$).

*кореляції*. Розрахунок цього показника провадиться за допомогою теореми складання дисперсій.

У восьмому розділі наведено правило складання дисперсій, згідно з яким загальна дисперсія ознаки ($\sigma^2$) завжди дорівнює сумі середній внутрішньогрупових дисперсій ($\overline{\sigma_j^2}$) і міжгрупової дисперсії ($\delta^2$). Міжгрупова дисперсія відображає коливання результативної ознаки, обумовлену дією факторної ознаки, а внутрішньогрупова — варіацію результативної ознаки під впливом всіх інших факторів (залишкова дисперсія). Очевидно, що чим сильніший вплив факторної ознаки на результативну, тим у більшій мірі дисперсія цієї ознаки наближається до загальної дисперсії.

Розрахунок коефіцієнта кореляції заснований на показниках, що характеризують варіацію результативної ознаки *y*. Варіація ознаки може бути виміряна дисперсією цієї ознаки, тобто величиною $\sigma_y^2 = \frac{\sum(y-\bar{y})^2}{n}$ $\left(\sigma_y^2 = \frac{\sum(y-\bar{y})^2 f}{\sum f}\right)$. Ця дисперсія називається загальною і є варіацією результативної ознаки *y* від дії всіх факторів. Якщо ж вирішити рівняння зв'язку і знайти вирівняні значення $\bar{y}_x$, то дисперсія ознаки за вирівняними значеннями факторної ознаки ($\bar{y}_x$) дорівнюватиме залишковій дисперсії (внутрішньогруповій): $\sigma^2_{(y-\bar{y}_x)} = \frac{\sum(y-\bar{y}_x)^2}{n}$ $\left(\overline{\sigma_j^2} = \frac{\sum \sigma_j^2 f_j}{\sum f_j}\right)$. Залишкова дисперсія характеризує варіацію результативної ознаки за рахунок інших факторів, крім *x* (тобто при виключеному *x*).

Різниця між $\sigma_y^2 - \sigma^2_{(y-\bar{y}_x)}$ вимірює результативної ознаки *y* за рахунок дії факторної ознаки *x*. Ця різниця дорівнює дисперсії факторної ознаки (міжгруповій дисперсії) $\delta_x^2 = \frac{\sum(\bar{y}_x-\bar{y})^2}{n}$ $\left(\delta_x^2 = \frac{\sum(x_j-\bar{x})^2 f_j}{\sum f_j}\right)$. Міжгрупова дисперсія показує розсіювання, що виникає за рахунок факторіальної ознаки. Отже, чим більше факторна ознака впливає на результативну, тим більша частина коливань буде припадати за рахунок фактора *x*, тобто на величину $\delta_x^2$. Тоді за правилом додавань дисперсій $\sigma_y^2 = \sigma^2_{(y-\bar{y}_x)} + \delta_x^2$. Різниця між загальною дисперсією, що вимірює відхилень *y* від $\bar{y}$, і дисперсією, що вимірює відхилення *y* від $\bar{y}_x$ дозволить нам обчислити дисперсію, що вимірює варіацію, обумовлену фактором *x*.

Отже, виміряти частку варіації, викликаної даним фактором $x$ у загальній величині результативної ознаки, можна відношенням факторної дисперсії до загальної дисперсії. Ця частка являтиме собою відношення $D = \frac{\delta_x^2}{\sigma_y^2}$, яке називають *індексом* (*коефіцієнтом*) *детермінації*. Коефіцієнт детермінації представляє відношення дисперсії, обумовленої впливом факторної ознаки, що лежить в основі групування, тобто міжгрупової дисперсії $\delta_x^2$, до загальної дисперсії, що вимірює всі коливання. Це відношення показує, яку частину загальної варіації досліджуваної ознаки становить варіація міжгрупова, тобто обумовлена групувальною ознакою.

Розрахунок загальної дисперсії, внутрішньогрупової та міжгрупової дисперсій дозволяє зробити деякі висновки про міру впливу факторної ознаки на коливання ознаки результативної. В якості показника міри тісноти зв'язку застосовується квадратний корінь із цього відношення. Корінь квадратний з цього відношення називається *індексом кореляції*, або *теоретичним кореляційним відношенням*. Зіставивши величину групової та загальної дисперсії, можна визначити ступінь впливу факторної ознаки, покладеної в основу групування, на коливання результативної ознаки.

Теоретичне кореляційне ставлення ($\eta$) розраховується за такою формулою:

$$\eta = \sqrt{\frac{\delta_x^2}{\sigma_y^2}}. \tag{9.67}$$

де $\delta_x^2$ — характеризує варіацію результативної ознаки під впливом варіації ознаки фактора $\left(\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2\right)$;

$\sigma_y^2$ — характеризує варіацію результативної ознаки під впливом усіх факторів $\left(\sigma_{y_x}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_x)^2\right)$.

Виходячи з правила складання дисперсій ($\delta_x^2 = \sigma_y^2 - \sigma_{(y-\bar{y}_x)}^2$) індекс кореляції, або теоретичне кореляційне відношення, можна подати у вигляді

$$\eta = \sqrt{\frac{\sigma_y^2 - \sigma_{(y-\bar{y}_x)}^2}{\sigma_y^2}}, \quad \text{або} \quad \eta = \sqrt{1 - \frac{\sigma_{(y-\bar{y}_x)}^2}{\sigma_y^2}}, \tag{9.68}$$

де $\sigma_{y-\bar{y}_x}^2$ — характеризує варіацію результативної ознаки під впливом інших неврахованих факторів;

$\sigma_y^2$ — характеризує варіацію результативної ознаки під впливом усіх факторів.

Найчастіше індекс кореляції обчислюють саме за цією формулою.

Кореляційне відношення показує, яку частку загальної міри розсіювання (дисперсії) займає дисперсія, що виникає за рахунок факторіальної ознаки. Цей коефіцієнт застосовується для всіх випадків зв'язку.

Якщо при розрахунку кореляційного відношення буде взята дисперсія факторного ознаки, обчислена не за теоретичним значенням $\bar{y}_x$ $\left(\frac{\sum(\bar{y}_x-\bar{y})^2 f}{\sum f}\right)$, а за груповими середніми $\bar{y}_i$ $\left(\frac{\sum(\bar{y}_i-\bar{y})^2 n_i}{\sum n_i}\right)$, тобто за варіантом формули, даної в дев'ятому розділі, то такий показник буде називатися *емпіричним кореляційним відношенням*.

Величина кореляційного відношення змінюється в межах від 0 до 1. Якщо даний фактор *x* не впливає на результативну ознаку *y*, то дисперсія $\delta_x^2$ виявиться рівною нулю, тобто $\sigma^2_{y-\bar{y}_x}=\sigma_y^2$ і індекс кореляції дорівнюватиме нулю. Це свідчить про відсутність зв'язку між *x* та *y*. Графічно лінія $\bar{y}_x$ буде збігатися з лінією $\bar{y}$, тобто приймає горизонтальне положення і ніяк не реагує на зміну величини *x*. Якщо ж фактор *x* повністю визначає значення зміни результативної ознаки *y*, тобто між ними існує функціональний зв'язок, то вся його варіація визначається зміною даного фактора. Це відбувається, коли $\sigma^2_{y-\bar{y}_x}$ =0, тобто лінія на графіці повністю зіллється з лінією *y*. У цьому випадку загальна дисперсія ($\sigma_y^2$) виявиться рівною дисперсії за рахунок фактора *x* ($\delta_x^2$), а коефіцієнт кореляції — одиниці, тобто $\eta=\sqrt{\frac{\sigma_y^2-0}{\sigma_y^2}}$. Рівність коефіцієнта одиниці вказує на функціональний зв'язок між результативною і факторною ознакою. Звідси випливає, що *чим ближче індекс кореляції до 1*, тим зв'язок між ознаками тісніше.

Розрахуємо величину кореляційного відношення за даними залежності вироблення цукру *y* від величини основних виробничих засобів *x* (розрахунок $\sigma_x$ і $\sigma_y$ зробимо за формулами $\sigma_x=\sqrt{\overline{x^2}-\bar{x}^2}$; $\sigma_y=\sqrt{\overline{y^2}-\bar{y}^2}$ — див главу восьму). Для розрахунку за відповідними формулами слід знайти $\bar{x}$, $\bar{y}$, $\overline{xy}, \sigma_x, \sigma_y, \sigma_y^2, \sigma^2_{(y-\bar{y}_x)}$. Побудуємо розрахункову таблицю, у якій введемо і необхідні розрахунку додаткові графи (табл. 9.25).

Таблиця 9.25

**Розрахункова таблиця для знаходження коефіцієнта кореляції між вартістю основних фондів та виробленням цукру**

| $x$ | $y$ | $x^2$ | $y^2$ | $y-\bar{y}$ | $(y-\bar{y})^2$ | $\bar{y}_x$ | $y-\bar{y}_x$ | $(y-\bar{y}_x)^2$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 2 | 210 | 4 | 44 100 | –104,4 | 10 899,36 | 147,5 | +62,50 | 3 906,2500 |
| 2,8 | 135 | 7,84 | 18 225 | –179,4 | 32 184,36 | 186,77 | –51,77 | 2 680,1329 |
| 4 | 240 | 16 | 57 600 | –74,4 | 5 535,36 | 245,68 | –5,68 | 32,2624 |
| 4,5 | 264 | 20,25 | 69 696 | –50,4 | 2 540,16 | 270,22 | –6,22 | 38,6884 |
| 5 | 300 | 25 | 90 000 | –14,4 | 207,36 | 294,76 | +5,24 | 27,4576 |
| 5,7 | 310 | 32,49 | 96 100 | –4,4 | 19,36 | 329,13 | –19,13 | 365,9569 |
| 6,5 | 365 | 42,25 | 133 225 | +50,6 | 2 560,36 | 368,4 | –3,40 | 11,5600 |
| 7 | 380 | 49 | 144 400 | +65,6 | 4 303,36 | 392,94 | –12,94 | 167,4436 |
| 7,8 | 440 | 60,84 | 193 600 | +125,6 | 15 775,36 | 432,21 | +7,79 | 60,6841 |
| 8,7 | 500 | 75,69 | 250 000 | +185,6 | 34 447,36 | 476,39 | +23,61 | 557,4321 |
| $\sum$54 | 3144 | 333,36 | 1 096 946 | 0 | 108 472,40 | 3 144,00 | –99,14<br>+99,14 | 7 847,8680 |

Звідси

$$\bar{x} = \frac{\sum x}{n} = \frac{54}{10} = 5{,}4;$$

$$\overline{x^2} = \frac{\sum x^2}{n} = \frac{333{,}36}{10} = 33{,}336;$$

$$\bar{y} = \frac{\sum y}{n} = \frac{3144}{10} = 314{,}4;$$

$$\overline{y^2} = \frac{\sum y^2}{n} = \frac{1096946}{10} = 109694{,}6;$$

$$\bar{x}^2 = 5{,}4^2 = 29{,}16; \quad \bar{y}^2 = 314{,}4^2 = 98847{,}36;$$

$$\overline{xy} = \frac{\sum xy}{n} = \frac{19027{,}5}{10} = 1902{,}75;$$

$$\sigma_x = \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{33{,}336 - 29{,}16} = \sqrt{4{,}176} \approx 2{,}044;$$

$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{109694{,}6 - 98847{,}36} = \sqrt{10847{,}24} \approx 104{,}15;$$

$$\sigma_y^2 = 10847{,}24; \text{ або } \sigma_y^2 = \frac{\sum (y-\bar{y})^2}{n} = \frac{108472{,}4}{10} = 10847{,}24;$$

$$\sigma^2_{(y-\bar{y}_x)} = \frac{\sum (y-\bar{y}_x)^2}{n} = \frac{7847{,}868}{10} = 784{,}7868.$$

Варіація виробництва цукру за рахунок всіх факторів виявилася рівною $\sigma_y$ =10847,24, варіація за рахунок всіх факторів, крім вартості основних промислово-виробничих засобів, дорівнює $\sigma^2_{(y-\bar{y}_x)}$ =784,7868. Отже, варіація виробництва цукру за рахунок зміни середньорічної вартості основних промислово-виробничих засобів $\delta_x^2 = \sigma_y^2 - \sigma^2_{(y-\bar{y}_x)}$ дорівнює 10062,4532.

І нарешті, підставляємо отримані вихідні дані у формулу $\eta$ :

$$\eta = \sqrt{1 - \frac{\sigma^2 (y-\bar{y}_x)}{\sigma_y^2}} = \sqrt{1 - \frac{784{,}7868}{10847{,}22}} = \sqrt{1-0{,}0723} =$$

$$= \sqrt{0{,}9277} = 0{,}963.$$

Обчислений індекс кореляції свідчить про досить тісний зв'язок змінних величин *x* і *y*. Коливання вироблення цукру на 96,3 % залежить від коливання середньорічної вартості основних промислово-виробничих засобів. Таким чином, зміна виробництва цукру значною мірою визначається зміною середньорічної вартості основних промислово-виробничих засобів.

Кореляційне відношення розраховуємо за такою формулою:

$$\eta = \sqrt{\frac{\delta_x^2}{\sigma_y^2}}. \qquad (9.69)$$

Дисперсія $\delta_x^2$ визначається за формулою:

$$\delta_x^2 = \frac{\sum (\bar{y}_x - \bar{y})^2}{n}. \qquad (9.70)$$

Дисперсія $\sigma_y^2$ визначається за формулою:

$$\sigma_y^2 = \frac{\sum (y - \bar{y})^2}{n}. \qquad (9.71)$$

Для розрахунку кореляційного відношення складемо розрахункову таблицю (табл. 9.26).

Дисперсія $\delta_x^2$, яка характеризує варіацію результативної ознаки під впливом варіації ознаки-фактора, складе:

$$\delta_x^2 = \frac{\sum (\bar{y}_x - \bar{y})^2}{n} = \frac{100626{,}53}{10} \approx 10062{,}6.$$

Дисперсія $\sigma_y^2$, що характеризує варіацію результативної ознаки

Таблиця 9.26

| Номер п/п | $x$ | $y$ | $y_x = 49{,}324 + 49{,}088x$ | $\bar{y}_x - \bar{y}$ | $(\bar{y}_x - \bar{y})^2$ | $y - \bar{y}$ | $(y - \bar{y})^2$ | $y - \bar{y}_x$ | $(y - \bar{y}_x)^2$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 210 | 147,50 | –166,90 | 27 855,61 | –104,40 | 10 899,36 | +62,50 | 3 906,25 |
| 2 | 2,8 | 135 | 186,77 | –127,63 | 16 289,42 | –179,40 | 32 184,36 | –51,77 | 2 680,13 |
| 3 | 4 | 240 | 245,68 | –68,72 | 4 722,44 | –74,40 | 5 535,36 | –5,68 | 32,26 |
| 4 | 4,5 | 264 | 270,22 | –44,18 | 1 951,87 | –50,40 | 2 540,16 | –6,22 | 38,69 |
| 5 | 5 | 300 | 294,76 | –19,64 | 385,73 | –14,40 | 207,36 | +5,24 | 27,46 |
| 6 | 5,7 | 310 | 329,13 | +14,73 | 216,97 | –4,40 | 19,36 | –19,13 | 365,96 |
| 7 | 6,5 | 365 | 368,40 | +54,00 | 2 916,00 | +50,60 | 2 560,36 | –3,40 | 11,56 |
| 8 | 7 | 380 | 392,94 | +78,54 | 6 168,53 | +65,60 | 4 303,36 | –12,94 | 167,44 |
| 9 | 7,8 | 440 | 432,21 | +117,81 | 13 879,20 | +125,60 | 15 775,36 | +7,79 | 60,68 |
| 10 | 8,7 | 500 | 476,39 | +161,99 | 26 240,76 | +185,60 | 34 447,36 | +23,61 | 557,43 |
| $\sum$ | 54 | 3 144 | 3 144,00 | 0 | 100 626,53 | | 108 472,00 | 0 | 7 847,87 |

під впливом всіх факторів, складе:

$$\sigma_y^2 = \frac{\sum (y - \bar{y})^2}{n} = \frac{108472}{10} = 10847{,}2.$$

Кореляційне відношення складе:

$$\eta = \sqrt{\frac{\delta_x^2}{\sigma_y^2}} = \sqrt{\frac{10062{,}6}{10847{,}2}} \approx 0{,}963.$$

Розрахуємо коефіцієнт детермінації за формулою:

$$D = \frac{\delta_x^2}{\sigma_y^2} = \frac{10062{,}4532}{10847{,}24} = 0{,}93.$$

Коефіцієнт детермінації, рівний 0,93, показує, що варіант результативної ознаки (вироблення цукру) на 93 % відбувається під впливом варіації середньорічної вартості основних промислово-виробничих засобів, а 7 % — під впливом інших неврахованих факторів.

Коефіцієнт кореляції застосовується для оцінки тісноти зв’язку за будь-якої форми зв’язку як прямолінійної, так і криволінійної.

**Лінійний коефіцієнт кореляції**. Кореляційне відношення застосовується для всіх випадків зв’язку. При лінійній залежності між ознаками визначення тісноти кореляційної зв’язку застосовується лінійний коефіцієнт кореляції. Коефіцієнт кореляції є окремим випадком загального коефіцієнта кореляції. За величиною лінійного коефіцієнта кореляції можна судити про рівень впливу факторної ознаки на результативний показник. Коефіцієнт кореляції встановлює ступінь відповідності між коливаннями двох змінних близько їх середніх рівнів. Крім того, коефіцієнт кореляції вказує напрямок досліджуваних залежностей між двома ознаками. Обчислення коефіцієнта кореляції передбачає, що між змінними існує прямолінійна залежність (прямолінійна регресія). При іншій формі рівняння регресії коефіцієнт кореляції втрачає свою значущість, чим менше рівняння регресії може бути представлено прямий.

*Лінійний коефіцієнт кореляції* визначається за такою формулою:

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y}, \tag{9.72}$$

або

$$r = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\sqrt{\left[\frac{\sum (x-\overline{x})^2}{n}\right] \cdot \left[\frac{\sum (y-\overline{y})^2}{n}\right]}}, \tag{9.73}$$

де $\overline{xy}$ — середнє значення добутку ознаки $x$ на ознаку $y$;

$\overline{x}$ і $\overline{y}$ — середні значення відповідних ознак;

$\sigma_x$ і $\sigma_y$ — середні квадратичні відхилення ознак $x$ і $y$.

На відміну від кореляційного відношення коефіцієнт кореляції показує не тільки тісноту, а й напрям зв'язку. Коефіцієнт кореляції може приймати будь-які значення від – 1 до +1. Якщо величина лінійного коефіцієнта кореляції позитивна, це говорить про прямий зв'язок між досліджуваними ознаками; якщо вона від'ємна — про зворотній. Якщо коефіцієнт кореляції дорівнює ±1, значить фактор $y$ повністю залежить від $x$, тобто по суті кореляційна залежність збігається з функціональною. У разі, якщо коефіцієнт кореляції дорівнює 0, то зв'язок між ознаками відсутній, або кажуть, що $x$ і $y$ не корелюють. В цьому випадку $\overline{xy} = \overline{x} \cdot \overline{y}$. Якщо $\overline{xy} > \overline{x} \cdot \overline{y}$, зв'язок прямий і лінійний коефіцієнт кореляції $r$ має знак плюс. При $\overline{xy} < \overline{x} \cdot \overline{y}$ — зв'язок зворотний, $r$ приймає знак мінус. Якщо $\overline{xy} - \overline{x} \cdot \overline{y} = \sigma_x \cdot \sigma_y$ — зв'язок функціональний, повний і $r$ дорівнює 1 (+1 — повний зв'язок; –1 — повний зворотний зв'язок). При $r = +1$ або $r = -1$ всі точки лежать на прямій лінії. У цьому випадку має місце строга пропорційність у зміні $y$ і $x$. Чим ближче лінійний коефіцієнт кореляції до ±1, тим тісніше зв'язок між $x$ і $y$ і, навпаки, чим ближче коефіцієнт кореляції до 0, тим менше $x$ впливає на $y$. Знак (+) при коефіцієнті кореляції означає пряму залежність, знак (–) зворотну.

Для визначення тісноти зв'язку за допомогою лінійного коефіцієнта кореляції немає необхідності в знаходженні рівняння регресії, що характеризує зв'язок між $y$ і $x$ і знаходження $\bar{y}_x$. Досить просто підрахувати потрібні суми та поставити їх у формулу коефіцієнта кореляції.

Розрахуємо величину лінійного коефіцієнта кореляції за даними про залежність випуску продукції $y$ від енергоозброєності праці $x$. Для цього побудуємо таблицю залежності випуску продукції $y$ залежно від енергоозброєності праці (кВт·год.) $x$. Для цього побудуємо таблицю, в яку введемо необхідні для розрахунку додаткові графи.

Таблиця 9.27

| $x$ | $y$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 4 | 10 | 40 | 16 | 100 |
| 5 | 12 | 60 | 25 | 144 |
| 6 | 13 | 78 | 36 | 169 |
| 7 | 15 | 105 | 49 | 225 |
| 8 | 17 | 136 | 64 | 289 |
| 30 | 67 | 419 | 190 | 927 |

$$\overline{yx} = \frac{\sum yx}{n} = \frac{419}{5} = 83{,}8;$$

$$\bar{x} = \frac{\sum x}{n} = \frac{30}{5} = 6;$$

$$\bar{y} = \frac{\sum y}{n} = \frac{67}{5} = 13{,}4;$$

$$\sigma_x^2 = \frac{\sum x^2}{n} - \bar{x}^2 = \frac{190}{5} - 6^2 = 38 - 36 = 2; \quad \sigma_x = \sqrt{2};$$

$$\sigma_y^2 = \frac{\sum y^2}{n} - \bar{y}^2 = \frac{927}{5} - 13{,}4^2 = 185{,}4 - 179{,}56 = 5{,}84; \quad \sigma_x = \sqrt{5{,}84};$$

$$r = \frac{83{,}8 - 6 \cdot 13{,}4}{\sqrt{2} \cdot \sqrt{5{,}84}} = \frac{3{,}4}{3{,}42} = 0{,}99 \text{ або } 99\ \%.$$

Величина лінійного коефіцієнта кореляції показує досить тісний прямий зв'язок, що існує між $x$ і $y$.

Кореляційне відношення, будучи універсальним показником оцінки тісноти зв'язку, розраховується для будь-якої форми (як прямолінійного, так і криволінійного) зв'язку між ознаками, лінійний коефіцієнт кореляції — тільки для лінійної. Значення цих коефіцієнтів, обчислених для лінійної форми зв'язку, є однаковими.

Для нашого попереднього прикладу:

$$\overline{xy} = 1902{,}75;\ \ \bar{x}=5{,}4;\ \bar{y}=314{,}4;\ \ \sigma_x=2{,}044;\ \ \sigma_y=104{,}15.$$

Підставимо вихідні дані у формулу та отримаємо:

$$r = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x \cdot \sigma_y} = \frac{1902{,}75 - 5{,}4\cdot 314{,}4}{2{,}044\cdot 104{,}15} = \frac{1902{,}75 - 1697{,}76}{212{,}883} = \frac{204{,}99}{212{,}883} \approx 0{,}963.$$

Лінійний коефіцієнт кореляції показує тісний прямий зв'язок між вартістю основних виробничих засобів ($x$) та виробництвом цукру ($y$). Його значення відповідає кореляційному відношенню, обчисленому раніше.

Розрахуємо коефіцієнт кореляції та за формулою 9.72. Для нашого прикладу необхідні суми є в табл. 9.15.

$$\overline{xy} = 1902{,}75;\quad \sum x = 54;\quad \sum y = 3144;$$

$$\sigma_y^2 = \frac{108472{,}4}{10} = 10847{,}24.$$

Дисперсія ознаки $x$, тобто $\delta_x^2 = \dfrac{\sum(x-\bar{x})^2}{n}$ може бути розрахована також за формулою:

$$\delta_x^2 = \overline{x^2} - (\bar{x})^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{333{,}36}{10} - \left(\frac{54}{10}\right)^2 = 4{,}176.$$

Підставивши у формулу вихідні дані, отримаємо:

$$r = \frac{\dfrac{19027{,}5}{10} - \dfrac{54}{10}\cdot\dfrac{3144}{10}}{\sqrt{4{,}176\cdot 10847{,}24}} = \frac{1902{,}75 - 1697{,}76}{\sqrt{45298{,}074}} = \frac{204{,}99}{212{,}83} = 0{,}963.$$

Відповідь, як і слід очікувати, та сама, що і при розрахунку коефіцієнта кореляції (кореляційного відношення).

В даному випадку при розрахунку коефіцієнта кореляції було використано всього 10 спостережень. Однак така кількість спостережень недостатня для того, щоб зробити висновок про тісноту зв'язку між двома ознаками. Для більш точного судження про ступінь взаємозв'язку між ознаками необхідно провести велику кількість спостережень.

Як говорилося раніше, при великій кількості спостережень вихідні дані доцільно звести в кореляційну таблицю. У такому випадку необхідно кожне значення ознаки знаходити як значення, зважене на його повторення. У цьому випадку розрахунок показників здійснюється за такими формулами:

$$\overline{yx} = \frac{\sum yxn_{xy}}{N}; \qquad \bar{x} = \frac{\sum xn_x}{N}; \qquad \overline{yx} = \frac{\sum yn_y}{N};$$

$$\sigma_x^2 = \frac{\sum x^2 n_x}{N} - \bar{x}^2; \qquad \sigma_y^2 = \frac{\sum y^2 n_y}{N} - \bar{y}^2. \qquad (9.74)$$

Використовуючи дані кореляційної таблиці, можна обчислити коефіцієнт кореляції.

Лінійний коефіцієнт кореляції може бути обчислений і за такою формулою:

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{n\sigma_x \cdot \sigma_y}, \qquad (9.75)$$

де $n$ — число спостережень.

$$r = \frac{2049{,}9}{10 \cdot 2{,}044 \cdot 104{,}15} = 0{,}963,$$

що дорівнює раніше обчисленому коефіцієнту кореляції.

Якщо в рівності $r = \dfrac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\sqrt{\left[\frac{\sum (x-\bar{x})^2}{n}\right] \cdot \left[\frac{\sum (y-\bar{y})^2}{n}\right]}}$ зробити деякі перетворення, то формула лінійного коефіцієнта кореляції набуде наступного вигляду:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{\sum (x)^2}{n}\right)\left(\sum y^2 - \frac{\sum (y)^2}{n}\right)}}. \qquad (9.76)$$

Наведена формула дозволяє дещо спростити розрахунок лінійного коефіцієнта кореляції. У цьому випадку для розрахунку немає потреби обчислювати дисперсії ознак $x$ та $y$. Достатньо звести в квадрат безпосередні значення $x$ і $y$, перемножити значення $x$ на $y$, а потім підрахувати необхідні суми.

Усі дані для розрахунку коефіцієнта кореляції для нашого прикладу за цією формулою ми вже маємо (див. табл. 9.15).

$$\sum xy = 19027{,}5;$$

$$\frac{\sum x \cdot \sum y}{n} = \frac{54 \cdot 3144}{10} = 16977{,}6;$$

$$\frac{\left(\sum x\right)^2}{n} = \frac{(54)^2}{10} = 291{,}6;$$

$$\sum x^2 - \frac{\left(\sum x\right)^2}{n} = 333{,}36 - 291{,}6 = 41{,}76;$$

$$\sum y^2 - \frac{\left(\sum y\right)^2}{n} = 1096946 - \frac{(3144)^2}{10} = 109646 - 988473{,}6 = 108472{,}4;$$

Обчислимо лінійний коефіцієнт кореляції за наведеною вище формулою:

$$r = \frac{19027{,}5 - \frac{54 \cdot 3144}{10}}{\sqrt{\left(333{,}36 - \frac{(54)^2}{10}\right)\left(1096946 - \frac{(3144)^2}{10}\right)}} = \frac{19027{,}5 - 16977{,}6}{\sqrt{41{,}76 \cdot 108472{,}4}} =$$

$$= \frac{2049{,}9}{\sqrt{4529807{,}424}} = \frac{2049{,}9}{2128{,}334} = 0{,}963.$$

Значення коефіцієнта кореляції збігається з раніше обчисленими.

З цієї формули можна отримати іншу, яка також часто застосовується на практиці

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}}. \qquad (9.77)$$

Використовуючи дані розрахункової таблиці, обчислимо лінійний коефіцієнт кореляції:

$$r = \frac{10 \cdot 19027{,}5 - 54 \cdot 3144}{\sqrt{\left[10 \cdot 333{,}36 - 54^2\right] \cdot \left[10 \cdot 1096946 - 3144^2\right]}} =$$

$$= \frac{190275 - 169776}{\sqrt{\left[3333{,}6 - 2916\right] \cdot \left[10969460 - 9884736\right]}} =$$

$$= \frac{20499}{\sqrt{452980742}} = \frac{20499}{21283{,}34} = 0{,}963.$$

Як видно, отримане значення лінійного коефіцієнта кореляції також відповідає раніше обчисленим його значенням.

Коефіцієнт кореляції може бути обчислений і за іншими формулами:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}. \qquad (9.78)$$

Якщо визначено форму кореляційного зв'язку і обчислено коефіцієнт регресії $a_1$, то коефіцієнт кореляції можна визначити за формулою:

$$r = a_1 \frac{\sigma_x}{\sigma_y}. \qquad (9.79)$$

З наведеної формули коефіцієнта кореляції (9.79) можна визначити коефіцієнт регресії, не рахуючи рівняння зв'язку. Коефіцієнт регресії дорівнює ($a_1$):

$$a_1 = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = \frac{19027{,}5 - 16977{,}6}{333{,}36 - 291{,}6} = 49{,}088.$$

Оцінка тісноти зв'язку провадиться за наступною приблизною схемою (табл. 9.28).

Зі збільшенням $x$ має місце позитивна кореляція; при $r$ від'ємному $\bar{y}_x$ знижується. Нарешті, $r = 0$ пряма регресія паралельна осі $x$ і кореляція відсутня.

Отже, показник тісноти зв'язку між середньорічною вартістю основних промислово-виробничих засобів і виробництва цукру становить 0,963 ($\eta$ і $r$). Це свідчить про дуже тісний зв'язок між досліджуваними ознаками.

Таблиця 9.28

| Тіснота зв'язку | Розмір коефіцієнта кореляції за наявності | |
|---|---|---|
| | прямого зв'язку | зворотного зв'язку |
| Слабий | 0,1–0,3 | (–0,1)–(–0,3) |
| Середній | 0,3–0,7 | (–0,3)–(–0,7) |
| Тісний | 0,7–1,0 | (–0,7)–(–1,0) |

Квадрати кореляційного відношення та лінійного коефіцієнта кореляції відповідно дорівнюють 0,927. Це означає, що питома вага середньорічної вартості основних промислово-виробничих засобів у загальній сумі факторів, що визначають виробництво цукру, становить 92,7 %. Інша частина зростання виробництва цукру (різниця між 100 % і 92,7 %) пояснює вплив інших факторів.

Для оцінки *надійності* отриманого *коефіцієнта кореляції* визначають похибку його за такою формулою:

$$\sigma_r = \frac{1 - r^2}{\sqrt{n}}, \tag{9.80}$$

де $n$ — число пар $x_i$ і $y_i$, що спостерігаються.

Після обчислення $\sigma_r$ знаходять співвідношення $\frac{r}{\sigma_r}$. Якщо величина цього відношення перевищує 3 при числі спостережень, більшому, ніж 50, можна вважати, що отриманий коефіцієнт кореляції відображає дійсну залежність між ознаками.

Величини $r - 3\sigma_r$ є гарантійним мінімумом, а величина $r + 3\sigma_r$ — гарантійний максимум коефіцієнта кореляції.

Здійснення вірогідної оцінки лінійного коефіцієнта кореляції необхідна в тому випадку, коли робиться вибірка із генеральної сукупності.

В прикладі отриманий лінійний коефіцієнт кореляції, рівний 0,963, має помилку:

$$\sigma_r = \frac{1 - 0{,}963^2}{\sqrt{10}} \cong 0{,}02.$$

Виникнення помилки пов’язано із невеликою кількістю вимірів.

Гарантійний мінімум дорівнює 0,963–0,02 = 0,943, або 94,3 %. Гарантійний максимум становить 0,963+0,02=0,983, або 98,3%. Це означає, що за умов даної кількості спостережень слід очікувати вплив основних виробничих фондів на випуск цукру не менше, ніж на 94,3 %.

Оскільки кількість спостережень тут зовсім невелика, то знаходити відношення немає сенсу.

При невеликій кількості спостережень знаходиться помилка коефіцієнта регресії за формулою:

$$\sigma_{yfx} = t\frac{\sigma_y}{\sigma_x} \cdot \frac{1 - r^2}{\sqrt{n}}, \tag{9.81}$$

де зберігаються раніше прийняті позначення.

Лінійний коефіцієнт кореляції ($r$) є мірою оцінки тісноти зв’язку і в тому випадку, коли рівняння регресії близьке до лінійної. Однак у деяких випадках виникає необхідність досліджувати дані з точки зору лінійності регресії. Необхідною і достатньою умовою лінійності є вираз $(R^2 - r^2)$ відрізнявся від нуля на величину більшу, ніж величина випадкових коливань. Загальноприйнятим критерієм є порівняння цієї різниці з її ймовірною помилкою.

Якщо ($R^2 - r^2$) мало в порівнянні з $r$ або якщо $R$ і $r$ обидва малі, то можна отримати критерій, що легко обчислюється:

$$\frac{\sqrt{n}}{0{,}6745} \cdot \frac{1}{2}\sqrt{R^2 - r^2} < 2{,}5,$$

або

$$n(R^2 - r^2) < 11{,}37.$$

Якщо рівність виконується, зв'язок можна вважати лінійною.

Слід мати на увазі, що якщо показники кореляції (коефіцієнт кореляції або кореляційне відношення) обчислені за малою кількістю пар значень ознак, то необхідно оцінювати з точки зору їх суттєвості (невипадковості). Для оцінки наявності кореляції в нормально розподіленій сукупності застосовується $t$-критерій Стьюдента:

$$t = \frac{r\sqrt{n-2}}{1-r^2}. \qquad (9.82)$$

де $n$ — число спостережень;

$n-2$ — число ступенів свободи при заданому рівні значимості.

Обчислене за цими формулами значення $t$ порівнюється з табличним (критичним) значенням. Якщо розраховане значення $t$ більше критичного, коефіцієнт кореляції визнається значущим (відхиляється гіпотеза про його випадковість).

У разі $n$=10, відшукуємо по таблицям (див. робота Ф. Мілс, 1958, с. 776). Для економічних розрахунків вважається достатнім гарантійний рівень 5 %. Для ймовірності 0,95 ≥1,96.

У прикладі з вартістю основних засобів та виробництвом цукру 10 підприємств:

$$t = \frac{0{,}963\sqrt{10-2}}{1-0{,}963^2} = \frac{0{,}963\sqrt{8}}{1-0{,}9274} = \frac{2{,}7238}{0{,}0726} \approx 37{,}5.$$

Табличне значення $t$ при 8 степенях свободи та значущості 5 % дорівнює 2,307. Отже, розрахована величина $t$ значно перевищує табличну, тому коефіцієнт кореляції невипадковий.

Цей спосіб застосовується і для оцінки значущості кореляційного відносини.

Істотність показника тісноти зв'язку можна надзвичайно просто визначити за спеціальною таблицею, в якій дано випадкові величини коефіцієнтів кореляції для різних рівнів значимості та різного числа ступенів свободи.

Знаючи коефіцієнт кореляції, можна оцінювати значущість коефіцієнта регресії $a_1$. Ця оцінка здійснюється також за допомогою $t$-критерія Стьюдента:

$$t_{a_1} = \frac{a_1 \sigma_x \sqrt{n-2}}{\sigma_y \sqrt{1-r^2}}. \tag{9.83}$$

В нашому прикладі ($a_1$=49,088; $\sigma_x$=2,044; $\sigma_y = 104{,}15$):

$$t_{a_1} = \frac{49{,}088 \cdot 2{,}044\sqrt{8}}{104{,}15\sqrt{1-0{,}963^2}} \approx \frac{283{,}79}{28{,}07} \approx 10{,}1.$$

Порівнявши це значення з табличним (2,307), приходимо до висновку, що $a_1$ не випадково.

Вище зазначалося, завдяки кореляційного відношення ($\eta$) вимірюється тіснота зв'язку будь-якого типу, а за допомогою лінійного коефіцієнта кореляції ($r$) — лише за лінійної (прямолінейною). При прямолінійній залежності між факторним і результативним ознаками значення $\eta$ і $r$ збігаються. Якщо значення $\eta$ і $r$ не збігаються, це свідчить про те, що залежність між факторною і результативною ознакою не прямолінійна, а криволінійна.

Кореляційне відношення та лінійний коефіцієнт кореляції використовуються в статистиці для вимірювання ступеня тісноти зв'язку між ознаками, що має велике практичне значення.

Кореляційні зв'язки, що досліджуються статистикою, як відомо, характеризуються тим, що зміна (варіювання) тієї чи іншої ознаки відбувається під впливом цілого комплексу факторів. Серед цих факторів одні мають основне значення для сукупності, інші — другорядне, хоча в окремих випадках вони мають сильний вплив на зміну ознаки окремих одиниць сукупності. Встановивши, які саме фактори здійснюють найбільший вплив на розвиток даного явища, можна потім розробити і впровадити практичні заходи щодо забезпечення оптимальних умов його розвитку, посилити вплив позитивних факторів і усунути або зменшити вплив причин, гальмують розвиток.

Застосування лінійного коефіцієнта кореляції для вимірювання тісноти зв'язку обмежено лінійною формою зв'язку між ознаками $x$ та

$y$ ($\overline{y}_x = a_0 + a_1 x$). У разі нелінійної залежності між явищами коефіцієнт кореляції як міра тісноти зв'язку втрачає своє значення. Для вимірювання тісноти зв'язку застосовують так зване кореляційне відношення, відоме під назвою «індекс кореляції».

Кореляційне відношення є більш досконалою мірою оцінки тісноти стохастичного зв'язку між випадковими змінними, ніж коефіцієнт кореляції. Лінійний коефіцієнт кореляції є цілком надійною мірою оцінки тісноти зв'язку між результативною ознакою і факторами-аргументами, що впливають на нього, лише тоді, коли він чисельно дорівнює кореляційному відношенню. Це відбувається лише за прямолінійної залежності між ознаками. Якщо тіснота зв'язку між ознаками при криволінійній залежності визначається за допомогою коефіцієнта кореляції, то в цьому випадку ми більш-менш недооцінюємо тісноту зв'язку.

## 9.12. Багатофакторний кореляційно-регресійний аналіз

**Множинна кореляція**. Визначення кореляційних зв'язків із використанням лише одного фактора припускають дуже високий ступінь узагальнення. Вони неявно передбачається, що варіація результативного ознаки залежить від варіації лише одного фактора.

Суспільні явища характеризуються складним комплексом взаємозв'язків, мають складну ієрархічну структуру та розвиваються під впливом не одного, а багатьох факторів. Тому в практиці кореляційного аналізу виникає необхідність побудови такої виробничої функції, в якій могла бути врахована залежність результативної ознаки від безлічі незалежних між собою факторів. З огляду на це статистки зіштовхуються з необхідністю вивчення як парної, а й множинної кореляцією. Під *множинною кореляцією* розуміється дослідження залежності результативної ознаки від кількох факторних ознак.

В загальному вигляді математична задача формулюється наступним чином. Необхідно знайти рівняння регресії, яке найкращим чином відображає зв'язок факторних ознак з результативними, тобто знайти функцію

$$y = f(x_1, x_2, \ldots, x_n). \quad (9.84)$$

Методологія множинної кореляції грунтується на загальних принципах, які коротко викладені стосовно парної кореляції. Однак при множинній кореляції значно зростає складність математичного апарату аналізу. Зупинимося тих проблемах, які зустрічаються щодо множинних кореляційних зв'язків. Слід зазначити, що з введення у функцію кількох змінних чинників з'являється додаткова передумова незалежності їх один від одної. Надалі будуть розглянуті математичні наслідки запровадження такої передумови.

Складання рівняння множинної кореляції щодо будь-якого явища включає такі основні етапи: 1) вибір виду рівняння; 2) відбір факторів-аргументів; 3) визначення числа спостережень, який буде необхідний отримання несмещенных оцінок.

1. Подібно до парних множинні кореляційні зв'язки можуть зображуватися різними функціями: функцією лінійного типу, степеневою функцією, параболічною функцією і т. п.; б) обрана функція повинна мати наскільки можна простий вид; в) кількість факторів має бути обмеженою. Багатофакторна регресійна модель може бути представлена функціями лінійного типу, параболічними або степеневими, що зводяться за допомогою логарифмування до лінійних. Стосовно цих функцій розробляється методика кореляційного аналізу.

Використання багатофакторної регресійної моделі дозволяє знайти з певної міри точності теоретичне (певною мірою очищене від впливу випадкових причин) значення результативної ознаки за будь-яких можливих поєднань рівнів факторів.

2. Головна складність при дослідженні множинних кореляційних зв'язків полягає у відборі факторів та складанні рівняння множинної регресії. На основі теоретичного аналізу статистик встановлює логічні зв'язки та визначає можливе коло взаємопов'язаних факторів. У це коло факторів включаються як найбільш суттєві та важливі фактори, що визначають той чи інший результат, так і несуттєві, другорядні ознаки. Забігаючи вперед, відзначимо, що до рівняння множинної регресії мають бути включені всі основні фактори, а другорядні фактори можуть бути використані при подальших дослідженнях. Введення в модель великої кількості факторів ускладнює розв'язання задачі. У той самий час їх непродумане виключення призводить до того, що отримане рівняння регресії не буде достатньо точно відтворювати залежності між явищами.

Ознаки, що включаються до рівняння множинної кореляції, повинні мати певні властивості: суттєвість впливу на результативний показник та наявність повної або відносної незалежності факторних ознак.

З багатьох зв'язків факторних ознак виключаються, по-перше, такі, які в процесі попереднього аналізу були віднесені до малозначущих, тобто до несуттєвих. По-друге, потрібно виключити дублюючі фактори, тобто ті, які перебувають з досліджуваною залежністю у функціональному або близькому до функціонального зв'язку. Це призводить до спотворення впливу інших факторів, що входять до рівняння. Наприклад, для відображення втрат робочого часу достатньо включити в модель або втрати в людино-днях, або в людино-годинах. Не слід включати до рівняння регресії комплексні, сукупні чинники та його складові. Скажімо, якщо в модель включена вся заробітна плата працівників, то немає потреби включати до неї окремі її складові. Дублювання факторів може виникати у разі, коли чинники виявляються взятими з різних причинно-наслідкових зв'язків. Так, якщо при вивченні трудомісткості промислової продукції як фактор беруться витрати праці робітників підприємств, то дублюючим виявиться фактор витрат праці допоміжних робітників, оскільки він знайде відображення у трудомісткості через витрати праці робітників. По-третє, виключаються чинники, щодо яких немає повних вихідних даних.

Практика показує, що після виключення малозначущих, дублюючих факторних ознак, а також ознак, за якими немає достатньої впевненості у достовірності вихідних даних, чисельність ознак залишається все ще громіздкою та потребує скорочення. З цією метою встановлюють, наскільки істотно фактори, що залишилися, впливають на результативний показник і чи немає серед них тих факторів, які знаходяться у функціональному або близькому до функціонального зв'язку один з одним. Для цього обчислюють та аналізують коефіцієнти регресії, парні коефіцієнти кореляції, їх критерії надійності.

Щоб виключити фактори-аргументи, що знаходяться між собою у функціональному або близькому до функціонального зв'язку, можна скласти матрицю парних коефіцієнтів кореляції. Матриця є таблицею, на перетині рядків і стовпців якої знаходяться коефіцієнти кореляції між відповідними значеннями. У цій матриці показується тіснота зв'язку кожного фактора з результативним показником та з кожним з інших факторів (див. стор. 450).

Звернемо увагу, що це елементи діагоналі матриці ($r_{00}, r_{11}, \ldots, r_{nn}$) рівні одиниці, оскільки зв'язок тотожних ознак завжди функціональна, а $r_{01} = r_{10}, r_{02} = r_{20}$ і т. п., оскільки елементи симетричної матриці ($r_{ij} = r_{ji}$).

Для вивчення тісноти взаємозв'язку кожного фактора з результативною ознакою обчислюють та порівнюють коефіцієнти кореляції. Спочатку на підставі значень коефіцієнтів кореляції виявляють несуттєві та малозначущі фактори. Фактори, зв'язок яких з результативною ознакою виявиться несуттєвим ($r_{ij}$<0,2), можуть бути виключені з подальшого аналізу.

Потім по матриці парних коефіцієнтів кореляції виявляють фактори, які знаходяться між собою у функціональному або близькому до функціонального зв'язку. Наявність великих за модулем значень коефіцієнт парної кореляції ($\left| r_{ij} \right|$>0,7–0,8) свідчить про те, що при подальшому дослідженні один із факторів, що знаходиться в такому тісному зв'язку, відкидається. При введенні їх у рівняння регресії спостерігається явище *колінеарності* — у разі двох факторів та мультиколінеарності — якщо факторів більше. *Мультиколінеарність* — лінійна залежність або кореляція між аргументами множинної кореляції. Якщо повна колінеарність (наявність функціонального зв'язку) призводить до невизначеності значень параметрів, то часткова мультиколінеарність (наявність сильної кореляції між факторами) призводить до нестійкості їх оцінок. Мультиколлінеарність часто зустрічається при моделюванні роботи виробничих підприємств, де у великій залежності між собою виявляються такі показники, як основні та оборотні фонди, персонал підприємства тощо. співвідношення між основними та оборотними засобами, чисельністю персоналу. Внаслідок чого зміна одного фактора призводить, як правило, до пропорційного зростання іншого (інших) фактора.

Таблиця 9.29

**Матриця парних коефіцієнтів кореляції**

| | $y$ | $x_1$ | $x_2$ | . . . | $x_j$ | . . . | $x_n$ |
|---|---|---|---|---|---|---|---|
| $y$ | 1 | $r_{01}$ | $r_{02}$ | . . . | $r_{0j}$ | . . . | $r_{0n}$ |
| $x_1$ | $r_{10}$ | 1 | $r_{12}$ | . . . | $r_{1j}$ | . . . | $r_{1n}$ |
| $x_2$ | $r_{20}$ | $r_{21}$ | 1 | . . . | $r_{2j}$ | . . . | $r_n$ |
| . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . |
| $x_i$ | $r_{i0}$ | $r_{i1}$ | $r_{i2}$ | . . . | $r_{ij}$ | . . . | $r_{in}$ |
| $x_n$ | $r_{n0}$ | $r_{n1}$ | $r_{n2}$ | . . . | $r_{nj}$ | . . . | 1 |

За наявності мультиколлінеарності із системи взаємопов’язаних ознак необхідно виключити ті з колінеарних ознак, вплив яких на результативний результат менший. Виключення дублюючих мультиколлінеарних ознак — важливий крок до виконання вимог, що пред’являються до множинної кореляції (звільнення від факторів-аргументів, що відображають одну і ту ж особливість або властивість аналізованого процесу; які знаходяться між собою у функціональному або близькому до функціонального зв’язку; наявність повного або відносного зв’язку; наявність повного або відносного) незалежності факторних ознак).

У зв’язку з використанням кореляційного методу в дослідженні, що вимагає нормального розподілу досліджуваних ознак, вихідні дані аналізуються також з точки зору підпорядкування закону нормального розподілу. Для цього будуються ряди розподілу та розраховуються їх основні характеристики: середня арифметична, медіана, мода, варіаційний розмах, середньоквадратичне відхилення, дисперсія, коефіцієнт варіації. Ці показники, що характеризують коливання ознак, дозволяють більш кваліфіковано здійснити як сам змістовний аналіз, так і все кореляційне дослідження в цілому.

Криві розподіли, побудовані за фактичними даними, порівнюють із кривими нормального розподілу. Для побудови кривих нормального розподілу слід використати формулу:

$$f(t)=\frac{Nk}{\sigma}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}, \tag{9.85}$$

де $N$ — сума частот розподілу;

$k$ — величина інтервалу дроблення емпіричного ряду розподілу;

$\sigma$ — середнє квадратичне відхилення ряду;

$t$ — нормоване відхилення, яке рівне $t=\frac{x-\bar{x}}{\sigma}$;

$e, \pi$ — постійні.

Крім того, при включенні факторів необхідно зважати на те, що рішення рівняння множинної регресії вимагає великих витрат праці; так, при переході від трьох до чотирьох факторів витрати праці трудомісткість розрахунків збільшується вдесятеро.

3. При вивченні багатьох зв’язків повинні виконуватися обмеження, що стоять перед кореляційним аналізом, зокрема, дотримання однорідності, достовірності, порівнянності та повноти вихідних даних. При відборі вихідних даних необхідно враховувати такі особли-

вості. З одного боку, надійність рівняння множинної регресії значною мірою залежить від кількості вихідних даних. Адже сутність кореляційних розрахунків полягає в усередненні впливу як окремих чинників на результативний показник, і загального впливу всіх інших неврахованих причин. А середні теми надійніші, чим за більшим обсягом даних вони розраховуються. Але з іншого боку, виникає небезпека у відборі та включення до регресійної моделі факторів без змістовного аналізу. Насправді часто трапляються випадки, коли проведення множинного регресійного аналізу складає основі малої вибірки. Якщо число аналізованих факторів велике, це призводить до того, що сильно збільшуються довірчі інтервали коефіцієнтів регресії і зменшується їх достовірність. Практикою встановлено грубий критерій визначення масиву вихідних даних: чисельність сукупності має перевищувати кількість чинників 6–7 раз.

Вибір форми рівняння починається з обгрунтування кількох видів рівнянь множинної регресії. Після того, як обрано форму рівняння множинної регресії, настає наступний етап: обчислення параметрів.

Розглянемо основні етапи побудови рівняння множинної регресії за динамічними рядами.

**Лінійна множинна кореляція**. На практиці найбільшого поширення набули лінійні форми зв'язку, оскільки такі моделі простіші, легко інтерпретуються та вимагають меншого обсягу обчислень; нерідко криволінійну, особливо слабку, залежність можна порівняно точно описати як прямолінійної залежності, при цьому крива лінія може бути замінена ламаною.

Застосування лінійного рівняння множинної кореляції передбачає дотримання певних вимог: а) аналізований показник може набувати будь-яких значень; б) він змінюється пропорційно до зміни показників, включених до рівняння регресії, на будь-якому інтервалі їх зміни.

Рівняння кореляційної залежності п змінних представлено функцією лінійного типу:

$$\bar{y} = a + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n, \tag{9.86}$$

де $\bar{y}$ — середнє значення, що відповідає заданим значенням аргументів $x_1, x_2, \ldots, x_n$;

$a_1, a_2, \ldots, a_n$ — невідомі параметри.

Лінійне рівняння використовується в тих випадках, коли результативний показник являє собою алгебраїчну суму декількох факторних показників, кожен з яких здійснює самостійний вплив на результат; сама модель називається адитивною.

Кожен із коефіцієнтів рівняння регресії показує, наскільки збільшується в середньому величина результативної ознаки зі збільшенням відповідної факторної ознаки на одиницю при постійному значенні решти всіх факторів, включених у модель.

Визначення параметрів рівняння множинної регресії можуть бути знайдені різними способами: методом найменших квадратів, точковим методом, графічним методом, методом розбиття сукупності досліджуваних об'єктів на групи за кількістю відшукуваних параметрів і подання потім залежності між середньогруповими показниками як функціональної і т.п. лінійного рівняння кореляційної залежності проводиться у спосіб найменших квадратів. Параметри множинної регресії зазвичай визначаються за алгоритмом Гаусса, але можуть бути обчислені методом Чебишева-Дулітля, в основі якого лежить спосіб найменших квадратів. Цей метод найкраще відповідає ідеї усереднення як одиничного впливу врахованих чинників, і загального впливу неврахованих.

Нехай $(y_i, x_{h1}, x_{h2}, \ldots, x_{Nn})$ — результат $i$-го спостереження, $i$=1, 2, . . . , $n$, де $n$ — обсяг вибірки. Вихідний статистичний матеріал може бути представлений наступною матрицею:

$$A_1 = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \ldots x_{14} & \ldots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \ldots x_{24} & \ldots & x_{2n} \\ . & . & . & . & . & . \\ x_{h1} & x_{h2} & x_{h3} & \ldots x_{h4} & \ldots & x_{hn} \\ . & . & . & . & . & . \\ x_{N1} & x_{N2} & x_{hN3} & \ldots x_{N4} & \ldots & x_{Nn} \end{pmatrix}. \qquad (9.87)$$

В процесі рішення знаходяться параметри рівняння $a_i$. З цією метою мінімізується сума квадратів відхилень спостережених значень результативної ознаки $y$ від модельного значення $\overline{y}$, розрахованого за рівнянням регресії $\overline{y} = a_0 + a_1 x_1 + a_2 x_2 + ... + a_n x_n$ для заданих значень аргументу $x_1, x_2, \ldots, x_n$.

При методі найменших квадратів умова, що оптимізує, конкретизується в наступну функцію:

$$S = \sum_{h=1}^{N} (y_i - \overline{y}_{1,2,3,\ldots n})^2 = \min, \qquad (9.88)$$

де $S$ — сума квадратів відхилень значень результативної ознаки, обчислених за рівнянням регресії, від фактичних;

$y_i$ — фактичні значення результативної ознаки;
$\bar{y}_{1,2,3,\ldots,n}$ — розрахункові значення цієї ознаки.

Якщо підставити замість $\bar{y}_{1,2,3,\ldots,n}$ його значення, що випливає з лінійного рівняння кореляційної залежності, то функція набуде наступного вигляду:

$$S=\sum_{h=1}^{N}(y_i-a_0-a_1x_1-a_2x_2-\ldots-a_nx_n)^2=\min. \tag{9.89}$$

Якщо сума квадратів відхилень розрахункових даних від фактичних дорівнює нулю, то обраний чинник повністю пояснює коливання результативного ознаки.

Сума квадратів відхилень досягає мінімуму при таких значеннях $a_1, a_2, a_3, \ldots, a_n$, при яких перетворюється на нуль всі перші часткові похідні:

$$\frac{\partial S}{\partial a_i}=0, \tag{9.90}$$

для всіх $i$=1, 2, 3, . . ., $k$.

Наприклад, похідна по $a_1$

$$\frac{\partial S}{\partial a_1}=\sum_{h=1}^{N}2(y_i-a_1-a_2x_2-\ldots-a_nx_n)(-x_2)=0. \tag{9.91}$$

Розкривши дужки та виносячи за знак суми постійні множники, отримаємо:

$$-2\sum x_2x_1+2a_1\sum x_2+2a_2\sum x_2^2+\ldots+2a_n\sum x_2x_n=0, \tag{9.91}$$

звідки

$$a_0\sum x_2+a_1\sum x_2^2+a_2\sum x_2x_3+\ldots+a_n\sum x_2x_n=\sum y_ix_2. \tag{9.92}$$

В результаті виходить система з так званих нормальних рівнянь з $n$ невідомими (за кількістю параметрів $a_i$):

$$\begin{cases} a_0N+a_1\sum x_1+a_2\sum x_2+\ldots+a_i\sum x_i+\ldots+a_n\sum x_n=\sum y; \\ a_0\sum x_1+a_1\sum x_2^2+a_2\sum x_1x_2+\ldots+a_i\sum x_1x_i+\ldots+a_n\sum x_1x_n=\sum yx_1; \\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ . \\ a_n\sum x_n+a_1\sum x_1x_n+a_2\sum x_3x_n+\ldots+a_i\sum x_ix_n+\ldots+a_n\sum x_n^2=\sum yx_n. \end{cases} \tag{9.93}$$

Найпростіше рівняння кореляційної залежності результативної ознаки ($y$) від двох факторних ознак ($x$, $z$) можна виразити формулою:

$$\bar{y}_{xz}=a_0+a_1x+a_2z. \tag{9.94}$$

Параметри рівняння лінійної регресії можуть бути знайдені різними способами: методом найменших квадратів, точковим методом, графічним методом, методом розбиття сукупності досліджуваних об’єктів на групи за кількістю параметрів, що відшукуються, і подання потім залежності між середньогруповими показниками як функціональної і т. д.

Параметри цього рівняння знаходяться при розв’язанні системи нормальних рівнянь, які отримують для способу найменших квадратів:

$$\begin{cases} a_0 n + a_1 \sum x + a_2 \sum z = \sum y; \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum xz = \sum yx; \\ a_0 \sum z + a_1 \sum xz + a_2 \sum z^2 = \sum zy, \end{cases} \quad (9.95)$$

де $n$ — число одночасних спостережень за трьома ознаками;

$\sum x, \sum y, \sum z$ — суми відповідних значень за цими ознаками.

Розглянемо залежність випуску продукції ($y$) від вартості основних фондів ($x$) та фонду заробітної плати ($z$) (див. графи 1–3 табл. 9.30).

Для розрахунків параметрів рівняння становлять розрахункову таблицю (табл. 9.30).

Таблиця 9.30

| $y$ | $x$ | z | $yx$ | $yz$ | $xz$ | $x^2$ | $z^2$ | $\bar{y}_{xz}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 0,7 | 30,4 | 8 | 21,28 | 5,60 | 243,20 | 924,16 | 64,00 | 0,75 |
| 0,8 | 31,8 | 8,1 | 25,44 | 6,48 | 257,58 | 1 011,24 | 65,61 | 0,87 |
| 0,9 | 30,5 | 8,6 | 27,45 | 7,74 | 262,30 | 930,25 | 73,96 | 0,86 |
| 1 | 31,9 | 8,2 | 31,90 | 8,20 | 261,58 | 1 017,61 | 67,24 | 0,90 |
| 1,1 | 33,2 | 8,8 | 36,52 | 9,68 | 292,16 | 1 102,24 | 77,44 | 1,10 |
| 1,2 | 33,7 | 9,2 | 40,44 | 11,04 | 310,04 | 1 135,69 | 84,64 | 1,21 |
| 5,7 | 191,5 | 50,9 | 183,03 | 48,74 | 1 626,86 | 6 121,19 | 432,89 | 5,69 |

Використовуючи дані таблиці, складемо систему трьох рівнянь:

$$\begin{cases} 6a_0 + 191{,}5a_1 + 50{,}9a_2 = 5{,}7; \\ 191{,}5a_0 + 6121{,}19a_1 + 1626{,}86a_2 = 183{,}03; \\ 50{,}9a_0 + 1626{,}86a_1 + 432{,}89a_2 = 48{,}74. \end{cases}$$

Поділимо всі члени рівнянь на коефіцієнти при $a_0$, отримаємо:

$$\begin{cases} a_0 + 31{,}9167\,a_1 + 8{,}4833\,a_2 = 0{,}95; \\ a_0 + 31{,}9644\,a_1 + 8{,}4954\,a_2 = 0{,}9558; \\ a_0 + 31{,}9619a_1 + 8{,}5047a_2 = 0{,}9576. \end{cases}$$

Віднімаючи з другого рівняння спочатку перше, а потім третє, отримаємо 2 рівняння з двома невідомими:

$$0{,}0477\,a_1 + 0{,}0121a_2 = 0{,}0058;$$
$$0{,}0025\,a_1 - 0{,}0093\,a_2 = -0{,}0018.$$

Ділимо кожен член обох рівнянь на коефіцієнти при $a_1$:

$$a_1 + 0{,}2537\,a_2 = 0{,}1216$$
$$a_1 - 3{,}72a_2 = -0{,}72$$
$$-3{,}4663\,a_2 = -0{,}5984$$

Звідки

$$a_2 = 0{,}1726.$$

Підставимо коефіцієнт при $a_2$ в перше рівняння:

$$a_1 + 0{,}2537 \cdot 0{,}1726 = 0{,}1216$$
$$a_1 = 0{,}0778.$$

У перше рівняння підставимо значення $a_1$ і $a_2$:

$$a_0 + 31{,}9167 \cdot 0{,}0778 + 8{,}4833 \cdot 0{,}1726 = 0{,}95.$$
$$a_1 = -2{,}9973.$$

Рівняння зв’язку, що визначає залежність результативної ознаки ($y$) від двох факторіальних ($x$ і $z$):

$$\overline{y}_{xz} = -2{,}9973 + 0{,}0778x + 0{,}1726z.$$

Обчислимо за цим рівнянням за відповідних $x$ і $z$ теоретичні величини $\overline{y}_{xz}$ та занесемо до графи 9 табл. 9.30. Аналіз показників рівняння множинної регресії дозволяє зробити висновок про рівень впливу двох факторів на показник обсягу реалізованої продукції. Так, зі збільшенням вартості основних виробничих фондів на 1 млн грн вартість продукції збільшується на 779 шт.; зі зростанням витрат на оплату праці на 1 млн грн обсяг реалізованої продукції скорочується на 173 шт. Зазначимо, що суми фактичних даних ($y$) та розрахункових даних ($\overline{y}_{xz}$) практично збігаються (5,7≈5,69), а окремі значення їх незначно відрізняються один від одного.

Цю систему можна вирішити різними способами, у тому числі через коефіцієнти парної кореляції та середні квадратичні відхилення:

$$a_0 = \overline{y} - a_1\overline{x}_1 - a_2\overline{z}; \qquad (9.96)$$

$$a_1 = \frac{r_{yx} - r_{yz} \cdot r_{xz}}{1 - r_{xz}^2} \cdot \frac{\sigma_y}{\sigma_x}; \qquad (9.97)$$

$$a_2 = \frac{r_{yz} - r_{yx} \cdot r_{xz}}{1 - r_{xz}^2} \cdot \frac{\sigma_y}{\sigma_z}. \qquad (9.98)$$

Обчислимо за цими формулами відповідні параметри та складемо рівняння множинного кореляційного зв'язку між випуском продукції ($y$) та двома факторами — основними засобами ($x$) та фондом заробітної плати ($z$). Як видно з формул $a_0, a_1, a_2$, для обчислення параметрів слід визначити коефіцієнт парної кореляції, середні та середні квадратичні відхилення. Усі необхідні вихідні дані та початкові розрахунки наводяться в табл. 9.31.

Розрахуємо спочатку три коефіцієнти парної кореляції з відповідними середніми та середніми квадратичними відхиленнями.

1. Розрахуємо парну кореляцію $r_{xy}$

$$\overline{x} = \frac{\sum x}{n} = \frac{191{,}5}{6} = 31{,}9167; \quad \overline{x^2} = \frac{\sum x^2}{n} = \frac{6121{,}19}{6} = 1020{,}2;$$

$$\overline{y} = \frac{\sum y}{n} = \frac{5{,}7}{6} = 0{,}95; \quad \overline{y^2} = \frac{\sum y^2}{n} = \frac{5{,}59}{6} = 0{,}9317;$$

$$\overline{x}^2 = 31{,}9167^2 = 1018{,}6757; \quad \overline{y}^2 = 0{,}95^2 = 0{,}9025;$$

$$\overline{xy} = \frac{\sum xy}{n} = \frac{183{,}03}{6} = 30{,}505;$$

$$\sigma_x = \sqrt{\overline{x^2} - \overline{x}^2} = \sqrt{1020{,}2 - 10186757} = \sqrt{1{,}5243} \approx 1{,}23463;$$

$$\sigma_y = \sqrt{\overline{y^2} - \overline{y}^2} = \sqrt{0{,}9317 - 0{,}9025} = \sqrt{0{,}0292} \approx 0{,}17088;$$

$$r_{yx} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y} = \frac{30{,}505 - 31{,}9167 \cdot 0{,}95}{1{,}23463 \cdot 0{,}17088} = \frac{0{,}18414}{0{,}21097} = 0{,}873.$$

2. Розрахуємо коефіцієнт парної кореляції $r_{zy}$

$$\overline{z} = \frac{\sum z}{n} = \frac{50{,}9}{6} = 8{,}4833; \quad \overline{z^2} = \frac{\sum z^2}{n} = \frac{432{,}89}{6} = 72{,}148;$$

$$\overline{zy} = \frac{\sum zy}{n} = \frac{48{,}74}{6} \approx 8{,}1233;$$

Таблиця 9.31

| Основні засоби, млн грн $x$ | Фонд заробітної плати, млн грн $z$ | Обсяг виробленої продукції (тис. шт.) $y$ | $x^2$ | $yx$ | $y^2$ | $z^2$ | $yz$ | $xz$ | $\bar{y}_{xz} = -2{,}99824 + 0{,}0681x + 0{,}2092z$ | $y - \bar{y}_{xz}$ | $(y - \bar{y}_{xz})^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 30,4 | 8 | 0,7 | 924,16 | 21,28 | 0,49 | 64 | 5,6 | 243,2 | 0,75 | –0,05 | 0,0025 |
| 31,8 | 8,1 | 0,8 | 1 011,24 | 25,44 | 0,64 | 65,61 | 6,48 | 257,58 | 0,86 | –0,06 | 0,0036 |
| 30,5 | 8,6 | 0,9 | 930,25 | 27,45 | 0,81 | 73,96 | 7,74 | 262,3 | 0,88 | +0,02 | 0,0004 |
| 31,9 | 8,2 | 1 | 1 017,61 | 31,9 | 1 | 67,24 | 8,2 | 261,58 | 0,89 | +0,11 | 0,0121 |
| 33,2 | 8,8 | 1,1 | 1 102,24 | 36,52 | 1,21 | 77,44 | 9,68 | 292,16 | 1,10 | +0,00 | 0 |
| 33,7 | 9,2 | 1,2 | 1 135,69 | 40,44 | 1,44 | 84,64 | 11,04 | 310,04 | 1,22 | –0,02 | 0,0004 |
| $\sum$191,5 | 50,9 | 5,7 | 6 121,19 | 183,03 | 5,59 | 432,89 | 48,74 | 1 626,86 | 5,70 | – | 0,0190 |

$$\sigma_z = \sqrt{\overline{z^2} - \bar{z}^2} = \sqrt{72{,}148 - 8{,}4833^2} = \sqrt{0{,}1816} \approx 0{,}4261;$$

$$r_{yz} = \frac{\overline{zy} - \bar{z} \cdot \bar{y}}{\sigma_z \cdot \sigma_y} = \frac{8{,}1233 - 8{,}4833 \cdot 0{,}95}{0{,}4261 \cdot 0{,}17088} = \frac{0{,}06417}{0{,}07274} = 0{,}881.$$

3. $$\overline{xz} = \frac{\sum xz}{n} = \frac{1626{,}86}{6} = 271{,}143.$$

$$r_{xz} = \frac{\overline{xz} - \bar{x} \cdot \bar{z}}{\sigma_x \cdot \sigma_z} = \frac{271{,}143 - 31{,}917 \cdot 8{,}4833}{1{,}2261 \cdot 0{,}4262} = \frac{0{,}3815}{0{,}5226} = 0{,}730$$ [1].

Звідки

$$a_1 = \frac{r_{yx} - r_{yz} \cdot r_{xz}}{1 - r_{xz}^2} \cdot \frac{\sigma_y}{\sigma_x} = \frac{0{,}873 - 0{,}881 \cdot 0{,}73}{1 - 0{,}73^2} \cdot \frac{0{,}17088}{1{,}23463} = \frac{0{,}03928}{0{,}57670} = 0{,}0681;$$

$$a_2 = \frac{r_{yz} - r_{yx} \cdot r_{xz}}{1 - r_{xz}^2} \cdot \frac{\sigma_y}{\sigma_z} = \frac{0{,}881 - 0{,}873 \cdot 0{,}73}{1 - 0{,}73^2} \cdot \frac{0{,}17088}{0{,}4261} = \frac{0{,}04165}{0{,}19903} \approx 0{,}2092;$$

$$a_0 = \bar{y} - a_1 \bar{x}_1 - a_2 \bar{z} = 0{,}95 - 0{,}0681 \cdot 31{,}9167 - 0{,}2092 \cdot 8{,}4833 =$$

$$= 0{,}95 - 2{,}17353 - 1{,}77471 \approx -2{,}99824.$$

Отже, рівняння множинної кореляції випуску продукції з основними фондами та фондом заробітної плати буде таким:

$$\bar{y}_{xz} = -2{,}99824 + 0{,}0681x + 0{,}2092z.$$

Параметр $a_1$=0,0681 означає, що зі збільшенням основних засобів на 1 млн грн випуск продукції при фіксованому значенні фонду заробітної плати збільшується в середньому на 68 шт. Параметр $a_2$=0,2092 показує, що зі збільшенням фонду заробітної плати на 1 млн грн і при фіксованому значенні основних фондів випуск продукції зростає в середньому на 209 шт. Параметр $a_0$, у якому певною мірою відображається вплив неврахованих у моделі факторів, у множинних рівняннях регресії змістовно не інтерпретується.

Підставляючи значення $x$ і $z$ рівняння множинної регресії, отримуємо вирівняні (теоретичні) значення обсягу виробленої продукції (графа 10). Так, при $x$=30,4 та $z$=8 теоретичний обсяг випуску складе:

---

[1] Ознаки $x$ і $z$ виявилися колінеарними, тому за існуючими правилами один з них повинен бути виключений з подальшого аналізу. Цього тут не робимо, оскільки методика аналізу від величини $r$ тут не залежить.

$$\overline{y}_{xz} = -2{,}99824 + 0{,}0681 \cdot 30{,}4 + 0{,}2092 \cdot 8 = 0{,}75 \text{ і т. д.}$$

У попередньому прикладі ми розглянули порядок знаходження лінійного рівняння кореляційної залежності результативної ознаки від двох факторних. Однак на практиці на результативну ознаку впливає значно більше факторів.

Розглянемо порядок знаходження параметрів рівняння кореляційної залежності між результативною ознакою та трьома факторами. За даними статистичної звітності дослідної станції за 1998–2022 рр. потрібно побудувати регресійну модель зв'язку між урожайністю пшениці на дослідній станції та трьома факторами ($x$, $z$, $v$). Мета моделювання полягала у визначенні та кількісному вимірі резервів зростання врожайності пшениці. На стадії апріорного аналізу виділено такі кореляційні змінні: $y$ — урожайність гречки; $x$ — дози азотних добрив; $z$ — дози калійних мінеральних добрив; $v$ — кількість опадів у період колошения.

Якщо число факторіальних ознак зростає, зростає число членів рівняння зв'язку. Так, для трьох факторіальних ознак лінійне рівняння зв'язку буде записано формулою:

$$\overline{y}_{xz} = a_0 + a_1 x + a_2 z + a_3 v, \tag{9.99}$$

де параметри рівняння $a_0; a_1; a_2; a_3$ $a_0$ знаходяться шляхом вирішення системи чотирьох нормальних рівнянь:

$$\begin{cases} a_0 n + a_1 \sum x + a_2 \sum z + a_3 \sum v = \sum y; \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum xz + a_3 \sum xv = \sum yx; \\ a_0 \sum z + a_1 \sum xz + a_2 \sum z^2 + a_3 \sum zv = \sum yz, \\ a_0 \sum v + a_1 \sum xv + a_2 \sum zv + a_3 \sum v^2 = \sum yv. \end{cases} \tag{9.100}$$

Побудуємо відповідну таблицю, отримаємо у ній необхідні сумарні дані для наведеної системи рівнянь (табл. 9.31).

Статистичні дані, отримані в результаті спостережень, та розрахунки представлені в табл. 9.32, звідки візьмемо необхідні дані для складання системи нормальних рівнянь:

$$\begin{cases} 25 a_0 + 761 a_1 + 1379 a_2 + 4879 a_3 = 233{,}5; \\ 761 a_0 + 34553 a_1 + 47078 a_2 + 131010 a_3 = 8696{,}8; \\ 1379 a_0 + 47078 a_1 + 96165 a_2 + 238816 a_3 = 14973{,}8; \\ 4879 a_0 + 131010 a_1 + 238816 a_2 + 1043091 a_3 = 40843{,}8. \end{cases}$$

Поділимо всі члени рівнянь на коефіцієнти при $a_0$, отримаємо:

Таблиця 9.32

**Урожайність гречихи і метеорологічні фактори**

Розрахунок допоміжних величин

| Роки | $x$ | $z$ | $v$ | $y$ | $xz$ | $xv$ | $zv$ | $x^2$ | $z^2$ | $v^2$ | $y^2$ | $xy$ | $zy$ | $vy$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1998 | 23 | 74 | 212 | 10,6 | 1 702 | 4876 | 15 688 | 529 | 5 476 | 44 944 | 112,36 | 243,8 | 784,4 | 2 247,2 |
| 1999 | 3 | 47 | 269 | 6,7 | 141 | 807 | 12 643 | 9 | 2 209 | 72 361 | 44,89 | 20,1 | 314,9 | 1 802,3 |
| 2000 | 46 | 40 | 164 | 10,6 | 1 840 | 7544 | 6 560 | 2 116 | 1 600 | 26 896 | 112,36 | 487,6 | 424 | 1 738,4 |
| 2001 | 27 | 34 | 223 | 6,1 | 918 | 6021 | 7 582 | 729 | 1 156 | 49 729 | 37,21 | 164,7 | 207,4 | 1 360,3 |
| 2002 | 39 | 56 | 127 | 11,2 | 2 184 | 4953 | 7 112 | 1 521 | 3 136 | 16 129 | 125,44 | 436,8 | 627,2 | 1 422,4 |
| 2003 | 53 | 56 | 146 | 14,8 | 2 968 | 7738 | 8 176 | 2 809 | 3 136 | 21 316 | 219,04 | 784,4 | 828,8 | 2 160,8 |
| 2004 | 3 | 16 | 312 | 2,5 | 48 | 936 | 4 992 | 9 | 256 | 97 344 | 6,25 | 7,5 | 40 | 780 |
| 2005 | 27 | 87 | 176 | 12,2 | 2 349 | 4752 | 15 312 | 729 | 7 569 | 30 976 | 148,84 | 329,4 | 1061,4 | 2 147,2 |
| 2006 | 29 | 52 | 155 | 10,1 | 1 508 | 4495 | 8 060 | 841 | 2 704 | 24 025 | 102,01 | 292,9 | 525,2 | 1 565,5 |
| 2007 | 38 | 50 | 172 | 10,8 | 1 900 | 6536 | 8 600 | 1 444 | 2 500 | 29 584 | 116,64 | 410,4 | 540 | 1 857,6 |
| 2008 | 88 | 94 | 158 | 17,8 | 8 272 | 13904 | 14 852 | 7 744 | 8 836 | 24 964 | 316,84 | 1 566,4 | 1673,2 | 2 812,4 |
| 2009 | 16 | 91 | 142 | 8,3 | 1 456 | 2272 | 12 922 | 256 | 8 281 | 20 164 | 68,89 | 132,8 | 755,3 | 1 178,6 |
| 2010 | 11 | 18 | 212 | 5,5 | 198 | 2332 | 3 816 | 121 | 324 | 44 944 | 30,25 | 60,5 | 99 | 1 166 |
| 2011 | 16 | 87 | 125 | 11,9 | 1 392 | 2000 | 10 875 | 256 | 7 569 | 15 625 | 141,61 | 190,4 | 1 035,3 | 1 487,5 |
| 2012 | 19 | 100 | 150 | 10,3 | 1 900 | 2850 | 15 000 | 361 | 10 000 | 22 500 | 106,09 | 195,7 | 1030 | 1 545 |
| 2013 | 34 | 57 | 215 | 9,4 | 1 938 | 7310 | 12 255 | 1 156 | 3 249 | 46 225 | 88,36 | 319,6 | 535,8 | 2 021 |
| 2014 | 3 | 3 | 364 | 1,4 | 9 | 1092 | 1 092 | 9 | 9 | 132 496 | 1,96 | 4,2 | 4,2 | 509,6 |
| 2015 | 21 | 47 | 172 | 9,7 | 987 | 3612 | 8 084 | 441 | 2 209 | 29 584 | 94,09 | 203,7 | 455,9 | 1 668,4 |
| 2016 | 42 | 53 | 179 | 8,8 | 2 226 | 7518 | 9 487 | 1 764 | 2 809 | 32 041 | 77,44 | 369,6 | 466,4 | 1 575,2 |
| 2017 | 10 | 16 | 310 | 3,1 | 160 | 3100 | 4 960 | 100 | 256 | 96 100 | 9,61 | 31 | 49,6 | 961 |
| 2018 | 14 | 106 | 168 | 11,5 | 1 484 | 2352 | 17 808 | 196 | 11 236 | 28 224 | 132,25 | 161 | 1219 | 1 932 |
| 2019 | 68 | 72 | 180 | 13,3 | 4 896 | 12240 | 12 960 | 4 624 | 5 184 | 32 400 | 176,89 | 904,4 | 957,6 | 2 394 |
| 2020 | 22 | 16 | 230 | 5,6 | 352 | 5060 | 3 680 | 484 | 256 | 52 900 | 31,36 | 123,2 | 89,6 | 1 288 |
| 2021 | 68 | 69 | 136 | 14,2 | 4 692 | 9248 | 9 384 | 4 624 | 4 761 | 18 496 | 201,64 | 965,6 | 979,8 | 1 931,2 |
| 2022 | 41 | 38 | 182 | 7,1 | 1 558 | 7462 | 6 916 | 1 681 | 1 444 | 33 124 | 50,41 | 291,1 | 269,8 | 1 292,2 |
| | 761 | 1 379 | 4 879 | 233,5 | 47 078 | 131 010 | 238 816 | 34 553 | 96 165 | 1 043 091 | 2 552,73 | 8 696,8 | 14 973,8 | 40 843,8 |
| Символ | $\sum x$ | $\sum z$ | $\sum v$ | $\sum y$ | $\sum xz$ | $\sum xv$ | $\sum zv$ | $\sum x^2$ | $\sum z^2$ | $\sum v^2$ | $\sum y^2$ | $\sum xy$ | $\sum zy$ | $\sum vy$ |

$$\begin{cases} a_0 + 30{,}44a_1 + 55{,}16a_2 + 195{,}16a_3 = 9{,}34; \\ a_0 + 45{,}4a_1 + 59{,}52a_2 + 172{,}16a_3 = 11{,}43; \\ a_0 + 34{,}14a_1 + 69{,}74a_2 + 173{,}18a_3 = 10{,}86; \\ a_0 + 26{,}85a_1 + 48{,}95a_2 + 213{,}79a_3 = 8{,}37. \end{cases}$$

Віднімемо від другого перше, третє та четверте, отримаємо:

$$\begin{aligned} &\text{II–I} \quad 14{,}96a_1 + 4{,}36a_2 - 23a_3 = 2{,}09; \\ &\text{II–III} \quad 11{,}26a_1 - 10{,}22a_2 - 1{,}02a_3 = 0{,}57; \\ &\text{II–IV} \quad 18{,}55a_1 + 10{,}57a_2 - 41{,}63a_3 = 3{,}06; \end{aligned}$$

Поділимо всі члени рівнянь на коефіцієнти при $a_1$, отримаємо:

$$a_1 + 0{,}2914a_2 - 1{,}5374a_3 = 0{,}1397;$$
$$a_1 - 0{,}9076a_2 - 0{,}0906a_3 = 0{,}0507;$$
$$a_1 + 0{,}5698a_2 - 2{,}2442a_3 = 0{,}1650;$$

Віднімемо від першого рівняння друге і від третього друге, отримаємо:

$$\begin{aligned} &\text{I–II} \quad 1{,}199a_2 - 1{,}4468a_3 = 0{,}089; \\ &\text{III–II} \quad 1{,}4774a_2 - 2{,}1536a_3 = 0{,}1143; \end{aligned}$$

Поділимо всі члени рівнянь на коефіцієнти при $a_2$, отримаємо:

$$a_2 - 1{,}2067a_3 = 0{,}0742$$
$$a_2 - 1{,}4577a_3 = 0{,}0774$$
$$0{,}251a_3 = -0{,}0032;$$
$$a_3 = -0{,}0128;$$
$$a_2 = 0{,}0742 + 1{,}2067 \cdot (-0{,}0128) = 0{,}0742 - 0{,}0155;$$
$$a_2 = 0{,}0587;$$
$$a_1 + 0{,}2914 \cdot 0{,}0587 - 1{,}5374 \cdot (-0{,}0128) = 0{,}1397;$$
$$a_1 + 0{,}0171 + 0{,}0197 = 0{,}1397;$$
$$a_1 = 0{,}1029;$$
$$a_0 + 30{,}44 \cdot 0{,}1029 + 55{,}16 \cdot 0{,}0587 + 195{,}16 \cdot (-0{,}0128) = 9{,}34;$$
$$a_0 + 3{,}1323 + 3{,}2379 - 2{,}4980 = 9{,}34;$$
$$a_0 = 5{,}4678;$$

Отже, кореляційне рівняння буде:

$$\overline{y}_{xz} = 5{,}4678 + 0{,}1029x + 0{,}0587z - 0{,}0128v.$$

Мірою суттєвості впливу тієї чи іншої факторіальної ознаки на результативну є показники тісноти зв'язку.

**Криволінійна множинна кореляція**. При множинні кореляційної залежності значно рідше для визначення рівняння регресії можуть бути використані інші типи функцій. Так, рівняння кореляційної залежності можуть бути представлені степеневою функцією наступного виду:

$$y = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot \ldots \cdot x_n^{a_n}. \quad (9.101)$$

Ця модель застосовується тоді, коли результативний показник є добутком кількох факторів; сама модель називається мультиплікативною.

Наприклад, якщо форма залежності між ознакою та кожним із двох факторів виражена параболою другого порядку, то рівняння регресії можна виразити у вигляді наступної формули:

$$\overline{y} = a_0 + a_1 x + a_2 x^2 + a_3 z + a_4 z^2. \quad (9.102)$$

Рівняння регресії може іноді складатися з поєднання рівняння прямої зі степеневої рівнянням або лінійних рівнянь множинної залежності.

**Множинний коефіцієнт кореляції**. Залежність результативної ознаки від рядів ознак-факторів може бути охарактеризована в цілому. Для оцінки їх тісноти застосовуються лінійний коефіцієнт множинної кореляції та теоретичне множинне кореляційне відношення. Для оцінки тісноти звязку, що встановлюється між результативною і двома або більше факторними ознаками, є *сукупний коефіцієнт множинної кореляції r*. Він служить основним показником лінійного кореляційного звязку. Величина сукупного коефіцієнту множинної кореляції змінюється від 0 до 1. Чим менше спостерігаєме значення показника відхиляється від лінії множинної регресії, тим кореляційний зв'язок є більш інтенсивним, а отже, величина $r$ ближча до одиниці.

У випадку лінійного зв'язку двох факторів ($x$ і $z$) на результативний показник ($y$) *сукупний коефіцієнт множинної кореляції* може бути обчислений за формулою:

$$r_{yxz} = \sqrt{\frac{r_{yx}^2 + r_{yz}^2 - 2r_{yx} \cdot r_{yz} \cdot r_{xz}}{1 - r_{xz}^2}}, \quad (9.103)$$

де $r$ — парні коефіцієнти кореляції, а підрядкові знаки показують, якими ознаками вони обчислюються.

Лінійні коефіцієнти кореляції між парами взаємопов'язаних значень можуть бути знайдені за формулами:

$$r_{xz} = \frac{\sum(x-\bar{x})(z-\bar{z})}{n\sigma_x\sigma_z}; \tag{9.104}$$

$$r_{yx} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y}; \tag{9.105}$$

$$r_{yz} = \frac{\sum(y-\bar{y})(z-\bar{z})}{n\sigma_y\sigma_z}. \tag{9.106}$$

Обчислимо лінійні коефіцієнти кореляції між парами ознак та сукупний коефіцієнт кореляції за даними про залежність між випуском продукції *y*, основними виробничими засобами *x* та заробітною платою *z* (табл. 9.33).

$$\sigma_x^2 = 1{,}5247; \quad \sigma_z^2 = 0{,}1814; \quad \sigma_y^2 = 0{,}0292;$$

$$\sigma_x = 1{,}235; \quad \sigma_z = 0{,}426; \quad \sigma_y = 0{,}171;$$

$$\bar{x} = 31{,}9167; \quad \bar{z} = 8{,}4833; \quad \bar{y} = 0{,}95.$$

Таблиця 9.33

| $x$ | $z$ | $y$ | $x-\bar{x}$ | $z-\bar{z}$ | $y-\bar{y}$ | $(x-\bar{x})^2$ | $(z-\bar{z})^2$ | $(y-\bar{y})^2$ | $(x-\bar{x})\cdot(z-\bar{z})$ | $(x-\bar{x})\cdot(y-\bar{y})$ | $(z-\bar{z})\cdot(y-\bar{y})$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 30,4 | 8 | 0,7 | –1,5167 | –0,4833 | –0,25 | 2,3004 | 0,2336 | 0,0625 | 0,7330 | 0,3792 | 0,1208 |
| 31,8 | 8,1 | 0,8 | –0,1167 | –0,3833 | –0,15 | 0,0136 | 0,1469 | 0,0225 | 0,0447 | 0,0175 | 0,0575 |
| 30,5 | 8,6 | 0,9 | –1,4167 | 0,1167 | –0,05 | 2,0070 | 0,0136 | 0,0025 | -0,1653 | 0,0708 | –0,0058 |
| 31,9 | 8,2 | 1 | –0,0167 | –0,2833 | 0,05 | 0,0003 | 0,0803 | 0,0025 | 0,0047 | –0,0008 | –0,0142 |
| 33,2 | 8,8 | 1,1 | 1,2833 | 0,3167 | 0,15 | 1,6469 | 0,1003 | 0,0225 | 0,4064 | 0,1925 | 0,0475 |
| 33,7 | 9,2 | 1,2 | 1,7833 | 0,7167 | 0,25 | 3,1802 | 0,5137 | 0,0625 | 1,2781 | 0,4458 | 0,1792 |
| 191,5 | 50,9 | 5,7 | – | – | – | 9,1483 | 1,0883 | 0,1750 | 2,3017 | 1,1050 | 0,3850 |

З н а й д е м о коефіцієнти кореляції між *x* і *z*, *y* і *x*, *y* та *z* за наведеними вище формулами:

$$r_{xz} = \frac{2{,}3017}{6\cdot1{,}235\cdot0{,}426} = 0{,}729;$$

$$r_{yx} = \frac{1{,}1050}{6\cdot1{,}235\cdot0{,}171} = 0{,}872;$$

$$r_{yz} = \frac{0{,}3850}{6\cdot0{,}171\cdot0{,}426} = 0{,}881.$$

Високі значення парних коефіцієнтів свідчать про сильний вплив (окремо, ізольовано) основних виробничих засобів та фонду заробітної плати на випуск продукції.

Обчислимо тепер сукупний коефіцієнт множинної кореляції:

$$r_{yxz} = \sqrt{\frac{0{,}872^2 + 0{,}881^2 - 2 \cdot 0{,}872 \cdot 0{,}881 \cdot 0{,}729}{1 - 0{,}729^2}} = 0{,}943.$$

Сукупний коефіцієнт кореляції показує ступінь тісноти зв'язку між результативною ознакою та сукупним впливом двох факторів. У цьому прикладі цей зв'язок дуже тісний.

Розрахунок множинного кореляційного відношення (зване і сукупним індексом кореляції) складає основі дисперсій. У загальному вигляді формула для розрахунку сукупного індексу кореляції може бути записана так:

$$R_{1,2,\ldots,n} = \sqrt{1 - \frac{\sigma^2_{1,2,\ldots,n}}{\sigma^2}}, \tag{9.107}$$

де $R_{1,2,\ldots,n}$ — множинне кореляційне відношення, що відображає тісноту зв'язку між першим показником і всіма іншими, починаючи з другого і закінчуючи $n$-м;

$\sigma^2$ — загальна дисперсія фактичних даних результативної ознаки;

$\sigma^2_{1,2,\ldots,n}$ — залишкова дисперсія.

Як зазначалося, кореляційне відношення змінюється в межах $0 \le R \le 1$. Якщо $R = 0$, то при встановленій множинні кореляції даний набір факторів взагалі не пояснює коливання результативної ознаки, при $R = 1$ — зв'язок функціональний. У багатьох випадках при значній кількості змінних розрахунки кореляційного відношення дуже громіздкі.

Якщо коефіцієнт множинної кореляції незначний за величиною, це може говорити про три факти: 1) у рівняння множинної кореляції не включені чинники, що надає значний вплив зміну результативного ознаки; виправити таке положення можна шляхом включення до регресійної моделі деяких додаткових факторів; 2) неправильно обраний тип функції, що характеризує кореляційну залежність між ознаками; 3) наведений теоретичний аналіз виявився неспроможним.

Для двофакторної моделі розрахунок множинного кореляційного відношення здійснюється за формулою:

$$R = \sqrt{1 - \frac{\sigma^2_{(y - \bar{y}_{xz})}}{\sigma^2_y}}, \tag{9.108}$$

Обчислимо $r$ та $R$ за попередніми даними (інформація для розрахунку в табл. 9.31, гр. 10–12).

$$r_{yxz} = \sqrt{\frac{r_{yx}^2 + r_{yz}^2 - 2r_{yx} \cdot r_{yz} \cdot r_{xz}}{1 - r_{xz}^2}} = \sqrt{\frac{0{,}873^2 + 0{,}881^2 - 2 \cdot 0{,}873 \cdot 0{,}881 \cdot 0{,}730}{1 - 0{,}730^2}} =$$

$$= \sqrt{\frac{0{,}4154}{0{,}4671}} = \sqrt{0{,}8893} \approx 0{,}943.$$

$$R_{yxz} = \sqrt{1 - \frac{\sigma^2_{(y - \bar{y}_{xz})}}{\sigma_y^2}} = \sqrt{1 - \frac{\left(\sum y - \bar{y}_{xz}\right)^2 : n}{\sigma_y^2}} =$$

$$= \sqrt{1 - \frac{0{,}00317}{0{,}17088^2}} = \sqrt{0{,}8914} \approx 0{,}944.$$

Значення множинного коефіцієнта кореляції та множинного кореляційного відношення, що дорівнює відповідно 0,943 та 0,944, говорить про тісний зв'язок між розміром основних засобів, фондом заробітної плати та випуском продукції.

Лінійний множинний коефіцієнт кореляції та множинне кореляційне відношення мають той же сенс і змінюються в тих же межах, що й у парній кореляції. Наявність кількісного розбіжності з-поміж них свідчить у тому, що зв'язок між ознаками є криволінійною. За прийнятим вище критерієм розбіжності в нашому прикладі ($R - r \leq 0{,}1$) можна прийняти незначним.

**Коефіцієнт множинної (сукупної) детермінації**. Величина $R^2$ називається *сукупним коефіцієнтом множинної детермінації*. Вона показує, яка частка варіації досліджуваного показника зумовлена впливом факторів, які включені до рівняння множинної регресії. Значення сукупного коефіцієнта детермінації знаходиться у межах від 0 до +1. Чим ближче значення коефіцієнта множинної кореляції до 1, тим більша варіація результативної ознаки визначається впливом відібраних факторів. У нашому прикладі сукупний коефіцієнт множинної детермінації $R^2 = 0{,}891$ (або 89,1 %) показує, що варіація виробництва продукції на 89,1 % обумовлюється двома факторами. Отже, вибрані фактори суттєво впливають на обсяг виробництва продукції. Решта коливання виробництва продукції (10,9 %) зумовлена іншими факторами (забезпеченістю запасами, якістю підготовки робітників, рівнем енерго і механоозброєності праці та ін.

**Часткові коефіцієнти кореляції**. Поряд із сукупним коефіцієнтом кореляції, формою якої виступають парні коефіцієнти, обчислюються коефіцієнти так званої часткової кореляції. *Часткові коефіцієнти кореляції* — міра лінійної залежності між двома випадковими величинами з деякої сукупності випадкових величин у разі, коли виключено (елміновано) вплив інших. Ці коефіцієнти оцінюють ступінь (силу) впливу одного фактора з іншим фактором за умов елімінування (виключення) впливу третіх факторів. Коефіцієнт часткової кореляції, як і парної, змінюється не більше від +1 до –1. Порядок обчислення коефіцієнтів приватної кореляції у тому, що послідовно усувається вплив кожного чинника.

Коефіцієнт часткової кореляції можна обчислити за допомогою коефіцієнтів парної кореляції за формулою:

$$r_{1,2,3,4,\ldots,k} = \frac{r_{1,2,3,4,\ldots,k-1} - r_{1,k,3,4,\ldots,k-1} \cdot r_{2,k,3,4,\ldots,k-1}}{\sqrt{(1 - r_{1,k,3,4,\ldots,k-1})(1 - r_{2,k,3,4,\ldots,k-1})}}, \qquad (9.109)$$

де $r_{1,2,3,4,\ldots,k}$ — коефіцієнт часткової кореляції, обчислений за умови, що на результативну ознаку діють усі фактори;

$r_{1,2,3,4,\ldots,k-1}$ — коефіцієнт часткової кореляції, обчислений за умови, що на результативну ознаку діють усі фактори, крім *k*-го.

При лінійному зв'язку між *y*, *x*, *z* формула коефіцієнта часткової кореляції результативного фактора з факторною ознакою *x* при виключенні впливу фактора *z* буде така:

$$r_{yx(z)} = \frac{r_{yx} - r_{yz} \cdot r_{xz}}{\sqrt{(1 - r_{yz}^2)(1 - r_{xz}^2)}}. \qquad (9.110)$$

Частковий коефіцієнт кореляції між *y* і *x* не залежить від значень фактора *z*. На цей коефіцієнт впливає лише те, що за фактором *z* закріплюються постійні значення, а не самі значення фактора *z*.

Якщо виключити вплив фактора *x*, то частковий коефіцієнт кореляції обчислюється за формулою:

$$r_{yz(x)} = \frac{r_{yz} - r_{yx} \cdot r_{xz}}{\sqrt{(1 - r_{yx}^2)(1 - r_{xz}^2)}}. \qquad (9.111)$$

Ці значення не залежать вже від значень фактора *x*, які приймаються за постійні величини, та визначає залежність між *y* та *z*.

Можна розрахувати залежність факторних ознак при усуненні впливу результативної ознаки:

$$r_{xz(y)} = \frac{r_{xz} - r_{yx} \cdot r_{yz}}{\sqrt{(1 - r_{yx}^2)(1 - r_{yz}^2)}}. \qquad (9.112)$$

Як бачимо, коефіцієнти часткової кореляції обчислюються з урахуванням коефіцієнтів парної кореляції. Перевагою коефіцієнта часткової кореляції у тому, що є об'єктивна оцінка суттєвості його відмінності від нуля у генеральній сукупності.

Розглянемо для прикладу рівняння регресії трьох змінних:

$$y = a_0 + a_1 x + a_2 z.$$

і зробимо розрахунок за попередніми прикладами:

$$r_{yx} = 0{,}873; \quad r_{yz} = 0{,}881; \quad r_{xz} = 0{,}730.$$

$$1)\ r_{yx(z)} = \frac{r_{yx} - r_{yz} \cdot r_{xz}}{\sqrt{(1 - r_{yz}^2)\ (1 - r_{xz}^2)}} = \frac{0{,}873 - 0{,}881 \cdot 0{,}730}{\sqrt{(1 - 0{,}881^2)\,(1 - 0{,}730^2)}} =$$

$$= \frac{0{,}2299}{0{,}3234} \approx 0{,}711;$$

$$2)\ r_{yz(x)} = \frac{r_{yz} - r_{yx} \cdot r_{xz}}{\sqrt{(1 - r_{yx}^2)\ (1 - r_{xz}^2)}} = \frac{0{,}881 - 0{,}873 \cdot 0{,}730}{\sqrt{(1 - 0{,}873^2)\ (1 - 0{,}730^2)}} =$$

$$= \frac{0{,}2437}{0{,}3333} \approx 0{,}731;$$

$$3)\ r_{xz(y)} = \frac{r_{xz} - r_{yx} \cdot r_{yz}}{\sqrt{(1 - r_{yx}^2)\,(1 - r_{yz}^2)}} = \frac{0{,}730 - 0{,}873 \cdot 0{,}881}{\sqrt{(1 - 0{,}873^2)\,(1 - 0{,}881^2)}} =$$

$$= \frac{-\,0{,}0391}{0{,}2320} \approx -0{,}169.$$

Частковий коефіцієнт кореляції між $y$ і $x$ ($r_{yx(z)}$) залежить від значень фактора $z$, які приймаються за постійні значення, а не самі значення фактора $z$. Частковий коефіцієнт $r_{yx(z)}$ залежить від значень фактора $x$. На величину цього коефіцієнта впливає лише те, що за фактором $x$ закріплюються постійні значення, а самі значення фактора $x$.

Коефіцієнт часткової кореляції показує, що вплив основних засобів на випуск продукції при виключенні впливу фонду заробітної плати менше, ніж при парній кореляції (коефіцієнт кореляції 0,711 замість 0,873). Коефіцієнт часткової кореляції означає, що при виключенні впливу основних засобів вплив фонду заробітної плати на випуск продукції також дещо нижчий (коефіцієнт кореляції 0,731 замість 0,881). Коефіцієнт часткової кореляції між факторними ознаками за винятком впливу результативної ознаки виявляється негативним (негативний коефіцієнт кореляції 0,169 замість позитивного

0,730). Значення часткового коефіцієнта кореляції є невисоким (0,169), що свідчить про низький рівень зв'язку факторних ознак.

Коефіцієнт часткової кореляції не дорівнює відповідному коефіцієнту парної кореляції. Перший вимірює зв'язок між ознаками, взятими окремо, за винятком впливу інших ознак другий — тісноту зв'язку з урахуванням взаємодії з іншими явищами.

Середню помилку вибіркового коефіцієнта множинної кореляції визначають за формулою:

$$m_R = \frac{1 - R^2}{\sqrt{n - m - 1}}, \tag{9.113}$$

де $n$ — кількість спостережень;

$m$ — кількість факторів.

Згідно з даних нашого прикладу помилка коефіцієнта множинної кореляції складе:

$$m_R = \frac{1 - 0{,}944^2}{\sqrt{6 - 2 - 1}} = 0{,}063.$$

Фактичне значення $t$-критерію:

$$t_Ф = \frac{R}{m_R} = \frac{0{,}944}{0{,}063} = 14{,}98.$$

Табличне значення $t$-критерію при рівні ймовірності 0,95 і 3 ступенях свободи дорівнює 3,1825 (додаток 4). У звязку з тим, що фактичне значення $t$ значно перевищує табличне, то можна зробити висновок про вірогідність коефіцієнта множинної кореляції.

## 9.13. Коефіцієнти кореляції рангів та коефіцієнт Фехнера

Застосування лінійного коефіцієнта кореляції для оцінки тісноти зв'язку між двома ознаками є обґрунтованим лише стосовно положення про нормальний характер розподілу ознак у досліджуваній сукупності. Крім того, для обчислення лінійного коефіцієнта кореляції необхідно знати чисельні значення всіх значень факторної та результативної ознак. Якщо генеральна сукупність, з якої береться вибірка, не підпорядковується нормальному або близькому до нормального закону розподілу, то застосовуються методи рангової кореляції. Методи рангової кореляції були спочатку розроблені К. Пірсоном і потім розвинені Хетелінгом, Пабстом, Спірменом, Кенделом та ін.

В основу рангової кореляції покладено принцип нумерації варіант статистичного розподілу. Кожному члену варіаційного ряду надається ранг — номер об'єкта в порядку зростання градацій. Ранг вказує на те місце, яке займає $i$-й об'єкт серед інших об'єктів, розташованих відповідно до ознаки ($i$=1, 2, …, $n$). Ранжування індивідуальних порядкових номерів за зростанням або зменшенням дозволяє отримати попереднє уявлення про наявність або відсутність зв'язку між двома ознаками. Якщо ранги значень одиниць факторного ознаки виявляють тенденцію до збільшення, можна припустити про наявність прямого зв'язку; якщо ж зі збільшенням величини рангів факторного ознаки величини рангів результативної ознаки зменшуються, то має місце наявність зворотного зв'язку.

Рангові коефіцієнти кореляції є найбільш простими показниками, які використовуються для вимірювання тісноти кореляційної залежності. Для ознак, виміряних у порядкових шкалах, показником зв'язку служить коефіцієнт рангової кореляції.

Розглянемо два коефіцієнт рангової кореляції, розроблені К. Спірменом і М. Кенделлом, так як вони мають найбільше застосування.

**Коефіцієнт кореляції рангів Спірмена**. Коефіцієнт кореляції рангів Спірмена є звичайним коефіцієнтом між векторами рангів, відповідних аналізованим показникам.

*Коефіцієнт кореляції рангів* (*коефіцієнт Спірмена*) розраховується не за значеннями двох взаємопов'язаних ознак, а за їх рангами наступним чином[1]:

$$\rho_{x/y} = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \qquad (9.114)$$

де $\rho_{x/y}$ — коефіцієнт кореляції рангів;

$d_i^2$ — квадрати різниці рангів;

$n$ — кількість спостережень (кількість пар рангів).

---

[1] Запишемо вираз, аналогічний формулі коефіцієнта лінійної кореляції

$$\rho = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \cdot \sqrt{\sum(y-\bar{y})^2}}}.$$

Так як $x$ и $y$ приймають лише кінцеве число цілих значень, тоді $\bar{x} = \bar{y} = \frac{n+1}{2}$ і

$$\sum(x-\bar{x})^2 = \sum(y-\bar{y})^2 = n\frac{n^2-1}{12};$$

Для розрахунку коефіцієнта рангів Спірмена може використову ватися й інша формула:

$$\rho_{x/y} = 1 - \frac{6}{n^3 - n}\sum_{i=1}^{N}(R_i - S_1)^2, \qquad (9.115)$$

де $R_i, S_i$ — ранги об'єкта $i$ за ознаками $R$, $S$;

$N$ — число об'єктів, що порівнюються з кожною ознакою.

Якщо є так звані пов'язані ранги, тобто еквівалентні об'єкти, то їм приписується один і той же ранг, рівний середньому арифметичному значенню номерів цих об'єктів в ранжованому варіаційному ряду.

Коефіцієнт кореляції рангів Спірмена, так само як і лінійний коефіцієнт кореляції, може змінюватися в межах від +1 до –1 і дорівнює нулю за відсутності зв'язку між рядами рангів. Якщо зв'язок повний і прямий, то $\frac{6\sum d_i^2}{n(n^2-1)} = 0$, так як $\sum d_i^2$ =0, і $\rho = 1$. Якщо ж зв'язок повний, але зворотний, то $\frac{6\sum d_i^2}{n(n^2-1)} = 2$, так як $\sum d^2 = \frac{n(n-1)}{3}$, і $\rho = -1$. Чим ближче значення $\rho$ до 1 або до –1, тим тісніше зв'язок між значеннями ознак.

Смислове значення цього показника та його обчислення розглянемо на наступному прикладі. Маються дані щодо однотипних підприємств про обсяг реалізованої продукції ($x$, млн грн) і витрат на збут з реалізації цієї продукції ($y$, тис. грн).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Випуск продукції ($x$), млн грн | 13,2 | 29,5 | 12,1 | 19,4 | 16,8 | 24,1 | 14,8 | 36,5 | 25,4 | 20,5 |
| Витрати на збут ($y$), тис. грн | 524 | 1206 | 574 | 915 | 792 | 854 | 984 | 1271 | 987 | 802 |

За результатам наведених даних необхідно в и з н а ч и т и рівень

---

$$\sum(x-\bar{x})(y-\bar{y}) = n\frac{n^2-1}{12} - \frac{1}{2}\sum d^2,$$

де $d = x - y$;

таким чином:

$$\rho_{x/y} = 1 - \frac{6\sum d_i^2}{n(n^2-1)}.$$

впливу випуску продукції на рівень витрат на збут. Розрахунок здійснюється у наступній послідовності.

Зобразимо на графіку висхідні дані про випуск продукції і витрати на збут (рис. 9.19).


*Рис. 9.19.* Зміна сум витрат на збут у зв'язку зі зростанням випуску продукції.

Розглядаючи ці дані, можна назвати, що зі збільшенням випуску продукції сума витрат на збут загалом також зростає. Це зрозуміло: більший оборот вимагає великих витрат. Разом з тим зростання суми витрат на збут, що спостерігався в цілому, не слідує строго за зростанням випуску продукції. Ламана лінія на графіці (рис. 9.19) показує, що у ряді випадків збільшення випуску тягне за собою не збільшення, а зменшення витрат на збут. Це відбувається тому, що, крім випуску продукції, сума витрат на збут визначалася та іншими причинами, які тут не враховані і які під кутом зору поставленого завдання мають другорядний, випадковий характер. Ці порушення загальної тенденції зростання суми витрат на збут можуть бути враховані, якщо розмістити ранги за обома зіставними ознаками, тобто порядкові номери за зростанням цих ознак. Якщо зустрічаються однакові дати, то кожній із них приписується середній ранг. Наприклад, однакові суми випуску продукції, що стоять на 5-му і 6 місцях, однакові, тому кожній з них потрібно приписувати ранг 6,5.

Для обчислення коефіцієнта рангів Спірмена треба впорядкувати ряд рангів. Маємо індивідуальні величини обсягу реалізованої продукції в порядку зростання (або спадання) і встановлюємо ранг (порядкові номери величин ознаки):

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Випуск продукції ($x$), млн грн | 13,2 | 29,5 | 12,1 | 19,4 | 16,8 | 24,1 | 14,8 | 36,5 | 25,4 | 20,5 |
| Ранги $R_x$ | 2 | 9 | 1 | 5 | 4 | 7 | 3 | 10 | 8 | 6 |

Маємо індивідуальні величини витрат на збут у порядку зростання і визначаємо ранги (порядкові номери) величин отриманого ряду:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Витрати на збут (у), тис. грн | 524 | 1206 | 574 | 915 | 792 | 854 | 984 | 1271 | 987 | 802 |
| Ранги $R_y$ | 1 | 9 | 2 | 6 | 3 | 5 | 7 | 10 | 8 | 4 |

Ці порядкові номери, або ранги, обох ознак можна використовувати для конструювання показника тісноти зв'язку між ними. І тому необхідно обчислити різниці між рангами і звести їх в квадрат.

Складемо розрахункову таблицю для обчислення рангового коефіцієнта кореляції.

Таблиця 9.34

| Номер підприємств | $x$ | $y$ | Ранжування | | | | Порівняння рангів | | Різниця рангів $d_i = R_x - R_y$ | $d_i^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $x$ | $y$ | ранг | | | | | |
| | | | | | $R_x$ | $R_y$ | $R_x$ | $R_y$ | | |
| 1 | 13,2 | 524 | 12,1 | 524 | 1 | 1 | 2 | 1 | 1 | 1 |
| 2 | 29,5 | 1 206 | 13,2 | 574 | 2 | 2 | 9 | 9 | 0 | 0 |
| 3 | 12,1 | 574 | 14,8 | 792 | 3 | 3 | 1 | 2 | –1 | 1 |
| 4 | 19,4 | 915 | 16,8 | 802 | 4 | 4 | 5 | 6 | –1 | 1 |
| 5 | 16,8 | 792 | 19,4 | 854 | 5 | 5 | 4 | 3 | 1 | 1 |
| 6 | 24,1 | 854 | 20,5 | 915 | 6 | 6 | 7 | 5 | 2 | 4 |
| 7 | 14,8 | 984 | 24,1 | 984 | 7 | 7 | 3 | 7 | –4 | 16 |
| 8 | 36,5 | 1 271 | 25,4 | 987 | 8 | 8 | 10 | 10 | 0 | 0 |
| 9 | 25,4 | 987 | 29,5 | 1 206 | 9 | 9 | 8 | 8 | 0 | 0 |
| 10 | 20,5 | 802 | 36,5 | 1 271 | 10 | 10 | 6 | 4 | 2 | 4 |

$\sum = 28$

Коефіцієнт кореляції рангів Спірмена обчислимо за формулою:

$$\rho_{x/y} = 1 - \frac{6\sum d_i^2}{n\,(n^2-1)} = 1 - \frac{6 \cdot 28}{10\,(10^2-1)} = 1 - \frac{168}{990} = 0{,}830.$$

Величина коефіцієнта кореляції рангів Спірмена свідчить про наявність досить тісного прямого тісного зв'язку між обсягом реалізованої продукції та витратами на збут.

Ранговий коефіцієнт кореляції має перевагу перед лінійним. Він може бути обчислений при будь-якій формі розподілу досліджуваних ознак, тоді як $r$ пристосований і дає хороші результати лише при їх нормальному розподілі. Подібно до лінійного ранговий коефіцієнт може бути оцінений на суттєвість.

Такий один із найпростіших показників тісноти кореляційної залежності, пропонований математичною статистикою.

**Коефіцієнт кореляції рангів Кендела**. М. Кендел запропонував ще одну міру зв'язку між змінними $x$ і $y$ (коефіцієнт кореляції рангів Кендела — $\tau$):

$$\tau = \frac{2S}{n\,(n-1)}, \text{ где } S=P+Q. \qquad (9.116)$$

Для обчислення $\tau$ треба впорядкувати ряд рангів змінної $x$ та змінної $y$.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Ранги $R_x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Ранги $R_y$ | 2 | 1 | 7 | 3 | 6 | 4 | 5 | 8 | 9 | 10 |

Для знаходження $S$ необхідно обчислити два доданки: $P$ і $Q$. При визначенні доданку $P$ потрібно встановити, скільки чисел, що знаходяться праворуч від кожного з елементів послідовності рангів змінної $y$, мають величину рангу, що перевищує ранг об'єкта, що розглядається. Так, наприклад, першому значенню в послідовності рангів змінної $y$, тобто числу 2, відповідає вісім чисел (7+3+6+4+5+8+9+10), які перевищують ранг 2; другому значенню 1 відповідає вісім чисел (7+3+6+4+5+8+9+10), що перевищує одиницю і т. д. Підсумовуючи отримані таким чином числа, ми отримуємо доданок $P$, яке можна розглядати як міру відповідності послідовності рангів змінної $x$. Для нашого прикладу $P$ = 39 (8 +8 +3 +6 +4 +4 +3 +2 +1).

Другий доданок $Q$ характеризує ступінь невідповідності послідовності рангів змінної у послідовності рангів $x$. Щоб визначити $Q$, підраховуємо, скільки чисел, що входять праворуч від кожного з членів послідовності рангів змінної у має ранг менше, ніж цей фактор. Такі величини беруться із знаком мінус. У прикладі $Q$= –7 (–1–0–4–0–2–0–0–0–0). Отже, $S = P + Q$ = 39–7 = 32.

Коефіцієнт кореляції рангів Кендела в нашому прикладі складе:

$$\tau = \frac{2S}{n\,(n-1)} = \frac{2 \cdot 32}{10 \cdot 9} = 0{,}711.$$

Згідно з отриманим коефіцієнтом кореляції Кендела між випуском продукції і витратами на збут існує тісний прямий зв'язок. Коефіцієнт кореляції рангів Кендела також, як і коефіцієнт рангів Спірмена, змінюється від – 1 до + 1.

**Коефіцієнт конкордації**. Для визначення тісноти зв'язку між трьома і більше ознаками застосовується ранговий коефіцієнт згоди — *коефіцієнт конкордації*, який обчислюється за формулою:

$$W = \frac{12S}{m^2\,(n^3 - n)}, \tag{9.117}$$

де $m$ — кількість факторів;
$n$ — число спостережень;
$S$ — сума квадратів відхилень рангів.

Розглянемо приклад розрахунку коефіцієнта конкордації.

Маються такі дані, що характеризують обсяг реалізованої продукції, витрати на збут, собівартість одиниці продукції та місячна заробітна плата робітників (табл. 9.35).

Таблиця 9.35

| Номер підприємства | Реалізація продукції, млн грн ($y$) | Витрати на збут, тис. грн ($x$) | Собівартість одиниці продукції, грн ($z$) | Середня місячна заробітна плата, тис. грн ($w$) |
|---|---|---|---|---|
| 1 | 13,2 | 524 | 69,7 | 16,9 |
| 2 | 29,5 | 1 206 | 71,3 | 15,8 |
| 3 | 12,1 | 574 | 72,1 | 17,2 |
| 4 | 19,4 | 915 | 79,6 | 18,3 |
| 5 | 16,8 | 792 | 66,2 | 16,1 |
| 6 | 24,1 | 854 | 69,3 | 16,6 |
| 7 | 14,8 | 984 | 65,4 | 16,3 |
| 8 | 36,5 | 1 271 | 73,8 | 17,5 |
| 9 | 25,4 | 987 | 72,9 | 16,2 |
| 10 | 20,5 | 802 | 70,5 | 16,4 |

В и з н а ч и м о тісноту зв'язку між обсягом реалізації продукції, сумою витрат на збут, собівартістю одиниці продукції та середньою заробітною платою робітників.

Тісноту зв'язку розрахуємо за допомогою коефіцієнта конкордації (табл. 9.36).

Таблиця 9.36

| $R_y$ | $R_x$ | $R_z$ | $R_w$ | Сума рядків | Квадрати сум |
|---|---|---|---|---|---|
| 2 | 1 | 1 | 7 | 11 | 121 |
| 9 | 9 | 6 | 1 | 25 | 625 |
| 1 | 2 | 7 | 8 | 18 | 324 |
| 5 | 6 | 10 | 10 | 31 | 961 |
| 4 | 3 | 3 | 2 | 12 | 144 |
| 7 | 5 | 4 | 6 | 22 | 484 |
| 3 | 7 | 2 | 4 | 16 | 256 |
| 10 | 10 | 9 | 9 | 38 | 1 444 |
| 8 | 8 | 8 | 3 | 27 | 729 |
| 6 | 4 | 5 | 5 | 20 | 400 |
| | | | | $\sum = 220$ | $\sum = 5488$ |

$$S = 5488 - \frac{(220)^2}{10} = 5\,488 - 4\,840 = 648;$$

$$W = \frac{12S}{m^2(n^3 - n)} = \frac{12 \cdot 648}{16\,(10^3 - 10)} = \frac{7\,776}{15\,840} = 0{,}49.$$

Величина коефіцієнта конкордації свідчить, що між досліджуваними величинами є тісний зв’язок. Ця залежність пояснюється тим, що аналізовані величини характеризують ефективність виробничо-господарської діяльності підприємств.

**Коефіцієнт Фехнера**. Перші спроби вимірювання зв’язку між факторами зробив Г. Фехнер, який запропонував найпростіший показник тісноти зв’язку — коефіцієнт кореляції рангів (коефіцієнт Фехнера).

*Коефіцієнт знаків* (*коефіцієнт Фехнера*[1]) обчислюється на підставі визначення знаків відхилень варіантів двох взаємопов’язаних ознак від їх середньої величини. Він оцінює зв’язок з урахуванням порівняння знаків відхилень значень ознак від середніх арифметичних. Для розрахунку цього показника обчислюють середні значення результативної та факторної ознак. Слідом за визначенням лінійних відхилень від середньої зіставляють знаки відхилень для всіх значень взаємопов’язаних ознак.

---

[1] Цей показник був запропонований німецьким статистиком *Г. Фехнером* у 1897 р. та отримав його ім’я.

Коефіцієнт Фехнера ($K_ф$) обчислюється за формулою:

$$K_ф = \frac{\sum a - \sum b}{\sum a + \sum b}, \tag{9.118}$$

де $a$ — співпадіння знаків відхилень індивідуальних значень від середньої;

$b$ — неспівпадіння знаків відхилень індивідуальних значень від середньої.

Коефіцієнт Фехнера може набувати різних значень в межах від +1 до –1. Якщо знаки всіх відхилень співпадуть, тобто якщо $\sum b$=0, то $K_ф$ виявиться рівним 1, що свідчить про наявність повного прямого зв'язку між ознаками, що зіставляються. Навпаки, якщо знаки відхилень індивідуальних відхилень збігатимуться, тобто якщо $\sum a$=0, то $K_ф$ буде рівним –1, що говорить про повний, але зворотний зв'язок. Проміжні значення $K_ф$ характеризують ступінь близькості до функціональної зв'язку.

Розрахуємо коефіцієнт знаків Фехнера за даними табл. 9.14 про виробництво цукру, побудувавши при цьому табл. 9.38.

Таблиця 9.38

| Середньорічна вартість промислово-виробничих засобів, млн грн ($x$) | Виробництво цукру, тис. ц ($y$) | $x-\bar{x}$ | $y-\bar{y}$ | Співпадіння або неспівпадіння знаків |
|---|---|---|---|---|
| 2,0 | 210 | –3,4 | –104,4 | *a* |
| 2,8 | 135 | –2,6 | –179,4 | *a* |
| 4,0 | 240 | –1,4 | –74,4 | *a* |
| 4,5 | 264 | –0,9 | –50,4 | *a* |
| 5,0 | 300 | –0,4 | –14,4 | *a* |
| 5,7 | 310 | 0,3 | –4,4 | *b* |
| 6,5 | 365 | 1,1 | 50,6 | *a* |
| 7,0 | 380 | 1,6 | 65,6 | *a* |
| 7,8 | 440 | 2,4 | 125,6 | *a* |
| 8,7 | 500 | 3,3 | 185,6 | *a* |
| $\sum$54 | $\sum$3144 | | | |

Попередньо розрахуємо середні арифметичні $\bar{x}$ і $\bar{y}$. Середня вартість промислово-виробничих фондів ($\bar{x}$) становить 5,4 млн грн (54:10); середнє вироблення цукру ($\bar{y}$) — 314,4 тис. ц (3144:10).

Обчислимо суму знаків відхилень значень ознак від середніх арифметичних. Знак мінус означає, що значення ознаки менше середньої, знак плюс — більше. Збіг знаків за обома ознаками означає узгоджену варіацію, розбіжність — порушення узгодженості. У прикладі число збігів знаків становить $\sum a = 9$, а число розбіжностей ($\sum b$) — відповідно 1.

Коефіцієнт Фехтера складе:

$$K_{ф} = \frac{\sum a - \sum b}{\sum a + \sum b} = \frac{9-1}{9+1} = \frac{8}{10} = 0{,}8.$$

Отримана величина коефіцієнта Фехнера $K_{ф}$ свідчить про те, що між середньорічною вартістю основних промислово-виробничих засобів і середнім виробництвом цукру існує прямий тісний зв'язок. Однак коефіцієнт Фехнера дуже примітивний: він уловлює тільки напрям варіації і не враховує величину відхилень факторного і результативного ознак. Тому не можна говорити про ступінь тісноти кореляційного зв'язку, а тим більше про оцінку її суттєвості на підставі лише коефіцієнта Фехнера.

## ЛІТЕРАТУРА


1. *Горкавий В. К.* Статистика: Підручник. Третє вид., переробл. і доповн. Київ: Алерта, 2019. С. 239–87.
2. Статистика: Підручник / *С. С. Герасименко, А. В. Головач, А.М. Єріна* та ін.; За наук. ред. д-ра екон. наук С. С. Герасименка. 2-ге вид., перероб. і доп. Київ: КНЕУ, 2000. С. 100–120.
3. *Ткач Є. І., Сторожук В. П.* Загальна теорія статистики: підручник. 3-тє вид. Київ: Центр учбової літератури, 2009. С. 196–241.


## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники, словники*


4. *Мармоза А. Т.* Теорія статистики: підручник. 2-ге вид., перероб. і доп. Київ: «Центр учбової літератури», 2013. 333–396.
5. *Опря А. Т.* Статистика (модульний варіант з програмованою формою контролю знань). Навч. посіб. Київ: Центр учбової літератури, 2012. С. 158–203.
6. *Педченко Г. П.* Статистика: Навчальний посібник. Мелітополь: Колор Принт, 2018. С. 32–57.
7. Статистичний словник / [*О. Г. Осауленко*, *О. О. Васєчко*, *М. В. Пугачова* та ін.]; за ред. д-ра держ. упр., проф., член-кор. НАН України О. Г. Осау-

ленка; НТК статистичних досліджень. Київ: ДП «Інформ.-аналіт. агентство», 2012. 498 с.
8. *Шапочка М. К.*, *Маценко О. М.* Теорія статистики: навч. посібник. Суми: Університетська книга, 2014. С. 203–218.

*Монографії та статті*

9. *Львівський Е. Н.* Статистичні методи побудови емпіричних формул. Київ: Вища школа, 1988.
10. Ольвінська Ю. О., Самотоєнкова О. В. Використання методу рангової кореляції при аналізі розвитку малого підприємництва. *Вісник соціально-економічних досліджень*. 2015. Випуск 3(58). С. 160–169.

Г Л А В А 10

# РЯДИ ДИНАМІКИ

## 10.1. Ряди динаміки та їх види

**Поняття про ряди динаміки**. При вивченні теми «Зведення та групування статистичних матеріалів» вивчалися ряди розподілу, в яких визначалися окремі варіанти варійованої ознаки і чисельності одиниць сукупності, що несуть ці варіанти. Наприклад, обсяг виробленої продукції, чисельність працівників, вартість основних виробничих засобів і т. п. Для таких рядів не має значення, в якій послідовності виникли ці явища.

Керуючись діалектичним методом пізнання, статистика розглядає суспільні явища у безперервному розвитку. При вивченні розвитку явища або процесу суспільного життя в часі найсуттєвіше значення набуває послідовність виникнення значень показника. У зв’язку з цим найважливішим завданням статистики є постійне вивчення цих змін у просторі та на території.

Для дослідження та аналізу суспільно-економічних явищ у статистиці складають таблиці, в яких показники розташовані в хронологічному порядку. Ці показники, які у своїх змінах відображають хід розвитку досліджуваного явища і утворюють ряди динаміки, або часові ряди. Отже, *динамічним рядом* називають ряд показників, що характеризують величину будь-якого явища станом на певні моменти або періоди часу (інтервали)[1].

---

[1] У зарубіжній статистичній літературі використовується термін (в англомовній літературі термін time series, у німецькій — zeitreihen analyze), що у перекладі означає «часові ряди» або «аналіз часових рядів». Однак з урахуванням особливостей перекладу ми будемо використовувати класичну назву «динамічний ряд» або «ряд динаміки», які більш точно відображають сутність цієї теми.

Ряд динаміки складається з двох елементів: у ньому вказуються *моменти часу* (зазвичай дати) або періоди часу (місяці, квартали, роки), до яких відносяться статистичні дати, що наводяться, і *статистичні показники*, що характеризують стан явища або групи явищ в ці моменти або періоди, називають *рівнями ряду*. Обидва елементи — час і рівень ряду — називаються *членами ряду динаміки*. Особливо виділяють *початковий рівень* — перший рівень ряду, *кінцевий рівень* — останній рівень ряду і *середній рівень* ряду динаміки, який розраховується як середня з рівнів ряду.

У статистиці рівні ряду частіше позначаються через $y$, моменти або періоди часу, яких ставляться рівні, через $t$.

| | | | | |
|---|---|---|---|---|
| Роки | $t_1$ | $t_2$ | $t_3$ | $t_4 \ldots t_n$ |
| Рівні | $y_1$ | $y_2$ | $y_3$ | $y_4 \cdots y_n$ |

Крім того, можуть бути ще й похідні аналітичні показники.

Наприклад, у табл. 10.1 є наступні два динамічні ряди економічних показників торгівлі.

Таблиця 10.1

**Динаміка товарообігу та товарних запасів на підприємствах оптової торгівлі в Україні за 2015–2020 роки.**

(млрд грн)

| Показник | 2015 р. | 2016 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|---|---|
| Загальний обсяг товарообігу за рік | 1 244,2 | 1 556,0 | 1 908,7 | 2 215,4 | 2 322,2 | 2 462,6 |
| Товарні запаси в оптовій торгівлі на кінець року | 156,6 | 203,1 | 249,6 | 280,7 | 288,0 | 326,4 |

За допомогою динамічних рядів досліджується процес розвитку явища (зростання, зменшення, стабільність), визначається середній рівень показника та середня інтенсивність зміни, встановлюється закономірність і тенденція в зміні показників.

Використовуючи дані, що є у звітності підприємств та організацій, українська статистика дає вичерпну характеристику динаміці розвитку народного господарства, культури та науки.

**Види рядів динаміки**. Ряди динаміки можна класифікувати за різними ознаками. Розглянемо найбільш головні з їх.

Ряди динаміки абсолютних величин характеризують зміну суспільних явищ або певні періоди часу, або певні моменти часу. В залежності від цього вони поділяються на моментні, чи інтервальні ряди динаміки.

У *моментних рядах* динаміки рівні ряду динаміки виражають величину будь-яких суспільно-економічних явищ станом на певні моменти часу (наприклад, на початок або кінець місяця, кварталу, року тощо). В них час позначається момент, до якого належить кожен рівень ряду. У моментних рядах динаміки період між датами називається *інтервалом ряду*. Величина інтервалу залежить від характеру змін явищ: для явищ, які швидко змінюються, ряди повинні мати більш короткі інтервали.

Прикладом моментних рядів динаміки можуть бути дані про чисельність населення України (табл. 10.2).

Таблиця 10.2

**Чисельність населення України**
(на 1 січня)

| Рік | 2013 р. | 2014 р. | 2015 р. | 2016 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|---|---|---|---|
| Млн осіб | 45,6 | 45,4 | 42,9 | 42,8 | 42,4 | 42,2 | 41,9 | 41,6 |

Це моментний ряд динаміки абсолютних величин з рівновідстоюючими рівнями в часі. Рівні цього ряду характеризують узагальнені підсумки обліку населення на певну дату (на 1 січня кожного року). Оскільки окремі особи, враховані, наприклад, у 2013 р. живуть і зараз, то вони були одиницями сукупності та в подальшому обліку руху населення. На певні моменти часу (початок року, початок кварталу або місяця) характеризується динаміка чисельності персоналу підприємства, запасів товарно-матеріальних цінностей, вартості основних засобів, числа підприємств, вартості та кількість насіння сільськогосподарської техніки, насіння, худоби в сільському господарстві, рухомого складу на транспорті, запасів товарів у торгівлі тощо.

Як приклад моментних рядів наведемо дані про розвиток транспортних шляхів співвідношення нашій країні (табл. 10.3).

Ряд динаміки, наведений у табл. 10.3, містить два елементи: час та рівень. Рівні в цьому ряді динаміки наведені у вигляді абсолютних чисел. З таблиці видно, що для багатьох видів транспорту характерна тенденція до стабілізації або скорочення розвитку транспортних шляхів сполучення. Особливо скорочувалася експлуатаційна довжина залізниць. Зазначені в таблиці ряди динаміки є моментними, оскільки

Таблиця 10.3

**Розвиток транспортних шляхів сполучення в Україні**
(на конець року; тис. км)

| Показатели | 2016 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|---|
| Експлуатаційна довжина залізничних колій | 21,0 | 19,8 | 19,8 | 19,8 | 19,8 |
| Довжина автомобільних доріг загального користування | 163,0 | 163,1 | 161,9 | 161,9 | 162,2 |
| Експлуатаційна довжина тролейбусних ліній (в однопутному обчисленні) | 3,3 | 3,4 | 3,4 | 3,4 | 3,4 |
| Експлуатаційна довжина метрополітенівських колій (у двоколійному обчисленні), км | 113,4 | 113,4 | 113,4 | 113,4 | 113,4 |

вони характеризують стан розвитку транспортних шляхів на певні моменти часу (кінець року). У прикладі інтервал річний.

У *інтервальних рядах* рівні ряду характеризують розміри суспільно-економічних явищ за певні відрізки (інтервали) часу. Вони відображаються результати діяльності людей, і навіть витрати продуктивних сил суспільства, діяльності, і навіть витрати, пов'язані з цією діяльністю. Ці результати і витрати спостерігають не в порядку одноразових обліків, а шляхом постійного їх обліку в часі. Так, кількість виготовленої продукції визначають шляхом щоденного її підрахунку. Таким же чином визначають витрати матеріалів, праці та інших факторів виробництва на виготовлення цієї продукції. Прикладом інтервального ряду можуть бути дані про виробництво електроенергії в Україні за 2000–2019 рр. (табл. 10.4).

Таблиця 10.4

**Виробництво електроенергії в Україні**

| Рік | 2000 р. | 2005 р. | 2010 р. | 2015 р. | 2016 р. | 2017 р. | 2018 р. | 2019 р. |
|---|---|---|---|---|---|---|---|---|
| Млрд кВт·год | 171,4 | 186,1 | 188,6 | 163,7 | 164,6 | 156,0 | 159,9 | 154,1 |

Це інтервальний ряд динаміки абсолютних величин з нерівновіддаленими рівнями в часі. У цьому ряду кожен рівень є результатом будь-якого процесу саме за той період, до якого він віднесений. Цифри ряду характеризують сумарний результат виробленої електроенергії в Україні за певний проміжок часу (за рік). Так, виробництво електроенергії в кількості 154,1 млрд кВт·год. здійснено повністю (і лише) у 2019 р. Рік є інтервалом у цих рядах. Інтервалами можуть бути інші інтервали: дні, місяці, квартали, п'ятирічки тощо.

В економічній статистиці інтервальні ряди набули широкого поширення. Їх можна побудувати для аналізу: кількості та вартості виробленої продукції, інвестицій, обсягів закупівель сировини та матеріалів, витрат на виробництво продукції, браку продукції та ін.

Через різницю понять «інтервал» в інтервальних і моментних рядів випливають деякі особливості рівнів відповідних рядів.

В інтервальному ряді інтервал — проміжок часу, за який узагальнено наведені відомості. Величина рівня інтервального ряду є результатом будь-якого процесу за той чи інший період часу. Зазначена властивість інтервального динамічного ряду дозволяє підсумовувати рівні цього ряду, укрупнюючи його інтервали. В результаті отримані нові рівні рівні мають економічний зміст. Сума показників періодичного ряду характеризує величину явища за триваліший (сумарний) період. Істотною особливістю інтервальних рядів і те, що рівень інтервального ряду за інших рівних умов тим більше, що більше довжина інтервалу, якого належить цей рівень. Так, обсяг випущеної продукції за рік більший, ніж за який-небудь квартал, квартальні — більше місяцях і т. д. Підсумки, отримані в результаті підсумовування складових їх даних, мають реальний зміст. Це означає, що порядок величин в інтервальному ряду визначається розміром інтервалу. Тому в інтервальному ряду динаміки рівні за послідовні (наступні один за одним) періоди часу можна підсумовувати, отримуючи підсумки (рівні) за більш тривалі періоди.

У моментному ряді інтервал — це проміжок часу між датами, на які зроблено відомості. Відмінною особливістю моментних рядів і те, що сума рядів членів низки немає реального сенсу. Тут теж є інтервали — проміжки часу, до яких відносяться рівні ряду. Наприклад, чисельність персоналу підприємства на початок 2022 р. не залежить від того, чи приурочені інші рівні до кінця кожного року або до кінця кожного кварталу. Скажімо, велика частина персоналу підприємства на 1 січня 2018 р. увійде наступний суміжний ряд — на 1 січня 2019 р.; частина персоналу, яка працювала на 1 січня 2019 р., увійде в наступний рівень — на 1 січня 2020 р. і т. п. Тому при підсумовуванні рівнів моментного ряду здебільшого одні одиниці сукупності увійдуть у результаті двічі, інші тричі і більше разів. Це означає, що показника ряду динаміки не можна безпосередньо підсумовувати, оскільки отримані при цьому підсумки позбавлені самостійної економічної значущості; якщо воно здійснюється, то не з метою отримання реальних сум, а лише як проміжний етап при обчисленні середніх рівнів. Має значення лише підрахунок різниці між рівнями динамічного моментного ряду. Вона характеризує зміну рівнів ряду за певний період, наприклад, між чисельністю населення на 1 січня 2021 і 1 січня 2022 р. Але щоб моментні ряди динаміки правильно відображали динаміку явищ, потрібно прагнути до того, щоб перерви, тобто інтервали ряду були коротшими, особливо якщо стан явища швидко змінюється.

Якщо рівні інтервального ряду залежать від тривалості спостережуваних і досліджуваних періодів, то рівні моментного ряду від періодичності спостереження не залежать.

В залежності від способу вираження рівнів ряди динаміки поділяються на ряди *абсолютних* (як наведений вище ряд), *відносних* і *середніх величин*.

Вихідними, початковими з метою оцінки динаміки розвитку явищ і процесів є ряди динаміки абсолютних величин. На основі динаміки абсолютних величин можуть бути отримані ряди динаміки відносних і середніх величин. Ці ряди утворюються не з безпосередніх спостережень, а з рядів динаміки з абсолютними показниками. Тому ряди динаміки, в яких рівні виражені абсолютними показниками, називають *основними*, а якщо рівні ряду виражені середніми і відносними величин, то такі ряди називаються *допоміжними*, або правильніше *похідними*. Наведені досі ряди показують зміну в часі абсолютних показників, в табл. 10.2 — чисельність населення, в табл. 10.4 — виробництво електроенергії. Прикладом низки абсолютних величин можуть бути дані про виробництво нафти, вугілля, нафти, зерна, верстатів, деталей, вузлів і т. п., чисельності персоналу, вартості основних засобів, собівартості продукції, кількості шлюбу за кілька років.

Ряд динаміки середніх величин наведено у табл. 10.5.

Таблиця 10.5

**Середня тривалість життя в Україні,** років

| | 2000 р. | 2005 р. | 2010 р. | 2015 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|---|---|---|---|
| Обидві статі | 67,72 | 67,96 | 70,44 | 71,38 | 71,98 | 71,76 | 72,01 | 72,01 |
| Жінки | 73,53 | 62,23 | 75,50 | 76,25 | 76,78 | 76,72 | 76,98 | 76,22 |
| Чоловіки | 62,10 | 73,97 | 65,25 | 66,37 | 67,02 | 66,69 | 66,92 | 66,39 |

Цей ряд динаміки отримано з основного ряду, у якому середня чисельність жителів дана певні моменти. Таблиця показує зміну середньої тривалості життя чоловіків і жінок в Україні за 2000–2020 рр. Слід зазначити, що зростання середньої тривалості життя було нерівномірним. Прикладами середніх величин є дані про урожайність сільськогосподарських культур, про середню заробітну плату на одного працівника, про середню вироблення продукції на одного робітника, тривалість обробки виробу, про середній удій на одну корову, про собівартість одиниці продукції і т. п. Подібно до інтервальних рядів, вони характеризують періоди часу, але підсумовування їх рівнів самостійного значення не має.

У табл. 10.6 даний ряд відносних показників.

Таблиця 10.6

**Зростання валового внутрішнього продукту України**
(в постійних цінах 2016 р.)

| | 2010 р. | 2013 р. | 2015 р. | 2016 р. | 2017 р. | 2018 р. | 2019 р. |
|---|---|---|---|---|---|---|---|
| Валовий внутрішній продукт, в % до 2016 р. | 109,5 | 115,8 | 97,6 | 100,0 | 102,5 | 106,0 | 109,4 |

Це найпоширеніший тип ряду динаміки відносних величин, що містить темпи зростання досліджуваного явища. І цей ряд динаміки також отримано з основного ряду, в якому відображається вартість валового внутрішнього продукту за певний рік. Зростання валового внутрішнього продукту змінюється його скороченням. У 2019 р. обсяг валового внутрішнього продукту перевищив рівень 2016 р. на 9,4 %.

*За повнотою часу*, що відображається у рядах динаміки, їх можна розділити на ряди *повні* та *неповні*. У *повному ряді* динаміки одноіменні моменти або періоди часу суворо слідують один за одним у календарному порядку через певні проміжки часу. якщо ж у рядах даються переривальні періоди або нерівномірні проміжки часу між датами, то ряд динаміки називається *неповним*. У зв'язку з цим перші іноді називають *рядами динаміки з рівновіддаленими рівнями*, другий — з *нерівновіддаленими рівнями*.

Наведені у табл. 10.2 ряди динаміки чисельності населення України мають рівну відстань рівнів один від одного і є рівновіддаленим рядом. Ряд динаміки, що характеризує виробництво електроенергії в Україні (табл. 10.4) є різновіддаленим (неповним) рядом, оскільки його рівні мають різну відстань один від одного.

За складністю утворення рівнів ряди динаміки поділяються на одновимірні та багатовимірні. *Одновимірний ряд* — це ряд, рівні якого складаються з одновимірних величин, а рівні багатовимірного — з багатовимірних величин. З одновимірними рядами ми познайомилися вище. Багатовимірні динамічні ряди можуть відображати зміну: одного і того ж явища, що відноситься до різних об'єктів, зокрема, до країн або районів країни, організації або підприємству і т. п.; двох чи кількох взаємопов'язаних явищ, які стосуються одного й того ж об'єкту — до однієї країні, підприємству, цеху, відділу і т. п. Вказані показники отримуються у результаті розгляду явища як зміни всієї його структури.

У табл. 10.2 та 10.4 ми маємо одновимірний ряд динаміки, в табл. 10.7 — багатовимірний.

Таблиця 10.7

**Індекс споживчих цін в 2010–2020 рр., %**

| | 2010 р. | 2015 р. | 2017 р. | 2018 р. | 2019 р. |
|---|---|---|---|---|---|
| Україна | 9,4 | 48,7 | 14,4 | 10,9 | 7,9 |
| Австрія | 1,8 | 0,9 | 2,1 | 2,0 | 1,5 |
| Бельгія | 2,2 | 0,6 | 2,1 | 2,1 | 1,4 |
| Данія | 2,3 | 0,5 | 1,1 | 0,8 | 0,8 |
| Італія | 1,5 | 0,0 | 1,2 | 1,1 | 0,6 |
| Швеція | 9,3 | 3,2 | 4,3 | 2,4 | 2,2 |

Багатомірні ряди набули широкого поширення при дослідженні соціально-економічних явищ, складних за своєю структурою.

За способом отримання статистичного показника, в якому виражаються рівні, динамічні ряди поділяються на *первинні* та *вторинні*. До перших відносяться динамічні ряди, рівні яких отримані в результаті безпосереднього зведення даних статистичного спостереження (чисельність працюючих, обсяг виробленої продукції), до других — динамічні ряди, рівні яких отримані на основі рівнів тих чи інших первинних динамічних рядів (або інших вторинних рядів) і, як правило, є відношенням двох рівнів (питома вага, середня заробітна плата, відсотки виконання плану). Наведені у табл. 10.2 і 10.3 ряди абсолютних величин є вихідними, первісними, тому що абсолютні показники, що лежать в їх основі, виходять безпосередньо при підрахунку підсумків статистичного спостереження. Динамічні ряди середніх і відносних величин (табл. 10.5 і табл. 10.6) є похідними, так як показники, що лежать в їх основі, обчислюються шляхом відповідних перетворень.

У статистиці ряди динаміки піддають різноманітній обробці. Так, обчислюються абсолютні прирости, темпи зростання і темпи приросту, середні показники, індекси сезонності та ін. Для виявлення тенденції розвитку соціально-економічних явищ, їх взаємозв'язку в динаміці застосовуються різні методи (метод ковзної середньої, регресійний аналіз, кореляційний аналіз ).

### 10.2. Правила формування динамічних рядів

**Порівнянність рівнів ряду — основна передумова аналізу рядів динаміки**. Найважливішою передумовою дослідження динаміки та зрушень у структурі тих або інших соціально-економічних явищ є порівнянність (співставність) рівнів динамічного ряду. *Співставність* (*порівнянність*) *статистичних даних* — можливість порівняння, зіставлення даних з метою виявлення тенденцій, закономірностей розвитку відображуваних ними явищ, змін, що відбуваються в просторі та в часі. Вона означає, що рівні динамічного ряду повинні мати загальний зміст, відноситися до однакової території, включати однакове коло об'єктів, виражатися в однакових одиницях виміру, реєструватися за однаковий період часу, підраховуватися за єдиною методологією і т. д. Проблема забезпечення сумісності даних особливо актуальна для динамічних рядів, тому що вони відображають розвиток будь-якого явища за тривалий період часу, за які могли відбутися зміни, що призводять до непорівнянності статистичних даних. Особливо гостро проблема сумісності динамічних рядів стоїть в рядах великої протяжності, оскільки в них часто надзвичайно важко встановити всі зміни, що відбулися у окремі періоди часу.

У дослідженні явищ і процесів виділяють такі основні види їх сумісності: по одиниці виміру, територіальну, хронологічну, відомчу, по об'єкту і одиниці спостереження, за методологією обчислення рівнів. *Порівнянність за одиницю виміру* зводиться до виконання наступного умови: всі рівні динамічного ряду повинні бути виражені однією і тією ж одиницею виміру (у грошових одиницях, в натуральних одиницях та ін.). Однак при використанні грошових одиниць виміру показники, виражені в тих самих грошових одиницях (гривнях), можуть бути непорівнянні. *Територіальна порівнянність рівнів динамічного ряду* забезпечується єдністю території (країни, області, регіону, економічного району і т. п.), за якими визначені всі рівні ряду. Це означає, що тільки ряди, всі рівні яких визначені в одних і тих же територіальних межах, можуть бути порівнянними (сумісними). *Відомча сумісність* означає порівнянність даних у межах того чи іншого відомства (міністерства, суду, прокуратури). У той самий час розукрупнення (реорганізація) відомств, передача окремих функцій іншим відомствам може вплинути на порівнянність рядів динаміки. *Хронологічна сумісність*, т. е. порівнянність рівнів динаміки низки у часі, характеризується одиницею часу, відносно якого у ньому представлені ряди динаміки.

Перш ніж показники ряду динаміки піддавати обробці та аналізу, необхідно встановити, наскільки дотримані правила порівнянності всіх рівнів, що до нього входять. Не можна обробляти та аналізувати ряди динаміки, якщо в них наведені непорівнянні дані.

Неспівставність даних ряду динаміки може бути обумовлена різними причинами. Найважливіші з них такі: зміна територіальних кордонів, різний ступінь охоплення одиниць об’єкта, відмінності в часі реєстрації даних спостереження, недосконалість методології статистичного спостереження, зміна одиниць вимірювання. Можливі й інші причини несумісності рівнів динамічного ряду.

Часто в статистичній практиці виявляється непорівнянність даних, що виникає внаслідок адміністративно-територіальних змін. Так, внаслідок адміністративної реформи територію м. Харкова збільшено на 4450,58 га[1]. Розширення кордонів Харкова в 2013 р. призвело до того, що статистичні відомості (про кількість населення, житлової площі, кількість торгових підприємств і т. д.) в колишніх і нових кордонах міста стали непорівнянні. Цими показниками не можна використовувати для характеристики динаміки населення Харкова, динаміки житлової площі, валового регіонального продукту та інших показників. Щоб забезпечити сумісність, потрібно дані щодо Харкова за роки перерахувати у нових кордонах. Аналогічні проблеми виникають і в окремих районах, областях і навіть у державі в цілому, якщо їх межі

---

[1] Про зміну і встановлення меж міста Харків, Дергачівського і Харківського районів Харківської області: Постанова Верховної Ради України від 6.09.2012 р. № 5215-VI. https://zakon.rada.gov.ua/laws/show/5215-VI#Text.

змінюються.

Однак у деяких випадках розширення територіальних кордонів не призводить до несумісності цифр. Наприклад, якщо стоїть завдання визначити, як змінилася територія м. Харкова в результаті зміни його кордонів, то потрібно дані після зміни кордонів зіставити з даними Харкова в колишніх кордонах. Отже, забезпечення сумісності значною мірою залежить від поставленого задачі.

Неспівставність статистичних даних виникає також у результаті зміни одиниць виміру, або одиниць рахунку. Не можна порівнювати та аналізувати валовий внутрішній продукт України у 1993–1997 р., оскільки за одні роки вартість виробленої продукції враховувалася у карбованцях, а в інші — у гривнях. Неспівставні рівні валового збору сільськогосподарських культур, якщо за одні роки показники наведені в центнерах, а в інші — у тоннах. Для забезпечення сумісності ряди динаміки необхідно попередньо перерахувати в одні і ті ж одиниці вимірювання. Із введенням національної валюти макроекономічні показники 1996 р. і в наступні роки стали незрівнянні з минулими роками. У зв'язку з цим знадобилася велика робота, щоб усі макроекономічні показники 1991–1995 рр., виражених у карбованцях, перерахувати на нову національну валюту.

Неспівставність статистичних даних у часі виникає також відмінностями в ступені охоплення явища спостереженням. Особливо це характерно для тих випадків, коли для отримання вихідних даних організується несуцільне спостереження за методом основного масиву. В цьому випадку розширення чи звуження колі одиниць, охоплених спостереженням, часто може породжувати несумісність результатів.

Статистичні дані можуть бути непорівнянні і через застосування різної методології підрахунку статистичних показників. Так, показники амортизаційних відрахувань на промисловому підприємстві до 2020 р. обчислювалися на основі прямолінійного методу розрахунку. Проте, починаючи з 2021 р., методологію розрахунку амортизаційних відрахувань було змінено. Цим порушилася порівнянність загального розміру амортизаційних відрахувань, а також середніх величин амортизаційних відрахувань, що приходяться на 1 грн. товарної продукції. Щоб її відновити, потрібно дані за минулі роки перерахувати.

Наведені приклади показують, що порівнянність рівнів ряду — основна передумова науково обґрунтованої побудови та дослідження динамічних рядів, їх використання у практичній роботі. Статистичні дані, представлені в рядах динаміки, повинні бути співставні по території, колу об'єктів, що охоплюються, одиницям вимірювання, моменту реєстрації, методиці розрахунку. Порівняння непорівнянних статистичних даних призводить до спотворення дійсних співвідношень між явищами. Тільки ряди, рівні яких можна порівняти, можуть бути базою для наукового дослідження. Слід пам'ятати, що завдання побудови рядів динаміки, що відображають зміну явищ та процесів, може бути успішно вирішено лише при виконанні добре розроблених наукових правил. Наукова побудова динамічних рядів є необхідною передумовою їхнього наукового аналізу.

Які ж основні правила формування динамічних рядів?

**Періодизація динаміки**. Динамічні ряди відображають розвиток масових суспільних явищ за тривалий період часу. Внаслідок безперервного розвитку відбуваються постійні кількісні зміни явищ, а на визначених ступенях — і якісні зміни, які призводять до змін закономірності явища. Тому причиною наукового аналізу рядів динаміки є необхідність розчленування великих періодів часу на такі, які об'єднували лише одноякісні періоди. Такі періоди, у яких відбуваються якісні стрибки, характеризуються одним законом (закономірністю) розвитку явищ.

Процес виділення однорідних етапів розвитку, розчленування динамічних рядів на однорідні етапи зветься *періодизація динаміки*. Це свого роду типове групування рівнів ряду у часі. Виділення в рядах однорідних періодів не тільки забезпечує дослідника важливою інформацією про досліджуване явище, але і закладає основи для подальшого аналізу динаміки, так як справді наукову характеристику динамічних процесів можна дати лише в рамках однорідних періодів.

Необхідність формувати рівні динамічного ряду по однорідним за характером періодах або етапах не означає заперечення можливості побудови та вивчення динамічних рядів за тривалий часовий інтервал, що включає різні етапи розвитку явища. Відомо, що й саме поняття однорідності періодів досить відносно, оскільки воно залежить від рівня абстракції, прийнятої в дослідженні.

Відправною точкою в періодизації рядів динаміки є теоретичний аналіз внутрішніх причин і зовнішніх умов розвитку явища.

**Одноякісність окремих рівнів динамічного ряду**. При побудові динамічного ряду потрібно прагнути того, щоб рівні ряду характеризували явище однієї якості. З метою забезпечення одноякісності показників динамічного ряду попередньо необхідно розчленувати рівні ряду на певні групи. Для цього попередньо має бути проведене типологічне або структурне групування. Після виділення однорідних груп або типів явищ можуть бути утворені рівні динамічного ряду. Наприклад, для характеристики грошових доходів населення не можна користуватися динамічним рядом середнього доходу одного жителя, так як між різними верствами і групами населення існують великі відмінності (наприклад, серед жителів є група мільярдерів, мільйонерів, великих чиновників, які отримують величезні доходи; у той же час значна частина населення отримує доходи на рівні або нижче прожиткового мінімуму). Висока диференціація доходів населення може стати причиною ілюзорності збільшення доходів, коли їх фактично їх зростання спостерігатиметься лише у окремій групі, а в усіх інших — незмінність або навіть зменшення.

Тому правильну характеристику середніх доходів можна дати лише за однорідними класами та групами населення.

**Послідовність та безперервність у часі рівнів динамічного ряду**. При побудові ряду динаміки необхідно дотримуватися послідовності і безперервності ряду в часі. Іншими словами, рівні ряду динаміки повинні послідовно охоплювати період розвитку явища від початку до кінця. Відсутність даних за певний період часу (або на конкретну дату) може спотворити уявлення про динаміку при наступному аналізі. У разі «розриву» ряду динаміки, тобто відсутності даних за ті чи інші проміжки часу часто вдаються до приблизного розрахунку цих показників методами інтерполяції та екстраполяції.

Виникає питання, якої величини має бути період чи інтервал динамічного ряду?

Вирішення цього питання залежить від змісту досліджуваного явища. Величина інтервалу в принципі повинна бути відповідати тривалості циклів в досліджуваних процесах: якщо виробництво машини триває протягом місяця, немає потреби вимагати звітів про вироблену продукцію по п’ятиденках; основні показники роботи сільськогосподарських підприємств даються в річних інтервалах, пов’язаних з періодом вирощування сільськогосподарських культур.

Величина інтервалів залежить від ступеня мінливості (коливання) ознак досліджуваного явища. Чим більше змінюються показники, які характеризують розвиток явища в часі, тим меншу величину інтервалу (або відстань між моментами) потрібно брати. І навпаки, чим повільніше змінюються ряди величин, що характеризують зміну будь-якого явища, тим ширше потрібно взяти інтервал.

Малі інтервали можуть ускладнити обробку рядів, великі — призвести до того, що відбудеться усереднення відхилень і не буде виявлено дійсну тенденцію зміни.

**Застосування єдиної методології підрахунку статистичних показників**. Істотним правилом побудови динамічних рядів виступає застосування єдиної методології підрахунку статистичних показників. Статистична методологія є основою для проведення статистичних спостережень та розрахунку статистичних показників. Вона базується на результатах наукових досліджень, міжнародних рекомендаціях та досвіду практики з урахуванням соціально-економічних особливостей країни.

На відміну від ознак, що фіксуються у процесі спостереження, статистичні показники виходять розрахунковим шляхом. Методологія розрахунку статистичних показників не є незмінною. Зі зміною дійсності змінюються і системи статистичних показників, удосконалюється методологія їх розрахунку.

Методологія статистичних показників має відображати ті зміни, які відбуваються у досліджуваних явищах. Наприклад, в Україні на початку 1990-х років розрахунок макроекономічних показників здійснювався на основі балансу народного господарства. Розвиток ринкових відносин спричинило перебудову всієї системи макроекономічних показників, перехід на систему національних рахунків (СНР). До особливостей системи балансу народного господарства можна віднести, по-перше, цільову орієнтацію на облік та аналіз матеріального продукту і широке використання натуральних показників, по-друге, нерозвиненість блоку фінансового обороту, динаміки цін та рівня життя. Національне рахівництво як система обліку економічних взаємозв’язків відрізняється від балансу народного господарства насамперед ступенем охоплення процесу суспільного виробництва. У СНР враховуються всі види діяльності, пов’язані з виробництвом матеріальних товарів та послуг. Для оцінки обсягів виробництва в СНР обчислюється валовий внутрішній продукт, тоді як у балансі народного господарства — національний дохід. Істотно відрізняються ці показники за методикою обчислення. Відповідно пряме зіставлення цих показників зважаючи на відмінності в їх побудові неможливо.

Статистичні показники можуть бути використані для аналізу та висновків тільки тоді, коли вони засновані на єдиній методології розрахунку. Тільки за повної впевненості у сумісності показників низки динаміки можна приступати до його аналізу.

Аналіз рядів розподілу та рядів динаміки починається в обчисленні різних узагальнюючих показників, таких, як середні величини, відносні величини, показники коливання та ін. Однак у рядах динаміки середні величини зазвичай вже характеризують динаміку явища, а відносні величини виражають темп зміни, а не стан, як у рядах розподілу.

### 10.3. Графічне зображення рядів динаміки

Для наочного зображення у часі явищ та процесів часто використовують графічний метод. За допомогою графічного методу швидко, логічно і візуально можна зробити загальне судження про розвиток досліджуваного явища.

У рядах динаміки використовуються для наочного зображення явища багато діаграм: стовпчикові, стрічкові, кругові, радіальні, квадратні та інші. Вибір виду діаграми залежить переважно від особливостей вихідних даних, від цілей дослідження.

Наприклад, якщо є ряд динаміки з декількома нерівновіддаленими рівнями часу (1990, 2000, 2005, 2010, 2010, 2015, 2020), то найбільш поширеною є стовпчикова діаграма.

На рис. 10.1 зображено діаграму виробництва електроенергії в Україні. На підставі даної діаграми чітко видно тенденцію зміни виробництва електроенергії в країні за останні 30 років. Стовпчикові діаграми добре запам'ятовуються. Але вони дуже громіздкі для зображення великої кількості рівнів ряду динаміки.

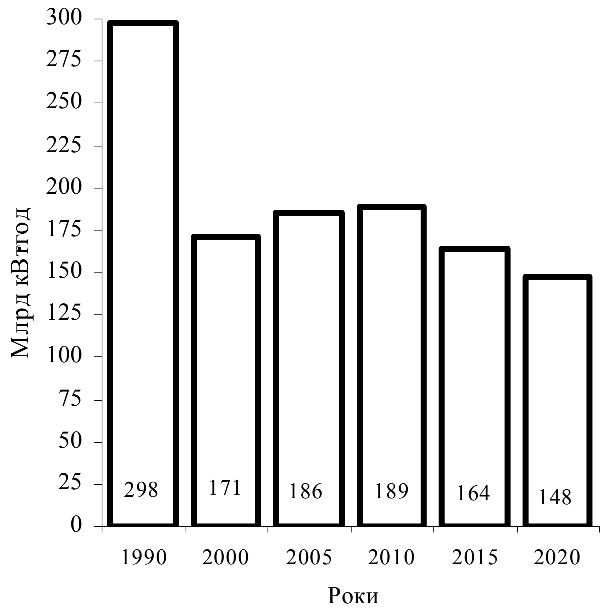



*Рис. 10. 1.* Динаміка виробництва електроенергії в Україні за 1990–2020 рр.

Лінійні діаграми зручно використовувати, коли метою дослідження є зображення загальної тенденції та характеру розвитку явищ, коли необхідно зобразити кілька рівнів ряду динаміки.

Для побудови лінійних діаграм використовують систему прямокутних координат. Зазвичай по осі абсцис відкладається час, а по осі ординат — міри зображуваних явищ або процесів. У статистичній практиці найчастіше зображають лінійні діаграми з рівними періодами часу. Разом з тим, в окремих випадках застосовуються і діаграми з нерівними інтервалами.

Зобразимо графічно динаміку посівних площ зернових культур в Україні за 2010–2020 роки. (табл. 10.8).

Таблиця 10.8

**Динаміка посівних площ зернових культур
в Україні за 2010–2020 рр.**

| Роки | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Тис. га | 15 090 | 15 724 | 15 449 | 16 210 | 14 801 | 14 739 | 14 401 | 14 624 | 14 839 | 15 318 | 15 392 |

Зображення посівних площ зернових культур в Україні за координатною сіткою з нерозривною шкалою, що починаються з нуля, недоцільно, оскільки 2/3 поля діаграми залишиться невикористаним і нічого не дасть для виразного зображення. Тому в даних умовах рекомендується будувати шкалу з розривом недалеко від нульової лінії. Це не призводить до спотворень у зображенні динаміки посівних площ зернових культур в Україні, і процес зміни цього показника малюється діаграмою більш чітко (рис. 10.2).

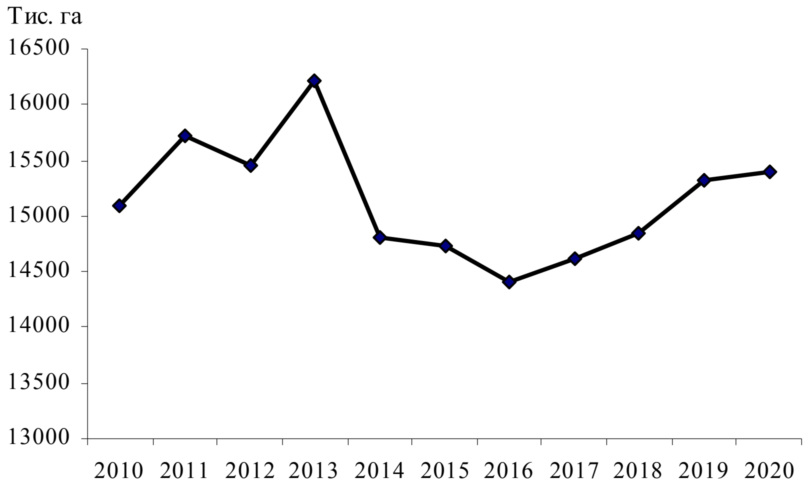



*Рис. 10.2.* Динаміка посівних площ зернових культур
в Україні за 2010–2020 рр.

Нерідко на одному полі графіка доводиться зображати кілька кривих ліній, які дають порівняльну характеристику динаміці різних показників або одного і того ж показника але в різних країнах.

Зобразимо динаміку товарообігу підприємств оптової торгівлі в Україні за 2013–2020 рр. (табл. 10.9).

Таблиця 10.9

**Динаміка оптового товарообігу підприємств оптової торгівлі (у % до 2013 р.)**

| Рік | весь товарообіг | продовольчі товари | непродовольчі товари |
|---|---|---|---|
| 2013 | 100,0 | 100,0 | 100,0 |
| 2014 | 91,9 | 97,5 | 90,7 |
| 2015 | 125,9 | 115,2 | 128,6 |
| 2016 | 157,5 | 129,8 | 164,3 |
| 2017 | 193,2 | 161,3 | 201,0 |
| 2018 | 224,2 | 182,9 | 234,4 |
| 2019 | 235,0 | 201,4 | 243,3 |
| 2020 | 249,3 | 224,8 | 255,3 |

При цьому обсяг товарообігу 2013 р. приймемо за 100 %. Приклад такого зіставлення наведено в рис. 10.3.

Іноді необхідно порівняти на графіку динаміку двох показників, що мають різні одиниці вимірювання. У таких випадках знадобиться не одна, а дві масштабні шкали. Одну їх розміщують з права, іншу — зліва (рис. 10.4).

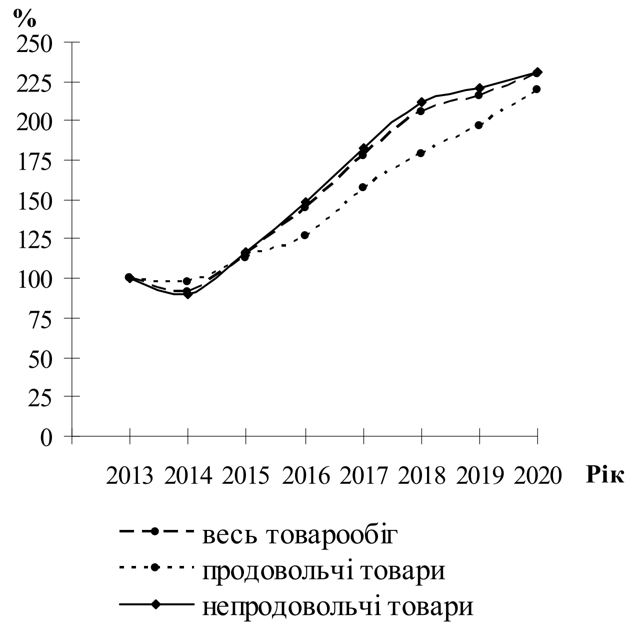



*Рис. 10.3.* Динаміка оптового товарообороту підприємств оптової торгівлі України за 2013–2020 рр.


*Рис. 10.4.* Динаміка виробництва м’яса курей і натуральної мінеральної негазованої води в Україні за 2012–2020 рр.

Однак таке порівняння кривих не дає достатньо повної картини динаміки цих показників, так як масштаби довільні. Тому, якщо потрібно порівняти динаміку двох різнорідних показників, краще здійснювати на основі використання одного масштабу після перетворення

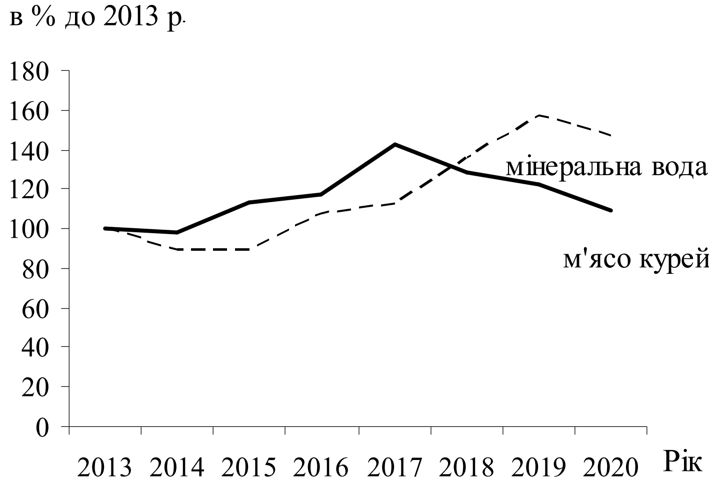



*Рис. 10.5.* Динаміка виробництва м’яса курей і мінеральної води в Україні за 2013–2020 рр.

абсолютних величин у відносні. Так, щодо виробництва м’яса курей та води після перерахунку абсолютних величин у відносні, де 2013 р. взято за 100 %, отримаємо інший графік (див. рис. 10.5).

У деяких випадках нанесення на один графік двох кривих дає можливість одночасно зобразити динаміку третього показника, якщо він є різницею або сумою перших двох. Наприклад, при зображенні динаміки експорту та імпорту товарів (рис. 10.6) відстань між кривими показує величину торговельного балансу (чистого експорту) зовнішньоекономічної діяльності.

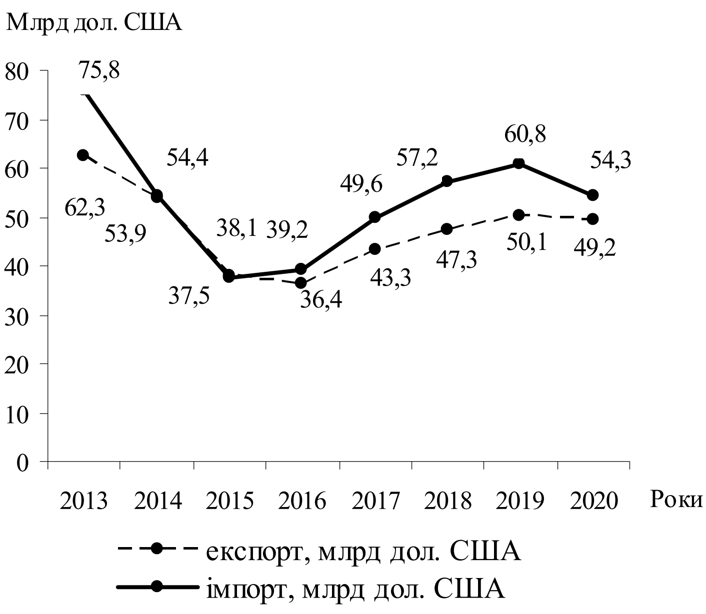



*Рис. 10.6.* Динаміка експорту і імпорту товарів в Україні за 2013–2020 рр. (млрд дол. США).

## 10.4. Статистичні характеристики (показники) ряду динаміки

**Початковий, кінцевий та середній рівні динаміки**. Процесу обробки піддаються всі рівні ряду динаміки. Розрізняють початковий рівень ($y_1$), що показує величину першого члена ряду, кінцевий рівень ($y_n$), що показує величину останнього члена ряду, та середній рівень ряду ($\bar{y}$), який розраховується за середньою хронологічною.

Середньою хронологічною називається середня, обчислена за сукупністю значень показників, що змінюються у часі (в різні моменти чи періоди часу). Такі середні узагальнюють коливання ознак ряду у часі. У хронологічній середній відображається сукупність тих умов, у яких розвинулося досліджуване явище в даному проміжку часу.

При дослідженні динаміки явищ у галузі статистики використовуються деякі статистичні характеристики, які дозволяють виміряти зміну явищ у часі.

Більшість статистичних характеристик засноване на абсолютному або відносному порівнянні рівнів динамічних рядів показників динаміки: абсолютний приріст, темпи зростання і приросту, абсолютне значення одного відсотка приросту. Порівнюваний рівень називається *поточним*, а рівень, з яким порівнюється — *базисним*. За базисний рівень часто приймається чи попередній рівень, або початковий у цьому динамічному ряду.

Для аналізу рядів динаміки обчислюють наступні показники: темпи зростання ($T_p$), абсолютні прирости ($\Delta y$) і відносні прирости, які називають темпами приросту ($T_{пр}$). Обчислюється абсолютна та відносна величина одного відсотка ($A$).

**Абсолютний приріст**. *Абсолютним приростом* називається різниця двох рівнів ряду динаміки. Якщо проводиться порівняння кожного даного рівня ($y_i$) з безпосередньо йому попереднім ($y_{i-1}$), то отримують

*ланцюгові* абсолютні прирости ($\Delta y$):

$$\Delta y = y_i - y_{i-1}. \tag{10.1}$$

Якщо кожен рівень ($y_i$) динамічного ряду порівнюється з початковим ($y_1$) або будь-яким іншим, прийнятим за постійну базу порівняння, то отримуються

*базисні* абсолютні прирости ($\Delta y$):

$$\Delta y = y_i - y_1 . \tag{10.2}$$

Сума ланцюгових приростів за певний період часу дорівнює базисному абсолютному приросту за весь період. Вона дорівнює також різниці між кінцевим і початковим рівнем ряду динаміки:

$$\sum_{i=1}^{n-1} \Delta_i = \sum_{i=1}^{n-1} (y_i - y_{i-1}) = y_n - y_1. \tag{10.3}$$

Абсолютний приріст може мати позитивний або від'ємний знак. Якщо показники рівня ряду, що характеризують явища, не зростають, а зменшуються, абсолютну різницю рівнів показують зі знаком мінус. І тут абсолютна різниця характеризує не абсолютний приріст, а абсолютне зменшення обсягу явища. Таким чином, різниця показує, *наскільки* даний рівень перевищує рівень, взятий для порівняння. Отже, абсолютний приріст характеризує швидкість (в абсолютному вираженні) виміру рівнів ряду динаміки в одиницю часу.

Для ілюстрації розрахунків абсолютного приросту табл. 10.10 наведено ряд динаміки, що характеризує обсяг виробництва продукції підприємства. Абсолютні рівні цього показника за 2015–2020 рр. зросли, проте зростання мало нестійкий характер. Абсолютний приріст також мав загальну тенденцію до зростання.

Таблиця 10.10

**Обсяг виробництва продукції за 2016–2020 рр., тис. т**

| Рік | Виробництво продукції, тис. т | Річний абсолютний приріст, тис. | |
|---|---|---|---|
| | | базисний | ланцюговий |
| 2015 | 116,8 | – | – |
| 2016 | 120,6 | 3,8 | 3,8 |
| 2017 | 125,7 | 8,9 | 5,1 |
| 2018 | 132,4 | 15,6 | 6,7 |
| 2019 | 136,2 | 19,4 | 3,8 |
| 2020 | 142,3 | 25,5 | 6,1 |

Техніка обчислення наведених у табл. 10.10 показників наступна.

Абсолютний приріст ($\Delta y$):

| *базисний* | *ланцюговий* |
|---|---|
| $\Delta y_{2016}$ =120,6–116,8=+3,8 (тис. т); | $\Delta y_{2016}$ =120,6–116,8=+3,8 (тис. т); |
| $\Delta y_{2017}$ =125,7–116,8=+8,9 (тис. т); | $\Delta y_{2017}$ =125,7–116,8 = +8,9 (тис. т); |
| $\Delta y_{2018}$ =132,4–116,8=+15,6 (тис. т); | $\Delta y_{2018}$ =132,4–125,7=+6,7 (тис. т); |
| $\Delta y_{2019}$ =136,2–116,8=+19,4 (тис. т); | $\Delta y_{2019}$ =136,2–132,4=+3,8 (тис. т); |
| $\Delta y_{2020}$ =142,3–116,8=+25,5 (тис. т); | $\Delta y_{2020}$ =142,3–136,2=+6,1 (тис. т). |

Сума ланцюгових приростів за 2015–2020 рр. дорівнює базисному приросту за весь період:

3,8+5,1+6,7+3,8+6,1=25,5 (тис. т) або (142,3–116,8)=25,5 (тис. т).

**Темпи зростання**. *Темп зростання* — відносний показник динаміки, представляє собою відношення одного рівня ряду динаміки до іншого його рівня, прийнятого за базу порівняння. За допомогою темпів зростання вимірюється, у скільки разів рівень поточного періоду вище або рівня базисного періоду, або на скільки відсотків (яку частку) даний рівень становить по відношенню до іншого рівня, прийнятого за базу порівняння. У ряді динаміки можуть бути обчислені як базисні, коли всі рівні ряду відносяться до одного рівня одного якогось періоду, прийнятого за базу, і як ланцюгові, коли рівень кожного періоду відноситься до рівня попереднього періоду. У тому й іншому випадку темп зростання може бути виражений у вигляді коефіцієнтів, коли визначається безпосередньо відношення абсолютних розмірів рівнів, і у відсотках, коли він показує, скільки відсотків поточний рівень становить по відношенню до базисного, прийнятого за 100 %

Темп зростання як коефіцієнтів обчислюється за формулою:

$$T_p = \frac{y_i}{y_{y-1}} \text{ — ланцюгові темпи зростання;} \quad (10.4)$$

$$T_p = \frac{y_i}{y_1} \text{ — базисні темпи зростання;} \quad (10.5)$$

$$T_p = \frac{y_n}{y_1} \text{ — темп зростання за весь період,} \quad (10.6)$$

де $T_p$ — темп зростання;

$y_i$ — рівень ряду динаміки;

$y_{i-1}$ — рівень ряду, що безпосередньо передує рівню $y_i$.

$y_1$ — рівень ряду, прийнятий за базу.

Для вираження темпу зростання у відсотках досить його величину, виражену у вигляді коефіцієнта, помножити на 100. Тому можливі наступні варіанти обчислення темпів зростання (табл. 10.11).

На основі даних табл. 10.11 розрахуємо темпи зростання. Базисні темпи зростання обчислюються шляхом ділення рівнів ряду кожного року на рівень ряду якогось року, прийнятого за базу. Зазвичай, це початковий рівень (у даному випадку — це 2015 р.).

Для розрахунку базисних темпів зростання виробництва продук-

Таблиця 10.11

**Темпи зростання виробництва продукції у 2015-2020 рр.**

| | 2015 р. | 2016 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|---|---|
| Виробництво продукції ($y$ — рівні ряду динаміки) тис. т | 116,8 | 120,6 | 125,7 | 132,4 | 136,2 | 142,3 |
| Темпи базисні ($T_б = y_i : y_1$): | | | | | | |
| коефіцієнти | 1,0 | 1,033 | 1,076 | 1,134 | 1,166 | 1,218 |
| проценти | 100,0 | 103,3 | 107,6 | 113,4 | 116,6 | 121,8 |
| Темпи ланцюгові ($T_б = y_i : y_{i-1}$): | | | | | | |
| коефіцієнти | 1,0 | 1,033 | 1,042 | 1,053 | 1,029 | 1,045 |
| проценти | 100,0 | 103,3 | 104,2 | 105,3 | 102,9 | 104,5 |

ції необхідно обсяг виробництва кожного наступного року розділити на обсяг виробництва 2015 р. та частку помножити на 100:

2016 $\frac{120{,}6}{116{,}8} \cdot 100 = 103{,}3$ %;

2017 $\frac{125{,}7}{116{,}8} \cdot 100 = 107{,}6$ %;

2018 $\frac{132{,}4}{116{,}8} \cdot 100 = 113{,}4$ %;

2019 $\frac{136{,}2}{116{,}8} \cdot 100 = 116{,}6$ %;

2020 $\frac{142{,}3}{116{,}8} \cdot 100 = 121{,}8$ %.

Базисні темпи характеризують безперервну лінію розвитку. По них для будь-якого року можна відповісти на питання, як зросло виробництво в кожному році в порівнянні з 2015 р., прийнятим за базу порівняння. При цьому, якщо користуються коефіцієнтом, кажуть, що обсяги виробництва у 2020 р. зросли порівняно з 2015 р. у 1,218 разів. Якщо ж користуються темпами, вираженими у відсотках, кажуть, що виробництво 2020 р. становило 121,8 % проти 2015 р.

Якщо необхідно з’ясувати, якими темпами зростало виробництво продукції в порівнянні з попередніми роками, то розрахунок проводиться за ланцюговому індексу, тобто обсяг виробництва кожного наступного року слід розділити на обсяг виробництва попереднього року і частку помножити на 100:

2016 $\frac{120,6}{116,8}\cdot 100 = 103,3\ \%;$

2017 $\frac{125,7}{120,6}\cdot 100 = 104,2\ \%;$

2018 $\frac{132,4}{125,7}\cdot 100 = 105,3\ \%;$

2019 $\frac{136,2}{132,4}\cdot 100 = 102,9\ \%;$

2020 $\frac{142,3}{136,2}\cdot 100 = 104,5\ \%.$

Ланцюгові темпи показують інтенсивність розвитку явища в кожному окремому періоді.

Показники темпів зростання полегшують аналіз, показують напрямок розвитку. Величина темпу зростання, більша 1 чи 100 %, показує збільшення рівня поточного періоду проти базисного. Величина темпу зростання, що дорівнює 1 або 100 %, показує, що рівень поточного періоду в порівнянні з базисним не змінився, а величина темпу зростання, менша 1 або 100 % показує зменшення рівня поточного періоду, але темп зростання завжди має позитивний знак .

Добуток ланцюгових темпів дорівнює кінцевому базисному ($T_p^б$):

$$T_p^б = \frac{y_2}{y_1}\cdot\frac{y_3}{y_2}\cdot\frac{y_4}{y_3}\cdots\frac{y_n}{y_{n-1}} = \frac{y_n}{y_1}. \qquad (10.7)$$

Якщо темпи виражені в коефіцієнтах, то легко перейти від ланцюгових темпів до базисних і назад, користуючись такими правилами:

1) добуток ланцюгових темпів дорівнює базисному;

2) частка від поділу базисних темпів зростання дорівнює проміжному ланцюговому.

Так, перемноживши ланцюгові темпи зростання 2016 та 2017 рр., отримаємо базисний темп 2017 р.: $1,033\cdot 1,042=1,076$ і, навпаки, розділивши базовий темп зростання 2020 р. (1,218) на базисний темп 2016 р. отримаємо проміжний ланцюговий темп 2020 р. (1,045).

При обчисленні базисних темпів зростання виникає питання, який рік взяти за основу порівняння. Це питання вирішують залежно від задач дослідження та особливостей розвитку досліджуваного явища.

Показники темпів зростання дуже часто використовуються для аналізу динаміки виробництва продукції, обсягів продажів, чисельності персоналу підприємства і т. п.

Користуючись відносними показниками для характеристики інтенсивності розвитку, ніколи не можна забувати про абсолютні рівні розвитку, які приховані за темпами. За одними й тими ж самими темпами зростання може бути прихована різна абсолютна величина показника. Зростання обсягу виробництва, скажімо, в 1,5 рази для тієї галузі промисловості, яка знаходиться ще на початковій стадії розвитку, матиме невеликий обсяг виробництва. Зростання виробництва, скажімо, на 5 % для високорозвиненої галузі промисловості матиме великі абсолютні обсяги виробництва. Тому темпи зазвичай доповнюються даними про абсолютні і відносні прирости.

**Темпи приросту**. Відносну оцінку значення абсолютного приросту в порівнянні з початковим рівнем дають температури приросту. *Темпи приросту* — відношення абсолютного приросту показника ряду динаміки до рівня показника, прийнятого за базу порівняння. Найчастіше темп приросту розраховується до попереднього рівня і виражається у відсотках, рідше — в коефіцієнтах. Темп приросту, як і абсолютний приріст, то, можливо і позитивним, і від’ємним.

Темп приросту в залежності від того, в яких одиницях виражений темп зростання (у вигляді коефіцієнта або у відсотках) може бути розрахований за такими формулами:

в процентах

$$T_{пр} = \frac{y_i - y_{i-1}}{y_{i-1}} \times 100, \tag{10.8}$$

або

$$T_{пр} = \frac{\Delta_i}{y_{i-1}} \times 100, \tag{10.9}$$

або

$$T_{пр} = \frac{y_i}{y_{i-1}} \times 100 - 100, \tag{10.10}$$

або

$$T_{пр} = (T_p - 1) \times 100, \tag{10.11}$$

у вигляді коефіцієнта

$$T_{пр} = \frac{y_i}{y_{i-1}} - 1 = \frac{y_i - y_i}{y_{i-1}} = \frac{\Delta_i}{y_{i-1}}, \qquad (10.12)$$

де $y_i$ — рівень ряду динаміки;

$y_{i-1}$ — рівень ряду, що безпосередньо передує рівню $y_i$.

Темп приросту за весь період, що охоплюється рядом динаміки, розраховується за формулою:

$$T_{пр} = \frac{y_n - y_1}{y_1} \times 100\,\%, \qquad (10.13)$$

де $y_n$ — кінцевий член ряду динаміки;

$y_1$ — початковий його член.

Темп приросту показує, наскільки процентів збільшилися розміри явища за досліджуваний період часу. Якщо рівень явища не збільшується, а скорочується, темпи приросту будуть зі знаком мінус. Вони характеризуватимуть не відносний приріст, а відносне зменшення рівня явищ.

На основі даних табл. 10.11 розрахуємо темпи приросту:

б а з и с н і

$$\Delta T_{p\ 2016} = \frac{120{,}6 - 116{,}8}{116{,}8} \times 100 = 3{,}8\ \%;$$

$$\Delta T_{p\ 2017} = \frac{125{,}7 - 116{,}8}{116{,}8} \times 100 = 8{,}9\ \%;$$

$$\Delta T_{p\ 2018} = \frac{132{,}4 - 116{,}8}{116{,}8} \times 100 = 15{,}6\ \%;$$

$$\Delta T_{p\ 2019} = \frac{136{,}2 - 116{,}8}{116{,}8} \times 100 = 19{,}4\ \%;$$

$$\Delta T_{p\ 2020} = \frac{142{,}3 - 116{,}8}{116{,}8} \times 100 = 25{,}5\ \%;$$

л а н ц ю г о в і

$$\Delta T_{p\ 2016} = \frac{120{,}6 - 116{,}8}{116{,}8} \times 100 = 3{,}8\ \%;$$

$$\Delta T_{p\ 2017} = \frac{125{,}7 - 120{,}6}{120{,}6} \times 100 = 5{,}1\ \%;$$

$$\Delta T_{p\ 2018} = \frac{132{,}4 - 125{,}7}{125{,}7} \times 100 = 6{,}7\ \%;$$

$$\Delta T_{p\ 2019} = \frac{136{,}2 - 132{,}4}{132{,}4} \times 100 = 3{,}8\ \%;$$

$$\Delta T_{p\ 2020} = \frac{142{,}3 - 136{,}2}{136{,}2} \times 100 = 6{,}1\ \%.$$

*Абсолютне значення одного відсотка приросту* — це відношення абсолютного приросту до темпу приросту, вираженого у процентах. Показує, яка абсолютна величина прихована за відносною — одним відсотком приросту. Абсолютне значення одного відсотка приросту розраховується за формулою:

$$|\ A\ | = \frac{\Delta y}{T_{np}\ \%} = \frac{y_i - y_{i-1}}{\dfrac{y_i - y_{i-1}}{y_{y-1}} \cdot 100} = \frac{y_{i-1}}{100} = 0{,}01\ y_{i-1}, \qquad (10.14)$$

де $A$ — абсолютне значення одного процента приросту;

$y_i$ — рівні ряду динаміки ($i$=1, 2, 3 …, …, $n$);

$n$ — число рівнів;

$T_{np}$ — темп приросту;

$\Delta y$ — абсолютний приріст.

Кількісно абсолютне значення одного відсотка дорівнює 0,01 попереднього рівня (або іншого рівня, прийнятого за основу порівняння).

Потрібно підкреслити, що розрахунок абсолютного значення 1 % приросту має значення тільки для ланцюгових приростів та темпів приросту. Для базисних темпів приросту значення цього показника за період буде однаковим, оскільки цей показник для всіх років буде один і той же самий, оскільки рівень ряду, безпосередньо попередній рівню $y_i$, залишається незмінним. Наведені дані говорять про те, що всі показники динаміки (рівні, темпи розвитку, абсолютні і відносні прирости) потрібно аналізувати комплексно, спільно.

У статистичній практиці нерідко використовуються ще деякі похідні показники динамічних рядів, наприклад, коефіцієнти випередження.

*Коефіцієнт випередження* ($K_{\text{вип}}$) обчислюється як відношення темпів зростання за однакові відрізки часу за двома динамічними рядами:

$$K_{\text{вип}} = \left(\frac{y_i'}{y_0'}\right) : \left(\frac{y_i''}{y_0''}\right), \qquad (10.15)$$

де $y_i'$ і $y_i''$ — рівні двох динамічних рядів.

$y_0'$ і $y_0''$ — базисні рівні двох динамічних рядів.

В якості рівнів динамічних рядів можуть виступати продуктивність і оплата праці, чисельність робітників за видами економічної діяльності, обсяг виробництва продукції окремими підприємствами і т. п. Так, за 2016–2020 рр. темп зростання реалізованої промислової продукції підприємств переробної промисловості склав 103,5%, а продукції машинобудування — 111,0%. Отже, коефіцієнт випередження зростання реалізованої продукції машинобудування порівняно зі зростанням реалізованої продукції переробної промисловості за 2016–2020 рр. складе:

$$K_{\substack{\text{вип. за} \\ 2016– \\ 2020\,\text{рр.}}} = \frac{\text{темп зростання продукції машинобудування}}{\text{темп зростання продукції переробної промисловості}} =$$

$$= \frac{1{,}110}{1{,}035} = 1{,}07.$$

Характерною особливістю динаміки обсягу реалізованої промислової продукції було випереджальний розвиток машинобудування в порівнянні з усією переробною промисловістю.

Порівнювані при обчисленні коефіцієнта випередження динамічні ряди можуть відноситься до двох просторових об’єктів (країн, областей, районів, видів економічної діяльності, підприємств і т. п.), різних видів продукції, різних показників тощо.

## 10.5. Середні показники динаміки

Як ми бачили раніше, статистичні характеристики динаміки соціально-економічних явищ, розраховані за рівнями ряду, змінюються у часі. Вони варіюються за окремими періодами, що викликає необхідність розрахунку узагальнюючих показників рядів динаміки. Ряди динаміки можуть бути складені як на основі абсолютних показників, так і на основі середніх та відносних показників. Найважливішими узагальнюючими показниками динамічного ряду виступають різного роду середні, що обчислюються як за основними рівнями, так і за похідними показниками ряду. Дуже часто в статистиці застосування середніх є просто необхідним. Наприклад, денний виробіток одного робітника підприємства істотно відрізняється за окремими днями місяця. Правильніше порівнювати середньомісячний (денний) виробіток на одного робітника, середньомісячні (середньорічні) темпи зростання за певні проміжки часу.

До середньорічних показників доводиться дуже часто прибігати і при неможливості зіставити абсолютні дані. Наприклад, щоб визначити продуктивність праці на підприємстві, необхідно обсяг продукції, виробленої за певний період на даному підприємстві, розділити на чисельність працівників, яка для даного проміжку часу не є постійною. Так, на початок місяця чисельність працівників складала 210 осіб, а на кінець — 220 чол. Щоб визначити продуктивність праці в цілому по підприємству за місяць, необхідно віднести загальний обсяг продукції до середньої чисельності персоналу, яке складе: $\frac{210+220}{2}=$ 215 осіб.

Головна мета методу середніх полягає в тому, щоб дати кількісну характеристику загальної закономірності розвитку явища або процесу. За допомогою обчислення середньої величини можна визначити типовий рівень розвитку явища з урахуванням міри різноманітних змін масових індивідуальних елементів, оскільки випадкові коливання позначаються на середній менше, ніж на окремих елементах.

Середні, обчислені за суміжними рівнями ряду, називають *динамічними* або *хронологічними*. Такі середні узагальнюють зміну величини варійованої ознаки якісно однорідної сукупності. У хронологічній середній відображається рівень досліджуваного явища в даному проміжку часу.

Розглянемо розрахунок середніх абсолютних рівнів, середніх абсолютних приростів, середніх темпів зростання та приросту.

**Розрахунок середнього абсолютного рівня інтервальному ряду динаміки**. Середній рівень ряду розраховується по-різному для моментних та інтервальних рядів динаміки.

*Середній рівень інтервального ряду* принципово можна розрахувати за такою формулою:

$$\bar{y}=\frac{\sum y}{\sum t}, \qquad (10.16)$$

де $y$ — рівні;

$t$ — тривалість окремих інтервалів часу.

Щоб знайти *середній рівень інтервального ряду*, в якому всі інтервали рівні, достатньо суму рівнів цього середнього ряду поділити на число періодів, до якого відноситься, тобто

$$\bar{y}=\frac{y_1+y_2+y_3+\ldots+y_n}{n}=\frac{\sum y}{n}, \qquad (10.17)$$

де $\sum y$ — сума рівнів ряду;

$n$ — число рівних проміжків чи інтервалів.

Таким чином, середня хронологічна інтервального ряду динаміки обчислюється за формулою середньої арифметичної простої (незваженої).

Обчислимо середній рівень виробництва шоколадних цукерок в Україні за 2015–2020 рр. за даними табл. 10.12.

Таблиця 10.12

| Показник | 2015 р. | 2016 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|---|---|
| Виробництво шоколадних цукерок, тис. т | 59,5 | 52,3 | 53,3 | 53,4 | 66,6 | 70,7 |
| Абсолютний приріст, тис. т | … | –7,2 | +1,0 | +0,1 | +13,2 | +4,1 |
| Темп зростання (коефіцієнт) | 1,00 | 0,879 | 1,019 | 1,002 | 1,247 | 1,062 |

За цими даними зробимо розрахунок середньорічного рівня виробництва шоколадних цукерок за 2016–2020 рр.:

$$\bar{y} = \frac{\sum y}{n} = \frac{59{,}5 + 52{,}5 + 53{,}3 + 53{,}4 + 66{,}6 + 70{,}7}{6} =$$

$$= \frac{355{,}8}{6} = 59{,}3 \text{ тис. т.}$$

Склавши шість річних рівнів, отримаємо виробництво шоколадних цукерок за шість років у розмірі 355,8 тис. т, а розділивши цю величину на шість, дізнаємося, скільки шоколадних цукерок виробляли за 2015–2020 рр. щорічно у середньому. При розрахунку середнього рівня в інтервальному ряду динаміки необхідно вказувати період, який прийнятий як одиниця розрахунку (у нашому прикладі рік). Можна змінити одиницю часу, наприклад, місяць, і обчислити середній рівень ряду. І тому загальний результат виробництва 355,8 тис. т необхідно розділити не так на 6, але в 72 (число місяців за 6 років). Отримаємо:

$$\bar{y} = \frac{355{,}8}{72} = 4{,}94 \text{ тис. т.}$$

Ця величина показує, скільки в середньому за 2015–2020 рр. в середньому протягом місяця вироблялося шоколадних цукерок. Для розрахунку середнього рівня інтервального ряду необхідно здійснити підсумовування рівнів ряду. Це підсумовування означає укрупнення інтервалу іноді має самостійне значення для вивчення закономірності розвитку явищ. Наведемо приклад (табл. 10.13).

Таблиця 10.13

**Виробництво цукру у вересні**

| Число місяця | Обсяг виробництва, т | Число місяця | Обсяг виробництва, т | Число місяця | Обсяг виробництва, т | Число місяця | Обсяг виробництва, т | Число місяця | Обсяг виробництва, т |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 045 | 7 | 3 060 | 13 | 3 217 | 19 | 3 352 | 25 | 3 382 |
| 2 | 3 068 | 8 | 3 128 | 14 | 3 228 | 20 | 3 421 | 26 | 3 461 |
| 3 | 3 075 | 9 | 3 164 | 15 | 2 984 | 21 | 3 328 | 27 | 3 427 |
| 4 | 2 986 | 10 | 3 108 | 16 | 2 860 | 22 | 3 276 | 28 | 3 428 |
| 5 | 2 915 | 11 | 3 085 | 17 | 3 186 | 23 | 3 389 | 29 | 3 512 |
| 6 | 3 024 | 12 | 3 196 | 18 | 3 271 | 24 | 3 391 | 30 | 3 507 |

Тепер перейдемо від ряду динаміки щоденних даних до ряду даних по п'ятиденках. Для цього підсумовуємо щоденні дані про виробництво за кожну п'ятиденку:

| | П'ятиденки | | | | | | За місяць в цілому |
|---|---|---|---|---|---|---|---|
| | перша | друга | третя | четверта | п'ята | шоста | |
| Виробництво, т | 15 089 | 15 484 | 15 710 | 16 090 | 16 766 | 17 335 | 96 474 |

За допомогою укрупнення інтервалів в інтервальних рядах динаміки прагнуть чіткіше виявити тенденцію розвитку явища. У прикладі інтервал укрупнений, тобто проведено перехід від денного інтервалу до п'ятиденному. При укрупненні інтервалів тенденцію послідовного наростання показника видно чіткіше, ніж у ряді щоденних даних.

Перетворимо тепер інтервальний ряд по п'ятиденках в ряд середньодобових рівнів виробництва шляхом поділу загального обсягу виробництва за кожні п'ять днів на величину інтервалу (5). Середньодобове виробництво цукру в листопаді по п'ятиденках була рівна:

| | П'ятиденки | | | | | | За місяць в цілому |
|---|---|---|---|---|---|---|---|
| | перша | друга | третя | четверта | п'ята | шоста | |
| Виробництво, т | 3017,8 | 3096,8 | 3142 | 3218 | 3353,2 | 3467 | 3215,8 |

Середньодобове виробництво протягом місяця загалом отримано шляхом поділу загального обсягу виробництва протягом місяця число днів на місяці (30). Воно може бути отримано також шляхом підсумовування середньодобових даних по п'ятиденках та поділу отриманої суми на шість.

Якщо окремі періоди інтервального ряду динаміки, що мають неоднакову довжину, то для визначення середнього рівня середнього зваженого ряду слід скористатися середньою арифметичною зваженою, тобто розрахувати його, зважуючи рівні за кількістю рівних періодів.

Середні в інтервальному ряді динаміки теоретично можна обчислити за будь-який інтервал часу (п'ятиріччя, рік, квартал, місяць і т. д.). Однак практично розраховують такі середні, які необхідні для вирішення теоретичних і практичних завдань.

**Розрахунок середнього рівня моментного ряду динаміки**. Перейдемо тепер до обчислення середнього рівня в моментному ряді динаміки. Обчислити середній рівень моментного ряду значно складніше, рівні ряду відносяться тут не до певних проміжків часу, а до моментів його, не пов'язаних із тривалістю періоду. Середній рівень моментного ряду динаміки розраховувати на основі середньої арифметичної не можна. Неточний буде і розрахунок, при якому береться напівсума рівнів на початок і кінець всього періоду. При цьому в розрахунках середнього рівня моментного ряду можуть зустрітися принаймні три варіанти.

*Перший варіант*. У практиці економічної роботи частини доводиться визначати середні рівні низки моментних величин із нерівновіддаленими датами. Наприклад, є такі дані про рух робітників на заводі за квітень. На 1-е квітня було 65 осіб, 7 квітня було прийнято 9 осіб, 12 квітня звільнено 6 осіб, 22 квітня прийнято 4 особи. Ці дані можна представити у вигляді такого моментного ряду:

| | |
|---|---|
| 1/IV | — 65 осіб |
| 7/IV | — 74 » |
| 12/IV | — 68 » |
| 22/IV | — 72 » |

Необхідно визначити середньоденну облікову чисельність працівників заводу за квітень.

Для вирішення цього завдання чисельність працюючих на певні моменти часу зважуємо нерівновіддаленими періодами часу, припускаючи, що чисельність працюючих змінювалася тільки на ці моменти.

У цьому прикладі чисельності працівників тут «зважено» за тривалістю періодів часу, протягом яких вони працювали на підприємстві. З цією метою необхідно встановити, як змінювалася чисельність робітників протягом місяця. Протягом перших шести днів — з 1 по 6 квітня включно — було 65 осіб, у наступні п’ять днів — з 7 по 11 квітня — було 74 особи, з 12 по 21 квітня, тобто 10 днів, складалося 68 осіб і наступні дні — з 22 по 30 квітня — 72 осіб. Маючи такі дані про рух робітників протягом місяця, середню їх чисельності можна обчислити за формулою середньої арифметичної зваженої:

$$\bar{y} = \frac{\sum yt}{\sum t} = \frac{65 \cdot 6 + 74 \cdot 5 + 68 \cdot 10 + 72 \cdot 9}{30} = \frac{2\,088}{30} = 70 \text{ осіб.}$$

Розрахунок можна подати у вигляді наступної таблиці.

Таблиця 10.14

| Календарні періоди квітня | Число робітників $y$ | Довжина періоду, днів $t$ | Число людино-днів $ty$ |
|---|---|---|---|
| 1–6 | 65 | 6 | 390 |
| 7–11 | 74 | 5 | 370 |
| 12–21 | 68 | 10 | 680 |
| 22–30 | 72 | 9 | 648 |
| Всього | – | 30 | 2 088 |

Звідси середньоденна облікова чисельність працівників у квітні:

$$\bar{y} = \frac{\sum ty}{\sum t} = \frac{2\,088}{30} \approx 70 \text{ осіб.}$$

Слід мати на увазі, що чим менші буде інтервали часу між рівнями ряду, тим менше буде допущена помилка при обчисленні середнього рівня. У разі точної буде така середня, яка обчислена за рівнями ряду динаміки за кожний день. Однак такий розрахунок середньої можливий лише у випадку, коли ми маємо вичерпні дані про зміну досліджуваних явищ за певний період часу. У більшості випадків статистики не мають таких вичерпних даних про зміну явища, вони зазвичай мають тільки дані про стан явищ на певні дати, що не збігаються з часом зміни явищ. У таких випадках розрахунок середнього рівня моментного ряду доводиться робити з припущенням деяких умовностей.

*Другий варіант*. Розглянемо приклад розрахунку середньої для випадку, коли відомі лише початковий і кінцевий рівні ряду динаміки і невідомо, як змінювалося явище протягом цього часу. Відомо, що в Україні чисельність працівників, задіяних на виконанні наукових досліджень та розробок, на кінець 2019 р. склала 79,2 тис., а на кінець 2020 р. — 78,9 тис. Виникає питання, скільки працівників було задіяно на виконанні наукових досліджень у середньому в 2020 р.?

Нам не відомо, як змінювалася чисельність працівників протягом року. Тому ми змушені прийняти, що їхнє скорочення відбувалося протягом року рівномірно. Це передбачається так: 79,2 тис. працівників протягом першої половини 2020 року; з 1 липня їх кількість скоротилася на 0,3 тис. Отже, з другої половини року чисельність працівників становила 78,9 тис. людина. Користуючись попередньою формулою, запишемо:

$$\bar{y} = \frac{\sum yt}{\sum t} = \frac{79{,}2 \cdot \frac{1}{2} + 78{,}9 \cdot \frac{1}{2}}{\frac{1}{2} + \frac{1}{2}} = \frac{39{,}6 + 39{,}45}{1} = 79{,}05 \text{ тис. осіб.}$$

Як і першому варіанті, зважування чисельності працівників різні моменти часу здійснюється у днях місяця. З розрахунку також видно, що для знаходження середнього рівня, коли відомі початковий та кінцевий рівні, достатньо обчислити середню арифметичну з останніх:

$$\bar{y} = \frac{y_1 + y_2}{2}. \tag{10.18}$$

Розглянемо приклад розрахунку середньої для випадку, коли крім початкового і кінцевого рівнів досліджуваному періоді відомі проміжні рівні, віддалені один від одного на рівній відстані.

Візьмемо для прикладу моментний ряд, що відображає зміну кількості запасів палива на складі за перше півріччя станом на 1-е число кожного місяця (тонн):

| 1/I | 1/II | 1/III | 1/IV | 1/V | 1/VI | 1/VII |
|---|---|---|---|---|---|---|
| 12 | 18 | 22 | 25 | 15 | 20 | 8 |

Якщо скористатися напівсумою крайніх значень цього ряду, то отримаємо середню кількість палива: $\frac{12+8}{2} = 10$ тонн. Але ми бачимо, що кількість палива на початок і кінець періоду була меншою, ніж на інші дати.

Перетворимо цей ряд на ряд середньомісячних залишків шляхом поділу на 2 суми даних на початок і кінець місяця. Знайдемо середню кількість палива за січень. Вона становите: $\frac{12+18}{2}$=15 тонн. За лютий — $\frac{18+22}{2}$=20 тонн і т. п.

Ряд набуде такого вигляду:

| | Січень | Лютий | Березень | Квітень | Травень | Червень |
|---|---|---|---|---|---|---|
| Середньомісячні залишки | 15 | 20 | 23,5 | 20 | 17,5 | 14 |

Обчислимо тепер середній залишок за I півріччя. Він буде рівний простій середній арифметичній із середньомісячних залишків палива:

$$\frac{15+20+23{,}5+20+17{,}5+14}{6} = 18{,}33 \text{ тонн.}$$

Якщо визначити середній залишок за I півріччя за допомогою простої середньої арифметичної з початкового моментного ряду, результат отримаємо такий:

$$\frac{12+18+22+25+15+20+8}{7} = 17{,}1 \text{ тонн.}$$

Розмір середнього рівня залишків, розрахованого другим методом, вийшов інший, ніж у першому розрахунку. Постає питання: який із них правильний? Правильним є перший розрахунок і чому. У нашому прикладі ми мали у своєму розпорядженні дані про рух запасів, тобто про надходження і вибуття запасів палива на перше число кожного місяця. Але насправді рух запасів протягом місяця відбувається безперервно. Навіть наприкінці кожного робочого дня через різні обсяги надходжень і використання їх залишки можуть істотно відрізнятися. Другий спосіб розрахунку середнього рівня ряду використовують у тому випадку, якщо будуть відомі дані, що характеризують безперервну зміну залишків. Але оскільки в умові нашого прикладу такі дані відсутні, та й на практиці дуже часто їх отримати дуже важко, то слід застосувати наближений метод, заснований на припущенні (в даному випадку залишки палива) змінювалися рівномірно в інтервалі, тобто в проміжках між моментами, котрим відомі значення ряду. Ясно, що чим коротші ці проміжки, тим точніше буде обчислено середній рівень.

Весь хід обчислення представимо в такий спосіб. Середня кількість палива за перше півріччя дорівнюватиме всім напівсумам, поділеним на число цих напівсум:

$$\left(\frac{12+18}{2}+\frac{18+22}{2}+\frac{22+25}{2}+\frac{25+15}{2}+\frac{15+20}{2}+\frac{20+8}{2}\right):6=18{,}33\text{ тонн.}$$

Перетворюючи чисельник, отримаємо

$$\frac{\frac{12}{2}+\frac{18+18}{2}+\frac{22+22}{2}+\frac{25+25}{2}+\frac{15+15}{2}+\frac{20+20}{2}+\frac{8}{2}}{6}=$$

$$=\frac{\frac{12}{2}+18+22+25+15+20+\frac{8}{2}}{6}=18{,}33\text{ тонн.}$$

Цю ж середню можна і зручніше обчислювати безпосередньо з рівнів моментного ряду динаміки за формулою середньої хронологічної для моментного ряду динаміки. Оскільки всі члени ряду, крім першого і останнього, в суму входять по два рази і ділять на два, чисельник дробу можна записати так:

$$\bar{y}=\frac{\frac{1}{2}y_1+y_2+y_3+\ldots+y_{n-1}+\frac{1}{2}y_n}{n-1}, \qquad (10.19)$$

де $n$ — число рівновіддалених один від одного рівнів ряду динаміки.

Таким чином, середня хронологічна з моментного ряду динаміки дорівнює сумі рівнів цього ряду, в якій початковий і кінцевий рівні взяті в половинному розмірі, поділеній на кількість рівнів без одного. Пояснюється дана особливість тим, що крайні рівні в моментному ряду динаміки відносяться не тільки до характеризується періоду, але і до суміжного, тому їх і враховують у половинному розмірі.

У знаменнику вираження середнього рівня моментного ряду в умовах, коли рівні відстоять від одного на рівній відстані, береться число рівнів без одиниці, оскільки в числі доданків перший і останній рівні беруться в половинному розмірі. Величина знаменника збігається з числом інтервалів, що охоплюються поруч; у прикладі цифра 6 — число місяців у півріччі. У чисельнику всі рівні «зважені» за величинами відповідних періодів (0,5 періоду — перший і останній рівні, 1,0 — усі проміжні рівні).

За формулою середньої хронологічної в моментному ряду динаміки обчислюється середня вартість основних засобів, середня вартість товарних запасів, середня чисельність працівників, середня чисельність підприємств та ін. показники, які зазвичай характеризуються на рівновіддалені моменти часу.

*Третій варіант*. У практиці статистичної роботи іноді трапляються випадки, коли рівні моментного ряду відстоять один від одного на різній відстані (у часі), тобто інтервали в ряду нерівні. Як у таких моментних рядах динаміки обчислити середній рівень? У цих випадках розрахунок середнього рівня принципово не відрізняється від попереднього: спочатку обчислюються середні для кожного з окремих інтервалів, а потім середня з цих середніх, зважених за величинами інтервалів.

Наприклад, на нафтобазі в процесі інвентаризації встановлено такі залишки бензину: на 1 січня — 70 т, на 1 квітня — 30 т, на 1 серпня — 120 т, на 1 січня наступного року — 50 т. Отже, є три інтервали: = 3 місяці, = 4 місяці, = 5 місяців. Визначимо середні рівні за відповідними інтервалами: $y_1 = \frac{70+30}{2}$=50 т; $y_2 = \frac{30+120}{2}$=75 т; $y_3 = \frac{120+50}{2}$=85 т.

Звідси

$$\bar{y} = \frac{50\cdot 3 + 75\cdot 4 + 85\cdot 5}{3+4+5} = \frac{875}{12} = 72{,}9 \text{ т.}$$

Розрахунок середнього рівня моментного ряду з нерівновіддаленими моментами часу обчислюється за формулою середньої арифметичної зваженої:

$$\bar{y} = \frac{\sum_{1}^{n} y_i t_i}{\sum_{1}^{n} t_i}, \qquad (10.20)$$

де $y_i$ — середні окремих інтервалів;

$t_i$ — перерви інтервалів часу нерівно віддалених моментів;

$n$ — кількість інтервалів.

Отже, якщо в моментному ряді інтервали нерівні, доводиться застосовувати середню зважену. Зважувати рівні доводиться інтервалами нерівних моментів часу.

Слід відмітити, що формула середнього рівня, обчисленого по вичерпним даним про зміну явищ (перший варіант розрахунку) і формула середнього рівня моментного ряду з рівними інтервалами (другий варіант розрахунку) є окремими випадками формули середнього рівня ряду з нерівними інтервалами.

**Ряд динаміки з наростаючими підсумками**. Інтервальний ряд, у якому показники можуть підсумовуватися, можна представити як нагромаджений результат. Для цього рівні ряду послідовно сумуються. Наведемо такий приклад.

Обсяг виробництва цукру на одному із цукрових заводів за тиждень характеризується такими показниками (табл. 10.15).

Таблиця 10.15

**Обсяг виробництва цукру за п'ятиденках, т**

| П'ятиденки | Обсяг виробництва, т | В наростаючих підсумках з початку місяця |
|---|---|---|
| Перша | 3 017,8 | 3 017,8 |
| Друга | 3 096,8 | 6 114,6 |
| Третя | 3 142,0 | 9 256,6 |
| Четверта | 3 218,0 | 12 474,6 |
| П'ята | 3 353,2 | 15 827,8 |
| Шоста | 3 467,0 | 19 294,8 |
| Всього за місяць | 19 294,8 | |

Наростаючими підсумками зручно користуватися, коли потрібно бачити узагальнені підсумки з початку місяця, наприклад, для аналізу виконання місячного плану.

**Розрахунок середнього абсолютного приросту**. При аналізі розвитку явищ часто виникає необхідність дати узагальнюючу характеристику інтенсивності розвитку за тривалий період. Узагальнюючим показником швидкості абсолютної зміни рівнів динаміки є середній абсолютний приріст.

*Середній абсолютний приріст* є середнім з абсолютних приростів за рівні проміжки часу одного періоду.

Середній абсолютний приріст характеризує збільшення (зменшення) рівня явища в середньому за ту чи іншу одиницю часу (місяць, квартал, рік тощо) аналізованого періоду. Обчислюється середній абсолютний приріст за формулою середньої арифметичної простий їх ланцюгових приростів за послідовні та рівні за тривалістю періоди часу.

Так, якщо є динамічний ряд: $y_1; y_2; y_3; \ldots; y_{n-1}; y_n$, прирости отримуються як різниця між наступними рівнями ряду та попередніми:

Таблиця 10.16

| Приросты | Разности |
|---|---|
| $\Delta_1$ | $y_2 - y_1$ |
| $\Delta_2$ | $y_3 - y_2$ |
| $\Delta_3$ | $y_4 - y_3$ |
| . . . | . . . |
| $\Delta_n$ | $y_n - y_{n-1}$ |

Знайдемо середній приріст ($\Delta\bar{y}$):

$$\Delta\bar{y} = \frac{(y_2 - y_1) + (y_3 - y_2) + \ldots + (y_n - y_{n-1})}{n-1}. \tag{10.21}$$

Якщо послідовні ланцюгові абсолютні прирости позначити через $\Delta_1$, $\Delta_2$, $\Delta_3$, . . . , то середній абсолютний приріст, що позначаються через може бути знайдений за формулою:

$$\Delta\bar{y} = \frac{\Delta_1 + \Delta_2 + \ldots + \Delta_n}{n} = \sum_{n=1}^{n} \Delta_i, \tag{10.22}$$

де $n$ — число абсолютних приростів в досліджуваному періоді.

Зважаючи на те, що ланцюгові та базисні абсолютні прирости пов’язані між собою, середній абсолютний приріст можна обчислювати і через базисний. Оскільки сума всіх ланцюгових приростів ($\sum\Delta_i$) дорівнює різниці між кінцевим і початковим рівнем ряду ($y_n - y_1$) за періоди, що розглядаються, то середній абсолютний приріст можна знайти за формулою:

$$\Delta\bar{y} = \frac{y_n - y_1}{n} = \frac{\Delta y_б}{n}. \tag{10.23}$$

де $\Delta y_б = y_n - y_1$; $n$ — число ланцюгових приростів; якщо ж під $n$ розуміти кількість одиниць часу аналізованого періоду, для якого обчислюється середній абсолютний приріст, формула дещо зміниться:

$$\Delta\bar{y} = \frac{y_n - y_1}{n-1}. \tag{10.24}$$

Ця формула використовується у випадку, якщо вихідні рівні динамічного ряду не відомі, а дані початковий і кінцевий рівень ряду динаміки. Число абсолютних приростів менше числа рівнів на одиницю.

Середній абсолютний приріст принципово можна обчислити за формулою:

$$\bar{\Delta} = \frac{\sum \Delta_i}{\sum t} = \frac{y_n - y_1}{\sum t}, \qquad (10.25)$$

де $y_1$ — базисний рівень, яким виступає кінцевим рівень попереднього етапу;

$t_i$ — тривалість окремих інтервалів часу;

$\Delta_i$ — окремі абсолютні прирости.

При обчисленні середнього абсолютного приросту потрібно вказувати період часу, за який він обчислюється, а також і одиницю часу, в розрахунку за який обчислюється середній абсолютний приріст.

Розглянемо приклад розрахунку середнього абсолютного приросту за ланцюговими абсолютними приростами.

Якщо повернуться до динамічного ряду виробництва шоколадних цукерок в Україні (табл. 10.12), середній приріст отримаємо з наступного розрахунку:

$$\Delta \bar{y} = \frac{(-7{,}2) + 1{,}0 + 0{,}1 + 13{,}2 + 4{,}1)}{5} = \frac{11{,}2}{5} = 2{,}2 \text{ тис. т.}$$

За даними табл. 10.12 знайдемо середній абсолютний виробництва шоколадних цукерок в Україні за 2016–2020 рр. на основі базисних абсолютних приростів:

$$\Delta \bar{y} = \frac{y_n - y_1}{n} = \frac{70{,}4 - 59{,}5}{5} = 2{,}2 \text{ тис. т.}$$

В результаті розрахунків за базисними абсолютними приростами ми отримали такий же результат, що і при визначенні $\Delta\bar{y}$ за ланцюговими значеннями ланцюгових абсолютних ознак.

Отже, у нашому прикладі абсолютний приріст складе: 70,4 – 59,5 = 11,2 = (– 7,2) + 1,0 + 0,1 + 13,2 + 4,1. Абсолютні прирости, розраховані за базисним принципом, підсумовуванню не підлягають, шукана різниця $y_n - y_1$ представляє безпосередньо абсолютний приріст кінцевого рівня поточного періоду.

За окремі роки абсолютний приріст виробництва шоколадних цукерок був значно більшим за середній, у деякі роки виробництво цукерок знижувалося в порівнянні з попередніми роками, а в середньому щорічно протягом усього досліджуваного періоду їх виробництво збільшувалося на 2,2 тис. т. Таким образом, абсолютна величина приросту (а ще більше — середня) істотно доповнюють аналіз рівнів явища відносними величинами.

Величина середньорічного приросту залежить від рівнів, які стоять на кінцях ряду, і зовсім залежить від проміжних рівнів. Цей показник може давати спотворене уявлення про динаміку, тому що цілком залежить від того, на які роки доведеться початок і кінець періоду.

**Середній темп зростання**. Серед статистичних показників велике поширення набув середній темп зростання. Відомі численні випадки його застосування в економічному аналізі та прогнозуванні. Закономірність зміни темпів динаміки є найважливішим аспектом дослідження динамічного ряду. Це змушує нас охарактеризувати цю найважливішу узагальнюючу характеристику динаміки докладніше.

*Середній темп зростання* або *коефіцієнт зростання* ($\overline{K}_p$) — середня величина відношення наступного рівня ряду динаміки до попереднього, виражена (як правило) у процентах. Він дає уявлення про середню швидкість зміни рівнів явища в часі і показує, у скільки разів у середньому збільшується (зменшується) рівень явища за ту чи іншу одиницю часу аналізованого періоду. При обчисленні середнього темпу зростання слід враховувати, що швидкість розвитку явища відбувається за правилами складних процентів, де нагромаджується приріст на приріст. Тому середній темп приросту прийнято обчислювати за формулою середньої геометричної із ряду послідовних (ланцюгових) темпів зростання за складові періоду проміжки часу. Це диктується зв'язком між окремими темпами зростання і загальним темпом зростання показника за весь період (останній дорівнює добутку окремих темпів). Ланцюговий темп зростання характеризує відношення будь-якого рівня динаміки до попереднього рівня і виражається у процентах або частках одиниці. В останньому випадку його називають коефіцієнтом зростання (надалі ми застосовуватимемо лише термін темпи зростання незалежно від цього, виражено це ставлення у частках одиниці чи відсотках, оскільки це несуттєво для характеристики динаміки).

Середній коефіцієнт зростання для спрощення розрахунків іноді обчислюється за середньою арифметичною простою з ланцюгових коефіцієнтів зростання. У прикладі ця величина дорівнює

$$\overline{K}_p = \frac{\sum K}{n} = \frac{0{,}879 + 1{,}019 + 1{,}002 + 1{,}247 + 1{,}062}{5} = \frac{5{,}209}{5} = 1{,}042.$$

Але цей розрахунок середнього коефіцієнта неправильний. Розрахунок середнього коефіцієнта за середньою арифметичною неправомірно, оскільки сума ланцюгових коефіцієнтів зростання (у нашому прикладі (5,209) не виражає собою ніякої реальності. Якщо ми візьмемо початковий рівень динаміки (59,5) і помножимо його п'ять разів на отриманий темп зростання, то отримаємо обсяг виробництва цукерок в 2020 р.:

$$59{,}5 \times 1{,}042 \times 1{,}042 \times 1{,}042 \times 1{,}042 \times 1{,}042 = 73{,}1 \text{ тис. т.}$$

Цей результат (73,1) не відповідає фактичному значенню ряду динаміки (70,7). Таким чином, розрахунок здійснено невірно. Це сталося тому, що визначальним показником коефіцієнтів зростання є добуток їх, бо коефіцієнт зростання показує, у скільки разів рівень явища даного періоду більше (або менше) рівня явища в періоді, прийнятому за базу порівняння. Якщо ж визначальним показником є добуток, то середню величину їм потрібно обчислювати за іншою формулою.

Мається динамічний ряд, що складається з рівновіддалених рівнів: $y_1, y_2, y_3, \ldots, y_{n-1}, y_n$. Ланцюгові темпи зростання (у вигляді коефіцієнтів) являють собою відношення:

$$K_{p_1} = \frac{y_2}{y_1}; \quad K_{p_2} = \frac{y_3}{y_2}; \quad K_{p_3} = \frac{y_4}{y_3}; \quad \ldots, \quad K_{p_n} = \frac{y_n}{y_{n-1}}. \qquad (10.26)$$

Якщо тенденція розвитку у відомому часовому інтервалі охарактеризована за допомогою деякого постійного темпу зростання, то як останній приймається середній темп за відповідний період.

Середній коефіцієнт (темп) зростання обчислюють за формулою середньої геометричної:

$$K_p = \sqrt[n]{K_{p1} \cdot K_{p2} \cdot K_{p3} \ldots K_{pn}} = \sqrt[n]{\prod_{i=1}^{n} K_{pi}}, \qquad (10.27)$$

або

$$\overline{T}_p = \sqrt[n]{K_{p1} \cdot K_{p2} \cdot K_{p3} \ldots K_{pn}} \cdot 100, \qquad (10.28)$$

де $\overline{T}_p$ — середній темп зростання;

$\prod$ — знак множення;

$n$ — число ланцюгових коефіцієнтів зростання.

У випадку нерівних інтервалів формула розрахунку прийме наступний вид:

$$\overline{T}_p = \sqrt[\sum t_i]{T_{p1}^{t1} \cdot T_{p2}^{t2} \cdot T_{p3}^{t3} \cdot \ldots \cdot T_{pn}^{tn}}. \qquad (10.29)$$

Середній темп зростання виробництва шоколадних цукерок за 2015–2020 роки. (див. табл. 10.12) можна отримати з наступного розрахунку:

$$\overline{T}_p = \sqrt[5]{0{,}879 \cdot 1{,}019 \cdot 1{,}002 \cdot 1{,}247 \cdot 1{,}062} = \sqrt[5]{1{,}188} = 1{,}035.$$

$$\lg \overline{T}_p = \frac{1}{5} \lg 1{,}188 = \frac{1}{5} 0{,}07482 = 0{,}01496.$$

По антилогарифмам[1] знаходимо середньорічний коефіцієнт зростання. Для виробництва шоколадних цукерок у нашому прикладі він буде за 2015–2020 рр. дорівнює $\overline{T}_p$ = 1,035 або 103,5%.

Отже, середній річний темп зростання виробництва шоколадних цукерок за 2015–2020 рр. становитиме 1,035 або 103,5 %. Правильність обчислення підтверджується визначенням теоретичного рівня виробництва шоколадних цукерок. Якщо послідовно перемножити щорічні рівні на середній коефіцієнт зростання, виходить рівень останнього року (табл. 10.17).

Таблиця 10.17

**Емпіричні та теоретичні дані рівня виробництва шоколадних цукерок в Україні за 2015–2020 роки. (тис. т)**

| Роки | Обсяг виробництва, тис. т | Теоретичний рівень виробництва шоколадних цукерок, тис. т |
|---|---|---|
| 2015 | 59,5 | 59,5 |
| 2016 | 52,3 | 59,5·1,035=61,6 |
| 2017 | 53,3 | 61,6·1,035=63,8 |
| 2018 | 53,4 | 63,8·1,035=66,0 |
| 2019 | 66,6 | 66,0·1,035=68,3 |
| 2020 | 70,7 | 68,3·1,035=70,7 |

З таблиці видно, що середньорічний темп зростання обчислено правильно, оскільки емпіричний рівень останнього року дорівнює теоретичному.

Середній коефіцієнт зростання, обчислений за середньою арифметичною (у нашому прикладі — 1,042) завжди буде більше середнього коефіцієнта зростання, обчисленого за середньою геометричною. Але в тих випадках, коли коливання індивідуальних значень

[1] Для обчислення середніх темпів зростання в теперішній час користуються спеціальними таблицями.

коефіцієнтів зростання є незначними, обчислювати середній темп зростання можна і за середньою арифметичною, так як вона наочна і обчислюється дуже просто. Різниця між коефіцієнтами зростання, численними за середньою арифметичною і за середньою геометричною буде тим менше, чим менші коливання індивідуальних коефіцієнтів. Якщо рівні коливання індивідуальних коефіцієнтів зростання значні, то обчислювати середній коефіцієнт зростання за середньою арифметичною не можна, оскільки це призведе до великих спотворень.

При повторюваності варіант застосовують середню зважену геометричну. Оскільки повторювані варіанти виявляються під коренем помноженими на власну величину, це означає їх зведення у відповідний ступінь, і числові значення вагою виявляться рівними числовим значенням показників ступеня, тобто

$$K_p = \sqrt[\Sigma y]{\prod (x^y)}. \tag{10.30}$$

де $y$ — ознака, що служить вагою.

Тому виваженою геометричною середньою називається корінь ступеня, що дорівнює сумі ступенів у добутку варіант, що знаходяться під коренем.

Наприклад, середньорічний темпи зростання виробництва за перші п’ять років становив 110 %, наступні три роки — 115 %, а останні два роки — 125 %. Потрібно обчислити середній річний темп зростання виробництва продукції за всі 10 років[1].

У цьому випадку слід застосувати формулу середньої виваженої геометричної, так як наведені відсотки зростання відносяться до різних за тривалістю періодів часу і вимагають зважування: 110 % на 5 років, 115 % на 3 роки і 125 % на 2 роки. Отже:

$$K_p = \sqrt[10]{110^5 \cdot 115^3 \cdot 125^2};$$

$$\lg K_p = \frac{1}{10}(5\lg 110 + 3\lg 115 + 2\lg 125) = \frac{1}{10}(5 \cdot 2{,}0414 + 3 \cdot 2{,}0607 +$$

$$+ 2 \cdot 2{,}0969) = \frac{1}{10}(10{,}2070 + 6{,}1821 + 4{,}1938) =$$

$$\frac{1}{10} \cdot 20{,}5829 = 2{,}0583.$$

---

[1] Приклад запозичений в кн.: *Гранков В. П.* 1957, с. 47.

Звідки $K_p$=114,4 %.

Коли відомі рівні динамічного ряду, розрахунок середнього темпу можна спростити. Підставивши в підкорене вираз замість $K_{p1}$, $K_{p2}$, $K_{p3}$, . . . їх вихідні значення $\frac{y_2}{y_1}, \frac{y_3}{y_2} \ldots$, отримаємо формулу для розрахунку середнього темпу зростання $\overline{K}_p$ (у вигляді коефіцієнта):

$$\overline{K}_p = \sqrt[n]{\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \ldots \frac{y_n}{y_{n-1}}} = \sqrt[n]{\frac{y_n}{y_1}}, \qquad (10.31)$$

або в процентах

$$\overline{T}_p = \sqrt[n]{\frac{y_n}{y_1}} \cdot 100, \qquad (10.32)$$

де $y_n$ — кінцевий член низки динаміки;

$y_1$ — початковий його член;

$n$ — число членів ряду динаміки.

Число коефіцієнтів зростання завжди на одиницю менше рівня безперервного ряду динаміки. Якщо під $n$ розуміти кількість рівнів безперервного ряду динаміки або кількість одиниць часу (місяців, кварталів, років) протягом досліджуваного періоду, то формула матиме вигляд:

$$\overline{T}_p = \sqrt[n-1]{\frac{y_n}{y_1}} \cdot 100, \qquad (10.33)$$

Ця формула набула широкого поширення. У прикладі з аналізом ряду шоколадних цукерок середній геометричний темпи зростання виробництва за п'ять років складе:

$$\overline{T}_p = \sqrt[6-1]{\frac{70{,}7}{59{,}7}} = \sqrt[5]{1{,}188}. \quad \lg \overline{K} = \frac{\lg 1{,}188}{5} = 0{,}0748.$$

Таким же чином за антилогарифмами визначаємо величину середнього коефіцієнта зростання. У прикладі $\overline{T}_p$=1,035, або 103,5 %. Це означає, що середньорічний темп приросту виробництва шоколадних цукерок за 2015–2020 роки. становив 103,5%.

У разі, якщо порівняння починається не з початковим рівнем поточного періоду ($y_1$), а з кінцевим рівнем попереднього періоду ($y_0$), формула середнього темпу зростання буде мати такий вигляд:

$$\overline{T}_p = \sqrt[n]{\frac{y_n}{y_0}} \cdot 100, \qquad (10.34)$$

Для розрахунку середніх темпів зростання за формулою (10.34) на відміну від розрахунку за формулою (10.27) не потрібно знати річні темпи.

Викладений вище порядок опосередкування темпів зростання відноситься до темпів, розрахованих за ланцюговою ознакою. Якщо окремі темпи обчислені за базовим принципом, то середній темп можна отримати, отримуючи корінь відповідної міри з останнього базисного темпу.

Мається можливість всі наведені вище формули середніх об'єднати *єдиною загальною середньою*[1].

Середній геометричний темп зростання дозволяє перейти від рівня ряду за рік, який передує періоду, що розглядається, до його рівня за останній рік цього періоду. Повернемося до нашого прикладу. Виробництво шоколадних цукерок у 2015 р. склало 59,5 тис. т. За середньорічного темпу зростання 3,5 % за п'ять років обсяг виробництва у 2020 р. становитиме 70 тис. т. ($59{,}5 \cdot 1{,}035^5 = 70{,}7$). Ця величина відповідає значенню рівня динаміки за останній період.

**Середній темп приросту** показує, наскільки процентів у середньому збільшується (зменшується) рівень явища за ту чи іншу одиницю часу (зазвичай за рік) аналізованого періоду.

Середній темп приросту за безпосередніми даними про темпи приросту не можна обчислити, оскільки загальний та індивідуальні темпи приросту можуть мати від'ємні значення. Тут довелося б як і для середніх темпів зростання, користуватися геометричною середньою, тобто отримувати корінь високих ступенів, і отже, логарифмувати. Але, як відомо, негативні числа логарифмів немає. Тому для розрахунку середніх темпів приросту користуються відомим нам з попереднього співвідношення: $\overline{T}_{пр} = \overline{T}_p - 1$ (у вигляді коефіцієнтів) і $\overline{T}_{пр}$ (%) = $\overline{T}_p$ (%) – 100 (у вигляді процентів). Це співвідношення справедливе і для розрахунку середніх темпів приросту.

---

[1] Загальну формулу середньої можна представити наступним чином: $\bar{x} = \left( \dfrac{\sum_{i=1}^{i=n} x_i^k y_i}{\sum_{i=1}^{i=n} y_i} \right)^{\frac{1}{k}}$.

Середньорічний темп приросту обчислюється через середньорічний коефіцієнт і темпи зростання як різниця середнього темпу зростання і 100:

$$\overline{T}_{пр} = T - 100 \quad \text{або} \quad \overline{T}_{пр} = K \cdot 100 - 100. \qquad (10.35)$$

Отже, для обчислення середнього темпу приросту спочатку треба визначити середній темп зростання, та потім середній темп приросту.

Вище ми виявили, що середній темп зростання виробництва шоколадних цукерок за 2016–2020 рр. становив 1,035, або 103,5%. Значить, середній темп приросту виробництва шоколадних цукерок становить

$$\overline{T}_{пр} = T - 100 = 103{,}5 - 100 = 3{,}5\ \%.$$

Якщо величина темпу зростання за весь період менше одиниці (правильний дріб), то і середній темп зростання буде менше одиниці, а середній темп приросту буде величиною від’ємною.

Наприклад, постачання електроенергії, виробленої на атомних станціях, в Україні за 2016–2020 рр., знизилося з 72,9 млрд кВт·год. до 55,4 млрд кВт·год., або на 24,0 % (загальний темп приросту становив — 24,0 %). Отже, темп зміни (зростання) виробництва електроенергії в Україні за цей період становитиме 0,760 $\left(\frac{55{,}4}{72{,}9}\right)$ або 76,0 %. Знайдемо спочатку середньорічний темп зміни (коефіцієнт зростання). Він складе

$$\sqrt[5-1]{\frac{55{,}4}{72{,}9}}, \text{ або } \overline{T}_{р} = \sqrt[4]{0{,}760}.$$

Звідки $\overline{T}_{р}$ =0,934 або 93,4 %.

Отже, середньорічний темп зміни постачання електроенергії, виробленої на атомних станціях, в Україні за 2016–2020 рр. становитиме 76,0 %, а середній розмір зниження — 6,4 % на рік.

Як видно з формули (10.34), величина середнього темпу зростання залежить від співвідношення кінцевого і початкового рівнів. Дуже важливо у зв’язку з цим обґрунтовано вибирати періоди, що узагальнюють середні темпи зростання. Ці періоди повинні мати, як правило, один напрямок бути в цьому відношенні якісно однорідними.

В економічному аналізі доводиться іноді визначати темп зростання за даними рівнів рядів динаміки, які різко коливаються. Як приклад скористаємося такими даними про урожайність гороху (ц з 1 га) в одному фермерському господарстві:

За 2010–2015 рр.

| 2010 р. | 2011 р. | 2012 р. | 2013 р. | 2014 р. | 2015 р. |
|---|---|---|---|---|---|
| 17,8 | 27,3 | 24,1 | 28,2 | 26,3 | 30,2 |

За 2015–2020 рр.

| 2015 р. | 2016 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. |
|---|---|---|---|---|---|
| 30,2 | 30,8 | 27,8 | 35,1 | 35,6 | 21,7 |

Якщо для розрахунку середнього темпу зростання урожайності за 2016–2020 рр. взяти за базу 2015 р. та обчислити середньорічний темп за формулою (10.34), то розрахунок покаже зниження урожайності:

$$\overline{T} = \sqrt[5]{\frac{21,7}{30,2}} = \sqrt[5]{0,718} = 0,936.$$

Таким чином, можна дійти невтішного висновку, що урожайність за 2016–2020 рр. знизилася на 28,2 %, а середньорічний темп зниження становив 7,4 %. Але це неправильно характеризує динаміку урожайності за досліджуваний період. Справді, обчислимо середню річну врожайність за 2010–2015 роки. Вона дорівнює 27,2 ц з га $\left(\frac{27,3+24,1+28,2+26,3+30,2}{5}\right)$. Обчислимо середню урожайність за 2016–2020 рр. Вона складе 30,2 ц з га $\left(\frac{30,8+27,8+35,1+35,6+21,7}{5}\right)$. Як бачимо, за 2016–2020 рр. відбулося зростання середньорічної урожайності в порівнянні з 2011–2015 рр. Темп зростання середньорічної урожайності становив 11,0 % (30,2 : 27,2 $\times$ 100 – 100).

Уявне зниження урожайності гороху за 2016–2020 рр. відбулося тому, що кінцевий рік п’ятирічки (2020 р.) був неурожайним. Отже, розрахунки показників динаміки в рядах з рівнями, що різко коливаються, правильніше зводити на основі розрахунку темпу зростання середньорічних рівнів, а не на основі розрахунку середньорічних темпів.

Розглянуті відносні показники динаміки мають широке застосування в практичній і науковій роботі при дослідженні соціально-економічних явищ і процесів.

Застосування перелічених показників динаміки є першим етапом аналізу динамічних рядів, що дозволяє виявити швидкість та інтенсивність розвитку явищ, які представлені рядом. Подальший аналіз рядів динаміки соціально-економічних показників пов’язаний з складнішими узагальненнями, з визначенням основний тенденції.

## 10.6. Методи вивчення тенденцій у розвитку явищ

*Тенденція динамічного ряду* — основна спрямованість динаміки розвитку явища. Якщо встановлено, що ряд має виражену тенденцію, то проводять аналіз компонентів ряду динаміки: 1) основної тенденції, або тренду; 2) довготривалих циклічних коливань; 3) короткочасних або сезонних коливань (регулярних змін усередині року); 4) випадкових коливань (вплив зовнішніх факторів).

Важливим завданням при аналізі рядів динаміки є задача визначення основної тенденції розвитку, властивої тому чи іншому ряду динаміки. Таке завдання виникає при прогнозуванні даного явища на майбутнє або при вивченні сезонних коливань, а також в інших випадках. Наприклад, за коливаннями рівнів урожайності будь-якої сільськогосподарської культури в певні роки може не проглядатися безпосередньо тенденція зростання урожайності і тому вона повинна бути виявлена статистичними методами. На розвиток явища в часі можуть впливати різні за своїм характером і напрямом впливу фактори. Одні з них істотно впливають на динаміку явища і формують у рядах динаміки певну тенденцію розвитку. Дія інших факторів є незначним випадковим.

У статистиці виявлення основної тенденції (тренду) у розвитку ряду динаміки називають також вирівнюванням часового ряду, а відповідно, методи виявлення основної тенденції називають методами вирівнювання.

Під час дослідження у рядах динаміки загальної тенденції розвитку застосовуються різні прийоми. Суть різних прийомів, за допомогою яких здійснюється згладжування і вирівнювання, зводиться до заміни фактичних рівнів динамічного ряду розрахунковими, що мають значно менші коливання. Зменшення коливання дозволяє тенденції розвитку проявити себе наочно.

Тенденція в рядах динаміки виявляється різними методами: метод укрупнених інтервалів, механічне згладжування, метод ковзної середньої, аналітичне вирівнювання методом найменших квадратів.

**Метод укрупнених інтервалів**. Одним з найбільш елементарних способів вивчення загальної тенденції в ряді динаміки є *укрупнення інтервалів*. Сутність цього способу полягає в укрупненні інтервалів, до яких відносяться рівні низки динаміки.

Дуже часто при розгляді динаміки зміни соціально-економічних явищ за короткі проміжки часу внаслідок впливу різних факторів спостерігається зниження або підвищення їх. У результаті дуже складно визначити основну тенденцію розвитку досліджуваного явища.

Техніку розрахунку методом укрупнених інтервалів покажемо на прикладі продажу молока та молочних виробів у торговельній мережі міста *N* за п'ять років (табл. 10.18).

Таблиця 10.18

**Продаж молока та молочних виробів у торговій мережі міста *N***

| Роки, квартали | | Продаж молока | Продаж молочних виробів |
|---|---|---|---|
| 1-й рік | I | 46,8 | 43,2 |
| | II | 45,3 | 39,1 |
| | III | 46,8 | 42,8 |
| | IV | 56,1 | 52,6 |
| 2-й рік | I | 57,5 | 46,1 |
| | II | 46,8 | 43,5 |
| | III | 44,6 | 43,0 |
| | IV | 53,4 | 50,4 |
| 3-й рік | I | 53,7 | 45,2 |
| | II | 42,5 | 47,7 |
| | III | 53,6 | 48,6 |
| | IV | 58,1 | 54,8 |
| 4-й рік | I | 56,1 | 47,6 |
| | II | 47,8 | 51,3 |
| | III | 45,7 | 51,9 |
| | IV | 61,2 | 59,4 |
| 5-й рік | I | 59,8 | 54,2 |
| | II | 50,4 | 55,6 |
| | III | 46,7 | 57,1 |
| | IV | 69,5 | 60,2 |
| Всього | | 1 042,4 | 994,3 |

З поквартальних даних продажу молока та молочних виробів важко визначити загальну тенденцію розвитку торгівлі цими продуктами. Розглядаючи наведені дані про продаж молока і молочних продуктів, можна спостерігати значні їх коливання. У силу дії різних факторів, що діють у різних напрямках, у рядах динаміки спостерігається підвищення та зниження їх. Через це не видно основну тенденцію розвитку досліджуваного явища. Для виявлення загальної тенденції збільшення обсяг продажу молока та молочних продуктів зробимо укрупнення інтервалів. Якщо укрупнити інтервали часу, поквартальні дані об'єднати в річні шляхом складання, то отримаємо нові ряди динаміки обсягів продажів, що показує їх послідовне збільшення (табл. 10.19). Після укрупнення інтервалів загальна тенденція збільшення обсягів реалізації молока і молочних продуктів виступає чітко.

Таблиця 10.19

**Укрупнені (річні) дані про продаж молока та молочних виробів у місті *N* за п'ять років**

| Рік | Продаж молока | Продаж молочних виробів |
|---|---|---|
| 1-й | 195,0 | 177,7 |
| 2-й | 202,3 | 183,0 |
| 3-й | 207,9 | 196,3 |
| 4-й | 210,8 | 210,2 |
| 5-й | 226,4 | 227,1 |

З метою кращого вирівнювання загальної тенденції розвитку за тривалі проміжки часу інколи розраховують середній рівень, що відповідає різним періодам часу. В цьому випадку весь період розбивається на рівні відрізки, для кожного з яких обчислюється середня арифметична проста. Ряд, побудований на основі обчислених середніх, буде характеризувати розвиток досліджуваного явища.

Як правило, укрупнення інтервалів розпочинають із найменшого інтервалу, тобто з інтервалу, що об'єднує два рівня. У випадку, якщо при такому об'єднанні не має ясної картини щодо загальної закономірності розвитку, переходять до наступного можливого варіанту (до об'єднання трьох рівнів, чотирьох рівнів і т. д.). Якщо в досліджуваній сукупності спостерігається циклічність у коливанні, то укрупнені інтервали беруть з урахуванням періоду коливання (циклу).

Метод середніх дозволяє наглядно визначити загальну тенденцію розвитку, але він не враховує зміни рівнів кожного періоду. Крім того, для застосування методу середніх необхідно мати ряд динаміки за тривалий проміжок часу, тоді як економісту дуже часто приходиться мати справу з рядом, який складається із 10–20 членів. За такого числа рівнів ряду визначити загальну тенденцію розвитку шляхом укрупнення інтервалів і обчислення середніх рівнів практично неможливо. Недоліком метода середніх є те, що при його використанні з поля зору випадає процес зміни всередині укрупнених інтервалів. В цих випадках необхідно застосовувати методи, які б дали змогу виключити вплив на загальну тенденцію другорядних і випадкових факторів і враховували б вплив лише основного (основних) факторів.

В статистиці розроблено ряд математико-статистичних методів, які застосовуються для визначення загальної тенденції розвитку. Зокрема, до них відносять метод механічного вирівнювання, метод аналітичного вирівнювання і спосіб ковзної середньої.

**Механічне вирівнювання**. Одним з найбільш простих способів визначення загальної тенденції розвитку є механічне вирівнювання. При *механічному вирівнюванні* необхідно попередньо визначити середній приріст ($\Delta\overline{y}$), який обчислюється шляхом поділу розмаху коливань ($y_n - y_1$) на число членів ряду без одного. Обчислення членів вирівняного ряду проводиться з початкового ряду шляхом послідовного додавання середнього приросту за аналізований період.

Розглянемо застосування цього методу за даними про продаж молока та молочних виробів у торговельній мережі міста *N* за п’ять років (табл. 10.20).

Таблиця 10.20

**Визначення способом механічного вирівнювання загальної тенденції продажу молока та молочних виробів у місті *N* за п’ять років** (млн грн)

| Роки, квартали | | Продаж молока | | Продаж молочних виробів | |
|---|---|---|---|---|---|
| | | Звітні дані | Вирівняні дані | Звітні дані | Вирівняні дані |
| 1 | | 2 | 3 | 4 | 5 |
| 1-й рік | I | 46,8 | 46,8 | 43,2 | 43,2 |
| | II | 45,3 | 48,0 | 39,1 | 44,1 |
| | III | 46,8 | 49,2 | 42,8 | 45,0 |
| | IV | 56,1 | 50,4 | 52,6 | 45,9 |
| 2-й рік | I | 57,5 | 51,6 | 46,1 | 46,8 |
| | II | 46,8 | 52,8 | 43,5 | 47,7 |
| | III | 44,6 | 54,0 | 43,0 | 48,6 |
| | IV | 53,4 | 55,2 | 50,4 | 49,5 |
| 3-й рік | I | 53,7 | 56,4 | 45,2 | 50,4 |
| | II | 42,5 | 57,6 | 47,7 | 51,3 |
| | III | 53,6 | 58,8 | 48,6 | 52,2 |
| | IV | 58,1 | 60,0 | 54,8 | 53,1 |
| 4-й рік | I | 56,1 | 61,2 | 47,6 | 54,0 |
| | II | 47,8 | 62,4 | 51,3 | 54,9 |
| | III | 45,7 | 63,6 | 51,9 | 55,8 |
| | IV | 61,2 | 64,8 | 59,4 | 56,7 |
| 5-й рік | I | 59,8 | 66,0 | 54,2 | 57,6 |
| | II | 50,4 | 67,2 | 55,6 | 58,5 |
| | III | 46,7 | 68,4 | 57,1 | 59,4 |
| | IV | 69,5 | 69,6 | 60,2 | 60,3 |
| Всього | | 1 042,4 | 1 164,0 | 994,3 | 1 035,0 |

У наведеному прикладі середній приріст за 5 років складе:
по молоку

$$\overline{\Delta} = \frac{y_n - y_1}{n-1} = \frac{69{,}5 - 46{,}8}{20-1} = \frac{22{,}7}{19} = 1{,}2 \text{ млн грн;}$$

за молочними виробами

$$\overline{\Delta} = \frac{y_n - y_1}{n-1} = \frac{60{,}2 - 43{,}2}{20-1} = \frac{17}{19} = 0{,}9 \text{ млн грн.}$$

Розрахунок членів вирівняного ряду проводиться послідовним додаванням середнього приросту, починаючи з першого члена емпіричного ряду динаміки. По молоку: 46,8+1,2=48,0; 48,0+1,2=49,2; 49,2+1,2=50,4 і т. п. По молочним виробам: 43,2+0,9=44,1; 44,1+0,9= =45,0; 45,0+0,9=45,9 і т. п. Результати розрахунків приведені в графах 3 и 5 табл. 10.20.

Вирівняний за допомогою способу механічного вирівнювання ряд графічно зображується у вигляді прямої лінії (рис. 10.7).

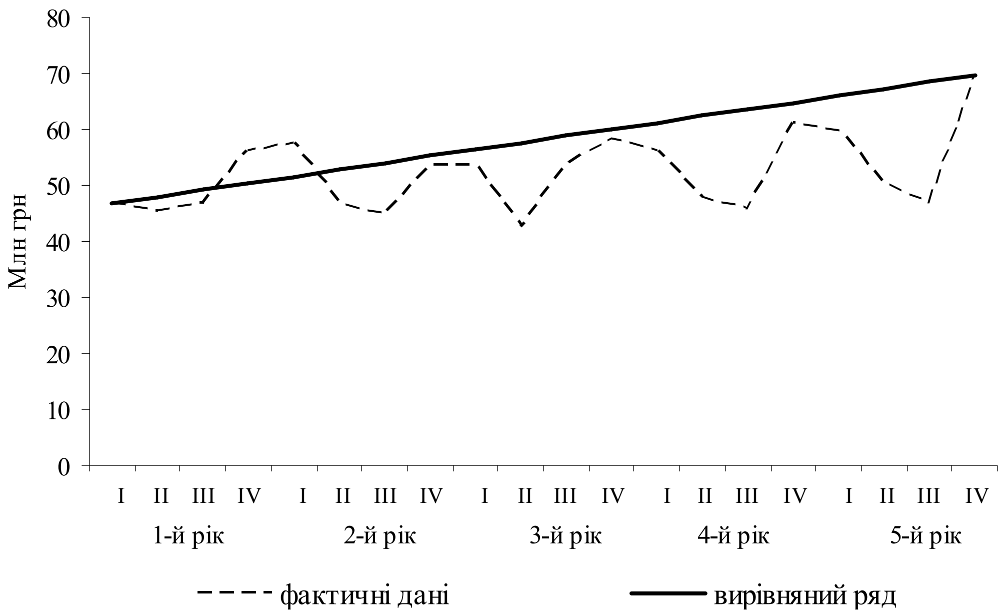



*Рис. 10.7.* Динаміка продажу молока, обчислена за даними табл. 10.20 способом механічного вирівнювання.

Другим недоліком способу механічного вирівнювання є те, що при розрахунку середнього приросту беруться тільки крайні члени ряду. Це часто призводить до перебільшення або зменшення загальної тенденції. Як очевидно з розрахунків, підсумки з продажу за звітними даними не збігаються з підсумками по вирівняному ряду. Наприклад, за молоком ці числа становлять 1042,4 та 1164,0 млн грн, по молочних продуктах — відповідно 994,3 та 10,35 млн грн.

Отже, застосовувати спосіб механічного вирівнювання для визначення загальної тенденції можна лише для динамічних рядів з відносно рівномірною зміною загального рівня ряду. Очевидно, що цей спосіб не прийнятний для визначення загальної тенденції в рядах з відносно значними коливаннями рівнів ряду.

Для виявлення загальної тенденції в рядах зі значними коливаннями в розвитку явища необхідно застосовувати способи побудови вирівняного ряду, в яких швидкість зміни була б розрахована з усіх наявних показників рівня ряду динаміки.

**Метод ковзної середньої**. Одним з найбільш широко відомих прийомів для виявлення загальної тенденції ряду динаміки служить метод *ковзної середньої*, або як іноді його називають, *рухливих середніх*. Сутність цього прийому полягає в заміні фактичних даних середніми арифметичними (ковзними середніми) з кількох рівнів рядів динаміки (трьох, п'яти, шести і т. д. — інтервал ковзання). *Ковзна середня* — рухлива динамічна середня, яка підраховується по динамічному ряду при послідовному пересуванні на один інтервал. Розрахунок середніх ведеться способом ковзання, тобто поступово виключаються з періоду перші рівні і включаються наступні.

Використання ковзної середньої дозволяє усунути в ряду динаміки випадкові коливання і цим виявити основну тенденцію у розвитку явища краще, ніж укрупнені інтервали чи ступінчаста середня. При цьому методі виявлення загальної тенденції використовуються всі члени динамічного ряду. Згладжування за допомогою ковзної середньої засноване на тому, що в середніх величинах погашаються індивідуальні відмінності одиниць сукупності в значеннях (величинах) середньої ознаки, зумовлені випадковими обставинами. Це відбувається внаслідок того, що початкові рівні ряду динаміки замінюються середньою арифметичною величиною всередині обраного інтервалу часу.

Для виявлення основної тенденції зміни явища методом згладжування рядів динаміки необхідно перш за все визначити за емпіричним (вихідним) даними *рухомі* (*ковзаючі*) *середні*. Розрахунок рухомих (ковзаючих) середніх здійснюють з такого числа рівнів ряду, яке відповідає тривалості спостережуваних у ряді динаміки циклів.

Розрізняють методи зважених та незважених ковзних середніх. При використанні методу незважених ковзних середніх період ковзання може включати парне і непарне число рівнів ряду. Залежно від цього дещо змінюється техніка згладжування ряду динаміки. Розглянемо порядок виявлення основної тенденції для рядц динаміки, що складається з непарного числа рівнів.

Для виявлення загальної тенденції протягом тривалого періоду часу застосовують п'ятичленові ковзні середні. Нехай ряд динаміки за певний період часу представлений такими значеннями: $y_1, y_2, y_3, \dots, y_n$. Розрахунок полягає у визначенні середньої з п'яти рівнів ряду. Потім при обчисленні кожної нової середньої відкидають одне значення ряду зліва і приєднують одне значення справа.

Для визначення першого члена згладженого ряду за п'ятичленною середньою ковзною складаються перші п'ять рівнів ряду і сума їх ділиться на п'ять:

$$\bar{y}_1 = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5}. \qquad (10.36)$$

Отримані значення відносять до середини обраного періоду. Для визначення наступного згладженого рівня підрахунок сум п'яти членів ряду починається з другого члена та закінчується шостим:

$$\bar{y}_2 = \frac{y_2 + y_3 + y_4 + y_5 + y_6}{5}. \qquad (10.37)$$

і т. д.

Таким чином, знайдені ковзні середні є центрованими, тобто відносяться до середини періоду згладжування. Тривалість того періоду, який включається до розрахунку при обчисленні ковзної середньої, називається *періодом ковзної середньої*. У нашому прикладі взято п'ятирічний період, але міг бути взятий дворічний, трирічний і т. д.

При використанні ковзної середньої число знайдених значень (рівнів) виявляється менше початкового числа членів ряду. Згладжений ряд коротший за початковий ряд ($p$–1) спостережень ($p$ — величина інтервалу згладжування). Якщо середня розраховується за трьома показниками, то ряд виходить коротшим на два члени (один зникає спочатку, інший — з кінця ряду), за п'ятьма показниками ряд виходить коротшим на чотири члени (два з початку та два з кінця) тощо.

Розглянемо динамічний ряд, що характеризує валовий збір зернових та зернобобових культур (табл. 10.21). Яка тенденція у зміні валового збору зернових та зернобобових культур? Мабуть, виразно відповісти на це питання важко, якщо оцінювати зміну валового збору за перші десять років.

Стосовно вихідних даних отримуємо 22 середні. Підрахуємо суму валового збору за перші п'ять років (1995–1999 рр.). Вона стано-

Таблиця 10.21

**Валовий збір зернових і зернобобових в Україні за 1995–2020 рр.,** млн т

| Рік | Валовий збір | 5-річна ковзна сума | 5-річна ковзна середня | Рік | Валовий збір | 5-річна ковзна сума | 5-річна ковзна середня |
|---|---|---|---|---|---|---|---|
| 1995 | 33,9 | – | – | 2007 | 29,3 | 163,6 | 40,2 |
| 1996 | 24,6 | – | – | 2008 | 53,3 | 196,7 | 40,4 |
| 1997 | 35,5 | – | 29,0 | 2009 | 46,0 | 200,9 | 44,9 |
| 1998 | 26,5 | – | 27,1 | 2010 | 39,3 | 202,2 | 48,3 |
| 1999 | 24,6 | 145,1 | 30,1 | 2011 | 56,7 | 224,6 | 50,3 |
| 2000 | 24,4 | 135,6 | 30,8 | 2012 | 46,2 | 241,5 | 53,8 |
| 2001 | 39,7 | 150,7 | 29,5 | 2013 | 63,1 | 251,3 | 58,0 |
| 2002 | 38,8 | 154 | 33,0 | 2014 | 63,8 | 269,1 | 59,9 |
| 2003 | 20,2 | 147,7 | 35,7 | 2015 | 60,1 | 289,9 | 63,0 |
| 2004 | 41,8 | 164,9 | 34,6 | 2016 | 66,1 | 299,3 | 64,4 |
| 2005 | 38,0 | 178,5 | 32,7 | 2017 | 61,9 | 315,0 | 66,7 |
| 2006 | 34,3 | 173,1 | 39,3 | 2018 | 70,1 | 322,0 | 67,6 |
| | | | | 2019 | 75,1 | 333,3 | – |
| | | | | 2020 | 64,9 | 338,1 | – |

вить 145,1 млн т (33,9+24,6+35,5+26,5+24,6), і розділимо її на 5, внаслідок чого отримуємо середню 29,0 млн т. Цю середню відносимо до року, що стоїть посередині взятих 5 років, тобто до 1997 р.

Далі, опускаючись у ряду однією ряд, обчислимо середню для наступної п'ятирічки (1996–2000 рр.). Вона становитиме 27,1 млн т $\left(\frac{24{,}6+35{,}5+26{,}5+24{,}6+24{,}4}{5}\right)$. Цю середню відносимо до відповідного серединного року — 1998 р. Потім обчислимо середню для ряду динаміки 1997–2001 рр. Її величина становитиме 30,1 млн т $\left(\frac{35{,}5+26{,}5+24{,}6+24{,}4+39{,}7}{5}\right)$ і т. д. до кінця ряду. Таким чином, отримаємо ряд середніх за всі п'ятиріччя, які наведені в табл. 10.21. У четвертому стовпці таблиці вийшла низка ковзних середніх, яка показує, що валовий збір 1996–2000 рр. дещо скоротився, а починаючи з 2000 р. знову став зростати, причому в останні роки більш швидкими темпами.

Специфічний для даного явища характер коливань рівнів ряду можна бачити з графічного представлення  вихідних (фактичних) даних (рис. 10.8). Аналіз даних дозволяє зробити висновок, що ковзна середня дає більш менш плавне зміна рівнів.

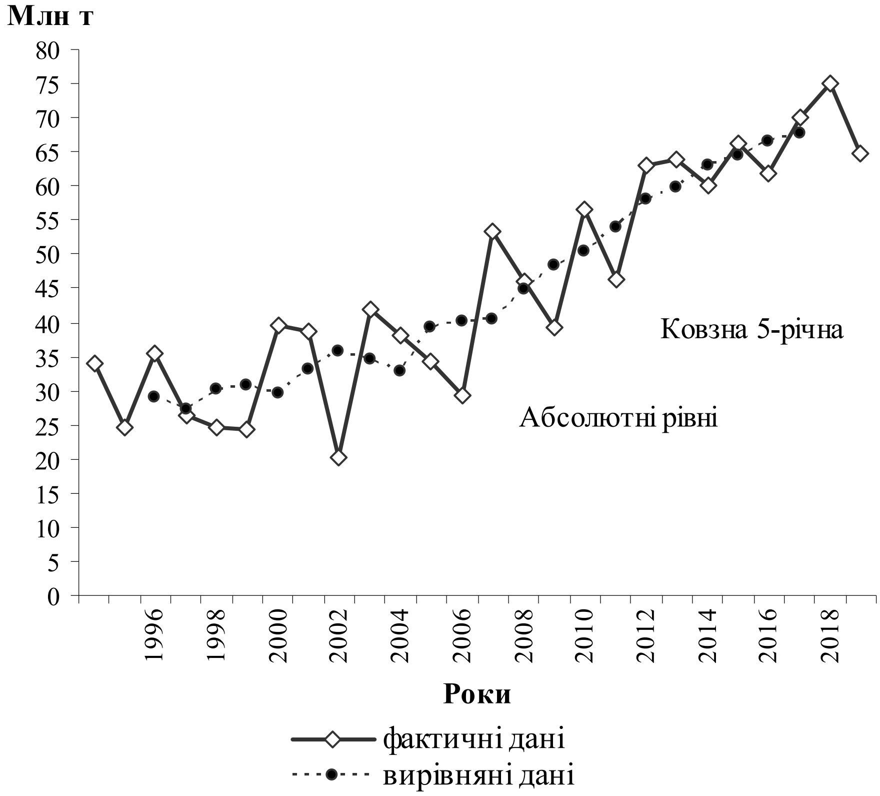



*Рис. 10.8.* Згладжування ряду динаміки валового збирання зернових та зернобобових культур в Україні ковзною середньою.

Виявлення основний тенденції ряду динаміки, що складається з парного рівня, має особливості. Знаходження ковзної середній по парному числу членів ускладнюється тим, що ковзні середні припадають, строго кажучи, на проміжок між двома датами, що знаходяться в середині інтервалу ковзання. При виявленні основної тенденції по парному числу рівнів ряду по-особливому укрупнюються інтервали часу. При парному числі членів періоду знайдені ковзаючі нецентровані середні центрують, тобто знаходять із них середню арифметичну. Спочатку обчислюють середні величини з відкиданням при обчисленні кожної нової середньої одного рівня ряду зліва і приєднанням одного рівня праворуч. Потім з цих середніх знаходять потім ковзні двочленні середні, які вже припадають на певні дати. Цей прийом називається центруванням. Він знайшов широке поширення у тривалих рядах динаміки.

Розглянемо порядок виявлення основної тенденції ряду динаміки, що складається з парного рівня ряду.

Наприклад, є такі дані про товарообіг продуктами харчування магазинами району (виручка; тис. грн):

| Квартал | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|
| I | 178 | 245 | 415 | 421 |
| II | 261 | 284 | 435 | 443 |
| III | 324 | 367 | 456 | 481 |
| IV | 296 | 340 | 397 | 462 |

Стосовно вихідних даних отримуємо тринадцять середніх:

$$\text{перша } \bar{y}_1 = \frac{175+261+324+296}{4} = 264{,}75;$$

$$\text{друга } \bar{y}_2 = \frac{261+324+296+245}{4} = 281{,}5;$$

$$\text{третя } \bar{y}_3 = \frac{324+296+245+284}{4} = 287{,}25;$$

. . . . . . . . . . . . . . . . . . . . .

$$\text{тринадцята } \bar{y}_{13} = \frac{421+443+481+462}{4} = 451{,}75.$$

Особливість згладжування по парному числу рівнів полягає у тому, що кожна із обчислених середніх відноситься до відповідних проміжків між двома суміжними кварталами. Так, перша середня ($\bar{y}_1$=264,75) відноситься до проміжку між II и III кварталами 2018 р., друга середня ($\bar{y}_2$=281,5) — до проміжку між III и IV кварталами.

На основі отриманих середніх необхідно провести центування. Для визначення згладженого середнього рівня III кварталу 2018 р. проведемо центрування першої середньої $\bar{y}_1$ і другої середньої $\bar{y}_2$:

$$\bar{y}_{c\,\text{III кв.}} = \frac{\bar{y}_1 + \bar{y}_2}{2} = \frac{264{,}75+281{,}5}{2} = 273{,}13.$$

Для визначення згладженого середнього рівня IV кварталу 2018 р. проведемо центрування другої $\bar{y}_2$ і третьої середньої $\bar{y}_3$

$$\bar{y}_{c\,\text{IV кв.}} = \frac{\bar{y}_2 + \bar{y}_3}{2} = \frac{281{,}5+287{,}25}{2} = 284{,}38.$$

Порядок розрахунку представимо в табл. 10.22 згладжування ряду по чотирьохчленній змінній (ковзній) середній. На основі проведених розрахунків видно досить чітку тенденцію зростання товарообігу за 2018–2021 рр.

Специфікою способу згладжування рядів динаміки і те, що отримані середні не дають теоретичних рядів, основу яких лежала б математично виражена закономірність.

Таблиця 10.22

| Період | Висхідні рівні | Середня із суми чотирьох рівнів ряду | Згладжений середній рівень (з центруванням, $\bar{y}_c$) |
|---|---|---|---|
| | 178 | – | |
| 2018 р. I, II | 261 | – | |
| | | 1059:4=264,75 | |
| III | 324 | | 273,13 |
| IV | | 1126:4=281,5 | |
| | 296 | | 284,38 |
| | | 1149:4=287,25 | |
| | 245 | | 292,63 |
| 2019 р. I | | 1192:4=298,0 | |
| II | 284 | | 303,5 |
| | | 1236:4=309,0 | |
| III | 367 | | 330,25 |
| IV | | 1406:4=351,5 | |
| | 340 | | 370,38 |
| | | 1557:4=289,25 | |
| | 415 | | 400,38 |
| 2020 р. I | | 1646:4=411,5 | |
| II | 435 | | 418,63 |
| | | 1703:4=425,75 | |
| III | 456 | | 426,5 |
| IV | | 1709:4=427,25 | |
| | 397 | | 428,25 |
| | | 1717:4=429,25 | |
| | 421 | | 432,38 |
| 2021 р. I | | 1742:4=435,5 | |
| II | 443 | | 443,63 |
| | | 1807:4=451,75 | |
| III | 481 | | |
| IV | | – | |
| | 462 | – | |

При використанні ковзної середньої згладжений ряд у нас виявився коротшим фактичного числа членів ряду. Якщо середня розраховується за трьома показниками, то ряд виходить коротшим на два члени (один зникає спочатку, інший — з кінця ряду), за п’ятьма показниками ряд виходить коротшим на чотири члени (два з початку та два з кінця) тощо.

При застосуванні методу ковзної середньої значення має вибір періоду або інтервалу ковзання. Він повинен відповідати періоду коливань в даному динамічному ряду. Якщо, наприклад, є динамічний ряд місячних даних, можна припустити, що періодичність повторюється через рік, і тому період ковзання необхідно взяти 12-місячний. Якщо періодичність коливань встановлена в 6 місяців, то береться 6-членна ковзна середня.

Метод ковзної середньої в аналітичній роботі застосовується досить часто і практично є одним з найпоширеніших методів виявлення основної тенденції розвитку соціально-економічних явищ і процесів. Застосування цього способу дозволяє згладити періодичні та випадкові коливання і тим самим виявити існуючу тенденцію у розвитку. Завдяки усередненню фактичних даних загальна тенденція явища виражається у вигляді плавної лінії. За допомогою методу ковзної середньої мається можливість виявити тенденції в розвитку краще, ніж при укрупненні інтервалів або ступінчатої середньої.

Ковзна середня *має недоліки*: 1) неможливість одержання рівнів для кінців ряду, що згладжується (згладжений ряд «укорочується» в порівнянні з фактичним на $(n-1)/2$ члена з одного та іншого кінця (під $n$ мається на увазі число членів, за якими розраховується ковзна середня); 2) довільність вибору інтервалу згладжування.

Метод ковзної середньої передбачає, що коливання, які бажано виділити, погашають один одного на досліджуваному відрізку часу. Очевидно, що такий ефект може бути отриманий лише за порівняно дуже нескладної структури ряду та його коливань. Метод ковзних середніх дає хороші результати в динамічних рядах з лінійною тенденцією. У хвилеподібних циклових рядах цей метод призводить нерідко до викривлень основної тенденції.

## 10.7. Аналітичне вирівнювання ряду

Більш ефективним способом виявлення основної тенденції розвитку явлень в рядах динаміки є вирівнювання рівнів динамічного ряду за способом найменших квадратів (*аналітичне вирівнювання*). За допомогою цього способу здійснюється перетворення динамічного ряду, встановлюються його основні закономірності зміни (тренди).

Характер зміни рівнів динамічного ряду може бути різним. Задача аналітичного вирівнювання є знаходженням простої математичної формули, яка дозволяє вирахувати теоретичні рівні. Основна вимога, що пред'являється до формули, складається в тому, що рівні, обчисленні за нею, повинні досить точно відновити загальну тенденцію розвитку явища. Знаходження цієї формули називається *аналітичним вирівнюванням*, що представляє собою не що інше, як опис емпіричних даних у вигляді математичної формули (рівнянням). Для опису характеру зміни рядів динаміки на практиці використовують обмежену кількість типів формул: прямої лінії, параболи другого або третього порядку, гіперболи та степеневої функції.

В основі вибору типу функції, за якою проводиться аналітичне вирівнювання рядів динаміки, лежить характер змін їх абсолютних приростів або рівнів.

**Вирівнювання по прямій.** На практиці найбільше поширення отримало аналітичне вирівнювання рядів динаміки по рівнянню прямої. Аналітичне рівняння прямої лінії має вид:

$$y_t = a_0 + a_1 t, \tag{10.38}$$

де $y_t$ — рівень результативної ознаки, знайдений за рівнянням;
$a_0$ і $a_1$ — параметри рівняння, які при застосуванні способу найменших квадратів знаходяться з рішення системи нормальних рівнянь;
$t$ — показник часу або інший аргумент.

Показник часу (дні неділі, декади, місяці, роки і т. д.) позначається порядковими номерами, починаючи з одиниць, двох і т. д.

Наприклад:

| Рік | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|
| $t$ | 1 | 2 | 3 | 4 | 5 |

Параметри рівняння прямої $a_0, a_1$ визначаються шляхом вирішення системи нормальних рівнянь, отриманих за методом найменших квадратів:

$$\begin{cases} a_0 n + a_1 \sum t = \sum y; \\ a_0 \sum t + a_1 \sum t^2 = \sum ty, \end{cases} \tag{10.39}$$

де $y$ — висхідні (емпіричні) рівні ряду динаміки;
$n$ — число членів ряду.

Розглянемо застосування методу аналітичного вирівнювання за прямою для вираження основної тенденції.

Вирівняємо ряд виробництва м’яса (у забійній вазі) за 2008–2022 рр. (табл. 10.23). Попередній аналіз цього ряду показує, що рівень його щорічно зростає приблизно на одне і те саме число, (стабільний абсолютний приріст або швидкість) і тому вирівнювати його потрібно

Таблиця 10.23

| Роки | Виробництво м'яса (у забійній вазі), тонн | Роки | Виробництво м'яса (у забійній вазі), тонн |
|---|---|---|---|
| 2008 | 710 | 2016 | 750 |
| 2009 | 730 | 2017 | 770 |
| 2010 | 720 | 2018 | 760 |
| 2011 | 740 | 2019 | 780 |
| 2012 | 730 | 2020 | 770 |
| 2013 | 760 | 2021 | 790 |
| 2014 | 740 | 2022 | 800 |
| 2015 | 750 | | |

за прямою.

Для вирівнювання всіх необхідних вихідних даних будуємо розрахункову таблицю (см. табл. 10.24).

Складаємо систему нормальних рівнянь:

$$\begin{cases} 15a_0 + 120a_1 = 11300 \\ 120a_0 + 1240a_1 = 91920 \end{cases} \quad \left| \begin{array}{c} \\ 1/8 \end{array} \right|$$

Урівнюємо коефіцієнти при $a_0$. Для цього поділимо друге рівняння на 8 $\left(\frac{120}{15}\right)$ і віднімемо з другого рівняння перше:

$$\begin{array}{c} \begin{cases} 15a_0 + 155a_1 = 11490 \\ — \\ 15a_0 + 120a_1 = 11300 \end{cases} \\ \hline 35a_1 = 190. \end{array}$$

Отже, $a_1$=5,42857≈5,43.

Із першого рівняння знаходимо $a_0$:

$$15a_0 + 120 \cdot 5{,}43 = 11300;$$

$$15a_0 = 10648{,}4$$

$$a_0 = \frac{10\,608{,}4}{15} = 709{,}8933 = 709{,}9.$$

Перевіримо правильність отриманих значень $a_0$ і $a_1$:

$$15 \cdot 709{,}9 + 120 \cdot 5{,}43 = 10648{,}5 + 651{,}6 = 11300{,}1;$$

$$120 \cdot 709{,}9 + 1240 \cdot 5{,}43 = 85188 + 6733{,}2 = 91921{,}2.$$

Таким чином, параметри рівняння визначені точно; відмінність

Таблиця 10.24

| Роки | Виробництво м’яса, т ($y$) | $t$* | $t^2$ | $t \cdot y$ | $\bar{y}_t$=709,9+ +5,43$t$ | $y-\bar{y}_t$ |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2008 | 710 | 1 | 1 | 710 | 715,3 | –5,3 |
| 2009 | 730 | 2 | 4 | 1 460 | 720,7 | +9,3 |
| 2010 | 720 | 3 | 9 | 2 160 | 726,2 | –6,2 |
| 2011 | 740 | 4 | 16 | 2 960 | 731,6 | +8,4 |
| 2012 | 730 | 5 | 25 | 3 650 | 737,0 | –7,0 |
| 2013 | 760 | 6 | 36 | 4 560 | 742,5 | +17,5 |
| 2014 | 740 | 7 | 49 | 5 180 | 747,9 | –7,9 |
| 2015 | 750 | 8 | 64 | 6 000 | 753,3 | –3,3 |
| 2016 | 750 | 9 | 81 | 6 750 | 758,8 | –8,8 |
| 2017 | 770 | 10 | 100 | 7 700 | 764,2 | +5,8 |
| 2018 | 760 | 11 | 121 | 8 360 | 769,6 | –9,6 |
| 2019 | 780 | 12 | 144 | 9 360 | 775,1 | +4,9 |
| 2020 | 770 | 13 | 169 | 10 010 | 780,5 | –10,5 |
| 2021 | 790 | 14 | 196 | 11 060 | 785,9 | +4,1 |
| 2022 | 800 | 15 | 225 | 12 000 | 791,4 | +8,6 |
| Всього | 11 300 | 120 | 1 240 | 9 1920 | 11 300,0 | 0 |

* Для заповнення колонки 3 зі значеннями $t$ здійснимо нумерацію дат. Так, якщо б у прикладі були дні неділі (понеділок, вівторок і т. д.), ми, прибігаючи до нумерації, вели б відлік з першої дати, прийнятою рівним одиницею, двом і т. д.

** Підсумок стовпця 3 можна знайти без сумування за формулою:

$$\sum t = \frac{n(n+1)}{2}.$$

Перевіримо наш приклад:

$$\sum t = \frac{15(15+1)}{2} = 120.$$

між підсумковими значеннями виникло із-за округлень.

Після визначення $a_0$ і $a_1$ складаємо аналітичне рівняння:

$$y_t = 709{,}9 + 5{,}43 \cdot t.$$

Підставляючи значення $t$, отримаємо:

$$y_1 = 733{,}5 + 59{,}9 \cdot 1 = 793{,}4;$$

$$y_2 = 733{,}5 + 59{,}9 \cdot 2 = 853{,}3;$$

$$y_3 = 733{,}5 + 59{,}9 \cdot 3 = 913{,}2,$$

і т. д. (див. колонку 6 табл. 10.24).

Отримані вирівняні рівні, а також фактичні дані наносимо на графік (рис. 10.9).

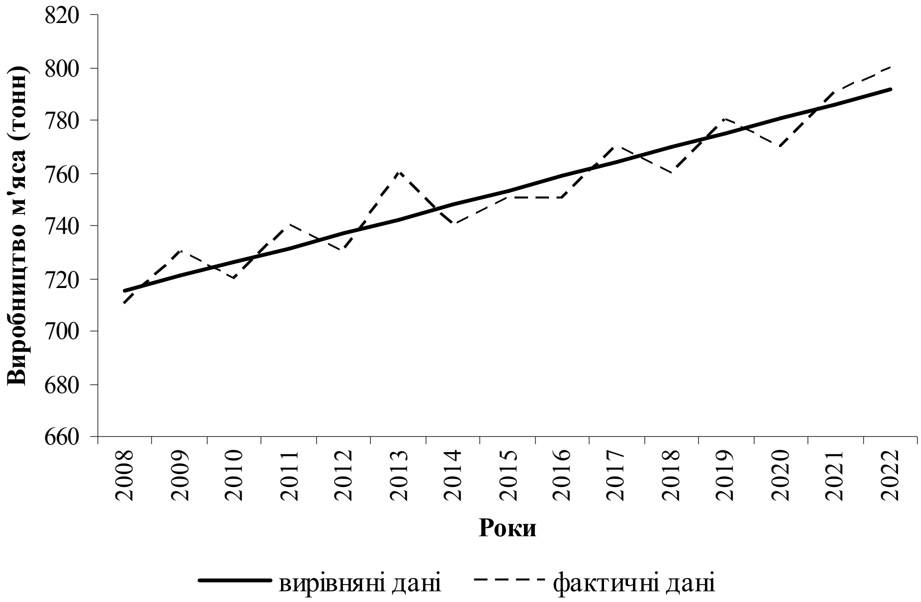



*Рис. 10.9.* Фактичні та вирівняні дані виробництва м’яса за 2008–2022 рр.

**Спрощений метод вирівнювання по прямої лінії**. У рядах динаміки техніка розрахунку параметрів рівняння по прямій лінії може бути спрощена. Для цієї мети показникам часу *t* надають такі значення, щоб їх сума була рівною нулю, тобто $\sum t = 0$. По ліву сторону значення часу приймаються негативними, по праву — позитивними. При цьому розрізняють два випадки:

1) Коли число членів динамічного ряду непарне, то слід відраховувати *t* від середини ряду. При такому відліку серединне значення дати (або періоду) динамічного ряду приймається рівним 0, ранні дати мають негативні значення (–1; –2; –3 і т. д.), а пізні дати — позитивні значення (1; 2; 3 і т. д.);

2) Коли число членів ряду парне, то в цьому випадку для дотримання вимоги про рівні інтервали між усіма значеннями *t* і в тому, щоб сума всіх значень *t* дорівнювала нулю, підбір проводиться так: знаходиться серединна пара дат (або періодів), і значення для неї набувають – 1 і + 1, а далі йдуть нагору – 3; –5; –7 і т. д., і вниз +3; + 5; +7 і т. д. В обох випадках $\sum t = 0$.

За умови, що $\sum t = 0$, спрощена *система нормальних рівнянь* набуде наступного вигляду:

$$na_0 = \sum y;$$

$$a_1 \sum t^2 = \sum ty, \tag{10.40}$$

Звідки:

$$a_0 = \frac{\sum y}{n}; \qquad a_1 = \frac{\sum yt}{\sum t^2}. \tag{10.41}$$

Для фактичного використання цього методу розглянемо наступний приклад. Вихідні дані наведено у табл. 10.24.

Стосовно даного прикладу, в якому число вихідних (емпіричних) рівнів ряду — непарне ($n$=15), це виконується за наступних позначень років:

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
| –7 | –6 | –5 | –4 | –3 | –2 | –1 | 0 | +1 | +2 | +3 | +4 | +5 | +6 | +7 |

З наведених формул видно, що знаходження параметрів рівняння прямої необхідно знайти величини $\sum y$, $\sum t^2$ і $\sum ty$. Для обчислення всіх необхідних вихідних даних збудуємо розрахункову таблицю (див. табл. 10.25).

Підставивши знайдені значення формулу для $a_0$ і $a_1$, отримуємо:

$$a_0 = \frac{\sum y}{n} = \frac{11300}{15} = 753{,}3333 \approx 753{,}33;$$

$$a_1 = \frac{\sum yt}{\sum t^2} = \frac{1520}{280} = 5{,}42857 \approx 5{,}43.$$

В результаті отримуємо наступне рівняння основної тенденції виробництва м’яса за 2008–2022 рр.:

$$y_t = 753{,}33 + 5{,}43t.$$

Підставляючи рівняння $y_t$=753,33+5,43$t$ прийняті позначення $t$, обчислимо вирівняні (теоретичні) рівні низки динаміки:

2008 р. $y_{t=-7}$=753,33+5,43 (–7)=715,3;

2009 г. $y_{t=-6}$=753,33+5,43 (–6)=720,8;

2018 г. $y_{t=-5}$=753,33+5,43 (–5)=726,2,

і т. д. (див. колонку 6 табл. 10.25).

Для перевірки розрахунку значень $y_i$ використовується формула

$$\sum y_i = \sum y_t. \tag{10.42}$$

Таблиця 10.25

| Роки | Виробництво м'яса, т ($y$) | $t$ | $t^2$ * | $ty$ | $\bar{y}_t$=753,33+ 5,43$t$ | $y-\bar{y}_t$ | $(y-\bar{y}_t)^2$ |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2008 | 710 | –7 | 49 | –4 970 | 715,3 | –5,3 | 28,09 |
| 2009 | 730 | –6 | 36 | –4 380 | 720,8 | +9,2 | 84,64 |
| 2010 | 720 | –5 | 25 | –3 600 | 726,2 | –6,2 | 38,44 |
| 2011 | 740 | –4 | 16 | –2 960 | 731,6 | +8,4 | 70,56 |
| 2012 | 730 | –3 | 9 | –2 190 | 737,0 | –7,0 | 49,00 |
| 2013 | 760 | –2 | 4 | –1 520 | 742,5 | +17,5 | 306,25 |
| 2014 | 740 | –1 | 1 | –740 | 747,9 | –7,9 | 62,41 |
| 2015 | 750 | 0 | 0 | 0 | 753,3 | –3,3 | 10,89 |
| 2016 | 750 | 1 | 1 | 750 | 758,8 | –8,8 | 77,44 |
| 2017 | 770 | 2 | 4 | 1 540 | 764,2 | +5,8 | 33,64 |
| 2018 | 760 | 3 | 9 | 2 280 | 769,6 | –9,6 | 92,16 |
| 2019 | 780 | 4 | 16 | 3 120 | 775,1 | +4,9 | 24,01 |
| 2020 | 770 | 5 | 25 | 3 850 | 780,5 | –10,5 | 110,25 |
| 2021 | 790 | 6 | 36 | 4 740 | 785,9 | +4,1 | 16,81 |
| 2022 | 800 | 7 | 49 | 5 600 | 791,3 | +8,7 | 75,69 |
| Всього | 11 300 | 0 | 280 | 1 520 | 11 300,0 | 0 | 1 080,30 |

* Математично доведено, що в таких рядах $\sum t^2 = \frac{n(n^2-1)}{12}$.

У прикладі $\sum y_i = 11300 = \sum y_t$; отже, значення $y_t$ визначені правильно. Збіг підсумків фактичних даних (графа 2) і теоретичних рівнів (графа 6) свідчить про правильність проведених обчислень. Зіставлення графи 2 і графи 6 таблиці за кожним роком показує незначні відхилення розрахункових рівнів від фактичних, що підтверджує правильність вибору математичного рівняння.

Слід зазначити, що у разі вирівнювання по прямій лінії спрощеним методом параметр $a_1$ при непарному числі членів періоду продовжує грати ту ж роль, що і при звичайному вирівнюванні. Якщо при звичайному вирівнюванні $a_0$ — це початковий рівень, то при спрощеному методі вирівнювання — це середній рівень явища.

**Вирівнювання по параболі другого порядку** проводиться у рядах динаміки, рівень яких змінюється при рівномірному прискоренні (стабільному прискоренні) чи уповільненні. Більше поширення вирівнювання по параболі другого порядку отримало в практиці економічних досліджень, оскільки загальна тенденція розвитку в рядах динаміки економічних явищ майже ніколи не відповідає параболам вищого порядку. Крім того, вирівнювання по параболі вищого порядку відрізняється порівняно більшою трудомісткістю обчислень.

Рівняння параболи другого порядку має такий вигляд:

$$\overline{y}_t = a_0 + a_1 t, + a_2 t. \qquad (10.43)$$

Покажемо вирівнювання емпіричних даних по рівнянню параболи другого порядку на прикладі поквартальних даних про продаж овочів у торговельній мережі міста *N* за шість років.

Поквартальні дані продажу овочів, складені в ряд (графа 3 табл. 10.26) та особливо графічне зображення (рис. 10.10) виявляють збільшення продажу овочів з тенденцією до уповільнення. Тому краще для виявлення характеру зміни явища використовувати рівняння параболи другого порядку.

Обчислення параметрів рівняння здійснюється за способом найменших квадратів. Параметри параболи другого порядку $a_0$, $a_1$ і $a_2$ знаходяться з наступної системи рівнянь:

$$\left.\begin{aligned} a_0 n + a_1 \sum t + a_2 \sum t^2 &= \sum y; \\ a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3 &= \sum yt; \\ a_0 \sum t^2 + a_1 \sum t^3 + a_2 \sum t^4 &= \sum yt^2. \end{aligned}\right\} \qquad (10.44)$$

Для обчислення параметрів рівняння параболи складемо розрахункову таблицю (табл. 10.26).

Відрізки часу, умовно позначені значеннями *t*, представляють собою ряд послідовних чисел, починаючи з одиниці і закінчуючи *n*-м значенням, де *n* — число членів безперервного ряду динаміки. Для обчислення параметрів рівняння необхідно визначити значення $\sum t, \sum t^2, \sum t^3, \sum t^4.$ Ці значення можна визначити за формулами:

$$\sum t = \frac{n(n+1)}{2} = \frac{24(24+1)}{2} = 300,$$

$$\sum t^2 = \frac{2n+1}{3} = \frac{2 \cdot 24 + 1}{3} = 4\,900,$$

$$\sum t^3 = \left(\sum t\right)^2 = 300^2 = 90\,000,$$

$$\sum t^4 = \frac{3n^2 + 3n - 1}{5} \cdot \left(\sum t^2\right) = \frac{3 \cdot 24^2 + 3 \cdot 24 - 1}{5} \cdot 4900 =$$

$$= \frac{3 \cdot 576 + 72 - 1}{5} \cdot 4900 = \frac{1728 + 71}{5} \cdot 4900 = 1\,763\,020.$$

Таблиця 10.26

**Розрахунок параметрів рівняння параболи другого порядку**

| Роки, квартали | | $t$ | $y$ (млн грн) | $ty$ | $t^2$ | $t^2y$ | $t^3$ | $t^4$ | $y_t$ | $y-y_t$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1-й год | I | 1 | 23,4 | 23,4 | 1 | 23,4 | 1 | 1 | 24,3 | –0,9 |
| | II | 2 | 24,8 | 49,6 | 4 | 99,2 | 8 | 16 | 24,8 | 0 |
| | III | 3 | 26,7 | 80,1 | 9 | 240,3 | 27 | 81 | 25,2 | +1,5 |
| | IV | 4 | 26,4 | 105,6 | 16 | 422,4 | 64 | 256 | 25,6 | +0,8 |
| 2-й год | I | 5 | 24,1 | 120,5 | 25 | 602,5 | 125 | 625 | 25,9 | –1,8 |
| | II | 6 | 24,8 | 148,8 | 36 | 892,8 | 216 | 1296 | 26,2 | –1,4 |
| | III | 7 | 28,2 | 197,4 | 49 | 1381,8 | 343 | 2401 | 26,4 | +1,8 |
| | IV | 8 | 25,6 | 204,8 | 64 | 1638,4 | 512 | 4096 | 26,5 | –0,9 |
| 3-й год | I | 9 | 23,7 | 213,3 | 81 | 1919,7 | 729 | 6561 | 26,6 | –2,9 |
| | II | 10 | 30,1 | 301 | 100 | 3010,0 | 1000 | 10000 | 26,7 | +3,4 |
| | III | 11 | 29,5 | 324,5 | 121 | 3569,5 | 1331 | 14641 | 26,6 | +2,9 |
| | IV | 12 | 26,1 | 313,2 | 144 | 3758,4 | 1728 | 20736 | 26,5 | –0,4 |
| 4-й год | I | 13 | 25,3 | 328,9 | 169 | 4275,7 | 2197 | 28561 | 26,4 | –1,1 |
| | II | 14 | 25,1 | 351,4 | 196 | 4919,6 | 2744 | 38416 | 26,2 | –1,1 |
| | III | 15 | 30,7 | 460,5 | 225 | 6907,5 | 3375 | 50625 | 25,9 | +4,8 |
| | IV | 16 | 25,2 | 403,2 | 256 | 6451,2 | 4096 | 65536 | 25,6 | –0,4 |
| 5-й год | I | 17 | 22,8 | 387,6 | 289 | 6589,2 | 4913 | 83521 | 25,2 | –2,4 |
| | II | 18 | 22,5 | 405 | 324 | 7290,0 | 5832 | 104976 | 24,7 | –2,2 |
| | III | 19 | 23,9 | 454,1 | 361 | 8627,9 | 6859 | 130321 | 24,2 | –0,3 |
| | IV | 20 | 24,1 | 482 | 400 | 9640,0 | 8000 | 160000 | 23,7 | +0,4 |
| 6-й год | I | 21 | 20,3 | 426,3 | 441 | 8952,3 | 9261 | 194481 | 23,0 | –2,7 |
| | II | 22 | 20,5 | 451 | 484 | 9922,0 | 10648 | 234256 | 22,4 | –1,9 |
| | III | 23 | 26,4 | 607,2 | 529 | 13965,6 | 12167 | 279841 | 21,6 | +4,8 |
| | IV | 24 | 20,8 | 499,2 | 576 | 11980,8 | 13824 | 331776 | 20,8 | 0 |
| Всього | | 300 | 601 | 7338,6 | 4900 | 117080 | 90000 | 1763020 | 601,0 | 0 |

Підставимо до системи нормальних рівнянь числові значення з табл. 10.26, тоді рівняння набудуть наступного вигляду:

$$24a_0 + 300a_1 + 4900a_2 = 601;$$
$$300a_0 + 4900a_1 + 90000a_2 = 7338,6;$$
$$4900a_0 + 90000a_1 + 1763020a_2 = 117080.$$

Найбільш зручним, правильним і простим способом обчислення параметрів рівняння параболи будь-якого порядку і складніших систем нормальних рівнянь є застосування кратності рівнянь на абсолютну величину при одному з невідомих. Розглянемо методику обчислення параметрів рівнянь.

Розділимо кожне рівняння на абсолютну величину при параметрі $a_0$ і отримаємо в результаті нову систему рівнянь:

$$a_0 + 12{,}5a_1 + 204{,}18a_2 = 25{,}04;$$
$$a_0 + 16{,}333a_1 + 300a_2 = 24{,}46;$$
$$a_0 + 18{,}37a_1 + 359{,}8a_2 = 23{,}89.$$

Віднімемо від другого рівняння перше і отримаємо:

$$\begin{array}{ll} & \text{II. } 16{,}333a_1 + 300a_2 = 24{,}46 \\ - & \\ & \text{I. } a_0 + 12{,}5a_1 + 204{,}18a_2 = 25{,}04 \\ \hline A & 3{,}833a_1 + 95{,}82a_2 = -0{,}58. \end{array}$$

Віднімемо з третього рівняння друге та отримаємо:

$$\begin{array}{ll} & \text{III. } a_0 + 18{,}37a_1 + 359{,}8a_2 = 23{,}89 \\ - & \\ & \text{II. } a_0 + 16{,}333a_1 + 300a_2 = 24{,}46 \\ \hline Б & 2{,}037a_1 + 59{,}8a_2 = -0{,}57. \end{array}$$

В результаті отримаємо два рівняння із двома невідомими:

$$3{,}833a_1 + 95{,}82a_2 = -0{,}58.$$
$$2{,}037a_1 + 59{,}8a_2 = -0{,}57.$$

Кожне з цих рівнянь ділимо на абсолютну величину при параметрі $a_1$

$$a_1 + 25{,}0a_2 = -0{,}151;$$
$$a_1 + 29{,}36a_2 = -0{,}28;$$

віднімемо з другого рівняння перше, отримаємо:

$$4{,}36a_2 = -0{,}129.$$

Звідси

$$a_2 = -\frac{0{,}129}{4{,}36} = -0{,}0296.$$

Підставимо параметр $a_2$ в рівняння з двома невідомими, визначаємо параметр $a_1$:

$$a_1 = -0{,}28 - 29{,}36 \cdot (-0{,}0296) = 0{,}5891.$$

Потім підставимо параметри $a_1$ і $a_2$ в рівняння з трьома невідомими, обчислимо параметр $a_0$

$$a_0 + 18{,}37 \cdot (0{,}5891) + 359{,}8 \cdot (-0{,}0296) = 23{,}89.$$

Звідси

$$a_0 = 23{,}72.$$

Якщо підставити ці значення в рівняння параболи, воно прийме вид:

$$y_x = 23{,}72 + 0{,}5891t - 0{,}0296t^2.$$

Підставляючи в зазначене рівняння різні значення $t$ і $t^2$, обчислимо теоретичну лінію, що відображає загальну тенденцію розвитку даного ряду динаміки.

$$y_1 = 23{,}76 + 0{,}5811 \cdot 1 - 0{,}0293 \cdot 1^2 = 24{,}3;$$

$$y_2 = 23{,}76 + 0{,}5811 \cdot 2 - 0{,}0293 \cdot 2^2 = 24{,}8;$$

$$y_3 = 23{,}76 + 0{,}5811 \cdot 3 - 0{,}0293 \cdot 3^2 = 25{,}2 \text{ і т. д.}$$

Рівняння параболи другого порядку такого типу ($a_1$ — позитивне, $a_2$ — від'ємне) відображає збільшення рівнів ряду з тенденцією до уповільнення (обсяг продажу овочів у нашому прикладі збільшувався до певного періоду, а потім починає поступово знижуватися), а те, що параметр $a_0$ виражається малим числом, говорить про незначне відхилення параболічної кривої від прямої.

Порівняння фактичних даних (графа 3 табл. 10.26) з рядами, вирівняними по параболі другого порядку (графа 10 табл. 10.26, рис. 10.10) показує, що останні більш плавно відображають загальну тенденцію розвитку і не спотворює коливання цього ряду.


*Рис. 10.10.* Динаміка та загальна тенденція, обчислені за даними табл. 10.26 способом аналітичного вирівнювання за рівнянням параболи другого порядку.

Визначення параметрів рівняння параболи другого порядку із системи нормальних рівнянь є трудомістким процесом. Для деякого спрощення обчислень параметрів параболи другого порядку можуть бути використані наступні формули:

$$a_0 = \frac{\left[\sum t^2 \sum t^4 - (\sum t^2)^2\right]\sum y - \left(\sum t^2 \sum t^3 - \sum t \sum t^4\right)\sum yt + \left[\sum t \sum t^3 - (\sum t^2)^2\right]\sum t^2 y}{m\sum t^2 \sum t^4 + 2\sum t \sum t^2 \sum t^3 - n\left(\sum t^3\right)^2 - \left(\sum t\right)^2 \sum t^4 - \left(\sum t^2\right)^3},$$

$$a_1 = \frac{\left(\sum t^2 \sum t^3 - \sum t \sum t^4\right)\sum y + \left[n\sum t^4 - \left(\sum t^2\right)^2\right]\sum t + \left(\sum t \sum t^2 - n\sum t^3\right)\sum t^2 y}{m\sum t^2 \sum t^4 + 2\sum t \sum t^2 \sum t^3 - n\left(\sum t^3\right)^2 - \left(\sum t\right)^2 \sum t^4 - \left(\sum t^2\right)^3},$$

$$a_2 = \frac{\left[t\sum t^3 - \left(\sum t^2\right)^2\right]\sum y + \left(\sum t \sum t^2 - n\sum t^3\right)\sum yt + \left[n\sum t^2 - \left(\sum t\right)^2\right]\sum yt^2}{m\sum t^2 \sum t^4 + 2\sum t \sum t^2 \sum t^3 - n\left(\sum t^3\right)^2 - \left(\sum t\right)^2 \sum t^4 - \left(\sum t^2\right)^3}.$$

**Вирівнювання за показовою (експонентною) кривою**. Якщо рівні динамічного ряду змінюються з більш менш постійним відносним приростом (стабільний темп зростання), то *вирівнювання* такого ряду слід проводити *за показовою* (*експонентною*) *функцією*. Вирівнювання проводиться за такою формулою: $\overline{y}_t = a_0 a_1^t$, де $a_0$ і $t$ мають колишній зміст, $a_1$ — темп зростання за одиницю часу.

Прологарифмуємо показову криву, отримаємо рівняння прямої, в якому рівні та параметри замінені їх логарифмами:

$$\lg \overline{y}_t = \lg a_0 + t \lg a_1 . \tag{10.45}$$

Параметри $a_0$ і $a_1$ визначаються за методом найменших квадратів відхилень логарифмів шляхом вирішення системи нормальних рівнянь:

$$\begin{cases} n \cdot \lg a_0 + \lg a_1 \sum t = \sum \lg y \\ \lg a_0 \cdot \sum t + \lg a_1 \cdot \sum t^2 = \sum t \cdot \lg y \end{cases} \tag{10.46}$$

Техніка вирівнювання показової кривої, вираженої в логарифмах, аналогічна викладеної вище техніки вирівнювання по прямій, за винятком того, що вирівнюванню тут піддаються не самі члени ряду, а їх логарифми.

При способі відліку $t$ таким чином, щоб $\sum t$=0 ми отримуємо:

$$n \lg a_0 = \sum \lg y, \tag{10.47}$$

звідки

$$\lg a_0 = \frac{\sum \lg y}{n}; \qquad (10.47)$$

і

$$\lg a_1 \sum t^2 = \sum t \lg y, \qquad (10.48)$$

звідки

$$\lg a_1 = \frac{\sum t \lg y}{\sum t^2}. \qquad (10.49)$$

Вирівняємо низку виробництва соняшникової олії маслоекстраційними районами області за 2016–2020 рр. Первинний аналіз вихідних даних свідчить, що його рівні щорічно зростають приблизно на ту саму величину. Внаслідок чого вирівнювання ряду слідує за показовою кривою. Необхідні розрахунки зробимо в табл. 10.27.

Таблиця 10.27

| Рік | Виробництво соняшникової олії, тис. ($y$) | $\lg y$ | $t$ | $t^2$ | $t \cdot \lg y$ | $\bar{y}_t = 558{,}08 \cdot 1{,}07^t$ |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2016 | 490,1 | 2,69028 | –2 | 4 | –5,38056 | 487,4 |
| 2017 | 520,0 | 2,71600 | –1 | 1 | –2,71600 | 521,6 |
| 2018 | 557,7 | 2,74640 | 0 | 0 | – | 558,1 |
| 2019 | 596,7 | 2,77576 | +1 | 1 | 2,77576 | 597,1 |
| 2020 | 638,3 | 2,80502 | +2 | 7 | 5,61004 | 638,9 |
| Всього | 2 802,8 | 13,73346 | 0 | 10 | 0,28924 | – |

Отримаємо:

$$\lg a_0 = \frac{\sum \lg y}{n} = \frac{13{,}73346}{5} = 2{,}74669;$$

$$\lg a_1 = \frac{\sum t \lg y}{\sum t^2} = \frac{0{,}28924}{10} = 0{,}02892.$$

Звідки

$$a_0 \approx 558{,}08; \quad a_1 \approx 1{,}07.$$

Аналітичне рівняння показової кривої прийме вид:

$$\bar{y}_t = 558{,}08 \cdot 1{,}07^t.$$

Підставляємо значення $t$ в отримане рівняння показової кривої і получимо (див. колонку 7 табл. 10.27):

$$\bar{y}_{t=-2} = 558{,}08 \cdot 1{,}07^{-2} = \frac{558{,}08}{1{,}07^2} = \frac{558{,}08}{1{,}1449} \approx 487{,}4;$$

$$\bar{y}_{t=-1} = 558{,}08 \cdot 1{,}07^{-1} = \frac{558{,}08}{1{,}07} \approx 521{,}6;$$

$$\bar{y}_{t=-0} = 558{,}08 \cdot 1{,}07^{0} \approx 558{,}1;$$

$$\bar{y}_{t=1} = 558{,}08 \cdot 1{,}07^{1} = 558{,}08 \cdot 1{,}07 \approx 597{,}1;$$

$$\bar{y}_{t=2} = 558{,}08 \cdot 1{,}07^{2} = 558{,}08 \cdot 1{,}1449 \approx 638{,}9.$$

Тепер можна побудувати графік, зобразивши фактичні дані і вирівняні дані показової (експонентної) функції аналогічно тому, як це було показано при аналітичному вирівнюванні параболі другого порядку.

**Рівняння гіперболи**. Зворотний зв’язок вказує на спад результативної ознаки при зростанні факторіального. Такий лінійний зв’язок при від’ємному значенні. У ряді випадків зворотний зв’язок може бути виражений рівнянням гіперболи:

$$\bar{y}_x = a_0 + a_1 \frac{1}{x}. \tag{10.50}$$

Параметри рівняння гіперболи $a_0$ і $a_1$ цього рівняння знаходяться із системи нормальних рівнянь: $na_0 + a_1 \sum \frac{1}{x} = \sum y;$

$$\begin{cases} na_0 + a_1 \sum \frac{1}{x} = \sum y; \\ a_0 \sum \frac{1}{x} + a_1 \sum \left(\frac{1}{x}\right)^2 = \sum y \frac{1}{x}. \end{cases} \tag{10.51}$$

Розглянемо приклад розрахунку зворотного зв’язку за рівнянням гіперболи.

Маються дані щодо 10 однорідних підприємств:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Випуск продукції ($x$), тис. шт. | 1,2 | 0,61 | 0,08 | 0,36 | 0,3 | 0,41 | 0,16 | 0,1 | 0,26 | 0,71 |
| Витрати матеріалу на одиницю виробу ($y$), грн | 2,1 | 1,3 | 11,1 | 6,5 | 5,7 | 2,6 | 7,8 | 9,8 | 4,9 | 1,8 |

Для визначення параметрів рівняння гіперболи будуємо розрахункову таблицю (табл. 10.28).

Таблиця 10.28

| Номер п/п | Випуск продукції, тис. шт. ($x$) | Витрати матеріалу на одиницю продукції, грн ($y$) | $\frac{1}{x} = x_1$ | $x_1^2$ | $yx_1$ | $\bar{y}_x$ |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 38 | 11,0 | 0,0263 | 0,0006925 | 0,2895 | 11,2 |
| 2 | 45 | 10,1 | 0,0222 | 0,0004938 | 0,2244 | 10,3 |
| 3 | 60 | 8,9 | 0,0167 | 0,0002778 | 0,1483 | 9,1 |
| 4 | 75 | 8,6 | 0,0133 | 0,0001778 | 0,1147 | 8,3 |
| 5 | 90 | 8,7 | 0,0111 | 0,0001235 | 0,0967 | 7,8 |
| 6 | 110 | 7,7 | 0,0091 | 0,0000826 | 0,0700 | 7,4 |
| 7 | 150 | 6,7 | 0,0067 | 0,0000444 | 0,0447 | 6,9 |
| 8 | 225 | 6,4 | 0,0044 | 0,0000198 | 0,0284 | 6,4 |
| 9 | 300 | 5,8 | 0,0033 | 0,0000111 | 0,0193 | 6,1 |
| 10 | 400 | 5,5 | 0,0025 | 0,0000063 | 0,0138 | 5,9 |
| Всього | – | 79,4 | 0,1156 | 0,0019296 | 1,0498 | 79,4 |

У складену систему рівнянь підставимо дані табл. 10.28 і складемо рівняння для обчислення параметрів $a_0$ і $a_1$ рівняння гіперболи:

$$10a_0 + 0{,}1156\,a_1 = 79{,}4;$$
$$0{,}1156a_0 + 0{,}0019296a_1 = 1{,}0498.$$

Помножимо друге рівняння на 86,505 $\left(\frac{10}{0{,}1156}\right)$, отримаємо

$$10a_0 + 0{,}1669a_1 = 90{,}8.$$

Віднімемо з отриманого рівняння перше, маємо:

$$0{,}0513a_1 = 11{,}4.$$

Звідки

$$a_1 = \frac{11{,}4}{0{,}0513} = 222.$$

Підставимо значення $a_1$ у перше вихідне рівняння, отримуємо:

$$10a_0 + 0{,}1156 \cdot 222 = 79{,}4, \qquad 10a_0 = 53{,}7,$$

звідки

$$a_0 = 5{,}37.$$

Отже, рівняння зв’язку між обсягом виробництва та витратами матеріалу на одиницю продукції прийме такий вид:

$$y_x = 5{,}37 + 222\frac{1}{x}.$$

Підставивши у це рівняння конкретні значення $x$, знаходимо для 10 підприємств $\overline{y}_x$, тобто теоретичні значення витрат матеріалу на одиницю продукції (гр. 7 табл. 10.28). Відхилення $\overline{y}_x$ від $y$ незначні, що підтверджує гіпотезу про наявність зв'язку, що описується гіперболою. Підсумки $\overline{y}_x$ і $y$ збігаються (79,4 і 79,4), що свідчить про правильність розрахунків.

Обчислимо параметри рівняння можна за такими формулами:

$$a_1 = \frac{\sum yx_1 - \dfrac{\sum x_1 \sum y}{n}}{\sum {x_1}^2 - \dfrac{\left(\sum x_1\right)^2}{n}} = \frac{1{,}0498 - \dfrac{0{,}1156 \cdot 79{,}4}{10}}{0{,}0019296 - \dfrac{0{,}1156^2}{10}} = 222;$$

$$a_0 = \overline{y} - a_1\overline{x} = \frac{79{,}4}{10} - 222 \cdot \frac{0{,}1156}{10} = 5{,}37.$$

Показники гіперболи, отримані за формулами, відповідають показникам, отриманим, розрахунковим шляхом, що підтверджує правильність розрахунків.

## 10.8. Прийоми дослідження сезонних коливань

**Поняття про сезонність**. При аналізі квартальних та місячних даних рядів динаміки часто виявляються регулярні повторювані коливання явищ, пов'язаних зі змінною пори року (зима, весна, літо, осінь), явищами природи (період дощів, період зібраних рослин), виконанням певних робіт і видів зайнятості (сезон полювання, лікувальний сезон, сезон збирання врожаю) та ін. Такі коливання носять назву *сезонних*.

Вивчення сезонних коливань має самостійне значення при дослідженні особливого типу динаміки. Сезонність характеризується змінами складових динамічного ряду, що описує внутрішньогосподарські періодичні коливання. Такі коливання спостерігаються в будь-якій сфері діяльності людини.

Сезонність і сезонні коливання в народному господарстві визначаються як соціальними, так і природно-кліматичними умовами. Суть сезонності в сфері виробництва полягає в розриві між періодом виробництва і робочим періодом, чим більше цей розрив, тим вище показник сезонності. З сезонністю пов'язано нерівномірність виробничої діяльності в галузях, що пов'язані з переробкою сільськогосподарської сировини. Періодичні коливання на взуття та одяг також пов'язані в значній мірі з сезонністю.

У багатьох випадках сезонність негативно впливає на результати роботи сільськогосподарських підприємств, підприємств торгівлі, транспорту, зв'язку, будівництва, видобутку корисних копалин відкритим способом. Істотний вплив сезонність здійснює на процеси міграції населення, розвиток епідемій, повторюваності надзвичайних метеорологічних та гідрологічних явищ та ін. Збитки від сезонності пов'язані з нерівномірним використанням обладнання та робочої сили, з нерівномірним навантаженням транспорту, нерівномірною поставкою сиру для підприємств, пов'язаних як із сезонним виробництвом, так і з сезонними відростками. Тому від них намагаються звільнитися чи послабити свої негативні наслідки.

У процесі статистичного аналізу сезонності суспільних явищ вирішуються такі завдання: визначення величини сезонних коливань та встановлення їх сили; аналіз факторів, що визначають зміну сезонних коливань; оцінка наслідків, до яких наводиться нерівномірність змін явищ.

Типи колінь показників різноманітні, але їх можна об'єднати в такі групи: маятникові; циклічні довгострокові коливання; випадкові розподіли у часі коливання.

Для виявлення сезонних коливань зазвичай беруться дані за кілька років, які розподілені зазвичай по місяцях. Дані за кілька років (зазвичай за три роки) беруться, щоб усунути сезонні випадкові коливання, що виникли в окремі роки.

Розмах сезонних коливань місячних даних прийнято вимірювати спеціальними показниками, які називаються індексами сезонності.

*Індексом сезонності* називають показники інтенсивності сезонних коливань. У загальному вигляді він визначається як відношення кожного рівня ряду динаміки у вигляді помісячних (рідше поквартальних) даних до якогось теоретичного або середнього рівня, прийнятого в якості бази порівняння. Індекс сезонності зазвичай розраховується у відсотках.

У статистиці для розрахунку індексу сезонності застосовуються різні способи розрахунку. За своєю суттю всі методи діляться на дві групи. Суть методу першої групи полягає у попередньому визначенні та виключенні загальної тенденції розвитку у рядах динаміки та у наступному дослідженні та кількості вимірювань сезонності. Аналіз сезонності явищ методами першої групи слід проводити у тих рядах динаміки, рівень яких має виражену тенденцію збільшення (або зниження) протягом досліджуваного періоду. Якщо ряд має виражену тенденцію, то перш ніж вирахувати індекс сезонних коливань, емпіричні дані обробляють, щоб виявити загальну тенденцію зміни. Для цієї цілі прибігають до аналітичного вирівнювання. Після цього вираховують сезонну хвилю (індекс сезонності), але не від постійної (середньорічної) середньої, а від вирівняних даних. Індекс сезонності вираховують як середнє з цих значень. Формула для розрахунку буде наступною:

$$I_s = \left[\sum \frac{y_i}{y_t} \cdot 100\right] : n\,, \qquad (10.52)$$

де $I_s$ — індекс сезонності;

$y_i$ — фактичні рівні ряду динаміки;

$y_t$ — згладжені (теоретичні) рівні внутрішньорічних періодів;

$n$ — число років.

Своє найменування ці способи отримали від назв методів визначення загальної тенденції. До цієї групи належать такі методи: метод механічного вирівнювання, метод аналітичного вирівнювання та метод ковзної середньої.

До другої групи належать методи, які застосовуються для обчислення сезонності в рядах динаміки безпосередньо за емпіричним даними, без їх попередньої обробки. До цієї групи входять: метод простої середньої, метод відносних чисел і метод У. Персонса. Сутність цих методів полягає у визначенні сезонної хвилі (індексу сезонності) як відсоткового відношення середніх місячних (квартальних) середніх до загальної середньої за розглянутий період. У загальному вигляді розрахунок індексу сезонності здійснюють за наступною формулою:

$$I_s = \frac{y_i}{\bar{y}_0} \times 100, \qquad (10.53)$$

де $\bar{y}_0$ — середня величина з рівнів ряду динаміки.

**Розрахунок індексів сезонності у стабільних рядах динаміки**. Розглянемо техніку обчислення сезонної хвилі (індексів сезонності) у рядах динаміки, які мають різко вираженої тенденції до зростання чи спадання, коли внутрішньорічні коливання відбуваються навколо деякого постійного рівня ($y_0$).

Маються наступні дані про чисельність робітників заводу за місяцями року (чол.):

Таблиця 10.29

| Місяці | Чисельність робітників | Місяці | Чисельність робітників |
|---|---|---|---|
| Січень | 625 | Липень | 992 |
| Лютий | 647 | Серпень | 987 |
| Березень | 712 | Вересень | 975 |
| Квітень | 736 | Жовтень | 878 |
| Травень | 883 | Листопад | 740 |
| Червень | 921 | Грудень | 631 |

Використовуючи формулу середньоарифметичної, розрахуємо середню з фактичних рівнів за однойменними місяцями:

$$\bar{y}_0 = \frac{\sum y}{n} = \frac{625+647+712+736+883+921+992+987+}{12}$$

$$\frac{+975+878+740+631}{12} = \frac{9727}{12} = 810{,}6 \text{ осіб.} \approx 811 \text{ осіб.}$$

Сезонна хвиля (індекс сезонності) визначається відсотковим ставленням місячного рівня ряду до середньої протягом року. Індекс сезонності для січня становить $\left(\frac{625}{811}\times 100\right)$=77,1%; для лютого — $\left(\frac{647}{811}\times 100\right)$=79,8 % і т. п. Відношення середніх фактичних до середніх із вирівняних даних (індекси сезонності) наведено у табл. 10.30.

Таблиця 10.30

| Місяць | Індекс | Місяць | Індекс | Місяць | Індекс |
|---|---|---|---|---|---|
| Січень | 77,1 | Травень | 108,9 | Вересень | 120,2 |
| Лютий | 79,8 | Червень | 113,6 | Жовтень | 108,3 |
| Березень | 87,8 | Липень | 122,3 | Листопад | 91,2 |
| Квітень | 90,8 | Серпень | 121,7 | Грудень | 77,8 |

Як видно за даними цієї таблиці, мінімум чисельності працівником має місце в січні, максимум — у липні.

Середнє відсоткове відношення передбачає, що ізольовані випадкові фактори, які впливають зміну чисельності цього року.

Показником коливання динамічного ряду через сезонний характер виробництва служить середнє квадратичне відхилення індексів сезонності (виражених у відсотках) від 100 %, тобто $\sigma = \sqrt{\frac{\sum (i-100)^2}{12}}$. Розрахунок середнього квадратичного відхилення наведених у табл. 10.30 індексів сезонності подано в табл. 10.31.

Таблиця 10.31

| Місяць | $i_{сез}$ | $i - 100$ | $(i-100)^2$ | Місяць | $i_{сез}$ | $i - 100$ | $(i-100)^2$ |
|---|---|---|---|---|---|---|---|
| I | 77,1 | –22,9 | 524,41 | VII | 122,3 | +22,3 | 497,29 |
| II | 79,8 | –20,2 | 408,04 | VIII | 121,7 | +21,7 | 470,89 |
| III | 87,8 | –12,2 | 148,84 | IX | 120,2 | +20,2 | 408,04 |
| IV | 90,8 | –9,2 | 84,64 | X | 108,3 | +8,3 | 68,89 |
| V | 108,9 | +8,9 | 79,21 | XI | 91,2 | –8,8 | 77,44 |
| VI | 113,6 | +13,6 | 184,96 | XII | 77,8 | –22,2 | 492,84 |

$$\sum (i_{сез} - 100)^2 = 3445{,}49; \quad \sigma = \sqrt{\frac{3445{,}49}{12}} = \sqrt{287{,}1} \approx 17\ \%.$$

Порівняння середніх квадратичних відхилень, обчислених за два періоди, показує зрушення сезонності. Якщо величина $\sigma$ зменшується, то сезонний характер досліджуваного явища спадає.

Однак місячні дані одного року, в силу елемента випадковості, занадто ненадійні для виявлення закономірностей коливань ряду динаміки. Тому на практиці для обчислення закономірностей коливань користуються місячними даними за ряд років (в основному за три роки). У цьому випадку обчислюються середньомісячний рівень для всього ряду і на закінчення визначається відсоткове відношення середніх для кожного місяця до загального середньомісячного рівня ряду:

$$I_s = \frac{\bar{y}_i}{\bar{y}_0} \times 100, \tag{10.54}$$

де $\bar{y}_i$ — середня для кожного місяця за три роки;

$\bar{y}_0$ — загальний середньомісячний рівень за три роки.

Припустимо, маються такі дані про внутрішньорічну закупівлю кондитерських виробів в одному з районів за 2020–2022 рр.:

У табл. 10.32 у перших трьох колонках наведено фактичні дані про закупівлю кондитерських виробів в одному з районів ($y_i$). Ряд не містить яскраво вираженої тенденції у розвитку. Тому індекс

Таблиця 10.32

**Закупівля кондитерських виробів в одному з районів за 2019–2021 рр.**

| Місяць | Закуплено кондитерських виробів, кг | | | | | Місячні дані в процентах до середньорічної ($\bar{y}_i : \bar{y}_0$) |
|---|---|---|---|---|---|---|
| | 2020 р. ($y_i$) | 2021 р. ($y_i$) | 2022 р. ($y_i$) | всього за три роки (гр. 1 + гр. 2 + гр. 3) ($\sum y_i$) | в середньому за 3 роки ($\bar{y}_i$) | |
| А | 1 | 2 | 3 | 4 | 5 | 6 |
| Січень | 1 531 | 1 587 | 1 762 | 4 880 | 1 627 | 62,9 |
| Лютий | 1 924 | 2 446 | 2 564 | 6 934 | 2 311 | 89,4 |
| Березень | 2 764 | 3 382 | 3 231 | 9 377 | 3 126 | 120,9 |
| Квітень | 3 276 | 3 595 | 4 028 | 10 899 | 3 633 | 140,5 |
| Травень | 2 751 | 3 974 | 4 015 | 10 740 | 3 580 | 138,4 |
| Червень | 3 284 | 3 279 | 4 478 | 11 041 | 3 680 | 142,3 |
| Липень | 2 593 | 2 840 | 3 152 | 85 85 | 2 862 | 110,7 |
| Серпень | 2 152 | 2 271 | 2 543 | 69 66 | 2 322 | 89,8 |
| Вересень | 2 257 | 2 534 | 2 658 | 7 449 | 2 483 | 96,0 |
| Жовтень | 1 875 | 2 283 | 2 214 | 6 372 | 2 124 | 82,1 |
| Листопад | 1 492 | 1 942 | 1 645 | 5 079 | 1 693 | 65,5 |
| Грудень | 1 458 | 1 793 | 1 524 | 4 775 | 1 592 | 61,6 |
| Всього за рік | 27 357 | 31 926 | 33 814 | 93 097 | 2 586 | 100,0 |

сезонності обчислюємо безпосередньо за емпіричними даними без їхнього попереднього вирівнювання. Для усунення випадкових сезонних коливань необхідно обчислити середню для кожного місяця. Спочатку підсумуємо протягом трьох років помісячні дані ($\sum y_i$). У графі 4 таблиці надано суми закупівель кондитерських виробів по місяцях за три роки. У цих сумах випадкові особливості кожного окремого року не погашаються. На цій основі у графі 5 обчислено середні за три роки обсяги закупівель кондитерських виробів по місяцях ($\bar{y}_i$).

Для отримання значень $\bar{y}_i$ здійснимо за способом середньої простої (незваженої) усереднення рівнів однойменних періодів:

$$\text{січень } \bar{y}_1 = \frac{y_{\text{січ. 2020}} + y_{\text{січ. 2021}} + y_{\text{січ. 2022}}}{3},$$

$$\text{лютий } \bar{y}_2 = \frac{y_{\text{лют. 2020}} + y_{\text{лют. 2021}} + y_{\text{лют. 2022}}}{3},$$

. . . . . . . . . . . .

$$\text{грудень } \bar{y}_2 = \frac{y_{\text{груд.2020}} + y_{\text{груд.2021}} + y_{\text{груд.2022}}}{3}.$$

Визначимо середні значення рівнів ряду $\bar{y}_i$ для кожного місяця річного циклу:

$$\text{січень } \bar{y}_1 = \frac{1531 + 1587 + 1762}{3} = \frac{4880}{3} = 1627;$$

$$\text{лютий } \bar{y}_2 = \frac{1924 + 2446 + 2564}{3} = \frac{6934}{3} = 2311 \text{ і т. п.}$$

Далі за обчисленими місячними середніми визначаємо загальний середній рівень ($\bar{y}_0$):

$$\bar{y}_0 = \frac{\sum \bar{y}_i}{12} = \frac{31033}{12} = 2586 \text{ кг.}$$

Значення загального середнього рівня можна обчислити також за підсумковими даними за окремі роки:

$$\bar{y} = \frac{\sum (\bar{y}_i)}{m} = \frac{2280 + 2660 + 2818}{3} = \frac{7758}{3} = 2586 \text{ кг.}$$

де $m$ — число років;

$\sum (\bar{y}_i)$ — сума середньорічних рівнів ряду динаміки.

У графі 6 надаються відносні величини (відсотки) помісячних даних до середньорічного обсягу закупівель ($\bar{y}_i : \bar{y}_0$):

$$\text{січень: } \frac{1627}{2586} \cdot 100 = 62{,}9\ \%;$$

$$\text{лютий: } \frac{2311}{2586} \cdot 100 = 89{,}4\ \%;$$

$$\text{березень: } \frac{3126}{2586} \cdot 100 = 120{,}9\ \%.$$

і т. д.

Це і є індекси сезонності, або показник сезонних коливань. Сукупність обчислених індексів сезонності характеризує сезонну хвилю закупівлі кондитерських виробів у внутрішньорічній динаміці. Для наочного уявлення про сезонну хвилю зобразимо отримані дані у вигляді лінійної діаграми (рис. 10.11).

Як видно з графіка, для кондитерських виробів характерна сезонність у їх закупівлі: найбільше зростання закупівель відбувалося


*Рис. 10.11.* Сезонна хвиля закупівель кондитерських виробів.

у березні-квітні, деякий спад у травні, подальший підйом у червні, що досягає максимального рівня, з липня і серпня відбувалося швидке скорочення, а з вересня по грудень — більш повільне зниження.

Застосування цього методу дає можливість нейтралізувати випадкові коливання показників досліджуваного ряду динаміки і визначити сезонні коливання загалом протягом року. Однак, чим триваліший період аналізу, тим суттєвіше в рядах динаміки проявляється тенденція до збільшення або зменшення рівнів ряду. Тому на показники сезонності здійснює суттєвий вплив загальна тенденція розвитку, а не сезонність. У таких випадках виявлення сезонності здійснюють попереднє вирівнювання, і після цього обчислюють сезонну хвилю.

**Розрахунок індексів сезонності у рядах з тенденцією розвитку.** Сутність обчислення сезонності в рядах динаміки з тенденцією розвитку полягає в попередньому визначенні та виключенні загальної тенденції розвитку.

Перш ніж обчислювати сезонність у рядах з тенденцією розвитку, необхідно провести попередній аналіз явища, метою якого є встановлення наявності сезонності, її періодичності та циклічності. Для попереднього встановлення сезонності у рядах динаміки використовуються емпіричні дані ряду динаміки, а також графічний спосіб їхнього зображення.

Визначення загальної тенденції розвитку в рядах динаміки може здійснюватися або методом механічного вирівнювання, або методом аналітичного вирівнювання за рівнянням регресії (лінійної або нелінійної форми), або за способом ковзної (рухливої) середньої. Після визначення загальної тенденції сезонні коливання спостерігаються чіткіше.

При використанні способу аналітичного вирівнювання порядок проведення розрахунків наступний:

1) за відповідним аналітичним рівнянням обчислюють для кожного місяця (кварталу) вирівняні рівні на момент часу $t$;

2) беруть відношення фактичних місячних (квартальних) даних ($\bar{y}_t$) до відповідних вирівняних даних ($\bar{y}_i$) (у процентах): $(y_i/\bar{y}_t)\cdot 100\ \% = U_t$;

3) знаходять середню з цих відношень для однойменних місяців (кварталів) у процентах: $\overline{U}_i = \frac{U_1 + U_2 + U_3 + \ldots + U_n}{U_n}$, де $n$ — число однойменних місяців;

4) з одержаних 12 помісячних відносних величин ($\overline{U}_i$) обчислюють загальний середньомісячний рівень ($\overline{U}_t$);

5) визначають індекси сезонності:

$$I_s = \frac{\overline{U}_i}{\overline{U}_t}\cdot 100, \quad \text{или} \quad I_s = \left[\sum \frac{y_i}{y_t}\cdot 100\right] : n\,. \tag{10.55}$$

Покажемо розрахунок сезонної хвилі на наступному прикладі. У табл. 10.33 у перших трьох колонках наведено фактичні дані про продаж овочів та фруктів за три роки в одному із районів. Можна відзначити, що обсяги продажів мають тенденцію до зростання. Тому спосіб розрахунку сезонної хвилі на основі середньої не придатний. Для виявлення сезонності необхідно спочатку застосувати попереднє вирівнювання фактичних даних. Загальна тенденція визначена способом аналітичного вирівнювання прямої лінії (табл. 10.34).

Здійснюємо розрахунок параметрів прямої:

$$a_0 = \frac{\sum y}{n} = \frac{2\,617}{36} = 72{,}69 \approx 73;$$

$$a_1 = \frac{\sum yt}{\sum t^2} = \frac{7\,778}{3\,885} = 2{,}002 \approx 2{,}0.$$

Таким чином, рівняння прямої має наступний вид:

$$y_t = 73 + 2t.$$

На основі цих параметрів обчислимо вирівняні дані (табл. 10.34). Підставляючи в рівняння прямої $y_t = 73 + 2t$ значення $t$, обчислимо вирівняні (теоретичні) рівні ряду динаміки:

Таблиця 10.33

**Реалізація овочів та фруктів у магазинах району за місяцями та роками (т)**

| Місяці | Фактичні дані | | | Вирівняні дані | | | Фактичні дані в процентах до вирівняних | | | Індекси сезонності |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2019 р. | 2020 р. | 2021 р. | 2019 р. | 2020 р. | 2021 р. | 2019 р. | 2020 р. | 2021 р. | |
| А | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Січень | 38 | 53 | 73 | 38 | 62 | 86 | 100,0 | 85,5 | 84,9 | 90,1 |
| Лютий | 35 | 47 | 60 | 40 | 64 | 88 | 87,5 | 73,4 | 68,2 | 76,4 |
| Березень | 34 | 45 | 55 | 42 | 66 | 90 | 81,0 | 68,2 | 61,1 | 70,1 |
| Квітень | 30 | 41 | 47 | 44 | 68 | 92 | 68,2 | 60,3 | 51,1 | 59,9 |
| Травень | 27 | 43 | 58 | 46 | 70 | 94 | 58,7 | 61,4 | 61,7 | 60,6 |
| Червень | 41 | 51 | 70 | 48 | 72 | 96 | 85,4 | 70,8 | 72,9 | 76,4 |
| Липень | 57 | 75 | 95 | 50 | 74 | 98 | 114,0 | 101,4 | 96,9 | 104,1 |
| Серпень | 90 | 98 | 124 | 52 | 76 | 100 | 173,1 | 128,9 | 124,0 | 142,0 |
| Вересень | 96 | 120 | 149 | 54 | 78 | 102 | 177,8 | 153,8 | 146,1 | 159,2 |
| Жовтень | 85 | 107 | 135 | 56 | 80 | 104 | 151,8 | 133,8 | 129,8 | 138,4 |
| Листопад | 80 | 99 | 126 | 58 | 82 | 106 | 137,9 | 120,7 | 118,9 | 125,8 |
| Грудень | 55 | 80 | 98 | 60 | 84 | 108 | 91,7 | 95,2 | 90,7 | 92,5 |

Таблиця 10.34

**Вирівнювання по прямої реалізації овочів та фруктів**

| $t$ | $y$ | $t^2$ | $yt$ | $y_t$ | $t$ | $y$ | $t^2$ | $yt$ | $y_t$ |
|---|---|---|---|---|---|---|---|---|---|
| –17,5 | 38 | 306,25 | –665 | 38 | 0,5 | 75 | 0,25 | 37,5 | 74 |
| –16,5 | 35 | 272,25 | –577,5 | 40 | 1,5 | 98 | 2,25 | 147 | 76 |
| –15,5 | 34 | 240,25 | –527 | 42 | 2,5 | 120 | 6,25 | 300 | 78 |
| –14,5 | 30 | 210,25 | –435 | 44 | 3,5 | 107 | 12,25 | 374,5 | 80 |
| –13,5 | 27 | 182,25 | –364,5 | 46 | 4,5 | 99 | 20,25 | 445,5 | 82 |
| –12,5 | 41 | 156,25 | –512,5 | 48 | 5,5 | 80 | 30,25 | 440 | 84 |
| –11,5 | 57 | 132,25 | –655,5 | 50 | 6,5 | 73 | 42,25 | 474,5 | 86 |
| –10,5 | 90 | 110,25 | –945 | 52 | 7,5 | 60 | 56,25 | 450 | 88 |
| –9,5 | 96 | 90,25 | –912 | 54 | 8,5 | 55 | 72,25 | 467,5 | 90 |
| –8,5 | 85 | 72,25 | –722,5 | 56 | 9,5 | 47 | 90,25 | 446,5 | 92 |
| –7,5 | 80 | 56,25 | –600 | 58 | 10,5 | 58 | 110,25 | 609 | 94 |
| –6,5 | 55 | 42,25 | –357,5 | 60 | 11,5 | 70 | 132,25 | 805 | 96 |
| –5,5 | 53 | 30,25 | –291,5 | 62 | 12,5 | 95 | 156,25 | 1 187,5 | 98 |
| –4,5 | 47 | 20,25 | –211,5 | 64 | 13,5 | 124 | 182,25 | 1 674 | 100 |
| –3,5 | 45 | 12,25 | –157,5 | 66 | 14,5 | 149 | 210,25 | 2 160,5 | 102 |
| –2,5 | 41 | 6,25 | –102,5 | 68 | 15,5 | 135 | 240,25 | 2 092,5 | 104 |
| –1,5 | 43 | 2,25 | –64,5 | 70 | 16,5 | 126 | 272,25 | 2 079 | 106 |
| –0,5 | 51 | 0,25 | –25,5 | 72 | 17,5 | 98 | 306,25 | 1 715 | 108 |
| | | | | | | 2 617 | 3 885,0 | 7 778,5 | 2 628 |

$y_{-17,5}$=73+2·(–17,5)=38;

$y_{-16,5}$=73+2·(–16,5)=40 . . . і т. п.

Результати розрахунків наведено в графах 4–6. В нашому прикладі підсумки $\sum y_i$ і $\sum y_t$ не співпадають із-за округлень.

Отримані величини теоретичних рівнів ряду $y_t$ нанесемо у вигляді лінії на графік поряд з емпіричними даними (рис. 10.12).

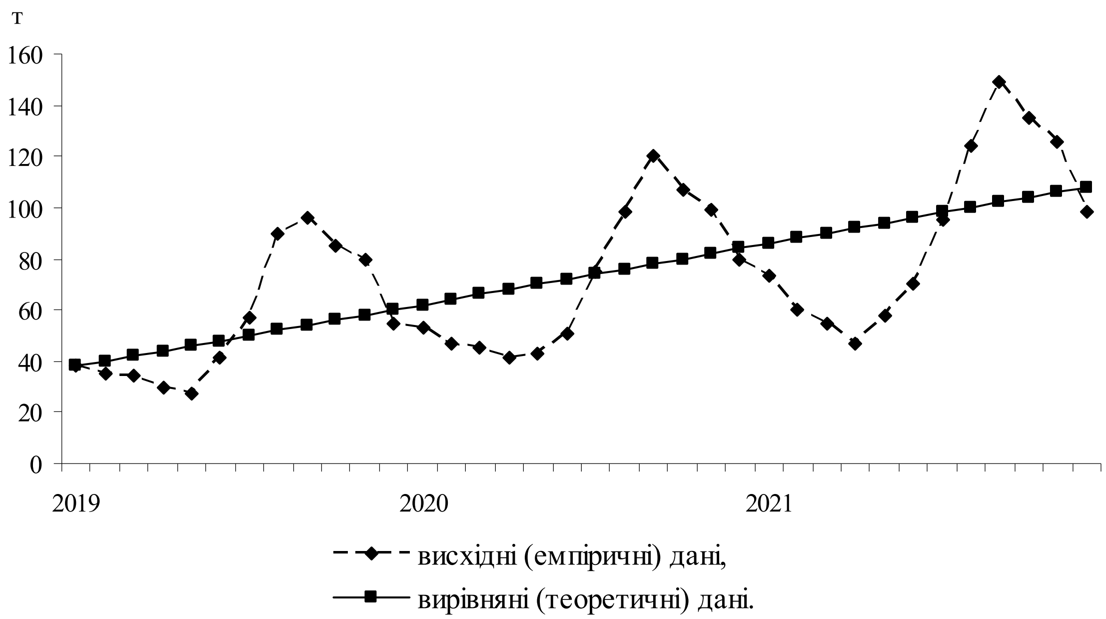



*Рис. 10.12*. Обсяги реалізації овочів та фруктів у магазинах району в 2019–2021 рр.

В наступних трьох графах (7–9) фактичні дані обчислені в процентах до вирівняних. Індекс сезонності для січня 2019 р. становить $\frac{38}{38}\times 100$=100 %; для лютого — $\frac{35}{40}\times 100$=87,5 % і т. п.

В останніх графі обчислені середні за три роки із процентних чисел.

$$I_{січень} = \frac{100{,}0 + 85{,}5 + 84{,}9}{3} = 90{,}1\ \%;$$

$$I_{лютий} = \frac{87{,}5 + 73{,}4 + 68{,}2}{3} = 76{,}4\ \%\ \text{і т. п.}$$

Це і є помісячні індекси сезонності, або показники сезонної хвилі. Таким чином, якщо у першому прикладі ми обчислювали індекси сезонності на основі трьохрічної середньої, то в даному випадку ми обчислюємо до вирівняних даних, а потім із процентних чисел обчислюємо трьохрічні місячні середні.

## 10.9. Інтерполяція та екстраполяція. Статистичні методи прогнозування

**Інтерполяція**. Вирівнюванням рядів динаміки користуються для того, щоб знайти проміжне значення функції в області її визначення. Такий спосіб називається *інтерполяцією* (*інтерполірування*) рядів динаміки. При вивченні часових рядів проводиться інтерполяція проміжних рівнів. Основна ідея інтерполювання полягає в заміні функції $y = f(x)$, для якої відома таблиця значень, інтерполяційним багаточленом, що розглядається як наближений аналітичний вираз для ($x$).

Інтерполяція може здійснюватися різними способами: 1) шляхом використання лише двох рівнів і 2) шляхом використання кількох рівнів. Різні припущення про динаміку розвитку явища призводять, природно, до різних результатів. Тому при виборі того чи іншого способу інтерполяції необхідно визначити тенденцію розвитку ряду динаміки.

Розглянемо застосування інтерполяції на конкретних прикладах.

1. Наприклад, маються такі дані про виробництво цукру в області за 2016–2021 рр.: 2016 р. — 19,8 тис. т, 2017 р. — 22,6, 2018 р. — 25,8, 2019 р. — 28,7, 2020 р. — 31,8, 2021 р. — 35,2 тис. т. Припустимо, що за 2019 р. нам не відомі дані про виробництво цукру. Як дані за цей рік можна отримати приблизно?

Оскільки нам відомі рівні динамічного ряду по обидві сторони невідомого значення, то невідоме значення рівня динаміки можна отримати способом інтерполяції. При інтерполяції (як і екстраполяції) рядів динаміки виходять з припущення, що зміни в межах періоду, що виражають закономірність розвитку, щодо стійкі. Значить, шляхом розрахунку недостатнього рівня можна приблизно встановити характер динаміки, тобто знайти відносно стійкий похідний показник, за яким змінюються рівні ряду (абсолютний приріст, темп зростання та ін.). О б ч и с л и м о похідні показники по ряду виробництва цукру в області (ланцюговим способом) у табл. 10.35.

Таблиця 10.35

| | 2016 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. | 2021 р. |
|---|---|---|---|---|---|---|
| Виробництво цукру, тис. т | 19,8 | 22,6 | 25,8 | 28,7 | 31,8 | 35,2 |
| Абсолютний приріст (швидкість) | … | +2,8 | +3,2 | +2,9 | +3,1 | +3,4 |
| Темп приросту, % | 100,0 | 114,1 | 114,2 | 111,2 | 110,8 | 110,7 |

Щоб визначити, за яким похідним показником динаміки знаходити невідомий рівень ряду, виявимо характер динаміки, тобто знайдемо відносно стійкий з похідних показників даного ряду.

Аналіз приростів за 2016–2021 рр. дозволяє зробити висновок про їхню відносну стійкість. Отже, інтерполювання треба провести за цим похідним показником.

І так, за допомогою інтерполяції знайдемо рівень виробництва цукру за 2019 р. Розрахунок невідомого рівня ряду за 2019 р. зробимо і на основі абсолютних приростів за суміжними рівнями, і на основі середнього абсолютного приросту.

А. Знаходимо абсолютний приріст за 2019 та 2020 р. як різницю рівнів 2020 та 2018 рр.; у прикладі цей рівень становить 6,0 тис. т (31,8 – 25,8). Вважаючи, що зміни приросту аналізованого ряду в цілому були відносно стійкими, то приріст за 2019 р. був таким самим, як і за 2020 р. У результаті отримуємо приріст за 2019 р. у розмірі 3,0 тис. т (6, 0: 2). Тоді рівень 2019 р. виходить як сума рівня 2018 р. та приросту за 2019 р., тобто 25,8+3,0=28,8 тис. т.

Цей же результат можна отримати як середнє арифметичне з прилеглих рівнів 2018 т 2020 рр. $\left(\frac{25{,}8+31{,}8}{2}\right)$. Як видно, отримані розрахунковим шляхом дані за 2019 (28,8) і фактичні дані (28,7) відрізняються один від одного вельми незначно, трохи більше, ніж на 0,3 %.

Б. Обчислимо середній абсолютний приріст за 2016–2021 роки. Його величина складе 3,08 тис. т $\left(\frac{35{,}2-19{,}8}{5}\right)$. Щоб отримати рівень ряду за 2019 р., необхідно до рівня за 2018 р. підсумовувати середній приріст. Він становитиме 29,6 тис. т (25,8+3,08). Розбіжність між цим і фактичним рівнями також є незначною, хоча й трохи вища, ніж у першому випадку. Очевидно, що інтерполяція на основі абсолютних приростів за суміжними рівнями тут є кращою за інтерполяції за середнім абсолютним рівнем.

В. Припустимо, що виробництво цукру за ці роки змінювалося із постійним (середнім) коефіцієнтом зростання. Обчислимо його:

$$\overline{K}_p = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[3-1]{\frac{y_3}{y_1}} = \sqrt{\frac{31{,}8}{25{,}8}} = \sqrt{1{,}2326} \approx 1{,}11.$$

Тоді $y_2 = y_1 \cdot \overline{K}_p = 25{,}8 \cdot 1{,}11 = 28{,}6$.

2. Нехай поряд із двома відомими рівнями випуску продукції за

2018 та 2020 рр. відомий ще один рівень, за 2017 р., рівний 22,6 тис. т. Тоді, припускаючи, що зміна виробництва цукру відбувалася за параболою другого порядку, отримаємо:

Таблиця 10.36

| Рік | $t$ | Виробництво цукру (тис. т) ($y$) | $\overline{y}_t = a_0 + a_1 t + a_2 t^2$ |
|---|---|---|---|
| 2017 | 1 | 22,6 | $\overline{y}_t = a_0 + a_1 \cdot 1 + a_2 \cdot 1$ |
| 2018 | 2 | 25,8 | $\overline{y}_t = a_0 + a_1 \cdot 2 + a_2 \cdot 4$ |
| 2019 | 3 | ? | $\overline{y}_t = a_0 + a_1 \cdot 3 + a_2 \cdot 9$ |
| 2020 | 4 | 31,8 | $\overline{y}_t = a_0 + a_1 \cdot 4 + a_2 \cdot 16$ |

Складаємо систему трьох рівнянь із трьома невідомими:

$$\begin{array}{ll} \text{I} & a_0 + a_1 \cdot 1 + a_2 \cdot 1 = 22{,}6 \\ \text{II} & a_0 + a_1 \cdot 2 + a_2 \cdot 4 = 25{,}8 \\ \text{III} & a_0 + a_1 \cdot 4 + a_2 \cdot 16 = 31{,}8 \end{array}$$

Віднімаючи з другого рівняння перше, отримаємо нове рівняння вже не з трьома, а з двома невідомими:

$$- \begin{array}{lr} \text{I} & a_0 + a_1 + a_2 = 22{,}6 \\ \text{II} & a_0 + 2a_1 + 4a_2 = 25{,}8 \\ \hline & a_1 + 3a_2 = 3{,}2 \end{array}$$

Віднімемо з третього рівняння друге:

$$- \begin{array}{lr} \text{II} & a_0 + 2a_1 + 4a_2 = 25{,}8 \\ \text{III} & a_0 + 4a_1 + 16a_2 = 31{,}8 \\ \hline & 2a_1 + 12a_2 = 6 \end{array}$$

В результаті перетворень маємо систему двох рівнянь їх двома невідомими:

$$\begin{cases} a_1 + 3a_2 = 3{,}2 \\ 2a_1 + 12a_2 = 6 \end{cases} \left| \begin{array}{c} -2 \\ +1 \end{array} \right|$$

Розв'яжемо цю систему рівнянь:

$$+ \begin{cases} -2a_1 - 6a_2 = -6{,}4 \\ 2a_1 + 12a_2 = 6 \end{cases}$$
$$\overline{\qquad 6a_2 = -0{,}4 \qquad}$$
$$a_2 \approx -0{,}067.$$

Тоді

$$a_1 + 3a_2 = 3{,}2;$$
$$a_1 = 3{,}2 - 3 \cdot (-0{,}067) \approx 3{,}4;$$

Підставляючи значення $a_1$ і $a_2$ перше рівняння, отримуємо:

$$a_0 + 3{,}4 - 0{,}067 = 22{,}6;$$
$$a_0 = 19{,}267.$$

Отже, $\bar{y}_{t=3}$ = 19,267+3,4·3 – 0,067·9≈28,9 тис. грн.

Інтерполяція шляхом використання декількох рівнів, що передують невідомим і наступних за ними, може бути зроблена за допомогою вирівнювання способом найменших квадратів.

Наприклад, маються рівні за 2015, 2016, 2017, 2020, 2021 та 2022 рр. (табл. 10.37).

Таблиця 10.37

| Роки | Рівні | Роки | Рівні |
|---|---|---|---|
| 2015 | 4 | 2019 | ? |
| 2016 | 6 | 2020 | 35 |
| 2017 | 14 | 2021 | 54 |
| 2018 | ? | 2022 | 66 |

Р о з р а х у є м о методом інтерполяції рівні за 2018 та 2019 рр. Передбачаючи, що рівні 2018 та 2019 рр. укладаються у загальну тенденцію зміни рівнів, зробимо вирівнювання за прямим способом найменших квадратів. Складемо розрахункову таблицю:

Таблиця 10.38

| Роки | Рівні $y$ | $t$ | $t^2$ | $y \cdot t$ |
|---|---|---|---|---|
| 2015 | 4 | –7 | 49 | –28 |
| 2016 | 6 | –5 | 25 | –30 |
| 2017 | 14 | –3 | 9 | –42 |
| 2018 | | –1 | 1 | |
| 2019 | | +1 | 1 | |
| 2020 | 35 | +3 | 9 | 105 |
| 2021 | 54 | +5 | 25 | 270 |
| 2022 | 67 | +7 | 49 | 469 |
| Всього | 180 | 0 | 168 | –100<br>+844<br>+744 |

$$a_0 = \frac{\sum y}{n} = \frac{180}{6} = 30;$$

$$a_0 = \frac{\sum yt}{\sum t^2} = \frac{744}{168} \approx 4{,}43.$$

Таким чином, рівняння прямої має вигляд:

$$\overline{y}_t = 30 + 4{,}43 \cdot t.$$

Отже, для 2018 р. при ($t$=–1)

$$\overline{y}_{t=-1} = 30 + 4{,}43 \cdot (-1) \approx 25{,}6;$$

а для 2019 р. при ($t$=+1)

$$\overline{y}_{t=-1} = 30 + 4{,}43 \cdot (+1) \approx 34{,}4.$$

Даним методом користуються в тих випадках, коли рівні певних періодів відомі, але спотворені дією будь-яких факторів, вплив яких необхідно оцінити.

Для інтерполяції використовують спеціальні формули: Ньютона, Лагранжа. Наведемо інтерполяційну формулу Ньютона:

$$y = y_0 + n\Delta y_0 + \frac{n(n-1)}{2!}\Delta^2 y_0 + \frac{n(n-1)(n-2)}{3!}\Delta^3 y_0 + \ldots, \qquad (10.56)$$

де $y_0$ — початковий рівень;

$n = \frac{x - x_0}{h}$

$x$ — аргумент; $x_0$ — початкове значення аргументу;

$h$ — інтервал;

$\Delta y_0, \Delta^2 y_0, \Delta^3 y_0$ і т. п. — перша, друга, третя і т. п. різниці початкового рівня.

Так, потрібно інтерполюванням визначити чисельність наявного населення України на 2005 р. Для цього використовуємо наявні дані про чисельність наявного населення України на початок 1998, 2003, 2008, 2013 та 2018 рр., приймаємо їх за умовне позначення часу величини $x$, 1, 2, 3, 4, 5 та знаходимо відповідні різниці (див. табл. 10.39).

$$x_0 = 2; \quad x{=}2{,}4; \quad y_0 = 48{,}0;$$

$$n = \frac{2{,}7 - 2}{1} = 0{,}4;$$

$$\Delta y_0 = -1{,}6; \quad \Delta^2 y_0 = 0{,}8; \quad \Delta^2 y_0 = -3{,}0.$$

Тоді

Таблиця 10.39

**Динаміка чисельності населення та різниці перших чотирьох порядків**

| Рік | | Чисельність населення України на початок року, млн осіб $y$ | Різниця | | | |
|---|---|---|---|---|---|---|
| фактично | умовно $x$ | | перша $\Delta y$ | друга $\Delta^2 y$ | третя $\Delta^3 y$ | четверта $\Delta^4 y$ |
| 1998 | 1 | 50,4 | –2,4 | 0,8 | 0 | –3,0 |
| 2003 | $\boxed{2}$ | $\boxed{48,0}$ | $\boxed{-1,6}$ | $\boxed{0,8}$ | $\boxed{-3,0}$ | |
| 2008 | 3 | 46,4 | –1,8 | –2,2 | | |
| 2013 | 4 | 45,6 | –3,0 | | | |
| 2018 | 5 | 42,6 | | | | |

$$y = 48{,}0 + 0{,}4 \cdot (-1{,}6) + \frac{0{,}4(-0{,}6)}{2} \cdot (0{,}8) +$$

$$+ \frac{0{,}4(-0{,}6)(-1{,}6)}{6} \cdot (-3{,}0) \approx 47{,}1 \text{ млн осіб.}$$

Значить, за інтерполяційною формулою Ньютона чисельність населення на початок 2005 р. дорівнює 47,1 млн осіб, що майже збігається з фактичною чисельністю населення, що дорівнює 47,3 млн осіб.

**Екстраполяція**. Другий прийом, заснований на вирівнюванні рядів динаміки, отримав назву екстраполяція рядів динаміки. *Екстраполяцією* називається знаходження наступних рівнів динамічного ряду, коли попередні рівні відомі. Сутність цього прийому полягає в знаходженні значень функції за межами її області визначення з використанням інформації про поведінку цієї функції в деяких точках, що належать області її визначення. На основі виявленої закономірності зміни рівнів у досліджуваному відрізку часу ми продовжуємо рівень динаміки ряду на майбутнє. Таким чином, ми як би прогнозуємо подальший розвиток явищ.

Для встановлення наступних рівнів ряду динаміки може бути використане вирівнювання рівнів динамічного ряду за способом найменших квадратів і підстановка в отримане аналітичне рівняння відповідних значень $t$.

Розглянемо приклад екстраполяції по рядах динаміки. Маються дані про виробництво тканини за 2016–2021 рр. (табл. 10.40). Припустимо, що нам невідомо виробництво тканини за останній рік цього

Таблиця 10.40

| | 2016 р. | 2017 р. | 2018 р. | 2019 р. | 2020 р. | 2021 р. |
|---|---|---|---|---|---|---|
| Виробництво синтетичних тканин, тис. $м^2$ | 138,0 | 147,8 | 158,4 | 171,4 | 184,3 | 198,7 |
| Абсолютний приріст (швидкість) | … | 9,8 | 10,6 | 13,0 | 12,9 | 14,4 |
| Темп приросту, % | 100,0 | 107,1 | 107,2 | 108,2 | 107,5 | 107,8 |

періоду, тобто за 2021 р. На відміну від попереднього прикладу, всі відомі рівня ряду динаміки лежать тільки по одну сторону невідомого рівня ряду (ліворуч). Для відшукання невідомого рівня ряду використовуємо прийом екстраполяції. Щоб вирішити це завдання, слід визначити, за яким показником буде виробляти екстраполяцію, розрахуємо показники динаміки.

Як видно з наведених даних, відносно стійким показником в аналізованому ряду динаміки є темп приросту. По цьому показнику ми зробимо екстраполяцію (як і за суміжним конкретним темпом зростання, так і за середнім темпом за період).

1. Темп зростання за суміжним 2020 становив 107,8 %. Обчислений по ньому рівень виробництва тканини за 2021 р. дорівнюватиме 198,7 тис. $м^2$ $\left(\frac{184{,}3 \cdot 107{,}8}{100}\right)$. У даному випадку отримане шляхом екстраполяції теоретичне значення за 2021 р. збіглося з фактичним рівнем. Це сталося внаслідок того, що фактичний темп 2021 р. був дуже близьким до темпу зростання за суміжний (2020 р.) період.

2. Середній темп зростання за 2016–2021 рр. складає 107,6 %. Виходячи з цього рівень виробництва тканини за 2021 р. дорівнюватиме 198,3 тис. $м^2$ $(184{,}3 \cdot 1{,}076)$. Розбіжність між отриманим теоретичним значенням рівня ряду за 2019 р. і фактичним значенням незначне і становить 0,2 %. Як видно, екстраполяція за суміжним рядом дозволяє отримати більш точні результати, ніж за середнім темпом зростання.

В якості інструмента для екстраполяційних розрахунків використовують виробничу функцію, яка в найбільш загальній формі може бути визначена як функція залежності результативної ознаки від ознак-факторів. Для знаходження невідомого рівня ряду динаміки може бути використане вирівнювання рівнів динамічного ряду за способом найменших квадратів і підстановка в отримане аналітичне рівняння відповідних значень *t*.

Так, наприклад, якщо ми, вирівнюючи рівні якого-небудь ряду по параболі другого порядку, з 2016 по 2021 р., отримали аналітичне рівняння $\overline{y}_t = 25{,}4 + 1{,}8 \cdot t + 0{,}4 \cdot t^2$ і якщо у нас є підстави припускати, що в 2022 і 2023 рр. дана тенденція не зміниться, то, підставивши в аналітичне рівняння значення *t*, рівні 7 і 8, знайдемо передбачувані рівні:

2022 г. — $\overline{y}_t$=25,4+1,8·7+0,4·7$^2$=57,6;

2023 г. — $\overline{y}_t$=25,4+1,8·8+0,4·8$^2$=65,4.

Таким чином, за допомогою екстраполяції нами були отримані дані ряду динаміки на майбутнє за 2022 та 2023 рр. Їх величина склала відповідно 57,6 і 65,4.

Пошук невідомого рівня ряду динаміки на основі виявленої закономірності вимірювання рівнів у досліджуваному відрізку часу шляхом екстраполяції можна здійснювати як у бік майбутнього (перспективна екстраполяція), так і в бік минулого (ретроспективна екстраполяція). Однак принциповим умовою застосування інтерполяції та екстраполяції є те, що вони повинні проводитися лише в межах однорідних періодів, тобто в межах періодів, в яких діє одна закономірність розвитку. Якщо вихідні рівні динамічного ряду належать до різнорідних періодів, то застосовувати екстраполяцію та інтерполяцію потрібно дуже обережно, враховуючи всі зміни в межах різнорідного періоду часу.

В сучасних умовах екстраполяція на майбутнє використовується в прогнозуванні багатьох соціально-економічних явищ і процесів, регресійному аналізі. Метод екстраполяції є одними із найбільш поширеними і найбільш розробленими методами серед всієї сукупності методів прогнозування. Застосування екстраполяції для прогнозування ґрунтується на інерційності соціально-економічних процесів, яка обумовлена стійкою дією основних факторів, що визначають закономірності розвитку: обсяги капітальних вкладень за минулі роки, технологія і структура виробництва, що склалася, наявні виробничі потужності і т. п. Метод екстраполяції часто використовують для прогнозування зростання населення, попиту населення та інших суспільних явищ. Головна умова прогнозу полягає у правильному об’єктивному формулюванні умов, що діяли в минулому. Разом з тим необхідно відзначити ту обставину, що оскільки рівняння підбираються не на основі якісного аналізу, тобто механічно, то при екстраполяції (виході за межі крайніх значень експериментальних даних) потрібна обережність.

## ЛІТЕРАТУРА


1. *Горкавий В. К.* Статистика: Підручник. Третє вид., перероблений і доповн. Київ: Алерта, 2019. С. 288–320.
2. Статистика: підручник / *С. І. Пирожков*, *В. В. Рязанцева*, *Р. М. Моторин* та ін. Київ: Київ. нац. торг.-екон. ун-т, 2020. С. 235–256.
3. Статистика: Підручник / *С. С. Герасименко*, *А. В. Головач*, *А.М. Єріна* та ін.; За наук. ред. д-ра екон. наук С. С. Герасименка. 2-ге вид., перероб. і доп. Київ: КНЕУ, 2000. С. 121–138.


## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники, словники*


4. *Козирєва О. В.*, *Федорова В. О*. Статистика: навчальний посібник. Харків: Видавництво Іванченка І.С., 2021. С. 83–111.
5. *Мармоза А. Т.* Теорія статистики: підручник. – 2-е вид. перероб. та доп. Київ: Центр учбової літератури, 2013. С. 409–479.
6. *Педченко Г. П.* Статистика: Навчальний посібник. Мелітополь: Колор Принт, 2018. С. 92–99.
7. Статистичний словник / [*О. Г. Осауленко*, *О. О. Васєчко*, *М. В. Пугачова* та ін.]; за ред. д-ра держ. упр., проф., член-кор. НАН України О. Г. Осауленка; НТК статистичних досліджень. Київ: ДП «Інформ.-аналіт. агентство», 2012. 498 с.
8. *Шапочка М. К.*, *Маценко О. М*. Теорія статистики: навч. посібник. Суми: Університетська книга, 2014. С. 221–238.

Г Л А В А 11

# ІНДЕКСИ

## 11.1. Поняття і значення індексів

**Визначення індексу та сфера його застосування**. Для характеристики явищ і процесів суспільного життя статистика застосовує узагальнюючі показники у вигляді середніх, відносних величин і всякого роду коефіцієнтів. До таких узагальнюючих показників належить і індекси. Їх широко використовують для характеристики динаміки, оцінки ступеня виконання планового завдання та порівняльної характеристики територіальних відмінностей та змін.

У широкому розумінні слово індекс «Index» — латинське, що в перекладі означає покажчик, список; часто це слово тлумачать і як «показник». Індекси, що застосовуються в статистиці, є різновидом відносних величин. Вони, як і звичайні відносні показники динаміки, просторового порівняння та виконання плану, дають кількісну характеристику зміни соціально-економічного явища у просторі або ступінь відхилення показника від певного стандарту (нормативу, середньої).

В індексі, як і будь-якій іншій відносній величині, ми абстрагуємося від абсолютного розміру того чи іншого явища як такого. Однак у статистиці це слово набуває специфічного значення.

*Індекс* у статистиці — це узагальнюючий показник порівняння двох сукупностей, які безпосередньо не піддаються підсумовуванню. Більшість статистичних сукупностей, з якими має справу соціально-економічна статистика, складаються з окремих одиниць, що володіють загальними ознаками. В результаті окремі первинні неподільні елементи, або індивідуальні об’єкти статистичної сукупності, можна підсумовувати. У цьому випадку порівняння статистичних сукупностей досягається порівнянням їх обсягів (сум). Так, наприклад, динаміку чисельності населення країни ми отримуємо, порівнявши загальну чисельність населення країни в поточному році з базисним. Динаміку собівартості продукції ми отримуємо, порівнюючи суму витрат за виробництво продукції цього року проти базисним. В основі таких порівнянь лежить сумарність елементів сукупностей, які порівнюються між собою. Однак на практиці при вивченні складних суспільних явищ у часі і просторі виникає необхідність охарактеризувати складне суспільне явище, одиниці якого безпосередньо непорівнянні за своїми споживчими властивостями і якісними характеристиками.

З такого роду сукупностями ми стикаємося, коли ставиться завдання охарактеризувати динаміку зміни обсягу виробленої, проданої або спожитої продукції в її натурально-речовій формі як маси матеріальних благ. Одиниці такої сукупності представлені різною натуральною, споживчою вартістю (тканина — у метрах, сталь — у тоннах, машини — у штуках, у потужностях і т. п.). Природно, неможливо підсумувати тонни металу, скажімо, з метрами тканини, нафти або вугілля і зіставити їх у просторі або часі. А тим часом виникає необхідність дати узагальнюючу характеристику динаміки обсягу виробництва в натурально-речовій формі.

З такими сукупностями ми стикаємося і тоді, коли виникає необхідність охарактеризувати зміну загального рівня цін. Як у попередньому випадку, ціни на окремі товари складати не можна. Не допускають узагальнення (підсумовування) урожайність окремих сільськогосподарських культур, тарифний розряд, норми витрати матеріалів і т. п. У всіх різноманітних прикладах основою несумісності елементів сукупності виступає натурально-речова форма. Як же в таких випадках дати узагальнюючу характеристику розвитку того чи іншого явища?

Проте, ця важка, здавалося б не виконувана задача, може бути вирішена за допомогою індексів. В даному випадку ця характеристика може бути дана за допомогою спеціально побудованих показників — індексів фізичного обсягу продукції, індексів цін і т. п. Для подолання цих труднощів в індекси вводяться *співвимірники* — ціни, фізичний обсяг, витрати праці і т.п. п. Такі співвимірники називаються *вагами*. Наприклад, щодо індексу заробітної плати вагами служить чисельність працівників за окремими категоріями.

Величина, зміна якої визначається індексах, називаються *індексованою величиною* (індексованою ознакою). Ці ознаки можуть бути виражені об’ємною, середньою, відносною величиною і т. п. Конкретні індекси отримують найменування виходячи з назви індексованої величини. Наприклад, індекс фізичного обсягу виробництва (індексована величина — обсяг продукції в натуральному вираженні), індекс цін (індексована величина — ціни реалізації) і т. п.

Індексовані ознаки і ваги не є постійними величинами. Вони можуть змінюватися в залежності від цілей дослідження. Наприклад, щодо динаміки фізичного обсягу продукції індексованої величиною є кількість проданих товарів, а вагами — ціна товару. Якщо ставиться завдання досліджувати динаміку цін, то при такому ж індексному наборі ціна вже виступає індексованою величиною, а кількість проданих товарів — вагами.

Таким чином, індекси є особливими відносними показниками зміни явища у часі чи просторі. Основною відмінністю індексів від відносних і середніх величин полягає в тому, що вони характеризують співвідношення абсолютних рівнів складного явища, окремі елементи якого не сумірні, тобто їх не можна безпосередньо підсумовувати. Це не можна робити при обчисленні середніх та відносних величин. З допомогою індексів можна виміряти незрівнянні складні явища, і навіть кількісно оцінити роль окремих чинників, які формують складне соціально-економічне явище.

Специфіка індексного методу полягає в тому, що в індексі кількісно незрівнянні величини приводяться до деякої спільної єдності, що робить їх порівнянними, порівнянними. К. Маркс писав: «...різні речі стають кількісно порівнянними лише після того, як вони зведені до одного і того ж єдності. Тільки як вираз однієї й тієї ж єдності вони є сумірними, а отже, порівнянними величинами»[1]. Такою єдністю може бути, наприклад, грошова оцінка (вартість) несумірних елементів явища. Таким способом зазвичай приводять до єдності різні за споживчими властивостями види продукції — шляхом множення кількості одиниць продукції на ціну одиниці її.

Отже, під *індексами в статистиці розуміють відносні числа, що характеризують середню зміну в часі або просторі окремого безпосередньо несумірного складного явища або процесу*.

У статистиці індексний метод застосовується при вирішенні різноманітних економіко-статистичних задач. Головними з них є: вивчення динаміки явищ, зіставлення в просторі, вимірювання ступеня

---

[1] *Маркс К.*, *Энгельс Ф.* Соч. Т. 23. С. 58–59.

виконання планових завдань. Індексний метод використовується при вимірі структурних змін у складних явищах і при визначенні впливу факторів на зміну досліджуваного явища, а також для характеристики зв’язків між окремими ознаками досліджуваних явищ.

В умовах економіки індекси стали важливим інструментом регулювання. Багато показників виражаються у вигляді індексів. В даний час інструментарій індексного методу складається з багатьох класів індексів, що охоплюють десятки родів, сотні видів і тисячі різновидів індексів, що обчислюються в економічній, статистичній та плановій практиці. За допомогою індексів можна визначити зміну обсягу виробництва продукції, зростання продуктивності праці, зниження собівартості та цін, зростання товарообігу та ін. Аналіз виконання планів також вимагає застосування індексів.

Державною службою статистики України даються найважливіші індекси, що характеризують підсумки розвитку нашої економіки. Серед них: індекси промислової продукції, індекси товарообігу (реалізації) продукції добувної та переробної промисловості, індекси будівельної продукції, індекси споживчих цін, індекси цін виробників промислової продукції, індекси цін у будівництві, індекси фізичного обсягу, середніх цін у зовнішній торгівлі та ін.

**Класифікація індексів**. Індекси класифікують за низкою ознак. Усі індекси, які застосовуються у соціально-економічній статистиці, діляться на два класи: *індивідуальні* (*елементарні*) і *складні* (*зведені*).

Під *індивідуальними індексами* розуміють відносну величину, що характеризує зміну в часі окремих (індивідуальних) елементів складного соціально-економічного явища. Індивідуальний індекс характеризує зміну окремих (індивідуальних) явищ у часі. При розрахунку індивідуального індексу досліджувана ознака береться без урахування зв’язку його з іншими ознаками явищ, що вивчаються. Наприклад, в 2020 р. в Україні було вироблено 70,7 тис. т шоколадних цукерок, 75,0 млн м$^2$ тканини з синтетичних ниток. У 2015 р. виробництво шоколадних цукерок становило 59,5 тис. т, тканини — 45,4 млн м$^2$. Отже, в 2020 р. в порівнянні з 2015 р. виробництво шоколадних цукерок збільшилося в 1,183 рази більше (70,7:59,5), тканини — 1,652 рази більше (75,0:45,4). Показники 1,183 (або 118,3 %) та 1,652 (або 165,2 %) — це індивідуальні індекси обсягу виробництва шоколадних цукерок та тканини за 2015–2020 рр. Якщо вартість товару в магазині роздрібної торгівлі в березні склала 200 грн, а в квітні — 190 грн, то показник 0,95 (190:200) буде індивідуальним індексом зміни ціни цього товару.

Для обчислення індивідуального індексу величину явища у *звітному періоді* ділять на величину їх у базисному періоді. Якщо потрібно визначити індивідуальний індекс обсягу виробництва, собівартості, цін, врожайності, продуктивності праці, то беруть величину значення ознаки поточного періоду і ділять її на величину порівнюваного періоду.

Виразимо елементарні індекси в формі алгебри. Наприклад, індивідуальний індекс фізичного обсягу виробництва ($i$) певного товару в поточному періоді порівняно з базисним може бути записаний так:

$$i = \frac{q_1}{q_0}, \tag{11.1}$$

де $q_1, q_0$ — фізичний обсяг виробленого товару відповідно у поточному та базисному періодах.

Для того, щоб дізнатися, як змінилися ціни проданих товарів, слід також обчислити індивідуальний індекс. Ціну в базовому періоді позначимо через $p_0$, а ціну товару в поточному періоді — через $p_1$. Індивідуальний індекс цін розраховується за такою формулою:

$$i = \frac{p_1}{p_0}. \tag{11.2}$$

Легко помітити, що за своїм змістом та технікою розрахунку наведені індекси нічим не відрізняються від відносних величин. Це означає, що поняття індивідуального індексу і відносної величини по суті однаково.

Обов'язковою умовою для обчислення індивідуального індексу є вимога однорідності одиниць сукупності того об'єкта, для якого він обчислюється. Практично це досягти дуже важко, оскільки навіть такий простий продукт, як хліб буває різних сортів.

Індивідуальний індекс використовується для зміни явища в динаміці, тому він може бути *базовим* і *ланцюговим*. Виражаються індекси у вигляді коефіцієнтів, відсотків і проміле.

Найбільшого поширення у статистиці набули зведені індекси, які є подальшим розвитком методу середніх.

*Зведений індекс* — індекс, розрахований для сукупності явищ. Зведені індекси служать для порівняння безпосередньо непорівнянних, різнорідних явищ. Щоб зробити такі явища порівнянними, необхідно непорівнянні явища (і складові елементи) зробити сумірними, виразити їх загальною мірою: вартістю, трудовими витратами і т. п.

До зведених індексів відносяться індекс споживчих цін, індекс фізичного обсягу виробництва, індекс продуктивності праці, індекс чисельності працівників підприємства, індекс сортності промислової продукції, індекс виконання плану і т. д. Зведені індекси позначаються буквою *I*, їх застосування є подальшим розвитком методу середніх величин

Зведений індекс може бути *індексом груповим* та *індексом загальним*.

*Загальний індекс* — індекс, розрахований для всієї сукупності досліджуваних явищ, що складаються як з однорідних, так і різнорідних, безпосередньо несумірних елементів. Так, відомо, що обсяг виробництва сільськогосподарської продукції у 2020 р. порівняно з 2019 р. склав 89,9 та 90,0 % порівняно з 2018 р. Показники 89,9 (0,899), 90,0 (0,900) та є узагальнюючі індекси продукції сільського господарства Із загальної сукупності явищ можуть бути виділені окремі їх однорідні групи, для яких обчислюють субіндекс, або груповий індекс.

*Груповий індекс* — індекс, що розраховується для окремих більш менш однорідних груп явищ, що становлять частину загальної сукупності досліджуваних явищ. Групові індекси іноді називають *субіндексами*. Так, індекс валової продукції сільського господарства у 2020 р. порівняно з 2019 р. становив 89,9 %, індекс продукції рослинництва — 87,9 %, а продукції тваринництва — 97,5 %. Індекси продукції рослинництва і тваринництва (89,9 і 97,5) виступають груповими індексами, що входять до складу зведеного індексу валової продукції сільського господарства. Наприклад, у загальному обсязі споживчих товарів та послуг, на основі яких розраховується загальний індекс цін, можуть бути виділені окремі групи товарів: продукти харчування, одяг та взуття, зв’язок, охорона здоров’я і т. п. І для цих окремих груп товарів можуть бути обчислені групові індекси. Зокрема, обчислюють індекс цін продуктів харчування, індекс цін на одяг та взуття, індекс цін на послуги зв’язку тощо. Груповими можуть бути індекси окремих галузей промисловості, складових частин валового внутрішнього продукту, доходів населення і т. п. Поділ індексів на загальні та групові не змінює їхньої суті: як і перші, так і другі є власне індексами.

В залежності від об’єктів дослідження індекси поділяються на індекси кількісних (об’ємних) та індекси якісних показників.

До першої групи належать *індекси фізичного обсягу* промислової продукції, індекс сільськогосподарської продукції, індекс фізичного обсягу роздрібного товарообігу, валового внутрішнього продукту, споживання та ін. У всіх цих індексах кількості оцінюються в однакових, незмінних цінах. Окремі індекси обчислюються щомісяця (наприклад, індекс фізичного обсягу промислової продукції), а більшість — як квартальні та річні індекси.

До другої групи *індексів якісних показників* належать індекси цін (роздрібних, оптових, виробників промислової продукції, продукції сільського господарства, тарифів на послуги зв'язку тощо), індекси собівартості, продуктивності праці (у промисловості, сільському господарстві ), індекси урожайності та ін.

Розподіл індексів на кількісні та якісні має велике значення для методології їх розрахунку.

В залежності від методології розрахунку загальні та групові індекси поділяються на агрегатні та середні із індивідуальних індексів.

*Агрегатні індекси* є основною формою зведених індексів. Вони характеризують відносні зміни індексованої величини в поточному періоді. Чисельники і знаменники агрегатних індексів представляють собою суми добутків індексованої величини на її ваги за два порівнювані періоди.

*Середні з індивідуальних індексів* — похідні індекси, які отримуються в результаті перетворення агрегатних індексів. Середній індекс має бути тотожним агрегатному індексу. При обчисленні середніх індексів використовують дві форми середніх: арифметичну і гармонійну.

Якщо ми будемо вимірювати розвиток економічних та соціальних явищ за деякий період часу, то слід розрізняти *базисну* і *ланцюгову системи розрахунку індексів*. Так, в Україні систематично кожен місяць, квартал, рік обчислюються індекси цін, індекси промислової продукції, індекси капітальних інвестицій, індекси будівельної продукції та ін. У цих випадках утворюються ряди індексів за певний період часу. Якщо як база порівняння для всіх індексів приймемо, скажімо, 2018 р., то такий індексний ряд буде базовим. При ланцюговому методі розрахунку база порівняння у кожному індексі постійно змінюється: вона завжди буде періодом, попереднім звітному. Якщо, наприклад, обчислюються індекси обсягу продукції за п'ять років (2018–2022 рр.), то при ланцюговому способі розрахунку індекс 2022 буде обчислений в порівнянні з 2021, індекс 2021 — в порівнянні з 2020 р. і т. д. У практиці застосовуються обидва способи розрахунку в залежності від поставленої мети.

## 11.2. Агрегатний індекс як висхідна форма індексу

Статистична практика виробила два способи побудови загальних індексів. Особливістю першого із них є те, що загальний індекс отримується в результаті співвідношення абсолютних рівнів складного явища, елементи якого безпосередньо непорівнянні. Такі індекси в статистиці називають агрегатними[1].

Інший спосіб побудови полягає у обчисленні загального індексу як середньої величини з індивідуальних індексів. Наприклад, знаючи індивідуальні індекси витрат праці виготовлення одиниці виробленої продукції і витрати на виробництво окремих видів продукції можна обчислити продуктивність праці в цілому. При обчисленні середніх використовуються дві форми середніх: середня арифметична та середня гармонійна. Тому ці загальні індекси незалежно від способу побудови є середніми показниками.

Надалі загальні індекси називатимемо просто «індексами».

Агрегатна форма індексу виступає як вихідна форма будь-якого загального індексу. *Агрегатний індекс* характеризує відносні зміни індексованої величини в поточному періоді в порівнянні з періодом базисним. Чисельник і знаменник агрегатного індексу представлені сумою добутку двох показників або двох сум добутків. При цьому в кожному добутку один із співмножників представляє фактичний рівень того явища, зміна якого виражає агрегатний індекс, тобто виступає в ролі індексованої величини, а другий співмножник залишається незмінним для обох зіставних періодів і виступає в ролі фіксованої величини (співвимірника).

Різниця між чисельником і знаменником агрегатного індексу відображає зміну складного показника в абсолютному вираженні за рахунок індексованої величини.

Незважений агрегатний індекс обчислюється як частка від ділення суми числових значень (агрегат) індексованої величини $p$, тобто $p_1^I + p_1^{II} + p_1^{III} + ... = \sum p_1$, обчисленої для поточного періоду «1», на аналогічну суму (агрегат) $p_0^I + p_0^{II} + p_0^{III} + ... = \sum p_0$, обчислену для базисного періоду «0»:

$$ {}_{ag}I_{1/0}^{p} = \frac{p_1 + p_1^{II} + p_1^{III} + ...}{p_0^I + p_0^{II} + p_0^{III} + ...} = \frac{\sum p_1}{\sum p_0}. \tag{11.3} $$

[1] Від лат. aggregatus — приєднаний.

Цей індекс виражає сукупну (інакше середню) зміну величини $p$ у часі.

На практиці безпосереднє підсумовування $\sum p_1$ і $\sum p_0$ або взагалі неможливо (внаслідок речової відмінності і несумірності елементів агрегату $\sum p$), або невизначено по самій постановці задачі. Тому практично єдино можливий розрахунок агрегатного індексу, де як ваги використовується інша величина $f$. Ця величина витікає із самої задачі агрегатного індексу і економічно відображає «вагу» (ступінь значення) окремих складових у всьому агрегаті.

У загальному вигляді агрегатні індекси обчислюють за такими формулами:

$I_x = \frac{\sum x_1 f_1}{\sum x_0 f_1}$ — агрегатний індекс із вагами поточного періоду; (11.4)

$I_x = \frac{\sum x_1 f_0}{\sum x_0 f_0}$ — агрегатний індекс із вагами базисного періоду; (11.5)

$I_x = \frac{\sum x_1 f_1}{\sum x_0 f_0}$ — агрегатний індекс спільної зміни обох величин (множників). (11.6)

де $x$ — значення індексованої величини;

$f$ — вага індексу;

0 і 1 — позначення відповідно базисного та поточного періодів.

При побудові індексів у такий спосіб представляється можливим отримати їх безпосередньо з абсолютних даних. Тим самим досягається наочність змісту індексу, правильність економічного змісту його. Основна складність при побудові агрегатного індексу полягає у виборі співвимірника (фіксованої величини). Фіксована величина повинна забезпечити теоретичну сутність індексу і можливість на його основі обчислювати абсолютні суми економічного ефекту динаміки.

Агрегатна форма являє собою основну вихідну форму індексів у статистиці, так як в ній прямо і безпосередньо виражається взаємозв'язок явищ. Так, в агрегатному індексі заробітної плати виражається взаємозв'язок рівнів заробітної плати, чисельності працівників та фонду заробітної плати на підприємстві. У чисельнику дробу цього індексу представлений фактичний фонд заробітної плати у звітному періоді, у знаменнику — умовний фонд заробітної плати тієї ж кількості працівників звітного періоду за збереження базисних рівнів заробітної плати. Різниця чисельника та знаменника виражає розмір зміни фонду оплати праці внаслідок зміни заробітної плати окремих працівників.

Розглянемо побудову агрегатного індексу з прикладу цін товарів, наведених у табл. 11.1. Припустимо, що ми маємо в наявності відомості про кількість реалізованих товарів та ціни на ці товари за два суміжні періоди. Позначимо ціну латинської літерою $p$, кількість реалізованих виробів — $q$, а цифрами знизу справа — період, до якого ціна відноситься: 0 — базисний, 1 — звітний.

Таблиця 11.1

| Виріб | Одиниця виміру | Базисний період | | Поточний період | |
|---|---|---|---|---|---|
| | | кількість виробів | ціна за одиницю в грн | кількість виробів | ціна за одиницю в грн |
| *А* | кг | 100 | 50 | 110 | 55 |
| *Б* | шт. | 500 | 30 | 600 | 27 |
| *В* | м | 1 000 | 25 | 900 | 20 |

Задача формулюється наступним чином: визначити зміну цін окремих видів виробів у поточному періоді порівняно з базисним. Лише для простоти обчислення ми взяли сукупність, що складається з трьох товарів. У статистичній практиці йдеться про сукупності, які складаються з десятків, сотень і тисяч товарів. Складність розв’язання цієї задачі полягає в тому, що потрібно знайти середній, узагальнюючий показник для ознак з різною натуральною формою. Неможливо скласти ціни різних продуктів, як це обчислюють щодо середньої ціни бензину, середньої ціни молока т. п.

Таким чином, просте порівняння середніх рівнів явищ, у нашому прикладі цін, неможливе з тієї простої причини, що ми маємо сукупність різнорідних елементів. У той же час існує загальна закономірність у зміні цін. Ціни на одні товари знижуються, на інші — ростуть, на треті — не змінюються. Крім того, інтенсивність зміни цін різних товарів істотно відрізняється. Але як би не змінювалися індивідуальні ціни на окремі товари, існує деяка загальна тенденція, що характеризує спрямованість цього руху.

Розв’язання цієї задачі здійснюється у декілька прийомів. Для початку потрібно визначити, як змінилися ціни на окремі товари в поточному періоді порівняно з базисним. З цією метою необхідно ціну в поточному періоді ($p_1$) розділимо на ціну в базисному періоді ($p_0$). Це співвідношення становитиме індивідуальний індекс: $i_p = \frac{p_1}{p_0}$. Обчислимо за наведеними даними індивідуальні індекси цін на вироби:

Виріб *А* — $i_A = \frac{55}{50} = 1{,}1$.

Виріб *Б* — $i_Б = \frac{27}{30} = 0{,}9$.

Виріб *В* — $i_В = \frac{20}{25} = 0{,}8$.

Отже, по виробу *А* ціни зросли в 1,1 разу, або на 10 %, по виробу *Б* і *В* ціни знизилися відповідно на 10 і 20 %.

Проте недостатньо знати зміну ціни окремі види виробів. Необхідно визначити динаміку цін на різні товари однією відносною величиною[1]. Очевидно, що скласти ціни окремих видів виробів та отриману суму поділити на три не можна, оскільки вироби відрізняються споживчою вартістю і тому непорівнянні. Незважаючи на те, що загальний індекс є завжди середнім з індивідуальних індексів, не можна складати індивідуальні індекси та обчислювати їх середню величину. Для обчислення загальних індексів завжди використовують строго певну систему співвимірників, інакше кажучи ваг, що надаються кожному індивідуальному індексу. Для того, щоб виміряти зміну цін на різні види виробів, що володіють різною споживчою вартістю, необхідно знайти те спільне, що дозволить їх перейти від незрівнянного до порівняного.

Та обставина, що ці три види виробів безпосередньо непорівнянні за своїми споживчими властивостями і якісними характеристиками, ще не означає їх повної несумірності. Якщо знайти яку-небудь загаль-

---

[1] Однією з перших спроб вираження динаміки ціни різні товари однією відносною величиною вважається запропонований 1738 р. Дюто (Франція) показник $\frac{\sum p_1}{\sum p_0}$, тобто відношення суми цін (*p*) різних товарів у звітному періоді до суми цін тих самих товарів у базисному періоді. Лише у другій половині XIX ст. спочатку Лайспейрес, а потім Паше (Німеччина) ввели в практику розрахунків агрегатну форму індексу цін, замінивши просте підсумовування цін різних товарів підрахунком вартості певної маси товарів ($q$). Лайспейрес вперше у 1871 р. запропонував агрегатну форму індексу цін $\left(\frac{\sum p_1 q_0}{\sum p_0 q_0}\right)$. Наведена формула Пааше була запропонована в 1874 р. і прийняла вид: $\left(\frac{\sum p_1 q_1}{\sum p_0 q_1}\right)$ Цим було покладено початок індексним обчисленням.

ний внутрішній зв'язок між явищами, їх можна виміряти. Так як вартості, як вираження людської праці, витраченої на їх виробництво, ці вироби однакові. Отже, зіставляючи сукупну вартість різних товарів у період, можна отримати показник, характеризує зміна різнорідної сукупності загалом.

Сукупну вартість продукції поточного періоду $\sum p_1 q_1$ можна отримати, помноживши ціну за виріб на фактично вироблену кількість продукції. Зіставивши фактичну вартість продукції з вартістю фактичної продукції, але обчисленої за цінами базисного періоду, ми отримуємо показник динаміки. Оскільки визначальним показником служить товарообіг поточного періоду, то і кількість поточного періоду можна взяти як ваги індексу. Оскільки у співвідношенні змінюються тільки ціни, а кількість вироблених виробів залишається незмінною, отриманий показник буде не що інше, як індекс цін. А саме:

$$I_p = \frac{55 \times 110 + 27 \times 600 + 20 \times 900}{50 \times 110 + 30 \times 600 + 25 \times 900} = \frac{40\ 250}{46\ 000} = 0{,}875.$$

Таким чином, ціни в поточному періоді в порівнянні з базовим періодом знизилися на 12,5 % (0,875·100 – 100), хоча по одним товарам вони підвищувалися, а по інших — знижувалися. Зміна цін за окремими товарами або товарними групами може бути різною, індекс же дає узагальнюючу характеристику цієї зміни.

Використовуючи наведені символи, весь розрахунок можна представити в наступній формі:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}. \tag{11.7}$$

Отримана формула називається агрегатною формулою індексу цін. Ця формула по імені її автора називається формулою Пааше. Агрегатним індекс називається тому, що береться сукупна вартість різних товарів, а як ваги тут використовується фізичний обсяг продукції. При розрахунку індексу цін використовуються поточні ваги, оскільки ціни зважувалися за кількістю поточної продукції. Щоб одержати цей індекс, необхідно фактичну вартість продукції ($\sum p_1 q_1$) зіставити з вартістю цієї продукції, але у базисних цінах ($\sum p_0 q_1$). Якщо різні за споживчими властивостями виробу не можна безпосередньо підсумовувати, то грошову масу можна не лише підсумовувати, а й зіставляти за два періоди.

Формула агрегатного індексу складається з двох елементів: індексованої величини і співвимірника або ваг.

*Індексованою величиною* називаються абсолютні рівні явищ, зміна якої характеризує індекс, а *співвимірником*[1] (*вагами*) — абсолютний рівень того постійного явища, за допомогою якого можна виміряти індексовані, безпосередньо не підсумовуючи величини. У індексі цін індексованою величиною буде ціна, а співвимірником (вагою) — вироблена продукція.

Агрегатна форма представляє собою вихідну форму індексів. Так, у наведеній формулі агрегатного індексу цін виражається взаємозв'язок рівень цін, фізичного обсягу продукції та суми товарообігу. У чисельнику дробу цього індексу представлена фактична величина товарообігу у звітному періоді, у знаменнику — умовна величина товарообігу на той же фізичний обсяг звітного періоду за збереження базисних цін. Різниця чисельника та знаменника виражає розмір зміни товарообігу (зниження чи збільшення) внаслідок динаміки рівнів цін на продукцію.

В інших формах індексу (індексі арифметичному та індексі гармонійному) взаємозв'язок явищ виражається не прямо, а побічно — у вагах так званих індивідуальних індексів.

Існує й інший агрегатний індекс — індекс із базисними вагами. На відміну від попереднього індексу, тут як ваги використовується показник базисного періоду.

Зміна суми товарообігу обумовлено також динамікою фізичного обсягу продукції, що виражається агрегатним індексом за наступною формулою (формула Лайспейреса):

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}. \qquad (11.8)$$

---

[1] Декілька слів з приводу других компонентів агрегатного індексу. Термін «ваги» безперечний по відношенню до середніх індексів, при обчисленні яких проводиться дійсно зважування, а при розрахунку індексів динаміки — окремих елементів індексу він не підходить. Термін «співвимірники» також не можна назвати вдалим. Так, при розрахунку індексу структури немає необхідності у співвимірюванні, оскільки ваги індексованої величини тут цілком співставні. Якщо виходити з визначення індексу, наведеного вище, і агрегатної форми індексу як основної форми будь-якого індексу, то більш правильно називати другі компоненти агрегатного індексу *фіксованими величинами*, так як суть їх саме в тому, що, будучи фіксованими (незмінними) величинами, вони дозволяють виміряти явища, які прямо не співвимірювані, тобто вони відносяться до будь-якого періоду і знаходяться на одному рівні в чисельнику і в знаменнику дробу.

У знаменнику дробу цього індексу — фактична сума товарообігу в базисному періоді, у чисельнику — умовна сума товарообігу звітного періоду за збереження рівнів цін базисного періоду. Різниця чисельника та знаменника виражає розмір зміни суми товарообігу внаслідок зміни її динаміки.

У нашому прикладі агрегатний індекс фізичного обсягу становитиме:

$$I_q = \frac{110 \times 50 + 600 \times 30 + 900 \times 25}{100 \times 50 + 500 \times 30 + 1000 \times 25} = \frac{46\ 000}{45\ 000} = 1{,}022, \text{ або } 102{,}2\ \%$$

Отже, фізичний (натуральний) обсяг реалізованої продукції, оцінений у цінах базисного періоду, збільшився з 45 000 грн у базисному періоді, до 46 000 грн у звітному періоді, або зріс у цілому (або в середньому для окремих елементів продукції) на 2,2 %; приріст продукції за цей період (різниця між чисельником та знаменником агрегатного індексу) склала за тими ж цінами 1 000 грн.

Виникає питання, за яким варіантом більш правильно обчислювати агрегатний індекс? Вибір ваг поточного або базисного періоду залежить від завдання, яке ставиться дослідником під час обчислення індексу. Якщо ставиться задача показати зміни цін на продукцію, яка була реалізована в базисному періоді, то як фіксована величина (ваги) необхідно брати кількість реалізованої продукції базисного періоду ($q_0$). Обчислений індекс характеризував б зміну цін на продукцію, реалізованої в базисному періоді, але з урахуванням зміни цін. Насправді керуються наступним положенням. Індекси об'ємних показників розраховуються за вагами (зазвичай цінами) базисного періоду, а індекси якісних показників (цін, собівартості, продуктивності праці) — за вагами (обсяг продукції) звітного періоду. Тому при побудові індексів у виборі співвимірника необхідно виходити з теоретичного дослідження явища, що вивчається.

Агрегатний індекс характеризує відносні зміни індексованої величини в поточному періоді в порівнянні з базисним періодом. В чисельнику і в знаменнику дроби містяться величини, економічний зміст яких виражений безпосередньо. Агрегатний індекс застосовують у тих випадках, коли важливо знати загальну (середню) відносну зміну всього агрегату $\sum p_i q$ або $\sum q_i p$, взятого в цілому. В інших формах індексів — індексі арифметичному і індексі гармонійному — взаємозв'язок явищ виражається не прямо, а побічно — за допомогою індивідуальних індексів.

## 11.3. Середні індекси

Агрегатна форма індексів, будучи вихідною формою будь-якого, побудованого відповідно до вимог економічного аналізу загального індексу, не є єдиною його формою. У практичній роботі не завжди є можливість мати абсолютні дані про елементи складного явища, тобто не завжди можна побудувати індекс за агрегатною формою. Якщо вихідних даних немає, обчислення нерідко можливе при перетворенні агрегатних індексів на середні форми індексу — середнього арифметичного та середнього гармонічного індексів. Так, якщо є дані про величину товарообігу в поточному періоді та відомі індивідуальні індекси цін за кожною продукцією, то загальний індекс цін можна обчислити як середній з індивідуальних.

Якщо агрегатний індекс характеризує зміна у часі та просторі індексованої величини, то середні форми індексу показують, що він є середньою величиною. Саме середні форми індексу, обчислені як тотожні індексу агрегатного шляхом перетворення останнього, характеризують сутність загального індексу як показника середньої зміни значень показника, що розглядається. Так, агрегатна форма цін представляє собою показник зміни рівня цін на суворо обмежовану сукупність товарів (визначений агрегат товарів), то середні форми індексу визначають, що загальний індекс цін є середня із показників зміни цін окремих товарів.

*Середній індекс* представляє собою індекс, обчислений як середня величина з індивідуальних індексів, тобто з індивідуальних відносних змін індексованої величини ($i^p_{1/0} = \frac{p_1}{p_0}$, $i^q_{1/0} = \frac{q_1}{q_0}$). Він представляє собою одну з форм агрегатних (зведених) індексів. Середні індекси отримані шляхом перетворення зведених (агрегатних) індексів.

В залежності від виду середньої величини розрізняють середні арифметичні індекси, геометричні, гармонічні, модальні, медіанні. При застосуванні тієї чи іншої формули необхідно виходити з теоретичного аналізу сутності досліджуваного явища. Найбільшого поширення набули середній індекс арифметичний та середній індекс гармонійний.

**Індекс середній арифметичний**. *Середній арифметичний індекс* — відносний показник зміни складного економічного явища, що обчислюється за формою середньої арифметичної зваженої з відносних показників зміни окремих елементів цього явища (з так званих індивідуальних індексів).

Середній арифметичний індекс є однією з основних форм зведеного індексу. Причому середній арифметичний індекс повинен бути зважений таким чином, щоб середній арифметичний був тотожній агрегатному індексу. Обчислюється середній арифметичний індекс як середня арифметична, зважена з індивідуальних індексів. Алгебраїчний арифметичний середньозважений індекс ($I$) може бути записаний наступним чином:

$$I = \frac{\sum i x_0 f_0}{\sum x_0 f_0}, \quad \text{або} \quad I = \frac{\sum i x_0 f_1}{\sum x_0 f_1}. \tag{11.9}$$

де $x$ — значення індексованої величини;

$f$ — вага індексів;

0 і 1 — позначення відповідно базисного та звітного періодів.

Вагами при обчисленні арифметичного індексу служать величини, сума яких становить знаменник дробу відповідного агрегатного індексу. Ця обставина дозволяє використовувати середній арифметичний індекс для обчислення загального відносного показниками зміни об'ємних величин (обсягу виробленої або реалізованої продукції, обсягу будівельних робіт і т. п.) в тих випадках, коли пряме використання індексованої величини поточного періоду в агрегатному індексі ускладено або недоцільно з будь-яких причин.

У статистиці середній арифметичний індекс обчислюється шляхом перетворення агрегатного індексу. Для цього індексовану величину поточного періоду, що стоїть у чисельнику агрегатного індексу, замінюють добутком значень індексу індивідуального та значень індексованої величини базисного періоду (виключення становить розрахунок продуктивності праці виходячи з трудомісткості виробництва продукції). Ця можливість випливає з індивідуальних індексів: $i = \frac{x_1}{x_0}$.

Так, агрегатний індекс фізичного обсягу можна представити у вигляді загальної формули:

$$I = \frac{\sum q_1 p_0}{\sum q_0 p_0}, \tag{11.10}$$

де $q_1$ і $q_0$ — фізичний обсяг реалізації відповідно у звітному та базисному періодах;

$p_0$ — ціна реалізації у базисному періоді.

Індивідуальний індекс фізичного обсягу реалізації складе $i = \frac{q_1}{q_0}$, звідси $q_1 = iq_0$. Щоб вивести формулу середнього арифметичного зваженого індексу фізичного обсягу реалізації, достатньо зробити заміну $q_1$ на добуток $iq_0$.

Таким чином:

$$I = \frac{\sum iq_0 p_0}{\sum p_0 p_0} = \frac{\sum \frac{q_1}{q_0} q_0 p_0}{\sum q_0 p_0}. \tag{11.11}$$

де $\frac{q_1}{q_0}$ — показники зміни обсягу окремих видів реалізованої продукції (індивідуальні індекси);

$q_0 p_0$ — вартість кожного виду продукції в базисному періоді (ваги індивідуальних індексів).

Отримана форма представляє собою середню з індивідуальних індексів фізичного обсягу реалізації, зважену за вартістю товарообігу базисного періоду.

Наприклад, за такими даними потрібно розрахувати індекс фізичного обсягу реалізації:

Таблиця 11.2

| Товар | Товарообіг, тис. грн | | Індивідуальні індекси | |
|---|---|---|---|---|
| | базисного періоду $q_0 p_0$ | звітного періоду $q_1 p_1$ | фізичного обсягу реалізації $i_q$ | цін $i_p$ |
| *А* | 5 000 | 6 050 | 1,1 | 1,1 |
| *Б* | 15 000 | 16 200 | 1,2 | 0,9 |
| *В* | 25 000 | 18 000 | 0,9 | 0,8 |

Індекс фізичного обсягу реалізації становитиме:

$$I_q = \frac{\sum i_q q_0 p_0}{\sum p_0 p_0} = \frac{5\,000 \times 1{,}1 + 15\,000 \cdot 1{,}2 + 25\,000 \cdot 0{,}9}{5\,000 + 15\,000 + 25\,000} = \frac{46\,000}{45\,000} =$$

$$=1{,}022, \text{ або } 102{,}2\ \%.$$

Фізичний обсяг реалізації по аналізованій групі товарів зріс на 2,2 %.

Арифметичний індекс призводить до того ж результату, що і агрегатний індекс, відрізняючись від нього лише порядком обчислень, що можна встановити наведеної вище формули:

$$\sum \frac{q_1}{q_0} q_0 p_0 = \sum q_1 p_0. \tag{11.12}$$

Це означає, що формула арифметичного індексу відповідає формулі індексу агрегатного. Середні арифметичні індекси частіше всього застосовуються на практиці для розрахунку зведених індексів кількісних показників.

**Індекс середній гармонійний**. *Середній гармонійний індекс* обчислюється у формі середньої гармонійної виваженої з відносних показників зміни окремих елементів цього явища (з так званих індивідуальних індексів). Розраховується у тих випадках, коли відсутні дані для розрахунку індексу в агрегатній формі.

Будь-який агрегатний індекс може бути перетворений на середній гармонійний з індивідуальних індексів. Для отримання середнього гармонійного індексу в знаменнику агрегатного індексу замінюють індексовану величину базисного періоду відношеннями значень індексованої величини поточного періоду до значень індивідуального індексу.

Так, агрегатний індекс ($I$) можна представити у загальному вигляді формулою:

$$I = \frac{\sum x_1 f_1}{\sum x_0 f_1}. \tag{11.13}$$

З виразу елементарного індексу $i = \frac{x_1}{x_0}$ легко визначити $x_0$: $x_0 = \frac{x_1}{i} = \frac{1}{i} x_1$. Підставляючи значення $\frac{1}{i} x_1$ замість $x_0$ формулу агрегатного індексу з поточними вагами, отримуємо середній гармонійний індекс:

$$I = \frac{\sum x_1 f_1}{\sum \frac{1}{i} x_1 f_1}. \tag{11.14}$$

Як показує формула, зважування проводиться добутками відповідних (залежно від конкретного змісту індексу) значень індексованої величини поточного періоду і значень показника, що служить в агрегатному індексі вагами, тобто в даному випадку $x_1 f_1$.

Середньою гармонійною називається зворотна величина середньої арифметичної із зворотних величин. Наприклад, індивідуальний індекс ціни зазвичай позначається як $i_p = \frac{p_1}{p_0}$, а зворотна йому величина як $1 : \frac{p_1}{p_0} = \frac{p_0}{p_1}$, причому $p_0$ і $p_1$ — ціна одиниці продукції в базовому і звітному періодах.

Так, для отримання середнього гармонійного індексу цін у знаменнику агрегатного індексу цін $I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$ індексовану величину базисного періоду ( $p_0$ ) замінюють рівним їй відношенням $\frac{p_1}{i_p} = \frac{1}{i_p} p_1$.

В результаті отримуємо наступну формулу середнього гармонійного індексу цін:

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{1}{i_p} p_1 q_1} \quad \text{або} \quad I_p = \frac{\sum p_1 q_1}{\sum \frac{p_0}{p_1} p_1 q_1}. \qquad (11.15)$$

Це середній індекс, але гармонійний, а не арифметичний. Нагадаємо формулу середньої гармонійної: $\bar{x} = \frac{\sum f}{\sum \frac{1}{x} f}$. Роль $x$ у нашому випадку виконує $i_p$, тобто індивідуальний індекс, а роль $f$ — фактична вартість виробленої або реалізованої у звітному періоді продукції ( $p_1 q_1$ ).

У знаменнику формули (11.15) $\frac{1}{i}$ — це результат змін цін конкретних виробів, тому в цілому знаменник виражає товарообіг звітного періоду в цінах базисного періоду, а чисельник свідчить про загальну вартість товарообігу у звітному періоді за діючими в цей період цінами. Це дозволяє визначати динаміку цін у порівнянних умовах.

У практиці економічних розрахунків середні гармонійні не застосовуються для розрахунку індексів фізичного обсягу і продуктивності праці, оскільки для їх розрахунків необхідно мати відомості про $p$ і $q$ або $t$ і $q$. Якщо маються ці значення, то розрахунок краще вести з використанням агрегатного індексу.

Покажемо розрахунок середнього гармонійного індексу цін за даними прикладу, наведеного вище. Як вихідна інформація для розрахунку необхідно мати індивідуальні індекси цін $\frac{p_1}{p_0}$ і товарообіг звітного періоду $p_1 q_1$ за окремими товарними групами.

За даними нашого прикладу, обчислимо середній гармонійний індекс цін:

$$I_p = \frac{6\,050 + 16\,200 + 18\,000}{\frac{6\,050}{1{,}1} + \frac{16\,200}{0{,}9} + \frac{18\,00}{0{,}8}} = \frac{40\,250}{5\,500 + 18\,000 + 22\,500} =$$

$$= \frac{40\,250}{46\,000} = 0{,}875, \text{ або } 87{,}5\ \%.$$

Отже, ціни групи товарів знизилися на 12,5 %. Економія від зниження цін становила 5 750 грн (46 000 – 40 250).

Гармонічний індекс відрізняється від відповідного агрегатного індексу тільки порядком обчислення, але призводить до того ж результату, що видно з формул:

$$\sum \frac{p_0}{p_1} p_1 q_1 = \sum p_0 q_1, \tag{11.16}$$

то

$$\frac{\sum p_1 q_1}{\sum \frac{p_0}{p_1} p_1 q_1} = \frac{\sum p_1 q_1}{\sum p_0 q_1}. \tag{11.17}$$

Середній гармонійний індекс застосовується в тих випадках, коли пряме використання базисної індексованої величини поточного періоду в агрегатному індексі утруднено у зв'язку з відсутністю висхідних даних або взагалі застосування середнього гармонічного індексу виявляється більш доцільним, дає економію сил і засобів, необхідних для обчислення.

**Індекс середній геометричний**. *Середній геометричний індекс* — середня геометрична з індивідуальних індексів. Обчислюється за формулами:

індекс середній геометричний незважений —

$$_{g}I = \sqrt[n]{\prod i}\ , \tag{11.18}$$

де $i$ — індивідуальний індекс;

$\prod$ — знак добутку;

$n$ — число індивідуальних індексів;

індекс середній геометричний зважений —

$$_{g}I = \sqrt[\sum f]{\prod i^{f}} \tag{11.19}$$

де $f$ — ваги.

Формули середнього геометричного індексу цін залежно від індексованої величини ($p$ або $q$) можуть бути зведені в таблицю (знаки 1/0 при всіх індексах опущені) (табл. 11.3).

Середній геометричний індекс у статистиці застосовується в ок-

Таблиця 11.3

| Індексована величина | Незважені середні індекси | Зважені середні індекси | |
|---|---|---|---|
| $p$ | $_{g}I^{p} = \sqrt[n]{\prod i^{p}} = \sqrt[n]{\prod\left(\frac{p_1}{p_0}\right)}$ | $_{g(w)}I^{p} = \sqrt[pq]{\prod (i^{p})^{pq}} = \sqrt[\sum pq]{\prod\left(\frac{p_1}{p_0}\right)^{pq}}$ | (1) |
| $q$ | $_{g}I^{q} = \sqrt[n]{\prod i^{q}} = \sqrt[n]{\prod\left(\frac{q_1}{q_0}\right)}$ | $_{g(w)}I^{q} = \sqrt[\sum pq]{\prod (i^{q})^{pq}} = \sqrt[\sum pq]{\prod\left(\frac{q_1}{q_0}\right)^{pq}}$ | (2) |

ремих випадках у міжнародних порівняннях.

## 11.4. Індекси з різною базою порівняння та з різними вагами

**Концепція системи індексів**. При побудові індексів є дуже важливим правильний вибір бази порівняння, так як вона значною мірою впливає на значення результату, що визначається при розрахунку індексу. У зв’язку з тим, що база порівняння служить масштабом для оцінки отриманого результату, незалежно від того, чи відбувається порівняння в територіальному чи динамічному розрізі, вона має бути обґрунтована завданням дослідження.

Усі раніше розглянуті індекси вимірювали динаміку соціально-економічних явищ шляхом порівняння даних за два періоди часу. При цьому в чисельнику індексу знаходився показник, що характеризує рівень явищ за поточний, а в знаменнику — за базовий період часу.

Часто індекси використовуються для аналізу розвитку економічних і соціальних явищ у часі, де є дані більше ніж за два періоди (або моменту часу), наприклад, за роки п’ятирічки. У цьому випадку виникає можливість застосування системи індексів, яка дозволяє охарактеризувати зміни, що відбуваються протягом вибраного інтервалу часу. Система індексів включає ряд індексів. Їх число дорівнює числу включених в аналіз періодів часу мінус одиниця.

Далі, якщо ми розглядатимемо не окремий індекс, а ряд індексів, які послідовно змінюються від одного періоду до іншого, то слід розрізняти *ланцюгову* та *базисну системи* розрахунку індексів.

Індекси, як і відносні показники динаміки, можуть бути ланцюговими та базовими.

1. *Базисні індекси* — система (ряд) послідовно обчислених індексів одного і того ж явища з постійною базою порівняння, тобто в знаменниках всіх обчислених індексів береться індексована величина базисного періоду. Базисні індекси можуть бути як індивідуальними, так і зведеними.

Наприклад, якщо ціну одиниці продукції за *n* періодів позначити як $p_0, p_1, p_2, ..., p_{n-1}, p_n$, то система базисних індивідуальних індексів буде мати такий вигляд:

$$i_1 = \frac{p_1}{p_0}; \quad i_2 = \frac{p_2}{p_0}; \quad ..., \quad i_{n-1} = \frac{p_{n-1}}{p_0}; \quad i_n = \frac{p_n}{p_0}. \tag{11.19}$$

Звернемося тепер до індексів зведеним (загальним). Зведені індекси можуть бути з постійними і змінними вагами, що визначається метою дослідження.

*Індекси з постійними вагами* — система індексів одного і того ж явища, обчислених з вагами, що не змінюються від одного індексу до іншого. Постійні ваги дозволяють виключити (елімінувати) вплив зміни структури на зміну величини, що індексується.

*Індекси зі змінними вагами* — система зведених індексів одного і того ж явища, послідовно обчислених з вагами, що змінюються від одного індексу до іншого. Змінні ваги — це ваги звітного періоду.

Вибір ваг індексів визначається теоретичним досліджуваного аналізом явища.

Зведені базисні індекси цін з постійними вагами ($q_n$):

$$I_1 = \frac{\sum p_1 q_n}{\sum p_0 q_n}; \quad I_2 = \frac{\sum p_2 q_n}{\sum p_0 q_n}; \quad I_{n-1} = \frac{\sum p_{n-1} q_n}{\sum p_0 q_n}; \quad I_n = \frac{\sum p_n q_n}{\sum p_0 q_n}; \tag{11.20}$$

із змінними вагами:

$$I_1 = \frac{\sum p_1 q_1}{\sum p_0 q_1}; \quad I_2 = \frac{\sum p_2 q_2}{\sum p_0 q_2}; \quad I_{n-1} = \frac{\sum p_{n-1} q_{n-1}}{\sum p_0 q_{n-1}}; \quad I_n = \frac{\sum p_n q_n}{\sum p_0 q_n}. \tag{11.21}$$

2. *Ланцюгові індекси* — система (ряд) індексів одного і того ж явища, обчислених із змінною від індексу до індексу базисною величиною. Ланцюгові індекси можуть бути як індивідуальними, так і зведеними.

$$i_1 = \frac{p_1}{p_0}; \quad i_2 = \frac{p_2}{p_1}; \quad ..., \quad i_{n-1} = \frac{p_{n-1}}{p_{n-2}}; \quad i_n = \frac{p_n}{p_{n-1}}. \tag{11.22}$$

Зведені індекси також можуть бути обчислені з постійною та змінною базою.

Ланцюгові індекси цін ряду продуктів (зведені) з постійними вагами ($q_n$):

$$I_1 = \frac{\sum p_1 q_n}{\sum p_0 q_n}; \quad I_2 = \frac{\sum p_2 q_n}{\sum p_1 q_n}; \quad I_{n-1} = \frac{\sum p_{n-1} q_n}{\sum p_{n-2} q_n}; \quad I_n = \frac{\sum p_n q_n}{\sum p_{n-1} q_n}; \tag{11.23}$$

із змінними вагами:

$$I_1 = \frac{\sum p_1 q_1}{\sum p_0 q_1}; \quad I_2 = \frac{\sum p_2 q_2}{\sum p_1 q_2}; \quad I_{n-1} = \frac{\sum p_{n-1} q_{n-1}}{\sum p_{n-2} q_{n-1}}; \quad I_n = \frac{\sum p_n q_n}{\sum p_{n-1} q_n}. \tag{11.24}$$

Індекси з постійними вагами мають ту перевагу, що для них діє правило, згідно з яким *добуток цепних індексів дорівнює індексу базисному*. Наприклад, розрахунок індексів цін з постійними вагами за три місяці дозволяє отримати базовий індекс цін:

$$\frac{\sum p_1 q_3}{\sum p_0 q_3} \times \frac{\sum p_2 q_3}{\sum p_1 q_3} \times \frac{\sum p_3 q_3}{\sum p_2 q_3} = \frac{\sum p_3 q_3}{\sum p_0 q_3}. \tag{11.25}$$

Розглянемо приклад розрахунку. Маються дані про обсяг виробленої продукції та ціни за чотири періоди (табл. 11.4).

Таблиця 11.4

| Види продукції | Грудень | | Січень | | Лютий | | Березень | |
|---|---|---|---|---|---|---|---|---|
| | $p_0$ | $q_0$ | $p_1$ | $q_1$ | $p_2$ | $q_2$ | $p_3$ | $q_3$ |
| *А* | 2,5 | 100 | 2,6 | 110 | 2,9 | 140 | 3,3 | 160 |
| *Б* | 5,1 | 200 | 5,4 | 220 | 5,2 | 250 | 5,8 | 280 |
| *В* | 6,3 | 300 | 6,2 | 280 | 6,5 | 314 | 6,8 | 320 |

Попередньо розрахуємо:

$\sum p_0 q_3 = 2{,}5 \cdot 160 + 5{,}1 \cdot 280 + 6{,}3 \cdot 320 = 400 + 1428 + 2016 = 3844;$

$\sum p_1 q_3 = 2{,}6 \cdot 160 + 5{,}4 \cdot 280 + 6{,}2 \cdot 320 = 416 + 1512 + 1984 = 3912;$

$\sum p_2 q_3 = 2{,}9 \cdot 160 + 5{,}2 \cdot 280 + 6{,}5 \cdot 320 = 464 + 1456 + 2080 = 4000;$

$\sum p_3 q_3 = 3{,}3 \cdot 160 + 5{,}8 \cdot 280 + 6{,}8 \cdot 320 = 528 + 1624 + 2176 = 4328.$

В нашому прикладі:

$$\frac{3912}{3844} \cdot \frac{4000}{3912} \cdot \frac{4328}{4000} = \frac{4328}{3844}.$$

Звідки

$$1{,}126=1{,}126.$$

Правило, згідно з яким твір ланцюгових індексів дорівнює індексу базисному, для індексів з постійними вагами підтверджується.

Спробуємо перевірити це правило для індексів зі змінними вагами. Додатково розрахуємо:

$$\sum p_0q_1 = 2{,}5\cdot110+5{,}1\cdot220+6{,}3\cdot280=275+1122+1754=3161;$$

$$\sum p_1q_1 = 2{,}6\cdot110+5{,}4\cdot220+6{,}2\cdot280=286+1188+1736=3210;$$

$$\sum p_1q_2 = 2{,}6\cdot140+5{,}4\cdot250+6{,}2\cdot314=364+1350+1946{,}8=3660{,}8;$$

$$\sum p_2q_2 = 2{,}9\cdot140+5{,}2\cdot250+6{,}5\cdot314=406+1300+2041=3747;$$

$$\sum p_2q_3 = 2{,}9\cdot160+5{,}2\cdot280+6{,}5\cdot320=464+1456+2080=4000;$$

$$\sum p_3q_3 = 3{,}3\cdot160+5{,}8\cdot280+6{,}8\cdot320=528+1654+2176=4328.$$

Перевіримо правило для індексів зі змінною вагою:

$$\frac{3\,210}{3161}\cdot\frac{3\,747}{3\,660{,}8}\cdot\frac{4\,328}{4\,000}=\frac{4\,328}{3\,844}.$$

Звідки

$$1{,}0155\cdot1{,}0235\cdot1{,}082=1{,}126,$$

$$1{,}1246\neq1{,}1259.$$

Для індексів зі змінними вагами таке правило не діє. Отриманий результат містить певну помилку. Незважаючи на те, що добуток ланцюгових індексів не дає їх базисного рівня, з метою отримання наближеного підсумкового індексу ланцюгові індекси цін іноді перемножують, свідомо знаючи, що в такому розрахунку допускається помилка. Проф. Л. С. Казінєц показав, що величина цієї помилки визначається рівністю:

$$\frac{\sum p_1q_1}{\sum p_0q_1}:\frac{\sum p_1q_0}{\sum p_0q_0}=1+r_{i_pi_q}\upsilon_{i_p}\upsilon_{i_q}, \tag{11.26}$$

тобто помилка визначається добутком коефіцієнта кореляції тісноти зв’язку між індивідуальними індексами цін та кількостей товарів на коефіцієнти варіації індивідуальних індексів цін та індивідуальних індексів кількостей. Чим менше буде кожна з цих величин, тим мен-

Так, в Україні на основі звітних даних систематично обчислюються індекси цін, обсягу промислової продукції, заробітної плати, доходів населення і т. п. На багатьох підприємствах обчислюються індекси собівартості, продуктивність праці. У таких випадках утворюються ряди послідовних у часі індексів. Якщо при цьому база порівняння залишається для всіх індексів постійною, скажімо, 2015 р., то такий індексний ряд буде базовим. При ланцюговому методі розрахунку база порівняння у кожному індексі постійно змінюється: вона завжди буде періодом, попереднім звітному. Якщо, наприклад, обчислюються індекси обсягу продукції за п'ять років (2016–2020 рр.), то за ланцюгового способу розрахунку індекс 2020 р. буде обчислений порівняно з 2019 р., 2019 р. — порівняно з 2018 р. і т. д. У практиці застосовуються обидва способи розрахунку в залежності від поставленої задачі.

ше буде їхній добуток, а отже, і помилка від такого перемноження.

Базисний і ланцюговий методи індексування застосовуються до будь-яких формул індексів.

Таблиця 11.5

**Динаміка обсягу перевезених вантажів автомобільним транспортом в Україні за 2015–2020 рр.**

| Роки | Осяг перевезених вантажів, млн т | Базисні індекси (в процентах до 2015 р.) | Ланцюгові індекси (в процентах до попереднього року) |
|---|---|---|---|
| 2015 | 1 020,6 | – | – |
| 2016 | 1 085,7 | 106,4 | 106,4 |
| 2017 | 1 121,7 | 109,9 | 103,3 |
| 2018 | 1 205,5 | 118,1 | 107,5 |
| 2019 | 1 147,0 | 112,4 | 95,1 |
| 2020 | 1 232,4 | 120,8 | 107,4 |

Розрахуємо систему індивідуальних та базисних ланцюгових індексів за даними табл. 11.5.

У соціально-економічних дослідженнях та в статистичній практиці вибір методу розрахунку індексу визначається економічною сутністю досліджуваного явища і тими завданнями, які перед індексом ставлять. Базисні індекси дають більш чітке уявлення про загальну тенденцію досліджуваного явища, тоді як ланцюгові — дають більш детальну картину послідовної зміни рівнів у часі.

**Взаємозв'язок індексів**. **Зв'язок між ланцюговими та базисними індексами**. Між окремими явищами і процесами, як відомо, існує об'єктивний взаємозв'язок і взаємозалежність. Цей взаємозв'язок притаманний і індексам. Як відомо, за допомогою індексного методу можна розкласти складне явище на окремі частини, вимірюючи при цьому роль окремого компонента в загальній зміні складного явища. У даному випадку ми використовуємо індекси не ізольовано, а комплексі, де кожен індекс постає як частина індексної системи.

*Взаємозв'язок індексів* — зв'язок між певними індексами, обумовлений як реальними зв'язками соціально-економічних явищ, що відображаються ними, так і математичними властивостями індексів. Наприклад, індекс виробництва продукції дорівнює добутку індексу чисельності працівників на індекс продуктивності праці; індекс витрат дорівнює добутку індексу фізичного обсягу виробленої продукції індекс собівартості одиниці виробленої продукції.

Зв'язок між індивідуальними та зведеними індексами слід розглядати у трьох аспектах:

1. Взаємозв'язок ланцюгових та базисних індексів.
2. Взаємозв'язок конкретних економічних індексів.
3. Взаємозв'язок індексів змінного, фіксованого складу та структури.

Розглянемо основні взаємозв'язки.

Для індивідуальних індексів між ланцюговими та базисними індексами є наступний взаємозв'язок: добуток членів системи ланцюгових індивідуальних індексів дорівнює останньому базовому індексу. В загальному вигляді:

$$i_{n/0} = \frac{x_1}{x_0} \cdot \frac{x_2}{x_1} \cdot \frac{x_3}{x_2} \cdot \ldots \cdot \frac{x_n}{x_{n-1}} = \frac{x_n}{x_0}. \quad (1.27)$$

Частка від ділення двох базисних індексів (наступного та попереднього) дорівнюють відповідному ланцюговому:

$$\frac{x_2}{x_0} : \frac{x_1}{x_0} = \frac{x_2}{x_1}. \quad (1.28)$$

Ці взаємозв'язки зберігаються і для зведених індексів, розрахованих за співвимірними ознаками. В загальному вигляді:

$$\frac{\sum x_1 f_2}{\sum x_0 f_2} \cdot \frac{\sum x_2 f_2}{\sum x_1 f_2} = \frac{\sum x_2 f_2}{\sum x_0 f_2}; \quad (1.29)$$

$$\frac{\sum x_2 f_2}{\sum x_0 f_2} : \frac{\sum x_1 f_2}{\sum x_0 f_2} = \frac{\sum x_2 f_2}{\sum x_1 f_2}. \quad (1.30)$$

Для ланцюгових зведених індексів зі змінними вагами це правило не поширюється.

Взаємозв'язки індексів конкретних економічних явищ зумовлені взаємозв'язками явищ, що відображаються між ними. Так, між індексом цін (якщо йдеться про роздрібні ціни), індексом фізичного обсягу продукції (якщо йдеться про оптові ціни підприємства) та індексом загальної вартості існує наступне постійне співвідношення:

$$\frac{\sum p_1 q_1}{\sum p_0 q_1} \times \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum p_1 q_1}{\sum p_0 q_0}, \tag{1.31}$$

| | | |
|---|---|---|
| індекс цін | індекс фізичного обсягу | показник динаміки загальної вартості |

або

$$I_{\text{товарообороту}} = I_{\text{цін}} \cdot I_{\text{фізичного обсягу}}. \tag{1.32}$$

Добуток індексу цін на індекс фізичного обсягу товарообігу або продукції у фактичних цінах дає індекс фізичного обсягу товарообігу у фактичних цінах, або індекс вартості продукції. Наприклад, вартість товарної продукції поточних цінах збільшилася на 20 % за зниження цін на 5 %. Тоді індекс фізичного обсягу продукції складе 126,3%, тобто випуск продукції у звітному періоді збільшився на 26,3%.

Звідси індекс обсягу виробленої (реалізованої) продукції може бути отримано як окреме від ділення індексу загальної вартості (товарообігу) на індекс цін, тобто

$$I_{\text{фізичного обсягу}} = \frac{I_{\text{товарообороту}}}{I_{\text{цін}}}. \tag{1.33}$$

Між індексом фізичного обсягу продукції (за трудовими витратами), індексом трудових витрат та індексом продуктивності праці (за трудовими витратами) існує подібне співвідношення:

$$\frac{\sum q_1 t_0}{\sum q_0 t_0} : \frac{\sum t_1 q_1}{\sum t_0 q_0} = \frac{\sum t_0 q_1}{\sum t_1 q_1}, \tag{11.34}$$

| | | |
|---|---|---|
| індекс фізичного обсягу | індекс затрат робочого часу | індекс продуктивності праці |

або

$$I_{\text{фізичного обсягу}} : I_{\text{затрат праці}} = I_{\text{продуктивності праці}}. \tag{11.35}$$

Між індексом витрат виробництва, індексом кількості виробленої продукції та індексом собівартості існує таке співвідношення:

$$\frac{\sum C_1 q_1}{\sum C_0 q_1} \times \frac{\sum q_1 C_0}{\sum q_0 C_0} = \frac{\sum C_1 q_1}{\sum C_0 q_0}. \tag{11.36}$$

| | | |
|---|---|---|
| індекс собівартості | індекс кількості виробленої продукції | показник динаміки виробничих витрат |

Основним показником рівня продуктивності праці є виробіток товарної (валової) продукції на одного середньооблікового працівника промислово-виробничого персоналу за рік, квартал, місяць. У цьому випадку індекс чисельності працівників пов’язаний з індексом продуктивності праці, утворюючи наступну індексну систему:

$$I_{\substack{\text{чисельності} \\ \text{робітників}}} \cdot I_{\substack{\text{продуктивності} \\ \text{праці}}} = I_{\substack{\text{продукції} \\ \text{(фізичного обсягу)}}} . \qquad (11.37)$$

Крім цього, динаміку складного соціально-економічного явища можна вивчити за допомогою індексу змінного складу, постійного складу та індексу структури. У загальному вигляді взаємозв’язок між індексом змінного складу, індексом постійного складу та індексом структури представлений таким чином:

$$I_{\substack{\text{постійного} \\ \text{складу}}} \cdot I_{\text{структури}} = I_{\substack{\text{змінного} \\ \text{складу}}} , \qquad (11.38)$$

або

$$\left(\frac{\sum x_1 f_1}{\sum x_0 f_1}\right) \times \left(\frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}\right) = \left(\frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}\right). \qquad (11.39)$$

За допомогою індексного методу ми можемо аналізувати динаміку складного явища, розклавши його на окремі частини. Так, аналіз, наприклад, індексу собівартості, індексу витрат обігу, індексу витрат праці і т. п. Крім цього, в тих випадках, коли якісні показники є одночасно середніми показниками, за допомогою індексного методу можна ізольовано розглядати, з одного боку, вплив зміни структури складного явища, що має своєю ознакою якісні показники, з іншого боку — зміна самої ознаки, що виступає як фактор динаміки середніх показників.

## 11.5. Індекс структурних зрушень

*Індекс структурних зрушень* (*індекс впливу зміни структури*) — індекс, що характеризує вплив структурних зрушень (зміну структури досліджуваного явища) на динаміку середнього рівня цього явища.

За допомогою індексів можна відокремити вплив структури явищ від зміни індексованої ознаки при аналізі динаміки вторинних ознак (середньої собівартості, середньої матеріаломісткості, середньої продуктивності праці, середньої урожайності і т. п.). Це друге завдання, яке вирішується за допомогою індексів.

На зміну середнього значення ознаки впливають як зміна значень досліджуваної ознаки в одиниць сукупності, і зміна ваг, тобто зміна структури явища. Внаслідок цього може виявитися, що ознака, що вивчається у одиниць сукупності, припустимо, не зміниться або навіть збільшитися, а середня за цією ознакою в цілому за сукупністю зменшиться; досліджуваний ознака зменшиться, а середня за сукупністю збільшиться[1]; можуть траплятися і змішані випадки.

Звернемося до прикладу. Маються такі дані про валові збори, посівні площі та врожайність по сільськогосподарському підприємству за 2021 та 2022 рр. (табл. 11.6).

Таблиця 11.6

| Зернова культура | Валовий збір, ц | | Посівна площа, га | | Урожайність, ц/га | |
|---|---|---|---|---|---|---|
| | 2021 р. | 2022 р. | 2021 р. | 2022 р. | 2021 р. | 2022 р. |
| | $n_0 y_0$ | $n_1 y_1$ | $n_0$ | $n_1$ | $y_0$ | $y_0$ |
| Пшениця | 28 600 | 26 000 | 1 100 | 1 000 | 26 | 26 |
| Овес | 2 080 | 6 160 | 160 | 440 | 13 | 14 |
| Кукурудза | 10 080 | 2 700 | 240 | 60 | 42 | 45 |
| Всього | 40 760 | 34 860 | 1 500 | 1 500 | 27,2 | 23,2 |

За наступними даними слід розрахувати вплив структурних зрушень зміну середньої урожайності зернових культур.

Якщо обмежитися аналізом середньої врожайності окремих зернових культур, можна позитивно оцінити результати її зміни. Можна помітити, що за озимою пшеницею урожайність не змінилася, а за вівсом і кукурудзою урожайність збільшилася. Отже, очікується і збільшення середньої урожайності загалом зернових культур.

Обчислимо середню урожайність за 2021 та 2022 рр.:

середня урожайність за 2021 р.:

$$\bar{y}_0 = \frac{\sum n_0 y_0}{\sum n_0} = \frac{1100 \cdot 26 + 160 \cdot 13 + 240 \cdot 42}{1100 + 160 + 240} = \frac{40\,760}{1500} = 27{,}2 \text{ ц/га};$$

середня урожайність за 2022 р.:

$$\bar{y}_1 = \frac{\sum n_0 y_0}{\sum n_0} = \frac{1\,000 \cdot 26 + 440 \cdot 14 + 60 \cdot 44}{1\,000 + 440 + 60} = \frac{34\,860}{1500} = 23{,}2 \text{ ц/га}.$$

[1] Такі ситуації за їхню суперечливість іноді називають «статистичними парадоксами».

Індекс змінного складу дорівнює:

$$I_{\bar{y}} = \frac{\bar{y}_1}{\bar{y}_0} = \frac{23,2}{27,2} = 0,855, \text{ або } 85,5\ \%.$$

Таким чином, середня урожайність зернових культур у 2022 р. порівняно з 2021 р. знизилася на 4,0 ц або 14,5 %. Це, здавалося б, суперечливий результат («статистичний парадокс») пояснюється тим, у звітному періоді змінилася структура посівних площ: скоротилася питома вага високоврожайної культури (кукурудзи) і одночасно збільшилася питома вага із середньою урожайністю (озимою пшениці) та низьковрожайної культури (вівса).

На динаміку загальної середньої урожайності вплинула зміна двох факторів — власне урожайності окремих зернових культур та структури посівних площ. Для того, щоб усунути вплив змін структури посівних площ, необхідно обчислити показник, що відображає зміну тільки урожайності. Таким показником і виступає індекс, в якому як індексована величина береться урожайність за окремими культурами, а вагами — структура посівних площ. Важливою умовою побудови індексу урожайності є правильний вибір ваг. У цьому індексі первинним фактором є посівні площі, а вторинним — урожайність. З урахуванням вибору періоду ваг у цьому індексі вторинна ознака (урожайність) повинна була взята з вагами звітного періоду.

Для оцінки величини середньої урожайності по зерновим культурам разом без урахування впливу структурних зрушень обчислюється індекс урожайності постійного складу (у структурі звітного періоду):

$$I_y = \frac{\sum n_1 y_1}{\sum n_1 y_0} = \frac{1\,000 \cdot 26 + 440 \cdot 14 + 60 \cdot 45}{1\,000 \cdot 26 + 440 \cdot 13 + 60 \cdot 42} = \frac{34\,860}{34\,240} = 1,018, \text{ або } 101,8\ \%.$$

Таким чином, під впливом зміни урожайності окремих зернових культур середня урожайність 2022 р. проти 2021 р. збільшилася 1,8 %.

Оскільки зміна середньої урожайності відбувається під впливом зміни урожайності та структурних зрушень посівних площ, то можна записати це наступним чином:

$$I_{\bar{y}} = I_y \cdot I_{\text{стр}}. \qquad (11.40)$$

Звідси розрахунок індексу впливу зміни структури на динаміку середньої урожайності проводиться шляхом розподілу індексу змінного складу на індекс постійного складу. У нашому прикладі він дорівнює:

$$I_{\text{стр}} = \frac{I_{\bar{y}}}{I_y} = \frac{0{,}855}{1{,}018} = 0{,}840, \text{ або } 84{,}0\ \%.$$

Отже, зміна структури посівних площ призвела до значного скорочення середньої урожайності зернових культур.

Можливий і інший спосіб розрахунку індекс структурного зрушення. Підставляючи замість $I_{\bar{y}}$ і $I_y$ формули їх обчислення, можна вивести формулу розрахунку індексу структури:

$$I_{\text{стр}} = \frac{I_{\bar{y}}}{I_y} = \left( \frac{\sum n_1 y_1}{\sum n_1} : \frac{\sum n_0 y_0}{\sum n_0} \right) : \frac{\sum n_1 y_1}{\sum n_1 y_0}. \tag{11.41}$$

Після перетворень отримуємо формулу розрахунку індексу структури:

$$I_{\text{стр}} = \frac{\sum n_1 y_0}{\sum n_1} : \frac{\sum n_0 y_0}{\sum n_0}. \tag{11.42}$$

Підставивши значення у виведену формулу, обчислимо індекс структури:

$$I_{\text{стр}} = \frac{\sum n_1 y_0}{\sum n_1} : \frac{\sum n_0 y_0}{\sum n_0} = \frac{34\ 240}{1\ 500} : \frac{40\ 760}{1\ 500} = 22{,}8 : 27{,}2 = 0{,}84.$$

В загальному вигляді *індекс структури* може бути обчислений за такою формулою:

$$I_{\text{стр}} = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}, \tag{11.43}$$

або

$$I_{\text{стр}} = \frac{\sum x_0 f_1}{\sum x_0 f_0} : \frac{\sum f_1}{\sum f_0}, \tag{11.44}$$

де $x$ — це може бути собівартість одиниці продукції, а $f$ — в цьому випадку — кількість продукції, або $x$ — виробіток продукції на одного робітника, тоді $f$ — чисельність робітників; 0 і 1 — позначення відповідного періоду базисного і періоду звітного. Якщо $f$ — відносна величина структури ($f'$), то: $I_{\text{стр}} = \frac{\sum x_0 f_1'}{\sum x_0 f_0'}$.

## 11.6. Вивчення впливу різних факторів за допомогою індексного методу

Агрегатні індекси, як зазначалося вище, застосовуються для дослідження складних суспільних явищ у часі та просторі. Поряд з розкладанням повного індексу на окремі аналітичні індекси як співмножники величезне практичне значення набуло розкладання повного абсолютного приросту складного явища на окремі прирости ($\Delta w = w_1 - w_0$ або $\Delta W = \sum \Delta w$) на абсолютні прирости, отримані за рахунок впливу окремих факторів ($\Delta x, \Delta y, \Delta z...$). При цьому особливо важливим є визначення долей інтенсивного та екстенсивного факторів у цьому прирості.

Позначимо, наприклад, «фактор I» символом «$x$», а «фактор II» — символом «$y$», досліджувану величину — символом «$w$», розмір факторів та показника базисного періоду позначимо відповідно — $x_0, y_0, w_0$, а звітного (аналізованого) періоду — символами $x_1, y_1, w_1$, різниці цих величин (алгебраїчне збільшення за період) $\Delta x = x_1 - x_0$, $\Delta y = y_1 - y_0$, $\Delta w = w_1 - w_0$. Тоді обсяг показника за поточний період можна подати такою формулою:

$$w_1 = x_1 \cdot y_1 = (x_0 + \Delta x)(y_0 + \Delta y) = x_0 y_0 + \Delta x y_0 + \Delta y x_0 + \Delta x \Delta y. \qquad (11.45)$$

Віднімемо з нього обсяг «показника» за базовий період «$x_0 \cdot y_0$», отримаємо приріст ($\Delta w$):

$$\Delta w = \Delta x y_0 + x_0 \Delta y + \Delta x \Delta y. \qquad (11.46)$$

З рівняння видно, що перший доданок пов’язаний зі зміною «фактора I», другий доданок — зі зміною «фактора II», третій доданок — зі зміною «фактора I» і «фактора II». Це доданок ($\Delta x \Delta y$) у літературі отримало назву «залишковий член», інколи ж його називають навіть «нерозкладним залишком».

При трифакторній моделі зв’язку $w = x \cdot y \cdot z$ виходить розкладання:

$$\Delta w = w_1 - w_0 = y_0 z_0 \Delta x + x_0 z_0 \Delta y + x_0 y_0 \Delta z + z_0 \Delta x \Delta y + \\ + y_0 \Delta x \Delta z + x_0 \Delta y \Delta z + \Delta x \Delta y \Delta z. \qquad (11.47)$$

При розкладанні загального приросту складного явища на окремі абсолютні прирости виникає аналітична задача: кількісно оцінити вплив зміни кожного компонента на повну зміну (абсолютне або відносне) всього складного явища. Для вирішення цієї задачі використовується метод елімінування, тобто виключають вплив зміни всіх інших факторів, крім одного досліджуваного. В агрегатних індексах елімінування досягається умовним закріпленням числового значення кожного виключеного фактора на якомусь рівні (базисному або поточному). Це означає, що в індексному елімінуванні відбувається свого роду «статистичне абстрагування» не від самого факту наявності даного фактора або факторів, а лише від змін та його числових значень в досліджуваному складному явищі.

Існує два методи такого елімінування: 1) метод послідовного виключення (або, інакше, метод ланцюгових підстановок) компонентів, без розриву зв'язку чергового компонента, що виключається, з ще не виключеними); 2) метод відокремлення факторів, при якому кожен фактор береться від самого початку повністю ізольованим від усіх інших.

При першому методі розрахунок впливу окремих факторів на зміну складного здійснюється шляхом послідовної заміни окремих факторів. Недоліком цього методу і те, що об'єктивно необхідна послідовність відбору чинників майже ніколи не може бути логічною, ні економічно обґрунтованою. Це веде до перебільшення ролі одних факторів і применшення інших. Так, якщо $x$ — кількісний фактор, а $y$ — якісний фактор, то залежність результативного показника від цих двох факторів може бути представлена в наступному вигляді:

$$\Delta w = y_0 \Delta x + (x_0 + \Delta x) = y_0 \Delta x + x_1 \Delta y, \tag{11.48}$$

або

$$\Delta w = (y_0 + \Delta y)\Delta x + x_0 \Delta y = y_1 \Delta x + x_0 \Delta y. \tag{11.49}$$

У першому випадку ступінь впливу кількісного фактора $x$ на зміну результативного показника $w$ виявиться перебільшеним, а в другому випадку — применшеним. До ще більших спотворень і протиріч отримуваних результатів може призвести метод довільного послідовного виключення компонентів при розкладанні у випадку $\Delta w$ трьох факторів.

Аналогічне перебільшення або применшення впливу факторів характерно для агрегатних індексів. Наприклад, вплив цін можна обчислити двома способами: $I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$ або $I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}$. У першій формулі ціни $p$ зважені за фізичним обсягом продукції $q$ поточного періоду, а в другій — відповідно базисного періоду.

Таким чином, розрахунок впливу окремих факторів на зміну результативного показника значною мірою визначається послідовністю підстановки окремих факторів.

Згідно індексної теорії при визначенні долей факторів потрібно при розрахунку величини приросту інтенсивного фактора в якості ваг брати об’ємний фактор звітного періоду, а при численні приросту екстенсивного (об’ємного) фактора — якісний показник базисного періоду.

Внаслідок багатьох істотних недоліків методу послідовного виключення наукове визнання отримав метод відокремлення факторів, при якому числове значення будь-якого елімінованого фактора встановлюється на рівні базисного періоду ($x_0, y_0, z_0$). При цьому кількісно оцінюють ізольований («чистий») вплив окремих факторів ($x$, $y$, $z$) і вплив спільної зміни різних факторів у їх взаємозв’язку. Тому, наприклад, для $w=pq$ аналітичного індексу розчленовують не на два, а на три індекси:

$$I_w = \left( \frac{\sum p_1 q_0}{\sum p_0 q_0} \right) \cdot \left( \frac{\sum q_1 p_0}{\sum q_0 p_0} \right) \cdot \left( \frac{\sum p_1 q_1 \cdot \sum p_0 q_0}{\sum p_0 q_1 \cdot \sum p_1 q_0} \right). \qquad (11.50)$$

Третій індекс є результатом спільного впливу окремих факторів на результативний показник. Цей індекс отримав назву «індекс взаємозв’язку».

Передумовою для використання агрегатних індексів в економічному аналізі є можливість представити результативний показник як добуток двох або більше визначальних його факторів або як їх сума. Наприклад, обсяг випущеної (реалізованої) продукції можна бути представленим добутком середньооблікової чисельності працівників і виробленої продукції з розрахунку на одного працівника (тобто продуктивності праці). Виходячи з цього, зміна обсягу випущеної продукції залежить від зміни середньооблікової чисельності працівників і продуктивності праці.

З метою виявлення ролі окремих факторів на зміну явища доцільно використовувати розкладання загального індексу на субіндекси, які характеризують роль кожного досліджуваного фактора. Аналіз факторів, які впливають обсяг випуску продукції, показує, на які з них у якому напрямі необхідно впливати, аби досягти бажаного збільшення обсягу випуску продукції.

Для обчислення ролі окремих факторів у зміни явища необхідно побудувати систему взаємозалежних індексів. Завдання полягає в тому, щоб розрахувати зміну складного показника при зміні величини тільки одного показника так, що величина інших факторів зберігається на фіксованому рівні. Вирішити таку задачу індексним методом можна тільки використовуючи загальноприйнятий принцип елімінування впливу всіх інших факторів, крім досліджуваного. Принцип у тому, що у формулі визначення впливу окремого чинника змінюється величина лише цього чинника, тоді як величини інших показників-факторів умовно приймаються однаковими.

Розглянемо розкладання приросту складного показника (вартість виробленої продукції) за факторами на прикладі табл. 11.7.

Таблиця 11.7

**Вплив чисельності робітників та продуктивності праці на приріст продукції**

| Показник | Базисний рік | Поточний рік |
|---|---|---|
| Продукція, млн грн | 307,2 | 323,0 |
| Середньорічна чисельність робітників, осіб | 1 200 | 1 212 |
| Продуктивність праці, тис. грн | 256 | 266,5 |

Для простоти візьмемо явище (або результативну ознаку), що визначається двома факторами, один з яких завжди буде первинною ознакою, другий — вторинним. У цьому прикладі результативною ознакою є обсяг виробленої продукції, а факторами — чисельність робітників і продуктивність їх праці. Між чисельністю робітників, продуктивністю праці та обсягом продукції існує наступний зв'язок: $N = R \cdot w,$ де $R$ — чисельність робітників (кількісний фактор), $w$ — продуктивність праці (якісний фактор), $N$ — вартість продукції.

При вивченні залежності обсягу випуску продукції на підприємстві від зміни чисельності працюючих та продуктивності праці можна скористатися наступною системою взаємопов'язаних індексів:

$$I_N = \frac{R_1 w_1}{R_0 w_0}, \tag{11.51}$$

$$I_N = \frac{R_1 w_0}{R_0 w_0} \cdot \frac{R_1 w_1}{R_1 w_0}. \tag{11.52}$$

Індекс середнього виробітку можна як $\frac{R_1 w_1}{R_1 w_0}$ (індекс якісного показника), а індекс чисельності працівників як $\frac{R_1 w_0}{R_0 w_0}$. Наведені формули показують, що загальна відносна зміна обсягу випуску продукції утворюється як добуток відносних змін двох факторів: чисельності працюючих та продуктивності праці. У першому співмножнику елімінoвано вплив чисельності працівників, у другому — продуктивності праці.

Добуток двох факторних індексів дорівнює загальному індексу виробництва продукції:

$$I_N = I_R \cdot I_w. \tag{11.53}$$

Обчислимо загальну зміну виробництва. Обсяг виробництва зріс у поточному періоді на 15,8 млн. грн. (323,0–307,2), або на 5,1 %, $I_{\text{пт}}$=(323,0 : 307,2)·100. Але зміна вартості продукції зумовлюється спільною зміною чисельності робітників та продуктивності праці.

Це відхилення утворилося під впливом змін середньої чисельності працюючих та продуктивності їхньої праці. Для визначення впливу чисельності працівників та продуктивності праці на вартість продукції необхідно розкласти загальну суму приросту вартості продукції $\Delta N$ на два доданки: на приріст продукції за рахунок збільшення чисельності робітників ($\Delta N_R$) і приріст продукції за рахунок зростання продуктивності праці ($\Delta N_w$). Щоб визначити, яка частина загальної зміни обсягу випуску продукції досягнута за рахунок зміни кожного їх факторів окремо, необхідно при розрахунку вплив одного з них елімінувати вплив іншого фактора.

У базисному періоді чисельність робітників склала 1 200 осіб. У звітному періоді чисельність робітників становила 1 212 осіб. Якби у звітному періоді продуктивність праці залишилася на рівні базисного періоду, то обсяг продукції становив би 310,3 млн грн (1212·226:1 000). Отже, за рахунок зміни чисельності робітників при колишньому рівні продуктивності праці вартість продукції зросла у звітному періоді порівняно з базисним періодом на 3,1 тис. грн, або в 1,01 рази:

$$\Delta N_R = 1212 \cdot 256\ 1200 \cdot 256 = 310{,}3 - 307{,}2 = 3{,}1 \text{ млн грн.}$$

Фактично у звітному періоді вартість продукції становила 323,0 млн грн. Різниця між фактичним обсягом виробництва продукції та умовним обсягом продукції, виробленою фактичним числом робітників з продуктивністю базисного періоду, дозволить визначити роль продуктивності праці в прирості виробництва продукції:

$$\Delta N_w = 1\,212 \cdot 266{,}5 - 1\,212 \cdot 256 = 323{,}0 - 310{,}3 = 12{,}7 \text{ млн грн.}$$

Таким чином, за рахунок збільшення чисельності робітників вартість продукції зросла на 3,1 млн грн, а за рахунок зростання продуктивності праці — на 12,7 млн грн, що в сумі дорівнює загальному приросту продукції:

$$\Delta N = N_R + N_w = 3{,}1 + 12{,}7 = 15{,}8 \text{ млн грн.}$$

Формули, що характеризують зміну виробництва продукції ( $\Delta N$ ) за факторами, мають такий вигляд:

а) за рахунок зміни кількісного фактора (чисельності працівників):

$$\Delta N_R = R_1 w_0 - R_0 w_0 = w_0 (R_1 - R_0); \qquad (11.54)$$

тобто зміна результативного показника (виробництва продукції) за рахунок зміни кількісного фактора (числа робітників) дорівнює приросту цього фактора, помноженому на базисний рівень якісного фактора (продуктивність праці);

б) за рахунок зміни якісного фактора (продуктивності праці):

$$\Delta N_w = w_1 R_1 - w_0 R_1 = R_1 (w_1 - w_0), \qquad (11.55)$$

тобто зміна результативного показника з допомогою якісного чинника дорівнює приросту цього чинника, помноженому на звітний рівень кількісного чинника.

Розглянемо (за даними табл. 11.7) розкладання за факторами темпу приросту вартості продукції (табл. 11.8).

Таблиця 11.8

**Вплив чисельності робітників та продуктивності праці на приріст продукції**

| Показник | Темп зростання, % | Темп приросту, % |
|---|---|---|
| Продукція, млн грн | 105,1 | 5,1 |
| Середньорічна чисельність робітників, осіб | 101,0 | 1,0 |
| Продуктивність праці, тис. грн | 104,1 | 4,1 |

Виходячи з наявних даних, визначимо, на скільки відсотків зростає обсяг виробництва продукції за рахунок середньооблікової чисельності працюючих і на скільки — за рахунок продуктивності праці.

Для цього обчислимо індивідуальний індекс зміни чисельності працюючих: $I_R = \frac{R_1}{R_0}$, звідси $R_1 = R_0 I_R$. Аналогічно обчислимо індекс зміни продуктивності праці: $I_w = \frac{w_1}{w_0}$, звідси $w_1 = w_0 I_w$.

Перетворимо вирази, що характеризують величину абсолютної зміни вартості продукції за рахунок кожного з фактором, наступним чином:

$$\Delta N_R = R_1 w_0 - R_0 w_0 = w_0(R_1 - R_0) = w_0(R_0 I_R - R_0) = N_0(I_R - 1);$$
$$\Delta N_w = w_1 R_1 - w_0 R_1 = R_1(w_1 - w_0) = R_0 I_R (w_0 I_w - w_0) =$$
$$= R_0 I_R w_0 (I_w - 1) = N_0 I_R (I_w - 1) = N_0 (I_R I_w - I_R) = N_0 (I_N - I_R). \qquad (15.56)$$

За рахунок зміни чисельності працівників вартість продукції збільшилася на 3,1 млн грн [307,2·(1,01 – 1)], а за рахунок зміни продуктивності праці — відповідно на 12,7 [307,2·(1,051 – 1,01)].

Розглянутий приклад розкладання абсолютного приросту (відхилення) узагальнюючого показника за факторами придатний для випадку, коли число факторів дорівнює двом (один з них — кількісний, інший — якісний), а аналізований показник представлений як їх добуток. Але такі випадки практично зустрічаються дуже рідко. Найчастіше на результативний показник впливають багато факторів. Відомо, що рівень таких узагальнюючих показників діяльності підприємства, як продуктивність праці, матеріаломісткість, фондовіддача, рентабельність, прибуток та ін., складається під впливом ряду виробничих факторів.

Наприклад, місячна продуктивність праці може бути представлена у вигляді наступних факторів-співмножників:

$$W_т = w_р \cdot d_р = w_д \cdot д \cdot d_р = w_ч \cdot ч \cdot д \cdot d_р. \qquad (11.57)$$

де $W_т$ — середній валовий виробіток одного працівника промислово-виробничого персоналу;
$w_р$ — середній валовий виробіток одного робітника промислово-виробничого персоналу;
$d_р$ — частка робітників у всьому промислово-виробничому персоналі;
$w_д$ — середній денний валовий виробіток на одного робітника;
$д$ — середня кількість днів роботи одного робітника за місяць;
$w_ч$ — середній годинний валовий виробіток одного робітника;
$ч$ — середня тривалість робочого дня.

У даній моделі відображено вплив трьох груп факторів:

а) інтенсивних, що характеризують особливості конструкцій виробів, що випускаються, рівень техніки і технології виробництва (показник $w_ч$);

б) екстенсивних, що відображають екстенсивне використання ресурсів, тобто стан організації виробництва і праці (показники *д* і *ч*);

в) характеризують структурні чинники, у разі співвідношення чисельності робочих і загальна кількість працюючих (показник $d_p$).

Продуктивність праці завжди результат дії ряду факторів.

Зміна середнього місячного виробітку одного працівника за рахунок кожного з перерахованих вище факторів:

$$I_W = \frac{w_{ч1} ч_1 д_1 d_1}{w_{ч0} ч_0 д_0 d_0} = I_w \cdot I_ч \cdot I_д \cdot I_d. \tag{11.58}$$

В результаті аналізу впливу факторів на абсолютний приріст продуктивності праці повинен бути розкладений на доданки (число доданків дорівнює числу факторів):

Різниця між чисельником і знаменником кожного множника правої частини цієї формули ( $w_1 ч_1 д_1 d_1 - w_0 ч_0 д_0 d_0$ ) показує абсолютну зміну валового виробітку на одного працюючого за рахунок відповідного показника.

1. Зміна місячного виробітку одного працівника за рахунок зміни середнього годинного виробітку робітників:

$$I_w = \frac{w_{ч1} ч_1 д_1 d_1}{w_{ч0} ч_1 д_1 d_1}, \tag{11.59}$$

або в абсолютному вираженні:

$$\Delta W_w = w_1 ч_1 д_1 d_1 - w_0 ч_1 д_1 d_1. \tag{11.60}$$

2. Зміна середнього місячного виробітку одного робітника за рахунок зміни середньої тривалості робочого дня:

$$I_w = \frac{w_{ч0} ч_1 д_1 d_1}{w_{ч0} ч_0 д_1 d_1}, \tag{11.61}$$

або в абсолютному вираженні:

$$\Delta W_ч = w_{ч0} ч_1 д_1 d_1 - w_0 ч_0 д_1 d_1. \tag{11.62}$$

3. Зміна середнього місячного виробітку одного працівника за рахунок зміни середньої кількості днів роботи одного робітника за місяць:

$$I_w = \frac{w_{ч0} ч_0 д_1 d_1}{w_{ч0} ч_0 д_0 d_1}, \tag{11.63}$$

або в абсолютному вираженні:

$$\Delta W_д = w_{ч0} ч_0 д_1 d_1 - w_{ч0} ч_0 д_0 d_1. \tag{11.64}$$

4. Зміна середнього місячного виробітку одного працівника за рахунок зміни питомої ваги робітників у загальній чисельності працівників:

$$I_w = \frac{w_{ч0} ч_0 д_0 d_1}{w_{ч0} ч_0 д_0 d_0}, \tag{11.65}$$

або в абсолютному вираженні:

$$\Delta W_д = w_{ч0} ч_0 д_0 d_1 - w_{ч0} ч_0 д_0 d_0. \tag{11.66}$$

В результаті аналізу впливу факторів на абсолютний приріст продуктивності праці повинен бути розкладений на доданки (число доданків дорівнює числу факторів):

$$\Delta W_m = W_1 - W_0 = \Delta W_{w_ч} + \Delta W_ч + \Delta W_д + \Delta W_d. \tag{11.67}$$

Приріст, отриманий за допомогою впливу кожного фактора, повинен відповісти на питання: наскільки зросла продуктивність праці (результативний показник) внаслідок зміни лише даного фактора?

Кожен із множників формули (11.58) характеризує, у скільки разів зріс (зменшився) би валовий виробіток одного працюючого в результаті зміни відповідного показника-множника, наприклад, годинного виробітку одного робітника, якщо інші показники — середня тривалість робочого дня, середня кількість днів роботи одного робітника і частка робітників у загальній чисельності працюючих — залишилися б без змін.

Наприклад, по підприємству відомі такі дані (табл. 11.9).

Таблиця 11.9

| Показник | Період | |
|---|---|---|
| | базисний | звітний |
| Середній годинний виробіток одного робітника, грн | 200 | 225 |
| Середня фактична тривалість робочого дня, год. | 7,9 | 8 |
| Середня фактична кількість днів роботи одного робітника за місяць, днів | 22 | 21 |
| Частка робітників у загальній чисельності працівників | 0,78 | 0,76 |

Виходячи з наведених даних необхідно обчислити вплив зміни середнього годинного виробітку одного робітника, середньої тривалості робочого дня, тривалості робочого місяця та частки робітників у загальній чисельності працівників.

Розрахуємо виробіток продукції на одного працюючого в базисному та звітному періодах:

$$\Delta W_0 = w_0 ч_0 д_0 d_0 = 200 \cdot 7{,}9 \cdot 22 \cdot 0{,}78 = 27\,112{,}8 \text{ грн.}$$

$$\Delta W_1 = w_1 ч_1 д_1 d_1 = 225 \cdot 8{,}0 \cdot 21 \cdot 0{,}76 = 28728{,}0 \text{ грн.}$$

Загальний приріст середнього місячного виробітку одного працюючого складе:

$$\Delta W_т = W_1 - W_0 = 28728{,}0 - 27112{,}8 = 1615{,}2 \text{ грн,}$$

або на 6,0 % $\left( I_W = \frac{28\,728{,}0}{27\,112{,}8} = 1{,}06 \text{ або } 6{,}0\ \% \right)$.

Визначаємо зміну середнього місячного виробітку одного працівника за рахунок зміни середнього годинного виробітку робітників:

$$\Delta W_w = w_1 ч_1 д_1 d_1 - w_0 ч_1 д_1 d_1 = 225 \cdot 8 \cdot 21 \cdot 0{,}76 - 200 \cdot 8 \cdot 21 \cdot 0{,}76 = \\ = 28\,728{,}0 - 25\,536{,}0 = 3\,192{,}0 \text{ грн.}$$

Таким чином, за рахунок зростання середнього годинного виробітку на 25 грн (225 – 200) середній місячний виробіток одного працівника збільшився на 3 192,0 грн. Відносна величина приросту середнього місячного виробітку працівників за рахунок відповідного фактора визначається відношенням абсолютного приросту за рахунок цього фактора до базисного рівня середнього місячного виробітку. Так, приріст продукції за рахунок збільшення годинного виробітку на одного робітника збільшилася на 11,8 % $\left( \frac{3192}{27112{,}8} \times 100 \right)$.

Визначаємо зміну середнього місячного виробітку одного працівника за рахунок зміни фактичної тривалості робочого дня:

$$\Delta W_ч = w_{ч0} ч_1 д_1 d_1 - w_0 ч_0 д_1 d_1 = 200 \cdot 8 \cdot 21 \cdot 0{,}76 - 200 \cdot 7{,}9 \cdot 21 \cdot 0{,}76 = \\ = 25\,536{,}0 - 25\,216{,}8 = 319{,}2 \text{ грн,}$$

тобто за рахунок покращення використання внутрішньозмінного робочого часу виробіток одного працівника збільшився на 319,2 грн, або на 1,2 % $\left( \frac{319{,}2}{27\,112{,}8} \times 100 \right)$.

Визначимо зміну середнього місячного виробітку одного працівника за рахунок зміни тривалості робочого місяця:

$$\Delta W_{\partial} = w_{ч0} ч_0 \partial_1 d_1 - w_{ч0} ч_0 \partial_0 d_1 = 200 \cdot 7{,}9 \cdot 21 \cdot 0{,}76 - 200 \cdot 7{,}9 \cdot 22 \cdot 0{,}76 =$$

$$= 25\,216{,}8 - 26\,417{,}6 = -1200{,}8 \text{ грн},$$

тобто за рахунок скорочення тривалості робочого місяця середній місячний виробіток одного працівника скоротився на 1 200,8 грн, або на 4,4 % $\left(\frac{-1\,200{,}8}{27\,112{,}8} \times 100\right)$.

Визначимо зміну середнього місячного виробітку одного працівника за рахунок зміни частки робітників у загальній чисельності працівників підприємства:

$$\Delta W_{\partial} = w_{ч0} ч_0 \partial_0 d_1 - w_0 ч_0 \partial_0 d_0 = 200 \cdot 7{,}9 \cdot 22 \cdot 0{,}76 - 200 \cdot 7{,}9 \cdot 21 \cdot 0{,}78 =$$

$$= 26\,417{,}6 - 27\,112{,}8 = -695{,}2 \text{ грн},$$

тобто за рахунок зменшення частки робітників у загальній чисельності працівників середньомісячний виробіток одного працівника скоротився на 695,2 грн, або на 2,6 % $\left(\frac{-695{,}2}{27\,112{,}8} \times 100\right)$.

Таким чином, загальний приріст виробітку одного працівника за рахунок усіх факторів склав:

$$\Delta W_m = \Delta W_{w_ч} + \Delta W_ч + \Delta W_{\partial} + \Delta W_d =$$

$$= 3\,192{,}0 + 319{,}2 + (-1200{,}8) + (-695{,}2) = 1615{,}2 \text{ грн}.$$

Отриманий результат збігається із загальною величиною фактичної зміни.

Таким чином, індексний метод дозволяє здійснити розкладання загального приросту узагальнюючого показника на складові фактори, що надають найбільший вплив на його зміну. У нашому прикладі обчислено величину абсолютного відхилення (приросту) узагальнюючого показника — середнього місячного виробітку одного працівника — на окремі фактори — зміну середнього годинного виробітку одного робітника, середньої тривалості робочого дня, тривалості робочого місяця і частки робітників у загальній чисельності працівників.

Тим не менш, використання індексів в економічному аналізі для більш-менш тривалого періоду наштовхується на серйозні перешкоди в умовах швидкої зміни асортименту продукції. В умовах швидкої зміни асортименту продукції важко визначити ваги для розрахунку агрегатних індексів.

## 11.7. Особливості територіальних індексів

Відносний показник, що характеризує співвідношення соціально-економічних явищ у просторі (по економічним районам, містам, державам і т. п.), носить назву *територіального індексу*, або *просторового індексу*. Подібно до індексів динаміки, територіальні індекси відображають відносну зміну рівня явища, його фізичного (натурального) обсягу і т. п. від однієї території до іншої (або до кількох інших).

Побудова територіальних індексів на відміну індексів динаміки має деякі особливості. Розвиток явища у часі відбувається в одному певному, поступальному напрямку. Наприклад, є два періоди: 2015 та 2020 рр. Для цього періоду можна обчислити індекси цін та індекси фізичного обсягу продукції. При цьому за базу порівняння буде прийнято 2015 р. Отримаємо такі індекси, як індекси цін $I^p_{20/15}$ та індекс фізичного обсягу $I^q_{20/15}$. Не можна при цьому побудувати зворотні індекси ($I^p_{15/20}$ або $I^q_{15/20}$) з тимчасовою базою 2020 р. Побудова таких індексів не має економічного змісту.

Навпаки, в територіальних індексах при зіставленні будь-якого явища на територіях *А* і *Б* за просторову базу порівняння можна з рівною підставою можна прийняти рівень явища (або його обсяг, структуру) на будь-якій території. Отже, має зміст індексів цін $I^p_{А/Б}$ та $I^p_{Б/А}$.

Питання про обчислення територіальних індексів стало особливо актуальним при розширенні прав місцевих органів влади. Багато статистичних показників (валовий внутрішній продукт, ціни, доходи населення та ін.) зараз обчислюються по областях, економічних районах. Значний інтерес представляє використання територіальних індексів при міжнародному співставленні.

При побудові територіальних індексів особливу складність становить вибір бази порівняння. Так, при порівнянні цін на товари ринків двох міст насамперед виникає питання про вибір ваг (бази порівняння) — кількості проданих товарів якого міста слід прийняти як ваги. Виникнення складнощів при побудові територіальних індексів пов'язане ще й тому, що проводиться зіставлення різних об'єктів, що мають зазвичай різко різну структуру і рівні якісних ознак на відміну від динамічних індексів, де вивчається розвиток одне і того ж самого об'єкта.

Ці питання вирішуються у кожному конкретному випадку. Але слід мати на увазі, що в територіальних індексах проводяться порівняння складних явищ, які мають різну структуру. Тому ваги мають певною мірою нейтралізувати наявні структурні відмінності. В силу цього для територіальних індексів як ваги правомірно використовувати деякі середні величини (галузеві і т. п.) за принципом стандартизованих коефіцієнтів. Наприклад, зіставимо собівартість продукції першої області з другої. Вихідна інформація наведена в табл. 11.10.

Таблиця 11.10

| Продукція | Область *А* | | Область *Б* | |
|---|---|---|---|---|
| | собівартість одиниці продукції, грн ($z_A$) | вироблено, шт. ($q_A$) | собівартість одиниці продукції, грн ($z_Б$) | вироблено, шт. ($q_Б$) |
| 1 | 10,4 | 160 | 10,6 | 670 |
| 2 | 9,7 | 220 | 9,2 | 210 |

Розрахуємо територіальний індекс собівартості продукції, використовуючи як ваги обсяги виробництва продукції в області *А*:

$$I_z = \frac{\sum z_A q_A}{\sum z_Б q_A} = \frac{10{,}4 \cdot 160 + 9{,}7 \cdot 220}{10{,}6 \cdot 160 + 9{,}2 \cdot 220} = \frac{3798}{3720} = 1{,}021, \text{ або } 102{,}1~\%.$$

Отже, собівартість продукції області *А* вище, ніж у області *Б*, на 2,1 % ($1{,}021 \cdot 100 - 100$), якщо вагами брати кількість продукції області $A$.

Територіальний індекс собівартості з використанням як ваги обсягу продукції, виробленої в області *Б*:

$$I_z = \frac{\sum z_Б q_Б}{\sum z_A q_Б} = \frac{10{,}6 \cdot 670 + 9{,}2 \cdot 210}{10{,}4 \cdot 670 + 9{,}2 \cdot 210} = \frac{9034}{9005} = 1{,}003, \text{ або } 100{,}3~\%.$$

Отримали суперечливий результат: перший індекс свідчить про те, що собівартість продукції області *А* вище, ніж області *Б*, другий — цілком протилежне. Це з тим, що вибір ваг (у першому випадку — $q_A$, у другому — $q_Б$) не нейтралізує структурних відмінностей у виробництві продукції, що у областях *А* і *Б*.

Існують різні точки зору на вибір баз при обчисленні територіальних індексів. Іноді як ваги пропонується використовувати і суми первинних ознак порівнюваних сукупностей. Наприклад, в індексі собівартості продукції — за сумою обсягу виробництва продукції обох територій, тобто $q_A + q_Б$.

Якщо як ваги використовуються обсяги відповідних видів продукції, територіальний індекс собівартості може бути обчислений за наступною формулою:

$$I_z = \frac{\sum z_A q_{A+Б}}{\sum z_Б q_{A+Б}} = \frac{10{,}4 \cdot (160 + 670) + 9{,}7 \cdot (220 + 210)}{10{,}6 \cdot (160 + 670) + 9{,}2 \cdot (220 + 210)} = \frac{12803}{12754} =$$

$$=1{,}004, \text{ або } 100{,}4 \ \%.$$

Крім загальноприйнятого порядку — порівнювати досліджувані дані з певним конкретним об'єктом, рекомендується в якості бази використання в індексах первинних ознак (скажімо, фізичного обсягу) середньої собівартості (ціни) кожного виробу за двома територіями, що порівнюються. У цьому випадку розрахунок здійснюється в наступній послідовності:

$$\bar{z}_1 = \frac{\sum z_1 q_1}{\sum q_1} = \frac{10{,}4 \cdot 160 + 10{,}6 \cdot 670}{160 + 670} = 10{,}56;$$

$$\bar{z}_2 = \frac{\sum z_2 q_2}{\sum q_2} = \frac{9{,}7 \cdot 220 + 9{,}2 \cdot 210}{220 + 210} = 9{,}45;$$

$$I_{z'} = \frac{\sum z_A q_A}{\sum \bar{z}_i q_A} = \frac{3\,798}{10{,}56 \cdot 160 + 9{,}45 \cdot 220} = 1{,}008, \text{ або } 100{,}8 \ \%;$$

$$I_{z''} = \frac{\sum z_Б q_Б}{\sum \bar{z}_i q_Б} = \frac{9\,034}{10{,}56 \cdot 670 + 9{,}45 \cdot 210} = 0{,}997, \text{ або } 99{,}7 \ \%.$$

Отримані результати не суперечать одне одному.

Є пропозиції індекси якісних ознак розраховувати за вагами того об'єкта, для якого обчислюється індекс, а індекси об'ємних (кількісних) ознак Ї за вагами середнього рівня відповідних якісних показників по всій сукупності даних об'єктів.

Слід зазначити, що на даний час методологію вибору бази при розрахунку територіальних індексів ще до кінця не вирішено. Це питання має вирішуватися відповідно до поставленого перед дослідженням завданням. Особливого аналізу вимагає проблема зіставлення територіальних індексів як уніфікованого інструменту для порівняння показників розвитку різних країн. Тут виникають складні питання зіставлення валют, порівнянності структур економіки та методології обчислення показників у різних країнах і т. п.

## 11.8. Особливості індексів виконання плану

*Індекс виконання плану* характеризує ступінь виконання плану і представляє собою зіставлення фактично досягнутого рівня показника до планового. Як ваги індексу виконання плану можуть бути як планові показники, так і фактичні.

Методологічні принципи обчислення індексів, розглянуті в попередніх розділах, відносяться до всіх індексів (до індексів динаміки, до територіальних індексів, до індексів структури). Однак переважно індексний метод розроблено для індексів динаміки. Разом з тим побудова індексів виконання плану має свої особливості. Зазначимо тут деякі з них.

Головна особливість побудови індексів виконання плану полягає в тому, що в планових індексах вторинних ознак (ціни, собівартості і т. п.) теоретично слід би використовувати не звітні ваги, як того вимагає природа вторинних ознак, а планові.

Індекс, що виражає відношення фактичної собівартості продукції до її планової собівартості, обчислюється за формулою:

$$I_{z_1 / z_{\text{пл}}} = \frac{\sum q_1 z_1}{\sum q_1 z_{\text{пл}}}, \qquad (11.68)$$

де $q_1$ — кількість одиниць продукції кожного виду у звітному періоді;

$z_1$ — фактична собівартість одиниці виробленої продукції у звітному періоді;

$z_{\text{пл}}$ — планова собівартість одиниці виробленої продукції.

Різниця між цим індексом (вираженим у відсотках) і 100 представляє індекс відхилення фактичної собівартості від планової, а різниця між чисельником і знаменником дробу — надпланову економію (або перевитрату) від зниження собівартості.

На зміну індексу виконання плану впливає як виконання плану собівартості за кожним виробом, так і зрушення в асортименті продукції (фактично проти планового). Оскільки план виробу обчислюється за плановим асортиментом, то для аналізу виконання планових зобов'язань обчислюють індекс виконання планового завдання. Цей індекс є різновидом індексу виконання плану та обчислюється за формулою:

$$\text{індекс планового завдання } I_{z_{\text{пл}} / z_0} = \frac{\sum q_{\text{пл}} z_{\text{пл}}}{\sum q_{\text{пл}} z_0}, \qquad (11.69)$$

де $q_{пл}$ — планова кількість одиниць продукції кожного виду.

Щоб усунути вплив на величину індексу виконання плану зміни планової структури (асортименту) продукції по собівартості, як ваги слід було б використовувати плановий обсяг виробництва продукції:

$$I_{z_1/z_{пл}} = \frac{\sum q_{пл} z_1}{\sum q_{пл} z_{пл}}. \quad (11.70)$$

Для оцінки впливу зрушень фактичного асортименту виробів ($q_1$) порівняно із запланованим ($q_{пл}$) на показник виконання плану за собівартістю обчислюють індекс впливу структурних зрушень. Цей індекс є відношенням двох різнозважених індексів виконання плану по собівартості:

$$I_{ас.\,зр} = \frac{\sum q_1 z_1}{\sum q_1 z_{пл}} : \frac{\sum q_{пл} z_1}{\sum q_{пл} z_{пл}}. \quad (11.71)$$

Якщо значення індексу більше одиниці, то зміна фактичного асортименту в порівнянні із запланованим вплинула у бік збільшення індексу собівартості, якщо він менше одиниці, то структурні зрушення в асортименті продукції призвели до зниження величини індексу собівартості.

## 11.9. Найважливіші економічні індекси та їх взаємозв’язок

Загальні основи методики індексного методу допомагають глибше зрозуміти особливості обчислення індексів окремих економічних показників. Почнемо розгляд найважливіших економічних індексів із індексу цін.

**Індекси цін**. Серед економічних індексів велике поширення отримали індекс цін. Ці індекси використовуються для оцінки динаміки цін, для розрахунку інфляції в країні, а також для визначення впливу цінового фактора на обсяг продукції (товарообіг) підприємства. Основне призначення індексу цін — показати зміну цін як таких, безвідносно до зміни структури.

*Індекс цін* характеризує середню зміну цін на товари та послуги. В Україні обчислюються індекси роздрібних цін (індекси споживчих цін), індекси оптових цін (промислових і сільськогосподарських), індекси цін послуги зв’язку і т. п. Використовується при аналізі динаміки цін на продукцію (товари, послуги).

Індекс цін обчислюється або за формулою агрегатного індексу:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}, \tag{11.72}$$

або за формулою гармонійного індексу:

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{1}{i} p_1 q_1}, \tag{11.73}$$

де $p_0$ і $p_1$ — ціни базисного та звітного періодів;

$q_1$ — кількість продуктів (товарів) у звітному періоді;

$i$ — індекси цін окремих товарів.

Чисельник індексу характеризує вартість товарів та послуг, виражених у цінах звітного періоду, знаменник — той же фізичний обсяг товарів та послуг у цінах базисного періоду. Вагами при обчисленні індексу цін виступає фізичний обсяг продуктів (товарів та послуг) звітного періоду.

Обидві формули дають тотожні результати. Вибір форми індексу визначається вихідними даними. Агрегатна форма індекс використовується, коли кількість товарів вимірюється в натуральних одиницях. Коли обсяг товарів враховується у вартісних одиницях, індекс цін розраховується за формулою гармонійного індексу. На практиці агрегатна форма цього індексу не використовується, так як облік вартості товарів проводиться в вартісних одиницях.

Індекс цін обчислюється окремих видів цін і відображає тим самим динаміку їх рівня у різних сферах грошового обігу. Основні з них: індекс споживчих цін, індекс цін виробників промислової продукції, індекс цін у будівництві, індекс тарифів на послуги пошти та зв'язку, індекс цін на житло, індекс цін зовнішньої торгівлі.

*Індекс споживчих цін (ІСЦ)* — показник, що характеризує зміни у часі загального рівня цін на товари та послуги, які купує населення для невиробничого споживання. Відображає зміни вартості фіксованого набору споживчих товарів та послуг у поточному періоді порівняно з базисним. Він не враховує структурні зміни у споживанні окремих товарів. Індекс споживчих цін займає центральне місце в системі показників статистики цін і розраховується в Україні, починаючи з серпня 1991 р. ІСЦ застосовується у багатьох сферах соціально-економічного життя країни. Зокрема, ІСЦ звичайно використовують у якості показника загального рівня інфляції в економіці, для аналізу і прогнозу динаміки цін, перегляду (індексації) розмірів грошових доходів, перерахунку господарської заборгованості за судовими спорами та ін.

Показник реальної заробітної плати, якому в нашій країні приділяється значна увага, тісно пов'язаний із індексом цін. Поняття реальної заробітної плати не тотожнім поняттю рівня життя, але вона є самим суттєвим фактором, що визначає його рівень. На величину та динаміку реальної заробітної плати впливають рівень номінальної заробітної плати, зміни споживчих цін, а також розмір податків і обов'язкових внесків, що сплачують наймані працівники. Перерахунок номінальної заробітної плати в реальну здійснюється наступним чином:

$$\text{Реальна заробітна плата} = \frac{\text{Номінальна заробітна плата}}{\text{Індекс споживчих цін}}. \qquad (11.74)$$

Динаміка реальної заробітної плати визначається за допомогою індексу. Індекс реальної заробітної плати розраховується за формулою[1]:

$$I_{t/0}^{\text{рзп}} = \frac{I_{t/0}^{\text{нзп}}}{I_{t/0}^{\text{ц}}},$$

де $I_{t/0}^{\text{нзп}}$ — індекс нарахованої номінальної заробітної плати «нетто» поточного періоду ($t$) порівняно з базисним (0);

$I_{t/0}^{\text{ц}}$ — індекс споживчих цін поточного періоду ($t$) порівняно з базисним (0).

Основною умовою підвищення реальної заробітної плати є випереджаюче зростання номінальної заробітної плати порівняно із зростанням цін.

*Індекс цін виробників промислової продукції (ІЦВ)* характеризує зміни цін у часі у сфері промислового виробництва, який дозволяє відслідковувати та визначати тенденції змін цін як за видами промислової діяльності, так і у виробництві конкретної продукції. Він використовується для інформаційного забезпечення прогнозування й управління процесами ціноутворення в промисловості, потреб статистики промисловості, оптової торгівлі та національних рахунків.

---

[1] Методика розрахунку індексів реальної заробітної плати: Наказ Державної служби статистики України від 11.09.2012 р. № 1950/22262. URL. https://zakon.rada.gov.ua/laws/show/z1950-12#Text.

Агрегатні індекси цін розраховуються на основі індивідуальних індексів цін та вагових коефіцієнтів за формулою типу Ласпейреса:

$$I_t^{t-1} = \frac{\sum_j \overline{cw}_{j,0} \times t_{j,t}^{t-1}}{\sum_j \overline{cw}_{j,0}}, \qquad (11.75)$$

де $I_t^{t-1}$ — агрегатний індекс цін у звітному місяці ($t$) порівняно з попереднім ($t-1$);

$\overline{cw}_{j,0}$ — вага продукції ($j$) в базисному періоді.

*Індекс цін продукції сільського господарства, реалізованої підприємствами* характеризує зміну середніх цін продукції сільського господарства, реалізованої підприємствами у часі. Розрахунок індексу здійснюють на основі даних про кількість та вартість реалізованої продукції власного виробництва, реалізованої сільськогосподарськими підприємствами. Перелік видів продукції, що включається до розрахунку індексу, визначають за Номенклатурою продукції сільського господарства. Основою для порівняння зміни середніх цін продукції сільського господарства, реалізованої підприємствами є індекси цін окремих видів продукції сільського господарства, реалізованої підприємствами у звітному місяці поточного року до середньорічних цін відповідних видів продукції, реалізованої у базисному періоді (році).

*Індекс цін у будівництві (ІЦБ)* характеризує динаміку загального рівня цін на будівництво у часі. Він є показником зміни вартості фіксованого набору матеріальних ресурсів у поточному періоді порівняно з базисним. ІЦБ розраховуються за ресурсно-технологічними моделями об'єктів-представників, відібраних за типами будівель і споруд, ресурсно-індексним методом.

*Індекс тарифів на вантажні перевезення залізничним транспортом* характеризує зміни у часі рівня тарифів на вантажні перевезення при незмінній структурі перевезення вантажів за різними ознаками: виду вантажу і його маси, швидкості доставки, відстані перевезення вантажів, типу рухомого складу, рівню використання його вантажопідйомності та ін. Він є показником зміни вартості фіксованого набору перевезених вантажів у поточному періоді порівняно з базисним. Розрахунки індексів тарифів здійснюється на державному рівні по країні в цілому за послугами-представниками відповідно до Класифікації видів вантажів.

*Індекс тарифів на послуги пошти та зв'язку* характеризує зміни в часі рівня тарифів на послуги пошти та зв'язку для підприємств, установ, організацій при незмінній структурі надання послуг пошти та зв'язку за певний період часу. Він є показником зміни вартості фіксованого набору послуг пошти та зв'язку в поточному періоді порівняно з базисним.

*Індекс цін зовнішньої торгівлі* характеризує зміну вартості експорту (імпорту) за рахунок зміни цін при постійному фізичному обсязі зовнішньої торгівлі. Для його розрахунку співставляється вартість експорту (імпорту) зі вартістю товарів, вивезених (ввезених) у звітному періоді, за цінами базисного періоду, тобто

$$I = \frac{\sum p_1 q_1}{\sum p_0 q_1}, \tag{11.76}$$

де $p_0$ — ціна одиниці експорту (імпорту) товару у базисному періоді;

$p_1$ — ціна одиниці експорту (імпорту) товару у звітному періоді;

$q_1$ — кількість експортованих (імпортованих) товарів у звітному періоді.

Співвідношення змін цін у зовнішній торгівлі товарами визначає «умови торгівлі». Цінові показники умов торгівлі розраховуються за формулою:

$$T_p = \frac{J_{pc}}{J_{pi}}, \tag{11.77}$$

де $T_p$ — індекс цінових умов торгівлі;

$J_{pc}$ — зведений індекс середньої вартості експорту за поточний період порівняно з базисним періодом;

$J_{pi}$ — зведений індекс середньої вартості імпорту за поточний період порівняно з базисним періодом.

Якщо ціни на експортовані товари ростуть швидше, ніж на імпортовані, то країна має можливість менше експортувати для оплати такого ж обсягу імпорту. Випередження зростання цін на імпортовані товари в порівнянні з цінами на експортовані товари та послуги веде до протилежного результату.

**Індекси собівартості продукції**. Важливим якісним показником роботи підприємств є собівартість своєї продукції. Від рівня собівартості продукції значною мірою залежить величина одержуваного прибутку, а також рентабельність підприємства. Собівартість продукції зрештою впливає на рівень оптових цін.

Узагальнюючим показником динаміки собівартості є індекс собівартості. *Індекс собівартості продукції* — відносний показник, що характеризує зміну собівартості продукції в даному періоді в порівнянні з її середнім рівнем у минулому періоді або відношення фактичної собівартості до її планової собівартості. В залежності від вибору бази порівняння використовують середні ціни, що фактично склалися в базисному періоді або середні планові ціни. Оскільки облік усередині підприємства завжди дозволяє безпосередньо чи опосередковано визначити собівартість одиниці окремого виду продукту, то для розрахунку індексу собівартості безпосередньо використується агрегатна форма.

Загальний індекс собівартості продукції ($I_c$) визначається за наступними формулами.

*Загальний індекс динаміки собівартості порівнянної продукції*, що виражає відношення собівартості продукції в даному періоді до її середнього рівня в минулому періоді, визначається як індекс постійного складу за наступною формулою:

$$I_c = \frac{\sum z_1 q_1}{\sum z_0 q_1}, \tag{11.78}$$

де $q_1$ — кількість одиниць продукції кожного виду у звітному періоді;
$z_1$ — фактична собівартість одиниці виробленої продукції у звітному періоді;
$z_0$ — собівартість одиниці виробленої продукції в базисному періоді.

Цей індекс розраховується лише за порівнянною продукцією. До порівнянної продукції належать ті її види, які вироблялися як у поточному році, так і в попередній (базисний) період у серійному або масовому порядку. Цей індекс відображає зміну собівартості, що обумовлена лише зміною собівартості одиниці продукції.

*Загальний індекс виконання плану за собівартістю порівнянної та незрівнянної продукції*:

$$I'_c = \frac{\sum z_1 q_1}{\sum z_{\text{пл}} q_{\text{пл}}}, \tag{11.79}$$

де $z_{\text{пл}}$ і $z_1$ — собівартість одиниці виробленої продукції відповідно до плану і фактична;
$q_{\text{пл}}$ і $q_1$ — плановий та звітний обсяг випуску продукції у відповідних одиницях.

Цей індекс охоплює всі види продукції, передбачені планом і які фактично випускалися у звітному періоді (а не лише порівнянну з минулим роком продукцію). Обчислення індексу виконання плану по собівартості здійснюють в аналітичних цілях. Економічний зміст цих індексів полягає у тому, що вони дозволяють визначити зміну собівартості продукції залежно від зміни собівартості одиниці виробленої продукції.

Різниця між чисельником і знаменниками характеризує надпланову економію (або перевитрату) від зниження собівартості. Проте розрахунок його наштовхується на серйозні перешкоди в умовах технічного прогресу та швидкого оновлення асортименту продукції. Особливістю індексу собівартості порівняно з індексом роздрібних цін є те, що він розраховується завжди за строго визначеним асортиментом продукції.

Процент зниження собівартості (планове завдання щодо зниження собівартості) на підприємстві обчислюється за формулою:

$$\frac{\sum z_{\text{пл}} q_{\text{пл}}}{\sum z_0 q_{\text{пл}}} \cdot 100 - 100. \qquad (11.80)$$

Індекс виконання плану за собівартістю характеризує ступінь виконання плану і є відношенням фактично досягнутого рівня показника до планового. Вагами індексу виконання плану можуть бути як планові показники, так і фактичні:

$$I_{\text{в. п}} = \frac{\sum z_1 q_1}{\sum z_{\text{пл}} q_1}; \qquad I_{\text{в. п}} = \frac{\sum z_1 q_{\text{пл}}}{\sum z_{\text{пл}} q_{\text{пл}}}. \qquad (11.81)$$

Найчастіше застосовується перша з наведених формул, що усуває вплив на величину індексу виконання плану зміни планової структури (асортименту) продукції.

За допомогою індексного методу ми можемо піддати аналізу динаміку якісних показників, розкласти відповідний індекс на два: на індекс змін за рахунок норм витрат на одиницю продукції та за рахунок зміни цін одиниці цих витрат. Так, індекс вартості основних матеріалів ($I_{\text{м}}$) може бути обчислений за наступною формулою:

$$I_{\text{м}} = \frac{\sum S_1 q_1}{\sum S_0 q_1}, \qquad (11.82)$$

де $S_0$ і $S_1$ — витрати на матеріали (на одиницю продукції).

Витрати на основні матеріали можна визначити як добуток норми витрати на ціну матеріалів:

$$S_0 = H_0 Ц_0; \quad S_1 = H_1 Ц_1; \tag{11.83}$$

де $H_0$ і $H_1$ — норми витрат матеріалів у базисному та звітному періодах на одиницю продукції;

$Ц_0$ і $Ц_1$ — ціни одиниці виробленої продукції в базисному і звітному періодах.

Тоді наведений вище індекс вартості основних матеріалів може бути представлений в наступному вигляді:

$$I_{\text{м}} = \frac{\sum q_1 H_1 Ц_1}{\sum q_1 H_0 Ц_0}. \tag{11.84}$$

Індекс вартості матеріалів можна розкласти на два взаємопов'язані індекси: індексу норм та індексу цін. Взаємозв'язок індексів представлений у наступному вигляді:

$$\frac{\sum q_1 H_1 Ц_0}{\sum q_1 H_0 Ц_0} \times \frac{\sum q_1 H_1 Ц_1}{\sum q_1 H_1 Ц_0} = \frac{\sum q_1 H_1 Ц_1}{\sum q_1 H_0 Ц_0}. \tag{11.85}$$

Важливим показником, який застосовується у вивченні динаміки собівартості продукції, є витрати на 1 грн товарної продукції. Рівень цих витрат обчислюється як відношення собівартості товарної продукції до її вартості оптових цін підприємства:

$$s = \frac{\sum qz}{\sum qp}, \tag{11.86}$$

де $p$ — оптова ціна одиниці продукції;

$q$ — кількість одиниць продукції кожного виду.

Витрати на одиницю продукції можуть бути обчислені за такими формулами:

планові витрати —

$$s_{\text{пл}} = \frac{\sum q_{\text{пл}} z_{\text{пл}}}{\sum q_{\text{пл}} p_{\text{пл}}}; \tag{11.87}$$

фактичні витрати:

$$s_1 = \frac{\sum q_1 z_1}{\sum q_1 p_1}; \tag{11.88}$$

фактичні витрати у планових оптових цінах:

$$s_1' = \frac{\sum q_1 z_1}{\sum q_1 p_{\text{пл}}};$$

витрати за планом, перерахованим на фактичний обсяг та асортимент продукції:

$$s'_{\text{пл}} = \frac{\sum q_1 z_{\text{пл}}}{\sum q_1 p_{\text{пл}}}; \tag{11.89}$$

Оцінка виконання плану витрат на одну гривню товарної продукції проводиться на підставі виконання плану по витратам на одну гривню товарної продукції, виходячи з планового завдання, перерахованого на фактичний випуск і асортимент продукції.

$$I_{\text{в. п}} = \frac{\sum q_1 z_1}{\sum q_1 p_{\text{пл}}} : \frac{\sum q_1 z_{\text{пл}}}{\sum q_1 p_{\text{пл}}}. \tag{11.90}$$

**Індекс фізичного обсягу продукції.** *Індекс фізичного обсягу продукції* характеризує зміну маси вироблених матеріальних благ у поточному періоді порівняно з базисним. Використовується при аналізі динаміки вартості фізичного обсягу продукції. Цей індекс відображає зміну вартості всієї продукції, що викликана зміною тільки обсягу продукції, і не пов’язане зі зміною цін. Вплив зміни цін на індекс виключається шляхом оцінки вартості продукції за звітний та базисний період за одними і тими ж цінами — базисного року. Відповідно до цього величина вартості продукції поточного і базисного періодів обчислюється в однакових, порівнянних, цінах, що становлять ціни будь-якого періоду, за якими проводиться оцінка фізичного обсягу продукції за ряд періодів.

Зведений індекс фізичного обсягу продукції обчислюється за такою формулою:

$$I_p = \frac{\sum p_0 q_1}{\sum p_0 q_0}, \tag{11.91}$$

де $q_0$ і $q_1$ — вироблена продукція в натуральному вираженні відповідно в поточному та базисному періодах;

$p_0$ — порівнянні ціни, якими проводиться оцінка продукції протягом кількох років.

Агрегатна форма індексу використовується, коли кількість товарів вимірюється у натуральних одиницях. При обчисленні індексу в якості «ваг» тут застосовуються ціни одиниці товарів у базисному періоді, а індексованою величиною є кількість товару. Для забезпечення взаємозв’язку індексів індекс фізичного обсягу розраховується за формулою з постійними «вагами» ( $p_0$ ).

Коли обсяг товарів обліковується у вартісних одиницях, індекс фізичного обсягу будується за формулою гармонійного індексу. У цьому випадку в чисельнику вартість продукції звітного періоду перераховується в базисні ціни за допомогою індивідуальних (або групових) індексів цін:

$$I_p = \frac{\sum \frac{1}{i_p} p_1 q_1}{\sum p_0 q_0}, \tag{11.92}$$

де $i_p$ — індивідуальний, а при розрахунку загального індексу — груповий індекс цін;

$p_1$ — ціна звітного періоду.

Різниця чисельника і знаменника цього індексу представляє собою абсолютний приріст вартості продукції в порівнянних цінах у поточному періоді в порівнянні з базисним, обумовлений зміною обсягу продукції.

Індекс промислової продукції характеризує собою зміну створеного у промисловості обсягу валової доданої вартості за факторною вартістю (тобто за виключенням будь-яких податків на виробництво) за періоди, що обрані для порівняння. Базовою формулою для цього індексу є стандартний індекс фізичного обсягу Ласпейреса:

$$I_q = \frac{\sum_{i=1}^{N} q_{it} \times p_{ib}}{\sum_{i=1}^{N} q_{ib} \times p_{ib}}, \tag{11.93}$$

де $t$ — звітний період;

$b$ — базисний період;

$q$ — кількість продукції в натуральних одиницях;

$p$ — ціни на готову продукцію;

$i$ — кількість товарів (товарних груп);

$N$ — кількість $i$-х товарів (товарних груп).

*Індекс сільськогосподарської продукції* відображає рівень змін фізичного обсягу виробництва продукції сільського господарства, виробленого за періоди, що обрані для порівняння. Принципи побудови індексу сільськогосподарської продукції аналогічні вищенаведеним. Його розрахунок здійснюється безпосередньо Державною службою статистики України, оскільки сільськогосподарські підприємства не ведуть подвійної оцінки своєї продукції.

*Індекс фізичного обсягу валового внутрішнього продукту*, який розглядається українською статистикою як валова додана вартість, створена всередині країни за певний період, також обчислюється на основі оцінки всієї маси валового внутрішнього продукту (у порівнянних цінах) шляхом віднімання з валового випуску промисловості , сільського господарства та інших галузей у порівнянних цінах матеріальних витрат у цих галузях, виражених у порівнянних цінах.

*Індекс фізичного обсягу роздрібного товарообороту* дає кількісну характеристику змін обсягів товарообороту. Для розрахунку індексу фізичного обсягу роздрібного товарообороту порівнюються товарообороти як звітного, так і базисного періоду в єдиних цінах (постійних), тобто в цінах одного й того ж періоду, як правило, базисного періоду. З цією метою товарооборот звітного періоду підлягає переоцінці в ціни базисного періоду.

Індекс фізичного обсягу роздрібного товарообороту (тобто у постійних цінах) визначається за формулою:

$$I_p = \frac{\sum q_1 p_1 / \frac{p_1}{p_0}}{\sum q_0 p_0} = \frac{\sum q_1 p_0}{\sum p_0 q_0}. \tag{11.94}$$

де $\frac{p_1}{p_0}$ — індивідуальний індекс цін.

Розрахунки індексів фізичного обсягу роздрібного товарообороту за рік і квартал здійснюються, виходячи з структури роздрібного товарообороту та індивідуальних індексів цін. Сума перерахованих оборотів звітного періоду (року, кварталу) за товарними групами у ціни базисного періоду утворює чисельник індексу фізичного обсягу товарообороту, а фактичний обсяг роздрібного товарообороту за базисний період — знаменник цього індексу. Індекс фізичного обсягу роздрібного товарообігу обчислюються за кожною окремою товарною групою, за продовольчими та непродовольчими товарами.

Різниця між цим індексом (вираженим у відсотках) та 100 становить відсоток зростання фізичного обсягу товарообігу.

*Індекс кількісних умов зовнішньої торгівлі* розраховується для оцінки відносної динаміки фізичного обсягу експорту та імпорту. Розрахунок здійснюють за формулою:

$$T_Q = \frac{J_{Qe}}{J_{Qi}}, \tag{11.95}$$

де $T_Q$ — індекс кількісних умов торгівлі;

$J_{Qe}$ — індекс фізичного обсягу експорту за поточний період порівняно з базисним періодом;

$J_{Qi}$ — індекс фізичного обсягу імпорту поточний період порівняно з базисним періодом.

**Індекси продуктивність праці**. *Індекс продуктивності праці* характеризує зміну рівня продуктивності праці за певний період часу. Обчислюється як відношення рівнів продуктивності праці звітного та базисного періодів:

Розрахунок індексу продуктивності праці значною мірою залежить від методу вимірювання продуктивності праці. Динаміку продуктивності праці, залежно від умов, можна висловити за допомогою натурального, трудового та вартісного (грошового) методу. Крім того, динаміка продуктивності праці характеризується однією і тією ж відносною величиною як шляхом зіставлення рівнів середнього виробітку, так і шляхом зіставлення (у зворотному порядку) рівнів трудомісткості.

Індекс продуктивності праці ($I_{\text{п}t}$) у загальному вигляді може бути розрахований за такою формулою:

$$I_{\text{п}t} = \frac{\sum q_1 t_0}{\sum q_1 t_1}, \tag{11.96}$$

де $q_1$ — обсяг продукції, виробленої у звітному періоді у відповідних одиницях виміру;

$t_0$ і $t_1$ — витрати робочого дня відповідно у базисному та звітному періодах на одиницю продукції.

Як відомо, витрати на одиницю продукції (трудомісткість) перебувають у протилежному відношенні до продуктивність праці. Звідси чисельник індексу характеризує кількість праці, яку потрібно витратити на виробництво продукції звітного періоду при рівні витрат праці на одиницю продукції базисного періоду. Знаменник цього індексу — фактичні витрати праці у звітному періоді на виробництво продукції. Різниця між чисельником та знаменником ($\sum q_1 t_0 - \sum q_1 t_1$) характеризує абсолютний розмір економії у витратах праці в результаті зростання його продуктивності.

Цей метод розрахунку індексу продуктивності праці зазвичай застосовується в первинних підрозділах підприємства (цеху, відділеннях і т. д.), де є можливість повного та достовірного обліку витрат робочого часу. У цьому випадку зіставлення проводиться у зворотному порядку — витрати праці при трудомісткості базисного періоду потрапляють у чисельник дробу, а при трудомісткості звітного періоду

— у знаменник дробу.

Найбільше поширення в галузях матеріального виробництва (у промисловості, сільському господарстві, будівництві) для визначення динаміки продуктивності праці набув індекс, заснований на зіставленні вироблення продукції на одного працівника в грошову вираженні у порівнянних цінах. Цей індекс продуктивності праці в загальному вигляді визначається за такою формулою:

$$I_{п} = \frac{\sum q_1 Ц}{\sum P_1} : \frac{\sum q_0 Ц}{\sum P_0}, \qquad (11.97)$$

де $q_0$ і $q_1$ — обсяг виробленої продукції у натуральному вираженні відповідно у базисному та звітному періодах;

$Ц$ — порівнянна вартість за одиницю продукції;

$P_1$ і $P_0$ — середньооблікова чисельність працівників у базисному та звітному періодах.

Тут в якості співвимірника різних видів продукції застосовуються інші співвимірники — ціни. У силу цього на показник продуктивності праці великий вплив надає зміна часток видів продукції, що відрізняється за ціною.

**Індекс якості продукції**. Найчастіше підприємства випускають різнорідну продукцію. Для досить повної характеристики якості різнорідної продукції обчислюють ряд індексів: індекс якості продукції, індекс (коефіцієнт) сортності, індекс дефектності продукції.

*Індексом якості продукції* $I_к$ називається середньозважене значення відносних значень показників якості різних видів продукції, випущених за певний проміжок часу. Як правило, при його обчисленні вагами служать обсяги продукції (робіт) у порівнянних цінах. Розрахунок здійснюється за формулою:

$$I_{к} = \frac{\sum i q_1 p_0}{\sum q_1 p_0}, \qquad (11.98)$$

де $i$ — індивідуальний індекс якості ($i = k_1 / k_0$; де $k_1, k_0$ — рівні якості будь-якого продукту в поточному та базисному періодах);

$q_1$ — випуск кожного виду продукції у поточному періоді у натуральних одиницях виміру;

$p_0$ — порівняна вартість одиниці.

Зведений індекс якості являє собою середній арифметичний індекс, зважений за вартістю продукції звітного періоду в порівнянних цінах підприємства.

*Індекс сортності продукції* — відносний показник, що характеризує середню зміну фактичної сортності різних видів продукції в порівнянні з сортністю в періоді, прийнятому за базу порівняння.

При виробництві одного виду продукції індекс сортності обчислюють як відношення сумарної вартості продукції, випущеної за певний період часу, і сумарної вартості цієї ж продукції в перерахунку на найвищий сорт:

$$k^s = \frac{\sum_{i=1}^{m} B_i Ц_i}{\sum_{i=1}^{m} B_i Ц_1} \quad \text{або} \quad k^s = \frac{\sum_{i=1}^{m} B_i Ц_i}{Ц_1 \sum_{i=1}^{m} B_i}, \tag{11.99}$$

де $B_i$ — обсяг випуску $i$-го виду продукції натуральному вираженні на підприємстві за цей період (місяць, рік);

$Ц_i$ — ціна одиниці $i$-го виду продукції відповідного сорту, грн;

$Ц_1$ — ціна одиниці $i$-го виду продукції першого (найвищого) сорту, грн;

$m$ — кількість сортів $i$-го виду продукції, встановлених на підприємстві.

У випадках, коли підприємство випускає кілька видів продукції, то коефіцієнт сортності розраховується за наступною формулою:

$$K_c = \frac{\sum_{i=1}^{s}\sum_{k=1}^{n} C_{ik} q_{ik}}{\sum_{i=1}^{s} C_{i1} \sum_{k=1}^{n} q_{ik}}, \tag{11.100}$$

де $s$ — кількість видів продукції;

$k$ — кількість сортів продукції;

$C_{ik}$ — вартість продукції $i$-го виду $k$-го сорту в грн;

$q_{ik}$ — обсяг випуску продукції $i$-го виду $k$-го сорту;

$C_{i1}$ — вартість продукції $i$-го виду найвищого сорту.

*Індекс дефектності продукції* $I_д$ дорівнює середньому зваженому значенню коефіцієнтів дефектності продукції, виготовленої за певний проміжок часу:

$$I_д = \frac{\sum_{i=1}^{n} B_i Ц_i I_{дi}}{\sum_{i=1}^{n} B_i Ц_i}, \tag{11.101}$$

де $B_i$ — обсяг $i$-го виду продукції в натуральному вираженні, випущеної в аналізованому періоді;

$Ц_i$ — оптово-відпускна ціна одиниці *i*-го виду продукції, грн;

$I_{дi}$ — індекс дефектності *i*-го виду продукції;

$n$ — загальна кількість видів продукції, випущених за розглянутий період.

Він характеризує кількість дефектів, що припадає на одиницю вартості продукції.

Статистика використовує індексний метод при вирішенні найрізноманітніших економічних завдань. Головними з яких є вивчення динаміки собівартості, цін, фізичного обсягу продукції, оцінка якості продукції, вимірі структурних змін складних явищ.

## ЛІТЕРАТУРА

1. *Горкавий В. К.* Статистика: Підручник. Третє вид., переробл. і доповн. Київ: Алерта, 2019. С. 321–354.
2. Статистика: підручник / *С. І. Пирожков*, *В. В. Рязанцева*, *Р. М. Моторин* та ін. Київ: Київ. нац. торг.-екон. ун-т, 2020. С. 264–296.
3. Статистика: Підручник / *С. С. Герасименко*, *А. В. Головач*, *А.М. Єріна* та ін.; За наук. ред. д-ра екон. наук С. С. Герасименка. 2-ге вид., перероб. і доп. Київ: КНЕУ, 2000. С. 139–159.

## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники, словники*

4. *Педченко Г. П.* Статистика: Навчальний посібник. Мелітополь: Колор Принт, 2018. С. 239–247.
5. Статистичний словник / [*О. Г. Осауленко*, *О. О. Васєчко*, *М. В. Пугачова* та ін.]; за ред. д-ра держ. упр., проф., член-кор. НАН України О. Г. Осауленка; НТК статистичних досліджень. Київ: ДП «Інформ.-аналіт. агентство», 2012. 498 с.
6. *Ткач Є. І.*, *Сторожук В. П.* Загальна теорія статистики: підручник. 3-тє вид. Київ: Центр учбової літератури, 2009. С. 302–332.
7. *Шапочка М. К.*, *Маценко О. М.* Теорія статистики: навч. посібник. Суми: Університетська книга, 2014. С. 269–298.

*Монографії та статті*

8. *Андрієнко В. Ю.* Статистичні індекси в економічних дослідженнях: монографія. Київ: Акдемперіодика, 2004.
9. *Задорожна Р. П.* Еволюція індексного методу аналізу. *Статистика України*». 2008. № 4 (43). С. 80–85.

ГЛАВА 12

# ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ

## 12.1. Теоретичні основи вибіркового спостереження

**Поняття про вибіркове спостереження**. У статистичному дослідженні масових суспільних явищ часто проводиться спостереження не всіх одиниць сукупності, а лише їх частини. З цією метою організують проведення вибіркового спостереження. *Вибірковим спостереженням* називається обстеження відібраного в порядку, як правило, випадкового відбору певної кількості одиниць генеральної сукупності з метою отримання її узагальнюючих характеристик. Вибірковий метод заснований на частковому спостереженні одиниць сукупності.

Вибіркове спостереження представляє собою один із найбільш широко застосовуваних видів несуцільного спостереження. Відмінною особливістю вибіркового спостереження від інших форм несуцільного статистичного обстеження полягає в тому, організація спостереження підпорядкована ідеї отримати в результаті відбору таку вибіркову сукупність одиниць, яка за своїми характеристиками представляла б ту генеральну (загальну) сукупність, з якої здійснено відбір. Це можливо при утворенні вибіркової сукупності в умовах стохастичного (імовірнісного) процесу. Перевага вибіркового спостереження у порівнянні з іншими видами несуцільного спостереження полягає в тому, що за результатами цього спостереження можна оцінити шукані параметри всіх одиниць досліджуваного явища, або, іншими словами, всіх одиниць генеральної сукупності.

Можливість екстраполяції (поширення) результатів вибіркових обстежень на генеральну сукупність — це один із головних факторів, що обумовлює ефективність вибіркових методів спостереження. Тому істотна вимога, що пред’являється до вибіркового статистичного спостереження, — це можливість оцінки точності його даних, або, іншими словами, відповідності відібраної частини цілому.

Науковою основою вибіркового спостереження є теорія ймовірностей і, зокрема, закон великих чисел[1]. Використання теорії закону великих чисел дозволяє вирішувати два взаємопов'язані і головні питання вибірки: розрахувати її обсяг при заданій точності дослідження і визначити помилку при даному обсягу вибірки.

Вибірковий метод знаходить широке застосування як самостійний метод статистичного дослідження в різних галузях народного господарства. У промисловості його застосовують для оцінки якості продукції, аналізу точності та стабільності технологічних процесів, вивчення використання робочого часу та часу роботи обладнання, у сільському господарстві – для визначення втрат урожаю при збиранні, продуктивності худоби, для контрольних перевірок під час перепису худоби та інших роботах. У торгівлі вибірковим шляхом спостерігаються ціни та обсяги продажів на ринку, попит населення і ступінь його задоволення і т. д. Особливо поширені вибіркові обстеження у вивченні рівня життя домогосподарств, економічної діяльності населення країни. Крім бюджетних обстежень сімей робітників і службовців проводяться одноразові вибіркові обстеження житлових умов та величини заробітної плати. Точно також неможливо було б оглянути з метою контролю якості всю кількість продукції, наприклад, зерна, деталей, що надходять у сховища або на склади. Вибірковий метод дозволяє за результатами дослідження деякої частини генеральної сукупності оцінювати цільові показники для національного рівня, для окремих регіонів, підприємств і різних верств населення.

Широке поширення вибіркових спостережень пов'язано тим, що він у порівнянні з суцільним спостереженням має низку досить істотних переваг.

До основних їх можна віднести:

*економічність*. Вибіркове спостереження вимагає значно менше витрат часу та коштів у порівнянні з суцільним спостереженням;

*оперативність*. При вибірковому спостереженні є можливість швидше отримувати інформацію, на підставі якої приймаються певні

---

[1] Докладніше з цим питанням можна ознайомитись з навчальних посібників: *Руденко В. М.* Математична статистика: Навч. посіб. Київ, 2012*; Барковский В*., *Барковска Н.*, *Лопатін О.* Теорія ймовірностей та математична статистика. Київ, 2019; Толок В. О., Киричевський В. В., Волкова Т. Д. Курс математики для економістів: [навч. посіб. для студ. екон. спец. ун-тів]. Ч. 3. Київ: Наукова думка, 2002.

управлінські рішення;

*поглибленість та деталізація*. При вибірковому спостереженні вивчається лише частина фактів, що дозволяє ретельніше і поглиблено провести спостереження і, отже, є можливість ширше і докладно вивчити різні властивості ознак досліджуваних об'єктів;

*підвищена точність у реєстрації фактів*. Внаслідок вивчення лише певної частини сукупності є можливість отримати більш точні результати, ніж при суцільному спостереженні.

Саме з цих мотивів дуже часто статистичне спостереження суцільного характеру замінюються вибірковими. Зрозуміло, якщо замість тисяч об'єктів обстежуються десятками або сотнями, то обстеження можна провести із значною економією сил, коштів і часу. Існують й інші випадки для обмеження статистичного спостереження вибірковими даними. Зокрема, при руйнівному контролі, тобто контролі, що супроводжується псуванням контрольованої одиниці продукції, суцільний контроль взагалі виключається (перевірка палива на калорійність, деталей на розрив і т. д.). Це відбувається при дослідженні, наприклад, якості продукції. Необхідність звертатися до вибіркового спостереженню виникає тоді, коли сукупність одиниць, що мають досліджувану ознаку, дуже велика і тому не може бути практично охоплена суцільним спостереженням. Неможливо оглянути з метою контролю якості всю кількість продукції, наприклад, шестерень або болтів, що надходять тисячам у склади на зберігання.

Вибіркове спостереження може бути використане також як засіб контролю для перевірки результатів суцільного спостереження. Вибірковими даними користуються і для того, щоб попередньо ознайомитися з підсумками господарської діяльності підприємств.

Таким чином, у статистичному дослідженні вибіркове спостереження займає дійсно важливу роль. До вибіркового методу вдаються економіст і соціолог, цими даними оперують лінгвіст і технолог, з ними мають справу представники різних галузей природознавства, що інтерпретують результати своїх експериментів і спостережень. Слід зазначити, що застосування вибіркового методу повинно ґрунтуватися на глибокому теоретичному знанні досліджуваних процесів і явищ. «Дослідник, який залучає вибірковий метод до вивчення соціально-економічних явищ і процесів, — зазначали І. Г. Венецький, І. В. Венецька, — повинен розуміти, що відсутність чіткого уявлення про природу явища, що вивчається, не може відшкодувати навіть велику кількість спостережень; лише наявність ясності в розумінні природи процесу призводить до того, що і мале число спостережень може бути достатнім для виявлення закономірностей перебігу процесу, судження про характерні зв'язки і т. д.».

Вибіркове спостереження, як і суцільне, має бути організовано за певним планом, в якому встановлені завдання, що вимагають рішення, досліджуваний об'єкт та інші питання проведення спостереження.

**Історія вибіркового методу**. В історії статистики не можна назвати будь-кого, кому належала ідея вибіркового методу. Деякі статистики пов'язують виникнення вибіркового методу з ім'ям Лапласа, який розвинув теорію помилок, настільки важливу для вибіркового методу. Однак у теорії статистичної практики можна знайти приклади, які свідчать про те, що вибірковий метод став застосовуватися набагато раніше того часу, коли його завдання почали теоретично осмислюватися. Зачатки практичного застосування вибіркового методу сягають далекого часу. Зокрема, до нас дійшли свідоцтва про застосування вибіркового методу ще в середині XVII століття в Російській державі. Документи, що збереглися, свідчать, наприклад, про те, що у вотчинах боярина Б. І. Морозова записували результати пробних ужинів і умолотів хліба.

Теоретична розробка вибіркового методу у широкому плані розпочалася наприкінці XIX ст. Як зазначав в «Нарисах з теорії статистики»[3] А. А. Чупров, навіть професійним статистикам було дійти висновку про можливість використання вибіркового методу в практичній діяльності. У 1885 р. на V сесії Міжнародного статистичного інституту А. Н. Кіер (Норвегія) поставив на обговорення питання про вибірковий метод. Однак після гарячих дебатів рішення відкладається до наступної сесії. У 1901 р. на VIII сесії було прийнято рішення рекомендувати вибірковий метод уваги статистиків. У Берліні 1903 р. дебати відновлюються — не з найкращим результатом. Протягом ще більше десяти років Міжнародний статистичний інститут, що представляє вищий науковий авторитет у статистиці, не знімав питання про вибірковий метод із програми своїх засідань.

Теоретичні основи вибіркового методу закладено у роботах видатних математиків та статистів Я. Бернуллі, А. Боулі, К. Гаусса, В. Госсета, П. Лапласа, О. М. Ляпунова, А. А. Маркова, К. Пірсона, С. Пуассона, П. Л. Чебишева, А. Чупрова та ін. Математичну основу вибіркового методу запропонував П. Л. Чебишев в праці «Про середні величини» (1867). За допомогою оригінального методу нерівностей великий математик довів теорему, що обгрунтовує закон великих чисел[1]. Пізніше учень П. Л. Чебишева А.М. Ляпунов довів, що у випадку великого числа випадкових величин їх розподіл підкоряється закону нормального розподілу, цей закон тепер називається законом Ляпунова — Лапласа — Гауса і він є основою більшості розрахунків вибіркового методу[2].

Теорія вибіркового методу складна, а практика застосування дуже широка. Автори у межах цієї роботи обмежилися розкриттям сенсу основних теоретичних положень.

**Характеристики генеральної та вибіркової сукупностей**. Для виявлення закономірностей у розвитку явищ і процесів суспільного життя необхідно вивчення великої кількості вихідних даних. Наприклад, щоб вивчити якість продукції, що виробляється, недостатньо визначити якість продукції декількох виробів. Для цього необхідно мати дані про якість всієї маси продукції, всієї сукупності виробів. У даному випадку, тобто у випадку статистичного спостереження, сукупність виробів буде називатися статистичної сукупністю. Взагалі статистичною сукупністю називається сукупність одиниць явищ суспільного життя, що об'єднуються якимось (якимись) загальною ознакою (загальним зв'язком).

При організації вибіркового спостереження розрізняють поняття генеральної сукупності та вибіркової сукупності. Під *генеральною сукупністю* розуміється безліч одиниць, характеристики яких необхідно дослідити, і щодо яких буде зроблено узагальнення за результатами вибіркового обстеження.

Предмети чи явища, що утворюють сукупність, називаються *одиницями сукупності*. Якщо сукупність містить обмежену кількість одиниць, вона називається *кінцевою*. Якщо число одиниць сукупності безмежно, то її називають *нескінченною сукупністю*.

---

[1] *Ніпорко Н. І.* Історія розвитку вибіркового методу. *Інвестиції: практика та досвід*. 2010. № 9. С. 49.

[2] Там же, С. 49.

*Вибіркова сукупність* — сукупність одиниць, відібраних за певними правилами з генеральної сукупності для статистичного спостереження. Число одиниць, у вибірці називають *обсягом вибірки*. Чисельність генеральної сукупності позначається $N$, а вибірковий через $n$. Тоді задачу вибіркового спостереження можна сформулювати так: на основі вивчення вибіркової сукупності отримати уявлення про показники генеральної сукупності.

Узагальнюючими характеристиками генеральної та вибіркової сукупностей є відносні величини, середні величини та показники варіації та ін.

Відносні величини застосовуються для зведеної характеристики досліджуваних ознак, вони виражаються у вигляді частки (питомої ваги) тих одиниць сукупності, які володіють певними ознаками у всій сукупності. Частка одиниць, що володіють тією чи іншою ознакою в генеральній сукупності, називається *генеральною часткою* та позначається $p$. Для вибіркової сукупності вона називається *вибірковою часткою* (або частістю) і позначається через $w$. Отже, завданням вибіркового спостереження в даному випадку є на основі вибіркової частки отримати правильне уявлення про частку в генеральній сукупності.

Припустимо, що в сукупності деталей, що складаються з 1 000 одиниць (генеральна сукупність), 120 деталей є нестандартними. Отже, частка нестандартних деталей у цій сукупності (генеральна сукупність) $p = \frac{120}{1000} = 0{,}12$.

Припустимо, що з усієї партії деталей (генеральної сукупності) відібрано у випадковому порядку 100 одиниць, а серед відібраних виявилося 8 нестандартних. Отже, вибіркова частка 0,08. Вибіркова частка називається також *частістю*.

Крім вимірювання частки, в процесі вибіркового дослідження необхідно обчислити середнє значення варійованої ознаки у всій сукупності. Узагальнюючими показниками для обчислення типової величини ознаки в одиниць сукупності є середні величини. Такі завдання виникають при вибірковому вимірі середнього рівня браку продукції, середньої зарплати робочих, втрат урожаю сільськогосподарських культур тощо.

Середнє значення варійованої ознаки в генеральній сукупності називається *генеральною середньою*, що позначається $\bar{x}$, а в вибірковій сукупності — *вибірковою середньою*, що позначається $\tilde{x}$. Отже, завданням вибіркового спостереження в даному випадку є на основі вибіркової середньої дати правильне уявлення про генеральну середню.

У теорії вибіркового спостереження велике значення мають показники варіації, тобто показники зміни (коливання) величини ознаки від одиниці до одиниці або в окремої одиниці в різні періоди часу. Як міра варіації (коливання) вивчається ознаки використовують такі показники, як дисперсію, середнє квадратичне відхилення, коефіцієнт варіації.

Для характеристики варіації досліджуваних ознак одиниць сукупності обчислюють середній квадрат відхилень — дисперсію. Дисперсію ознаки в генеральній сукупності позначається $\sigma^2$, вибіркову дисперсію — через $\sigma_0^2$.

При організації вибіркового спостереження статистик вправі очікувати, що вибірка спостережень із сукупності буде найкращим чином представляти генеральну сукупність, бути зменшеною копією її. У разі статистики кажуть, щоб вибірка буде представницькою, або репрезентативною. *Репрезентативність* – властивість вибіркової сукупності повно та адекватно відображати основні характеристики генеральної сукупності. При оцінці представницької вибірки необхідно враховувати і те, наскільки розподіл у виборці суттєвих для досліджуваного ознак характерно для аналізованої сукупності в цілому. Так, якщо серед 1 000 робітників заводу мають 100 осіб, що не виконують норми (середнє виконання норми становить 94%), 500 осіб, що виконують норми на 100 %, 250 осіб — на 110 % і 150 осіб на 124 %), то для забезпеченні представництва вибірки чисельністю 200 осіб необхідно, щоб у вибіркову сукупність увійшли представники всіх названих груп чисельністю, що становлять одну п’яту їх чисельності в генеральній сукупності.

**Помилки вибірки**. Головне завдання вибіркового спостереження полягає в тому, щоб на основі обстеження частини одиниць сукупності отримати достовірні дані для характеристики сукупності в цілому. Але що означає отримати досить достовірні дані щодо характеристик генеральної сукупності?

Оскільки йдеться про варійовані ознаки і обстеженню підлягають не всі одиниці сукупності, а лише їх частина, то можна заздалегідь сказати, що зведені показники за цими ознаками у вибіркової сукупності майже ніколи не будуть абсолютно збігатися зі зведеними показниками всіх одиниць сукупності. Отже, йдеться не про те, щоб досягти абсолютної відповідності загальних показників вибірки і генеральної сукупності. Таке завдання практично нерозв'язне.

Йдеться про те, щоб максимально наблизити показники вибіркової сукупності до показників генеральної, знати можливі межі відхилень цих показників і умови, від яких залежить величина цих відхилень.

Питання представницькості вибірки тісно пов'язані з питанням про помилки вибірки.

*Помилка вибірки*, або *помилка репрезентативності* (*представницькості*) представляє собою розбіжність між значеннями досліджуваної ознаки (показника) вибіркової сукупності і генеральної сукупності. Помилки вибірки, як показує сама їх назва, властиві лише вибірковому спостереженню. Чим більша величина цієї помилки, тим, отже, більшою мірою зведені показники вибіркової сукупності відрізняються від зведених показників сукупності в цілому. Із зменшенням величини помилки вибірки вибіркове спостереження дає більш правильне уявлення про зведені показники всієї сукупності.

Розрізняють помилки середньої та помилки частки. *Помилка середньої* ($\varepsilon$) обчислюється шляхом віднімання з вибіркової середньої ($\tilde{x}$) генеральної середньої ($\bar{x}$), тобто $\varepsilon = \tilde{x} - \bar{x}$. *Помилка частки* ($\varepsilon$) обчислюється як різниця між часткою ознаки у вибірковій сукупності ($w$) і часткою ознаки в генеральній сукупності ($p$), тобто $\varepsilon = w - p$.

Якщо $\Delta$ представляє собою межу, яку не перевищує абсолютна величина $\varepsilon$, то

$$\tilde{x} - \Delta \leq \bar{x} \leq \tilde{x} + \Delta, \tag{12.1}$$

$$w - \Delta \leq p \leq w + \Delta. \tag{12.2}$$

Помилки вибіркового спостереження відбуваються внаслідок помилок реєстрації та помилок вибірки.

*Помилки реєстрації*, як відомо, виникають через неправильні та неточні відомості в процесі спостереження. Вони властиві як суцільному, і вибірковому спостереженню. Однак при проведенні вибіркового спостереження ці помилки менші через те, що при проведенні вибіркового спостереження обстеженню підлягає менша кількість одиниць сукупності і є можливість залучити більш кваліфікований персонал. До помилок реєстрації належать також помилки, які є результатом описок, низької кваліфікації особи, яка проводить вимірювання, а також несподіваних зовнішніх впливів на вимірювання. Джерелами цих помилок можуть бути неправильне розуміння, неуважність реєстратора, пропуск або повторний рахунок, описки при заповненні бланків і т. п. У свою чергу, помилки реєстрації поділяються на систематичні діючі в якомусь одному напрямку (у бік перебільшення або применшення), і на випадкові помилки, що часто проявляються в різних напрямках, часто врівноважують один одного і не надають істотного впливу на показник, що вивчається.

До систематичних помилок реєстрації відносяться такі помилки, які за даних умов проведення мають цілком певне значення. До них відносяться помилка вимірювального приладу, помилки запам’ятовування (наприклад, пропуски записів у витратах при обстеженні сімей), помилки суб’єктивних вражень (наприклад, при вибірковому обстеженні посівних площ). Для уникнення систематичних помилок необхідно ретельно підбирати кадри, постійно підвищувати їх рівень кваліфікації, забезпечити суворий контроль за станом обліку та звітності, розробляти докладні та чіткі інструкції, ставити контрольні питання, що дозволяють виявити неточності, і т. д.

*Помилки вибірки* зумовлені відмінністю структури вибіркової сукупності від генеральної структури. Розрізняють помилку вибірки (представницькості) двох видів — систематичну (тенденційну) і випадкову. *Тенденційні помилки* (*помилки репрезентативності*) виникають внаслідок порушення правил відбору одиниць генеральної сукупності для вибіркового спостереження. До систематичних помилок репрезентативності відноситься проведення попереднього або подальшого відбору з превалюванням або виключенням деяких одиниць вибірки, тобто вибірки формується переважно з «кращих» (або «гірших») одиниць генеральної сукупності. Наприклад, для вибіркового вивчення ступеня виконання плану робітниками навмисне відібрали робітників, що перевиконують планове завдання. Зрозуміло, що результати вибіркового обстеження значно перебільшуватимуть дійсні показники виконання плану робітниками. Систематична помилка вибірки може спричинити повну непридатність результатів спостереження. В даному випадку результати втрачають сенс і не можуть бути використані для всієї сукупності робітників. Найчастіше причинами цього є сумлінна помилка, але може бути, звичайно, і різновидом наукового шахрайства, маніпуляцією даними.

Тому основне завдання організації проведення вибіркового спостереження полягає в тому, щоб не допустити порушення правил відбору одиниць з генеральної сукупності, тобто забезпечити суворе дотримання принципу випадкового відбору.

*Випадкові помилки репрезентативності* вибірки виникають навіть тоді, коли незважаючи на правильність відбору спостереження і точність реєстрації, все ж є розбіжності між характеристиками вибіркової і генеральної сукупностями. Це відбувається через те, що це статистичне спостереження є спостереженням несуцільним, вибіркова сукупність недостатньо точно відтворює (репрезентує) генеральну сукупність. Таким чином, випадкові помилки репрезентативності органічно властиві вибірковому спостереженню, так як ціле характеризується за його частинами. Обов'язкова наявність у кожній вибірці власне-репрезентативної помилки обґрунтована теоретично[1].

При організації вибіркового спостереження передбачається, що з відповідних умов помилки вибірки, яка складається з помилок реєстрації, може бути зведена нулю. Оскільки дотримання певних умов призводить до усунення помилок реєстрації, то помилку вибірки завжди ототожнюють з власне-репрезентативною помилкою.

Величина помилки в основному залежить:

по-перше, від обсягу вибірки, оскільки в міру збільшення числа обстежуваних одиниць результати вибіркового обстеження частини одиниць сукупності буде все менше і менше розходиться від показників генеральної сукупності. Між обсягом вибірки і величиною помилки існує зворотна залежність: зі збільшенням обсягу вибіркового спостереження зменшується помилка вибірки і навпаки;

по-друге, величина помилки вибірки залежить від зміни (коливання) величини досліджуваної ознаки. Чим більше варіює ознака, тим більше вибіркова середня (частка) відрізняється від генеральної середньої (частки). Оскільки основними показниками для характеристики варіації досліджуваного ознаки служать дисперсія і середнє квадратичне відхилення, то можна сказати, що помилка вибірки знаходиться в прямій залежності від величини цих показників: зі збільшенням дисперсії та середнього квадратичного відхилення збільшується і помилка вибірки;

по-третє, величина помилки вибірки залежить від прийнятого способу формування вибіркової сукупності (формування вибіркової сукупності пов'язане з вирішенням питання про одиницю відбору, про спосіб і вид відбору одиниць вибіркової сукупності).

В задачу вибіркового спостереження входить не тільки визначення характеристик вибіркової сукупності, а й вивчення і вимірювання

---

[1] Докладніше з цим питанням можна ознайомитись з навчальних посібників, зазначених у зносці 1 на початку глави 12.

випадкових помилок репрезентативності. Систематичні та випадкові помилки реєстрації не можна змішувати з систематичними та випадковими помилками репрезентативності.

**Принцип випадкового відбору**. Відбір одиниць сукупності, що підлягає вибірковому обстеженню, може здійснюватися різними способами. Але який би спосіб відбору не був застосований, необхідно суворо дотримуватися загального правила, яке полягає в тому, що *в окремих одиниць генеральної сукупності повинна бути абсолютно однакова можливість потрапити в число одиниць, що підлягають обстеженню*. Інакше кажучи, відбір конкретних одиниць може бути зроблено не за розсудом особи, що проводить спостереження, а у випадковому порядку. При такому підході до відбору одиниць, коли жодна з них не має переваги потрапити в відбирається сукупність в порівнянні з іншими, характеристики вибіркової сукупності при збільшенні обсягу вибірки прагнуть до характеристик генеральної сукупності. У цьому полягає найважливіший принцип вибіркового спостереження — *принцип випадкового відбору*.

Проте випадковий відбір — це безладний відбір. «Якщо безладно встромляти шпильки в карту, — писав англійський статистик Френк Йейтс, — це не дасть випадкового розподілу точок на карті. Якщо відібрати будинки для обстеження, просто йдучи вулицями міста, то це не буде випадковим відбором будинків» (див.: *Френк Й.*, 1965, с. 34). При такому способі неможливо усунути відому навмисність, хоча б і не усвідомлену.

Виникає питання, чому при організації вибіркового спостереження необхідно суворо дотримуватися принципу випадкового відбору? Тому, що дотримання цього принципу гарантує незалежність результатів вибіркового спостереження від навмисних дій обстежувача. При методі спрямованого відбору, що часто практикувалися до розробки теорії науково-організованих вибіркових досліджень, відбір одиниць сукупності проводиться на розсуд особи, яка проводить спостереження. Такі суб'єктивно-організовані, навмисні відбори мають ту відмінну особливість, що суб'єктивний фактор при відборі тієї чи іншої одиниці сукупності має більше значення, ніж випадковий фактор. Зрозуміло, похибку спостереження встановити неможливо. Навмисність дій обстежувача завжди тенденційна, бажає він цього чи ні, і призводить до систематичних (тенденційних) помилок. Ці помилки неможливо кількісно виміряти.

Таким чином, найбільш важливим принципом у використанні вибіркового методу є забезпечення рівної можливості кожної одиниці генеральної сукупності бути відібраною (принцип рівноможливості). У разі дотримання принципу рівноможливості вибір цілком певної одиниці має один шанс (випадок) з $N$ таких шансів. Порушення цього принципу, коли спостереженню піддаються одиниці, відібрані на підставі суб'єктивної думки дослідника, призводить до того, що результати не можуть бути застосовані до всієї генеральної сукупності.

Дотримання принципу випадкового відбору гарантує недопущення систематичних помилок. Помилки вибірки за дотримання принципу випадкового відбору носять випадковий характер, тобто мають рівну можливість перебільшувати чи применшувати характеристики генеральної сукупності. У математичній теорії закону великих чисел доводиться, що зі збільшенням чисельності вибірки розміри випадкових помилок скорочуються. Це означає, що з досить велику кількість спостережень результат вибірки не буде залежати від випадку. Так закон великих чисел доводить застосування вибіркового спостереження в статистиці.

Якщо принцип випадкового відбору недотримується, то застосувати теореми закону великих чисел не можна, оскільки ці теореми виведені із закономірностей, що у випадковому процесі. У випадку, якщо вибіркове спостереження проведено з усіх правил його наукової організації, то дані вибірки максимально близькі до характеристики всієї сукупності.

Вибіркова сукупність (вибірка) повинна бути представницькою, тобто повно і адекватно представляти властивості генеральної сукупності. Ступінь представницькості вибіркової сукупності визначається способом організації вибірки та її обсягом.

## 12.2. Методи та способи відбору одиниць у вибіркову сукупність

Процес вилучення чи складання вибірки називають *відбором вибірки*. Відбір одиниць з генеральної сукупності можна проводити порізному, залежно від низки умов. Систему організації відбору одиниць з генеральної сукупності називають *способом відбору*.

Питання відбору одиниць у вибіркову сукупність є найскладнішими питаннями організації вибіркового спостереження. Для того, щоб вибірка була представницькою (репрезентативною), відбір одиниць з генеральної сукупності повинен бути відповідним чином організований.

**Відбір власно-випадковий та відбір механічний**. Історично і логічно першим склався так званий власне-випадковий відбір. *Власне-випадковим* називається такий відбір, коли при його здійсненні суворо, без обмежень дотримується принцип випадковості (принцип «вільної гри випадку»). Випадковий відбір заснований на принципі випадковості, тобто всі одиниці потрапляють у вибіркову сукупність випадково. За такого відбору має бути забезпечене дотримання основного принципу вибірки — забезпечення рівної можливості опинитися серед відібраних усім одиницям, що входять до складу генеральної сукупності. Вибірка одиниць з генеральної сукупності без будь-якого розчленування на частини або групи.

При власне-випадковому відборі здійснюється відбір одиниць спостереження, тобто одиниця відбору дорівнює одиниці спостереження.

Власне-випадковий відбір дає лотерея, жеребкування або будь-який інший спосіб (таблиця випадкових чисел тощо). Методика відбору за жеребом не вимагає, мабуть, особливих роз’яснень. Відбір за жеребом полягає у відборі карток з ретельно перетасованої пачки, на які наносяться номери одиниць генеральної сукупності. Наприклад, прийнято рішення відібрати 100 із 2 000 робітників за допомогою жеребкування. Для цього на квитки (фішки), кількість яких відповідає чисельності робітників (2 000), наносяться прізвища робітників або наданий їм номер. Потім всі фішки перемішуються і навмання вибирається 100 фішок, що відповідає чисельності вибіркової сукупності.

Відбір із генеральної сукупності може здійснюватися і з допомогою таблиці випадкових чисел. У статистиці використовують кілька таблиць випадкових. Одна з них складена англійським статистом Типпетом. Вона містить 41 600 однозначних чисел, з яких отримано було 10 400 чотиризначних чисел. Нижче наведено перші 40 чисел цієї таблиці:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2952 | 6641 | 3992 | 9792 | 7979 | 5911 | 3170 | 5624 |
| 4167 | 9524 | 1545 | 1396 | 7203 | 5356 | 1300 | 2693 |
| 2370 | 7483 | 3408 | 2762 | 3563 | 1089 | 6913 | 7691 |
| 0560 | 5246 | 1112 | 6107 | 6008 | 8126 | 4233 | 8776 |
| 2754 | 9143 | 1405 | 9025 | 7002 | 6111 | 8816 | 6446 |

Якщо з сукупності, що складається з 4 000 одиниць, вирішено відібрати 15 номерів, то, перенумерувавши ці одиниці, дослідник може відібрати за таблицею 15 чисел, що не перевищують 4 000.

Для того, щоб організувати власне-випадковий відбір, необхідно мати повний перелік генеральної сукупності, який на практиці скласти дуже важко, а в окремих ситуаціях і неможливо. Однак, якщо сукупність нескінченна або настільки велика, то неможливо перерахувати всі одиниці та привласнити їм порядкові номери. Так, при дослідженні якості продукції, яка надходить безперервно, немає повної та вичерпної інформації про перелік усіх одиниць. Незручність власно-випадкового відбору полягає і в тому, що для жеребкування необхідно виготовляти фішки (білетики).

У статистичній практиці дотриматися принцип власно-випадкового відбору, особливо в економічних і соціологічних дослідженнях досягти буває складно. Цьому перешкоджає характер досліджуваних економістом і соціологом явищ, які відрізняються якісною неоднорідністю і складними та багатосторонніми зв'язками і взаємодіями. Тому випадковий відбір у чистому вигляді використовують дуже рідко.

У практиці вибіркового спостереження найчастіше застосовують *механічний відбір* — послідовний відбір одиниць через рівні проміжки з певного розташування їх у генеральній сукупності або в якомусь переліку. У цьому вигляді відбір одиниць відбувається не навмання. Суть його полягає в тому, що відбирається по одній одиниці від кожної з рівних частин, на які механічно поділяється вся статистична сукупність. При механічному відборі генеральну сукупність як би розбивають на стільки груп, скільки потрібно вибрати одиниць для вибіркового спостереження. У цьому механічний відбір завжди буває безповторним. Розглянемо приклад механічного відбору 100 деталей із сукупності, що складається з 10 000 одиниць.

Для проведення механічного відбору необхідно визначити чисельність груп, у тому числі буде обрано порядку власне-випадкового відбору одна одиниця. Кількість груп визначають як окреме від поділу чисельності генеральної сукупності на чисельність вибіркової сукупності. У прикладі генеральна сукупність має бути розчленована на 100 рівних частин (10 000:100). Отже, від кожної такої частини (групи) дослідник має механічно відібрати одиницю-представника. Вибір початку відліку зазвичай проводиться в порядку власне-випадкового відбору з одиниць першого інтервалу. Якщо почати відбір, наприклад, з 20-ї деталі за списком, то у вибірку будуть відібрані 20-а, 120-а, 220-а одиниці і т. д. до 9920. Якщо почнемо з середини інтервалу, то будуть відібраними 50-а, 150-а, 250-а одиниці тощо до 9950. Можуть бути, звичайно, взяті й інші порядкові номери. Це залежить від випадкової вибірки. Одиниці генеральної сукупності при механічному відборі розташовуються в такому порядку, що це не впливає на поведінку цікавої для нас ознаки.

На відміну від випадкового відбору, при якому виникає лише можливість потрапляння у вибірку представників усіх тих станів, якими характеризується досліджуваний ознака в загальній сукупності, механічний відбір вже прямо спрямований на те, щоб дійсно забезпечити попадання у вибірку таких представників. Тому його і називають одним із видів *спрямованого відбору*. Механічний відбір, як і районований, спрямований на підвищення репрезентативності вибіркових даних. Однак створення необхідних для цього умов носить механічний характер.

Механічний відбір набув широкого поширення в економічній статистиці. Зокрема, його використовують при оцінці якості продукції, особливо на підприємствах масового та серійного виробництва. При вибірковому обстеженні якості продукції беруть через певний інтервал (при 5 % відборі — кожен 20-й виріб, при 1 % відборі — кожен 100 виріб) виготовлені вироби та оцінюють їх якість. З цих оцінок роблять висновок про рівень якості всієї виготовленої продукції.

Механічний відбір може здійснюватися за списками на місці спостереження або безпосередньо на місцевості відповідно до природного розташування одиниць генеральної сукупності. Наприклад, при відборі 100 студентів із 2 000 учнів в університеті можна взяти кожного двадцятого студента, який виходить із вузу.

Випадкова помилка вибірки при механічному відборі, зазвичай, не перевищує помилку при власне-випадковому відборі.

**Відбір індивідуальний та відбір груповий (серійний)**. Найбільш поширеним практично є індивідуальний добір. *Індивідуальний відбір* — це відбір безпосередньо одиниць спостереження, тобто одиниця відбору дорівнює одиниці спостереження. При індивідуальному відборі за кожний акт відбору відбирається лише одна одиниця генеральної сукупності. Отже, відбір повторюється стільки разів, скільки необхідно вибрати одиниць з генеральної сукупності.

Однак в окремих випадках при проведенні вибіркового спостереження застосовують серійний, або гніздовий, відбір. При *серійній вибірці* відбору піддаються не окремі одиниці, а цілі серії (групи) одиниць. Серії, що потрапили у вибірку, піддаються суцільному спостереженню. У сільськогосподарській статистиці, де цей вид відбору набув поширення, серії господарств, що відбираються, називаються «гніздами», а вид відбору отримав назву «гніздового». Найчастіше відбір одиниць цілими партіями проводиться для перевірки якості виробів. При цьому якщо обстеженню піддаються всі одиниці відібраних серій, відбір називається серійним, а якщо обстежуються лише частина одиниць кожної серії, що відбираються в індивідуальному порядку, відбір називається *комбінованим*.

У порівнянні з іншими методами серійний відбір має ряд переваг. Серійний відбір значно простіше організувати і він дешевший за інші способи. Однак за всіх своїх переваг серійний відбір виявляється, як правило, менш задовільним у сенсі репрезентативності одержуваних даних, ніж інші способи відбору. Випадкова помилка вибірки в переважній більшості випадків більше, ніж за будь-якого іншого способу відбору.

З метою зменшення випадкової помилки вибірки вдаються до збільшення її обсягу. Крім того, при серійному відборі значно зменшується помилки реєстрації, що, у свою чергу, веде до зниження випадкової помилки вибірки.

**Відбір повторний та безповторний**. Залежно від того, чи беруть участь відібрані одиниці у подальшій вибірці чи ні, розрізняють методи відбору: повторний та безповторний.

*Повторним* називається такий метод відбору, при якому відібрана один раз одиниця повертається назад в генеральну сукупність і, отже, знову може бути відібрана при проведенні цього ж спостереження. При повторному відборі кожна одиниця має однакові шанси потрапити у вибірку. Теоретично ймовірностей повторний метод характеризується як схема «*шар, що повертається в урну*». Це означає, що один і той же шар може бути вийнятий, оскільки після попереднього вилучення від був повернутий в урну. Справді, якщо витягнутий з урни шар у неї повертається, то попереднє вилучення шару не позначається на ймовірності подальшого вилучення. При повторному відборі досліджувана сукупність до кінця спостереження не змінює свого кількісного складу.

*Безповторним* називається такий метод відбору, при якому кожна відібрана та обстежена одиниця в сукупність не повертається і в подальших випробуваннях не бере участі. Це означає, що один раз відібрана для спостереження одиниця — чи це людина, тварина, виріб або який-небудь інший обраний об'єкт — виключається з сукупності і тому не може бути відібрана повторно. Досліджувана сукупність після відбору кожної наступної одиниці змінює свій кількісний склад. При відборі кожної нової одиниці можливість потрапити у вибірку змінюється (збільшується). На мові теорії ймовірностей безповторний відбір отримав назву «*шар, що не повертається в урну*». Безповторний відбір застосовується, наприклад, при статистичному контролі якості продукції, а також при демографічних дослідженнях. Відбір серійний, як правило, завжди безповоротний. Випадковий відбір із поверненням часто буває важкий в організаційному плані.

Наприклад, стоїть завдання вивчити внутрішньозмінні втрати робочого часу, для чого вирішено було на машинобудівному заводі провести вибіркове спостереження. Вибіркове спостереження можна провести двояко: по-перше, верстат (працюючий або простоює) не виключається із загального їх числа, а може бути повторно відібраний; по-друге, кожен верстат більше не перевіряється з погляду його завантаження і цим він позбавлений можливості знову бути обстеженим у наступних відборах. Перший із цих варіантів — повторний відбір, другий — безповторний.

При повторному відборі ймовірність бути знову відібраною у окремої одиниці сукупності не змінюється протягом усього відбору, при безповторному ця ймовірність змінюється після вибору кожної одиниці.

Усі розглянуті види відбору (крім механічного) можуть бути повторними та безповторними. Механічний відбір завжди безповторний.

У практиці статистичних досліджень у переважній більшості випадків застосовується безповторний відбір. Повторний відбір застосовується лише в обмежених випадках, зокрема, при аналізі робочого часу та часу роботи обладнання, при обстеженні потоків пасажирів різних видів транспорту, пішоходів, покупців і т. д. Порядок здійснення повторного та безповторного відбору забезпечується за допомогою жеребкування або таблиць випадкових чисел.

**Відбір нерайонований та районований**. Перерахованими вище методами та способами відбір формування вибіркових сукупностей може проводитися з районованої (типової) або нерайонованої генеральної сукупності.

При *районованому*, або *типовому відборі* вибіркова сукупність формується шляхом відбору одиниць з груп або районів (якщо сукупність просторово розподілена), утворених за яким-небудь типовою для визначення величини досліджуваного предмета або явища ознакою (звідси і назва цього способу відбору). Потім обчислюються міжгрупові результати, які характеризують вже всю сукупність. В якості «районів» можуть бути взяті сформовані групи — області, підприємства, цехи і т. п. Кількість одиниць, що відбираються з кожної групи, як правило, береться пропорційно чисельності груп у генеральній сукупності. При цьому відбір може бути як випадковим, так і механічним. В останньому випадку обмежується принцип випадкового відбору. Прикладом районованого відбору є відбір підприємств за видами економічної діяльності, цін — за окремими регіонами країни. Районований відбір завжди є безповторним. Таким чином, при районованому відборі підвищення репрезентативності здійснюється шляхом свідомого групування даних.

Районований відбір значно підвищує точність результатів спостереження і тому є важливою умовою організації вибіркового спостереження. Розподіл обсягу спостереження між типовими групами може бути пропорційним (з урахуванням якихось умов) і непропорційним (рівномірно по всіх групах).

При *нерайонованому відборі* формування вибіркових даних здійснюється з усієї генеральної сукупності, не розділеної на частини. У цьому випадку відбір може бути повторним та безповторним. При такому відборі всі одиниці мають рівну можливість бути відібраними.

Організація типового чи районованого, відбору вимагає, зрозуміло, чіткого уявлення про характер і розподіл у загальній сукупності тих ознак, які мають бути покладені в основу утворення типових груп або виділення районів. Неправильний вибір ознаки при розподілі одиниць сукупності може призвести не до підвищення репрезентативності вибіркових даних, а, навпаки, до її зниження.

**Комбінування вибіркового спостереження із суцільним**. *Комбінованим* називається такий вид відбору, коли на різних ступенях відбору застосовуються різні його варіанти. Насамперед у практиці вибіркового спостереження можна поєднувати *суцільне* і *вибіркове* спостереження. Так, наприклад, за основною програмою спостереження обстежується вся генеральна сукупність, а за додатковою — вибіркова сукупність. Найчастіше для одиниць вибіркової сукупності розробляються особливі, більш розширені формуляри.

Цінність такого роду комбінації видів спостереження полягає в тому, що дані вибіркового спостереження тут тісно пов’язані з даними суцільного спостереження, що робить їх більш доказовими і значущими.

При вибірковому обстеженні особистих підсобних господарств спочатку відбувається відбір району із груп, тобто відбір типовий. Потім відбирається по одній одиниці від кожної з рівних частин, на які механічно поділяється вся сукупність особистих підсобних господарств, тобто в порядку механічного відбору.

На практиці різні способи відбору часто доповнюють один одного в тому самому дослідженні: спочатку відбір більших об'єктів проводиться одним способом, а потім вже іншим способом вибираються одиниці з тих, які входять до складу великих об'єктів.

Такі види відбору, у яких здійснюється відбір одиниць різних порядків (великих, а потім дрібніших), звуться багатоступінчастим відбором. *Багатоступінчастий* відбір характеризується тим, що формування вибірки здійснюється поступово, в кілька щабелів, і одиниця спостереження визначається тільки на останній сходинці. Багатоступінчатий відбір відбувається у кількох послідовних сходинок (етапів): на першій ступені з основи вибірки відбираються певні відносно великі одиниці; на другій ступені всередині кожної відібраної на першому ступені одиниці формується своя основа вибірки, з якої відбираються свої одиниці, і так далі за кількістю ступенів відбору. Дуже важливо, щоб наступні одна за одною вибірки проводилися із сукупностей, утворених на попередніх етапах, а не з генеральної сукупності. Відмінною особливістю багатофазного відбору є наявність тих самих одиниць відбору на всіх його фазах. Багатоступінчастий відбір може бути гніздовим чи змішаним.

*Змішаним* відбір називається в тому випадку, коли на різних ступенях відбору застосовуються різні його варіанти: наприклад, серійний та індивідуальний.

Усі форми відбору призводять до помилок, які не можна назвати випадковими. Якщо зі списку, елементи якого розташовані в порядку зростання або зменшення їх величини, проводиться механічний відбір, то виникає помилка, що носить вже системний характер. Якщо у списках існує якась правильна повторюваність і якщо інтервал відбору з нею пов'язаний, то ця обставина також призводить до систематичної помилки. Тому основне завдання статистика, що організує спрямований відбір, полягає в тому, щоб виключити з обчислюваної ним помилки ті елементи, які можуть викликати систематичне її зміщення і правильно визначити величину власне-випадкової помилки.

## 12.3. Середня і гранична помилки вибірки

**Формула середньої помилки вибірки**. Відбір одиниць з генеральної сукупності проводиться таким чином, щоб вибіркова сукупність булла представницькою (репрезентативною) і правильно характеризувала генеральну сукупність. Оскільки в умовах вибіркового спостереження немає можливості обчислити величину конкретної помилки, можна встановити, наскільки за даних умов в середньому узагальнюючі показники вибіркової сукупності відрізняються від узагальнюючих показників генеральної сукупності. Іншими словами, можна визначити величину середньої помилки.

Для вимірювання помилки вибірки встановлено спеціальний захід варіації, який називається середньою помилкою вибірки. *Середня* (або *стандартна*) *помилка вибірки* представляє собою середньоквадратичне відхилення можливих значень вибіркової середньої від генеральної середньої, зважених за ймовірностями їх виникнення. Вона позначається грецькою літерою $\mu$ (мю). Середня помилка вибірки для середньої показує середню величину всіх можливих розбіжностей вибіркової та генеральної середньої.

Уявлення про середню помилку відіграє значну роль у теорії вибіркового статистичного методу. У частках середньої помилки виражається передбачувана величина помилки даної конкретної вибірки, допускається гранична її помилка, тобто помилка, що не виходить за відомі передбачувані межі.

У вибірковому дослідженні користуються теоремами математиків П. Л. Чебишева та А. М. Ляпунова, сформульованими для випадкового відбору.

Величина середньої помилки вибірки різна для окремих різновидів випадкового відбору.

У математичній статистиці доводиться, що величина середньої квадратичної помилки простої випадкової вибірки може бути визначена за формулою:

$$\mu = \sqrt{\frac{\sigma^2}{n}}, \tag{12.3}$$

або

$$\mu = \frac{\sigma}{\sqrt{n}}, \tag{12.4}$$

де $\sigma$ — середнє квадратичне відхилення у генеральній сукупності;

$n$ — обсяг вибірки.

У формулі середньої квадратичної помилки простої випадкової повторної вибірки входить дисперсія ознаки у генеральній сукупності, величина якої, зазвичай, під час проведення вибіркового спостереження буває невідомим. Нам доводиться використовувати вибіркову дисперсію як оцінку генеральної дисперсії. Така заміна правомірна, оскільки між дисперсіями генеральної та вибіркової є певна залежність:

$$\sigma^2 = \sigma_0^2 \frac{n}{n-1}. \tag{12.5}$$

де $\sigma_0^2$ — вибіркова дисперсія.

Отже, генеральну дисперсію можна визначити за такою формулою:

$$\sigma^2 = \frac{\sum (x_i - \widetilde{x})^2}{n} \cdot \frac{n}{n-1} = \frac{\sum (x_i - \widetilde{x})^2}{n-1}. \tag{12.6}$$

При досить великих значеннях обсягах вибірки ($n$>30) різниця між дисперсією, обчисленої за формулою $\sigma^2 = \frac{\sum (x_i - \widetilde{x})^2}{n}$ і за формулою $\sigma^2 = \frac{\sum (x_i - \widetilde{x})^2}{n-1}$, дуже невелика (для $n$ = 50 отримаємо $\frac{50}{50-1}$=1,0204, тобто вибіркова дисперсія буде на 2,04 % менше генеральної). Тому можна допустити, що вибіркова дисперсія дорівнює генеральній дисперсії.

З урахуванням усіх цих зауважень *середня помилка* при власне випадковому повторному відборі визначається такими формулами:

$$\mu = \sqrt{\frac{\sigma_0^{\,2}}{n}}, \tag{12.7}$$

або

$$\mu = \frac{\sigma_0}{\sqrt{n}}, \tag{12.8}$$

де $\mu$ — середня помилка вибірки;

$\sigma_0^2$ і $\sigma_0$ — дисперсія і середнє квадратичне відхилення ознаки, що варіюється;

$n$ — чисельність одиниць вибіркової сукупності.

Таким чином, розмір середньої помилки вибірки дорівнює кореню квадратному з дисперсії ознаки, поділеної на чисельність вибіркової сукупності. За допомогою цієї формули встановлюється, що надійність *вибіркової середньої пропорційно до кореня квадратного з числа спостережень*.

Ця формула застосовується в тому випадку, коли вибіркове обстеження ставить своїм завданням виміряти *середнє значення* якої-небудь ознаки. *Середня помилка вибірки* для середньої показує середню величину всіх можливих розбіжностей вибіркової та генеральної середньої.

Формула (12.7) може бути використана і для встановлення помилки вибірки у разі, коли вибірково вимірюється частка певної ознаки (наприклад, відсоток браку в продукції). Але у зв'язку з тим, що дисперсія частки в генеральній сукупності дорівнює $pq$, а у вибірковій сукупності — $w(1-w)$, то формула (12.7) набуває вигляду:

$$\mu = \sqrt{\frac{w(1-w)}{n}}, \tag{12.9}$$

де $w$ — частка ознаки (наприклад, якщо відсоток браку в продукції становить від усієї продукції 0,5 %, то при w =0,005);

$n$ — чисельність одиниць вибіркової сукупності.

У цій формулі відмінний лише чисельник дробу. Добуток $w(1-w)$, як і $\sigma^2$, служить мірою варіації ознаки в генеральній сукупності. На відміну від $w(1-w)$ має максимум при $w$=0,5, рівний 0,25.

Проста випадкова повторна вибірка на практиці має обмежену сферу застосування. Насамперед практично недоцільно, інколи ж неможливо повторне спостереження тих самих одиниць. В результаті одиниця, яка одного разу потрапила у вибірку, повторного рахунку не піддається. Застосування безповторного відбору замість повторного диктується також вимогою підвищення ступеня репрезентативності вибірки, особливо при невеликому обсязі ($n$<30).

Середня помилка випадкової безповторної вибірки визначається за формулами[1]:

---

[1] Ці формули є спрощенням суворої формули $\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$. Оскільки чисельність генеральної сукупності ($N$) порівняно з чисельністю вибіркової сукупності ($n$) досить велика, то $\frac{N-n}{N-1} \approx 1 - \frac{n}{N}$. У багатьох випадках чисельність генеральної сукупності, навіть кінцевої, залишається невідомою. Тому зазначену вище форму взагалі неможливо застосувати і визначають середню помилку випадкового безповторного відбору як помилку випадкового повторного відбору.

для середньої

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}; \tag{12.10}$$

для частки

$$\mu_p = \sqrt{\frac{p(1-p)}{n}\left(1 - \frac{n}{N}\right)}, \tag{12.11}$$

де $N$ — обсяг генеральної сукупності.

У тих випадках, коли обсяг генеральної сукупності є достатньо великим у порівнянні з обсягом вибірки, то величина $\left(1 - \frac{n}{N}\right)$ буде близька одиниці, а тому її можна знехтувати. Тоді для підвищення надійності оцінок генеральних характеристик за вибірковими даними помилку випадкового повторного відбору визначають за формулою для простої випадкової повторної вибірки:

$$\mu = \sqrt{\frac{\sigma^2}{n}}\,. \tag{12.12}$$

Цією формулою зазвичай і користуються практично при випадковому відборі.

Таким чином, при власне-випадковому методі відбору середня помилка вибірки ($\mu$) визначається за формулами, наведеними в табл. 12.1.

Таблиця 12.1

**Формула розрахунку $\mu$**
**при власне-випадковому відборі**

| Спосіб відбору | Для середньої | Для частки |
|---|---|---|
| Повторний | $\mu = \sqrt{\frac{\sigma^2}{n}}$ | $\mu = \sqrt{\frac{w(1-w)}{n}}$ |
| Безповторний | $\mu = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}$ | $\mu = \sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)}$ |

Після всіх цих теоретичних міркувань розглянемо приклад розв’язання задачі на практиці.

При прийманні продукції електролампового заводу з усієї партії електроламп, що поставляється замовнику, контролером якості продукції було взято для перевірки в порядку випадкового відбору 100 ламп. Як основний показник якості електричних ламп було прийнято тривалість їх горіння.

Таблиця 12.2

| Тривалість горіння в годинах | Число зразків |
|---|---|
| 1 300–1 350 | 7 |
| 1 350–1 400 | 32 |
| 1 400–1 450 | 35 |
| 1 450–1 500 | 15 |
| 1 500–1 550 | 6 |
| 1 550–1 600 | 5 |
| Всього | 100 |

У табл. 12.2 обчислені вихідні дані, необхідних встановлення дисперсії середньої. Обчислимо середні помилки вибірки середньої тривалості горіння електричних ламп. Для цього складемо розрахункову таблицю (табл. 12.3).

Таблиця 12.3

**Обчислення середнього та середнього квадратичного відхилень для 100 електроламп**

| Групи електроламп за тривалістю горіння (у годинах) $x$ | $f$ | $xf$ | $(x-\bar{x})$ | $(x-\bar{x})^2$ | $(x-\bar{x})^2 f$ |
|---|---|---|---|---|---|
| 1 325 | 7 | 9 275 | –98 | 9 604 | 67 228 |
| 1 375 | 32 | 44 000 | –48 | 2 304 | 73 728 |
| 1 425 | 35 | 49 875 | 2 | 4 | 140 |
| 1 475 | 15 | 22 125 | 52 | 2 704 | 40 560 |
| 1 525 | 6 | 9 150 | 102 | 10 404 | 62 424 |
| 1 575 | 5 | 7 875 | 152 | 23 104 | 115 520 |
| Всього | 100 | 142 300 | – | – | 359 600 |

Середня тривалість горіння всіх відібраних електроламп буде рівною

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{142\,300}{100} = 1\ 423 \text{ години;}$$

а середнє квадратичне відхилення

$$\sigma_0^{\ 2} = \frac{\sum (x-\bar{x})^2 f}{\sum f} = \frac{359\,600}{100} = 3\ 596 \text{ годин.}$$

Застосовуючи формулу середньої помилки випадкового відбору, обчислимо середню помилку середньої:

$$\mu = \sqrt{\frac{3\,596}{100}} = \sqrt{35{,}96} \approx 6{,}0 \text{ годин.}$$

Вибіркове обстеження дозволяє зробити такий висновок: з ймовірністю 0,6827 можна стверджувати, що середній термін служби ламп у всій їх партії знаходиться в межах 1423 ±6 годин, тобто він не менше 1 417 годин ($\tilde{x} - \Delta$=1 423–6,0) і не більше 1 429 годин ($\tilde{x} + \Delta$=1423+ +6,0). Це можна записати нерівністю:

$$1417 \text{ год.} \leq \bar{x} \leq 1429 \text{ год.}$$

Те, що генеральна середня (або генеральна частка) не вийде за певні межі, можна стверджувати не абсолютно, а з певним ступенем ймовірності.

Доведено, що генеральна середня не вийде за межі, що дорівнює величині однієї середньої помилки, але не у всіх можливих випадках, а лише в 6827 вибірках з 10 000, тобто сформульоване вище положення про очікувані межі можна стверджувати лише з ймовірністю 0,6827. Величина ймовірності ($p$) 0,6827 є величиною відомого інтеграла ймовірності Лапласа ($F(t)$ при $t$=1.

З ймовірністю 0,6827 можна стверджувати, що в 6827 випадках з 10 000 ($t$=1) фактична помилка не перевищить 6 годин. Якщо ж ця величина ймовірності не влаштовує дослідника (з 10 000 всіх можливих випадків 3173 випадку не сприяють здійсненню даної події), то він може вказати з більшою ймовірністю граничну помилку, в межах якої може знаходитись шукана середня тривалість горіння в генеральній сукупності, т. е. у всіх партіях. Якби він прирівняв цю величину дворазової середньої помилки, тобто $t= \pm 2$, то генеральна середня не вийде вже за межі, що дорівнює двом середнім помилкам; ймовірність цього судження склала вже 0,9545. Це означає, що у 9545 випадків із 10 000 вибірок середня тривалість горіння електроламп по генеральної сукупності всієї партії не вийде межі 1423 години ±2·6,0 годин=1423 годин ±12 годин, тобто від 1411 годин (1423 – 12) та до 1435 годин (1423 + 12,0).

Достатньо часто на практиці дослідник і задовольняється такою величиною ймовірності. Але іноді величина можливої граничної помилки приймається рівною трьом середнім помилкам, тобто $t= \pm 3$, що дозволяє збільшити ймовірність судження про межі, в яких може бути значення середньої генеральної сукупності, до 0,9973. Так, з ймовірністю 0,9973 межі середнього терміну служби електроламп становитиме 1423±18,0 години, тобто 1405 годин (1423–18,0) до 1441

(1423+18,0). Остання ймовірність настільки велика, що надає судження дослідника вже практично достовірний характер.

Дослідник може далі поставити питання не про середні розміри покупки, а про частку електроламп відомої тривалості горіння. Наприклад, він може поставити питання: яка у всіх партіях частка електроламп, тривалість горіння яких не перевищує 1375 годин. В отриманій вибірці частка таких електроламп становить 0,23 (23:100). Отже, частка електроламп з тривалістю горіння більше 1375 години дорівнює 0,77 (1–0,23). Пам'ятаючи, що при альтернативному варіюванні дисперсія має вираз: $pq$=0,23·0,77=0,1771, то можна визначити величину середньої помилки:

$$\mu \approx \sqrt{\frac{0{,}1771}{100}} \approx 0{,}0421.$$

Таким чином, з ймовірністю 0,6827 дослідник може стверджувати, що у всій партії електроламп, що поставляються споживачеві, доля електроламп з тривалістю горіння не більше 1475 годин становить 0,23 ± 0,0421, або 23 % ± 4 21 %. Зрозуміло, що, як і в попередньому випадку, при ймовірностях 0,9545 і 0,9973 він повинен подвоїти і потроїти величину передбачуваної граничної помилки.

Питання про частку електроламп з різним терміном служби може бути вирішений аналогічно вирішенню питання про середній термін служби. У цьому випадку необхідно обчислювати середню помилку не вибірковою середньою, а вибіркової частки. Для її обчислення необхідно $\sigma^2$ замінити добутком часток $pq$ — як при варіації альтернативного ознаки. Припустимо, що термін служби електроламп менше 1350 год. неприйнятний для замовника. Таких ламп із терміном служби менше 1350 год. у вибірці виявилося 7 із 100, тобто $w$=0,07. Інші лампочки, що мають термін служби більше 1350 год., не будуть володіти даною ознакою. Їх кількість становить 93 (100 – 7), а частка $q$=0,93 (93:100). Як відомо раніше, дисперсія альтернативної ознаки становить $\sigma^2 = w(1-w) = 0{,}07 \times (1–0{,}07)=0{,}0651$. Звідси середня помилка вибірки для частки дорівнює

$$\mu = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{w(1-w)}{n}} = \sqrt{\frac{0{,}0651}{100}} = 0{,}026\ \%.$$

**Гранична помилка вибірки**. Гранична похибка — гранично припустима розбіжність між оцінкою показника за результатами вибіркового обстеження та дійсним значенням показника (невідомим) для генеральної сукупності, визначена на основі стандартної похибки вибірки та встановленого рівня довірчої ймовірності. Гранична помилка характеризує передбачувані межі, в яких може знаходитися невідома характеристика генеральної сукупності, її довірчі межі.

Теорія статистики встановлює співвідношення між межею помилки вибірки ($\Delta$), що гарантується з певною ймовірністю (*P*), величиною *t*, пов'язаною з цією ймовірністю і так званою середньою помилкою вибірки ($\mu$).

Середня помилка, помножена на прийнятий коефіцієнт *t*, має назву *граничної помилки вибірки*. Позначимо її літерою $\Delta$ (дельта).

*Гранична помилка вибірки* $\Delta$ розраховується за формулою:

$$\Delta = t\mu, \tag{12.13}$$

де $t$ — коефіцієнт кратності помилки (у літературі він називається коефіцієнт довіри), що залежить від ймовірності, з якою можна гарантувати, що гранична помилка не перевищить *t*-кратну середню помилку $\mu$, що вона не вийде за встановлені розрахунком межі.

Гранична помилка вибірки дорівнює *t*-кратному числу середніх помилок вибірки. Тому коли вимірюють величину помилки вибірки, завжди задаються певною ймовірністю. Ця ймовірність позначається буквою *t* («коефіцієнт довіри»).

Припустимо, що $t$=1. Тоді[1] $P(|\tilde{x}-\bar{x}|\leq 1\,\mu)=F(t=1)=0{,}6827$, тобто з ймовірністю, що дорівнює 0,68269, можна очікувати, що показники вибіркової сукупності будуть відрізнятися від показників генеральної сукупності не більше ніж на величину середньої квадратичної помилки вибірки ($\mu$); з ймовірністю 0,31731 показники можуть відрізнятися і більшою мірою.

Для $t$=2 ймовірність $p$=0,9545; для $t$=3 ймовірність $p$=0,9973. Як можна помітити, ймовірність появи помилки, рівної потрійної середньої помилки вибірки, тобто при $t = 3$, настільки мала, що її практично можна приймати за достовірність.

Значення *t* при даній ймовірності *P* визначається за таблицями значень функції, яка виражається інтегральною формулою Лапласа, і

---

[1] Значення інтегральної функції Лапласа для різних значень *t* обчислені та наводяться у таблицях Див.: *Ткач Є. І., Сторожук В. П.* Загальна теорія статистики: підручник. 3-тє вид. Київ: Центр учбової літератури, 2009. С. 420–421.

відображають залежність між $t$ і ймовірністю $P$. Чим більший обсяг вибірки, тим ймовірніше, що гранична помилка вибірки не перевищить потрійну середню помилку ($3\mu$).

Наводимо коротку витримку з таблиці значення ймовірностей у межах $t$ від 1 до 3,28 (значення $t$ взяті через 0,1, а ймовірності з точністю до 0,0001).

Таблиця 12.4

**Значення ймовірності за різної величини коефіцієнта довіри (*t*)**

| $t$ | Ймовірність $F(t)$ | $t$ | Ймовірність $F(t)$ | $t$ | Ймовірність $F(t)$ |
|---|---|---|---|---|---|
| 1,0 | 0,68269 | 1,8 | 0,92814 | 2,5 | 0,98758 |
| 1,1 | 0,72867 | 1,9 | 0,94257 | 2,58 | 0,99012 |
| 1,2 | 0,76986 | 1,96 | 0,94882 | 2,6 | 0,99068 |
| 1,3 | 0,80640 | 2,0 | 0,95450 | 2,7 | 0,99307 |
| 1,4 | 0,83849 | 2,1 | 0,96427 | 2,8 | 0,99489 |
| 1,5 | 0,86639 | 2,2 | 0,97219 | 2,9 | 0,99627 |
| 1,6 | 0,89040 | 2,3 | 0,97855 | 3,0 | 0,99730 |
| 1,7 | 0,91087 | 2,4 | 0,98360 | 3,28 | 0,99896 |

Джерело: *Ткач Є. І., Сторожук В. П.* Загальна теорія статистики: підручник. 3-тє вид. Київ: Центр учбової літератури, 2009. С. 406.

Як видно з таблиці, ймовірність помилки, що дорівнює або більше потрійної середньої помилки вибірки, тобто $\Delta \geq 3\mu$, вкрай мала і дорівнює 0,0027 (1 – 0,99730). Такі малоймовірні події вважаються практично неможливими, а тому величину $\Delta = 3\mu$ можна прийняти за межу можливої помилки вибірки.

Величину коефіцієнта довірчої ймовірності зазвичай пов’язують з коливанням досліджуваного ознаки. Чим менші коливання, тим більша впевненість у тому, що результати спостережень не вийдуть за межі заданої помилки, і тим самим береться коефіцієнт довірчої ймовірності.

Наведені вище формули відносяться до повторного власне-випадкового відбору. Стосовно безповторної вибірки в підкореневий вираз наведених вище формул величини помилки вибірки вводиться додатковий множник , $\left(1-\frac{n}{N}\right)$, де $n$ — загальна кількість відібраних одиниць сукупності, а $N$ — первісна кількість одиниць всієї генеральної сукупності. Таким чином, $\frac{n}{N}$ є відносною частиною відібраних одиниць. За цими ж формулами обчислюються помилки механічного відбору. (Помилки типового та гніздового відборів обчислюються за іншими формулами).

Отже, формули граничної помилки вибірки (позначимо її $\Delta$) для власне-випадкового та механічного відборів приймають такий вид:

Таблиця 12.5

| $\Delta = t\mu$ | При повторному відборі | При безповторному відборі |
|---|---|---|
| 1. Для середньої | $\Delta = t\sqrt{\frac{\sigma^2}{n}}$ | $\Delta = t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$ |
| 2. Для частки | $\Delta = t\sqrt{\frac{w(1-w)}{n}}$ | $\Delta = t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$ |

Виходячи з наведених формул граничних помилок, можна виявити фактори, що визначають їхню величину.

1. Коливання (варіація) досліджуваної ознаки, що позначається у формулах $\sigma^2$ і $w(1-w)$. Аналіз показує, що розмір помилки прямо пропорційний величині коливання ознаки.

2. Коефіцієнт кратності помилки (або коефіцієнт довіри), що залежить від ймовірності, з якою дослідник хоче одержати межі помилок. Чим вище задана ймовірність, тим більше $t$ і відповідно більше величина граничної помилки ($\Delta$).

3. Спосіб відбору одиниць у вибіркову сукупність (повторний чи безповторний). Множник $\left(1-\frac{n}{N}\right)$ завжди менше одиниці на величину, що характеризує відношення чисельності вибірки до обсягу одиниць у генеральній сукупності. Тому гранична помилка безповторного відбору завжди менше граничної помилки по-другого відбору, оскільки підкореневий вираз формули помилки безповторного відбору множиться на число, менше одиниці.

4. Чисельність вибірки. Оскільки $n$ входить у знаменник дроби, то між чисельністю вибірки і граничною помилкою існує назад пропорційний зв'язок: при збільшенні чисельності вибірки гранична помилка зменшується і навпаки, при зменшенні чисельності вибірки гранична помилка зростає.

Будь-яка формула граничної помилки вибіркової середньої (частки) дозволяє вирішувати три групи задач:

1) визначати межу можливої помилки середньої (частки), тобто величину можливих відхилень показників генеральної сукупності від

показників вибіркової сукупності. Слід зазначити, що фактична помилка, як правило, завжди менша за граничну;

2) визначати необхідну чисельність вибірки, за якої межі можливої помилки не перевищать деякої наперед заданої величини;

3) визначати ймовірність того, що у проведеній вибірці помилка перебуватиме в заданих межах.

Рішення тієї чи іншої з поставлених завдань залежатиме від того, які є вихідні дані, що входять до формули граничної помилки, і які невідомі. Так, для оцінки граничної помилки вибірки необхідно додатково задати певну ймовірність висновків, або величину коефіцієнта довіри, а також обсяг вибірки *n*. Якщо ж задані гранична помилка вибірки $\Delta$ і величину коефіцієнта *t*, то можна визначити обсяг вибірки, необхідний для забезпечення заданої точності. Для розрахунку коефіцієнта довіри *t* необхідно задати граничну помилку вибірки $\Delta$ та її обсяг *n*.

Наведемо приклад. У вибірці обсягом 1000 одиниць частка бракованих виробів склала 4 %. Яка можливість, що у всій партії виробів (10 000 прим.) частка бракованих виробів вбирається у 4,5 %. Довірча ймовірність, яку потрібно визначити, є функцією від *t*, величина якої може бути визначена з формули граничної помилки вибірки $\Delta_p = t\mu_p$, звідки $t = \frac{\Delta_p}{\mu_p}$. Величина середньої помилки може бути визначена як різниця між максимально допустимою генеральною часткою (за умовою вона дорівнює 4,5 %) і часткою бракованих деталей у вибірці (за умовою 4,0 %).

Таким чином, (4,5 % – 4,0 %)=0,5 %. Оскільки вибірка виробів для перевірки якості продукції є безповторною, величину середньої помилки вибірки визначаємо за формулою:

$$\mu_p = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0{,}04(1-0{,}04)}{1\,000}\left(1 - \frac{1000}{10\,000}\right)} = 0{,}59\ \%.$$

Знаходимо коефіцієнт кратності помилки (коефіцієнт довіри):

$$t = \frac{\Delta_p}{\mu_p} = \frac{0{,}5}{0{,}59} = 0{,}85.$$

По таблиці інтегральної функції Лапласа ймовірність, що відповідає даній величині коефіцієнта, дорівнює 0,60468.

Для вирішення всіх перелічених задач необхідно знати дисперсію ознаки у вибірковій сукупності ($\sigma^2$), а у разі безповторної вибірки —

чисельність генеральної сукупності ($N$). Чисельність генеральної сукупності ($N$) завжди відома; $\sigma^2$ і $w(1-w)$ можуть бути визначені за наведеними вище формулами. Лише $t$, $n$ і $\Delta$ виступають змінними величинами, значення яких можуть встановлюватися дослідником при організації проведення вибіркового спостереження.

**Оцінка за даними вибірки характеристик (параметрів) генеральної сукупності**. Наукова і практична цінність вибіркового статистичного спостереження полягає в тому, що воно, на відміну від інших видів несуцільного спостереження, дозволяє на основі деякої, певним чином відібраної частини сукупності судити про ряд властивостей всієї сукупності фактів або явищ з відомим ступенем точності. Однією з основних завдань є оцінка за даними вибірки узагальнюючих показників (параметрів) генеральної сукупності.

Розглянемо визначення величини середньої арифметичної генеральної сукупності на основі даних вибірки.

Вибіркове спостереження дає можливість визначити середню арифметичну вибіркову сукупність ($\tilde{x}$) і величину граничної помилки середньої арифметичної ($\Delta_x$). Обчислення цих показників дозволяє встановити (з певною ймовірністю), наскільки вибіркова середня може відрізнятися від генеральної середньої у більшу чи меншу сторону. Величина середньої арифметичної генеральної сукупності буде представлена в певних межах. Нижня межа дорівнюватиме $\tilde{x} - \Delta_{\tilde{x}}$, а верхня межа — $\tilde{x} + \Delta_{\tilde{x}}$. Межі, в яких з даним ступенем ймовірності буде заключена невідома величина параметра, що оцінюється, називають *довірчими*, а ймовірність $P$ — *довірчою ймовірністю*. Довірчий інтервал для генеральної середньої арифметичної може бути записаний так:

$$\tilde{x} - t\mu_{\bar{x}} < \bar{x} < \tilde{x} + t\mu_{\bar{x}}. \qquad (12.14)$$

При організації вибіркового спостереження найчастіше довірчу ймовірність встановлюють на рівні 0,95 або 0,99. Імовірність того, то величина середньої вийде за довірчі межі, дорівнюватиме 1 – $P$, тобто буде дорівнює 0,05 або 0,01. Подія, що має таку ймовірність, вважається практично неможливим.

Аналогічно визначаються довірчі інтервали для генеральної частки:

$$w - t\mu_p < p < w + t\mu_p. \qquad (12.15)$$

Таким чином, довірчі інтервали для генеральної середньої та генеральної частки залежить від граничної помилки вибірки $\Delta_x$ і $\Delta_p$. Оскільки величина граничної помилки вибірки дорівнює $t\mu$, то цьому випадку довірчий інтервал для середньої та вибірки буде залежати від величини середньої помилки вибірки та рівня довірчої ймовірності (коефіцієнта кратності помилки). Чим більша величина середньої помилки ($\mu_{\bar{x}}$ і $\mu_p$), тим нижче точність оцінки. Величина граничної помилки значною мірою визначається рівнем довірчого інтервалу, значення якого відповідає рівню довірчої ймовірності.

Приклад.

1. Припустимо, з партії продукції чисельністю 10 000 шт. ($N$=10 000) у порядку випадкової безповторної вибірки було відібрано 1 000 шт. ($n$ = 1 000). Частка браку за даними вибіркового обстеження була встановлена у розмірі 2,4 (0,024). Розмір стандартної помилки випадкової безповторної вибірки складе 0,0046:

$$\Delta = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{0{,}024(1-0{,}024)}{1000}\left(1-\frac{1000}{10\,000}\right)} = 0{,}0046.$$

З ймовірністю, що дорівнює 0,9545 ($t$=2), можна очікувати, що гранична помилка частки браку не перевищить 0,0092 ($\Delta_p = 2\mu_p$), і частка браку в генеральній сукупності буде в інтервалі 0,024±0,0092, і частка браку в генеральній сукупності буде перебувати в інтервалі 1,48 % ≤ $p$≤3,32 %. Якщо ймовірність, що дорівнює 0,9973, очікується, що гранична помилка частки не перевищить 0,0138, тобто частка браку в генеральній сукупності буде в інтервалі 1,02 %≤$p$≤3,78 %. Таким чином, з ймовірністю 99,73% можна очікувати, з ймовірністю 99,73 % очікується, що кількість бракованих деталей буде перебувати в інтервалі від 102 до 378 шт., тоді як з ймовірністю 95,45 % кількість бракованої продукції складе 148 ÷ 332 шт.

Незначне збільшення достовірності висновків (з 95,4 % до 99,7 %) вимагає включення у вибірку додаткових десятків і сотень одиниць і тим більших витрат вимагає подальше підвищення точності. При цьому в економічних розрахунках найчастіше рекомендується використовувати довірчу ймовірність $P$=0,95, або $P$=0,954 ($t$=2).

Розглянемо приклад розрахунку граничної помилки середньої та частки за різних видів відбору.

2. *Розрахунок граничної помилки вибіркової середньої при повторному відборі*. З сукупності 10 000 деталей відібрано власне випадковим поворотним методом 1 000 деталей, для яких середня вага деталі

виявився рівною 55 г, дисперсія дорівнює 54. З ймовірністю 0,954 потрібно визначити межі, в яких знаходиться середня вага деталі в генеральній сукупності.

Дано:

$N = 10\,000$; $n$=1 000; $\bar{x}$=55; $\sigma^2$ =49;

Генеральна середня $\bar{x}$ відрізняється від вибіркової середньої $\tilde{x}$ на величину помилки вибірки:

$$\bar{x} = \tilde{x} \pm \Delta_{\tilde{x}}.$$

Гранична помилка для середньої при безповторному відборі розраховується за формулою: $\Delta_{\bar{x}} = t\sqrt{\dfrac{\sigma^2}{n}}$. З ймовірністю 0,954 помилка вибірки не перевищить двох середніх помилок, так як значення $t$ при $P$=0,954 дорівнює 2.

Підставимо значення формулу помилки вибірки:

$$\Delta_{\bar{x}} = 2 \cdot \sqrt{\frac{54}{1000}} = 0{,}46.$$

Визначимо верхню межу генеральної середньої:

$$\bar{x} = \tilde{x} + \Delta_{\tilde{x}} = 55 + 0{,}45 = 55{,}46 \text{ г}.$$

Визначимо нижню межу генеральної середньої:

$$\bar{x} = \tilde{x} - \Delta_{\tilde{x}} = 55 - 0{,}46 = 54{,}54 \text{ г}.$$

З ймовірністю 0,954 можна стверджувати, що середня вага деталі в генеральній сукупності знаходиться в межах 54,54 г$\leq\bar{x}\leq$55,46 г.

Середня помилка вибірки для середньої при безповторному відборі становитиме:

$$\mu_{\bar{x}} = \sqrt{\frac{54}{1000} \cdot \left(1 - \frac{1000}{10000}\right)} = 0{,}22.$$

Підставимо значення у формулу помилки вибірки

$$\Delta = t\mu = 2 \cdot 0{,}22 = 0{,}44.$$

$$\bar{x} = \tilde{x} \pm \Delta_{\tilde{x}}.$$

3. *Розрахунок граничної помилки вибіркової середньої при повторному відборі*. На заводі з числом робітників 1 000 осіб було проведено 5 % вибіркове обстеження віку робочих методом випадкового безповторного відбору. В результаті обстеження отримано такі дані:

| Вік робітників, років | До 30 | 30–40 | 40–50 | 50–60 | Понад 60 |
|---|---|---|---|---|---|
| Число робітників | 7 | 23 | 10 | 6 | 4 |

З ймовірністю 0,997 визначити межі, в яких знаходиться вік робітників заводу.

Для визначення меж генеральної середньої необхідно розрахувати вибіркову середню і помилку вибіркової середньої. Розрахуємо середній вік робітників у вибірковій сукупності та дисперсію вибіркової сукупності:

Таблиця 12.6

| Середній вік робочих $x$ | Цент інтервалу х | Кількість робітників $f$ | $xf$ | $x-\tilde{x}$ | $(x-\tilde{x})^2$ | $(x-\tilde{x})^2 f$ |
|---|---|---|---|---|---|---|
| До 30 | 25 | 7 | 175 | –15,4 | 237,16 | 1 660,12 |
| 30–40 | 35 | 23 | 805 | –5,4 | 29,16 | 670,68 |
| 40–50 | 45 | 10 | 450 | +4,6 | 21,16 | 211,60 |
| 50–60 | 55 | 6 | 330 | +14,6 | 213,16 | 1 278,96 |
| Понад 60 | 65 | 4 | 260 | +24,6 | 605,16 | 2 420,64 |
| Всього | – | 50 | 2 020 | – | 1 105,80 | 6 242 |

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{2020}{50} = 40{,}4 \text{ роки;}$$

$$\sigma^2 = \frac{\sum (x-\bar{x})^2 f}{\sum f} = \frac{6242}{50} = 124{,}84.$$

Гранична помилка вибіркової середньої при випадковому відборі розраховується за формулою $\Delta = t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$. З ймовірністю 0,997 наша помилка вибірки не перевищує трьох середніх помилок ($t$=3):

$$\Delta = 3\sqrt{\frac{124{,}84}{50}\left(1-\frac{50}{1\,000}\right)} = 4{,}6 \text{ роки.}$$

Визначимо межі, в яких знаходиться середній вік робітників заводу:

$$\bar{x} = \tilde{x} \pm \Delta_{\tilde{x}} = 40{,}6 \pm 4{,}6.$$

З ймовірністю 0,997 можна стверджувати, що середній вік робітників заводу знаходиться в межах $36{,}0 \leq \bar{x} \leq 45{,}2$.

## 12.4. Визначення необхідної чисельності вибірки

Одним із визначальних етапів проектування та організації вибіркового спостереження є визначення обсягу вибіркової сукупності, яка необхідна для забезпечення визначеної точності розрахунку генеральних середніх. Встановлення обсягу вибірки необхідно для забезпечення репрезентативності та належної якості результатів вибіркового обстеження. Від обсягу вибірки залежатиме не тільки якість результатів, але і вартість обстеження. З одного боку, занадто великий обсяг вибірки призводить до невиправданих витрат та неефективного використання коштів, а з іншого боку — недостатній обсяг вибірки є причиною незадовільної якості результатів обстеження.

Обсяг вибірки тісно пов'язаний із заданою точністю вибірки, яка виражається величиною граничної помилки вибірки.

Формула для визначення величини помилки вибірки дає можливість не тільки визначати точність вибіркового обстеження, але і розраховувати попередньо необхідну чисельність вибірки, щоб помилка вибірки не перевищувала певні, заздалегідь встановлені розміри. Практично це навіть важливіше, оскільки проектування будь-якого вибіркового спостереження має обов’язково включати визначення його чисельності.

Шляхом нескладного перетворення можна отримати формули визначення *необхідної чисельності вибірки*.

Вище було зазначено, що величина середньої помилки при випадковому відборі змінюється обернено пропорційно до кореня квадратного з числа спостережень. З формули $\mu = \frac{\sigma}{\sqrt{n}}$ випливає, що при заданій помилці ($\mu$) і за відомою величиною $\sigma$ можна визначити і необхідний обсяг вибірки для отримання бажаної величини середньої випадкової помилки. Справді, якщо помилка вибірки при вибірковому вимірюванні середнього значення визначається за формулою:

$$\mu = \frac{\sigma}{\sqrt{n}}, \text{ то } \mu^2 = \frac{\sigma^2}{n}. \tag{12.16}$$

Звідси необхідна чисельність вибірки, при якій може бути забезпечена бажана величина середньої випадкової помилки, становить:

$$n = \frac{\sigma^2}{\mu^2}. \tag{12.17}$$

Отже, *необхідна чисельність вибірки при вимірюванні середньої дорівнює середньому квадрату відхилень, поділеному квадрат заданої точності*. Підставляючи в цю формулу дисперсію, визначену на основі вибірки (це цілком обґрунтовано, так як вибіркова дисперсія, як вказувалося раніше, приблизно, з малою ймовірністю великої випадкової похибки відтворює середню величину дисперсії всіх можливих за даних умов вибірок), і квадрат бажаної величини середньої помилки, дослідник зможе визначити необхідну чисельність вибірки.

Якщо у формулу запровадити коефіцієнт довіри *t*, вона прийме такий вид:

$$n = \frac{t^2 \times \sigma_0^2}{\mu^2}. \tag{12.18}$$

На практиці нерідко задаються не абсолютною величиною граничної помилки середньої $\Delta_{\bar{x}}$, а величиною відносної похибки, вираженої у відсотках. В цьому випадку відносна величина граничної помилки середньої обчислюється по формулою $\Delta_{відн} = \frac{\Delta_{\bar{x}}}{\bar{x}} \cdot 100\,\%$.

Звідки

$$\Delta_{\bar{x}} = \frac{\Delta_{відн} \cdot \bar{x}}{100}. \tag{12.19}$$

Підставимо величину $\Delta_{\bar{x}}$, виражену через відносну похибку, у формулу для визначення граничної помилки, отримаємо наступну форму розрахунку чисельності вибірки:

$$n = \frac{t^2 \times \sigma_0^2}{\Delta_{відн}^2 \cdot \bar{x}^2} \cdot 100^2. \tag{12.20}$$

Як відомо раніше, $\frac{\sigma}{\bar{x}} \cdot 100$ представляє собою коефіцієнт варіації за середньоквадратичним відхиленням. Тоді

$$n = \frac{t^2 \times V^2}{\Delta_{відн}^2}. \tag{12.21}$$

Наприклад, за даними вибіркового обстеження деталей коефіцієнт варіації становить 24 %. Скільки потрібно відібрати одиниць, щоб з ймовірністю 0,954 гранична помилка вибірки не перевищувала 4 %?

При *V*= 24%; $\Delta_{відн} = 4$ %; *t*=2 обсяг вибірки складе:

$$n = \frac{t^2 \times V^2}{\Delta^2_{відн}} = \frac{2^2 \cdot 24^2}{4^2} = \frac{4 \cdot 576}{16} = 144 \text{ деталей.}$$

Формула (12.18) застосовується у разі повторного відбору. При безповторному відборі формула визначення обсягу вибірки матиме вигляд[1]:

$$n = \frac{t^2 \sigma_0^2 N}{\Delta^2_{\bar{x}} N + t^2 \sigma_0^2}. \qquad (12.22)$$

До сих пір ми розглядали варіювання кількісних ознак. У разі якісних ознак варіація матиме більш простий вигляд: кожна варіанта буде або володіти даною ознакою, або не буде нею володіти (наприклад, населення складається з чоловіків і жінок). Як відомо, така варіація зветься альтернативною.

Для альтернативного варіювання середня помилка («помилка долі») розраховується аналогічно формулі і для кількісних показників:

$$\mu_p = \sqrt{\frac{pq}{n}}, \qquad (12.23)$$

де $p$ — частка одиниць у генеральній сукупності, які мають досліджувану ознаку;

$q$ — частка одиниць у генеральній сукупності, які не мають досліджуваної ознаки ($q = 1 - p$).

Але так як генеральна частка нам невідома, то, замінюючи її на вибіркову, отримаємо формулу обчислення середньої помилки вибірки для альтернативного варіювання:

$$\mu_p = \sqrt{\frac{w(1-w)}{n}}, \qquad (12.24)$$

де $w$ — частка ознаки (наприклад, якщо відсоток браку продукції становить від усієї продукції 0,5 %, то при $w$ =0,005);

---

[1] Якщо $\Delta = t\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}$, то звівши в квадрат обидві частини рівності, матимемо: $\Delta^2 = \frac{t^2\sigma^2}{n}\left(1 - \frac{n}{N}\right)$. Далі розділивши обидві частини рівності на добуток $t^2\sigma^2$, отримаємо: $\frac{\Delta^2}{t^2\sigma^2} = \frac{1}{n} - \frac{1}{N}$, звідки $\frac{1}{n} = \frac{\Delta^2}{t^2\sigma^2} + \frac{1}{N}$, або $\frac{1}{n} = \frac{\Delta^2 N + t^2\sigma^2}{t^2\sigma^2 N}$, $n = \frac{t^2\sigma^2 N}{\Delta^2 N + t^2\sigma^2}$. .

$n$ — чисельність одиниць вибіркової сукупності.

Звідси слідує, що обсяг вибірки за альтернативної варіації визначається за формулою:

$$n = \frac{w(1-w)}{\mu_p^2}. \qquad (12.25)$$

Отже, *необхідна чисельність вибірки дорівнює частці, помноженої на доповнення до чисельності і поділеної на квадрат заданої точності.*

Якщо у формулу ввести коефіцієнт довіри $t$, вона набуде такого вигляду:

$$n = \frac{t^2 w(1-w)}{\Delta^2{}_p}. \qquad (12.26)$$

За формулою (12.26) визначається обсяг для власне-випадкової повторної вибірки.

Для безповторної вибірки формула (12.26) доповнюється виразом $\frac{N-n}{N-1}$ (де $N$ — обсяг генеральної сукупності), яке при великому значенні $N$ можна записати як $1-\frac{n}{N}$.

Підставивши цей вираз у формулу (12.26), отримаємо

$$\Delta_p = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}, \qquad (12.27)$$

звідки обсяг вибірки при безповторному відборі дорівнює

$$n = \frac{t^2 w(1-w)N}{\Delta^2 N + t^2 w(1-w)}. \qquad (12.28)$$

Ця формула незручна для практичних розрахунків, оскільки часто обсяг генеральної сукупності ($N$) не відомий. Тому для обчислення чисельності вибірки використовують формулу (12.26).

Таким чином, залежно від способу відбору обсяг вибірки визначається за такими формулами:

для повторного відбору $$n = \frac{t^2 w(1-w)}{\Delta^2{}_p}; \qquad (12.29)$$

для безповторного відбору $$n = \frac{t^2 w(1-w)N}{\Delta^2 N + t^2 w(1-w)}. \qquad (12.30)$$

Дуже часто на практиці для зручності помилку виражають не в абсолютних величинах, а у відсотках. У зв'язку з тим, що на у практичній діяльності найчастіше необхідна не абсолютна, а відносна помилка[1], то для розрахунку чисельності вибірки у формулі (12.26) вносять певні зміни. Остаточна формула для розрахунку обсягу вибірки має вигляд:

$$n = \frac{t^2 w(1-w)100^2}{d^2 w}. \qquad (12.31)$$

Формула (12.31) часто використовується як основа для різних модифікацій.

Позначивши кількість досліджуваних елементів через $q$, маємо:

$$\bar{w}_q = \frac{1}{q}, \qquad (12.32)$$

де $\bar{w}_q$ — середнє значення частки досліджуваних елементів.

Підставивши формулу (12.31) $w$ через $\frac{1}{q}$, отримаємо

$$n = \frac{t^2 (q-1)100^2}{d^2}. \qquad (12.33)$$

В результаті отримана більш спрощена формула розрахунку обсягу вибірки. Великою зручністю формули (12.33) є те, що досліднику заздалегідь відомі всі її показники.

У зв'язку з тим, що досліднику найчастіше невідома питома вага досліджуваного елемента (частки), застосування формули (12.31) характеризується великою незручністю. Тому необхідну величину $w$ завжди беруть орієнтовно, посилаючись на минулі дослідження або на особистий досвід.

Для випадку, коли частість $w$ навіть приблизно невідома, можна зробити приблизний розрахунок чисельності вибірки, враховуючи максимальну величину дисперсії частки, що дорівнює 0,25. Тоді

для повторного відбору $$n = \frac{t^2}{4\Delta^2_p}; \qquad (12.34)$$

---

[1] $$d = \frac{\Delta}{w},$$

де $d$ — відносна помилка;

$\Delta$ — абсолютна помилка;

$w$ — питома вага досліджуваного елементу (частки).

для безповторного відбору $n = \dfrac{0{,}25t^2 N}{N\Delta^2_p + 0{,}25t^2}$. (12.35)

Розглянемо використання наведених формул.

Приклади.

*Приклад 1.* Проведено вибіркове обстеження товщини металічних листів. Для виміру товщини листів було взято 200 проб. В результаті обстеження середня товщина за вибірковими даними визначена в 245 *мм*, а середній квадрат відхилень ($\sigma^2$) дорівнює 484 *мм*. Потрібно визначити точність вибіркового спостереження.

Тут число проб $n$ = 200; $\sigma^2$ = 484 *мм*. Розрахунок ведемо за такою формулою:

$$\mu = \sqrt{\frac{\sigma_0^2}{n}};$$

$$\mu = \sqrt{\frac{484}{200}} = \sqrt{2{,}42} \approx \pm 1{,}56 \text{ мм}.$$

Помилка показує, у яких межах у той чи інший бік очікується відхилення середньої товщини листів у всій партії за вибірковими даними. У нашому прикладі це відхилення становило ±1,56 *мм*. Отже, можна з ймовірністю, що дорівнює 0,683 стверджувати, що генеральна середня лежить у межах: від 243,44 *мм* (245–1,56) до 246,56 *мм* (245+1,56).

Всі перераховані розрахунки виходили з того, що середня помилка встановлювалася з ймовірністю 0,683. Практично такий ступінь надійності буває достатнім. Проте є точніші прийоми розрахунку. Так, з ймовірністю, що дорівнює 0,954, середня товщина листа не вийде за межі: від 241,88 *мм* (245–3,12) до 248,12 *мм* (245+3,12).

*Приклад 2*. Потрібно визначити середню помилку, якщо взято 300 проб.

Якщо взято 300 проб, середня помилка становитиме:

$$\mu = \sqrt{\frac{484}{300}} = \sqrt{1{,}613} \approx \pm 1{,}27 \text{ мм}.$$

*Приклад 3.* Виходячи з умов попереднього прикладу потрібно визначити, скільки проб треба взяти, якщо помилка повинна бути не більше 1,4 *мм*.

На підставі формули $n = \dfrac{\sigma^2}{\mu^2}$ отримаємо:

$$n = \frac{484}{1{,}4^2} = \frac{484}{1{,}96} = 247 \text{ проб.}$$

*Приклад 4*. Треба визначити чисельність проб у разі, якщо товщина листів більш строката і $\sigma^2$ = 729 *мм*.

Чисельність проб становитиме:

$$n = \frac{\sigma^2}{\mu^2} = \frac{729}{1{,}4^2} \approx 372 \text{ проб.}$$

*Приклад 5*. Потрібно визначити за умов прикладу 3 кількість вибірки, якщо точність вибірки збільшимо вдвічі, тобто щоб розмір помилки не перевищував 0,7 *мм*.

$$n = \frac{484}{0{,}7^2} \approx 988 \text{ проб.}$$

Таким чином, при зменшенні помилки у 2 рази чисельність вибірки зростає у 4 рази ($247 \times 4 = 988$).

*Приклад 6*. Яка має бути чисельність вибірки, якщо розмір помилки, зазначеної в прикладі 3, гарантувати з ймовірністю 0,997 ($t$=3)?

Якщо брати до уваги зазначене кратне ставлення заданої середньої помилки до отриманої помилки вибіркової середньої, то вищенаведена формула повинна бути уточнена наступним чином:

$$\mu = \frac{t\sigma}{\sqrt{n}}.$$

Звідси ясно, що формула для визначення обсягу вибірки матиме такий вигляд:

$$n = \frac{t^2\sigma^2}{\mu^2}.$$

Для вищенаведеного прикладу: $\sigma^2$=484 *мм*, $\mu$=1,4 *мм*, $t$=3.

Обсяг вибірки складе:

$$n = \frac{3^2 \times 484}{1{,}4^2} = \frac{4356}{1{,}96} = 2222 \text{ проб.}$$

тобто при вимозі, щоб надійність результату вибіркового дослідження піднялася від ймовірності 0,68 практично повної достовірності, слід збільшити число вимірювань з 247 до 2222.

Іноді при проведенні вибіркового спостереження потрібно визначити чисельність вибірки, що забезпечує задану точність за двома ознаками одночасно. Крім того, для різних ознак може бути потрібна різна точність результатів спостереження. У таких випадках слід при визначенні необхідного обсягу вибірки орієнтуватися на ту ознаку, яка при найбільшому коливанні має найменшу величину припустимої помилки.

*Приклад 7.* Сукупність деталей становить 6 000 одиниць. У порядку випадкового безповторного відбору було відібрано партію одиниць зі 100 шт. За даними суцільного спостереження визначили дисперсію ваги деталей, рівну 42, і дисперсію частки виробів першого сорту, рівну 0,06. Необхідна точність результатів спостереження: щодо середньої ваги — 1,2 *г*, щодо частки виробів першого сорту — 5 %. Потрібно визначити необхідну чисельність випадкової безповторної вибірки, що забезпечує точність визначення середньої ваги виробу і частки виробів першого сорту з ймовірністю 0,954.

Орієнтуючись на задану точність результатів спостереження визначення середньої ваги виробів: $N$=6 000; $\sigma_x^2=42$; $\Delta_{\bar{x}}$=1,2; $P$ = 0,954; $t$=2. Тоді

$$n=\frac{t^2\sigma_x^2N}{\Delta_{\bar{x}}^2N+t^2\sigma_x^2}=\frac{2^2\cdot 42\cdot 6\,000}{1{,}2^2\cdot 6\,000+2^2\cdot 42}\approx 114 \text{ деталей.}$$

Отже, для забезпечення необхідної точності при визначенні середньої ваги необхідно відібрати 114 виробів.

Орієнтуючись на задану точність результатів спостереження визначення частки виробів першого сорту: $N$=6 000; $\sigma_p^2=0{,}06$; $\Delta_p$=0,05; $P$ = 0,954; $t$=2. Отже

$$n=\frac{t^2\sigma_p^2N}{\Delta_{\bar{x}}^2N+t^2\sigma_p^2}=\frac{2^2\cdot 0{,}06\cdot 6\,000}{0{,}05^2\cdot 6\,000+2^2\cdot 0{,}06}\approx 95 \text{ деталей.}$$

Таким чином, відбір 95 деталей, що забезпечують необхідну точність визначення частки виробів першого сорту, недостатній, щоб забезпечити необхідну точність визначення середньої ваги деталі. Тому слід провести відбір 114 деталей, щоб забезпечити необхідну точність за двома ознаками.

Визначаючи чисельність вибірки та її точність, необхідно враховувати, що чим більший абсолютний обсяг вибірки, тим менш відчутно впливає на точність результату включення у вибірку додаткових десятків і сотень одиниць і тим більших витрат вимагає подальше підвищення точності.

Для спрощення визначення обсягу вибірки можна скористатися заздалегідь розрахованими таблицями, в яких вказується число одиниць у вибірковій сукупності при даній величині довірчої ймовірності і припустимої помилки. Наприклад, для визначення частки ознаки в досліджуваній сукупності можна використовувати табл. 12.7.

Таблиця 12.7

| Довірча ймовірність | Гранична помилка, % | | | | Довірча ймовірність | Гранична помилка | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 5 | 10 | | 1 | 2 | 5 | 10 |
| 0,90 | 6 763 | 1 670 | 270 | 67 | 0,90 | 16 587 | 4 146 | 663 | 165 |
| 0,95 | 9 603 | 2 400 | 384 | 96 | 0,95 | 22 018 | 5 504 | 880 | 220 |

Наприклад, щоб при довірчій ймовірності 0,90 відхилення від вибіркової частки не перевищило 1 %, обсяг вибірки повинен становити 6 763 одиниць. Якщо граничну помилку збільшити до 5 %, потрібно обстежити 270 одиниць, тобто у 25 разів менше. А для того, щоб гарантувати з ймовірністю 0,90 величину відхилення від генеральної частки, що не перевищує 10 %, обсяг вибірки складе всього 67 одиниць. Аналіз таблиці показує необхідну чисельність вибірки при зміні граничної помилки, що дозволяє визначити доцільність додаткових витрат на збільшення точності.

## 12.5. Різні види вибіркового спостереження

Розглянуті вище поняття та категорії вибіркового спостереження відносилися до простої випадкової вибірки. Як вказувалося раніше, крім простої випадкової вибірки, при проведенні вибіркового спостереження використовують різні способи формування вибіркової сукупності: типову (районовану) вибірку, серійну (гніздову) вибірку, механічну вибірку, комбіновану вибірку. Ці форми вибіркового спостереження представляють собою подальший розвиток і видозміну простої випадкової вибірки. Застосування цих форм викликається або міркуваннями здешевлення, або полегшення процесу спостереження, або відсутністю необхідної інформації для складання списку одиниць спостереження. Різні способи відбору впливають на величину помилки вибірки.

Для кожного конкретного вибіркового спостереження величина середньої помилки вибірки може бути точно визначена за відповідними формулами.

**Розрахунок середньої помилки вибірки при типовому відборі**. Почнемо розгляд видів вибірки з організації *типової вибірки,* яка дає більш точні результати в порівнянні з іншими способами відбору одиниць у вибіркову сукупність. Як правило, більшість соціально-економічних явищ представляють собою неоднорідні сукупності. За наявності у складі генеральної сукупності різноманітних типів явищ з різними значеннями ознаки, необхідно організувати вибірку таким чином, щоб забезпечити представництво у всіх однорідних групах. Для цього всі одиниці генеральної сукупності розбиваються на кілька якісно однотипних груп за ознаками, які впливають на показники, що вивчаються. Як типи можуть використовуватися групи, що склалися в практиці планування та статистики. Наприклад, при вибірковому обстеженні доходів і витрат населення в Україні виділяють групи населення в сільській місцевості, в «великих» і малих містах. Крім того, населення групується за децільними (10%-ми) групами за рівнем середньодушових еквівалентних загальних доходів у місяць. При вивченні внутрішньозмінних втрат робочого часу можуть, наприклад, бути утворені групи висококваліфікованих та малокваліфікованих робітників. Потім з кожної типової групи власне-випадковою або механічною вибіркою відбирається деяка кількість одиниць, які включаються у вибіркову сукупність.

Таким чином, при проведенні типової вибірки необхідно розбити загальний обсяг вибірки $n$ між групами і визначити число одиниць, які потрібно відібрати з кожної групи.

При організації типового відбору варіанти розрахунку чисельності вибірки диференціюються залежно від прийомів відбору статистичних одиниць. Відбір одиниць з типів може здійснюється трьома методами: пропорційно чисельності типових груп, непропорційно чисельності одиниць типових груп, пропорційно коливання у групах.

Найбільш часто вживана форма типового відбору — так званий пропорційний відбір, коли кількість одиниць, що відбираються у вибірці, пропорційно питомій вазі цієї групи в генеральній сукупності.

Розрахунок обсягу вибірки з кожної типової групи проводиться за формулою:

$$n_i = n \cdot \frac{N_i}{N}, \qquad (12.36)$$

де $n_i$ — кількість спостережень з $i$-ї типової групи;

$n$ — чисельність одиниць вибіркової сукупності;

$N_i$ — обсяг $i$-ї типової групи у генеральній сукупності;

$N$ — обсяг генеральної сукупності.

Можливий й інший варіант (назвемо його умовно непропорційним обсягу груп), коли з кожної групи відбирається однакова кількість одиниць. При відборі, непропорційному обсягу типової вибірки, загальна кількість одиниць, що відбираються, ділиться на число типових груп, тобто

$$n_i = \frac{n}{k}, \tag{12.37}$$

де $k$ — кількість виділених типових груп.

Отримана величина дає чисельність відбору одиниць ознаки кожної типової групи.

Можливий і ще один варіант відбору одиниць у вибіркову сукупність (назвемо його найвигіднішим або умовно оптимальним), що враховує також і ступінь варіації ознаки в різних групах генеральної сукупності. При відборі з урахуванням рівня коливання ознаки процент вибірки з кожної типової групи повинен бути пропорційний середньому квадратичному відхиленні з цієї групи ($\sigma_i$).

Розрахунок чисельності одиниць, відібраних у вибірку за кожною групою ($n_i$), здійснюється за формулою:

$$n_i = \frac{nN_i\sigma_i}{\sum N_i\sigma_i} \text{ — для середньої;} \tag{12.38}$$

$$n_i = \frac{nN_i\sqrt{w_i(1-w_i)}}{\sum N_i\sqrt{w_i(1-w_i)}} \text{ — для частки,} \tag{12.39}$$

де $\sigma_i$ і $w_i$ — середнє квадратичне відхилення і частка ознаки досліджуваної ознаки в $i$-й групі.

При такому принципі відбору з групи з більшим рівнем коливання ознаки буде відібрано більше одиниць, що в свою чергу призводить до відповідного зменшення можливої випадкової помилки у визначенні групової середньої. Використання принципу оптимального відбору одиниць у вибіркову сукупність на практиці обмежується тим, що дослідник часто не має даних у генеральній сукупності.

При типовому відборі у вибірку потрапляють, таким чином, представники з кожної типової групи, тому ймовірність отримати більшу точність вибірки тут більше, ніж при організації інших видів відбору.

Для розрахунку середньої помилки вибірки всієї сукупності при власне випадковому методі відбору та меж генеральної середньої при цьому методі відбору потрібно знати загальну вибіркову середню і загальну дисперсію вибіркової сукупності.

На основі типової вибірки можна безпосередньо визначити середні за групами ($\tilde{x}_i$). Величина ж загальної середньої для всієї сукупності $\tilde{x}_0$ розраховується за формулою середньої арифметичної зваженої з вибіркових середніх за групами, де в якості ваги використовують обсяг відповідної типової групи в генеральній сукупності $N_i$, тобто

$$\tilde{x}_0 = \frac{\sum \tilde{x}_i N_i}{\sum N_i}. \tag{12.40}$$

Розрахунок загальної вибіркової середньої з груп групових вибіркових середніх при пропорційному відборі обчислюється шляхом зважування вибіркових середніх за чисельністю відібраних груп:

$$\tilde{x}_0 = \frac{\sum \tilde{x}_i n_i}{\sum n_i} = \frac{\sum \tilde{x}_i N_i}{\sum N_i}. \tag{12.41}$$

З наведеної формули видно, що точність вибірки буде залежати від ступеня точності визначення групових середніх, тобто від того, наскільки відібрані з типової групи одиниці репрезентують середню кожної окремої групи. Отже, випадкова помилка типової вибірки залежить від величини внутрішньогрупової дисперсії. Тому при визначенні помилки вибірки для середньої величини ознаки в сукупності береться не загальна дисперсія, а середня із внутрішньогрупових дисперсій, яку позначають $\overline{\sigma_i^2}$.

Формули для визначення стандартної помилки типової вибірки для різних прийомів відбору наведено в табл. 12.8.

Покажемо специфіку розрахунку помилки вибірки при різних варіантах типового відбору.

Приклад. На заводі працює 1 000 робітників. Вся чисельність робітників ділиться на 3 типові групи чисельністю 250, 350 і 400 робітників. Для визначення середньогодинної заробітної плати була проведена 10 %-на типова вибірка з відбором одиниць пропорційно чисельності одиниць типових груп. Відбір одиниць усередині типових груп провадиться випадковим безповоротним методом.

Таблиця 12.8

**Формула розрахунку $\mu$ при типовому відборі**

| Відбір | Відбір | | | |
|---|---|---|---|---|
| | повторний | | безповторний | |
| | для середньої | для частки | для середньої | для частки |
| Непропорційний обсягу груп | $\frac{1}{N}\sqrt{\frac{\sigma_i^2}{n_i}N_i^2}$ | $\frac{1}{N}\sqrt{\frac{w_i(1-w_i)}{n_i}N_i^2}$ | $\frac{1}{N}\sqrt{\frac{\sigma_i^2}{n_i}N_i^2\left(1-\frac{n_i}{N_i}\right)}$ | $\frac{1}{N}\sqrt{\frac{w_i(1-w_i)}{n_i}N_i^2\left(1-\frac{n_i}{N_i}\right)}$ |
| Пропорційний обсягу груп | $\sqrt{\frac{\overline{\sigma_i}^2}{n}}$ | $\sqrt{\frac{\overline{w(1-w)}}{n}}$ | $\sqrt{\frac{\overline{\sigma_i}^2}{n}\left(1-\frac{n}{N}\right)}$ | $\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$ |
| Пропорційний коливанням ознаки в групах до обсягу груп (є найвигіднішим) | $\frac{1}{N}\cdot\frac{\sum\sigma_i N_i}{\sqrt{n}}$ | $\frac{1}{N}\cdot\frac{\sum w_i(1-w_i)N_i}{\sqrt{n_i}}$ | $\frac{1}{N}\cdot\frac{\sum\sigma_i N_i}{\sqrt{n}}\sqrt{\left(1-\frac{n}{N}\right)}$ | $\frac{1}{N}\cdot\frac{\sum w_i(1-w_i)N_i}{\sqrt{n_i}}\sqrt{\left(1-\frac{n}{N}\right)}$ |

де $\overline{\sigma_i}^2$ — середня з вибіркових дисперсій типових групп;

$\overline{w(1-w)}$ — середня із добутку частот на доповнення їх до одиниці;

$\sigma_i^2$ — вибіркова дисперсія $i$-ї типової групи;

$\sigma_i$ — середнє квадратичне відхилення у вибірці з $i$-ї типової групи.

В результаті вибірки було отримано такі дані (табл. 12.9).

Таблиця 12.9

| Групи робітників з кваліфікацією ($i$) | Чисельність груп ($N_i$) | Середньогодинна заробітна плата, грн $\tilde{x}_i$ | Дисперсія годинної заробітної плати $\sigma_i^2$ |
|---|---|---|---|
| 1 | 250 | 120 | 105 |
| 2 | 350 | 135 | 176 |
| 3 | 400 | 154 | 78 |
| Всього | 1 000 | | |

Визначимо межі, в яких з ймовірністю 0,954 знаходиться генеральна середня.

Межі генеральної середньої з ймовірністю 0,954 дорівнюватимуть:

$$\bar{x} = \tilde{x} \pm 2\mu_{\bar{x}}. \qquad (12.42)$$

1. *При відборі одиниць пропорційно чисельності груп* чисельність відібраних одиниць за кожною типовою групою визначається за формулою:

$$n_i = n\frac{N_i}{N}. \qquad (12.43)$$

Розрахуємо обсяг вибірки в кожній типовій групі за умови, що чисельність вибіркової сукупності становить 100 робітників. Так, для першої типової групи маємо при заданому обсязі всієї вибірки, що дорівнює 2 000 одиниць:

$$n_1 = 100 \cdot \frac{250}{1\,000} = 25 \text{ одиниць;}$$

для другої типової групи:

$$n_2 = 100 \cdot \frac{350}{1\,000} = 35 \text{ одиниць;}$$

для третьої типової групи:

$$n_3 = 100 \cdot \frac{400}{1\,000} = 40 \text{ одиниць.}$$

Розрахуємо:

1) середню заробітну плату за даними типового відбору:

$$\tilde{x} = \frac{\sum \tilde{x}_i n_i}{\sum n_i} = \frac{120 \cdot 25 + 135 \cdot 35 + 154 \cdot 40}{100} = 138{,}85 \text{ грн;}$$

2) середню із внутрішньогрупових дисперсій:

$$\overline{\sigma_i^2} = \frac{\sum \sigma_i^2 n_i}{\sum n_i} = \frac{105 \cdot 25 + 176 \cdot 35 + 78 \cdot 40}{100} = 119{,}05;$$

3) середню помилку вибіркової середньої типової вибірки:

$$\mu_{\bar{x}} = \sqrt{\frac{\overline{\sigma_i}^2}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{119{,}05}{100}\left(1 - \frac{100}{1\,000}\right)} = 1{,}1 \text{ грн};$$

4) граничну помилку вибіркової середньої з ймовірністю 0,954:

$$\Delta = t\mu = 2 \cdot 1{,}1 = 2{,}2.$$

Таким чином, з ймовірністю 0,954 можна стверджувати, що середня заробітна плата робітників підприємства в генеральній сукупності буде знаходитися в наступних межах: $\bar{x} = 138{,}85 \pm 2{,}2$, тобто $136{,}65 \leq \bar{x} \leq 141{,}05$.

2. *При відборі одиниць пропорційно до чисельності груп*. Визначимо середню помилку вибірки за такими даними. З генеральної сукупності 20 000 одиниць була проведена вибірка 4 000 одиниць (10 % від їх загальної чисельності) типовим шляхом. Всі одиниці генеральної сукупності на 5 однотипних груп чисельністю 4 000, 3 500, 1 000, 7 500 і 4 000 одиниць. Відбір одиниць усередині типових груп проводиться випадковим безповторним методом.

При відборі одиниць непропорційно до обсягу груп з кожної групи відбирають однакову кількість одиниць, тобто

$$n_i = \frac{n}{k}. \tag{12.44}$$

Число спостережень по кожній групі дорівнюватиме загальному обсягу вибірки, поділеної на кількість виділених типових груп, тобто 4000:5 = 800 одиниць. За формулами для середньої помилки вибірки, чисельною якою непропорційна обсягу груп, здійснюємо відповідні розрахунки (див. табл. 12.10).

Число спостережень по кожній групі дорівнюватиме загальному об’єму вибірки, поділеної на кількість виділених типових груп, тобто 4000:5 = 800 одиниць. За формулами для середньої помилки вибірки, чисельною якою непропорційна обсягу груп, здійснює відповідні розрахунки (див. табл. 12.8).

Обчислюємо середню помилку вибірки всієї сукупності із табл. 12.10.

Таблиця 12.10

**Розрахункова таблиця**

| Номер групи $i$ | Чисельність групи $N_i$ | Чисельність вибірки $n_i$ | Вибіркова середня $\tilde{x}$ | Вибіркова дисперсія $\sigma_i^2$ | $\frac{\sigma_i^2}{n_i}$ | $\frac{\sigma_i^2}{n_i}N_i^2$ | $1-\frac{n_i}{N_i}$ | $\frac{\sigma_i^2}{n_i}N_i^2 \times \times\left(1-\frac{n_i}{N_i}\right)$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 4 000 | 800 | 20 | 8 | 0,0100 | 160 000 | 0,80000 | 128 000 |
| 2 | 3 500 | 800 | 22 | 12 | 0,0150 | 183 750 | 0,77143 | 141 750 |
| 3 | 1 000 | 800 | 30 | 20 | 0,0250 | 25 000 | 0,20000 | 5 000 |
| 4 | 7 500 | 800 | 36 | 40 | 0,0500 | 2 812 500 | 0,89333 | 2 512 500 |
| 5 | 4 000 | 800 | 40 | 32 | 0,0400 | 640 000 | 0,80000 | 512 000 |
| Всього | 20 000 | 4 000 | 29,6 | – | – | – | – | 3 299 250 |

$$\mu_{\text{непроп. числ}} = \frac{1}{N}\sqrt{\sum \frac{\sigma_i^2}{n_i}N_i^2\left(1-\frac{n_i}{N_i}\right)} = \frac{1}{20000}\sqrt{3299250} \approx$$

$$\approx 0{,}09.$$

3. *При відборі з урахуванням ступеня варіації ознаки різних груп.* При такому відборі частка відібраних одиниць з кожної типової групи має бути пропорційно не тільки обсягу групи, а й середньому квадратичному відхиленню в цій групі ($\sigma_i$). Число одиниць, що відбираються з кожної групи ($n_i$), визначається за формулою.

$$n_i = \frac{nN_i\sigma_i}{\sum N_i\sigma_i}. \tag{12.45}$$

Вихідні дані візьмемо із попереднього прикладу. Зробивши необхідні розрахунки для нашого прикладу, отримуємо (табл. 12.11).

Обчислюючи середню помилку вибірки за такою формулою з табл. 12.8 та залучаючи необхідні підсумки з табл. 12.10, отримуємо:

$$\mu_{\text{проп. чис. і колив.}} = \frac{1}{N}\cdot\frac{\sum\sigma_i N_i}{\sqrt{n_i}}\sqrt{\left(1-\frac{n}{N}\right)} =$$

$$= \frac{97971{,}81}{20000\cdot\sqrt{4000}}\sqrt{1-\frac{4000}{20000}} \approx 0{,}06.$$

Розрахунок помилки вибіркової частки дещо відрізняється від розрахунку помилки вибіркової середньої. Розглянемо умовний приклад.

Таблиця 12.11

**Розрахункова таблиця**

| Номер групи ($i$) | $N_i$ | $\sigma_i^2$ | $N_i\sigma_i$ | $n_i = \frac{N_i\sigma_i}{\sum N_i\sigma_i}$ | $n_i = \frac{nN_i\sigma_i}{\sum N_i\sigma_i}$ |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 4 | 5 |
| 1 | 4 000 | 2,82843 | 11313,72 | 0,11548 | 461 |
| 2 | 3 500 | 3,46410 | 12124,35 | 0,12375 | 495 |
| 3 | 1 000 | 4,47214 | 4472,14 | 0,04565 | 183 |
| 4 | 7 500 | 6,32456 | 47434,20 | 0,48416 | 1 937 |
| 5 | 4 000 | 5,65685 | 22627,40 | 0,23096 | 924 |
| Всього | 20 000 | – | 97971,81 | 1,00 | 4 000 |

Для виявлення частки простоїв через несвоєчасне надходження заготовок була проведена фотографія робочого дня 10 % робітників чотирьох різних цехів. Відбір робітників усередині цехів проводився шляхом механічного відбору. Результати вибірки представлені у табл. 12.12.

Таблиця 12.12

| Цех | Число робітників у вибірці | Питома вага простоїв через несвоєчасне надходження заготовок |
|---|---|---|
| № 1 | 22 | 6 |
| № 2 | 37 | 10 |
| № 3 | 15 | 14 |
| № 4 | 26 | 2 |

З ймовірністю 0,954 потрібно визначити межі, в яких знаходиться частка простоїв на заводі через несвоєчасне надходження заготовок.

Частка простоїв через несвоєчасне надходження заготовок у вибірці розраховується за формулою:

$$\bar{w} = \frac{\sum w_i n_i}{\sum n_i};$$

$$\bar{w} = \frac{6 \cdot 22 + 10 \cdot 37 + 14 \cdot 15 + 2 \cdot 31}{22 + 37 + 15 + 26} = \frac{764}{100} = 7{,}64 \text{ \%}.$$

Розрахуємо дисперсії для кожної з типових груп:

$$\text{I} \quad \sigma_1^2 = w_1(1 - w_1) = 6 \cdot 94 = 564;$$

$$\text{II} \quad \sigma_2^2 = w_2(1-w_2) = 10 \cdot 90 = 900;$$

$$\text{III} \quad \sigma_3^2 = w_3(1-w_3) = 14 \cdot 86 = 1204;$$

$$\text{IV} \quad \sigma_4^2 = w_4(1-w_4) = 2 \cdot 98 = 196;$$

Середня із внутрішньогрупових дисперсій визначається за формулою:

$$\overline{w(1-w)} = \frac{564 \cdot 22 + 900 \cdot 37 + 1204 \cdot 15 + 196 \cdot 26}{100} = 688{,}64.$$

Визначимо середню помилку вибіркової частки:

$$\mu_w = \sqrt{\frac{688{,}64}{100}\left(1 - \frac{100}{1000}\right)} = 2{,}49\ \%.$$

Розрахуємо граничну помилку вибірки для частки із ймовірністю 0,954 ($t$=2):

$$\Delta = t\mu = 2 \cdot 2{,}49 = 5{,}0.$$

З ймовірністю 0,954 можна стверджувати, що частка простоїв робітників через несвоєчасне надходження заготовок знаходиться в межах 2,64 % $\leq p \leq$ 12,64 %.

При типовому відборі середня величина помилки обчислюється не на основі загальної дисперсії $\sigma^2$, а середньої величини групових дисперсій $\overline{\sigma_j^2}$, тобто коливання ознаки всередині тих типових груп, із яких здійснювався відбір одиниць. Так як згідно з теорією складання дисперсій загальна дисперсія ознаки в генеральній сукупності дорівнює сумі середньої внутрішньогрупових дисперсій і дисперсії групових середніх величина стандартної помилки типової вибірки завжди буде меншою, ніж при простій випадковій вибірці. Відмінності між загальною дисперсією і середньою з внутрішньогрупових дисперсій будуть тим більше, чим більше відмінності у величині досліджуваної ознаки, що виникають під впливом ознаки-фактора, покладеного в основу групування (тобто відповідно тим більше величина міжгрупової дисперсії). Що ж досягається при заміні $\sigma^2$ на $\overline{\sigma_i^2}$? Відповідно до правила дисперсій, $\sigma^2 = \overline{\sigma_i^2} + \delta^2$. Якщо розрахунку середньої помилки береться лише середня величина групових дисперсій ($\overline{\sigma_i^2}$), то з аналізу виключається та частина загальної дисперсії, яка представлена дисперсією групових дисперсій ($\delta^2$). При типовому відборі генеральна сукупність поділяється на групи, з яких відбирається певна кількість одиниць. Утворюючи типові групи, дослідник ставить результати свого відбору в залежності від тієї ознаки, за якою ці групи були відібрані. Середні, обчислені для кожної з груп генеральної сукупності, утворюють щось на зразок емпіричної лінії регресії. Зміна цих середніх носить цілком певні риси неухильного зростання (зниження), які привносять у загальну коливання даних значний елемент «систематичності». Коливання групових середніх вже не носить випадковий характер, а значною мірою обумовлюється зміною ознаки-фактора, який дослідник вибрав і поклав в основу поділу генеральної сукупності на групи. Ця коливання середніх, що має велику частку в загальній коливання, позначиться, звичайно, і на величині помилки вибірки, спотворивши її.

Групові середні в цьому випадку розглядається тут як емпірична регресія, а відхилення від неї — як залишкова варіація, яка, очевидно, отримала своє вираження в середній величині внутрішньогрупових дисперсій.

Таким чином, обчислення помилки на основі середньої величини внутрішньогрупових дисперсій дозволяє очистити цю помилку від системних спотворень. Типова вибірка порівняно з простою випадковою вибіркою має переваги за наявності у складі генеральної сукупності груп, які різко різняться між собою, тоді як варіація ознаки всередині групи незначна.

**Розрахунок середньої помилки вибірки при серійному відборі**. При серійному відборі з рівновеликими серіями генеральну сукупність розбивається на однакові за обсягом групи, звані серіями. У випадковому порядку відбираються групи (серії, гнізда), які піддаються суцільному спостереженню. Серійна вибірка перед уже розглянутими видами вибірки має переваги організаційного характеру і широко використовується в тих випадках, де генеральна сукупність складається з певним чином відокремлених груп одиниць. Наприклад, при статистичному контролі якості готової продукції серійна вибірка знаходить широке поширення у випадках застосування тарної упаковки, наприклад, підшипники одного типорозміру упаковуються в коробки зі строго встановленою кількістю в кожній з них. У виробничій діяльності підприємств часто зустрічається серійний відбір з рівними серіями. Як і при типовому відборі, при серійній вибірці одиниці з кожної серії підлягають суцільному обстеженню. Оскільки всі входять у серії одиниці піддаються суцільному спостереженню, середні за серіями визначаються без випадкової помилки. Таким чином, загальна середня $\tilde{x}_0$, що розглядається як оцінка генеральної середньої, буде репрезентувати генеральну середню. Тим самим випадкова помилка середньої по всій вибірці залежить від величини міжсерійної дисперсії $\delta^2$, що характеризує ступінь коливання серійних середніх.

Точність серійної вибірки залежить не від величини загальної дисперсії, як від міжсерійної дисперсії, або, інакше, дисперсії групових середніх.

Як відомо, міжгрупова (міжсерійна) дисперсія обчислюється за формулою:

$$\delta_{\bar{x}}^2 = \frac{\sum(\tilde{x}_i - \bar{x})^2}{r}, \tag{12.46}$$

де $r$ — кількість відібраних серій;

$\tilde{x}_i$ — середня в окремих серіях;

$\bar{x}$ — загальна середня для всієї сукупності.

Міжсерійна (міжгрупова) дисперсія частки визначається за формулою:

$$\delta_p^2 = \frac{\sum(w_i - \bar{w}_i)^2}{r}, \tag{12.47}$$

де $w_i$ — частка ознаки $i$-й серії;

$\bar{w}_i$ — частка ознаки у всієї вибіркової сукупності.

Серійна вибірка може бути повторною та безповторною. Крім того, серії можуть бути рівновеликими та нерівновеликими.

Середня помилка вибірки ($\mu$) при серійному методі відбору з рівновеликими серіями підраховується за наступними формулами (див. табл. 12.13).

Таблиця 12.13

**Формула розрахунку $\mu$ при серійному методі відбору з рівновеликими серіями**

| Спосіб відбору | Для середньої | Для частки |
|---|---|---|
| Повторний | $\mu_{\bar{x}} = \sqrt{\frac{\delta_{\bar{x}}^2}{r}}$ | $\mu_p = \sqrt{\frac{\delta_p^2}{r}}$ |
| Безповторний | $\mu_{\bar{x}} = \sqrt{\frac{\delta_{\bar{x}}^2}{r}\left(\frac{R-r}{R-1}\right)} \approx \sqrt{\frac{\delta_{\bar{x}}^2}{r}\left(1-\frac{r}{R}\right)}$ | $\mu_p = \sqrt{\frac{\delta_p^2}{r}\left(\frac{R-r}{R-1}\right)} \approx \sqrt{\frac{\delta_p^2}{r}\left(1-\frac{r}{R}\right)}$ |

де $R$ — число серій у генеральній сукупності;

$r$ — кількість відібраних серій;

$\delta^2_{\bar{x}}$ — міжсерійна (міжгрупова) дисперсія середніх;

$\delta^2_p$ — міжсерійна (міжгрупова) дисперсія частки.

З наведених формул випливає, що випадкова помилка серійної вибірки повинна бути тим меншою, чим менша серійна дисперсія і чим, отже, неоднорідніші в собі серії.

Наведемо приклад розрахунку. З сукупності 4 000 одиниць проводиться вибірка типовим методом. Вся сукупність ділиться на 40 рівних за величиною серій (100 одиниць). Безповторним способом відібрано 10 серій. Результати вибірки представлені у таблиці:

Таблиця 12.14

| Номери серій | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Середній розмір ознаки ($\tilde{x}_i$) | 62 | 65 | 64 | 66 | 67 | 74 | 78 | 82 | 85 | 96 |
| Групова дисперсія $\left(\sigma^2_i\right)$ | 26,1 | 24,3 | 17,6 | 36,4 | 46,2 | 89,2 | 85,7 | 76,9 | 107,6 | 138,2 |

В и з н а ч и т и середню помилку серійної безповторної вибірки та межі генеральної середньої з ймовірністю 0,954.

Межі генеральної середньої з ймовірністю 0,954 дорівнюють:

$$\bar{x} = \tilde{x} \pm 2\mu_{\bar{x}}.$$

Розрахуємо:

а) загальну середню усієї вибіркової сукупності за серійними середніми:

$$\tilde{x} = \frac{\sum \tilde{x}}{n} = \frac{62+65+64+66+67+74+78+82+85+96}{10} = \frac{739}{10} = 73{,}9;$$

б) міжсерійну (міжгрупову) дисперсію:

$$\bar{\delta}^2_{\bar{x}} = \frac{\sum(\tilde{x}_i - \tilde{x})^2}{n} = \frac{(62-73{,}9)^2 + (65-73{,}9)^2 + \ldots + (96-73{,}9)^2}{10} =$$

$$\frac{1122{,}9}{10} = 112{,}29;$$

в) середню квадратичну (стандартну) помилку серійної безповторної вибірки:

$$\mu_{\bar{x}} = \sqrt{\frac{\delta^2_{\bar{x}}}{r}\left(\frac{R-r}{R-1}\right)} = \sqrt{\frac{112{,}29}{10}\cdot\left(\frac{50-10}{50-1}\right)} = \sqrt{9{,}1665} \approx 3{,}03.$$

Отже, з ймовірністю 0,954 можна стверджувати, що середня в генеральній сукупності перебуватиме в межах $70{,}87 \le \bar{x} \le 76{,}93$.

У цьому прикладі видно, що з дуже великою ймовірністю можна очікувати, що розбіжність у вазі (гранична помилка) буде невеликим, тому що проводиться відбір серій. Якщо в серіях досить велика кількість виробів мають приблизно однакові розміри, то міжгрупова дисперсія, а отже, і гранична помилка вибірки порівняно невеликі.

Наведемо ще один приклад, що ілюструє визначення помилки при серійному відборі. Для встановлення середніх розмірів доходу було відібрано на певній території у порядку випадкового відбору 5 серій, або «гнізд» підприємств по 5 підприємств у кожній серії. Результати вибірки наведено у табл. 12.1.

Таблиця 12.15

| Серії | Доход (млн грн) | В середньому на 1 підприємство |
|---|---|---|
| 1-а | 440, 430, 470, 420, 460 | 436 |
| 2-а | 410, 410, 380, 390, 400 | 398 |
| 3-а | 340, 350, 330, 360, 350 | 346 |
| 4-а | 290, 310, 320, 280, 270 | 294 |
| 5-а | 270, 280, 260, 240, 230 | 256 |

Середня помилка за таких умов завдання може бути визначена (якщо невідомо кількість серій у генеральній сукупності) за формулою:

$$\mu_{\bar{x}} = \sqrt{\frac{\delta_{\bar{x}}^2}{r}}.$$

У розглянутому випадку дисперсія серійних середніх величин дорівнює:

Таблиця 12.16

| $\bar{x}_c$ | $(\bar{x}_c - \bar{x})$ | $(\bar{x}_c - \bar{x})^2$ |
|---|---|---|
| 436 | +90 | 8 100 |
| 398 | +52 | 2 704 |
| 346 | 0 | 0 |
| 294 | –52 | 2 704 |
| 256 | –90 | 8 100 |
| $\bar{x} = 346$ | | 21 608 |

$$\sigma_{\bar{x}_c}^2 = \frac{21608}{5} = 4321{,}6.$$

$$\mu = \sqrt{\frac{4\,321{,}6}{5}} \approx 29{,}4.$$

Серійний, або «гніздовий», відбір, як зазначалося, має перевагу перед випадковим відбором одиниць першого порядку.

**Розрахунок середньої помилки вибірки при комбінованій вибірці**. Вище була розглянута так звана одноступенева вибірка, коли у випадково відібраних групах генеральної сукупності суцільному обстеженню піддавалися одиниці, що входять в них. У практиці організації вибіркового спостереження крім розглянутих вище способів відбору застосовується комбінована (багатоступінчата) вибірка. *Комбінована вибірка* (рівновеликі серії) передбачає використання кількох способів вибірки. Можна, наприклад, провести вибірку в два етапи: на першому етапі вся сукупність ділиться на однакові за обсягами групи (серії), а на другому етапі з відібраних серій у випадковому порядку відбирається певна кількість одиниць, що підлягають безпосередньому спостереженню. Це означає, що в даному випадку серійний відбір поєднується з індивідуальним відбором. Помилка такої вибірки залежатиме від помилки серійного відбору і від помилки індивідуального відбору, тобто помилка такої вибірки визначається ступінчастістю відбору. За такого відбору помилки, обчислювані кожному ступені, накладаються друг на друга. Таким чином, при комбінованій вибірці точність отриманих результатів порівняно з одноступеневим відбором значно нижче. Однак це загальне збільшення похибки компенсується вигідністю такого відбору в організаційному відношенні. У сільськогосподарській статистиці, наприклад, такий відбір дає очевидні вигоди в порівнянні з випадковим відбором такого ж числа окремих господарств.

Комбінована вибірка може проводитися в порядку повторного та безповторного відбору.

Середня помилка вибірки різних комбінаціях її способів обчислюється по-різному залежно від ступінчастості вибірки. Виділяють кілька ступенів вибірки. Відбір називається *одноступінчастим*, якщо з генеральної вибірки якимось способом відібрані одиниці, які піддаються спостереженню і з них робляться висновки. На цій стадії об’єкти репрезентації та одиниці спостереження збігатимуться. *Багатоступінчаста вибірка* передбачає вилучення з генеральної сукупності спочатку окремих груп (серій), а потім груп менших за обсягом, і так доти, доки не будуть відібрані ті одиниці, які піддаються безпосередньому спостереженню.

При комбінованій вибірці з рівновеликими серіями середня помилка вибірки ( $\mu$ ) підраховується за такими формулами (табл. 12.17).

Таблиця 12.17

**Формула розрахунку $\mu$ при комбінованій вибірці з рівновеликими серіями**

| Спосіб відбору | Для середньої | Для частки |
|---|---|---|
| Повторний | $\mu_{\tilde{x}} = \sqrt{\frac{\overline{\sigma}^2}{n} + \frac{\delta_{\tilde{x}}^2}{r}}$ | $\mu_p = \sqrt{\frac{\delta_p^2}{r}}$ |
| Безповторний | $\mu_{\bar{x}} = \sqrt{\frac{\overline{\sigma}^2}{n}\left(1 - \frac{n}{N_r}\right) + \frac{\delta_{\tilde{x}}^2}{r}\left(1 - \frac{r}{R}\right)}$ | $\mu_p = \sqrt{\frac{w_i(1 - w_i)}{n}\left(1 - \frac{n}{n_r}\right) + \frac{\delta_p^2}{r}\left(1 - \frac{r}{R}\right)}$ |

де $n_r$ — загальна кількість одиниць, що потрапили у вибірку при відборі серій; визначається за формулою $n_r = r \cdot \frac{N}{R}$;

$n$ — кількість одиниць, що потрапили у вибірку із серій.

Для ілюстрації комбінованого відбору розглянь наступний приклад. Для визначення середнього розміру грошового доходу підприємств було відібрано 5 серій, або «гнізд», а потім із кожної серії у випадковому порядку було відібрано по 2 підприємства з наступними розмірами грошового доходу (млн грн):

| | | |
|---|---|---|
| Із 1-ї серії | 430 | 420 |
| » 2-ї » | 410 | 380 |
| » 3-ї » | 350 | 330 |
| » 4-ї » | 290 | 270 |
| » 5-ї » | 270 | 260 |

Як визначається середня помилка при випадковому відборі серій, було показано вище. Залишається лише вирахувати середню помилку для відібраних із серій підприємств. Оскільки ці підприємства відбиралися з різних серій, то помилка такої вибірки повинна, очевидно, визначатися як при типовій (або районованій) вибірці через середню величину внутрішньогрупових дисперсій.

Для розрахунку середньої величини внутрішньогрупових дисперсій складе розрахункову таблицю.

Таблиця 12.18

| $x_1$ | $(x_1-\bar{x}_1)$ | $(x_1-\bar{x}_1)^2$ | $x_2$ | $(x_2-\bar{x}_2)$ | $(x_2-\bar{x}_2)^2$ | $x_3$ | $(x_3-\bar{x}_3)$ | $(x_3-\bar{x}_3)^2$ |
|---|---|---|---|---|---|---|---|---|
| 430 | 5 | 25 | 410 | 15 | 225 | 350 | 10 | 100 |
| 420 | –5 | 25 | 380 | –15 | 225 | 330 | –10 | 100 |
| $\bar{x}_1=425$ | | 50 | $\bar{x}_2=395$ | | 450 | $\bar{x}_3=340$ | | 100 |
| $\sigma_1^2=\frac{50}{2}=25$ | | | $\sigma_2^2=\frac{450}{2}=225$ | | | $\sigma_2^2=\frac{200}{2}=100$ | | |

Продовження табл. 12.18

| $x_4$ | $(x_4-\bar{x}_4)$ | $(x_4-\bar{x}_4)^2$ | $x_5$ | $(x_5-\bar{x}_5)$ | $(x_5-\bar{x}_5)^2$ |
|---|---|---|---|---|---|
| 290 | 10 | 100 | 270 | 5 | 25 |
| 270 | –10 | 100 | 260 | –5 | 25 |
| $\bar{x}_4=280$ | | 200 | $\bar{x}_5=265$ | | 50 |
| $\sigma_4^2=\frac{200}{2}=100$ | | | $\sigma_5^2=\frac{50}{2}=25$ | | |

$$\overline{\sigma_i^2}=\frac{25+225+100+100+25}{5}=95.$$

Таким чином, величина середньої помилки комбінованого відбору може бути визначена за формулою:

$$\mu_{\bar{x}}=\sqrt{\frac{\overline{\sigma}^2}{n}+\frac{\delta_{\bar{x}}^2}{r}}\,. \qquad (12.48)$$

Підставляючи значення розрахованих показників, отримуємо:

$$\mu=\sqrt{\frac{4321{,}6}{5}+\frac{95}{5}}\approx 29{,}7.$$

Розглянемо ще один приклад. З генеральної сукупності 100 000 одиниць провадиться комбінована вибірка. Вся сукупність розбивається на 100 рівних за обсягом серій. Зроблено безповторну вибірку 50 % серій та з кожної серії по 25 % одиниць. При відборі отримані такі результати: середня із серійних дисперсій дорівнює 14, а міжсерійна дисперсія — 8. В и з н а ч и т и середню помилку вибірки.

Запишемо вихідні дані:

$$N=100\,000;\ R=100;\ r=50;\ \overline{\sigma_i^2}=14;\ \delta_i^2=8;$$

$$n=12500;$$

Визначимо загальну кількість одиниць, що потрапили у вибірку:

$$n_r = 50 \cdot \frac{100\,000}{100} = 50\,000.$$

Визначаємо середню помилку вибірки:

$$\mu = \sqrt{\frac{14}{12\,500}\left(1-\frac{12\,500}{50\,000}\right)+\frac{8}{50}\left(1-\frac{50}{100}\right)} = 0{,}284.$$

Характерною особливістю багатоступінчастої вибірки є те, що на всіх ступенях, крім останньої, здійснюється тільки відбір, а спостереження одиниць здійснюється тільки на останній стадії. У зв'язку з цим на різних стадіях спостереження одиниці застосовуються різні одиниці відбору. Наприклад, при поточному спостереженні використання часу робочого часу одиницею спостереження є робочий. Формування вибіркової сукупності проводиться спочатку підприємств, а потім з робітників, що працюють на підприємстві.

При проведенні комбінованого відбору число ступенів вибірки може бути і більше трьох. Однак потрібно уникати надмірної ступінчастості проведення вибірки, ретельно продумуючи організацію вибіркового обстеження.

Якщо число ступенів відбору більше двох, то середня помилка вибірки визначається з виразу:

$$\mu = \sqrt{\mu_1^2 + \frac{\mu_2^2}{n_1} + \frac{\mu_3^2}{n_1 n_2} + \frac{\mu_4^2}{n_1 n_2 n_3} + \ldots}, \tag{12.49}$$

де $\mu_1, \mu_2\ \mu_3 \ldots$ — середні помилки вибірки на окремих ступенях;

$n_1, n_2\ n_3 \ldots$ — чисельності вибірок на відповідних його ступенях.

Наведена формула середньої помилки вибірки використовується для рівновеликих груп, із яких проводиться комбінована вибірка. Насправді таке становище зустрічається нечасто. Формула середньої вибірки для нерівновеликих на різних ступенях відбору занадто громіздка, а дає практично ті ж результати, що і наведена формула.

**Розрахунок середньої помилки вибірки при механічному відборі**. Однією з найпоширеніших форм вибіркового спостереження у статистиці є механічний відбір. Це тим, що з усіх прийомів відбору він є найпростішим. При механічному відборі включення одиниць у вибіркову сукупність (з основи вибірки) проводиться через якийсь рівний інтервал (крок відбору).

Часто до механічного відбору вдаються з метою підвищення репрезентативності вибіркового спостереження. У цьому випадку списки одиниць генеральної сукупності складаються в іншій формі, у формі ранжованого ряду, тобто одиниці заносяться в списки в порядку зростання або зменшення будь-якої ознаки, тісно пов'язаного з досліджуваними ознаками і до певної міри визначає їх величину. Наприклад, щодо кількості дітей у сім'ях у статистиці використовують механічний відбір зі списків, складених у порядку зростання кількості дітей у сім'ї. Механічним відбором користуються й у випадках, коли можна заздалегідь скласти список одиниць генеральної сукупності.

Під час проведення механічного відбору необхідно встановити розмір інтервалу («кроку» відбору). Під кроком відбору розуміється той проміжок, через який відбираються дані з генеральної сукупності. Інтервал відбору дорівнює зворотному значенню частки вибірки. Наприклад, з генеральної сукупності обсягом 1 000 одиниць обстеженню підлягають 50 одиниць (5% вибірка). Це означає, що з кожних 20 одиниць підлягає обстеженню одна одиниця (1000:50). Вибір початку відліку може здійснюватися шляхом випадкового відбору. Наприклад, на підставі жеребкування номер початкової одиниці дорівнює 7. Тоді у вибірку потраплять одиниці, що стоять під номерами 7, 27, 47, 67, …, 967, 987.

Якщо одиниці генеральної сукупності попередньо не ранжуються і передбачається, що розташування одиниць у ній є випадковим по відношенню до досліджуваних ознак, то механічний відбір можна розглядати як різновид простої випадкової безповторної помилки. Таким чином, величина випадкової помилки такого відбору визначається за такою формулою:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}. \qquad (12.50)$$

Якщо вибірка береться з нескінченної сукупності, то відношення $\frac{n}{N}$ дуже мале, к і множник $\left(1 - \frac{n}{N}\right)$ мало відрізняється від 1. Тому для дуже великих генеральних сукупностей величина помилки випадкової механічної вибірки може бути визначена за формулою:

$$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}}. \qquad (12.51)$$

З метою забезпечення пропорційного представництва у вибірці окремих груп генеральну сукупність подають у вигляді ранжованого ряду. У цьому випадку такий відбір можна уподібнити типовій вибірці з пропорційним розміщенням. Генеральна сукупність при цьому способі відбору як би розбивається на однорідних груп рівного обсягу і з кожної групи вибирається одна одиниця. Випадкову помилку вибірки при механічному відборі потрібно, очевидно, обчислювати як і, як і за типовому відборі, тобто виходячи з середньої величини внутрішньогрупових дисперсій. Однак визначення помилки за формулами типового відбору тут не цілком правомірно через відсутність даних про варіацію ознаки по групах (з групи відбирається лише один представник). У зв'язку з цим величина помилки механічного відбору з ранжованої сукупності приблизно оцінюється за формулами простої випадкової вибірки.

Механічний відбір поряд з перевагами має і ряд недоліків. Точність проведення механічного відбору значною мірою визначається правильністю вибору початкової одиниці. Невдалий вибір початкової одиниці може призвести до серйозних помилок усунення при механічному відборі. Справа в тому, що нерідко механічний відбір проводять для варіаційних рядів, що містять або певні тренди, або коливання. Крім того, при розміщенні одиниць генеральної сукупності в порядку зростання ознаки відбір перших може призвести до применшення помилки, а останніх — до її перебільшення. Помилка при виборі початкової одиниці не забезпечить пропорційне представництво у вибірці окремих груп, що неминуче знижує точність отриманих результатів.

## 12.6. Мала вибірка

Як зазначалося, для певного способу відбору одиниць величина стандартної помилки залежить від обсягу вибірки та ступеня коливання досліджуваної ознаки в генеральній сукупності. Причому, що менше обсяг вибірки, то вище величина стандартної помилки. Відомо, що для встановлення ймовірності граничної помилки вибірки використовують таблицю значень інтеграла ймовірності, яка передбачає наявність досить численної вибірки. Це пояснюється тим, що ця таблиця належить до нормальному розподілу ймовірностей, до якого наближається біномінальний їх розподіл зі збільшенням числа випробувань. Прийнято вважати, що таблицею можна користуватися за умови, якщо число одиниць вибірки становить більше 20 ($n$>20). Проте в деяких випадках чисельність вибіркової сукупності не завжди досягає цього значення. Іноді на практиці доводиться мати справу з «малими вибірками», за результатами якої необхідно зробити певні висновки про генеральну сукупність. Особливо ця необхідність виникає при перевірці якості продукції, пов'язаного зі знищенням одиниці продукції, що перевіряється.

Якщо вибіркова сукупність складається з невеликої кількості одиниць, то таку вибірку називають *малою вибіркою*. Часто число одиниць у малій вибірці не перевищує 20, а іноді складає всього 5–6 одиниць. До безумовно малих відносяться вибірки обсягом менш як двадцять одиниць[1]. У вибірках невеликого обсягу величина вибіркової дисперсії, яка використовується в якості оцінки дисперсії генеральної сукупності, може бути значною мірою схильна до дії випадкових факторів. Для характеристики генеральної сукупності за результатами малого числа спостережень формули для звичайної (великої) вибірки, засновані на нормальному розподілі ймовірностей, дають значні неточності. Тому у вибірках невеликого обсягу оцінка результатів вибіркового спостереження проводиться шляхом «виправлення» вибіркового середнього квадратичного відхилення та використання закону розподілу ймовірностей Стьюдента.

Середнє квадратичне відхилення ознаки у вибірковій сукупності у великих вибірках обчислюється за тими ж формулами, що і середнє квадратичне відхилення ознаки в генеральній сукупності, тобто:

$$\sigma = \sqrt{\frac{\sum (x - \tilde{x})^2}{n}}. \qquad (12.52)$$

Раніше наводилося співвідношення між генеральною та вибірковою дисперсією. Вказувалося, що при досить великій кількість спостережень коефіцієнт $\left(\frac{n}{n-1}\right)$, який потрібно помножити вибіркову

---

[1] Перші роботи в області теорії малої вибірки були зроблені англійським статистом *В. С. Госсетом* в 1908 (псевдонім Стьюдент) і продовжені в дослідженнях *Фішера*.

дисперсію, щоб отримати генеральної частки, не відіграє істотного значення. Але коли обсяг вибірки невеликий цей коефіцієнт потрібно брати до уваги.

Для малих вибірок сума квадратів відхилень ділиться не так на $n$, а на $n-1$. В цьому випадку вибіркове середнє квадратичне відхилення малої вибірки обчислюється за формулою:

$$\sigma_{\text{м. в}} = \sqrt{\frac{\sum (x-\tilde{x})^2}{n-1}}, \tag{12.53}$$

де $n-1$ — число степенів свободи варіації.

*Число ступенів свободи* варіації представляє собою кількість варіантів, що можуть приймати довільні значення, що не змінюють величини середньої. Наприклад, є дані трьох спостережень: $x_1$=10, $x_2$=12, $x_3$=17. Звідси $\bar{x} = \frac{10+12+17}{3} = 13$. При трьох спостереженнях вільно є дві величини, які вільно коливаються, третю можна визначити на підставі двох відомих і середньої. Так, якщо відомі $x_1$ і $x_3$, можна визначити $x_2$: $n\bar{x} - (x_1 + x_3)$. Використовуючи дані приклади, розрахуємо $x_2$: $3 \cdot 13 - (10 + 17) = 12$.

Таким чином, середнє квадратичне відхилення малої вибірки ($\sigma_{\text{м.в}}$) відрізняється від середнього квадратичного відхилення ($\sigma$) тим, що суму квадратів відхилень значень ознаки від вибіркової середньої ділять не на $n$, а на $n-1$. Знаючи вибіркове середнє квадратичне відхилення $\sigma$ можна шляхом його «виправлення» обчислити вибіркове середнє квадратичне відхилення малої вибірки $\sigma_{\text{м.в}}$ за формулою:

$$\sigma_{\text{м.в}} = \sigma\sqrt{\frac{n}{n-1}}. \tag{12.54}$$

*Середня помилка малої вибірки* ($\mu$) обчислюється за такою формулою:

$$\mu_{\text{м. в}} = \frac{\sigma_{\text{м. в}}}{\sqrt{n}} = \sqrt{\frac{\sum (x-\tilde{x})^2}{n(n-1)}}, \tag{12.55}$$

або шляхом використання невиправленого вибіркового середнього квадратичного відхилення

$$\mu_{\text{м. в}} = \frac{\sigma}{\sqrt{n-1}}. \tag{12.56}$$

*Гранична помилка малої вибірки* обчислюється за наступною формулою:

$$\Delta_{\text{м.в}} = t \cdot \mu_{\text{м.в}}. \tag{12.57}$$

Але тут величина $t$ інакше пов’язана з ймовірнісною оцінкою, ніж за нормальної вибірки. Розглянемо це докладніше.

Точність оцінки параметрів малої вибірки відрізняється від оцінки у великий вибірки тим, що при невеликому числі спостережень розподіл ймовірностей для середньої переважно залежить від кількості відібраних одиниць. Англійський математик-статист В. Госсен, який писав під псевдонімом «Стьюдент», в 1908 р. опублікував дослідження, в якому обґрунтував, що при невеликій кількості спостережень середнє квадратичне відхилення малої вибірки істотно відрізняється від середнього квадратичного відхилення генеральної сукупності. Стьюдент довів, що оцінка розбіжності між вибірковою середньою малою вибіркою та генеральною середньою має особливий закон розподілу. Займаючись питанням про ймовірної вибіркової середньої при невеликому числі спостережень, Стьюдент дійшов висновку, що в умовах малих вибірок потрібно розглядати розподіл не самих середніх, а величин їх нормованих відхилень від середньої генеральної сукупності, тобто значень *t*.

Нормоване відхилення або стандартна різниця малої вибірки (*t*) визначається аналогічно тому, як це виходить у здійснюється у звичайній вибірці:

$$t = \frac{\tilde{x} - \bar{x}}{\mu_{\text{м.в}}} = \frac{\Delta_{\text{м.в}}}{\mu_{\text{м.в}}}. \tag{12.58}$$

Знаменник наведеної формули представляє собою міру випадкових коливань вибіркової середньої в малій вибірці, тобто $\mu = \frac{\sigma_{м.в}}{\sqrt{n}}$. Розглянувши величину $t = \frac{\tilde{x} - \bar{x}}{\mu_{м.в}}$, де Стьюдент в 1908 р. знайшов закон розподілу *t*, який називається розподілом Стьюдента:

$$P_{(t)} = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)} \cdot \Gamma\left(\frac{n-1}{2}\right)} \cdot \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} = S(t), \tag{12.59}$$

де $P(t) = S(t)$ — ймовірності того, що стандартна різниця між вибірковою і генеральною середньою має величину *t*;

$\Gamma\left(\frac{n}{2}\right)$ і $\Gamma\left(\frac{n-1}{2}\right)$ — гамма-функції, які можна розглядати як узагальнення факторіалу натурального числа.

Закон розподілу величини $t$ носить назву «Закон розподілу Стьюдента».

Теоретичне нормоване відхилення для малих вибірок дістало назву критерію $t$-Стьюдента. Зі збільшенням обсягу вибірки розподіл Стьюдента наближається до нормального розподілу:

$$P(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}. \tag{12.60}$$

Відповідно до розподілу Стьюдента ймовірність того, що гранична помилка не перевищить $t$-кратну середню величину в малих вибірках, залежить і від величини $t$, і від чисельності вибірки. На основі встановленої закономірності розподілу помилок малих вибірок складено спеціальні таблиці, в яких наведено значення критерію $t$-Стьюдента. Складені таблиці розподілу Стьюдента полегшують застосування $t$ на практиці. Наведемо витримку з таблиць розподілу Стьюдента (табл. 12.19).

Таблиця 12.19

**Витяг з таблиці ймовірностей Стьюдента**
(ймовірності помножені на 1000)

| $t$ \ $n$ | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 15 | 20 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,5 | 348 | 356 | 362 | 366 | 368 | 370 | 372 | 376 | 378 | 383 |
| 1,0 | 608 | 626 | 636 | 644 | 650 | 654 | 656 | 666 | 670 | 683 |
| 1,5 | 770 | 792 | 806 | 816 | 832 | 828 | 832 | 846 | 850 | 865 |
| 2,0 | 860 | 884 | 908 | 908 | 914 | 920 | 924 | 936 | 940 | 954 |
| 2,5 | 933 | 946 | 955 | 959 | 963 | 966 | 968 | 975 | 978 | 988 |
| 3,0 | 942 | 960 | 970 | 976 | 980 | 938 | 984 | 992 | 992 | 997 |

Порівняння величини $t$ з таблиці нормального розподілу з коефіцієнтом $t$-Стьюдента показує, при збільшення $n$ це розподіл прагне до нормального, а при $n = 20$ вже мало чим відрізняється від нормального. Так, при $t$=1 імовірність дорівнює в звичайних вибірках 0,683, у малих — 0,670, при $t = 2$ відповідно — 0,954 і 0,940, при $t$=3 — 0,997 і 0,992. При великій кількості спостережень таблиця значень $t$ Стьюдента при ймовірностях 0,95 і 0,99 збігається з таблицею значень інтеграла ймовірностей. Так, виходячи з таблиці нормального розподілу при $t = 2$ ймовірність того, що гранична помилка вибірки не перевищить дворазову помилку, становить 0,95 (точніше, 0,9545). Це можна інтерпретувати інакше. Гранична помилка випадкової вибірки виявиться невірною при ймовірності, що дорівнює 0,05 (1–0,95). У таблиці Стьюдента при великій кількості ступенів свободи (60) і ймовірності 0,05 вказано значення $t$=2. Чим менше число спостережень, тим більше значення $t$ таблиці Стьюдента.

Таблицю ймовірностей Стьюдента часто викладають у короткій зручній для практичного використання формі: для кожного числа степенів свободи вказують величину відношення Стьюдента $t$, яка із заданою ймовірністю (0,95 і 0,99) не буде перевищувати по абсолютній величині в силу випадкового відбору. У таблиці «Значення $t$ Стьюдента для $p$=0,05 і $p$=0,01» наведено таблицю значень $t$ за різного числа ступенів свободи, які визначаються як $n - 1$ (додаток 4). Таблицю можна прочитати й інакше: при ймовірностях $(1 - p)$, тобто при ймовірностях 0,95 і 0,99, званих довірчими ймовірностями, і при даній чисельності вибірки розбіжність між вибірковою середньою і генеральною середньою (гранична випадкова помилка ) не перевищує зазначене за цих умов значення $t$.

Наприклад, при 6 степенях свободи та рівні ймовірності 0,05 теоретичне значення критерію $t$ дорівнює 2,447. Це означає, що тільки в 5 випадках із 100 значення нормованого відхилення $t$ через випадкові причини може перевищити зазначену величину (2,447), а в інших випадках воно буде меншим або таким, як у таблиці. Іншими словами, табличне значення показує максимальну величину відношення випадкових помилок до їхньої середньої помилки.

Слід враховувати, що розподіл Стьюдента застосовується тільки до оцінки помилок вибірок, взятих з генеральної сукупності з нормальним розподілом ознаки. При малому обсязі вибірки крива розподілу Стьюдента вона значно відхиляється від кривої нормального розподілу, повільніше спускаючись до осі абсцис (рис. 12.1). Це означає, що більші відхилення від криво середньої мають більшу ймовірність, ніж для нормального розподілу. Отже, при малих вибірках виникає більша ймовірність помилок, ніж при великих вибірках.

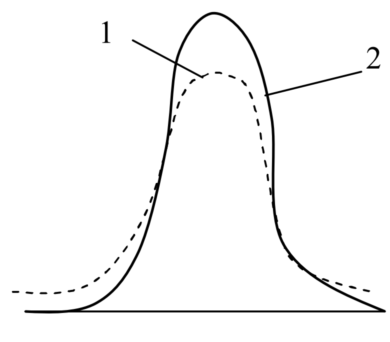



*Рис. 12.1.* 1 — крива розподілу Стьюдента; 2 — крива нормального розподілу.

Можна навести наступний приклад, що ілюструє застосування на практиці таблиці розподілу *t* Стьюдента при «малій вибірці». Для визначення ваги деталей було відібрано в порядку випадкової вибірки 15 виробів з наступною вагою (*г*):

| | | |
|---|---|---|
| 510 | 390 | 470 |
| 515 | 440 | 480 |
| 480 | 450 | 490 |
| 500 | 410 | 410 |
| 460 | 470 | 425 |

Вибіркова середня $\bar{x}$=460.

Для визначення середньої помилки необхідно розрахувати вибіркову дисперсію. Розрахуємо вибіркову дисперсію (див. табл. 12.61). При малому числі спостережень сума квадратів відхилень $\sum(x-\bar{x})^2$ ділиться не на *n*, а на (*n* – 1). Це випливає з того, що середня величина дисперсій всіх можливих за даних умов відбору вибіркових сукупностей пов'язана з дисперсією генеральної сукупності рівністю:

$$\sigma^2 = \frac{\sigma_0^2(n-1)}{n}, \quad \text{або} \quad \sigma_0^2 = \frac{n\sigma^2}{n-1}. \qquad (12.61)$$

Таблиця 12.20

| $x$ | $(x-\bar{x})$ | $(x-\bar{x})^2$ | $x$ | $(x-\bar{x})$ | $(x-\bar{x})^2$ |
|---|---|---|---|---|---|
| 510 | +50 | 2 500 | 410 | –50 | 2 500 |
| 515 | +55 | 3 025 | 470 | +10 | 100 |
| 480 | +20 | 400 | 470 | +10 | 100 |
| 500 | +40 | 1 600 | 480 | +20 | 400 |
| 460 | 0 | 0 | 490 | +30 | 900 |
| 390 | –70 | 4 900 | 410 | –50 | 2 500 |
| 440 | –20 | 400 | 425 | –35 | 1 225 |
| 450 | –10 | 100 | | | |
| | | | $\bar{x}$=460 | | 20 650 |

$$\sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{20\,650}{15-1} = 1475.$$

Розмір середньої помилки визначається за формулою випадкового відбору:

$$\mu \approx \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{1475}{15}} = 9{,}92.$$

Відповідно до таблиці значень $t$ Стьюдента для $p$ = 0,05, або для довірчої ймовірності 0,95, при 14 степенях свободи ($k$= 14), відповідних 15 спостереженням, $t$ = 2,145. Отже, з ймовірністю 0,95 можна стверджувати, що шукана генеральна середня знаходиться в довірчих межах: $\bar{x} \pm t\mu = 460 \pm 9{,}92 \times 2{,}145 = 460 \pm 21{,}28$, тобто межі середньої знаходяться від 438,72 до 421,28 г. Якщо ж судження робиться за довірчої ймовірності 0,99, то за 14 степенях свободи $t$=2,977. Таким чином, з цієї більшої величини довірчої ймовірності дослідник повинен вказати ширші довірчі межі для невідомої генеральної середовищі: 460±9,92$\times$2,977, тобто від 430,47 до 489,53 г.

Можна поставити інше питання: яка ймовірність того, що величина середньої ваги деталей не вийде за 10 г?

Виходячи з того, що межа розбіжностей між вибірковою та генеральною середньою становить 29,53 г ($\Delta_{\bar{x}} = \left|\tilde{x} - \bar{x}\right| = |460 - 489{,}53|$), визначимо величину $t_ф$: $t_ф = \frac{\Delta_{\bar{x}}}{\mu_{\bar{x}}} = \frac{29{,}53}{9{,}92} = 2{,}977$. По таблиці розподілу Стьюдента $S_{k(t)}$ знаходимо, що при $k$=14 значення $t_ф$=2,977 відповідає ймовірність 0,990.

Оскільки малі вибірки дають неточні результати, її рідко використовують задля встановлення дійсної величини середнього розміру чи частки генеральної сукупності. В основному малі вибірки використовуються для оцінки ймовірності розбіжностей між показниками вибіркових сукупностей, наприклад, при визначенні втрат під час збирання врожаю на двох ділянках.

Оскільки малі вибірки дають неточні результати, їх рідко використовують задля встановлення дійсної величини середнього розміру чи частки генеральної сукупності. В основному малі вибірки використовуються для оцінки ймовірності розбіжностей між показниками вибіркових сукупностей, наприклад, при визначенні втрат під час збирання урожаю на двох ділянках.

## 12.7. Практика застосування вибіркового методу у статистиці

Вибірковий метод в українській статистиці набув широкого поширення. Він застосовується як у галузях матеріального виробництва, так і в невиробничій сфері. Перелік сфер застосування вибіркового методу, які українська статистика проводила і проводить, досить великий. Наведемо лише деякі з цих робіт:

1) вибіркове вивчення доходів та умов життя населення (заробітної плати робітників і службовців за професіями, житлових умов та забезпеченості населення предметами культурно-побутового обслуговування);

2) статистичні методи контролю якості продукції, що набули широкого поширення на підприємствах;

3) вибіркове обстеження продуктивності худоби, що знаходиться в особистій власності жителів сільської місцевості;

4) вибірковий облік посівних площ у господарствах населення;

5) вивчення соціально-демографічних питань;

6) застосування вибіркового методу при переписі населення.

**Застосування вибіркового методу під час обстеження населення.** Важливою сферою використання вибіркового методу виступає *обстеження населення*, яке проводиться на постійній основі з 1 січня 1999 р.[1] і відповідає міжнародним стандартам. Вибіркові дослідження населення є найважливішим джерелом інформації для вимірювання багатьох соціально-економічних і демографічних показників. Зокрема, вибірковий метод використовується для всеохоплюючих аналізу бюджетів та умов життя населення України, комплексного аналізу стану ринку праці тощо.

У статистиці, що регулярно ведеться, можна вказати, наприклад, на статистику обстеження умов життя населення, що охоплює близько 13 тис. сімей, тобто приблизно 0,1 % числа сімей в Україні. Вибіркова сукупність репрезентує все населення України, за винятком військовослужбовців строкової служби, осіб, що знаходяться у місцях позбавлення волі, осіб, що постійно проживають у будинках-інтернатах, будинках для осіб похилого віку, а також маргінальних прошарків населення. Крім того, при формуванні вибіркової сукупності були також виключені території, які не можуть обстежуватися у зв’язку з радіоактивним забрудненням (зони відчуження та обов’язкового відселення). Як показує досвід, незважаючи на невеликий обсяг вибірки, результати обстеження виходять задовільними. Це досягається зав-

---

[1] Згідно з постановою Кабінету Міністрів України від 02 листопада 1998 року № 1725 «Про проведення обстеження умов життя у домогосподарствах» Держкомстатом з 01 січня 1999 року запроваджено вибіркове обстеження умов життя домогосподарств.

дяки ретельному багатоступінчастому відбору бюджетів населення та гарною організацією вибіркового обстеження, проведенням та контролем якості робіт.

По кожному обстеженню визначаються вимоги щодо таких основних напрямів: 1) вимоги щодо представлення у вибірці територіальних одиниць відповідно до адміністративно-територіального устрою України; 2) вимоги до надійності результатів вибіркових обстежень населення; 3) вимоги до терміну дії територіальної вибірки.

Для вибору сімей, що підлягають вибірковому обстеженню, застосовують багатоступеневу вибірку. При цьому у великих і малих містах регіонів застосовують принцип побудови триступеневої вибірки, у сільській місцевості — двоступеневої У територіальній вибірці представлені всі регіони України відповідно до адміністративно-територіального устрою України. Для обстежень у міських поселеннях та сільській місцевості відбір територіальних одиниць здійснюють з імовірністю, пропорційною розміру (кількості населення), відбір домогосподарств здійснюють з використанням методу систематичного відбору. Відібрані сім'ї перевіряють на репрезентативність.

Крім часткових спостережень, що регулярно повторюються, в українській статистиці робляться й різні спеціальні одноразові обстеження, які також носять частковий характер (наприклад, обстеження запасів торгових підприємств та ін.).

**Статистичні методи контролю якості продукції**. Ці методи контролю розвиваються у двох напрямках: 1) контроль якості готової продукції та 2) контроль ходу технологічних процесів.

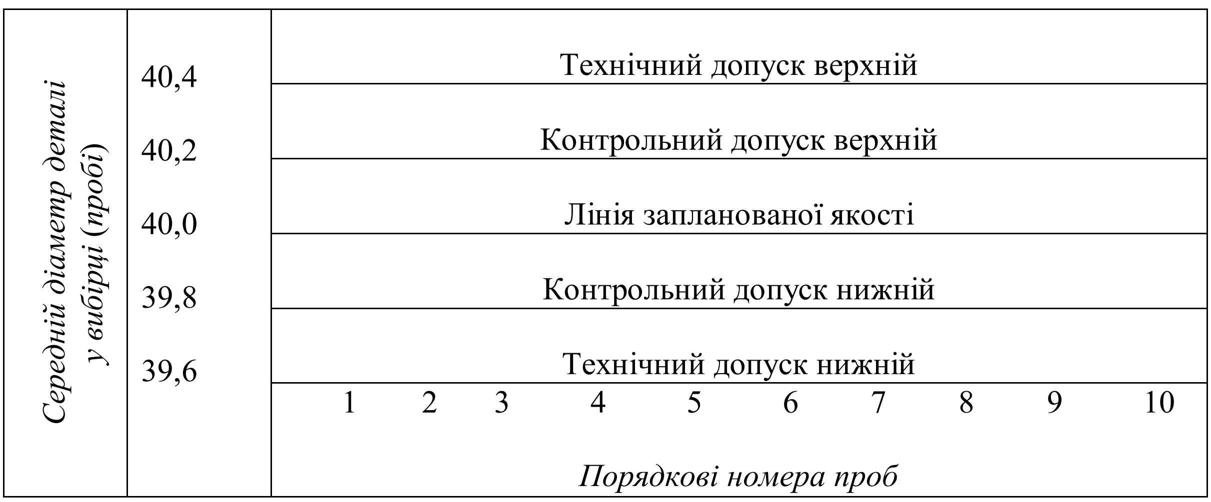
Статистичний контроль якості продукції з використанням вибіркового методу застосовується при вхідному, операційному, приймальному, інспекційному та інших видах контролю. Висновки про якість партії (потоку) продукції роблю на підставі контролю одиниць продукції, що входять у вибірку. Шляхом механічного відбору (кожен 5-й, 20-й, 100-й виріб) проводиться вибіркова перевірка придатності виробів та матеріалів, що надходять у виробництво. Якщо виріб упакований у тару, то найчастіше застосовується серійний відбір. Це так званий приймальний, або наступний контроль, заснований на перевірці якості виготовлених виробів; він не в змозі попередити появу браку. Тут застосування вибіркового методу вивчення якості продукції здійснюється за описаними раніше принципами. Завдяки застосуванню вибіркового методу значно скорочуються витрати на контроль якості продукції порівняно із суцільним контролем.

З метою попередження появи браку у виробництві застосовується попереджувальний поточний контроль технологічного процесу. Він набув широкого поширення в машинобудуванні, особливо у виробництвах з нестійкими апаратурними і безперервними процесами, в умовах високомеханізованих і автоматизованих виробництв, коли дотримання заданого параметра залежить не стільки від робочого оператора, скільки від стабільності режимів роботи та культури експлуатації діючих агрегатів

Попереджувальний поточний контроль має періодичну, за графіком, перевірку якості шляхом відбору проб або вибірок, завдяки чому виявляються та усуваються відхилення від нормального перебігу технологічних процесів ще до того, як ці відхилення можуть привести до браку. Відбір проб здійснюється під час виробничого процесу безпосередньо біля робочих місць. Такий контроль забезпечує систематичну перевірку як якості продукції, а й самого виробничого процесу, його стійкості. Головна увага при цьому приділяється дослідженню стабільності технологічного процесу, виявленню причин її порушення.

Попереджувальний контроль ходу технологічного процесу може бути організований по-різному, залежно від того, які характеристики (середня, мода, медіана) використовуються в кожній окремій вибірці.

Розглянемо один із способів запобіжного контролю ходу технологічного процесу з обчисленням середньої як статистичних характеристик проби. Припустимо, що у заводі у масовому порядку виготовляються вироби. Відповідно до ДСТУ, довжина виробу (найважливіший параметр якості) не повинен перевищувати 40±0,4 мм, тобто довжина не повинна бути меншою за 39,6 мм (нижня межа допуску) і не більше 40,4 мм (верхня межа допуску).



Спочатку готують спеціальну статистичну контрольну карту, на яку заздалегідь наносять контрольні лінії: дві характеризують величину технічних допусків (верхню і нижню межі), а інші дві визначають межі статистичних контрольних допусків. Нехай точність нашого виміру дорівнює 0,1 мм. Контролер періодично, скажімо через кожну годину, вибирає безпосередньо на робочому місці кілька пробних деталей, допустимо по 5 шт. у кожній пробі. На карту у вигляді точок наносяться контрольовані показники, наприклад, діаметр виробів.

Статистична контрольна карта відобразить весь хід технологічного процесу. Вихід точки за лінії верхнього чи нижнього технічних допусків свідчить про появу браку. Якщо ж точки виходять за межі контрольних статистичних допусків, то це служить сигналом про погіршення якості продукції, необхідності втручання у виробничий процес та усунення несправностей.

Допустимо, проба, взята після початку роботи верстата (8 год 00 хв.) показує довжину виробу 40,0 мм. Це означає, що верстат налаштований правильно і браку немає. Наступна проба взята о 8 год. 30 хв. — також свідчить про відсутність браку. Проба, взята о 9 год. 00 хв. показує 40,2 мм, свідчить про те, що середня довжина виробу увійшла в небезпечну зону і необхідно попередити наладчика. Так, у проміжку 11 год. 00 хв. — 11 год. 30 хв. середня довжина виробу перевищила верхній технічний допуск (40,5 мм). Це означає, що у цьому проміжку виготовлена продукція не відповідає вимогам.

Статистична контрольна карта вивішується у кожного верстата. На ній через певні проміжки часу відзначаються параметри виробу.

Вибірковий метод застосовується при профілактичному контролі і для визначення числа одиниць, які відбираються в пробу, і для визначення частоти взяття проб.

В останні роки у дослідженні використання робочого часу інтенсивно впроваджується так зване *моментно-вибіркове* спостереження. Суть моментно-вибіркового спостереження полягає в періодичній фіксації стану процесу на певні моменти години, які вибирають за схемою випадкової чи механічної вибірки (через певні інтервалі години). Воно застосовується для визначення середньої структури або середньої тривалості досліджуваних процесів. Так, за допомогою моментно-вибіркового спостереження в промисловості визначають склад змінного робочого часу робітників, фактичне навантаження обладнання і т.д.

Метод моментних спостережень характеризується відносно меншою трудомісткістю, ніж, наприклад, хронометраж або фотографія робочого дня, відрізняється простотою, дозволяє одному реєстратору фіксувати дані за великою кількістю об'єктів. При цьому результати (при правильній його організації) дає достатньо точні. Це дає можливість спостерігати за роботою обладнання на протязі тривалого проміжку часу (наприклад, на протязі місяця), оскільки обладнання за декадами буває завантаженим нерівно у зв'язку із неритмічністю виробництва.

Спосіб вибіркових спостережень робочого часу полягає у використанні випадкової вибірки із великої сукупності. Головна ідея способу вибіркових спостережень полягає в тому, що вірогідність зафіксувати при спостереженні даний різновид витрат часу при правильно розробленому плані дослідження повинна дорівнювати частці цієї різновидності в загальному підсумку затрат праці за період обстеження. За наявності достатнього обсягу вибірки результати поширюються на всю сукупність.

В програму вибіркового спостереження необхідно включати лише суттєві ознаки, які характеризують досліджуване явище. При укрупненому дослідження використання робочого часу суттєвими ознаками будь ознаки, що характеризують лише роботу і перериви (всі разом). Окремі причини перериві будуть несуттєвими, другорядними ознаками. Звісно, при більш детальному аналізу причин простоїв обладнання суттєвими ознаками будуть саме окремі причини, пов'язані із простоєм обладнання.

Спосіб вибіркових обстежень роботи чи простою заключається у проведенні спостережень через певні випадкові інтервали над одним або більше робітниками або верстатами з відмітками про роботу або простій. Якщо в момент перевірки верстат або робітник працює, в картці робиться відмітка «працює», якщо не працює – «простоює». Процент часу, на протязі якого робітник або верстат не працював, відповідає відношенню числа відміток в строчці «простій» до загальної кількості відміток.

На рис. 12.2 наведена картка спостережень; в ній сорок відміток «працює» і шість «простоює», всього сорок спостережень.

В нашому прикладі процент простою обладнання становить 15 % $\left(\frac{6}{40}\cdot 100\right)$ і час роботи — 85 % $\left(\frac{34}{40}\cdot 100\right)$. Якщо спостереження охоплювало одного робітника на протязі зміни (8 годин), то робітник простоював 72 хв. і працював 408 хв.

| Стан | Кількість випадків | Всього |
| --- | --- | --- |
| Працює | \|\|\|\|\| \|\|\|\|\| \|\|\|\|\| \|\|\|\|\| \|\|\|\|\| \|\|\|\|\| \|\|\|\| | 34 |
| Простоює | \|\|\|\|\| \| | 6 |

*Рис. 12.2.* Відмітки про час спостереження при способі миттєвих спостережень

Відбір у вибіркову сукупність моментних станів об’єктів, що виступають тут одиницями сукупності та одиницями спостереження, здійснюється рядом способів, але переважно механічно. Однак через незворотність часу він завжди є безповторним.

Найважливішим питанням моментно-вибіркового спостереження є визначення чисельності вибірки. Для визначення обсягу вибірки альтернативної ознаки може бути використана така формула:

$$n = \frac{t^2(1-w)100^2}{d^2 w}, \tag{12.61}$$

де $t$ — коефіцієнт довірчої ймовірності;
$w$ — частка у вибірковій сукупності (звичайно це коефіцієнт завантаженості робітників, що визначається на основі звітних даних);
$d$ — заздалегідь встановлена відносна гранична помилка спостереження.

Ця формула часто використовується як основа для різних модифікацій. Як правило, $t$ в умовах стабільного виробничого процесу приймається за 2, нестабільного — за 3.

Приведемо приклад виявлення непродуктивних втрат часу обладнання методом моментних спостережень на одній із ділянці промислового підприємства.

Встановимо, наприклад, кількість необхідних спостережень за роботою токарів в умовах нестабільного виробництва. Раніше було проведена фотографія часу роботи обладнання і було встановлено, що час використання роботи обладнання (основний і допоміжний) становив 1568 хв., час простоїв 718 хв. Отже коефіцієнт використання змінного фонду роботи обладнання ($w$) становить 68,59 % $\left(\frac{1\,568}{718+1\,568}\cdot 100\right)$. Будемо вважати, що в цеху виробничий процес є нестабільним, тому застосовуємо $t$=3. Будемо вважати достатньо точним результат спостереження при відносній граничній помилці не більше ±5 %. Звідси кількість необхідних моментів спостереження становитиме:

$$n = \frac{3^2 \times (1 - 0{,}6859) \times 100^2}{0{,}6859 \times 5^2} = 1650 \text{ спостережень.}$$

Якщо на ділянці встановлено 10 верстатів, то необхідно здійснити 165 обходів верстатів (1650:10). Загальний час спостереження, тобто тривалість зміни, — 480 хв. Звідси тривалість одного обходу 480:165=2,91 хв., тобто 2 хв. 55 сек.

Отримані в результаті вибіркового обстеження дані про використання обладнання наведені в табл. 12.21.

Таблиця 12.21

**Час роботи і простої обладнання**

| Номер станка | Всього спостережень | | Час використання обладнання | | | | Простої обладнання[1] | |
|---|---|---|---|---|---|---|---|---|
| | | | основний | | додатковий | | | |
| | моментів | процентів | моментів | процентів | моментів | процентів | моментів | процентів |
| 1 | 165 | 100 | 132 | 80,0 | 19 | 11,5 | 14 | 8,5 |
| 2 | 165 | 100 | 128 | 77,6 | 18 | 10,9 | 19 | 11,5 |
| 3 | 165 | 100 | 126 | 76,4 | 19 | 11,5 | 20 | 12,1 |
| 4 | 165 | 100 | 134 | 81,2 | 15 | 9,1 | 16 | 9,7 |
| 5 | 165 | 100 | 122 | 73,9 | 24 | 14,5 | 19 | 11,5 |
| 6 | 165 | 100 | 140 | 84,8 | 18 | 10,9 | 7 | 4,2 |
| 7 | 165 | 100 | 135 | 81,8 | 15 | 9,1 | 15 | 9,1 |
| 8 | 165 | 100 | 115 | 69,7 | 13 | 7,9 | 37 | 22,4 |
| 9 | 165 | 100 | 70 | 42,4 | 17 | 10,3 | 78 | 47,3 |
| 10 | 165 | 100 | 113 | 68,5 | 32 | 19,4 | 20 | 12,1 |
| Всього | 1 650 | × | 1 215 | × | 190 | × | 245 | × |

[1] Простої обладнання внаслідок різних причин, зокрема, порушення трудової дисципліни, відсутності сировини і т. п.

Зведемо отримані дані в табл. 12.22.

Простої обладнання були зафіксовані в 245 випадках, що становило 14,85 % від змінного фонду робочого часу обладнання. На основі відносних показників визначаємо абсолютні втрати робочого часу. В нашому випадку вони в середньому за зміну на один верстат склали 26,71 хв. (480× 14,84 : 100).

Після цього необхідно перевірити репрезентативність результатів

Таблиця 12.22

**Структура витрат змінного часу роботи обладнання**

| Елементи змінного фонду часу | Число моментів спостережень за зміну | Питома вага елементів (в процентів до підсумку) |
|---|---|---|
| Час використання обладнання: | | |
| основний | 1 215 | 73,64 |
| додатковий | 190 | 11,52 |
| Простої обладнання | 245 | 14,84 |
| Всього | 1 650 | 100,00 |

спостереження. Щоб робити висновок про точній моментного спостереження, необхідно із заданою вірогідністю визначити величину ознаки в генеральній сукупності. Для цього за формулою (12.61) розрахуємо граничну похибку моментного спостереження, причому *w* (питома вага часу у змінному фонді) прийнята рівною 0,8516 (0,7364+0,1152).

Звідси

$$\Delta_p = 3\sqrt{\frac{0{,}8516(1-0{,}8516)}{1650}} = 3\times 0{,}008752 = 0{,}026255, \text{ або } 2{,}6\ \%.$$

Таким чином, із вірогідністю 0,997 ($t$=3) можна стверджувати, що непродуктивні втрати в генеральній сукупності рівні 14,84 % ±2,6 %, або не менше 14,14 % і не більше 17,44 %.

Засіб вибіркового обстеження робочого часу, що пов’язаний із невеликою систематичною похибкою, може бути кращим інших способів, які можуть привести до великих випадкових помилок. У цілому нині вибірковий метод дозволяє проводити найбільш раціональний збір і обробку необхідної інформації про якість продукції та характеристики ходу технологічного процесу, що дозволяє скласти судження про перебіг всього виробничого процесу на підприємстві.

Тією чи іншою мірою, на тих чи інших етапах виробництва вибірковий метод давно і широко застосовується у вітчизняній економіці, проте найбільший ефект дає лише комплексне застосування на сучасній науковій базі та впровадженням математичних методів та ЕОМ, створенням АСДС (автоматизованої системи державної) статистики). Розумне поєднання суцільного та вибіркового спостережень може з’явитися великим кроком у досягненні оптимальної системи статистичної інформації в країні.

## ЛІТЕРАТУРА


1. *Горкавий В. К.* Статистика: Підручник. Третє вид., переробл. і доповн. Київ: Алерта, 2019. С. 145–180.
2. Статистика: підручник / *С. І. Пирожков*, *В. В. Рязанцева*, *Р. М. Моторин* та ін. Київ: Київ. нац. торг.-екон. ун-т, 2020. С. 162–177.
3. Статистика: Підручник / *С. С. Герасименко*, *А. В. Головач*, *А.М. Єріна* та ін.; За наук. ред. д-ра екон. наук С. С. Герасименка. 2-ге вид., перероб. і доп. Київ: КНЕУ, 2000. С. 81–99.


## ДОДАТКОВА ЛІТЕРАТУРА

*Навчальні посібники, словники*


4. Вибіркове спостереження: Термінологічний словник / [уклад. *О. І. Черняк*, *Є. М. Жуйкова*, *О. В. Гончар* та ін.]; під наук. кер. О. О. Васєчко. Київ: НТК статистичних досліджень, 2004. 140 с.
5. *Єріна А. М.* Організація вибіркових обстежень: навч. посіб. Київ: КНЕУ, 2004. 127 с.
6. *Козирєва О. В.*, *Федорова В. О.* Статистика: навчальний посібник. Харків: Видавництво Іванченка І.С., 2021. С. 34–42.
7. *Педченко Г. П.* Статистика: Навчальний посібник. Мелітополь: Колор Принт, 2018. С. 162–177.


*Монографії та статті*


8. *Гладун О. М.* Вибіркові обстеження населення: методологія, методика, практика: монографія. Ніжин: ТОВ Видавництво «Аспект-Поліграф», 2008.
9. *Непран А. В.*, *Тимченко І. Є.* Визначення внутрішньозмінних втрат робочого часу роботи обладнання методом миттєвих спостережень. *Бізнес Інформ*. 2022. №1. С. 159–164.

Додаток 1

**Таблиця значень функції**

$$f(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt \text{ при різних значеннях } t$$

| $t$ | $F(t)$ | $t$ | $F(t)$ | $t$ | $F(t)$ | $t$ | $F(t)$ |
|---|---|---|---|---|---|---|---|
| 0,00 | 0,00000 | 0,30 | 0,23582 | 0,60 | 0,45149 | 0,90 | 0,63188 |
| 01 | 00798 | 31 | 24344 | 61 | 45814 | 91 | 63718 |
| 02 | 01596 | 32 | 25103 | 62 | 46474 | 92 | 64243 |
| 03 | 02393 | 33 | 25860 | 63 | 47131 | 93 | 64763 |
| 04 | 03191 | 34 | 26614 | 64 | 47783 | 94 | 65278 |
| 05 | 03988 | 35 | 27366 | 65 | 48431 | 95 | 65789 |
| 06 | 04784 | 36 | 28115 | 66 | 49075 | 96 | 66294 |
| 07 | 05581 | 37 | 28862 | 67 | 49714 | 97 | 66795 |
| 08 | 06376 | 38 | 29605 | 68 | 50350 | 98 | 67291 |
| 09 | 07171 | 39 | 30346 | 69 | 50981 | 99 | 67783 |
| 0,10 | 0,07966 | 0,40 | 0,31084 | 0,70 | 0,51607 | 1,00 | 0,68269 |
| 11 | 08759 | 41 | 31819 | 71 | 52230 | 01 | 68750 |
| 12 | 09552 | 42 | 32552 | 72 | 52848 | 02 | 69227 |
| 13 | 10348 | 43 | 33280 | 73 | 53461 | 03 | 69699 |
| 14 | 11134 | 44 | 34006 | 74 | 54070 | 04 | 70166 |
| 15 | 11924 | 45 | 34729 | 75 | 54675 | 1,05 | 70628 |
| 16 | 12712 | 46 | 35448 | 76 | 55275 | 06 | 71086 |
| 17 | 13499 | 47 | 34164 | 77 | 55870 | 07 | 71538 |
| 18 | 14285 | 48 | 33877 | 78 | 56461 | 08 | 71986 |
| 19 | 15069 | 49 | 37587 | 79 | 57047 | 09 | 72429 |
| 0,20 | 0,15852 | 0,50 | 0,38292 | 0,80 | 0,57629 | 1,10 | 0,72867 |
| 21 | 16633 | 51 | 38995 | 81 | 58206 | 11 | 73300 |
| 22 | 17413 | 52 | 39694 | 82 | 58778 | 12 | 73729 |
| 23 | 18191 | 53 | 40389 | 83 | 59346 | 13 | 74152 |
| 24 | 18967 | 54 | 41080 | 84 | 59909 | 14 | 74571 |
| 25 | 19741 | 55 | 41768 | 85 | 60468 | 15 | 74986 |
| 26 | 20514 | 56 | 42452 | 86 | 61021 | 16 | 75395 |
| 27 | 21284 | 57 | 43132 | 87 | 61570 | 17 | 75800 |
| 28 | 22052 | 58 | 43809 | 88 | 62114 | 18 | 76200 |
| 29 | 22818 | 59 | 44481 | 89 | 52653 | 19 | 76595 |

| $t$ | $F(t)$ | $t$ | $F(t)$ | $t$ | $F(t)$ | $t$ | $F(t)$ |
|---|---|---|---|---|---|---|---|
| 1,20 | 0,76986 | 1,55 | 0,87886 | 1,90 | 0,94257 | 2,25 | 0,97555 |
| 21 | 77372 | 56 | 88124 | 91 | 94387 | 26 | 97618 |
| 22 | 77754 | 57 | 88358 | 92 | 94514 | 27 | 97679 |
| 23 | 78130 | 58 | 88589 | 93 | 94639 | 28 | 97739 |
| 24 | 78502 | 59 | 88817 | 94 | 94762 | 29 | 97798 |
| 25 | 78870 | 1,60 | 0,89040 | 95 | 94882 | 2,30 | 0,97855 |
| 26 | 79233 | 61 | 89260 | 96 | 95000 | 30 | 97911 |
| 27 | 79592 | 62 | 89477 | 97 | 95116 | 32 | 97966 |
| 28 | 79945 | 63 | 89690 | 98 | 95230 | 33 | 98019 |
| 29 | 80295 | 64 | 89899 | 99 | 95341 | 34 | 98072 |
| 1,30 | 0,80640 | 65 | 90106 | 2,00 | 0,95450 | 35 | 98123 |
| 31 | 80980 | 66 | 90309 | 01 | 95557 | 36 | 98172 |
| 32 | 81316 | 67 | 90508 | 02 | 95662 | 37 | 98221 |
| 33 | 81648 | 68 | 90704 | 03 | 95764 | 38 | 98269 |
| 34 | 81975 | 69 | 90897 | 04 | 95865 | 39 | 98315 |
| 35 | 82298 | 1,70 | 0,91087 | 05 | 95964 | 2,40 | 0,98360 |
| 36 | 82617 | 71 | 91273 | 06 | 96060 | 41 | 98405 |
| 37 | 82931 | 72 | 91457 | 07 | 66155 | 42 | 98448 |
| 38 | 83241 | 73 | 91637 | 08 | 96247 | 43 | 98490 |
| 39 | 83547 | 74 | 91814 | 09 | 96338 | 44 | 98531 |
| 1,40 | 0,83849 | 75 | 91988 | 2,10 | 0,96427 | 45 | 98571 |
| 41 | 84146 | 76 | 92159 | 11 | 96514 | 46 | 98611 |
| 42 | 84439 | 77 | 92327 | 12 | 96599 | 47 | 98649 |
| 43 | 84729 | 78 | 92492 | 13 | 96683 | 48 | 98686 |
| 44 | 85013 | 79 | 92655 | 14 | 96765 | 49 | 98723 |
| 45 | 85294 | 1,80 | 0,92814 | 15 | 96844 | 2,50 | 0,98758 |
| 46 | 85571 | 81 | 92970 | 16 | 96923 | 51 | 98793 |
| 47 | 85844 | 82 | 93124 | 17 | 96999 | 52 | 98826 |
| 48 | 86113 | 83 | 93275 | 18 | 97074 | 53 | 98859 |
| 49 | 86378 | 84 | 93426 | 19 | 97148 | 54 | 98891 |
| 1,50 | 0,86639 | 85 | 93569 | 2,20 | 0,97219 | 55 | 98923 |
| 51 | 86696 | 86 | 93711 | 21 | 97289 | 56 | 98953 |
| 52 | 87149 | 87 | 93852 | 22 | 97358 | 57 | 98983 |
| 53 | 87398 | 88 | 93989 | 23 | 97425 | 58 | 99012 |
| 54 | 87644 | 89 | 94124 | 24 | 97491 | 59 | 99040 |

| $t$ | $F(t)$ | $t$ | $F(t)$ | $t$ | $F(t)$ | $t$ | $F(t)$ |
|---|---|---|---|---|---|---|---|
| 2,60 | 0,99068 | 2,95 | 0,99682 | 3,30 | 0,99903 | 3,65 | 0,99974 |
| 61 | 99095 | 96 | 99692 | 31 | 99907 | 66 | 99975 |
| 62 | 99121 | 97 | 99702 | 32 | 99910 | 67 | 99976 |
| 63 | 99146 | 98 | 99712 | 33 | 99913 | 68 | 99977 |
| 64 | 99171 | 99 | 99721 | 34 | 99916 | 69 | 99978 |
| 65 | 99195 | 3,00 | 0,99730 | 35 | 99919 | 3,70 | 0,99978 |
| 66 | 99219 | 01 | 99739 | 36 | 99922 | 71 | 99979 |
| 67 | 99241 | 02 | 99747 | 37 | 99925 | 72 | 99980 |
| 68 | 99263 | 03 | 99755 | 38 | 99928 | 73 | 99981 |
| 69 | 99285 | 04 | 99763 | 39 | 99930 | 74 | 99982 |
| 2,70 | 0,99307 | 05 | 99771 | 3,40 | 0,99933 | 75 | 99982 |
| 71 | 99327 | 06 | 99779 | 41 | 99935 | 76 | 99983 |
| 72 | 99347 | 07 | 99786 | 42 | 99937 | 77 | 99984 |
| 73 | 99367 | 08 | 99793 | 43 | 99940 | 78 | 99984 |
| 74 | 99386 | 09 | 99800 | 44 | 99942 | 79 | 99985 |
| 75 | 99404 | 3,10 | 0,99806 | 45 | 99944 | 3,80 | 0,99986 |
| 76 | 99422 | 11 | 99813 | 46 | 99946 | 81 | 99986 |
| 77 | 99439 | 12 | 99819 | 47 | 99948 | 82 | 99987 |
| 78 | 99456 | 13 | 99825 | 48 | 99950 | 83 | 99987 |
| 79 | 99473 | 14 | 99831 | 49 | 99952 | 84 | 99988 |
| 2,80 | 0,99489 | 15 | 99837 | 3,50 | 0,99953 | 85 | 99988 |
| 81 | 99505 | 16 | 99842 | 51 | 99955 | 86 | 99989 |
| 82 | 99520 | 17 | 99848 | 52 | 99957 | 87 | 99989 |
| 83 | 99535 | 18 | 99853 | 53 | 99958 | 88 | 99990 |
| 84 | 99549 | 19 | 99585 | 54 | 99960 | 89 | 99990 |
| 85 | 99563 | 3,20 | 0,99863 | 55 | 99961 | 3,90 | 0,99990 |
| 86 | 99576 | 21 | 99867 | 56 | 99963 | 91 | 99991 |
| 87 | 99590 | 22 | 99872 | 57 | 99964 | 92 | 99991 |
| 88 | 99602 | 23 | 99876 | 58 | 99966 | 93 | 99992 |
| 89 | 99615 | 24 | 99880 | 59 | 99967 | 94 | 99992 |
| 2,90 | 0,99627 | 25 | 99855 | 3,60 | 0,99968 | 95 | 99992 |
| 91 | 99639 | 26 | 99889 | 61 | 99969 | 96 | 99992 |
| 92 | 99650 | 27 | 99892 | 62 | 99971 | 97 | 99993 |
| 93 | 99661 | 28 | 99896 | 63 | 99972 | 98 | 99993 |
| 94 | 99672 | 29 | 99900 | 64 | 99973 | 99 | 99993 |

Додаток 2

**Таблиця ймовірностей *P* ( $\chi^2$ )**

| $\chi^2$ \ $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,3173 | 0,6065 | 0,8013 | 0,9098 | 0,9626 | 0,9656 | 0,9948 | 0,9982 | 0,9994 | 0,9998 |
| 2 | 1574 | 3679 | 5724 | 7318 | 8491 | 9197 | 9598 | 9810 | 9915 | 9963 |
| 3 | 0833 | 2231 | 3916 | 5578 | 7000 | 8088 | 8850 | 9344 | 9643 | 9814 |
| 4 | 0455 | 1353 | 2615 | 4060 | 5494 | 6767 | 7798 | 8571 | 9114 | 9473 |
| 5 | 0254 | 0821 | 1718 | 2873 | 4153 | 5438 | 6600 | 7576 | 8343 | 8912 |
| 6 | 0143 | 0498 | 1116 | 1991 | 3062 | 4232 | 5398 | 6472 | 7399 | 8153 |
| 7 | 0081 | 0302 | 0719 | 1359 | 2206 | 3208 | 4289 | 5366 | 6371 | 7254 |
| 8 | 0047 | 0183 | 0460 | 0916 | 1562 | 2381 | 3326 | 4335 | 5341 | 6288 |
| 9 | 0027 | 0111 | 0293 | 0611 | 1091 | 1736 | 2527 | 3423 | 4373 | 5321 |
| 10 | 0016 | 0067 | 0186 | 0404 | 0752 | 1247 | 1886 | 2650 | 3505 | 4405 |
| 11 | 0009 | 0041 | 0117 | 0266 | 0514 | 0884 | 1386 | 2017 | 2757 | 3575 |
| 12 | 0005 | 0025 | 0074 | 0174 | 0348 | 0620 | 1006 | 1512 | 2133 | 2851 |
| 13 | 0003 | 0015 | 0046 | 0113 | 0234 | 0430 | 0721 | 1119 | 1626 | 2237 |
| 14 | 0002 | 0009 | 0029 | 0073 | 0156 | 0296 | 0512 | 0818 | 1223 | 1730 |
| 15 | 0001 | 0006 | 0018 | 0047 | 0104 | 0203 | 0360 | 0591 | 0909 | 1321 |
| 16 | 0001 | 0003 | 0011 | 0030 | 0068 | 0138 | 0251 | 0424 | 0669 | 0996 |
| 17 | 0000 | 0002 | 0007 | 0019 | 0045 | 0093 | 0174 | 0301 | 0487 | 0744 |
| 18 | – | 0001 | 0004 | 0012 | 0029 | 0062 | 0120 | 0212 | 0352 | 0550 |
| 19 | – | 0001 | 0003 | 0008 | 0019 | 0042 | 0082 | 0149 | 0252 | 0403 |
| 20 | – | 0000 | 0002 | 0005 | 0013 | 0028 | 0056 | 0103 | 0179 | 0293 |
| 21 | – | – | 0001 | 0003 | 0008 | 0018 | 0038 | 0071 | 0126 | 0211 |
| 22 | – | – | 0001 | 0002 | 0005 | 0012 | 0025 | 0049 | 0089 | 0151 |
| 23 | – | – | 0000 | 0001 | 0003 | 0008 | 0017 | 0034 | 0062 | 0107 |
| 24 | – | – | – | 0001 | 0002 | 0005 | 0011 | 0023 | 0043 | 0076 |
| 25 | – | – | – | 0001 | 0001 | 0003 | 0008 | 0016 | 0030 | 0053 |
| 26 | – | – | – | 0000 | 0001 | 0002 | 0005 | 0010 | 0020 | 0037 |
| 27 | – | – | – | – | 0001 | 0001 | 0003 | 0007 | 0014 | 0026 |
| 28 | – | – | – | – | 0000 | 0001 | 0002 | 0005 | 0010 | 0018 |
| 29 | – | – | – | – | – | 0001 | 0001 | 0003 | 0006 | 0012 |
| 30 | – | – | – | – | – | 0000 | 0001 | 0002 | 0004 | 0009 |

Продовження додатку 2

| $\chi^2$ \ $k$ | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,9999 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 2 | 9985 | 0,9994 | 0,9998 | 0,9999 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 3 | 9907 | 9955 | 9979 | 9991 | 0,9996 | 0,9998 | 0,9999 | 1,0000 | 1,0000 | 1,0000 |
| 4 | 9699 | 9834 | 9912 | 9955 | 9977 | 9989 | 9995 | 0,9998 | 0,9999 | 1,0000 |
| 5 | 9312 | 9580 | 9752 | 9858 | 9921 | 9958 | 9978 | 9989 | 9994 | 0,9997 |
| 6 | 8734 | 9161 | 9462 | 9665 | 9797 | 9881 | 9932 | 9962 | 9979 | 9989 |
| 7 | 7991 | 8576 | 9022 | 9347 | 9576 | 9733 | 9835 | 9901 | 9942 | 9967 |
| 8 | 733 | 7851 | 8436 | 8893 | 9238 | 9489 | 9665 | 9786 | 9867 | 9919 |
| 9 | 9219 | 7029 | 7729 | 8311 | 8775 | 9134 | 9403 | 9597 | 9735 | 9829 |
| 10 | 5304 | 6160 | 6939 | 7622 | 8197 | 8666 | 9036 | 9319 | 9539 | 9682 |
| 11 | 4433 | 5289 | 6108 | 6860 | 7526 | 8095 | 8566 | 8944 | 9238 | 9462 |
| 12 | 3626 | 4457 | 5276 | 6063 | 6790 | 7440 | 8001 | 8472 | 8856 | 9161 |
| 13 | 2933 | 3690 | 4478 | 5265 | 6023 | 6728 | 7362 | 7916 | 8386 | 8774 |
| 14 | 2330 | 3007 | 3738 | 4497 | 5255 | 5987 | 6671 | 7291 | 7837 | 8305 |
| 15 | 1825 | 2414 | 3074 | 3782 | 4514 | 5246 | 5955 | 6620 | 7226 | 7764 |
| 16 | 1411 | 1912 | 2491 | 3134 | 3821 | 4530 | 5238 | 5925 | 6573 | 7166 |
| 17 | 1079 | 1496 | 1993 | 2562 | 3189 | 3856 | 4544 | 5231 | 5899 | 6530 |
| 18 | 0816 | 1157 | 1575 | 2068 | 2627 | 3239 | 3888 | 4557 | 5224 | 5874 |
| 19 | 0611 | 0885 | 1231 | 1649 | 2137 | 2687 | 3285 | 3918 | 4568 | 5218 |
| 20 | 0453 | 0671 | 0952 | 1301 | 1719 | 2202 | 2742 | 3328 | 3946 | 4579 |
| 21 | 0334 | 0504 | 0729 | 1016 | 1368 | 1785 | 2263 | 2794 | 3368 | 3971 |
| 22 | 0244 | 0375 | 0554 | 0786 | 1078 | 1432 | 1847 | 2320 | 2843 | 3405 |
| 23 | 0177 | 0277 | 0417 | 0603 | 0841 | 1137 | 1498 | 1906 | 2373 | 2888 |
| 24 | 0127 | 0203 | 0311 | 0458 | 0651 | 0895 | 1194 | 1550 | 1962 | 2424 |
| 25 | 0091 | 0148 | 0231 | 0346 | 0499 | 0698 | 0947 | 1249 | 1605 | 2014 |
| 26 | 0065 | 0107 | 0170 | 0259 | 0380 | 0540 | 0745 | 0998 | 1302 | 1658 |
| 27 | 0046 | 0077 | 0124 | 0193 | 0287 | 0415 | 0581 | 0790 | 1047 | 1353 |
| 28 | 0032 | 0055 | 0090 | 0142 | 0216 | 0316 | 0449 | 0621 | 0834 | 1094 |
| 29 | 0023 | 0039 | 0065 | 0104 | 0161 | 0239 | 0345 | 0484 | 0660 | 0878 |
| 30 | 0016 | 0028 | 0047 | 0076 | 0119 | 0180 | 0263 | 0374 | 0518 | 0699 |

Продовження додатку 2

| $\chi^2$ \ $k$ | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 2 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 3 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 4 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 5 | 0,9999 | 0,9999 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 6 | 9994 | 9997 | 0,9999 | 0,9999 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 7 | 9981 | 9990 | 9995 | 9997 | 0,9999 | 0,9999 | 1,0000 | 1,0000 | 1,0000 |
| 8 | 9951 | 9972 | 9984 | 9991 | 9995 | 9997 | 0,9999 | 0,9999 | 1,0000 |
| 9 | 9892 | 9933 | 9960 | 9976 | 9986 | 9992 | 9995 | 9997 | 0,9999 |
| 10 | 9789 | 9863 | 9913 | 9945 | 9967 | 9980 | 9988 | 9993 | 9996 |
| 11 | 9628 | 9747 | 9832 | 9890 | 9929 | 9955 | 9972 | 9983 | 9990 |
| 12 | 9396 | 9574 | 9705 | 9799 | 9866 | 9912 | 9943 | 9964 | 9977 |
| 13 | 9086 | 9332 | 9520 | 9661 | 9765 | 9840 | 9892 | 9929 | 9954 |
| 14 | 8696 | 9015 | 9269 | 9466 | 9617 | 9730 | 9813 | 9872 | 9914 |
| 15 | 8230 | 8622 | 8946 | 9208 | 9414 | 9573 | 9694 | 9784 | 9850 |
| 16 | 7696 | 8159 | 8553 | 8881 | 9148 | 9362 | 9929 | 9658 | 9755 |
| 17 | 7111 | 7634 | 8093 | 8487 | 8818 | 9091 | 9311 | 9486 | 9622 |
| 18 | 6490 | 7060 | 7575 | 8080 | 8424 | 8758 | 9035 | 9261 | 9443 |
| 19 | 5851 | 6453 | 7012 | 7520 | 7971 | 8364 | 8700 | 8981 | 9213 |
| 20 | 5213 | 5830 | 6419 | 6968 | 7468 | 7916 | 8308 | 8645 | 8929 |
| 21 | 4589 | 5207 | 581 | 6387 | 6926 | 7420 | 7863 | 8253 | 8591 |
| 22 | 3995 | 4599 | 5203 | 5793 | 6357 | 6887 | 7374 | 7813 | 8202 |
| 23 | 3440 | 4017 | 4608 | 5198 | 5776 | 6329 | 6850 | 7330 | 7765 |
| 24 | 2981 | 3472 | 4038 | 4616 | 5194 | 5760 | 6303 | 6815 | 7289 |
| 25 | 2472 | 2971 | 3503 | 4058 | 4624 | 5190 | 5745 | 5278 | 5782 |
| 26 | 2064 | 2517 | 3009 | 3532 | 4076 | 4631 | 5186 | 5730 | 6255 |
| 27 | 1709 | 2112 | 2560 | 3045 | 3559 | 4093 | 4638 | 5182 | 5717 |
| 28 | 1402 | 1757 | 2158 | 2600 | 3079 | 3585 | 4110 | 4644 | 5179 |
| 29 | 1140 | 1449 | 1803 | 2201 | 2639 | 3111 | 3609 | 4125 | 4651 |
| 30 | 0920 | 1185 | 1494 | 1848 | 2243 | 2676 | 3142 | 3632 | 4140 |

Додаток 3

**Таблиця істотності $\chi^2$**

| $p$ | 0,99 | 0,98 | 0,95 | 0,90 | 0,80 | 0,70 | 0,50 | 0,30 | 0,20 | 0,10 | 0,05 | 0,02 | 0,01 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,00016 | 0,0006 | 0,039 | 0,0158 | 0,064 | 0,148 | 0,455 | 1,074 | 1,642 | 2,706 | 3,841 | 5,412 | 6,635 |
| 2 | 0,0201 | 0,0404 | 0,103 | 0,211 | 0,446 | 0,713 | 1,386 | 2,408 | 3,219 | 4,605 | 5,991 | 7,824 | 9,210 |
| 3 | 0,115 | 0,185 | 0,352 | 0,584 | 1,005 | 1,424 | 2,366 | 3,665 | 4,642 | 6,251 | 7,815 | 9,837 | 11,345 |
| 4 | 0,297 | 0,429 | 0,711 | 1,064 | 1,649 | 2,195 | 3,357 | 4,878 | 5,989 | 7,779 | 9,488 | 11,668 | 13,277 |
| 5 | 0,554 | 0,752 | 1,145 | 1,160 | 2,343 | 3,000 | 4,351 | 6,064 | 7,289 | 9,236 | 11,070 | 13,388 | 15,086 |
| 6 | 0,872 | 1,134 | 1,635 | 2,204 | 3,070 | 3,828 | 5,348 | 7,231 | 8,558 | 10,645 | 12,592 | 15,033 | 16,812 |
| 7 | 1,239 | 1,564 | 2,167 | 2,333 | 3,822 | 4,671 | 6,346 | 8,383 | 9,803 | 12,017 | 14,067 | 16,622 | 18,475 |
| 8 | 1,646 | 2,032 | 2,733 | 3,490 | 4,594 | 5,527 | 7,344 | 9,524 | 11,030 | 13,362 | 15,507 | 18,168 | 20,090 |
| 9 | 2,088 | 2,532 | 3,325 | 4,168 | 5,380 | 6,393 | 8,343 | 10,656 | 12,242 | 14,684 | 16,919 | 19,679 | 21,666 |
| 10 | 2,358 | 3,059 | 3,940 | 4,865 | 6,179 | 7,267 | 9,342 | 11,781 | 13,442 | 15,987 | 18,307 | 21,161 | 23,209 |
| 11 | 3,053 | 3,609 | 4,575 | 5,578 | 6,898 | 8,148 | 10,341 | 12,899 | 14,631 | 17,275 | 19,675 | 22,618 | 24,725 |
| 12 | 3,571 | 4,178 | 5,226 | 6,304 | 7,807 | 9,024 | 11,340 | 14,011 | 15,821 | 18,549 | 21,026 | 24,054 | 26,217 |
| 13 | 4,107 | 4,765 | 5,892 | 7,042 | 8,364 | 9,926 | 12,340 | 15,119 | 16,985 | 19,812 | 22,362 | 25,472 | 27,688 |
| 14 | 4,660 | 5,368 | 6,571 | 7,790 | 9,467 | 10,821 | 13,339 | 16,222 | 18,151 | 21,064 | 23,685 | 26,873 | 29,141 |
| 15 | 5,229 | 5,985 | 7,261 | 8,547 | 10,307 | 11,721 | 14,339 | 17,322 | 19,311 | 22,307 | 24,996 | 28,259 | 30,578 |
| 16 | 5,812 | 6,614 | 7,962 | 9,312 | 11,152 | 12,624 | 15,338 | 18,418 | 20,465 | 23,542 | 26,296 | 29,633 | 32,000 |
| 17 | 6,408 | 7,255 | 8,672 | 10,085 | 12,002 | 13,531 | 16,338 | 19,511 | 21,615 | 24,769 | 27,587 | 30,995 | 33,409 |
| 18 | 7,015 | 7,906 | 9,390 | 10,865 | 12,857 | 14,440 | 17,338 | 20,601 | 22,760 | 25,988 | 28,869 | 32,346 | 34,805 |
| 19 | 7,633 | 8,567 | 10,117 | 11,651 | 13,716 | 15,352 | 18,338 | 21,689 | 23,900 | 27,204 | 30,144 | 33,687 | 36,191 |
| 20 | 8,260 | 9,237 | 10,851 | 12,443 | 14,578 | 16,266 | 19,337 | 22,775 | 25,038 | 28,412 | 31,410 | 35,020 | 37,566 |
| 21 | 8,897 | 9,915 | 11,591 | 13,240 | 15,445 | 17,182 | 20,337 | 23,858 | 26,171 | 29,615 | 32,671 | 36,343 | 38,932 |
| 22 | 9,542 | 10,600 | 12,338 | 14,041 | 16,314 | 18,101 | 21,337 | 24,939 | 27,301 | 30,813 | 33,924 | 37,659 | 40,289 |
| 23 | 10,196 | 11,293 | 13,091 | 14,848 | 17,187 | 19,021 | 22,337 | 26,018 | 28,429 | 32,007 | 35,172 | 38,968 | 41,638 |
| 24 | 10,856 | 11,992 | 13,348 | 15,659 | 18,062 | 19,943 | 23,337 | 27,096 | 29,553 | 33,196 | 36,415 | 40,270 | 42,980 |
| 25 | 11,524 | 12,697 | 14,611 | 16,473 | 18,940 | 20,867 | 24,337 | 28,172 | 30,675 | 34,382 | 37,652 | 41,566 | 44,314 |
| 26 | 12,198 | 13,409 | 15,379 | 17,292 | 19,820 | 21,792 | 25,336 | 29,246 | 31,795 | 35,563 | 38,885 | 42,856 | 45,642 |
| 27 | 12,879 | 14,125 | 16,151 | 18,114 | 20,703 | 22,719 | 26,336 | 30,319 | 32,912 | 36,741 | 40,113 | 44,140 | 46,963 |
| 28 | 13,565 | 14,847 | 16,928 | 18,928 | 21,586 | 23,647 | 27,336 | 321,391 | 34,027 | 37,916 | 41,337 | 45,419 | 48,278 |
| 29 | 14,256 | 15,574 | 17,708 | 19,768 | 22,475 | 24,577 | 28,336 | 32,461 | 35,139 | 39,087 | 42,557 | 46,693 | 49,588 |

**Таблиця значень *t* Стьюдента для *p*=0,05 і 0,01**

| Число ступеней свободи | *p*=0,05 | *p*=0,01 | Число ступеней свободи | *p*=0,05 | *p*=0,01 |
|---|---|---|---|---|---|
| 1 | 12,69 | 63,655 | 28 | 2,049 | 2,764 |
| 2 | 4,302 | 9,924 | 29 | 2,045 | 2,757 |
| 3 | 3,183 | 5,841 | 30 | 2,042 | 2,750 |
| 4 | 2,776 | 4,604 | 32 | 2,037 | 2,739 |
| 5 | 2,571 | 4,032 | 34 | 2,032 | 2,728 |
| 6 | 2,447 | 3,707 | 36 | 2,027 | 2,718 |
| 7 | 2,364 | 3,500 | 38 | 2,025 | 2,711 |
| 8 | 2,307 | 3,356 | 40 | 2,020 | 2,704 |
| 9 | 2,263 | 3,250 | 42 | 2,017 | 2,696 |
| 10 | 2,227 | 3,169 | 44 | 2,015 | 2,691 |
| 11 | 2,200 | 3,138 | 46 | 2,012 | 2,685 |
| 12 | 2,179 | 3,055 | 48 | 2,010 | 2,681 |
| 13 | 2,161 | 3,012 | 50 | 2,007 | 2,678 |
| 14 | 2,145 | 2,977 | 55 | 2,005 | 2,668 |
| 15 | 2,131 | 2,946 | 60 | 2,000 | 2,661 |
| 16 | 2,119 | 2,921 | 65 | 1,997 | 2,653 |
| 17 | 2,110 | 2,898 | 70 | 1,994 | 2,648 |
| 18 | 2,100 | 2,877 | 80 | 1,990 | 2,638 |
| 19 | 2,093 | 2,860 | 100 | 1,985 | 2,615 |
| 20 | 2,086 | 2,846 | 125 | 1,980 | 2,615 |
| 21 | 2,078 | 2,832 | 150 | 1,977 | 2,610 |
| 22 | 2,074 | 2,818 | 200 | 1,972 | 2,600 |
| 23 | 2,069 | 2,807 | 400 | 1,965 | 2,588 |
| 24 | 2,064 | 2,796 | 1 000 | 1,962 | 2,581 |
| 25 | 2,059 | 2,787 | | | |
| 26 | 2,054 | 2,778 | | | |
| 27 | 2,052 | 2,771 | | | |

Додаток 5

**Критерії значення *F*-критерія**

| $k_1$ / $k_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| рівень значимості $a$=0,05 | | | | | | | | | |
| 1 | 161,4 | 199,5 | 215,7 | 224,6 | 230,2 | 234,0 | 238,9 | 242,0 | 248,0 |
| 2 | 18,51 | 19,00 | 19,16 | 19,25 | 19,30 | 19,33 | 19,37 | 19,39 | 19,44 |
| 3 | 10,13 | 9,55 | 9,28 | 9,12 | 9,01 | 8,94 | 8,84 | 8,78 | 8,66 |
| 4 | 7,71 | 6,94 | 6,59 | 6,39 | 6,26 | 6,16 | 6,04 | 5,96 | 5,80 |
| 5 | 6,61 | 5,79 | 5,41 | 5,19 | 5,05 | 4,95 | 4,82 | 4,74 | 4,56 |
| 6 | 5,99 | 5,14 | 4,76 | 4,53 | 4,39 | 4,28 | 4,15 | 4,06 | 3,87 |
| 7 | 5,59 | 4,74 | 4,35 | 4,12 | 3,97 | 3,87 | 3,73 | 3,63 | 3,44 |
| 8 | 5,32 | 4,46 | 4,07 | 3,84 | 3,69 | 3,58 | 3,44 | 3,34 | 3,15 |
| 9 | 5,12 | 4,26 | 3,86 | 3,63 | 3,48 | 3,37 | 3,23 | 3,13 | 2,93 |
| 10 | 4,96 | 4,10 | 3,71 | 3,48 | 3,33 | 3,22 | 3,07 | 2,97 | 2,77 |
| 11 | 4,82 | 3,98 | 3,59 | 3,63 | 3,20 | 3,09 | 2,95 | 2,86 | 2,65 |
| 12 | 4,75 | 3,88 | 3,49 | 3,26 | 3,11 | 3,00 | 2,85 | 2,76 | 2,54 |
| 14 | 4,60 | 3,74 | 3,34 | 3,01 | 2,96 | 2,85 | 2,70 | 2,60 | 2,39 |
| 16 | 4,49 | 3,63 | 3,24 | 3,01 | 2,85 | 2,74 | 2,59 | 2,49 | 2,28 |
| 18 | 4,41 | 3,55 | 3,16 | 2,93 | 2,77 | 2,66 | 2,51 | 2,41 | 2,19 |
| 20 | 4,35 | 3,49 | 3,10 | 2,87 | 2,71 | 2,60 | 2,45 | 2,35 | 2,12 |
| 30 | 4,17 | 3,32 | 2,92 | 2,69 | 2,53 | 2,42 | 2,27 | 2,16 | 1,93 |
| 40 | 4,08 | 3,23 | 2,84 | 2,61 | 2,45 | 2,34 | 2,18 | 2,12 | 1,84 |
| 60 | 4,00 | 3,15 | 2,76 | 2,52 | 2,37 | 2,25 | 2,10 | 2,04 | 1,75 |
| 120 | 3,92 | 3,07 | 2,68 | 2,45 | 2,29 | 2,17 | 2,02 | 1,90 | 1,65 |
| | 3,84 | 2,99 | 2,60 | 2,37 | 2,21 | 2,09 | 1,94 | 1,83 | 1,57 |
| рівень значимості $a$=0,01 | | | | | | | | | |
| 1 | 4052 | 4999 | 5403 | 5625 | 5764 | 5859 | 5981 | 6056 | 6208 |
| 2 | 98,49 | 99,00 | 99,17 | 99,28 | 99,30 | 99,33 | 99,36 | 99,40 | 99,45 |
| 3 | 34,12 | 30,81 | 29,46 | 28,71 | 28,24 | 27,91 | 27,49 | 27,23 | 26,69 |
| 4 | 21,20 | 18,00 | 16,69 | 15,98 | 15,52 | 15,21 | 14,89 | 14,54 | 14,02 |
| 5 | 16,26 | 13,27 | 12,06 | 11,39 | 10,97 | 10,67 | 10,27 | 10,05 | 10,55 |
| 6 | 13,74 | 10,92 | 9,78 | 9,15 | 8,75 | 8,47 | 8,10 | 7,87 | 7,39 |
| 7 | 12,25 | 9,55 | 8,45 | 7,85 | 7,46 | 7,19 | 6,84 | 6,62 | 6,15 |
| 8 | 11,26 | 8,65 | 7,59 | 7,01 | 6,63 | 6,37 | 6,03 | 5,82 | 5,36 |
| 9 | 10,56 | 8,02 | 6,99 | 6,42 | 6,02 | 5,80 | 5,47 | 5,26 | 4,80 |
| 10 | 10,04 | 7,56 | 6,55 | 5,99 | 5,64 | 5,39 | 5,06 | 4,85 | 4,41 |
| 11 | 9,65 | 7,20 | 6,22 | 5,64 | 5,32 | 5,07 | 4,74 | 4,54 | 4,10 |
| 12 | 9,33 | 6,93 | 5,95 | 5,41 | 5,06 | 4,82 | 4,50 | 4,30 | 3,86 |
| 14 | 8,86 | 6,51 | 5,56 | 5,03 | 4,69 | 4,46 | 4,14 | 3,94 | 3,51 |
| 16 | 8,58 | 6,23 | 5,29 | 4,47 | 4,44 | 4,20 | 3,89 | 3,69 | 3,25 |
| 18 | 8,28 | 6,01 | 5,09 | 4,58 | 4,25 | 4,01 | 3,71 | 3,51 | 3,07 |
| 20 | 8,10 | 5,85 | 4,94 | 4,43 | 4,10 | 3,87 | 3,56 | 3,37 | 2,94 |
| 30 | 7,56 | 5,39 | 4,51 | 4,02 | 3,70 | 3,47 | 3,17 | 2,98 | 2,55 |
| 40 | 7,31 | 5,18 | 4,31 | 3,83 | 3,51 | 3,29 | 2,99 | 2,80 | 2,37 |
| 60 | 7,08 | 4,98 | 4,13 | 3,65 | 3,34 | 3,12 | 2,82 | 2,63 | 2,20 |
| 120 | 6,85 | 4,79 | 3,95 | 3,48 | 3,17 | 2,96 | 2,66 | 2,47 | 2,03 |
| | 6,54 | 4,60 | 3,78 | 3,32 | 3,02 | 2,80 | 2,51 | 2,32 | 1,87 |

# ІНФОРМАЦІЯ ПРО АВТОРСЬКИЙ КОЛЕКТИВ


| | |
|---|---|
| **Дмитрієв** Ілля Андрійович | д-р екон. наук, заслужений діяч науки і техніки України, академік Транспортної академії України, проректор з наукової роботи Харківського національного автомобільно-дорожнього університету |
| **Дмитрієва** Оксана Іллівна | д-р екон. наук, завідувач кафедри економіки і підприємництва Харківського національного автомобільно-дорожнього університету |
| **Гіржева** Ольга Миколаївна | д-р екон. наук, доцент кафедри менеджменту, бізнесу і адміністрування Державного біотехнологічного університету |
| **Непран** Андрій Володимирович | канд. екон. наук, доцент кафедри економіки і підприємництва Харківського національного автомобільно-дорожнього університету |
| **Бірченко** Наталія Олександрівна | канд. екон. наук, доцент кафедри обліку, аудиту та оподаткування Державного біотехнологічного університету |
| **Воронкова** Алла Анатоліївна | канд. екон. наук, доцент кафедри менеджменту, бізнесу і адміністрування Державного біотехнологічного університету |
| **Чуйко** Наталія Василівна | канд. екон. наук, доцент кафедри менеджменту, бізнесу і адміністрування Державного біотехнологічного університету |



Відповідальний за випуск *А. В. Непран*
В авторській редакції
Комп'ютерна верстка *А. В. Непран*
Підписано до друку 30.05.2022 р.

Формат 60× 84 1/16. Папір офісний. Друк цифровий.
Ум. друк. арк. 16,8.
Тираж 100 прим. Зам №

---

Видавництво та друк
ФОП Іванченко І. С.
пр. Тракторобудівників, 89-а/62, м. Харків, Україна, 61135
тел.: +038(050/093) 40-243-50.
Свідоцтво про внесення суб'єкта видавничої діяльності до державного реєстру видавців, виготовників та розповсюджувачів видавничої продукції
ДК № 4388 від 15.08.2012 р.
61002, Харків, вул.


---