

**Київський національний торговельно-економічний
університет**

СТАТИСТИКА

Підручник

Київ 2020

**Розповсюдження і тиражування без офіційного дозволу
КНТЕУ заборонено**

УДК 311 (075)

С 78

Автори: С. І. Пирожков, д-р екон. наук, проф. (розд. 1); В. В. Рязанцева, канд. фіз.-мат. наук, доц. (розд. 9, 10); Р. М. Моторин, д-р екон. наук, проф. (розд. 2); Н. А. Головач, канд. екон. наук, доц. (розд. 11); Н. О. Іванченко, канд. екон. наук, доц. (розд. 3); Я. О. Міщенко, канд. екон. наук, доц. (розд. 6); А. Ю. Чорний, канд. екон. наук, доц. (розд. 7, 8); О. А. Антонюк, ст. викл. (розд. 4, 5)

Рецензенти: І. Г. Манцуров, д-р екон. наук, професор, член-кореспондент НАН України, директор НВК «Інститут системних статистичних досліджень», завідувач кафедри статистики ДВНЗ «Київський національний економічний університет ім. Вадима Гетьмана»; Н. В. Ковтун, д-р екон. наук, проф. професор кафедри статистики та демографії Київського національного університету імені Тараса Шевченка;
О. К. Щетініна, д-р фіз.-мат. наук, проф., завідувачка кафедри вищої та прикладної математики Київського національного торговельно-економічного університету

*Рекомендовано до друку вченою радою Київського національного
торговельно-економічного університету
(протокол № 2 від 31 жовтня 2019 р.)*

Статистика : підручник / С. І. Пирожков, В. В. Рязанцева,
С 78 Р. М. Моторин та ін. – Київ : Київ. нац. торг.-екон. ун-т,
2020. – 328 с.

ISBN 978-966-629-986-7

DOI DOI: <http://doi.org/10.31617/p.knute.2020-164>

У підручнику викладено теоретичні та прикладні засади статистики, які допомагають набутти навички щодо оцінювання та аналізу економічних явищ за допомогою статистичних показників і методів. Висвітлено 11 тем, які містять основні поняття статистичної науки, визначені основні статистичні показники та їх економічна інтерпретація, наведено приклади розв'язання задач.

Призначено для студентів, аспірантів, наукових працівників, які вивчають статистику.

УДК 311 (075)

ISBN 978-966-629-986-7

© Пирожков С. І., Рязанцева В. В., Моторин Р. М.,
Головач Н. А., Іванченко Н. О., Міщенко Я. О.,
Чорний А. Ю., Антонюк О. А., 2020

© Київський національний торговельно-
економічний університет, 2020

ЗМІСТ

ВСТУП	7
Розділ 1. ПРЕДМЕТ, МЕТОД І ЗАВДАННЯ СТАТИСТИКИ	8
1.1. Предмет статистики. Становлення статистики як науки. Завдання статистики.....	8
1.2. Основні категорії статистики.....	12
1.3. Організація державної статистики в Україні	13
1.3. Питання для самоконтролю.....	23
Розділ 2. МЕТОДИ СТАТИСТИЧНОГО СПОСТЕРЕЖЕННЯ	24
2.1. Поняття про статистичне спостереження	24
2.2. Форми, види і способи статистичного спостереження	26
2.3. Програмно-методологічні питання статистичного спостереження.....	32
2.4. Помилки спостереження	34
2.5. Програмні інструменти для вилучення даних із зовнішніх джерел і обробки величезних масивів даних (Big data).....	36
2.6. Завдання для самостійної роботи.....	43
2.7. Питання для самоконтролю.....	46
Розділ 3. ЗВЕДЕННЯ ТА ГРУПУВАННЯ СТАТИСТИЧНИХ ДАНИХ	47
3.1. Поняття статистичного зведення та групування	47
3.2. Основні завдання і види групувань	51
3.3. Групування за кількісною ознакою. Правила створення груп та інтервалів групування	57
3.4. Метод вторинного групування.....	61
3.5. Ряди розподілу	65
3.6. Статистичні класифікації.....	72
3.7. Статистичні таблиці	76
3.8. Задачі для самостійної роботи.....	78
3.9. Питання для самоконтролю.....	86

Розділ 4. АБСОЛЮТНІ ТА ВІДНОСНІ СТАТИСТИЧНІ ВЕЛИЧИНИ	88
4.1. Суть і види статистичних показників	88
4.2. Абсолютні статистичні величини, одиниці їх вимірювання.....	91
4.3. Відносні статистичні величини, їх види та форми виразу.....	94
4.4. Задачі для самостійної роботи.....	99
4.5. Питання для самоконтролю.....	104
Розділ 5. СЕРЕДНІ ВЕЛИЧИНИ ТА ЗАГАЛЬНІ ПРИНЦИПИ ЇХ ЗАСТОСУВАННЯ	105
5.1. Сутність і значення середніх величин	105
5.2. Середня арифметична та її властивості.....	107
5.3. Середня гармонічна і квадратична.....	113
5.4. Середня геометрична і хронологічна.....	117
5.5. Поняття моди, медіани та їх застосування.....	120
5.6. Задачі для самостійної роботи.....	123
5.7. Питання для самоконтролю.....	129
Розділ 6. АНАЛІЗ ВАРІАЦІЇ ТА ФОРМИ РОЗПОДІЛУ	130
6.1. Поняття варіаційного ряду розподілу. Характеристики центру розподілу.....	130
6.2. Характеристики варіації	137
6.3. Види та взаємозв'язок дисперсій	142
6.4. Характеристики форми розподілу. Порівняльний аналіз структур	147
6.5. Задачі для самостійної роботи.....	153
Питання для самоконтролю.....	161
Розділ 7. ВИБІРКОВИЙ МЕТОД	162
7.1. Суть вибіркового спостереження.....	162
7.2. Види і способи вибіркового спостереження.....	163
7.3. Похибки вибірки.....	166
7.4. Оцінювання параметрів генеральної сукупності	170

7.5. Визначення мінімально необхідного обсягу вбіркової сукупності	175
7.6. Задачі для самостійної роботи.....	178
7.7. Питання для самоконтролю.....	181
Розділ 8. СТАТИСТИЧНА ПЕРЕВІРКА ГІПОТЕЗ.....	182
8.1. Статистичні гіпотези. Принципи їх перевірки та формулювання статистичного висновку	182
8.2. Перевірка гіпотез про дисперсії.....	187
8.3. Перевірка гіпотез про середні значення.....	192
8.4. Перевірка гіпотез про частки одиниць сукупності	206
8.5. Задачі для самостійної роботи.....	210
8.6. Питання для самоконтролю.....	212
Розділ 9. КОРЕЛЯЦІЙНО-РЕГРЕСІЙНИЙ АНАЛІЗ.....	213
9.1. Загальне поняття кореляційно-регресійного методу аналізу	213
9.2. Рівняння регресії і визначення його параметрів	218
9.3. Множинна кореляційно-регресійна модель	226
9.4. Задачі для самостійної роботи.....	229
9.5. Питання для самоконтролю.....	233
Розділ 10. РЯДИ ДИНАМІКИ ТА ЇХ АНАЛІЗ	235
10.1. Поняття ряду динаміки. Зіставленість і змикання рядів динаміки.....	235
10.2. Показники аналізу рядів динаміки.....	237
10.3. Середні показники аналізу динаміки.....	241
10.4. Дослідження динамічних та кореляційних властивостей статистичних показників. Прогнози за моделями динаміки.....	249
10.5. Задачі для самостійної роботи.....	257
10.6. Питання для самоконтролю.....	262

Розділ 11. ІНДЕКСИ	264
11.1. Суть і функції індексів	264
11.2. Методологічні основи побудови зведених індексів	268
11.3. Агрегатна форма індексів	271
11.4. Середньозважені індекси	274
11.5. Системи взаємозв'язаних індексів	278
11.6. Індекси середніх величин	284
11.7. Індексні ряди	291
11.8. Територіальні індекси	294
11.9. Задачі для самостійної роботи	296
11.10. Питання для самоконтролю	307
ДОДАТКИ	309
<i>Додаток 1.</i> Наближені формули для обчислення значень законів розподілу	309
<i>Додаток 2.</i> Таблиця стандартизованого нормального закону розподілу	317
<i>Додаток 3.</i> Таблиця χ^2 -розподілу Пірсона	318
<i>Додаток 4.</i> Таблиця t -розподілу Стьюдента	319
<i>Додаток 5.</i> Таблиці F -розподілу Фішера	320
СПИСОК ЛІТЕРАТУРИ	324

ВСТУП

Статистика є однією з базових дисциплін сучасної економічної освіти. Призначення статистики полягає у дослідженні соціально-економічних явищ та закономірностей за допомогою математико-статистичних методів обробки первинної інформації з метою обґрунтування ефективних управлінських та економічних рішень.

Статистика належить до дисциплін загальної підготовки студентів, які здобувають економічну освіту, формує професійні навички проведення статистичного аналізу, забезпечує підготовку висококваліфікованих дослідників-аналітиків.

Засвоєння статистики та методів аналізу розширює знання студентів щодо використання статистичного інструментарію у дослідженнях реальних економічних об'єктів на мікро- та макрорівнях, сприяє більш глибокому розумінню особливостей їх функціонування, засвоєнню методів вибору найбільш ефективних рішень.

Статистика – теоретико-прикладна економічна дисципліна, що являє собою методологію статистичного аналізу, дослідження поведінки економічних систем будь-якої складності, рівня організації та управління. Тому в навчальному процесі вона виконує роль «цементуючого матеріалу» щодо здобутих знань практично з усіх дисциплін, які викладаються на економічних факультетах майже усіх університетів.

Вивчення статистики сприяє формуванню у студентів нового економічного мислення, спрямованого на підготовку конкурентоспроможних фахівців на ринку робочої сили.

Підручник, що пропонується, орієнтовано на студентів економічних спеціальностей та буде корисним широкому колу фахівців, що цікавляться статистичними дослідженнями. Він охоплює усі теми, що викладаються на лекціях зі статистики в університеті. У підручнику розглянуто приклади застосування методів статистичного аналізу даних. Структуру задач складено так, що дозволяє досить повно і всебічно розглянути на практичних заняттях питання практичного застосування методів статистичного аналізу.

Розділ 1

ПРЕДМЕТ, МЕТОД І ЗАВДАННЯ СТАТИСТИКИ

План

- 1.1. Предмет статистики. Становлення статистики як науки. Завдання статистики.
- 1.2. Основні категорії статистики.
- 1.3. Організація державної статистики в Україні.
- 1.4. Питання для самоконтролю.

1.1. Предмет статистики. Становлення статистики як науки. Завдання статистики

Термін «статистика» походить (від лат. status – стан речей) – це синонім сукупності фактів, певних відомостей про соціально-економічні явища та процеси.

Історично перші статистичні операції такі як облік населення, визначення кількості чоловіків, придатних для військової служби, відомості для оподаткування населення здійснювалися ще у Стародавньому Єгипті, Індії, Китаї, Персії, Римській імперії. Це були примітивні статистичні операції й стосувалися небагатьох суспільних процесів і явищ. Поступово статистичний облік став охоплювати все більше і більше об'єктів, явищ й процесів суспільного життя, що обумовило вироблення загальних правил статистичної роботи та розробки методології кількісного аналізу.

Означення. Статистика – це суспільна наука, яка об'єднує принципи та методи оброблення масових числових даних, що характеризують різні суспільні явища. Для цього статистика розробляє методи збирання, систематизації, аналізу, інтерпретації та відображення результатів спостережень масових явищ та процесів з метою виявлення наявних у них закономірностей.

Закономірності характеризують *повторюваність, послідовність* та *порядок* в явищах та процесах, які по-різному проявляються у природі та суспільстві. Якщо одні мають прояв у кожному окремо взятому *одиночному* явищі, то інші – лише *в масі*, у великій кількості одиниць спостереження – у великих сукупностях. Перший вид закономірностей має назву – *динамічних*, а другий – *статистичних*.

Глибокий аналіз передбачає вивчення не лише *кількісної* сторони явища, яка виражає його зовнішнє проявлення, але й дозволяє викрити причини, що формують закономірності та механізми дії цих причин, тобто пізнати природу та суть закономірностей.

У понятті «закономірність» знаходить відображення об'єктивний зв'язок між причинами та наслідками, який по-різному проявляється у динамічних та статистичних закономірностях. Для динамічної закономірності характерна жорстка, механічна причинність, яка конкретно визначає, детермінує поведінку кожної одиниці сукупності й підпадає під дію *функціонально* обумовленого зв'язку. Під це підпадають закони природи, які описують фізичні закони, закони механіки.

У статистичних закономірностях знаходить прояв лише *визначена* правильність, послідовність у явищах, яка не має функціонально обумовленого зв'язку, оскільки може бути обумовлена впливом цілої низки багатьох причин й складатися під впливом не одного, а сукупності законів.

Нагромадження масових даних й необхідність їх узагальнення, підвищення попиту на інформацію заклали підвалини статистики як науки, яка почала формуватися в країнах Західної Європи у XVII–XVIII ст. Із розвитком математики, передусім теорії ймовірностей, удосконалювалися методи статистичного аналізу й розширювалася сфера їх використання. У XX ст. статистичні методи було запроваджено майже в усі галузі знань. Статистика використовується під час відтворення населення та його життєвого рівня, громадської думки, економічного та зовнішньо-економічного

розвитку країн, фінансових ризиків, у страхуванні, метрології тощо.

Отже, об'єктами статистичного аналізу можуть бути найрізноманітніші явища й процеси суспільного життя. *Предметом* статистики є розміри і кількісні співвідношення між масовими суспільними явищами, закономірністю їх формування, розвитку, взаємозв'язку. Таким чином, статистика вивчає, по-перше, *кількісну* сторону суспільних явищ, а по-друге, вона вивчає *масові* явища в їх нерозривному якісному взаємозв'язку.

Вивчаючи кількісну сторону явищ, статистика відображає її у відповідних *числах-показниках* і саме цим характеризує конкретну міру явищ, встановлює загальні властивості, виявляє схожість й різницю окремих рис, об'єднує елементи в групи, виявляє структуру і певні типи явищ.

Важливо підкреслити, що вивчення кількісної сторони суспільних явищ нерозривно пов'язане з їх *якісним* змістом. Адже кількісна розмірність не існує без якісної визначеності.

Явища суспільного життя є динамічними, вони безперервно розвиваються і змінюються як *в часі*, так і *в просторі*. З часом змінюються розміри явищ, співвідношення й пропорції. Їх значення є різними для окремих об'єктів, регіонів тощо. А відтак кількісну сторону суспільних явищ статистика має вивчати в конкретних умовах простору і часу. Інша особливість предмета статистики пов'язана з масовістю суспільних явищ, які мають ймовірнісну основу й описуються математичними формулами теорії ймовірностей. Статистика вивчає явища, які повторюються у просторі або з плином часу.

Для масового явища характерна участь у ньому певної *множини елементів*, істотні властивості яких схожі. Наявність будь-яких властивостей у окремого, поодинокого елемента – *випадковість*. Тільки-но значна кількість елементів об'єднується в одне ціле, сукупна дія випадковостей дає результат, практично незалежний від випадку.

Розглядаючи суспільні явища як масові й спираючись на облік усієї сукупності фактів, що належать до цих явищ,

статистика за допомогою чисел показує ступінь їх розвитку, напрям і швидкість змін, щільність взаємозв'язків і взаємозалежностей. Усе це дає підстави стверджувати, що статистика – могутній засіб пізнання складного суспільного життя.

Завдання статистики – розроблення, узагальнення й аналіз достовірних статистичних даних. Щоб статистична інформація була достовірною, її потрібно збирати й узагальнювати на науковій основі. Наукова система статистики охоплює: статистичну теорію, статистичну методологію, підсумкові результати статистичних досліджень.

Статистична теорія – це загальне вчення про обсяг суспільних явищ і статистичні показники, які їх характеризують.

Статистична методологія – це сукупність статистичних методів дослідження, тобто способів вивчення зміни обсягу суспільних явищ.

Підсумкові результати статистичних досліджень – це сукупність науково обґрунтованих статистичних даних, одержаних в результаті їх збирання, зведення й аналізу.

Статистика розробляє свої, специфічні способи вивчення явищ суспільного життя, основні з них такі:

- масове статистичне спостереження;
- зведення матеріалів;
- групування;
- застосування узагальнюючих й аналітичних показників;
- побудова балансів.

У своїй сукупності та єдності ці способи складають *метод статистичної науки*.

Як суспільна наука статистика не може розвиватися окремо від теоретичних наук про суспільство, зокрема економічної теорії, галузевих економічних дисциплін та соціології. Спираючись на суть, якісну природу явищ, через узагальнення масових даних, статистика вивчає характер і дію основних законів у реальному житті, робить прогностичні розрахунки для обґрунтування напрямів економічної політики держави.

1.2. Основні категорії статистики

Об'єкт вивчення статистики – суспільство, явища та процеси суспільного життя. Для вивчення кількісного аспекту масових суспільних явищ у статистиці використовують поняття закономірності та статистичної сукупності.

Статистична закономірність – це форма закономірності, коли будь-яке правило, закон виявляються лише у великій кількості елементів сукупності, виражаються тільки в масі явищ не одним, а цілою низкою показників.

Статистична сукупність – це масова кількість явищ, процесів, об'єктів, однорідних за своїми якісними ознаками, які поділяють на *два види*: однорідні та різнорідні.

Однорідною статистичною сукупністю називається така сукупність об'єктів, у яких одна чи кілька досліджуваних ознак є спільними, наприклад, магазини (спільні ознаки – товарооборот, торговельна площа, витрати обігу).

Різнорідною називається така сукупність, яка складається з різних явищ, наприклад: магазини (продовольчі, непродовольчі); банки (іпотечні, ощадні, земельні тощо).

Статистична сукупність складається з окремих елементів, або одиниць.

Одиниця сукупності – це первинний неподільний елемент досліджуваного явища чи процесу, за яким можна виявити якісну визначеність процесу чи явища.

Кожна одиниця сукупності має свої властивості, аспекти, *ознаки*, за якими їх групують для визначення статистичної сукупності. Ознаки також поділяють на атрибутивні (якісні) та кількісні.

Атрибутивними називають ознаки, які не можна виразити числовою величиною (освіта, стать, національність, назва товару тощо), а *кількісними* – такі, які можна виразити числовою величиною (вага, ціна, вартість, вік, відстань,

тощо). У свою чергу, кількісні ознаки поділяють на дискретні (переривчасті) та безперервні.

Дискретними називають кількісні ознаки, які набувають лише цілих значень, *безперервними* – такі, що можуть набувати як цілих, так і дробових значень (вартість продукції, ціна, собівартість тощо).

Ознаки поділяють також на факторні та результативні.

Факторними називають ознаки, які за певних умов зумовлюють зміну іншої ознаки. Наприклад, середня ціна залежить від вартості товарів та їх кількості. У цьому разі вартість товарів та їх кількість – це факторні ознаки, зміна яких зумовлює зміну середньої ціни.

Результативною називають ознаку, зміна якої залежить від зміни іншої (собівартість одиниці продукції, рентабельність, продуктивність праці тощо).

1.3. Організація державної статистики в Україні

Інституційні аспекти

Завданням Державної служби статистики України (Держстат) є розвиток статистичної системи в країні, включаючи такі напрями:

- визначення державної політики у сфері статистики;
- координація статистичної діяльності в країні;
- збір, обробка, зберігання, захист, розповсюдження і аналіз статистичної інформації;
- забезпечення достовірності й об'єктивності статистичних даних; розроблення, впровадження і вдосконалення статистичної методології; розроблення, впровадження і вдосконалення системи статистичних класифікацій;
- розроблення і ведення Єдиного державного реєстра підприємств і організацій.

Держстат має належну законодавчу базу в сфері статистики. Ухваленням у 1992 році Закону України «Про державну статистику» було створено правову основу її розвитку і стало початком реформування національної статистики. Закон включає основні положення про незалежність статистики і реорганізацію статистичної системи з урахуванням потреб ринкової економіки, визначає відповідальність Держстату за організацію і функціонування статистичної системи країни загалом і забороняє втручання в діяльність Держстату інших урядових органів, юридичних осіб і громадськості.

Новий Закон України «Про внесення змін до Закону України «Про державну статистику», прийнятого в липні 2000 року, ґрунтується на основоположних принципах офіційної статистики ООН. У Законі враховані вимоги європейського модельного закону і містяться положення про незалежність Держстату і конфіденційність інформації. Положення про конфіденційність даних не завжди дотримуються на практиці і можуть порушуватися у випадках, коли органи державного управління запрошують індивідуальні дані для адміністративних цілей. Подальше вдосконалення законодавства у сфері статистики здійснюватиметься в напрямі гармонізації законів, що діють в контексті змін, внесених до Закону про статистику. Наприклад, Кабінет міністрів України видав Постанову від 23 вересня 2014 р. № 481 «Про затвердження Положення про Державну службу статистики України» (із змінами і доповненнями, внесеними постановою Кабінету Міністрів України від 27 грудня 2018 року № 1138).

Статистична система складається з *трьох рівнів*: центральний апарат, регіональні управління і місцеві відділи статистики, а також включає низку функціональних органів статистики, зокрема – Державна академія статистики, обліку і аудиту, Державне підприємство «Інформаційно-аналітичне агентство» (Держаналітінформ).

Розділ 1. ПРЕДМЕТ, МЕТОД І ЗАВДАННЯ СТАТИСТИКИ

На сьогодні структура центрального апарату Держстату складається з 17 департаментів, кожен з яких має декілька відділів.

Організаційна структура апарату Держстату



Рис. 1.1. Організаційна структура центрального апарату Державної служби статистики України

Джерело: <http://www.ukrstat.gov.ua/>

Другий рівень статистичної системи представлений 28 регіональними управліннями статистики і складається з Головного міжрегіонального управління статистики у м. Києві (реорганізований в кінці 2000 р. як колишній обчислювальний центр), Головного управління статистики в Автономній Республіці Крим, 24 обласних управлінь статистики, а також управлінь в містах Києві і Севастополі. Регіональні управління статистики відповідають за збір, перевірку і передачу даних в Держстат, а також за розповсюдження цих даних на територіальному рівні.

Третій рівень – це місцеві органи державної статистики, що складаються з 587 відділів статистики районного і міського рівня.

Держстат фінансується за рахунок державного бюджету, які виділяються на проведення обстежень, включених в щорічний план державних статистичних спостережень. Для впровадження нових статистичних методів і адаптації міжнародних стандартів залучаються зовнішні джерела фінансування і реалізації відповідних програм технічної допомоги. На тлі недостатнього виділення бюджетних коштів розширюється діяльність Держстату в сфері надання послуг на платній основі.

Держстат послідовно удосконалює свою орієнтацію на користувача. Це проведення семінарів з різних питань статистики, як, наприклад, національні рахунки, статистика промисловості і обстеження домогосподарств. Основним користувачем статистичної інформації залишається Уряд.

Разом з тим, інші групи користувачів також залучаються до участі в процесі планування статистичних робіт і до вивчення попиту на статистичну інформацію. Поступово поліпшується зміст публікацій, їх тематика, збільшується кількість публікацій, доступних для користувачів. Щорічно видається каталог статистичних публікацій Держстату, з якого користувачі отримують інформацію про майбутні випуски. Після створення Державного агентства по розповсюдженню статистичної інформації, що знаходиться у сфері управління Держстату, робота в цій сфері була покращена. Зроблені перші кроки в розповсюдженні інформації за допомогою електронних засобів. Розробляється політика в сфері маркетингу статистичної інформації. Крім Уряду увага приділяється також іншим користувачам як в плані вивчення попиту на інформацію, так і розповсюдження даних.

Статистична інформація ґрунтується на єдиній системі статистичних показників, які обчислюються за єдиними

рекомендаціями, розробленими і схваленими міжнародними організаціями як міжнародні статистичні стандарти.

Для координації статистичної діяльності агентств ООН використовується класифікація програм ООН (статистична частина). Вона була прийнята Адміністративним комітетом системи ООН на засіданні Підкомісії з координації статистичної діяльності у 1999 році (див. веб-сторінку Статистичної комісії ООН – <http://www.un.org>).

КЛАСИФІКАЦІЯ СТАТИСТИКИ ТА СТАТИСТИЧНОЇ ДІЯЛЬНОСТІ ООН

030 Теорія статистики.

031 Демографічна і соціальна статистика.

1. Демографічна статистика (4.11).
2. Статистика населення:
 - а) перепис населення (4.5);
 - б) статистика груп населення, які мають особливі інтереси (4.12).
3. Статистика житлового господарства:
 - а) перепис житла (4.5);
 - б) інші напрями статистики житлового господарства (4.12).
4. Статистика праці (4.10).
5. Статистика освіти і навчання (4.8).
6. Статистика культури, включаючи засоби масової інформації (4.9).
7. Статистика доходів і витрат домогосподарств та їх розподіл (4.13).
8. Статистика соціального страхування (4xx).
9. Статистика охорони здоров'я (4.6).
10. Інші соціальні, демографічні та пов'язані з ними статистики (4.xx).

032 Економічна статистика.

1. Національні рахунки (3.1).
2. Статистика сільського господарства (3.10).
3. Статистика лісового господарства і рибальства (3.10).

4. Статистика промисловості (оперативно-технічна статистика) (3.11).

5. Статистика енергетики (3.12).

6. Статистика внутрішньої торгівлі (оптово-роздрібної) (3.5).

7. Статистика міжнародної торгівлі:

а) торгівля товарами;

б) торгівля послугами.

8. Статистика транспорту (3.4).

9. Статистика зв'язку (3.4).

10. Статистика туризму (3.7).

11. Статистика послуг, яка не включена в інші розділи (3.8).

12. Статистика грошово-кредитна, фінансів і страхування (3.2).

13. Податкова статистика (3.2).

14. Статистика платіжного балансу (3.2).

15. Статистика цін (3.9).

16. Статистика науки, технологій і патентів (3.13).

17. Інші галузі економічної статистики (3xx).

033 Статистика природних ресурсів і довкілля.

1. Статистика природних ресурсів і довкілля (5.1 і 5.3).

2. Облік природних ресурсів і довкілля (5.2).

3. Метеорологічна статистика (5xx).

034 Види діяльності, які не класифіковані як сфера діяльності.

1. Організація і надання статистичних послуг (1).

2. Технічна інфраструктура та інші пов'язані з цим питання (2).

3. Поширення і підтримка секретарської діяльності (6).

4. Інші види діяльності, які не класифіковані як сфера діяльності (x.xx).

Запис у круглих дужках після кожного пункту найнижчого рівня класифікації відсилає до відповідного елементу в класифікації, прийнятій Конференцією європейських статистиків під егідою ООН для використання в інтегрованому

вигляді статистичних програм Економічної комісії для Європи (ECE), Організації економічного співробітництва і розвитку (OECD) і Статистичного управління Європейського Союзу (EUROSTAT).

Під час планування статистичних робіт враховуються різноманітні запити, що надходять як від Уряду, так і з боку інших користувачів. Цю діяльність координує департамент методології і планування Держстату а також проводяться міжвідомчі консультації. Річні плани ґрунтуються на довгостроковій програмі і затверджуються Кабінетом Міністрів після проведення широкого обговорення зі всіма міністерствами і відомствами, які складають статистичну систему країни. Проте, існують певні статистичні обстеження, що проводяться іншими міністерствами і відомствами. Основна проблема виконання річних планів робіт полягає в їх постійному недофінансуванні з державного бюджету. Це значно ускладнює процес планування, і впливає на його ефективність відносно чіткого визначення цілей, які повинні бути досягнуті за встановлений період часу.

Держстат виступає як координатор статистичної системи в Україні. В основному, координацію здійснює Колегія Держстату, куди входять співробітники Держстату і представники наукових організацій. Держстат також має доступ до реєстрів та інших джерел адміністративних даних. Проте, слабкі зв'язки і недостатня взаємодія між різними органами державного управління вимагають посилення координуючої ролі Держстату.

Всього в системі державної статистики працює майже 8 тисяч фахівців. Станом на 01.01.2019 штатними розписами органів державної статистики передбачено 8375 посад, фактично працює 7603 особи. Питома вага жінок, які працюють на посадах державних службовців в органах державної статистики, становить 92,4% їх загальної кількості. Щодо освітнього рівня персоналу, то вищу освіту ступеня магістра мають 5246 осіб, або 93,5% загальної кількості державних службовців. Вищу освіту ступеня молодшого бакалавра та

бакалавра мають 358 державних службовці, або 6,4%. Частка державних службовців із вищою освітою, які мають спеціальність за дипломом, що відповідає визначеним у посадових інструкціях вимогам, становить 92,1%. Однак питома вага фахівців зі статистики становить 14,8% загальної кількості держслужбовців із вищою освітою. При цьому 1075 держслужбовців (19,2% їх загальної кількості) мають дві або більше вищі освіти, 16 державних службовців мають науковий ступінь доктора філософії. Крім того, у функціональних органах державної статистики працює 23 доктори наук та 44 докторів філософії (кандидатів наук).

Загальна чисельність персоналу є досить значною порівняно із статистичними службами інших країн Центральної і Східної Європи, але це пов'язано з неадекватним рівнем розвитку інфраструктури в територіальних управліннях, із-за чого потрібно більше людських ресурсів для обробки даних. Скорочення персоналу системи органів державної статистики можливо тільки після створення необхідної технічної інфраструктури, рівень якої нині достатньо низький, особливо в регіонах. Необхідна тісна взаємодія між регіональним і центральним рівнями для якіснішого збору і передачі даних.

Держстат здійснює заходи у сфері навчання для розширення знань фахівців про статистичні методи і розвиток мовних навичок завдяки підготовці фахівців-мультиплікаторів з різних сфер статистики. Це дозволяє вирішити проблему охоплення більшої кількості статистиків найперше з регіонів і провести в найближчому майбутньому їх навчання або перенавчання.

За останні роки в практику роботи Держстату був упроваджений певний вид колективного ухвалення рішень, так званий «командний підхід», проте рішення в Держстаті все ще ухвалюються централізовано.

Відповідно до покладених на Держстат обов'язків щодо розроблення і розповсюдження об'єктивної, достовірної і своєчасної статистичної інформації, що відображає

різні процеси в економічному і соціальному житті країни, статистичні дані надаються різним групам користувачів: органам державного управління (найбільша група), науково-дослідним організаціям, засобам масової інформації і громадськості. Потреба користувачів у статистичній інформації постійно зростає.

Держстат щомісячно готує доповідь і статистичний бюлетень про соціально-економічний розвиток України, а також оперативну статистичну інформацію. Всі публікації видаються українською мовою. Є низка двомовних публікацій (англійський/український), деякі з яких містять резюме, зміст і заголовки таблиць на англійському (наприклад, «Статистичний щорічник», «Україна в цифрах», «Україна і країни СНД»). У 2000 році випущена сумісна публікація «Україна і Німеччина» на українській і німецькій мовах. Національна академія статистики, обліку та аудиту щокварталу видає журнал «Статистика України». Щорічно публікується збірка наукових праць «Проблеми статистики». Більшість публікацій містять актуальну інформацію. Дані по ВВП, безробіттю і інфляції публікуються щомісячно.

Статистична інфраструктура

У 1997 р. Світовий банк надав Уряду України кредит для інституційного розвитку. Частина коштів кредиту було виділено Держстату на реформування статистики, завдяки чому було придбано комп'ютерну техніку та інше устаткування. Протягом п'яти років був здійснений перехід від великих ЕОМ до використання персональних комп'ютерів і локальних мереж. Нині обробка статистичної інформації здійснюється на персональних комп'ютерах.

Такі елементи інфраструктури, як статистичний реєстр підприємств і класифікації ще не повністю впроваджені. Цей процес триває декілька років, що перешкоджає впровадженню нових методів збору даних у сфері статистики підприємств. Переведені і адаптовані до умов України найбільш важливі європейські класифікації, такі як: NACE, CPA і PRODCOM. З 2001 року використовується для проведення статистичних

обстежень Класифікацію видів економічної діяльності (КВЕД), розроблено на основі NACE; CPA і PRODCOM. У сфері статистики зовнішньої торгівлі з 1 липня 2001 року упроваджено в практику затверджену Парламентом Українську класифікацію товарів зовнішньоекономічної діяльності (УКТВЕД), що базується на версії Гармонізованої Системи 1996 року.

Створені і розвиваються короткострокова і структурна статистика. Достатньо розвинена соціальна статистика, що найперше стосується обстежень бюджетів домашніх господарств і робочої сили. Сформована основна сукупність домашніх господарств, у практику роботи запроваджуються вибіркові методи. По всій країні загалом налічується 510 інтерв'юерів. Для проведення обстеження робочої сили і нового обстеження умов життя домашніх господарств було створено єдину мережу інтерв'юерів і тим самим, одні і ті ж інтерв'юери беруть участь в обох обстеженнях. Цей підхід також гарантує ефективність підготовки інтерв'юерів. Ті ж самі інтерв'юери беруть участь і в деяких інших обстеженнях. Таким чином, Держстат використовує їх упродовж всього року при проведенні різних обстежень домашніх господарств.

Певний прогрес досягнуто у сфері національних рахунків, проте є потреба подальшого вдосконалення концепцій і підвищення якості даних. Надалі особливу увагу Держстат приділятиме таким сферам: обхвату і якості реєстра підприємств, розвитку інформаційних технологій, статистиці послуг і транспорту, вдосконаленню якості статистичних даних.

Основними елементами майбутнього розвитку інформаційних технологій є впровадження і використання сучасних інформаційних технологій, програмних продуктів і технічних засобів для автоматизації процесів збору, обробка зберігання, аналізу і розповсюдження статистичної інформації, а саме:

- створення і ведення баз даних статистичних показників;

- створення і ведення систем довгострокового зберігання статистичних даних і забезпечення доступу до них;
- збір даних в електронному вигляді;
- удосконалення системи обміну статистичною інформацією між різними рівнями її обробки;
- використання статистичних пакетів;
- розповсюдження статистичної інформації за допомогою Інтернету; забезпечення захисту і конфіденційності статистичних даних у процесі їх автоматизованої обробки і зберігання;
- забезпечення ефективного функціонування локальних обчислювальних мереж в органах державної статистики.

1.4. Питання для самоконтролю

1. Навести означення предмета статистики.
2. Що характеризують закономірності статистики?
3. Що таке функціонально обумовлений зв'язок?
4. Навести означення предмета статистики.
5. Дати визначення завдань статистики.
6. Що таке статистична теорія?
7. Що таке статистична методологія?
8. Що таке підсумкові результати статистичних досліджень?
9. Які бувають способи вивчення явищ суспільного життя?
10. Що таке статистична закономірність?
11. Що таке статистична сукупність?
12. Назвати види статистичної сукупності.
13. На які групи поділяються ознаки сукупності?
14. Що таке атрибутивна ознака?
15. Що таке факторна ознака?
16. Що таке дискретна ознака?
17. Що таке результативна ознака?

Розділ 2

МЕТОДИ СТАТИСТИЧНОГО СПОСТЕРЕЖЕННЯ

План

- 2.1. Поняття про статистичне спостереження.
- 2.2. Форми, види і способи статистичного спостереження.
- 2.3. Програмно-методологічні питання статистичного спостереження.
- 2.4. Помилки спостереження.
- 2.5. Програмні інструменти для вилучення даних із зовнішніх джерел і обробки величезних масивів даних (Big data).
- 2.6. Завдання для самостійної роботи.
- 2.7. Питання для самоконтролю.

2.1. Поняття про статистичне спостереження

Статистична інформація є необхідною частиною процесу прийняття управлінських рішень. Важливим етапом будь-якого статистичного дослідження є статистичне спостереження.

Означення. Статистичне спостереження – це плано-мірне, науково організоване спостереження за масовими явищами, яке полягає в реєстрації відібраних ознак у кожній одиниці сукупності.

Статистичне спостереження проводиться органами державної статистики, спеціалізованими інформаційними агентствами, науково-дослідними інститутами, підприємствами, банками, біржами, фірмами.

Під час підготовки спостережень як правило ставляться такі питання:

1. Якого типу повинна бути інформація?
2. Який спосіб потрібно використати для збору і зведення інформації?
3. У який спосіб опрацювати зібрану інформацію і зробити відповідні висновки?

Розділ 2. МЕТОДИ СТАТИСТИЧНОГО СПОСТЕРЕЖЕННЯ

Статистичні спостереження можна поділити умовно на:

- *первинні* спостереження спираються на реєстрації даних безпосереднього об'єкта (приклад продукції підприємства і т.п.)
- *вторинні* спостереження спираються на використанні раніше зібраних даних, наприклад банківські звіти, статистичні щорічники і т.п.

Головні вимоги до статистичних даних:

- правдивість і регулярність;
- комплектність;
- актуальність даних;
- однорідність, порівнюваність даних, тотожність одиниць виміру даних;
- доступність даних.

Отримання статистичних даних передбачає підготовку і планування спостереження, реєстрацію даних, складання бази даних. Кожен етап вимагає опрацювання певного методу. На початку роботи потрібно визначити:

- мету статистичного дослідження;
- об'єкт спостереження;
- властивості об'єкта спостереження;
- одиниці спостереження.

Послідовність статистичного дослідження наведено на рис. 2.1.

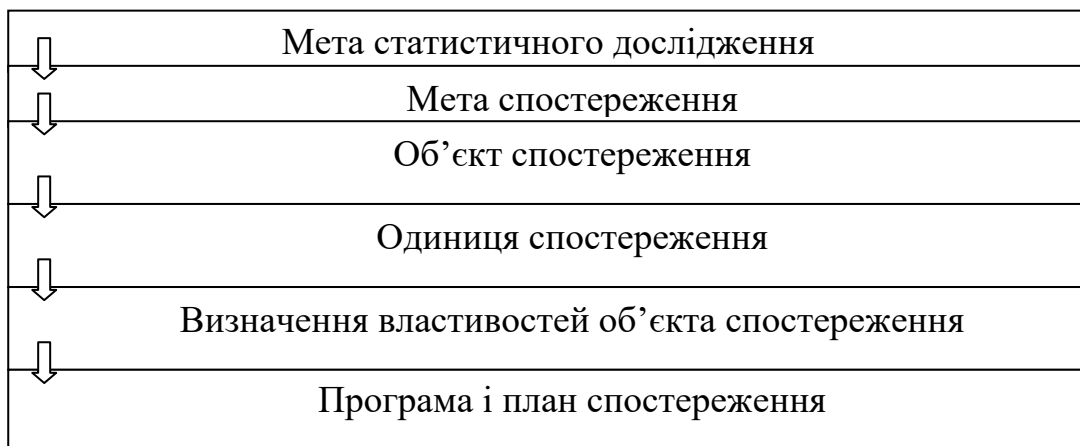


Рис. 2.1. Послідовність статистичного дослідження

Джерело: розроблено автором.

Істотним елементом дослідження є програма і план спостереження, де визначаються інструменти статистичних досліджень (збір законів, інструкції, технічні інструменти). Приклад організації статистичного дослідження наведено на рис. 2.2.

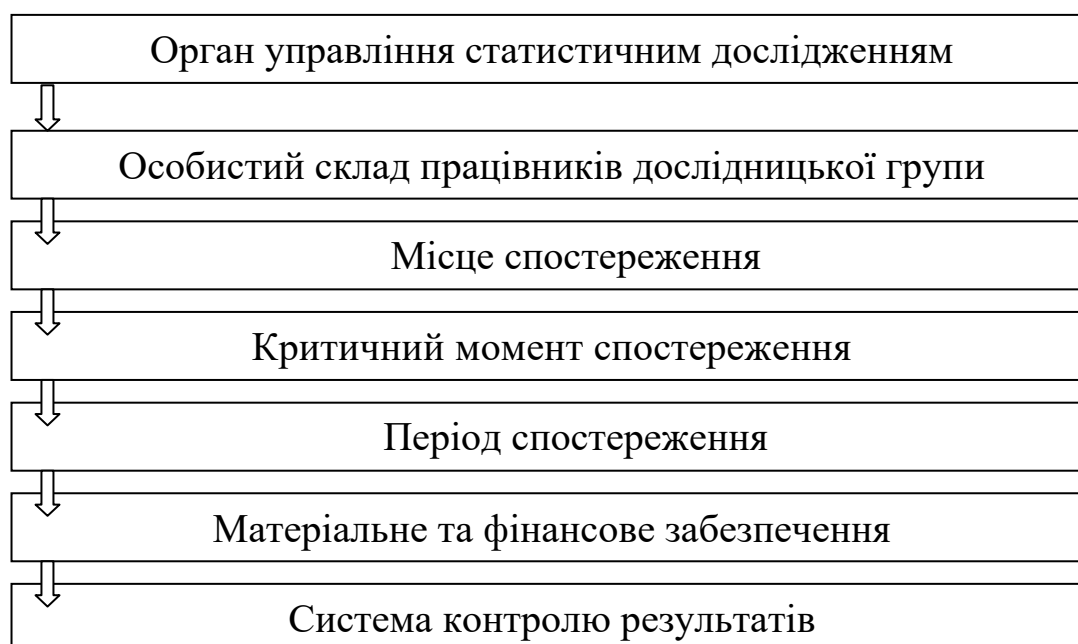


Рис. 2.2. Організація статистичного дослідження

Джерело: розроблено авторами.

Точність даних – це основна вимога, що висувається під час статистичного спостереження.

2.2. Форми, види і способи статистичного спостереження

Форми статистичного спостереження. У вітчизняній статистиці використовують три організаційні форми статистичного спостереження:

- Статистична звітність (підприємств, організацій, установ і т. п.);

- спеціально організовані статистичні спостереження суцільного і несуцільного характеру;
- реєстри.

Статистична звітність – це форма статистичного спостереження, за допомогою якої статистичні органи в певні терміни отримують від підприємств, установ і організацій необхідні дані у вигляді встановлених в законному порядку звітних документів, що містять статистичні відомості про роботу підприємства. Вона має обов'язковий характер, а також документальну обґрунтованість, оскільки всі дані базуються на документах первинного обліку.

Статистична звітність буває: щоденна, тижнева, двотижнева, місячна, квартальна і річна. Окрім річної звітності всі перераховані види є поточною звітністю.

Спеціально організоване спостереження проводиться з метою отримання відомостей, відсутніх у звітності, або для перевірки її даних. Найбільш простим прикладом такого спостереження є перепис. Держслужба статистики України проводить переписи населення, матеріальних ресурсів, багаторічних насаджень, невстановленого устаткування, незавершеного будівництва; устаткування та ін.

Перепис – це спостереження, що повторюється, як правило, через рівні проміжки часу з метою отримання даних певних ознак про чисельність, склад і стан об'єкта статистичного спостереження.

Характерними особливостями перепису є: одночасність проведення її на всій території, яка повинна бути охоплена обстеженням, єдність програми спостереження; реєстрація всіх одиниць спостереження за станом на один і той же критичний момент часу.

Окрім переписів статистика проводить й інші спеціально організовані спостереження, зокрема бюджетні обстеження, які характеризують структуру споживчих витрат і доходів домогосподарств.

Реєстри – це форма безперервного статистичного спостереження за довготривалими процесами у межах конкретного

часу. Воно ґрунтується на веденні статистичного реєстру. Реєстр є таким інструментом, що постійно стежить за станом одиниці спостереження. У реєстрі кожна одиниця спостереження характеризується сукупністю показників. Одні з них залишаються незмінними протягом всього часу спостереження і реєструються один раз; інші показники, періодичність зміни яких є невідомою, оновлюються з настанням зміни; треті – це часові ряди показників із заздалегідь відомим періодом оновлення. Всі показники зберігаються до повного завершення спостереження за одиницею обстежуваної сукупності.

У практиці статистики розрізняють реєстри населення і реєстри підприємств.

Реєстр населення – пойменованій перелік жителів країни, що регулярно актуалізується. Програма спостереження обмежена загальними ознаками, такими, як стать, дата і місце народження, дата вступу до шлюбу і шлюбного стану.

Реєстр підприємств нараховує всі види економічної діяльності і містить значення основних ознак по кожній одиниці спостережуваного об'єкта за певний період або момент часу. Реєстри підприємств містять дані про час створення (реєстрації) підприємства, його назва і адреса, телефон, про організаційно-правову форму, структуру, вид економічної діяльності, кількість зайнятих (цей показник відображає розмір підприємства) та ін.

Йому відводиться важливе значення у впровадженні системи національних рахунків в статистичну практику.

Єдиний державний реєстр підприємств і організацій всіх форм власності України (ЕДРПОУ) дає можливість організувати суцільне спостереження по обмеженому колу статистичних показників підприємств, зареєстрованих на території України, дозволяє отримати безперервні ряди показників у разі зміни територіальної, галузевої та інших структур сукупності.

У реєстр заносяться дані по всіх підприємствах, організаціях, установах і об'єднаннях незалежно від їх форм влас-

ності, включаючи підприємства з іноземними інвестиціями, банківські установи та інші юридичні особи.

Інформаційний фонд реєстру містить: реєстровий код суб'єкта; дані про галузеву, територіальну приналежність суб'єкта, його підпорядкування; довідкові відомості (прізвища керівників, адреси, номери телефонів, факсів і т. ін., відомості про засновників), економічні показники. Реєстр містить дані про такі показники: середньоспискова чисельність працівників; ресурси, що направляються на споживання; залишкова вартість основних засобів; балансовий прибуток (збиток); статутний фонд. Оскільки реєстр ведеться по окремих територіях, регіональні статистичні служби можуть розширювати склад економічних показників у разі потреби.

ЕДРПОУ дозволяє робити відбір і угруповання будь-якої сукупності одиниць поодинці або декількох ознаках.

Види статистичного спостереження. Статистичні спостереження можна поділити на групи за такими ознаками:

- час реєстрації фактів;
- обхват одиниць сукупності.

За часом реєстрації фактів буває поточне, періодичне і одноразове спостереження. При поточному спостереженні зміни відносно явищ, що вивчаються, реєструються у міру їх настання, наприклад при реєстрації підприємства. Таке спостереження проводиться з метою вивчення динаміки явища.

До періодичних спостережень відноситься реєстрація цін виробників по окремих товарах, яка проводиться щомісячно.

Одноразове обстеження дає зведення про кількісні характеристики якого-небудь явища або процесу у момент його дослідження. Повторна реєстрація проводиться через якийсь час (не визначений заздалегідь) або може не проводитися взагалі. Наприклад, маркетингові дослідження.

По кількості одиниць сукупності статистичне спостереження буває суцільне і не суцільне.

Суцільне спостереження полягає в отриманні інформації про всі одиниці досліджуваної сукупності.

Такий вид спостереження має недоліки: високу вартість отримання і обробки всього об'єму інформації, недостатню оперативність інформації.

Несуцільне спостереження припускає, що обстеженню підлягає лише частина одиниць сукупності, що вивчається. Однією з переваг не суцільних спостережень є можливість отримання інформації в короткі терміни і з меншими витратами ресурсів, ніж при суцільному спостереженні. Це пов'язано з меншим об'ємом збираної інформації, а отже, з нижчими витратами на її отримання, перевірку достовірності, обробку, аналіз.

Одним з видів несуцільного спостереження є *вибіркове спостереження*. Воно засноване на принципі випадкового відбору тих одиниць сукупності, що вивчається. Чисельність вибіркової сукупності залежить від природи (характеру) досліджуваного явища.

Різновидом вибіркового спостереження є метод *моментних спостережень*. Суть його полягає в тому, що інформація збирається шляхом реєстрації значень ознак у одиниць вибіркової сукупності в деякі заздалегідь певні моменти часу. Тому метод моментних спостережень припускає відбір не тільки одиниць досліджуваної сукупності (вибірку в просторі), але і моментів часу, в які проводиться реєстрація стану досліджуваного об'єкта (вибірка в часі). Цей вид спостереження застосовується під час проведення обстежень доходів і витрат домогосподарств.

За методом *основного масиву* обстеженню піддаються найістотніші, найбільш вагомі одиниці сукупності, що вивчається, які за основною ознакою мають найбільшу питому вагу в сукупності. Наприклад, спостереження за цінами на споживчих ринках.

Монографічне обстеження проводиться з метою виявлення наявних або таких, що намічаються тенденцій в розвитку цього явища. Це вид несуцільного спостереження, за якого ретельному обстеженню піддаються окремі одиниці сукупності, що вивчаються, зазвичай представники яких-

небудь нових типів явищ. Детальне монографічне дослідження одного банку, ферми і так далі дозволяє характеризувати пропорції і зв'язки, яких немає при масових спостереженнях. Це дозволяє уточнити програму масового спостереження, характерні риси і основні ознаки об'єкта дослідження.

Способи статистичного спостереження. Статистична інформація може бути отримана різними способами, найважливішими з яких є безпосереднє спостереження, документальний облік фактів і опит.

Безпосереднє спостереження здійснюють шляхом безпосереднього виміру, зважування, підрахунку і на цій підставі роблять записи у формулярі спостереження. Наприклад, цей спосіб застосовують під час спостереження за виробництвом продукції.

Документальний спосіб спостереження ґрунтується на різного роду документах облікового характеру з первинного обліку.

Опитування – спосіб спостереження, за якого інформацію отримують із слів респондента коли явища і процеси, непіддатливі безпосередньому прямому спостереженню.

У статистиці застосовуються такі види опитування: усний, самореєстрація, кореспондентський, анкетний і явочний.

Під час усного (експедиційного) опитування спеціально підготовлені працівники отримують необхідну інформацію на основі опитування відповідних осіб і самі фіксують відповіді у формулярі спостереження. За формою проведення усне опитування може бути прямим коли обліковувач зустрічається з кожним респондентом і опосередкованим, наприклад по телефону, ПК.

При самореєстрації формуляри заповнюються самими респондентами, а обліковувачі роздають їм бланки опитувального аркуша, роз'яснюють правила їх заповнення, а потім їх збирають.

Кореспондентський спосіб полягає в тому, що відомості отримують від добровільних кореспондентів.

Анкетне опитування – це збір інформації за допомогою анкет. Заповнення цих анкет має добровільний характер і здійснюється, як правило, анонімно. Цей спосіб збору інформації використовується під час не суцільного спостереження. Анкетне опитування застосовується коли не потрібна висока точність, а тільки наближені, орієнтовні результати, наприклад, при вивченні громадської думки про результати виборів, роботи банків і таке ін.

Явочний спосіб опитування передбачає збір інформації в явочному порядку, наприклад при реєстрації браків, народжень та інше.

Вибираючи той або інший вид опитування необхідно враховувати: з якою точністю треба провести спостереження, фінансові можливості.

2.3. Програмно-методологічні питання статистичного спостереження

Мета спостереження. Статистичні спостереження переслідують основну мету – отримання достовірної інформації для виявлення закономірностей розвитку явищ і процесів. Наприклад, метою перепису населення України було отримання даних про чисельність, склад населення. Нечітко поставлена мета може спровокувати, збір непотрібних даних або, навпаки, не будуть отримані відомості, необхідні для аналізу.

Для підготовки спостереження окрім мети потрібно встановити об'єкт спостереження. Під об'єктом спостереження розуміється деяка статистична сукупність, в якій протікають досліджувані масові явища і процеси. Об'єктом спостереження може бути сукупність населення окремої країни; житлові будинки; підприємства, комерційні банки, навчальні заклади.

Об'єкти статистичного спостереження складається з окремих елементів – одиниць спостереження.

Одиницею спостереження називають складовий елемент об'єкта, що є носієм ознак, що підлягають реєстрації. Наприклад, при бюджетних обстеженнях – домашнє господарство.

Програма статистичного спостереження це перелік ознак (або питань), що підлягають реєстрації в процесі спостереження. Для визначення складу реєстрованих ознак розробляють програму спостереження. Необхідно відібрати ті ознаки, які є істотними, основними для характеристики об'єкта, виходячи з мети дослідження.

Питання програми повинні бути точними і недвозначними, а також легкими для розуміння щоб уникнути зайвих труднощів при отриманні відповідей. Логічний порядок дослідження ознак допоможе отримати достовірні відомості про явища і процеси. У програму доцільно включати питання контрольного характеру для перевірки і уточнення зібраних даних.

Для забезпечення одноманітності отримуваних відомостей від кожної звітної одиниці (це важливо для подальшої обробки інформації) програма оформляється у вигляді документа – статистичного формуляру.

Статистичний формуляр – це документ єдиного зразка, що містить програму і результати спостереження. Обов'язковими елементами статистичного формуляру є титульна і адресна частини, які повинні містити найменування статистичного спостереження і органу, провідного спостереження, адресу звітної одиниці, її підлеглість.

Формуляри називають як: звіт, картка, переписний лист, опитувальний бланк, анкета і так далі.

До формуляру розробляється інструкція, що визначає порядок проведення спостереження і заповнення форми звітності, переписного листа, анкети. Формуляр і інструкція по його заповненню разом складають інструментарій статистичного спостереження.

Місце і час спостереження. Вибір місця проведення обстеження залежить головним чином від мети спостереження. Якщо необхідно отримати дані для вивчення складу

населення по країні, то в цьому випадку спостереження охопить територію всієї країни.

Як *критичний момент* розуміють конкретний день року, година дня, за станом на який повинна бути проведена реєстрація ознак по кожній одиниці досліджуваної сукупності. Критичний момент встановлюється з метою отримання зіставних статистичних даних. У разі дослідження варіювання біржових котирувань на торгах фондових бірж в різних містах України необхідно мати дані про курси акцій зареєстровані в один і той же день. Якщо ж треба проаналізувати зміну обсягу продажів акції якої-небудь фірми на біржовому ринку в звітному році порівняно з попереднім роком, то встановлюється не критичний момент, а інтервал часу, за який слід отримати статистичні дані. Вибір критичного моменту або інтервалу часу визначається перш за все метою дослідження.

Період спостереження – це час, протягом якого відбувається заповнення статистичних формулярів, тобто час, необхідний для проведення масового збору даних. Цей термін визначається виходячи з об'єму роботи (числа реєстрованих ознак і одиниць в обстежуваній сукупності), чисельності персоналу, зайнятого збором інформації.

2.4. Помилки спостереження

Помилкою спостереження називається розбіжність між розрахунковим і реальним значеннями величин, що вивчаються.

Щоб уникнути помилок спостереження, попередити, виявити і виправити їх, необхідно забезпечити якісне навчання персоналу, який проводитиме спостереження; організувати контрольні перевірки правильності заповнення статистичних формулярів; провести логічний і арифметичний контроль отриманих даних після закінчення збору інформації.

Залежно від причин виникнення розрізняють помилки реєстрації і помилки репрезентативності.

Помилки реєстрації – це відхилення між значенням показника, отриманого в ході статистичного спостереження, і фактичним його значенням. Цей вид помилок може бути і при суцільному, і при не суцільному спостереженнях.

Помилки реєстрації бувають випадкові й систематичні.

Випадкові помилки – це результат дії різних випадкових чинників (механічні при заповненні статистичного формуляру). Такі помилки мають різну спрямованість: вони можуть і підвищувати, і знижувати значення показників. При достатньо великій обстежуваній сукупності в результаті дії закону великих чисел ці помилки взаємно погашаються.

Систематичні помилки реєстрації завжди мають однакову тенденцію або до збільшення, або до зменшення значення показників по кожній одиниці спостереження, і тому величина показника по сукупності включатиме накопичену помилку.

Помилки репрезентативності характерні тільки для несуцільного спостереження. Вони виникають тому, що відібрана і обстежена сукупність недостатньо точно відтворює генеральну сукупність.

Помилки репрезентативності також бувають випадкові й систематичні.

Випадкові помилки виникають, якщо відібрана сукупність неповно відтворює генеральну сукупність.

Систематичні помилки репрезентативності з'являються внаслідок порушення принципів відбору одиниць з генеральної сукупності.

Арифметичний контроль ґрунтується на використанні кількісних зв'язків між значеннями різних показників. Наприклад, якщо серед зібраних даних є зведення про чисельність персоналу, заробітну плату в середньому на одного працівника і фонд заробітної плати, то множення перших двох показників повинне дати значення третього показника. Якщо арифметичний контроль покаже, що така залежність не виконується, це свідчатиме про невірність зібраних даних. Тому, в програму статистичного спостереження

доцільно долучати показники, які дають можливість провести арифметичний контроль.

Логічний контроль ґрунтується на знанні взаємозв'язків між показниками, але не кількісних, а логічних. Наприклад, якщо підприємство має стабільний прибуток, то воно не може мати збиток. Отже, якщо в звіті підприємства є одночасно обидва записи, то це свідчить що одна з них не відповідає дійсності. Для виправлення помилок, виявлених в ході логічного контролю, потрібно повторно звернутися до джерела відомостей.

2.5. Програмні інструменти для вилучення даних із зовнішніх джерел і обробки величезних масивів даних (Big data)

Великий вибух даних, що стосуються взаємодії людей, компаній та інших організацій дає нові можливості для їх аналізу. Щоб перетворити ці дані в знання про соціальну та економічну поведінку цих інституційних одиниць, дослідники повинні мати справу з великою кількістю неструктурованих і різнорідних даних. Для виконання цього завдання необхідно ретельно спланувати і організувати весь процес аналізу даних з урахуванням особливостей соціального та економічного аналізу, який складається з широкого спектра різнорідних джерел інформації. Ґрунтуючись на підході до життєвого циклу даних, необхідно мати архітектуру великих даних, яка належним чином інтегрує більшість нетрадиційних джерел інформації і методів аналізу даних, щоб забезпечити спеціально розроблену систему для прогнозування соціальної й економічної поведінки, тенденцій.

Big Data або «великі дані» – термін, що з'явився зовсім недавно, але саме явище з'явилося значно раніше. Великими даними прийнято називати величезні масиви інформації зі складною неоднорідною і / або невизначеною структурою. Іноді про Big Data кажуть як про неструктуровану інфор-

мацію, але це не так – великі дані завжди мають структуру, вона може бути складною через те, що дані надходять з різних джерел і містять абсолютно різні відомості або структура є зовсім невідомою. Як правило, привести ці масиви інформації в єдину форму є неможливим.

Великі дані хоча й існують вже кілька років, але раніше вони не мали великої цінності, оскільки обробка та аналіз були ускладнені. Для цього потрібні були значні обчислювальні потужності, трудові та фінансові витрати. Все змінилося, коли з'явилася технологія обробки багатогігабайтних масивів інформації з швидкою оперативною пам'яттю. Прорив у цій галузі пов'язують з виходом на ринок вільно розповсюджуваної платформи Hadoop, що налічує бібліотеки, утиліти і фреймворки для роботи з Big Data. Компоненти Hadoop використовуються сьогодні в більшості комерційних платформ і систем таких компаній, як SAP, Oracle, IBM і так далі.

Сьогодні термін Big Data, як правило, використовується для позначення не тільки самих масивів даних, але також інструментів для їх обробки і потенційної користі, яка може бути отримана в результаті кропіткого аналізу. Головні характеристики, що відрізняють Big Data від іншого роду даних – три V: volume, velocity, variety. Перша volume – великі обсяги, друга velocity – необхідність в швидкій обробці і висока швидкість накопичення цих даних, третє variety – різноманітність.

Великі дані, наприклад, в ритейлі можуть складатися з різних відомостей про споживачів, історії їх покупок, детальної інформації з кожного чека про знижки, факти відвідуванні різних магазинів і т. ін. Банки і страхові компанії також мають можливість збирати інформацію про клієнтів, їх дії, фінансові транзакції і навіть переміщення як по місту, так і по світу. Банк може визначити дати важливих подій у житті своїх клієнтів – весілля, народження дитини, зміна роботи, переїзд і т. ін. Ці відомості можуть використовувати виробники для збільшення продажів і роботи над лояльністю

клієнтів, державні установи для прийняття управлінських рішень при розробленні заходів щодо економічної та соціальної політики.

Великі дані характеризують також розвиток комунальної галузі. Можливість збирати й аналізувати інформацію, яку надають з лічильників обліку води, газу та електроенергії – це перший і головний крок на шляху до розумного використання ресурсів як на рівні домогосподарств, так і в масштабі ЖКГ-компаній. Так, наприклад, застосування великих даних дозволило естонській розподільчій компанії Elektrilevi, спільно з Ericsson реалізувати запуск інтелектуальної системи обліку електроенергії, всього за перші два роки проекту підвищити ефективність на 20% і уникнути помилок за рахунок оперативного виявлення несправності.

Поняття великих даних. Поняття великих даних має на увазі роботу з інформацією величезного обсягу і різноманітного складу, часто оновлюваної і, яка з'являється з різних джерел з метою збільшення ефективності роботи, створення нових продуктів і підвищення конкурентоспроможності компаній. Консалтингова компанія Forrester дає коротке формулювання: великі дані об'єднують техніки й технології, які витягують користь з даних на екстремальні межі практичності.

Великі дані призначені для обробки більш значних обсягів інформації, ніж бізнес-аналітика, і це, звичайно, відповідає традиційному визначенню великих даних.

Великі дані призначені для обробки швидко одержуваних і мінливих відомостей, що означає більш глибоке дослідження й інтерактивність. У деяких випадках результати формуються швидше, ніж завантажується веб-сторінка.

Великі дані призначені для обробки неструктурованих даних, способи використання яких ми тільки починаємо вивчати після того, як змогли налагодити їх збір і зберігання, та нам потрібні алгоритми і можливість діалогу для полегшення пошуку тенденцій, що містяться всередині цих масивів.

Робота з великими даними не схожа на звичайний процес бізнес-аналітики, де просте додавання відомих значень приносить результат: наприклад, результат складання даних про сплачені рахунки стає обсягом продажів за рік. При роботі з великими даними результат виходить в процесі їх очищення шляхом послідовного моделювання: висувається гіпотеза, будується статистична, візуальна або семантична модель, на її підставі перевіряється правильність висунутої гіпотези. Цей процес вимагає від дослідника або інтерпретації візуальних значень або складання інтерактивних запитів на основі власних знань, або розроблення адаптивних алгоритмів «машинної освіти», здатних отримати шуканий результат. Причому термін життя такого алгоритму може бути досить коротким.

Методики аналізу великих даних. Існує безліч різноманітних методик аналізу масивів даних, в основі яких лежить інструментарій, запозичений із статистики та інформатики (наприклад, машинне навчання). Список не претендує на повноту, проте в ньому відображені найбільш затребувані в різних галузях підходи. При цьому потрібно розуміти, що дослідники продовжують працювати над створенням нових методик і вдосконаленням існуючих. Крім того, деякі з перерахованих методик зовсім не обов'язково можуть бути застосовані виключно до великих даних і можуть з успіхом використовуватися для менших за обсягом масивів (наприклад, А / В-тестування, регресійний аналіз). Безумовно, чим більше об'ємний і диверсифікований масив піддається аналізу, тим точніші дані вдається отримати на виході.

A / B testing. Методика, завдяки якій контрольна вибірка по черзі порівнюється з іншими. Таким чином, вдається виявити оптимальну комбінацію показників для досягнення, наприклад, найкращої відповідної реакції споживачів на маркетингову пропозицію. Великі дані дозволяють провести величезну кількість ітерацій і таким чином отримати статистично достовірний результат.

Association rule learning. Набір методик для виявлення взаємозв'язків, тобто асоціативних правил між змінними величинами у великих масивах даних. Використовується в data mining.

Classification. Набір методик, які дозволяють передбачити поведінку споживачів в певному сегменті ринку (прийняття рішень про покупку, обсяг споживання та ін.). Використовується в data mining.

Cluster analysis. Статистичний метод класифікації об'єктів по групах за рахунок виявлення наперед невідомих загальних ознак. Використовується в data mining.

Crowdsourcing. Методика збору даних з великої кількості джерел.

Data fusion and data integration. Набір методик, який дозволяє аналізувати коментарі користувачів соціальних мереж і зіставляти з результатами продажів у режимі реального часу.

Data mining. Набір методик, який дозволяє визначити найбільш сприйнятливі для продукту, що просувається або послуги категорії споживачів, виявити особливості найбільш успішних працівників, передбачити поведінкову модель споживачів.

Ensemble learning. У цьому методі використовується безліч предикативних моделей за рахунок чого підвищується якість зроблених прогнозів.

Genetic algorithms. У цій методиці можливі рішення подаються у вигляді «хромосом», які можуть комбінуватися і мутувати. Як і в процесі природної еволюції, виживає найбільш пристосована особина.

Machine learning. Напрямок в інформатиці (історично за ним закріпилася назва «штучний інтелект»), яке має на меті створення алгоритмів самонавчання на основі аналізу емпіричних даних.

Natural language processing (NLP). Набір запозичених з інформатики та лінгвістики методик – це розпізнавання природної мови людини.

Network analysis. Набір методик аналізу зв'язків між вузлами у мережах. Стосовно до соціальних мереж дозволяє аналізувати взаємозв'язок між окремими користувачами, компаніями, спільнотами і т.п.

Optimization. Набір чисельних методів для редизайну складних систем і процесів для поліпшення одного або декількох показників. Допомагає в прийнятті стратегічних рішень, наприклад, складу виведеної на ринок продуктової лінійки, проведенні інвестиційного аналізу та ін.

Pattern recognition. Набір методик з елементами само-навчання для передбачення поведінкової моделі споживачів.

Predictive modeling. Набір методик, які дозволяють створити математичну модель наперед заданого ймовірного сценарію розвитку подій. Наприклад, аналіз бази даних CRM-системи на предмет можливих умов, які підштовхнуть абоненти змінити провайдера.

Regression. Набір статистичних методів для виявлення закономірності між зміною залежною змінною і однією або декількома незалежними. Часто застосовується для прогнозування. Використовується в data mining.

Sentiment analysis. В основі методик оцінювання настроїв споживачів лежать технології розпізнавання природної мови людини. Вони дозволяють вичленувати із загального інформаційного потоку повідомлення, пов'язані з предметом дослідження (наприклад, споживчим продуктом). Далі оцінити полярність судження (позитивне чи негативне), ступінь емоційності та ін.

Signal processing. Запозичений з радіотехніки набір методик, який має на меті розпізнавання сигналу на тлі шуму і його подальшого аналізу.

Spatial analysis. Набір частково запозичених з статистики методик аналізу просторових даних – топології місцевості, географічних координат, геометрії об'єктів. Джерелом великих даних у цьому випадку часто виступають геоінформаційні системи (ГІС).

Statistics. Наука про збір, організації та інтерпретації даних, включаючи розробку опитувальників і проведення експериментів. Статистичні методи часто застосовуються для оціночних суджень про взаємозв'язки між тими чи іншими подіями.

Supervised learning. Набір заснованих на технологіях машинного навчання методик, які дозволяють виявити функціональні взаємозв'язки в аналізованих масивах даних.

Simulation. Моделювання поведінки складних систем часто використовується для прогнозування і опрацювання різних сценаріїв при плануванні.

Time series analysis. Набір запозичених із статистики методів аналізу повторюваних з плином часу послідовностей даних. Наприклад, відстеження ринку цінних паперів або захворюваності пацієнтів.

Unsupervised learning. Набір заснованих на технологіях машинного навчання методик, які дозволяють виявити приховані функціональні взаємозв'язки в аналізованих масивах даних. Має спільні риси з Cluster Analysis.

Візуалізація. Методи графічного подання результатів аналізу великих даних у вигляді діаграм або анімації для спрощення інтерпретації і полегшення розуміння отриманих результатів.

Згідно зі звітом IDC, International Data Corporation (<https://www.idc.com/>) у зв'язку із зростанням обсягів даних, що генеруються підключеними до Інтернету пристроями, датчиками й іншими технологіями, доходи, пов'язані з великими даними, збільшаться з \$ 130 млрд у 2016 році до більше ніж \$ 203 млрд до 2020 року. Однак ті компанії, у яких немає ІТ-інфраструктури, необхідної для адаптації до революції великих даних, не зможуть отримати вигоду від цього зростання, вважають експерти компанії TmaxSoft. (<https://www.tmaxsoft.su/>).

Підприємствам зазвичай відомо, що в накопичених ними величезних обсягах даних міститься важлива інформація про їхній бізнес і клієнтів. Якщо компанія зможе успішно

застосувати цю інформацію, то у неї буде значна перевага в порівнянні з конкурентами, і вона зможе запропонувати кращі, ніж у них, продукти. Однак, як стверджують експерти компанії TmaxSoft, багато підприємств та організацій все ще не можуть ефективно використовувати великі дані через те, що їх успадкована ІТ-інфраструктура нездатна забезпечити необхідну ємність систем зберігання, процеси обміну даних, утиліти і додатки, необхідні для обробки і аналізу великих масивів неструктурованих даних для вилучення з них цінної інформації.

Крім того, збільшення процесорної потужності, необхідної для аналізу великих даних, може вимагати значних інвестицій в застарілу ІТ-інфраструктуру організації.

2.6. Завдання для самостійної роботи

Завдання 2.1. Зробити перелік питань переписного аркуша перепису населення України 2001 р., відповідь на які подати у вигляді чисел.

Завдання 2.2. Провести опитування клієнтів перукарень щодо якості надаваних послуг. Визначити для цього спостереження одиницю сукупності та одиницю спостереження. Яким має бути це спостереження за ступенем охоплення одиниць сукупності та часом реєстрації фактів?

Завдання 2.3. Для проведення обстеження з метою вивчення ефективності підприємницької діяльності підприємств торгівлі визначити об'єкт, одиницю спостереження та одиницю сукупності.

Завдання 2.4. Для проведення обліку осіб, що мають субсидії на комунальні послуги визначити об'єкт, одиницю спостереження та одиницю сукупності.

Завдання 2.5. Визначити об'єкт спостереження для перепису установ охорони здоров'я.

Завдання 2.6. Сформулювати найістотніші ознаки для таких одиниць статистичного спостереження:

- а) домогосподарства;
- б) школи;
- в) малі підприємства.

Завдання 2.7. Скласти перелік запитань, що входять до програми спостережень:

- а) поточного обліку емігрантів з України для вивчення мети виїзду та країни прибуття;
- б) опитування студентів щодо визначення їхнього репродуктивного здоров'я з урахуванням віку, статі респондентів.

Завдання 2.8. Скласти проект статистичного формуляра для обстежень:

- а) поточного обліку клієнтів страхової компанії з метою визначення виду та терміну страхування;
- б) опитування клієнтів банку з метою розширення сфери надання послуг.

Завдання 2.9. Які, на Вашу думку, ознаки слід було б реєструвати під час проведення:

- а) обстеження торговельних підприємств з метою вивчення сезонності їх роботи;
- б) обстеження роботи транспорту для вивчення ефективності роботи різних видів його;
- в) обстеження домогосподарств з метою вивчення бюджету часу?

Завдання 2.10. Визначити місце, об'єктивний та суб'єктивний час спостереження, а також критичний момент:

- а) перепис населення, що здійснюється за станом на 0.00 години з 17 до 18 грудня протягом десяти днів.

б) збір декларацій про доходи громадян у податкових інспекціях здійснюється не пізніше 1 березня від початку поточного року;

в) опитування учасників лотереї з 2 до 3 травня;

Завдання 2.11. Сформулювати об'єкт, одиницю і мету спостереження і розробити програму обстеження:

а) ліцеїв;

б) фармацевтичних підприємств;

в) супермаркетів.

Завдання 2.12. Визначити вид статистичного спостереження за ступенем охоплення одиниць та часом реєстрації фактів:

а) облік чисельності біженців;

б) перелік партій в Україні.

Завдання 2.13. Сформулювати питання для включення їх у формуляр спостереження за такими ознаками об'єктів спостереження:

а) стать і вік людини;

б) чисельний склад родини;

в) родинні зв'язки членів родини.

Завдання 2.14. Щоб вивчити думки студентів про причини використання шпаргалок на іспитах ректорат ЗВО вирішив провести спеціальне обстеження. Потрібно визначити: а) об'єкт і одиницю спостереження; б) ознаки, що підлягають реєстрації; в) вид і спосіб спостереження; а також г) розробити формуляр і написати стислу інструкцію до його заповнення; д) скласти організаційний план обстеження.

Завдання 2.15. Держкомстат України проводив одноразове обстеження підприємств з метою використання робочої сили у 2002 році. До якого виду статистичного спостереження за ознакою часу треба відносити це обстеження?

Завдання 2.16. Для перевірки якості експортного товару була відібрана у випадковому порядку десята частина партії. Кожну одиницю товару ретельно оглядали, визначали, фіксували якість. До якого виду спостереження і за якими ознаками можна віднести це обстеження?

Завдання 2.17. Відповіді на запитання формуляра спостереження записуються на підставі документів, що містять відповідні дані. Як називають такого виду спостереження?

Завдання 2.18. Видавництво, бажаючи з'ясувати думку читачів щодо виданої ним книжки, розіслало анкету з проханням відповісти на запитання, що містяться в ній, і повернути її в редакцію. Як називається таке спостереження?

2.7. Питання для самоконтролю

1. У чому полягає суть і завдання статистичного спостереження?
2. Які є види статистичного спостереження?
3. З чого складається план статистичного спостереження?
4. Які є програмно-методологічні питання плану статистичного спостереження?
5. Які є організаційні питання статистичного спостереження?
6. Які є організаційні форми статистичного спостереження?
7. Які є види статистичної звітності?
8. Які є спеціально організовані статистичні спостереження та їх види ?
9. Які є реєстри?
10. Які є помилки статистичного спостереження та методи перевірки достовірності зібраних даних?
11. Які є програмні засоби для розв'язання задач статистичного спостереження?

Розділ 3

ЗВЕДЕННЯ ТА ГРУПУВАННЯ СТАТИСТИЧНИХ ДАНИХ

План

- 3.1. Поняття статистичного зведення та групування.
- 3.2. Основні завдання і види групувань.
- 3.3. Групування за кількісною ознакою. Правила створення груп та інтервалів групування.
- 3.4. Метод вторинного групування.
- 3.5. Ряди розподілу.
- 3.6. Статистичні класифікації.
- 3.7. Статистичні таблиці.
- 3.8. Задачі для самостійної роботи.
- 3.9. Питання для самоконтролю.

3.1. Поняття статистичного зведення та групування

Матеріал, який було зібрано в результаті масового статистичного спостереження, являє собою розрізнені початкові дані про окремі одиниці сукупності досліджуваного явища. Такі дані ще не характеризують явище загалом, не дають уяви про його обсяг, склад, величину характерних ознак. Тому дані про кожну одиницю статистичного спостереження необхідно обробити належним чином, впорядкувати, систематизувати, узагальнити і за допомогою системи узагальнюючих показників дати характеристику досліджуваного явища. Така наукова обробка даних виконується на другому етапі статистичного дослідження, який називається *зведенням і групуванням статистичних даних*.

Означення. *Статистичне зведення* – це упорядкування, систематизація і наукова обробка первинних матеріалів статистичного спостереження з метою отримання узагальненої характеристики досліджуваного явища за деякими істотними ознаками.

Зведення – відповідальний етап статистичного дослідження. Від якості проведення зведення даних залежить зміст отриманих на його основі узагальнюючих показників і результатів дослідження.

Статистичні зведення відрізняються низкою ознак: за обсягом робіт і складністю обробки матеріалів, за способом організації робіт, за кількістю етапів проведення.

За обсягом робіт та їх складністю зведення поділяють на *просте і складне (групове)*.

Просте підсумкове зведення – це простий підрахунок підсумків первинних статистичних даних без попереднього розподілу одержаних відомостей на групи, яке передбачає визначення лише загальної кількості обстежених одиниць сукупності та узагальнюючих показників.

Наприклад, в результаті вивчення матеріалів статистичної звітності підприємств шляхом простого підсумування можна визначити показники загальної кількості підприємств України, кількість зайнятих працівників, загальний обсяг реалізованої продукції та ін. Значення простого зведення зростає в умовах ринкової економіки, тому що підсумкові дані можуть бути отримані швидко, що є основою для прийняття оперативних управлінських рішень в соціальній сфері, торгівлі, або іншій діяльності.

Групове статистичне зведення включає попередній розподіл одиниць статистичної сукупності на окремі групи, підрахунок групових і загальних показників, комплекс дій по систематизації та узагальненню статистичних даних.

Групове статистичне зведення передбачає послідовне виконання низки операцій з обробки первинних статистичних матеріалів, які здійснюються на основі науково розробленої програми:

- групування даних статистичного спостереження;
- розроблення системи статистичних показників для характеристики груп, підгруп і всієї сукупності загалом;

- підрахунок групових і загальних підсумків з метою отримання узагальнюючих статистичних показників;
- розроблення технологічних схем обробки інформації для виконання розрахунків з використанням комп'ютерних засобів;
- табличне і графічне оформлення результатів статистичного спостереження.

Статистичне зведення за способом організації робіт може бути *централізоване і децентралізоване*.

При централізованому зведенні всі первинні матеріали статистичного спостереження зосереджуються у центральному органі державної статистики, де вони й обробляються, тобто систематизуються і узагальнюються. Такий спосіб зведення використовується при переписах населення, соціологічних опитуваннях та інших статистичних обстеженнях.

При децентралізованому зведенні обробка статистичної інформації здійснюється на місцях, в районних, міських відділах статистики міністерств і відомств, за видами економічної діяльності, формами власності. Потім дані зводяться у межах області, регіону і їх надсилають до центрального державного органу, де дані узагальнюються по країні.

Наприклад, децентралізоване зведення застосовується під час вивчення показників статистичної звітності про діяльність підприємств. Воно особливо важливо в умовах децентралізації регіонів України, оскільки результати зведення можна оперативніше використовувати на місцях та приймати управлінські рішення, корисні для розвитку регіону. Підсумки по регіонах передаються в центральні органи статистики, де узагальнюються загалом по країні.

На практиці використовують різні технологічні схеми обробки первинної інформації. Наприклад, при розробці матеріалів перепису населення спочатку підсумовують чисельність і склад населення на місцях в окремих населених пунктах децентралізовано, а в подальшому кінцеві детальні результати перепису обробляють централізовано.

Раніше основною перевагою централізованого зведення було використання автоматизації обробки даних, застосування єдиної методології їх розроблення. Проте збільшувалися витрати на його організацію, створювався великий розрив у часі між збиранням даних і результатами їх обробки.

У сучасних умовах обов'язковим для децентралізованого та централізованого зведення є дві основні операції: кодування даних і формування автоматизованих банків даних, які дозволяють багаторазове використання інформації. Тому, з урахуванням використання комп'ютерних засобів та обробки інформації на місцях, децентралізоване зведення стало оперативніше і дешевше.

Найбільш складним у методологічному плані етапом зведення є *групування*.

Означення. *Статистичне групування* – це об'єднання одиниць сукупності в однорідні групи й підгрупи за певними ознаками, одержання системи узагальнюючих показників з метою всебічної характеристики груп і підгруп та сукупності загалом, табличне й графічне оформлення статистичних показників.

Методологічні засади групувань і обґрунтування системи показників залежать від мети дослідження, суті явища, яке вивчається, особливостей сукупності. Це передбачається заздалегідь при розробленні його програми.

Групування дозволяє одержати більш детальну характеристику досліджуваної сукупності за окремими групувальними ознаками. Наприклад, матеріали статистичного вивчення підприємств можна спочатку згрупувати за трьома видами підприємств: малими, середніми та великими, а потім кожну групу вивчати за низкою показників.

Разом з програмою складається план організаційних питань, які пов'язані з умовами та способом організації робіт, термінами виконання етапів робіт, їх послідовністю, способом зведення даних, проведення подальшої обробки даних, аналіз і оформлення отриманих результатів.

3.2. Основні завдання і види групувань

Групування є важливим методом статистичного дослідження, який впливає на кінцеві результати. В результаті групування сукупність поділяється на групи, однорідні в тому чи іншому розумінні. Виділені групи можна охарактеризувати різними показниками, дослідити залежність між ознаками, що покладені в основу групування та іншими показниками. Таким чином, здійснюється систематизація показників, виявляються певні тенденції та закономірності.

Основний принцип будь-якого поділу ґрунтується на двох положеннях:

- 1) в один клас, групу об'єднуються елементи певною мірою подібні між собою;
- 2) ступінь подібності між елементами, які належать до одного класу, значно вищий, ніж між елементами, що належать до різних класів.

У кожному конкретному дослідженні вирішуються три питання:

- 1) що взяти за основу групування;
- 2) скільки груп, позицій необхідно виокремити;
- 3) як розмежувати групи.

Зауваження. Найбільш важливим моментом групування є відбір типових для цього явища ознак, які дозволять відокремити суттєво відмінні групи одиниць сукупності.

Якщо в основу групувань будуть покладені другорядні ознаки, то групування не досягне мети дослідження, не вирішить покладених на нього завдань. Тому метод статистичних групувань є одним із найбільш ефективних методів обробки масових даних, який є основою дослідження й дозволяє виявити закономірності розвитку явищ і процесів.

Ознаки, за якими проводиться розподіл одиниць певної сукупності на групи, називаються *групувальними ознаками* або *основою групування*.

Основою групування може бути будь-яка *атрибутивна (якісна) або кількісна ознака*.

За допомогою методу групувань статистика вирішує різні завдання:

- відокремлення соціально-економічних типів явищ і процесів;
- виявлення складу (структури) досліджуваних явищ і структурних зрушень;
- дослідження взаємозв'язків і залежностей між різними суспільними явищами та їх ознаками.

Відповідно до цих завдань виділяють такі види групувань: *типологічні, структурні, аналітичні*.

Зауваження. Наведені види групувань за завданнями, які вирішуються ними, є певною мірою умовними, оскільки на практиці вони, як правило, використовуються в комплексі.

Типологічне групування – це розподіл якісно неоднорідної статистичної сукупності за певною ознакою на окремі однорідні групи, класи, соціально-економічні типи з подальшою їх характеристикою. Основне завдання типологічних групувань – визначення типів, однорідних груп, з яких складається статистична сукупність, істотних відмінностей між групами, а також спільних для всіх ознак.

Основою типологічного групування є, як правило, **атрибутивна ознака**, яка не має кількісного вираження, (наприклад, стать, професія, освіта, місце проживання тощо). Різновидом атрибутивної ознаки є **альтернативна ознака**, коли існує лише два варіанти цієї ознаки (наприклад, стать чоловіча або жіноча).

Кількість груп залежить від того, яку ознаку покладено в основу групування. При групуванні за атрибутивною ознакою групи відрізняються між собою не величиною, а характером ознаки, кількість груп визначається кількістю різновидностей атрибутивної ознаки. Прикладом цього виду групувань є групування населення за місцем проживання; групування господарюючих суб'єктів за формами власності, капітальних інвестицій за джерелами фінансування та ін. Типологічні групування широко застосовуються в економічних, соціологічних і демографічних дослідженнях. Вони забезпечують

також можливість аналізу специфіки й особливостей розвитку окремих типів, зміни їх співвідношень у межах загального економічного процесу. Приклад типологічного групування наведено в табл. 3.1.

Дані таблиці свідчать про те, що основним джерелом фінансування капітальних інвестицій за певний період є власні кошти підприємств та організацій. В динаміці збільшується частка капітальних інвестицій за рахунок коштів місцевих бюджетів й зменшується обсяг коштів іноземних інвесторів.

Таблиця 3.1

Капітальні інвестиції за джерелами фінансування, %

Джерела фінансування	2015 р.	2016 р.	2017 р.
Кошти державного бюджету	2,5	2,6	3,4
Кошти місцевих бюджетів	5,2	7,5	9,3
Власні кошти підприємств та організацій	67,5	69,3	69,1
Кредити банків та інші позики	7,6	7,5	6,6
Кошти іноземних інвесторів	3,0	2,7	1,4
Кошти населення на будівництво житла	11,7	8,3	7,3
Інші джерела фінансування	2,5	2,1	2,9
Усього	100	100	100

Джерело: Статистичний щорічник України за 2017 рік. Державна служба статистики України. С. 359

Структурними називаються групування, за допомогою яких виявляють структуру досліджуваної сукупності за будь-якою ознакою.

За допомогою структурних групувань можна аналізувати структуру сукупності та структурних зрушень у розвитку

соціально-економічних явищ і процесів у часі. Наприклад, структуру собівартості продукції (сировина, матеріали, зарплата, амортизаційні відрахування тощо), валового внутрішнього продукту (оплата праці, податки, валовий прибуток тощо), населення за статтю, віковими групами. *Приклад* структурного групування населення за місцем проживання наведено в табл. 3.2.

Таблиця 3.2

Розподіл наявного населення за місцем проживання

(на 1 січня 2018 року)

Наявне населення	Млн осіб.	У % до підсумку
Міське	29,4	69,3
Сільське	13,0	30,7
Усього	42,4	100

Структурні групування можуть здійснюватися як за атрибутивними, так і кількісними ознаками. При групуванні за кількісною ознакою необхідно визначати кількість груп та інтервал групування.

Необхідно зауважити, що типологічні й структурні групування дуже близькі один до одного: типологічні групування виділяють самі типи, а структурні – означають питому вагу окремих типів (груп) у загальному обсязі. Структурні групування є похідними від типологічних групувань. Завдання, які вирішуються типологічними та структурними групуваннями, тісно пов'язані між собою. Типологічні і структурні групування відрізняються лише за метою дослідження, на практиці вони, як правило, використовуються в комплексі.

Типологічні й структурні групування вважаються описовими. Відомо, що явища суспільного життя перебувають у певному взаємозв'язку та взаємозалежності. Важливим завданням статистичних групувань є вивчення взаємозв'язків і взаємозалежності між ознаками. Це завдання вирішується за допомогою аналітичних групувань.

Аналітичними називаються групування, за допомогою яких виявляються взаємозв'язки між окремими ознаками статистичної сукупності.

Аналітичні групування проводять за двома ознаками, з яких одна ознака розглядається як **фактор**, вона впливає на іншу, друга ознака, яка підлягає впливу, розглядається як **результат**. Аналітичне групування дозволяє виявити наявність чи відсутності залежності між двома ознаками. Основою групування, як правило, є факторна ознака. Приклад аналітичного групування за факторною ознакою наведено в табл. 3.3.

Особливістю аналітичних групувань є те, що кожна група факторної ознаки характеризується середніми значеннями результативної ознаки. Наприклад, від вартості основних засобів підприємства залежить обсяг виробленої продукції підприємства; від рівня кваліфікації робітника залежить його продуктивність.

Таблиця 3.3

Залежність обсягу виробленої продукції підприємств від вартості основних виробничих засобів

Вартість основних виробничих засобів, млн грн	Кількість підприємств, одиниць	Вартість основних виробничих засобів у середньому на одне підприємство, млн грн	Вироблено продукції в середньому на одне підприємство, млн грн
1–4	7	2,9	3,1
4–7	10	5,2	6,1
7–10	15	7,8	8,8
10–13	21	10,2	12,0
13 і більше	4	15,5	17,1
Усього	52	9,5	11,9

Тому вартість основних засобів є **факторною** ознакою, а обсяг продукції – **результативною**; кваліфікація робітника – **факторна** ознака, а продуктивність – **результативна** ознака.

За наявності зв'язку між факторною та результативною ознаками групові середні від групи до групи поступово змінюються – збільшуються або зменшуються.

Однак може бути і навпаки. Аналітичне групування може проводитись за результативною ознакою (урожайністю, собівартістю), а факторною ознакою може бути (якість ґрунту, кількість добрив, опадів та ін.). трапляються різноманітні взаємозв'язки між ознаками, які можуть виступати в ролі причини або наслідку явища. Наприклад:

1) факторною ознакою може бути кількісна ознака, а результативною – якісна ознака (стаж роботи і кваліфікація робітника);

2) факторною ознакою може бути якісна ознака, а результативною кількісна (кваліфікація працівника і продуктивність праці);

3) факторною і результативною ознакою можуть бути кількісні показники (продуктивність праці і середня заробітна плата).

Залежно від складності досліджуваного явища та мети дослідження може мати одну, дві й більше ознак. За кількістю групувальних ознак, покладених в основу групування, розрізняють **прості та комбінаційні** групування.

Простими називають групування, які здійснюються за якоюсь однією ознакою. Групування, які здійснюються за двома і більше ознаками (тобто кожна група поділяється на підгрупи) називають **комбінаційними**.

Комбінаційні групування дають можливість комплексно охарактеризувати досліджуване явища чи процес. Приклад, групування позичальників кредитів за обсягом кредитування та терміном надання позички; розподіл населення на міське та сільське, за статтю і віком, студентів за місцем народження та статтю. *Приклад* комбінаційного групування за двома ознаками наведено в табл. 3.4.

Таблиця 3.4

**Групування студентів академічної групи
за національністю та статтю**

Національність	Кількість студентів	Зокрема	
		юнаки	дівчата
Українці	15	9	6
Росіяни	8	2	6
Інші	2	1	1
Усього:	25	12	13

Зауваження. Під час побудови комбінаційних групувань необхідно починати групування за атрибутивними ознаками, значення яких мають суттєві якісні відмінності. Потім групи, утворені за першою ознакою поділяють на підгрупи за другою ознакою, далі ці підгрупи поділяють за третьою ознакою.

Крім того, на практиці складні групування будують не більше ніж за трьома ознаками, оскільки збільшення групвальних ознак ускладнює аналіз, а групові показники стають ненадійними внаслідок різкого зростання кількості утворених груп та зменшення кількості одиниць у групах.

3.3. Групування за кількісною ознакою.

Правила створення груп та інтервалів групування

Важливе значення при використанні методу групувань в економіко-статистичному аналізі має визначення кількості груп і величини інтервалів. Кількісні (варіаційні) ознаки можуть приймати різні цифрові значення в окремих одиницях досліджуваної сукупності. Вони можуть бути дискретними (перервними) або безперервними.

Дискретні ознаки, виражені окремими цілими числами без проміжних значень. Тому кількість груп для дискретної

ознаки є такою скільки існує варіантів ознаки, яка змінюється у відносно неширокому діапазоні. Наприклад, при групуванні сімей за кількістю дітей, квартир – за кількістю кімнат, працівників підприємств – за кількістю тарифних розрядів.

Безперервні ознаки можуть набувати будь-яких значень у певних межах. При групуванні за кількісною ознакою, яка змінюється безперервно (наприклад, сума прибутку, вартість основних засобів, собівартість продукції, зарплатня тощо), визначається число груп і величина інтервалу групування. При цьому число груп і величину інтервалу потрібно встановити такими, які б дозволили більш рівномірно розподілити одиниці сукупності по групах. Чим більший розмах варіації ознаки, тобто різниця між його максимальним і мінімальним значенням, тим, як правило, більше утворюється груп. Необхідно також врахувати чисельність досліджуваної сукупності, щоб в кожную групу потрапила достатня кількість одиниць спостереження.

Залежно від ступеня коливання групувальної ознаки, характеру розподілу статистичної сукупності та її чисельності, застосовують інтервали **рівні або нерівні, відкриті та закриті**.

Рівні інтервали мають однакову величину у всіх групах. Рівні інтервали використовують у випадках, коли варіація групувальної ознаки змінюється плавно.

Якщо застосовуються рівні інтервали, то оптимальну кількість груп можна визначити за формулою американського вченого Стерджесса:

$$n = 1 + 3.322 \lg N, \quad (3.1)$$

де n – кількість груп;

N – обсяг сукупності.

На основі цієї формули можна встановити оптимальне число груп для сукупностей, що мають різну чисельність.

Чисельність сукупності (N)	25–40	40–60	60–100	100–200	200–500	Понад 500
Чисельність груп (n)	5–6	6–7	7–8	8–9	9–10	10–15

Зауваження. Визначення числа груп за формулою Стерджесса при застосуванні рівних інтервалів обґрунтоване в тих випадках, коли розподіл одиниць сукупності за даною ознакою наближається до нормального при незначній варіації ознаки, яка має відносно плавний характер.

В інших випадках, спочатку потрібно утворити ранжований ряд, в якому одиниці сукупності розподілити у порядку зростання чи зменшення групувальної ознаки. Під час аналізу графічного зображення, можна виділити якісно відмітні групи.

На основі ранжированого ряду можна побудувати варіаційний ряд розподілу. Проміжне аналітичне групування і, проаналізувавши їх, визначити кількість істотно відмітних однорідних груп. Групи мають бути якісно однорідними, з достатньою кількістю одиниць в кожній групі, щоб не можна було дійти випадкових неправильних висновків.

Вирішивши питання про кількість груп, потрібно встановити розмір рівного інтервалу групування та визначити межі груп.

Інтервалом групування називають різницю між максимальним і мінімальним значенням ознаки в кожній групі.

Розмір рівного (рівновеликого) інтервалу залежить від розмаху варіації ознаки ($X_{\max} - X_{\min}$) та обґрунтованої кількості груп (n).

Величину рівного інтервалу визначають за формулою

$$h = \frac{X_{\max} - X_{\min}}{n} \quad (3.2)$$

Нерівні інтервали можуть бути зростаючими чи спадаючими.

Відкритими називають інтервали, в яких наперед невідомі мінімальне й максимальне значення. Відкритими можуть бути перший і останній інтервали з відсутньою однією із меж (наприклад, до 10; 100 і більше).

При групуванні за безперервною ознакою важливо правильно визначити нижню і верхню межі кожного інтервалу. Якщо верхня межа попереднього і нижня межа кожного наступного інтервалу, тобто суміжних інтервалів, збігаються, тоді порядок віднесення до груп межових значень ознаки про віднесення одиниць об'єкта спостереження до тієї чи іншої групи у практиці вирішується дwoяко.

Якщо в першому інтервалі є слово «до» або в останньому інтервалі «і більше», то це свідчить про те, що нижню межу інтервалу потрібно вважати належною до включення межових значень ознаки в інтервал, тобто «включно», а верхню – не належною інтервалу, тобто потрібно вважати «виключно».

Закриті інтервали – це інтервали, в яких обидві межі (мінімальна і максимальна) позначені відповідною числовою величиною.

При обчисленні величини інтервалу та розподілі одиниць спостереження на окремі групи при безперервній зміні ознаки можуть бути два варіанта.

Якщо ознака вимірюється тільки цілими числами, межі суміжних інтервалів будуть відрізнятися на одиницю. Наприклад, при групуванні малих підприємств за кількістю працівників межі інтервалів можуть позначатися так: 1–5 працівників; 6–10; 11–15; 16–20 і т.д., тоді обидва межові значення ознаки вважаються такими, що належать конкретному інтервалу. Такий підхід дозволяє правильно розподілити одиниці сукупності на групи.

Однак у практиці групувань за безперервною ознакою при закритих інтервалах нерідко трапляється інший варіант, коли одне і те саме число є верхньою і нижньою межею двох

суміжних груп. Наприклад, при групуванні підприємств за врожайністю сільськогосподарських культур. межі інтервалів позначають таким чином: 20–25 ц/га; 25–30; 30–35; 35–40 і т.д.

У цьому випадку можна застосувати правило, що ліва межа інтервалу вважається «включно», тобто охоплює межові значення ознаки в інтервалі, а права «виключно». Таким чином, підприємство, яке має урожайність 25 ц/га буде зараховано до другої групи.

3.4. Метод вторинного групування

Вторинним називається групування, яке виконане на основі первинного і передбачає створення нових груп. Його застосовують для зведення групувань із різними інтервалами до одного типу групування.

Завдання вторинного групування:

– забезпечення порівнянності двох або більше групувань з різними інтервалами до єдиного вигляду з метою порівняння та аналізу;

– утворення більш численних груп, в яких ясніше проявляється закономірність розподілу;

– формування на основі групувань, які побудовані за кількісними ознаками, якісно однорідних груп.

Розрізняють два способи вторинного групування:

1) просте укрупнення інтервалів;

2) перегрупування за часткою окремих груп в загальному їх підсумку.

Застосовуючи метод вторинного групування, виходять з припущення, що розподіл ознаки всередині інтервалів буде рівномірним.

Вторинне групування способом простого укрупнення інтервалів застосовують, якщо межі первинного та вторинного групувань збігаються. В цьому випадку частоти підсумовують. В іншому випадку, якщо межі інтервалів первинного і вторинного групувань розрізняються між собою, застосо-

вують часткове перегрупування одиниць сукупності за формулою

$$\Delta f_1 = f_i \cdot k, \quad (3.3)$$

де Δf_1 – кількість одиниць сукупності, яку потрібно включити до знову створеної групи вторинного групування;

f_i – кількість одиниць сукупності у відповідній групі первинного групування;

k – частка одиниць сукупності, яку потрібно включити до знову створеної групи, обчислюється за формулою

$$k = \frac{X_{\max} - Y_{\min}}{Y_{\max} - Y_{\min}}, \quad (3.4)$$

X_{\max} – максимальне значення ознаки в кожній групі вторинного групування;

Y_{\max}, Y_{\min} – відповідно максимальне і мінімальне значення ознаки у відповідній групі первинного групування, між якими вміщується X_{\max} .

Тоді, кількість одиниць цієї групи первинного групування, що залишилися, необхідно включити в наступну групу вторинного групування, визначимо за формулою

$$\Delta f_2 = f_i - \Delta f_1 \quad (3.5)$$

Δf_2 – залишок одиниць сукупності відповідної групи первинного групування, який потрібно включити в наступну групу вторинного групування.

Розглянемо дані первинного групування працівників двох підприємств за стажем роботи, яке наведено в табл. 3.5.

Дані про розподіл працівників за стажом роботи не є порівняльними, оскільки різними є самі інтервали та їх кількість.

Таблиця 3.5

**Первинне групування працівників підприємств
за стажом роботи**

№ групи	Підприємство 1		Підприємство 2	
	Стаж роботи, років	Частка працівників, %	Стаж роботи, років	Частка працівників, %
1	До 2	23	До 5	20
2	2–5	15	5–8	30
3	5–10	25	8–15	28
4	10–15	12	15–25	12
5	15–20	11	25–30	8
6	20–30	9	30 і більше	2
7	30–35	5	х	х
Усього		100	Усього	100

Для одержання порівняльних даних необхідно їх перегрупувати, утворивши по обох підприємствах однакову кількість груп з рівними інтервалами. В обох підприємствах найбільш раціонально створити по чотири групи зі стажом роботи – до 10, 10–20, 20–30, 30 і більше років.

Розв'язання 1. Перегрупування працівників першого підприємства *способом укрупнення інтервалів*.

Так, у першу групу зі стажом роботи до 10 років, було включено працівників перших трьох груп, в групу зі стажом 10–20 років – увійшли працівники двох наступних груп. Дві останні групи залишились без змін.

Розв'язання 2. Перегрупування працівників другого підприємства шляхом *часткового перегрупування одиниць сукупності* зробили за допомогою спеціальних розрахунків.

На другому підприємстві в першу групу увійшли працівники двох перших груп ($20 + 30 = 50$) і частка працівників

СТАТИСТИКА

третьої групи ($\Delta f_1 = 8$), яка була визначена частковим перегрупуванням. Так, кількість (частку) працівників третьої групи (28) було поділено на величину інтервалу ($15 - 8 = 7$) та помножено на кількість одиниць інтервалу, які повинні бути включені до цієї нової групи ($10 - 8 = 2$), тобто $\left(\frac{28 \cdot 2}{7} = 8\right)$.

Таблиця 3.6

Вторинне групування працівників підприємств за стажем роботи

№ групи	Стаж роботи, років	Частка працівників, %	
		Підприємство 1	Підприємство 2
1	До 10	$23 + 15 + 25 = 63$	$20 + 30 + 8 (28/7 \cdot 2) = 58$
2	10–20	$12 + 11 = 23$	$20 + 6 (12/10 \cdot 5) = 26$
3	20–30	9	$6 + 8 = 14$
4	30 і більше	5	2
Усього		100	100

У групу зі стажем роботи 10–20 років було додано залишок працівників третьої групи (Δf_2), тобто ($28 - 8 = 20$) та частку працівників четвертої групи (Δf_1), яка була визначена таким чином $\left(\frac{12 \cdot 5}{10} = 6\right)$.

До четвертої групи було включено залишок працівників (Δf_2) від попередньої групи ($12 - 6 = 6$) та працівники п'ятої групи в кількості 8%, а остання група залишилась без змін.

У результаті вторинного групування дані по кожному підприємству приведено до порівняльного вигляду, що дозволяє зробити висновки про структуру кадрів за стажем роботи. Так, на першому підприємстві на 5% більше працівників, які мають стаж роботи до 10 років, а на другому підприємстві на 5% більше працівників, які мають стаж

роботи від 20 до 30 років. В інших групах відхилення є менш істотними.

Зауваження. Зведення та групування статистичних даних, а також перегрупування даних, можна здійснювати за допомогою вбудованих функцій MS Excel. Для цього є спеціальні команди в головному меню. Основою групування даних в середовищі Excel є функція:

– **Сортування та фільтр**, які дозволяють впорядкувати дані за якісними (атрибутивними) ознаками, наприклад, підприємств – за назвою, формами власності, організаційними формами та ін.

– **Сортування від мінімального до максимального** або **Сортування від максимального до мінімального** – за кількісними ознаками (за зростанням або спаданням ознаки).

– Відсортовані дані можна підсумувати за допомогою функції – **Проміжні підсумки у** відповідному діапазоні, у стовпчиках за показниками, які знаходяться в них, або **СУММ** – для загальної суми.

3.5. Ряди розподілу

Особливим видом групувань в статистиці є ряди розподілу, які є раціональною формою подання результатів зведення і групування матеріалів. Це є найпростіший спосіб узагальнення статистичних даних для наочного зображення типологічних і структурних групувань.

Означення. **Ряд розподілу** – це впорядковані та (або) згруповані за певною ознакою одиниці статистичної сукупності, які характеризують розподіл явища в цей період часу (наприклад, розподіл підприємств району за чисельністю працівників, сімей за кількістю дітей, робітників за тарифним розрядом та ін.).

Статистичні ряди розподілу являють собою упорядкований розподіл одиниць сукупності на групи за групувальною

ознакою. Вони характеризують склад (структуру) досліджуваного явища, дають змогу судити про однорідність сукупності.

Залежно від групувальної ознаки ряди розподілу поділяють на *атрибутивні та варіаційні*.

Атрибутивні ряди розподілу утворені за якісною ознакою. Прикладом таких рядів можуть бути розподіли населення за статтю, освітою; юридичних осіб за організаційними формами та ін.

Різновидом атрибутивних рядів розподілу є *альтернативні* ряди. **Альтернативними** називають ряди якісних ознак, які набувають тільки два взаємовиключних значення. Наприклад, розподіл продукції підприємства на якісну та неякісну, розподіл підприємств на прибуткові та збиткові тощо.

Ряди розподілу, побудовані за кількісною ознакою, що розташована у визначеному (висхідному чи низхідному), порядку, називають **варіаційними**.

Варіаційні ряди розподілу поділяють на **дискретні** (перервні) та **інтервальні** (безперервні).

Дискретні – це такі варіаційні ряди розподілу, в яких досліджувані ознаки набувають значення тільки цілих чисел. *Приклад* дискретного ряду розподілу наведено в табл.3.7.

Таблиця 3.7

Групи магазинів за кількістю касових апаратів

Кількість касових апаратів, шт. (x)	Кількість магазинів, шт., частота (f)	У % до підсумку, частка (W)
1	5	10
2	8	16
3	10	20
4	15	30
5	9	18
6	3	6
Усього	50	100

Інтервальними називають ряди розподілу, в яких варіанти подані у вигляді інтервалів. *Приклад* інтервального ряду розподілу працівників підприємства за віком наведено в табл. 3.8.

Ряди розподілу складаються із двох елементів: **варіантів** тобто значень досліджуваної групувальної ознаки (x), та **частот** (f) – чисел, які показують чисельність одиниць, що увійшли до кожної групи. Частоти можуть бути виражені як абсолютними величинами, тобто числами, які показують скільки разів окремі значення варіант повторюються у сукупності, так і відносними величинами – частками (частинами), за допомогою коефіцієнтів або відсотків до підсумку. Сума всіх частот варіаційного ряду в абсолютних величинах дорівнює загальній чисельності одиниць досліджуваної сукупності і називається обсягом сукупності, що позначається через n , тобто

$$f_1 + f_2 + f_3 + \dots + f_k = \sum f_i = n \quad (3.3)$$

Частки (W_i), тобто модифіковані частоти, визначаються діленням частот окремих варіант на загальне число одиниць сукупності за формулою

$$W_i = \frac{f_i}{\sum f_i} = \frac{f_i}{n}. \quad (3.4)$$

Якщо частоти виражені в частках одиниці (коефіцієнтах), їх сума дорівнює одиниці, тобто $\sum w_i = 1$.

Частки, виражені в коефіцієнтах, досить важко сприймаються у практичній роботі. Тому їх подають у зручнішому для сприйняття вигляді, помноживши на 100, тобто у відсотках (%).

Крім того, обчислюють також **кумулятивні або нагромаджені частоти** (S_f) і **кумулятивні частки** (S_w).

Нагромаджені частоти одержують підсумуванням частоти даної варіанти з попередніми частотами шляхом послідовного їх додавання до частот (часток) першої групи частот наступних груп. У підсумку отримують суму частот або часток варіантів від мінімального значення до останнього тобто обсяг сукупності (див. табл. 3.8).

Якщо нагромаджені частоти виражені відносними величинами – *частками*, то їх сума дорівнює одиниці або 100%.

Варіанти (x) – це вік працівників, а *частоти* (f) – чисельність працівників, *частки* (W_i) – питома вага, визначені в % та коефіцієнтах, *кумулятивні частоти* (S_f).

Для наочності та підвищення якості аналізу рядів розподілу їх зображають графічно. Графічне зображення варіаційного ряду розподілу називають *кривою розподілу*. Варіаційний ряд може бути у вигляді: полігону розподілу, гістограми або кумуляти.

Таблиця 3.8

Інтервальний варіаційний ряд розподілу працівників підприємства за віком

Номер групи	Вік працівників, років (x)	Кількість працівників, осіб (f)	Питома вага, частки (W_i)		Кумулятивні частоти (S_f)
			% до підсумку	коефіцієнти	
I	20–25	4	8	0,08	4
II	25–30	9	18	0,18	13(4 + 9)
III	30–35	10	20	0,20	23(13 + 10)
IV	35–40	12	24	0,24	35(23 + 12)
V	40–45	8	16	0,16	43(35 + 8)
VI	45–50	5	10	0,10	48(43 + 5)
VII	50–55	2	4	0,08	50(48 + 2)
Всього	x	50	100	1,00	-

Дискретні варіаційні ряди розподілу звичайно зображають у вигляді *полігону* (многокутника розподілу) у прямокутній системі координат, де на осі абсцис відкладають

значення варіант x , а на осі ординат – частоти f , або частки W_i .

За даними табл. 3.7 побудуємо *полігон розподілу* групи магазинів за кількістю касових апаратів (рис. 3.1).

Полігон розподілу показує, що більшість магазинів – 15 мають по 4 касових апарати.

Для зображення *інтервальних варіаційних рядів* розподілу застосовується *гістограма* у вигляді прямокутників, що прилягають один до одного. При побудові гістограми на осі абсцис відкладають інтервали варіантів, а на осі ординат – частоти (частки).

Над віссю абсцис будуються прямокутники, площа яких відповідає величинам добутків інтервалів на їх частоти. При рівних інтервалах ширина стовпчиків однакова, при нерівних – неоднакова. Якщо середини верхніх сторін прямокутників (середини інтервалів) з'єднати, то одержимо полігон розподілу.



Рис. 3.1. Полігон розподілу магазинів за кількістю касових апаратів

Джерело: розроблено та складено авторами.

За даними табл. 3.8 побудуємо *гістограму розподілу* 50 працівників за віком (рис. 3.2).

Якщо варіаційний ряд має групи з нерівними інтервалами, то частоти в окремих інтервалах безпосередньо не зіставні, оскільки їх величина залежить від ширини інтервалу. Для їх зіставності необхідно додатково обчислювати *щільність розподілу* та *відносну щільність розподілу*. Вони визначаються відповідно відношенням частоти до величини інтервалу та відношенням частки до величини інтервалу.

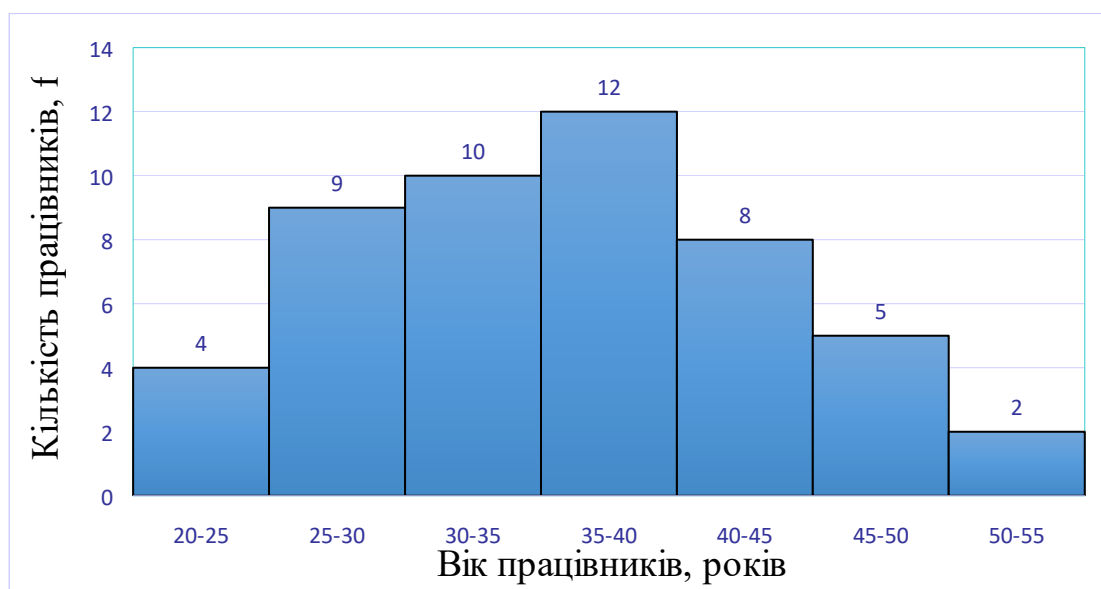


Рис. 3.2. Гістограма розподілу працівників підприємства за віком

Джерело: розроблено та складено авторами.

У деяких випадках зручніше користуватися нагромадженими (кумулятивними) частотами, якими замінюються значення чисельності окремих варіант. Варіаційний ряд із нагромадженими частотами на графіку зображується у вигляді кривої, яка одержала назву *кумуляти розподілу*.

Для побудови *кумуляти розподілу* спочатку підраховують нагромаджені частоти, послідовно підсумовуючи їх. *Приклад наведено* в табл. 3.9. Якщо розподіл має дискретний характер, то на графіку на осі абсцис відкладають значення варіант, а на осі ординат – нагромаджені частоти (частки).

Якщо розподіл має безперервний характер і має вигляд інтервального ряду, то на графіку будують точки, абсциси яких є праві (верхні) межі інтервалів, а ординати – відповідні їм нагромаджені (кумулятивні) частоти (частки).

Таблиця 3.9

Інтервальний варіаційний ряд розподілу 100 працівників за продуктивністю праці

Номер групи	Продуктивність праці працівників, шт. (x)	Кількість працівників, частота (f)	Нагромаджені (кумулятивні) частоти (S_f)
I	25–30	18	18
II	30–35	22	40
III	35–40	34	74
IV	40–45	18	92
V	45–50	8	100
Всього	x	100	-

За даними табл. 3.9 побудуємо кумулятивну криву розподілу 100 працівників за продуктивністю праці на одного працівника (рис. 3.3).

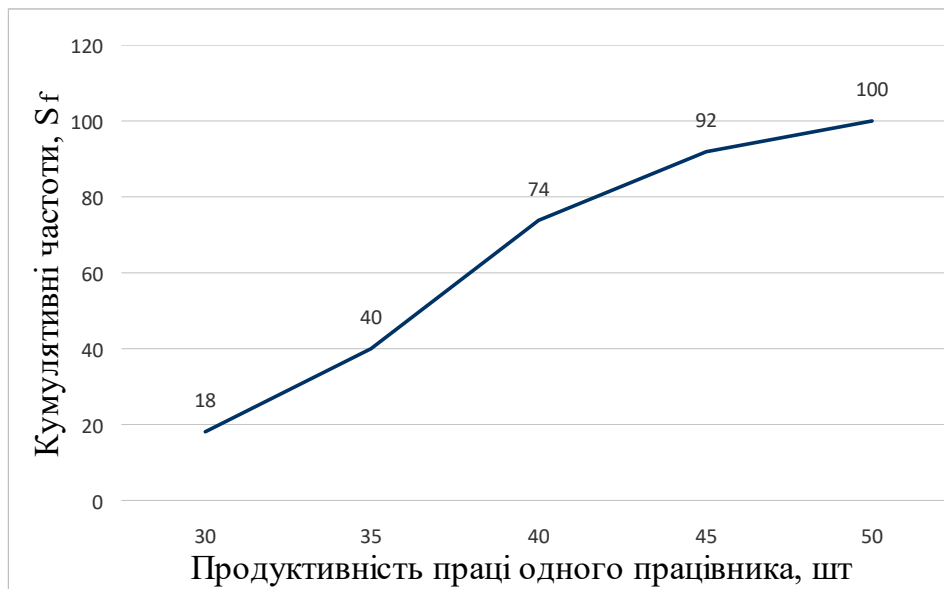


Рис. 3.3. Кумулята розподілу 100 працівників за продуктивністю праці

Зображення варіаційних рядів розподілу в табличній і графічній формах дає змогу одержати лише перше уявлення про найбільш загальні характеристики досліджуваного розподілу. Для всебічної характеристики рядів розподілу в подальшому вивчають найважливіші узагальнюючі характеристики: форму кривої розподілу, кількість вершин, симетричність (скошеність ряду розподілу), середні показники ряду, ступінь варіації ознаки.

3.6. Статистичні класифікації

Особливим різновидом групувань є *класифікації, класифікатори і номенклатури*, які входять до Державної системи класифікацій і кодування техніко-економічної та соціальної інформації України.

Означення Статистичні класифікації – це систематизований розподіл явищ і об'єктів на певні групи, класи, розряди на основі їх схожості або відмінностей.

Класифікації відрізняються від групувань тим, що групувальною основою класифікації є *атрибутивна (якісна)* ознака, вони більш *стійкі, сталі й стандартні*. В Україні їх затверджує Кабінет Міністрів чи міністерства та відомства. Класифікації мають історичний характер. Із часом з'являються нові класифікації, а в старі вносять якісь зміни.

У класифікації кожному значенню групувальної ознаки присвоюють код, який слугує засобом ідентифікації елемента з відповідним значенням ознаки. Процедура обробки даних за допомогою сучасних комп'ютерних технологій стала можливою завдяки використанню в класифікаціях кодів. Це суттєво прискорило час отримання кінцевих результатів статистичних досліджень.

Класифікатор – це систематизований перелік об'єктів, наприклад, (класифікатор об'єктів адміністративно-територіального устрою України – КОАТУУ, класифікатор професій –

КП, статистичний класифікатор органів державного управління – СКОДУ).

Номенклатура – це перелік систематизованих певним чином об’єктів. Іноді термін «номенклатура» вживається як синонім терміна «класифікація». Наприклад, номенклатура продукції промисловості (НПП), номенклатура продукції будівництва (НПБ), номенклатура продукції сільського господарства (НПСГ).

У своїй діяльності органи державної статистики використовують понад 20 класифікацій (класифікаторів), у тому числі такі:

1. Класифікація видів економічної діяльності (КВЕД);
2. Статистична класифікація продукції (СКП);
3. Українській класифікатор товарів зовнішньоекономічної діяльності (УКТЗД);
4. Класифікація організаційно-правових форм господарювання (КОПФГ);
5. Статистичний класифікатор органів державного управління (СКОДУ);
6. Класифікатор професій (КП).

Базовим інструментом аналізу структури національної економіки слугує класифікація видів економічної діяльності – **КВЕД** (коротка назва ДК 009:2010; КВЕД-2010).

Згідно з міжнародними стандартами у КВЕД усі види економічної діяльності поділені на добувні, обробні і такі, що надають послуги. Так, до добувних належать сільське господарство, мисливство та лісове господарство, рибне господарство та власне добувна промисловість. До обробних – обробна промисловість, виробництво електроенергії, газу та води. Решта видів діяльності, наприклад, будівництво, торгівля, готелі, ресторани, охорона здоров’я, наука, державне управління, вважаються такими, що надають послуги.

КВЕД – це статистичний інструмент для впорядкування економічної інформації, який призначений забезпечувати:

– статистичний облік підприємств і організацій за видами економічної діяльності;

– проведення державних статистичних спостережень економічної діяльності й аналізу статистичної інформації на макрорівні;

– зіставлення національної статистичної інформації з міжнародною через застосування єдиної статистичної термінології, статистичних одиниць і принципів визначення та змінення видів економічної діяльності підприємств і організацій.

Код виду економічної діяльності суб'єкта, який зазначається за КВЕД, є одним з основних показників стратифікації статистичної інформації під час організації суцільних та вибірових обстежень.

Ця класифікація базується на стандартній міжнародній класифікації видів діяльності Європейського Союзу. Об'єктами класифікації в КВЕД є види економічної діяльності статистичних одиниць.

КВЕД 2010 має ієрархічну систему кодування із застосуванням літерно-цифрового коду. Літерами позначаються секції від А до Q (21 секція). Подальша деталізація класифікаційних угруповань КВЕД – розділ, група, клас – здійснюється за допомогою цифрових кодів.

Загальне кодове позначення об'єктів КВЕД має таку схему:

Y XX.XX

де Y – Секція (від А до Q)

XX. – Розділ;

XX.X – Група;

XX.XX – Клас.

КВЕД гармонізовано за статистичною термінологією та визначенням статистичних одиниць, прийнятих у Європейському Союзі, та міжнародною класифікацією.

За КВЕД визначається основний вид економічної діяльності підприємств і організацій за методом «top down», у відповідності з яким це підприємство включають в ту чи іншу секцію (розділ, групу, клас).

Інші класифікатори (класифікації) також утворюють основу для збору, обробки, аналізу та поширення великих обсягів даних у стандартному, зрозумілому для користувача форматі. Вони забезпечують єдність понять, які використовуються для опису масових явищ і процесів, таких як економічна діяльність, продукція, витрати, зайнятість тощо, і дають змогу порівнювати дані у межах країни, між країнами, за регіонами. Класифікатори є невід'ємною частиною набору інструментів, за допомогою яких проводяться державні статистичні спостереження та комплексні статистичні роботи.

В Україні ведеться *Єдиний державний реєстр підприємств і організацій України (ЕДРПОУ)*.

Це автоматизована система збирання, накопичення та опрацювання даних про юридичних осіб усіх форм власності та організаційно-правових форм господарювання, що знаходяться на території України та провадять свою діяльність на підставі її законодавства. При включенні до ЕДРПОУ суб'єкти класифікуються відповідно до Національних (державних) статистичних класифікацій (класифікаторів).

ЕДРПОУ є важливим компонентом інфраструктури державної статистики України. Ведення реєстру здійснюється органами державної статистики з 1993 року відповідно до Закону України «Про державну статистику» та Закону України «Про державну реєстрацію юридичних осіб та фізичних осіб-підприємців».

На державному рівні ЕДРПОУ містить інформацію щодо суб'єктів загалом по країні, на територіальному рівні щодо суб'єктів на відповідній території. Підставою для внесення до ЕДРПОУ або виключення з нього даних щодо суб'єктів, а також внесення змін до ЕДРПОУ є надходження від державного реєстратора до органу державної статистики інформації про вчинення реєстраційних дій.

3.7. Статистичні таблиці

Результати статистичного зведення і групування оформлюються у вигляді статистичних таблиць.

Означення. Статистична таблиця – це форма систематизованого, раціонального, наочного і компактного викладання статистичних даних.

Макет таблиці (рис. 3.4) містить загальний заголовок, який коротко і точно характеризує зміст таблиці, боковий і верхній заголовки як сукупність рядків і граф. Статистичну таблицю можна порівняти з реченням: вона складається з *підмета і присудка*.

Підметом таблиці є складові елементи досліджуваного об'єкта або їх групи, які розміщують у лівій частині таблиці, і підлягають характеристиці й вивченню.

Присудком таблиці – є показники, які розміщують у вертикальних графах правої частини таблиці.

Складена і оформлена статистична таблиця повинна мати загальний заголовок, який розміщується над таблицею і виражає у стислій формі її зміст. Бокові заголовки, розміщуються зліва, вони розкривають зміст рядків підмета. Верхні заголовки розміщуються в графах і розкривають зміст статистичного присудка.

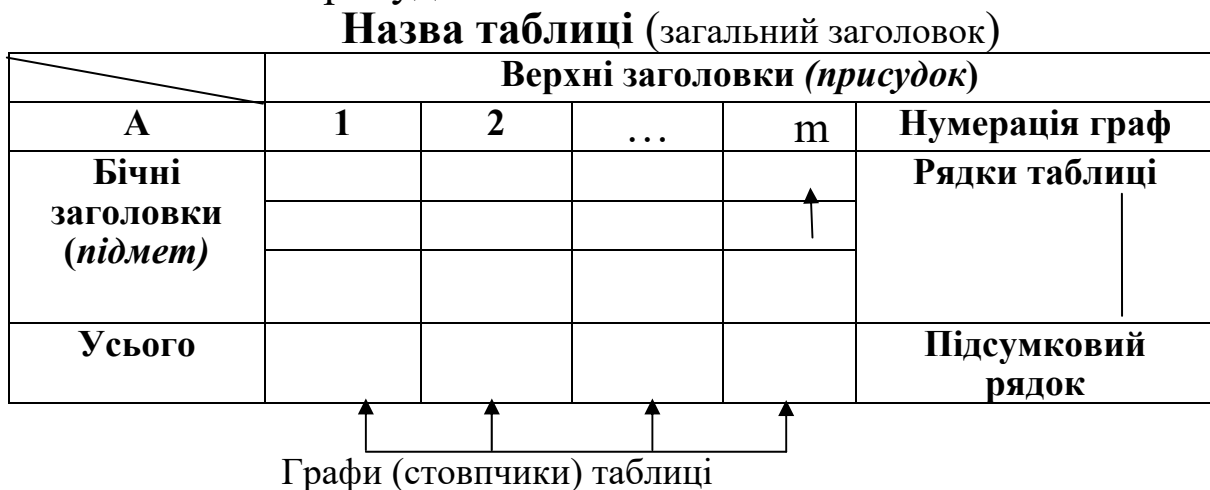


Рис. 3.4. Макет статистичної таблиці

Джерело: опрацьовано авторами на основі [2, 58].

Загальна кількість клітинок, які можна заповнити числами, визначає *розмір таблиці*. Він дорівнює добутку кількості рядків і кількості граф. Наприклад, 4 рядка і 5 граф складають розмір таблиці, що дорівнює 20.

Таблиці можуть бути *простими, груповими і комбінаційними*.

Простими називаються таблиці, *підмет яких* містить перелік об'єктів без групування їх за будь-якою ознакою.

Наприклад, перелік областей України, перелік підприємств, видів економічної діяльності. Прості таблиці є найбільш поширеними.

Групові таблиці – це таблиці, *підмет яких* розбито на групи за якоюсь істотною ознакою. Наприклад, групування магазинів за обсягом товарообороту, собівартістю продукції за статтями витрат, населення – за віком, статтю або національністю.

Комбінаційні таблиці – це таблиці, в яких підмет згруповано на окремі групи за двома і більше ознаками, пов'язаними між собою. Приклад, розподіл підприємств – за формами власності й обсягом виробленої продукції; населення – за областями України, за віком та статтю.

Для побудови статистичних таблиць важливим моментом є розроблення присудка, який складається із системи показників для повної характеристики виділених груп та їх істотних особливостей.

Правила складання таблиць:

1. Таблиця повинна бути компактною, наочною та виразною, щоб легше було читати і аналізувати її.

2. У таблиці має бути загальний заголовок (назва), розміщена над нею. Заголовок повинен відображати у стислій виразній формі зміст таблиці, включати об'єкт дослідження, територію і період дослідження, до яких відносяться наведені дані, а також одиницю виміру, якщо вона єдина для всієї сукупності.

3. У рядках та графах також мають бути заголовки. Кожний заголовок повинен бути точним, коротким і виразним. У графах проставляються одиниці виміру.

4 Рядки в підметі та графі у присудку потрібно нумерувати:

рядки – цифрами, колонки – літерами. Абсолютні дані в межах однієї графі потрібно округлювати з однаковим ступенем точності.

5. Відсутність значень ознаки у певних одиниць позначають знаком (тире), а відсутність відомостей – трьома крапками.

Якщо дані є, але їх числові значення менші за прийнятий у графі ступінь точності обчислення, записують 0,0 або 0,00.

6. У таблиці повинні бути усі необхідні підсумки: групові (проміжні) і загальні. Підсумковий рядок виділяють від основного тексту таблиці (подвійною чи товщою лінією, більшим інтервалом, шрифтом, кольором).

3.8. Задачі для самостійної роботи

Задача 3.1. За наведеними даними про дослідників наукових розробок визначити вид групування.

Таблиця 3.10

Групування дослідників, задіяних у виконанні наукових досліджень і розробок, за рівнем освіти (осіб)

Освіта дослідників наукових досліджень і розробок	Усього	
	2016 рік	2017 рік
Мають вищу освіту:		
- доктори наук	7071	6925
- доктори філософії (кандидати наук)	20085	19094
- магістри (спеціалісти)	34694	31789
Бакалаври (молодші бакалаври, молодші спеціалісти)	1844	1584
Усього	63694	59392

Задача 3.2. За наведеними даними про розподіл наявного населення України за місцем проживання визначте вид групування.

Таблиця 3.11

Населення України за місцем проживання (на 1 січня)

Наявне населення, млн осіб	2015	2016	2017	2018
Міське	29,7	29,6	29,5	29,4
Сільське	13,2	13,2	13,1	13,0
Усього	42,9	42,8	42,6	42,4

Задача 3.3. За наведеними даними про розподіл групи сімей за кількістю дітей визначте вид групування.

Таблиця 3.12

Кількість дітей у сім'ях	Кількість сімей	В % до підсумку
1	18	36
2	13	26
3	8	16
4	5	10
5	3	6
6	2	4
7	1	2
Всього	50	100

Задача 3.4. Підприємства мають таку чисельність працівників: 25; 30; 16; 50; 19; 40; 35; 80; 46; 58; 95; 39; 40; 68; 80; 48; 55; 75; 62; 84; 35; 65; 70; 48; 24; 18; 72; 64; 15; 22.

Побудуйте інтервальний варіаційний ряд розподілу підприємств за чисельністю працівників, створивши 4 групи з рівними інтервалами.

Задача 3.5. Є дані вибіркового обстеження земельної площі 20 ділянок фермерських господарств.

Таблиця 3.13

Площа ділянок, га

5,6	7,2	4,0	10,5	8,1
7,9	4,3	5,7	9,5	7,7
8,4	4,8	6,9	9,2	8,5
12,8	6,8	8,9	3,8	12,6

Побудуйте варіаційний ряд розподілу фермерських господарств за розміром земельної площі, для чого утворити три групи з рівними інтервалами. Зобразити результати графічно.

Задача 3.6. За даними вибіркового обстеження 24 домогосподарств у сільській місцевості району кількість членів домогосподарств становить:

2	5	6	5	4	4
6	3	4	3	2	5
4	2	3	4	3	6
3	5	4	6	5	3

Побудуйте варіаційний ряд розподілу домогосподарств за кількістю членів. Зобразити результати графічно.

Задача 3.7. Підприємства міста характеризуються такими даними, млн грн:

Таблиця 3.14

№ підприємства	Обсяг виробленої продукції	Прибуток від реалізації	№ підприємства	Обсяг виробленої продукції	Прибуток від реалізації
1	2	3	1	2	3
1	85,7	7,8	13	120,0	16,5
2	100,0	9,5	14	87,0	5,7
3	69,0	7,2	15	98,4	9,0
4	75,0	5,6	16	68,4	4,3
5	58,5	5,7	17	59,8	7,0
6	55,0	6,8	18	102,7	10,8
7	88,7	9,3	19	94,0	8,0

Розділ 3. ЗВЕДЕННЯ ТА ГРУПУВАННЯ СТАТИСТИЧНИХ ДАНИХ

Закінчення табл. 3.14

№ підприємства	Обсяг виробленої продукції	Прибуток від реалізації	№ підприємства	Обсяг виробленої продукції	Прибуток від реалізації
8	110,0	12,8	20	109,0	18,0
9	106,0	16,0	21	47,0	8,0
10	78,5	9,8	22	57,2	5,0
11	95,5	20,5	23	85,0	4,9
12	92,5	15,8	24	135,0	17,0

1. За наведеними даними складіть комбінаційне групування підприємств з рівними інтервалами, розподіливши їх на 4 групи за обсягом виробленої продукції, та на 3 групи за обсягом прибутку.

2. Складіть аналітичне групування за обсягом виробленої продукції (факторна ознака) та величиною прибутку (результативна ознака).

Результати кожного групування оформите у таблиці, проаналізуйте отримані показники та зробіть висновки.

Задача 3.8. Первинне групування юридичних осіб за організаційними формами характеризується такими даними:

Таблиця 3.15

Юридичні особи за організаційними формами

(на 01.01 2018 року, тис. осіб)

Організаційні форми	Кількість юридичних осіб	В % до підсумку
Фермерське господарство	45	3,6
Приватне підприємство	200,3	16,3
Державне підприємство	3,9	0,30
Комунальне підприємство	11,6	0,90
Акціонерні товариства	14,7	1,2

СТАТИСТИКА

Закінчення табл. 3.15

Організаційні форми	Кількість юридичних осіб	В % до підсумку
Товариство з обмеженою відповідальністю	576,6	46,7
Кооперативи	27	2,2
Державна організація (установа, заклад)	10,7	0,90
Громадська організація	80,5	6,5
Релігійна організація	25,2	2,04
Профспілка, об'єднання профспілок	27,6	2,20
Об'єднання співвласників багатоквартирного будинку	28	2,3
Інші організаційні форми	183,9	14,9
<i>Усього юридичних осіб</i>	<i>1235</i>	<i>100</i>

Джерело : Статистичний щорічник України за 2017 рік. Державна служба статистики України. – Київ, 2018. – С. 245.

1. За допомогою функцій MS Excel систематизувати дані за кількістю юридичних осіб у організаційних формах у порядку зростання.

2. Провести перегрупування юридичних осіб за організаційними формами, розподіливши їх на 3 групи за кількістю юридичних осіб в організаційних формах: малій, середньої та великої чисельності на ваш розсуд.

3. Визначити організаційні форми, які потрапили в кожену групу. Розрахувати суму і частку (в % до підсумку) кількості юридичних осіб організаційних форм за чисельністю в кожній групі та сукупності загалом.

Результати перегрупування оформити у вигляді таблиці. Проаналізувати отримані показники та зробити висновки.

Задача 3.9. Первинне групування регіонів за рівнем номінальної заробітної плати працівників характеризується такими даними:

Таблиця 3.16

Середньомісячна номінальна заробітна плата штатних працівників за регіонами в 2017 році

Регіони	Зарплата, грн	Регіони	Зарплата, грн
Вінницька	6121	Одеська	6542
Волинська	5849	Полтавська	6551
Дніпропетровська	6939	Рівненська	6013
Донецька	7764	Сумська	5946
Житомирська	5836	Тернопільська	5554
Закарпатська	6355	Харківська	6244
Запорізька	6863	Херсонська	5842
Івано-Франківська	6074	Хмельницька	5938
Київська	7188	Черкаська	6042
Кіровоградська	5792	Чернівецька	5621
Луганська	5862	Чернігівська	5636
Львівська	6391	Місто – Київ	11135
Миколаївська	6709	Усього:	7104
		Україна	

Джерело: Статистичний щорічник України за 2017 рік. Державна служба статистики України, с. 72.

1. За допомогою сортування в Ексел систематизувати дані за величиною номінальної заробітної плати за регіонами України у порядку зростання.

2. Здійснити групування регіонів України за величиною середньої заробітної плати, виділивши 3 групи з рівними інтервалами.

3. Визначити регіони, які потрапили в кожену групу, та розрахувати середній розмір заробітної плати у кожній групі та сукупності загалом.

Результати групування оформити у вигляді таблиці, проаналізувати отримані показники та зробити висновки.

Задача 3.10. Маємо дані про розподіл підприємств промисловості за вартістю основних засобів та обсягом реалізованої продукції за групами.

Таблиця 3.17

Розподіл підприємств промисловості за вартістю основних засобів та обсягом реалізованої продукції

Групи підприємств за вартістю основних засобів, млн грн	Кількість підприємств, %	Обсяг реалізованої продукції, %
1–5	5,2	2
5–10	7,8	10,6
10–20	26,9	15,8
20–40	48,5	58,2
40–60	11,3	13,4
Усього	100	100

Здійснити вторинне групування підприємств, виділивши такі групи підприємств за розміром основних засобів, млн грн: 1–10; 10–25; 25–50; 50 і більше. Проаналізувати отримані результати.

Задача 3.11. Побудувати типологічне групування торговельних підприємств за видами товариств. Дослідити їх за кількістю працівників та обсягом товарообороту. Проаналізувати отримані результати.

Таблиця 3.18

Товариства за організаційними формами

Організаційна форма товариств	Кількість працівників, осіб	Обсяг товарообороту, млн грн
ПАТ	110	68
ПрАТ	150	70
ТОВ	80	38
ПрАТ	170	95
ТОВ	55	26
ПрАТ	125	115
ТОВ	40	21
ТОВ	35	24
ПАТ	130	45
ТОВ	28	52
ПрАТ	128	179
ПрАТ	140	205
ТОВ	60	54
ТОВ	55	45
ТОВ	42	37
ПрАТ	180	196
ТОВ	47	29
ПрАТ	200	220
ПАТ	75	33

Задача 3.12. З метою аналізу охорони здоров'я за даними про кількість лікарняних ліжок за регіонами України:

1) за допомогою MS Excel побудувати **структурно-типологічне** групування;

2) виконати перегрупування даних за кількістю лікарняних ліжок за регіонами, розподіливши регіони на 3 групи;

3) розрахувати середні показники в кожній групі;

4) проаналізувати отримані результати.

Розрахунки виконати за допомогою MS Excel.

Таблиця 3.19

**Кількість лікарняних ліжок за регіонами України
в 2017 р.**

Регіони	Тис. одиниць	Регіони	Тис. одиниць
Вінницька	10,8	Одеська	18,5
Волинська	7,3	Полтавська	11,1
Дніпропетровська	28,3	Рівненська	8,6
Донецька	12,8	Сумська	9
Житомирська	8,5	Тернопільська	8,7
Закарпатська	8,5	Харківська	22,4
Запорізька	14,4	Херсонська	8
Івано-Франківська	10,5	Хмельницька	10
Київська	12,7	Черкаська	9,8
Кіровоградська	8,1	Чернівецька	6,6
Луганська	6,4	Чернігівська	9,4
Львівська	20,8	Місто – Київ	29,4
Миколаївська	8,1	Усього по Україні	308,7

3.9. Питання для самоконтролю

1. Що таке статистичне зведення, які його завдання?
2. Які види статистичного зведення Ви знаєте? Коротко їх охарактеризувати.
3. Що таке централізоване і децентралізоване зведення?
4. У чому полягає суть статистичного групування?
5. Які завдання вирішує статистика за допомогою методу групувань?
6. Назвати види статистичних групувань.
7. Що таке групувальна ознака? Назвати види групувальних ознак та навести приклади.

8. Яку групувальну ознаку покладено в основу типологічних групувань?
9. У чому полягає сутність аналітичних групувань?
10. Як визначається кількість груп в типологічних групуваннях?
11. Як визначається кількість груп в групуваннях за дискретною та безперервною ознаками?
12. Які бувають інтервали групувань і як позначають їх межі?
13. Які групування називають комбінаційними? Навести приклади.
14. Пояснити суть вторинного групування. Які є способи побудови вторинного групування?
15. Що таке ряди розподілу? Назвати складові елементи ряду розподілу.
16. Що називають кривою розподілу?
17. У яких випадках ряд розподілу зображають за допомогою полігону, гістограми, кумуляти?
18. Що таке макет статистичної таблиці? Назвати складові елементи статистичних таблиць.
19. Що таке статистичні класифікації (класифікатори)? Назвати основні класифікації України.
20. Охарактеризувати сферу призначення КВЕД та його систему кодування.

Розділ 4

АБСОЛЮТНІ ТА ВІДНОСНІ СТАТИСТИЧНІ ВЕЛИЧИНИ

План

- 4.1. Суть і види статистичних показників.
- 4.2. Абсолютні статистичні величини, одиниці їх вимірювання.
- 4.3. Відносні статистичні величини, їх види та форми виразу.
- 4.4. Задачі для самостійної роботи.
- 4.5. Питання для самоконтролю.

4.1. Суть і види статистичних показників

Інформація про суспільні явища і процеси створюється, передається і зберігається у вигляді **статистичних показників**. Вони є однією з основних економічних категорій, за допомогою яких відображають кількісну і якісну сторони явищ і процесів.

З філософського погляду *статистичний показник* – це міра, тобто єдність якісного і кількісного відображення певної властивості соціально-економічного явища чи процесу.

Якісний зміст показника визначається суттю явища і відбивається в його назві. Кількісна сторона подається числом та його вимірником. Оскільки статистика вивчає суспільні явища в конкретних умовах простору і часу, значення будь-якого показника визначається щодо цих атрибутів.

Сполучною ланкою між якісним змістом і числовим вираженням є правило побудови – *модель показника*, яка розкриває його статистичну структуру. У моделі обґрунтовуються одиниці, взяті для вимірювання, технологія отримання даних, обчислювальні операції.

Модель показника має надзвичайно важливе значення для забезпечення вірогідності статистичної інформації. Як видно з рис. 4.1, від обґрунтованості моделі залежать обидва

аспекти вірогідності такої інформації: адекватність відображення явища і точність вимірювання.

Адекватність розглядається як здатність показника відбити саме ту властивість, яка передбачена програмою дослідження.

Точність і повнота вимірювання залежать від можливостей обліку, організації збору та обробки даних. Отже, щоб показник відповідав своєму призначенню і виконував покладені на нього функції, на стадії його проектування потрібно з усіх боків логічно та статистично обґрунтувати модель.

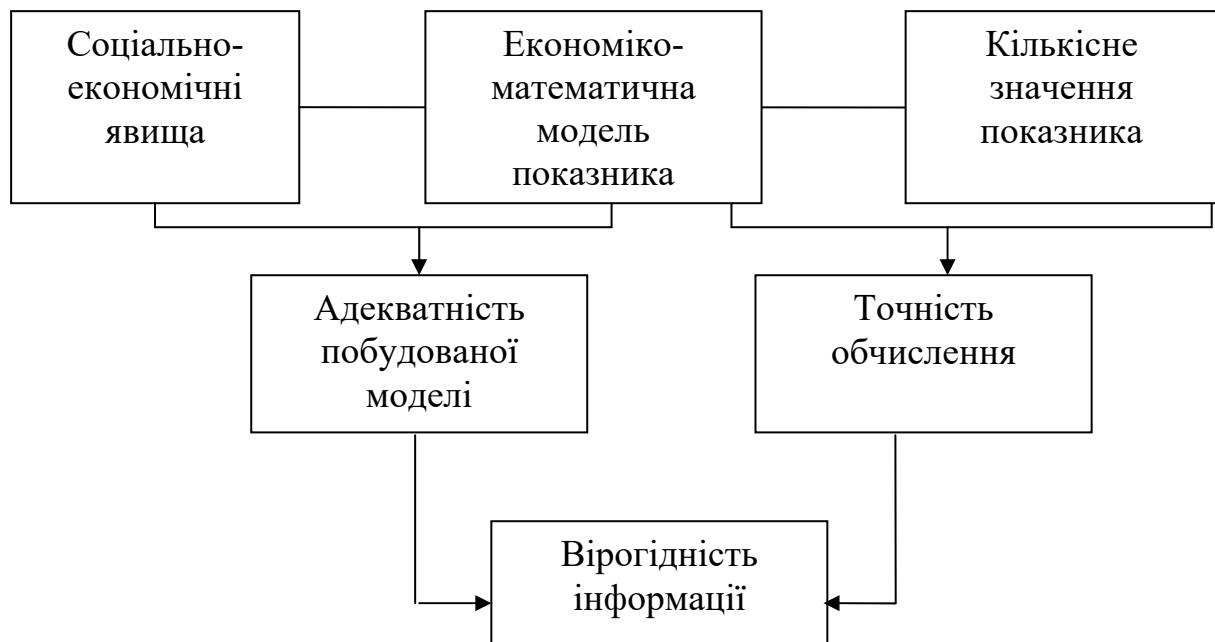


Рис. 4.1. Зв'язок статистичної моделі показника та вірогідності інформації

Джерело: розроблено автором.

Статистичні показники розрізняють за способом обчислення, ознакою часу та аналітичними функціями.

За способом обчислення розглядають *первинні і похідні* показники. Первинні визначаються зведенням даних статистичного спостереження й подаються у формі абсолютних величин. Похідні показники обчислюються на базі первинних

або похідних показників. Вони мають форму середніх або відносних величин.

Окрему групу становлять *взаємообернені показники* – пара характеристик одного і того ж явища, але *прямий показник* змінюється в напрямі зміни явища, а *обернений* – у протилежному напрямі. Прямий показник зростає з підсиленням явища, обернений, навпаки, зменшується. Наприклад, споживчі ціни (прямий) та купівельна спроможність грошової одиниці (обернений).

Означення. Статистичний показник – це узагальнююча характеристика явищ і процесів, в якій поєднується їх якісна та кількісна визначеність. Якісний бік показника визначається сутністю явища (що відбивається в його назві), а кількісний бік – числовим значенням та відповідною одиницею виміру. Якісний зміст показника поєднується з його числовим вираженням за допомогою моделі показника, що містить його назву, алгоритм розрахунку, одиницю, застосовувану під час вимірювання, а також місце та час, на який припадає явище чи процес.

Показники поділяються на види залежно від їх аналітичної функції, способу обчислення та ознаки часу.

За *аналітичною функцією* розглядають показники, що відбивають обсяг явища, його середній рівень, інтенсивність прояву, структуру, зміну в часі або порівняння у просторі.

За ознакою часу розрізняють *інтервальні* та *моментні показники*. *Інтервальні* характеризують явище за певний період часу (місяць, квартал, рік). Наприклад, річний експорт товарів та послуг, середньомісячні сукупні витрати на душу населення. *Моментні показники* характеризують явище за станом на певний момент часу: зовнішній борг держави на початок року, кількість банківських установ або туристичних фірм на кінець року.

Отже, такі показники подаються через абсолютні та відносні величини.

4.2. Абсолютні статистичні величини, одиниці їх вимірювання

Абсолютними величинами у статистиці називають первинні узагальнюючі показники, які характеризують суспільні явища і процеси в конкретних умовах місця і часу.

Абсолютні величини як узагальнюючі показники характеризують сукупність за її *чисельністю* (число працівників, кількість магазинів, лікарень) і *обсягом* (валовий випуск продукції, фонд заробітної плати, обсяг роздрібного товарообороту і т.д.).

У статистиці використовується велике число різноманітних одиниць виміру, серед яких найбільш поширеним є: натуральні, вартісні, трудові, одиниці часу та сукупності.

Натуральними називаються одиниці виміру, які виражають розміри конкретних явищ у фізичних вимірниках (тоннах, кілограмах, метрах, гектарах, літрах, кубометрах і ін.).

Натуральні одиниці виміру можуть бути простими, складними і умовно-натуральними. Прості величини згадані вище.

Складні натуральні одиниці виміру отримують шляхом перемноження двох величин різних розмірностей. Наприклад, потужність електродвигунів вимірюється в кіловатах, а спожита ними енергія в складних одиницях – кіловат-годинах, обсяг перевезених вантажів вимірюється в тоннах, а вантажообіг – в тонно-кілометрах, верстатний парк цеху обчислюється в штуках, а робота верстатів у верстато-днях, верстато-змінах і т.д.

В ряді випадків статистика використовує умовно-натуральні одиниці виміру. Такі одиниці виміру використовуються для зведення до купи декількох різновидностей однакової споживчої вартості. Одну з них приймають за еталон, а всі інші перераховують за допомогою спеціальних перевідних коефіцієнтів в одиниці виміру взятого еталону. Перерахунок в умовно-натуральні одиниці здійснюють за формулою:

СТАТИСТИКА

$$V_{ун} = V_0 + \sum V_i \times K_i, \quad (4.1)$$

де $V_{ун}$ – випуск продукції в умовно-натуральних одиницях виміру;

V_0 – випуск продукції, яка прийнята за еталон;

V_i – випуск інших видів продукції;

K_i – коефіцієнт переведу в умовно-натуральні одиниці виміру:

$$K_i = \frac{X_i}{X_0}, \quad (4.2)$$

де X_0 , X_i – якісна або кількісна характеристика відповідної продукції, яка прийнята за еталон, та інших видів продукції.

Приклад 4.1. За звітний період підприємство випустило таку кількість зошитів:

Таблиця 4.1

Зошити	Кількість аркушів у зошиті, с.	Випуск, шт.
Шкільні	12	4200
Загальні	36	3500
Загальні	60	1900
Загальні	84	1200
Загальні	96	950

Визначити загальну кількість виробленої продукції підприємством в умовно-натуральних одиницях виміру. За умовну одиницю обчислювання прийняти шкільний зошит обсягом 12 аркушів.

Розв'язання. Оскільки згідно з умовою задачі, умовною одиницею є шкільний зошит обсягом 12 аркушів, то його беремо за одиницю. Використовуючи вище наведену формулу, розра-

ховуємо загальний випуск зошитів в умовно-натуральних одиницях:

$$\begin{aligned} \text{Вун}_{(12 \text{ арк})} &= 4200 + 3500 \cdot \frac{36}{12} + 1900 \cdot \frac{60}{12} + 1200 \cdot \frac{84}{12} + 950 \cdot \frac{96}{12} = \\ &= 40200 \text{ шт.} \end{aligned}$$

Вартісними називаються одиниці виміру, які використовуються для характеристики в грошовому виразі багатьох різноманітних статистичних показників. Наприклад, собівартість і ціна одиниці продукції обліковується в гривнях і копійках, обсяг товарообороту продуктового магазину – в тисячах гривень, а валовий внутрішній продукт держави в мільйонах або мільярдах гривень.

Трудовими називаються одиниці виміру, які використовуються для обліку затрат робочого часу, для визначення рівня продуктивності праці, величини трудових ресурсів і раціонального їх використання та ін.

Трудові вимірники виражаються в людино-годинах, людино-роках, людино-днях, верстато-днях.

Існує певна множина абсолютних величин, які обліковуються у формі балансу. Така форма передбачає розрахунок показника за джерелами формування та напрямками використання, а це дає змогу визначити не лише сумарний показник, а й усі його компоненти. Складаються баланси підприємств, матеріальні баланси найважливіших продуктів, палива, трудових ресурсів і т. ін.

Абсолютні величини є основою для обчислення різних видів відносних і середніх величин, індексів та інших узагальнюючих показників.

У деяких випадках статистика використовує одиниці рахунку часу, життя в роках, строк служби будівель і таке інше.

Також одиниці сукупності відображають кількість або частоту відповідного показника в рядах розподілу.

Статистика виділяє два види абсолютних величин: індивідуальні і загальні.

Індивідуальні величини отримують в процесі статистичного спостереження і які виражають розміри кількісних ознак окремих одиниць досліджуваної сукупності.

Загальні величини отримують на стадії статистичного зведення і які виражають величину ознаки у всіх одиниць цієї сукупності, або окремих її груп.

4.3. Відносні статистичні величини, їх види та форми виразу

Абсолютні статистичні величини мають незаперечне значення в системі управління, проте поглиблений соціально-економічний аналіз фактів потребує різного виду порівнянь. Порівнюються значення статистичних показників у часі (за одним об'єктом), у просторі (між об'єктами), співвідносяться різні ознаки одного й того ж об'єкту. Результатом порівняння є відносні статистичні величини.

Означення. Відносними величинами називають статистичні показники, які виражають кількісні співвідношення між соціально-економічними явищами і процесами. Їх отримують шляхом порівняння (ділення) двох однойменних, або різнойменних величин.

Величина, за якою порівнюють, називається основою відносної величини, базою порівняння або базисною величиною. Величина, яку порівнюють, називається поточною, порівнюваною чи звітною величиною.

Відносні величини показують у скільки разів порівнювана величина більша (менша) за базисну, або яку частку перша займає в другій, або скільки одиниць однієї величини припадає на одиницю іншої.

Залежно від бази порівняння відносні величини можуть виражатись у вигляді:

а) коефіцієнтів – якщо база порівняння приймається за одиницю;

б) процентів (%) – якщо база порівняння береться за 100;

в) проміле (‰) – якщо за базу порівняння взято 1000;

г) продесиміле (‱) – якщо база порівняння становить 10 000;

д) просантиміле (‱‱) – якщо база порівняння прийнята за 100 000.

Різноманітність співвідношень і пропорцій реального життя для свого відображення потребує різних за змістом і статистичною природою відносних величин. Залежно від змісту і пізнавального значення відносні величини, що використовуються в статистиці, поділяють на такі основні види: відносні величини планового завдання, виконання плану, динаміки, структури, координації, інтенсивності і порівняння.

Усі суб'єкти фінансово-господарської сфери, незалежно від форми власності, як правило, планують результати своєї діяльності і порівнюють реально досягнуті результати з наміченими. У такому випадку для оцінювання напруженості планового завдання і міри його виконання застосовуються перші три види відносних показників: показник планового завдання (*ВВПЗ*), показник виконання плану (*ВВПП*) та показник динаміки (*ВВД*).

Відносна величина планового завдання (*ВВПЗ*) – показує на скільки відсотків передбачається змінити рівень показника у звітному періоді порівняно з базисним. Розраховується у формі коефіцієнта чи відсотка, тому і позначається як *коефіцієнт планового завдання*:

$$K_{\text{пз}} = \frac{U_{\text{пл}}}{U_0}, \quad (4.3)$$

де $U_{\text{пл}}$ – плановий рівень показника у звітному періоді.

U_0 – фактичний рівень показника у базисному періоді.

Відносна величина виконання плану (ВВВП) – показує на скільки відсотків фактично перевиконано або не довиконано планове завдання у звітному періоді. Тобто, це процентне відношення фактично досягнутого рівня до запланованого за відповідний період часу (місяць, квартал і т.д.). Розраховується у формі коефіцієнта чи відсотка.

$$K_{\text{вп}} = \frac{Y_n}{Y_{\text{пл}}}, \quad (4.4)$$

де Y_n – фактичний рівень показника у звітному періоді

Відносна величина динаміки (ВВД) – показує на скільки відсотків змінився рівень показника у звітному періоді порівняно з базисним. Залежно від характеру бази порівняння, розрізняють відносні величини динаміки із змінною базою порівняння, або ланцюгові, і відносні величини з постійною базою порівняння, або базисні. Розраховується у формі коефіцієнта чи відсотка.

$$K_d = \frac{Y_n}{Y_0} \quad (4.5)$$

Між відносними величинами планового завдання, виконання плану та динаміки існує взаємозв'язок:

$$K_d = K_{\text{пз}} \times K_{\text{вп}} \quad (4.6)$$

Приклад 4.2. Є такі дані по підприємству:

Таблиця 4.2

Випуск продукції, тис. од.

Базисний рік	Звітний рік	
	План	Факт
Y_0	$Y_{\text{пл}}$	Y_n
280,0	331,8	309,6

Визначити відносні величини планового завдання, виконання плану та динаміки. Зробити висновки.

Розв'язання.

1. Відносна величина планового завдання:

$$K_{\text{пз}} = \frac{U_{\text{пл}}}{U_0} = \frac{331,8}{280,0} = 1,185.$$

Висновок: планом на наступний (звітний) рік передбачалось збільшити випуск продукції на 18,5 %.

2. Відносна величина виконання плану:

$$K_{\text{вп}} = \frac{Y_n}{U_{\text{пл}}} = \frac{309,6}{331,8} = 0,933$$

Висновок: фактично план випуску продукції у звітному році було не довиконано на 6,7 %, оскільки мало б дорівнювати 100 %, а виготовили лише на 93,3 %.

3. Відносна величина динаміки:

$$K_d = \frac{Y_n}{U_0} = \frac{309,6}{280,0} = 1,106$$

Висновок: у звітному році порівняно з попереднім випуск продукції збільшився на 10,6 %.

Відносні величини структури (ВВС) характеризують питому вагу окремих частин досліджуваної сукупності в загальному її обсязі. Їх обчислюють шляхом відношення частини до цілого. Розраховуються у формі коефіцієнта чи відсотка.

$$K_{\text{стр}} = \frac{\text{Частина сукупності}}{\text{Сумарна сукупність}} = \frac{y_i}{\sum y_i} \cdot 100\% \quad (4.7)$$

Відносні величини координації (ВВК) характеризують співвідношення частин цілого між собою. За допомогою відносних величин координації визначають, скільки одиниць даної частини цілого припадає на 1, на 100, на 1000, на 10000 одиниць іншої частини, взятої за базу порівняння.

$$K_{\text{коорд}} = \frac{\text{Одна частина сукупності}}{\text{Друга частина сукупності}} = \frac{y_i}{y_k} \quad (4.8)$$

Відносні величини інтенсивності (ВВІ) характеризують ступінь поширення або розвитку явища в певному середовищі. Їх отримують шляхом зіставлення різнойменних абсолютних величин, пов'язаних між собою, які не є складовими цілого.

$$K_{\text{інт}} = \frac{\text{Обсяг певного явища}}{\text{Обсяг середовища, якому це явище властиве}} \quad (4.9)$$

У порівняльному аналізі використовуються кратні співвідношення не лише абсолютних величин. Комплексна й всебічна характеристика закономірностей суспільного життя передбачає порівняння середніх і відносних величин.

Відносна величина порівняння характеризує співвідношення однойменних величин, що стосуються одного й того ж періоду або моменту часу, різних об'єктів чи територій. Показує, у скільки разів порівнювана величина перевищує базисну.

Приклад 4.3. На початок року чисельність фахівців з вищою освітою, зайнятих на підприємстві, становила 53 особи, а чисельність фахівців із середньоспеціальною освітою – 106 осіб. Приймавши за базу порівняння чисельність фахівців з вищою освітою, розрахуємо відносну величину координації: 106:53 або 2,0:1,0. Тобто, на двох фахівців з середньою спеціальною освітою припадає один фахівець із вищою освітою.

Можна використовувати відносні величини порівняння для зіставлення рівня цін на один і той же товар, що реалізується через різні торгові точки.

Приклад 4.4. За даними The Big Mac Index журналу The Economist за 2018 рік ціна одного «Біг Мака» становила в Україні 54 грн (при курсі 27,8 грн/дол. США). У США «Біг Мак» можна купити за 5,58 долари. Тоді реальний курс за купівельною спроможністю становить $54 : 5,58 = 9,68$ грн за долар.

4.4. Задачі для самостійної роботи

Задача 4.1. Прибуток акціонерного банку за перший рік зріс на 25%, за другий – на 15%, за третій зменшився на 8%. На скільки процентів зріс прибуток акціонерного банку за три роки?

Задача 4.2. Є дані про віковий склад населення регіону, тис. осіб:

Таблиця 4.3

Вікова група, років	2010	2018
До 15	193,4	211,6
16–60	469,9	586,8
60 і старше	77,7	163,6

Обчислити усі можливі види відносних величин.

Задача 4.3. Обсяг товарообороту за планом повинен скласти 24 тис. грн. План перевиконано на 25%. У порівнянні з минулим роком товарооборот за звітний період збільшився у 1,6 раза. Обчислити відносні величин планового завдання, динаміки та виконання плану.

Задача 4.4. У порівнянні із минулим роком собівартість одиниці продукції передбачалось знизити на 16 %. Фактично вона знизилась на 18 %. Визначити на скільки перевиконано план зниження собівартості продукції.

Задача 4.5. Прибуток туристичної фірми за перший рік роботи зріс на 15%, за другий – на 8,5%, а за третій зменшився на 4,4%. На скільки процентів змінився прибуток туристичної фірми за три роки?

Задача 4.6. З початку року інфляція щомісячно зростала на 2 %. Яким був рівень інфляції на 1 жовтня?

Задача 4.7. За звітний період було видобуто: вугілля – 720 млн т; нафти – 700 млн т; газу – 650 млн м³. Обчислити видобуток палива в умовних одиницях. Коефіцієнти перерахунку: вугілля – 0,8, нафти – 1,3, газу – 1,2.

Задача 4.8. У звітному році прибуток підприємства досяг 5 млн грн проти 4,4 млн грн у минулому році. Планом же передбачалось у порівнянні з минулим роком обсяг прибутку збільшити на 5%. Обчислити рівень виконання плану (двома способами).

Задача 4.9. Супермаркет у квітні одержав прибуток на 14,8% більше, ніж у січні, у березні на 10,2% більше ніж у січні і в лютому – на 3,7 % менше ніж у січні. На скільки % змінився прибуток супермаркету у квітні порівняно із лютим? Із січнем?

Задача 4.10. Маємо такі дані з виробництва молока двома заводами:

Таблиця 4.4

Завод	План, виробництва, літрів	Фактично надійшло	
		літрів	% жирності
Полісся	1460	1280	3,6
Лісостеп	2540	2240	3,4

Обчислити рівень виконання плану продажу молока разом, якщо враховувати, що планова стандартна жирність молока становить 3,2%.

Задача 4.11. Із звітів трьох підприємств відомо, що фактичний обсяг виробленої продукції у звітному періоді становив 840, 960, 720 тис. грн. План по випуску продукції перше підприємство виконало на 105%, друге – на 100%, третє – на 94%.

Обчислити середній рівень виконання плану по трьох підприємствах разом.

Задача 4.12. Ціни на продовольчі товари зросли у березні на 2% порівняно з січнем, а порівняно із лютим – не змінились; у квітні зросли на 1,5% порівняно з березнем; у червні – на 0,5% порівняно з травнем. Обчислити, на скільки процентів зросли ціни за півроку, якщо в травні порівняно із квітнем ціни не змінились?

Задача 4.13. На 01.01.2013 в Україні проживало 46,0 млн осіб населення, із них у містах – 31,5 млн осіб. Територія країни 603600 км². Обчислити усі можливі відносні величини.

Задача 4.14. Консервний завод за звітний період виготовив 300 тис. консервних банок місткістю 440 см³, 220 тис. консервних банок місткістю 250 см³, 100 тис. консервних банок місткістю 500 см³. Обчислити випуск продукції в

СТАТИСТИКА

умовних банках, якщо умовно-натуральною одиницею виміру є банка місткістю 300 см³.

Задача 4.15. За даними реалізації товарів підприємством обчислити:

- 1) рівень виконання плану реалізації по кожному товару та разом;
- 2) динаміку продажу за кожний місяць;
- 3) відносну величину структури за планом і фактично.

Таблиця 4.5

Товари	План	Реалізовано наростаючим підсумком (тис. грн)		
		червень	липень	серпень
Чоловіче взуття	880	260	590	910
Жіноче взуття	650	190	325	632

Задача 4.16. Чисельність студентів факультету характеризується такими даними:

Таблиця 4.6

Стать	Форма навчання			Всього
	Денна	Заочна	Вечірня	
Чоловіки	520	820	300	1640
Жінки	1610	2200	850	4660
Всього	2130	3020	1150	6300

Обчислити усі можливі види відносних величин.

Задача 4.17. За наведеними даними по продовольчому магазину обчислити усі можливі види відносних величин.

Розділ 4. АБСОЛЮТНІ ТА ВІДНОСНІ СТАТИСТИЧНІ ВЕЛИЧИНИ

Таблиця 4.7

Товарні групи	Базисний рік, млн грн	Звітний рік, млн грн	
		За планом	Фактично
М'ясо	18	20	22
Рибні вироби	9	10	9,5
Бакалійні вироби	16	17	14,8

Задача 4.18. Ціна соняшникової олії на внутрішньому і світовому ринках становила, (дол. США за 1 т):

Таблиця 4.8

Місяць	Внутрішній ринок	Світовий ринок
Квітень	724	695
Серпень	759	690

Обчислити усі можливі види відносних величин. Зробити висновки.

Задача 4.19. На торгах фондової біржі зафіксовані ціни акцій банків (грн. за од.):

Таблиця 4.9

Банк	Номінальна ціна	Ринкова ціна	
		на 15.09.	на 18.09.
А	12	14,5	14,8
Б	10	10,4	10,1
В	9	9,3	8,7

Обчислити усі можливі види відносних величин. Зробити висновки.

Задача 4.20. Натуральний баланс цукру в регіоні (млн т) характеризується такими даними: запаси на початок року – 2,2; вироблено – 23,4; спожито – 16,6; у т. ч. виробниче споживання – 6,4, споживання населенням – 10,2; вивіз за межі регіону – 7,0. Визначити:

- 1) запаси цукру на кінець року;
- 2) співвідношення спожитого та вивезеного цукру;
- 3) структуру споживання.

4.5. Питання для самоконтролю

1. Що називається абсолютними та відносними величинами?
2. Які є види і способи одержання абсолютних і відносних величин?
3. Які Ви знаєте одиниці виміру абсолютних величин?
4. Що показує відносна величина динаміки?
5. Які є види та умови застосування відносної величини динаміки?
6. Що показує відносна величина планового завдання та виконання плану?
7. Що показує відносна величина структури і в яких випадках вона використовується?
8. Що визначає відносна величина координації та інтенсивності?
9. Який взаємозв'язок відносних величин динаміки, планового завдання та виконання плану?
10. Який принцип обчислення базисних і ланцюгових відносних величин динаміки?

Розділ 5

СЕРЕДНІ ВЕЛИЧИНИ ТА ЗАГАЛЬНІ ПРИНЦИПИ ЇХ ЗАСТОСУВАННЯ

План

- 5.1. Сутність і значення середніх величин.
- 5.2. Середня арифметична та її властивості.
- 5.3. Середня гармонічна і квадратична.
- 5.4. Середня геометрична і хронологічна.
- 5.5. Поняття моди, медіани та їх застосування.
- 5.6. Задачі для самостійної роботи.
- 5.7. Питання для самоконтролю.

5.1. Сутність і значення середніх величин

Означення. *Середня величина* – це узагальнююча міра варіаційної ознаки, що характеризує її рівень в розрахунку на одиницю сукупності.

Завдання середньої – одним показником охарактеризувати всю сукупність за досліджуваною ознакою. Основна властивість середньої величини полягає в тому, що вона віддзеркалює загальну, типову, притаманну кожній даній сукупності закономірність розвитку, яка складається під впливом основних факторів.

У статистичній практиці використовують кілька видів середніх: *середню арифметичну, середню гармонічну, середню геометричну, середню квадратичну* і ін. Кожна із зазначених середніх може набирати двох форм: **простої і зваженої**. *Проста* застосовується в разі обчислення середньої за первинними (не згрупованими) даними, а *зважена* – за вторинними (згрупованими) даними.

Використання того чи іншого виду середніх залежить від двох обставин. По-перше, від *характеру індивідуальних значень ознаки* (прямі, обернені, квадратичні, відносні).

СТАТИСТИКА

По-друге, від характеру зв'язку між індивідуальними значеннями ознаки та її загальним обсягом (сума, добуток, степінь, квадратичний корінь). Цей зв'язок є визначальною властивістю сукупності і відбивається в логічній формулі осереднювальної ознаки. На підставі логічної формули обирається вид середньої.

Правильний вибір виду середньої величини іноді досить складний.

Середні, що застосовуються в статистиці, відносяться до класу степеневих середніх, формула яких має вигляд:

$$\bar{x} = \sqrt[m]{\frac{\sum x^m}{n}}, \quad (5.1)$$

- де \bar{x} – степенева середня;
 x – рівень ознаки – варіанта;
 n – число одиниць сукупності;
 m – показник ступеня середньої.

- Зміна значення степені (m) середньої визначає її вид:
 при $m = 1$, середня арифметична;
 $m = 0$, середня геометрична;
 $m = -1$, середня гармонічна;
 $m = -2$, середня квадратична;

Таблиця 5.1

Степінь	Назва середньої	Формула розрахунку	
		Проста	Зважена
1	Середня арифметична	$\frac{\sum x}{n}$	$\frac{\sum xf}{\sum f}$
-1	Середня гармонічна	$\frac{n}{\sum \frac{1}{x}}$	$\frac{\sum w}{\sum \frac{w}{x}}$
0	Середня геометрична	$\sqrt[n]{\prod (x)}$	$\frac{\sum (Ln x) f}{\sum f}$

**Розділ 5. СЕРЕДНІ ВЕЛИЧИНИ ТА ЗАГАЛЬНІ ПРИНЦИПИ
ЇХ ЗАСТОСУВАННЯ**

Закінчення табл. 5.1

Степінь	Назва середньої	Формула розрахунку	
		Проста	Зважена
2	Середня квадратична	$\sqrt{\frac{\sum x^2}{n}}$	$\sqrt{\frac{\sum x^2 f}{\sum f}}$

Найбільш поширеною з середніх величин є середня арифметична.

5.2. Середня арифметична та її властивості

Середня арифметична застосовується у тих випадках, коли обсяг кількісної ознаки для всієї сукупності є сумою індивідуальних значень її окремих елементів. Розрізняють середню арифметичну просту і зважену.

Середню арифметичну просту визначають діленням суми індивідуальних значень на їх кількість, використовуючи при цьому формулу:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x}{n}, \quad (5.2)$$

де \bar{x} – середня;

x – індивідуальні значення ознаки;

n – кількість індивідуальних значень ознаки.

Приклад 5.1. Припустимо, що п'яти працівникам магазину нараховано заробітну плату 8400, 8440, 8480, 8580 і 8600 грн. Визначити середню заробітну плату працівників.

$$\begin{aligned} \bar{x} &= \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \\ &= \frac{8400 + 8440 + 8480 + 8580 + 8600}{5} = \frac{42500}{5} = 8500 \text{ грн} \end{aligned}$$

Це і буде формула середньої арифметичної простої, в якій в чисельнику записується сума варіант ($\sum x$), а в знаменнику – число варіант (n), де \bar{x} – середнє значення ознаки, x – варіанта (конкретне значення ознаки), n – число варіант (ознак), \sum – знак додавання (підсумовування).

Середня арифметична зважена: застосовується у випадках, коли окремі значення ознаки повторюються, зустрічаються по декілька разів або, іншими словами, інформація згрупована і являє собою варіаційний ряд розподілу – дискретний чи інтервальний.

Середня арифметична зважена визначається за формулою:

$$\bar{x} = \frac{\sum xf}{\sum f}, \quad (5.3)$$

де \bar{x} – середнє значення ознаки;

x – варіанта;

f – частота.

Приклад 5.2. Десяти працівникам нараховано таку заробітну плату (за даними табл. 5.2):

Таблиця 5.2

Розподіл працівників за величиною заробітної плати

№ гр.	Зарплата (x)	Кількість працівників (f)	Фонд зарплати (xf)
n	грн	осіб	грн
1	8400	2	16800=8400·2
2	8440	3	25320=8440·3
3	8480	2	16960=8480·2
4	8580	1	8580=8580·1
5	8600	2	17200=8600·2
Σ		10	84860

**Розділ 5. СЕРЕДНІ ВЕЛИЧИНИ ТА ЗАГАЛЬНІ ПРИНЦИПИ
ЇХ ЗАСТОСУВАННЯ**

Для визначення середньої заробітної плати потрібно фонд заробітної плати усіх працівників (24860 грн) розділити на чисельність працівників (10), тобто зробити обчислення:

$$\begin{aligned}\bar{x} &= \frac{\sum xf}{\sum f} = \frac{8400 \cdot 2 + 8440 \cdot 3 + 8480 \cdot 2 + 8580 \cdot 1 + 8600 \cdot 2}{2 + 3 + 2 + 1 + 2} = \\ &= \frac{84860}{10} = 8486 \text{ грн}\end{aligned}$$

В інтервальних рядах розподілу середня величина також визначається за зваженою середньою арифметичною. В інтервальних рядах розподілу ознака подана у вигляді інтервалу. У такому разі за значення ознаки в групах беруть середини інтервалів, тобто умовно перетворюють інтервальний ряд у дискретний.

У таких випадках потрібно для кожного інтервалу знайти його середину за простою середньою між верхньою і нижньою межами кожного інтервалу і після цього робити обчислення за формулою середньої арифметичної зваженої.

Середина інтервалу в групі дорівнює пів сумі його верхньої і нижньої межі. Для відкритих інтервалів такий же порядок визначення середини з умовою, що ширина відкритого інтервалу дорівнює її ширині у попередньому або наступному закритому інтервалі.

Приклад 5.3. За наведеними в таблиці даними обчислити середній товарооборот магазинів.

Таблиця 5.3

Розподіл магазинів за обсягом товарообороту

№ пор.	Обсяг товарообороту, млн грн	Число магазинів	Середина інтервалу	Обсяг товарообороту, млн грн
n	X	f	x	xf
1	До 8	10	$(6+8) : 2=7$	70
2	8–10	6	$(8+10) : 2=9$	54

СТАТИСТИКА

Закінчення табл. 5.3

№ пор.	Обсяг товарообороту, млн грн	Число магазинів	Середина інтервалу	Обсяг товарообороту, млн грн
3	10–12	4	$(10+12) : 2=11$	44
4	12–14	15	$(12+14) : 2=13$	195
5	14 і більше	5	$(14+16) : 2=15$	75
Σ	Усього	40	-	438

Насамперед закриємо відкриті інтервали у першій і п'ятій групах магазинів.

У першій групі відсутнє мінімальне значення ознаки. Величина інтервалу в наступній групі дорівнює 2 млн грн (10–8). Умовно приймається, що і в першій групі величина інтервалу дорівнює 2 млн грн. Тобто мінімальне значення товарообороту в першій групі буде дорівнювати (8–2) 6 млн грн. Отже, перша група включає магазини із товарооборотом від 6 до 8 млн грн.

Аналогічно з останнім інтервалом. Закривши відкриті інтервали, слід знайти середину інтервалу за середньою арифметичною простою.

Скориставшись формулою середньої арифметичної зваженої, обчислимо середній обсяг товарообороту на один магазин.

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{438}{40} = 10,95 \text{ млн грн.}$$

Звичайно, обчислений середній товарооборот – це досить умовна величина, приблизна.

При обчисленні середньої арифметичної замість частот (f), можна брати частоти (F – в % до підсумку, або D – коли

сума частот приймається за одиницю). Результат від цього не зміниться.

Властивості середньої арифметичної

Процедуру розрахунку середніх можна спростити і полегшити, якщо скористатися їх математичними властивостями. Знання цих властивостей середньої розширює розуміння її як статистичної категорії і має практичне значення для скорочення трудомісткості розрахунків.

Властивості арифметичної середньої величини:

1. Сума відхилень індивідуальних значень ознаки від її середнього значення дорівнює нулю.

2. Якщо кожен знак помножити або поділити на постійне число A , то і середня відповідно збільшиться або зменшиться у (стільки ж) A разів.

3. Збільшення або зменшення індивідуальних ознак на постійне число A , відповідно, збільшує або зменшує середню на цю ж величину.

4. Збільшення або зменшення частот (ваг) відповідної ознаки в A разів середню не змінює.

5. Сума квадратів відхилень індивідуальних значень ознаки від її середнього значення менша, ніж сума квадратів їх відхилень від будь-якої іншої довільної величини C .

6. Середня сталої величини дорівнює цій сталій.

Середню арифметичну можна обчислити методом моментів. Цей спосіб доцільно використовувати для обчислення середнього значення ознаки в інтервальному ряді розподілу, де величина інтервалів однакова.

Формула для обчислення середньої способом моментів із інтервального ряду розподілу.

$$\bar{x} = a + i \frac{\sum \left(\frac{x-a}{i} \right) f}{\sum f} \quad \text{або} \quad \bar{x} = a + iM_1, \quad (5.4)$$

$$\text{де } M_1 = \frac{\sum \left(\frac{x-a}{i} \right) f}{\sum f}$$

M_1 – момент першого порядку;
 x – варіанта;
 f – частота.

a – довільно взята величина (як правило мода або медіана);
 i – крок (ширина) інтервалу.

Приклад 5.4. За наведеними даними про розподіл магазинів за обсягом товарообороту обчислити середній товарооборот.

Таблиця 5.4

Робоча таблиця для обчислення середньої арифметичної способом моментів

Обсяг товарообороту, тис. грн	Кількість магазинів f	Середина інтервалу x	xf	$x - a$	$\frac{x-a}{i}$	$\left(\frac{x-a}{i}\right)f$
				$a=1878$	$i=18$	
1833–1851	3	1842	5526	-36	-2	-6
1851–1869	9	1860	16740	-18	-1	-9
1869–1887	21	1878	39438	0	0	0
1887–1905	15	1896	28440	18	1	15
1905–1923	6	1914	11484	36	2	12
Усього	54	x	101628	x	x	12

За довільну величину (a) приймається ознака, яка зустрічається найчастіше. У нашому прикладі $a = 1878$, оскільки повторюється найбільш часто ($f = 21$); $i = 18$. Обчислення зручніше виконати у вищенаведеній таблиці. Результати обчислення підставляємо у формулу:

$$\bar{x} = a + i \frac{\sum \left(\frac{x-a}{i} \right) f}{\sum f}; \quad \bar{x} = 1878 + 18 \frac{12}{54} = 1882 \text{ тис. грн}$$

Розв'язання задачі звичайним способом дає ті ж результати:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{101628}{54} = 1882 \text{ тис. грн.}$$

5.3. Середня гармонічна і квадратична

Середні арифметичні, як це було показано раніше, застосовуються у випадку, коли відомі ознаки і їх частоти. В ситуації, коли відомі добутки ознаки на частоту (xf) і ознака (x), а частота (f) – невідома, визначення середнього рівня ознаки у такому випадку здійснюється за середньою гармонійною зваженою. Введемо позначення: $xf = w$ – добуток ознаки на частоту, звідки:

$$f = \frac{w}{x}$$

Тепер у формулі середньої зваженої зробимо відповідні перетворення і вона стане іншою:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum w}{\sum \frac{w}{x}} \quad (5.5)$$

Нова формула розрахунку середньої називається *гармонійною середньою зваженою*.

Середня гармонійна зважена є перетвореною формою середньої арифметичної зваженої і тотожна їй. Гармонійна, як

видно з формули, є середньою арифметичною зваженою із обернених значень ознаки.

Зауваження до застосування гармонійної:

– у випадку, коли у вихідній інформації показник w для всіх значень ознаки однаковий, формула середньої гармонійної зваженої перетворюється в нову конструкцію гармонійної – *середню гармонійну просту*:

$$\bar{x} = \frac{\sum w}{\sum \frac{w}{x}} = \frac{w + w + w + \dots + w}{\frac{w}{x_1} + \frac{w}{x_2} + \dots + \frac{w}{x_n}} = \frac{1 + 1 + 1 + \dots + 1}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x}} \quad (5.6)$$

Приклад 5.5. Протягом години два магазини продавали картоплю по 8,0 та 12,0 грн за кілограм. Одержали однакову виручку по 4800 грн. Визначити середню ціну проданої картоплі.

Здавалося б, можна скористатися формулою середньої арифметичної простої $\bar{x} = \frac{\sum x}{n} = \frac{8+12}{2} = 10$ грн. Проте обчислена таким чином середня не буде відповідати дійсності, бо, наприклад, за 1000 кг картоплі, проданих двома магазинами, одержано 9600 грн (4800+4800), а не 10000 (1000 · 10) грн.

Середню ціну в такому разі визначають шляхом ділення спільної виручки двох магазинів (9600 грн) на загальну кількість проданої картоплі, обчислену діленням виручки кожного магазину на ціну.

Розв'язок має такий вигляд:

$$\bar{x} = \frac{\sum w}{\sum \frac{w}{x}} \quad \text{або} \quad \bar{x} = \frac{4800 + 4800}{\frac{4800}{8} + \frac{4800}{12}} = \frac{9600}{1000} = 9,6 \text{ грн.}$$

Це і є формула середньої гармонійної зваженої.

Оскільки виручка від реалізації картоплі в обох магазинах однакова ($w_1 = w_2$), цю величину можна винести за дужки у чисельнику і знаменнику і скоротити:

$$\bar{x} = \frac{w_1 + w_2}{\frac{w_1}{x_1} + \frac{w_2}{x_2}} = \frac{w(1+1)}{w\left(\frac{1}{x_1} + \frac{1}{x_2}\right)} = \frac{n}{\sum \frac{1}{x}} \quad (5.7)$$

Це і буде формула середньої гармонійної простої. Розв'язок буде мати вигляд:

$$\bar{x} = \frac{n}{\sum \frac{1}{x}} = \frac{2}{\frac{1}{8} + \frac{1}{12}} = 9,6 \text{ грн.}$$

Залежно від того, які вихідні дані є в умові задачі, використовується відповідний вид середньої:

□ Якщо є варіанти (x) і частоти (f) – використовується середня арифметична зважена.

□ Якщо у кожній групі частоти (f) однакові – можна використовувати середню арифметичну просту.

□ Якщо є варіанти (x) і добуток варіанти на частоту ($xf = w$) – використовується середня гармонійна зважена.

□ Якщо у кожній групі добуток варіанти на частоту ($xf = w$) однакові, можна використовувати середню гармонійну просту.

Середня квадратична використовується:

□ При обчисленні середніх сторін квадратів.

□ При обчисленні середніх діаметрів циліндричних тіл (труби, колоди, цистерни, банки, люки, колодязі тощо).

□ В електротехніці при визначенні середніх перерізів провідників електроенергії.

□ У медицині при визначенні середніх діаметрів пухлин, синців, гематом.

□ При обчисленні деяких показників варіації у статистиці (дисперсія, середнє квадратичне відхилення, коефіцієнт варіації).

Середня квадратична, як і всі степеневі середні, буває простою і зваженою:

Проста

$$\bar{x} = \sqrt{\frac{\sum x^2}{n}} \quad (5.8)$$

Зважена

$$\bar{x} = \sqrt{\frac{\sum x^2 f}{\sum f}} \quad (5.9)$$

Приклад 5.6. Потрібно визначити середню сторону квадратів, яка залишила б незмінною їх загальну площу. Припустімо, що сторона першого квадрата дорівнює 20 см, а другого – 30 см (рис. 5.1).

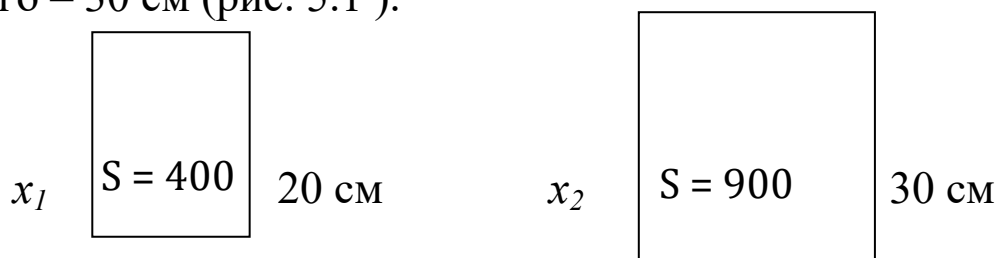


Рис. 5.1. Площі й сторони квадратів

Визначити середню сторону квадрата. Здавалося б, що для розв'язання задачі найбільш придатна формула середньої арифметичної простої

$$\bar{x} = \frac{\sum x}{n} = \frac{20 + 30}{2} = 25 \text{ см.}$$

Але ця середня (25 см) дає площу двох рівновеликих квадратів 1250 квадратних сантиметрів $[(25 \times 25)2 = 1250]$. А реальна площа двох рівновеликих квадратів $(20^2 + 30^2)$ становить 1300 квадратних сантиметрів.

Отже, 25 см – це середня сторона рівновеликих квадратів. Істинна середня сторона рівновеликих квадратів буде

дорівнювати кореню квадратному із половини загальної площі двох рівновеликих квадратів.

$$\bar{x} = \sqrt{\frac{20^2 + 30^2}{2}} = \sqrt{\frac{400 + 900}{2}} = \sqrt{\frac{1300}{2}} = \sqrt{650} \approx 25,5 \text{ см.}$$

На цьому і побудована логіка середньої квадратичної.

5.4. Середня геометрична і хронологічна

Середня геометрична у статистиці (економіці) використовується лише для обчислення середніх коефіцієнтів динаміки, темпів росту та приросту.

З'ясуємо поняття середньої геометричної на такому прикладі. Прибуток роздрібного підприємства зріс із 20 до 24 тис. грн у лютому і до 30 тис. грн у березні, або на 50%.

$$K_{\partial} = \frac{Y_n}{Y_0} = \frac{30}{20} = 1,5 = 150\% = +50\%$$

На скільки процентів в середньому щомісячно зростав прибуток роздрібного підприємства?

Здавалося б що на 25%. Якщо за два місяці прибуток зріс на 50%, то в середньому за місяць ($50 : 2$) – на 25%. Прості розрахунки спростовують це. Якби прибуток щомісячно зростав на 25%, то уже в лютому він досяг би (20 тис. грн + 25%) 25 тис. грн, а в березні (25 тис. грн + 25%) 31,25 тис. грн, а не 30 тис. грн, як задано в умові задачі. Отже, в середньому прибуток зростав не на 25%, а на дещо меншу величину.

Зробимо обчислення за середньою геометричною.

Середня геометрична із n варіант дорівнює кореню n -ної степені із добутку ланцюгових коефіцієнтів динаміки.

$$\bar{K}_\partial = \sqrt[n]{DK_{\partial l}}, \quad (5.10)$$

де \bar{K}_∂ – середній коефіцієнт динаміки,
 $K_{\partial l}$ – ланцюгові коефіцієнти динаміки,
 n – число ланцюгових коефіцієнтів.
 Підставимо дані задачі у формулу:

$$\bar{K}_\partial = \sqrt[n]{DK_{\partial l}} = \sqrt[2]{1,2 \cdot 1,25} = \sqrt{1,5} \approx 1,225$$

Таким чином, прибуток роздрібного підприємства в середньому щомісячно зростає у 1,225 разів, або на 22,5%.

Перевіримо: $Y_2 = Y_0 \cdot \bar{K}_\partial \cdot \bar{K}_\partial = 20 \cdot 1,225 \cdot 1,225 \approx 30$

З попередньої теми відомо, що добуток ланцюгових коефіцієнтів динаміки ($K_{\partial l}$) дорівнює останньому базисному.

Звідси формула може мати такий вигляд:

$$\bar{K}_\partial = \sqrt[m-1]{\frac{Y_n}{Y_0}}, \quad \bar{K}_\partial = \sqrt[3-1]{\frac{30}{20}} = \sqrt{1,5} \approx 1,225$$

де Y_n – рівень звітнього періоду;

Y_0 – рівень базисного періоду;

m – число рівнів ряду динаміки, або календарних дат у періоді.

За формулою середньої арифметичної простої обчислюються також середні у хронологічному ряду, якщо інтервали часу, за який подаються значення ознак, рівні.

Середня хронологічна використовується лише для обчислення середнього рівня моментного ряду динаміки.

Приклад. 5.7. Припустімо, що нам відома чисельність працівників на початок кожного місяця третього кварталу року (табл. 5.5.)

Таблиця 5.5

**Динаміка чисельності продавців овочевої продукції
міста у третьому кварталі року становила (на 1 число
кожного місяця):**

Дата	1.07	1.08	1.09	1.10
Число продавців	200	240	380	480
Символ	y_1	y_2	y_3	y_4

Визначимо середню чисельність продавців овочів у третьому кварталі за формулою середньої хронологічної для моментного ряду динаміки з рівними проміжками часу між датами.

$$\begin{aligned}\bar{y} &= \frac{\frac{1}{2}y_1 + y_2 + y_3 + \dots + \frac{1}{2}y_n}{n-1} = \\ &= \frac{\frac{1}{2} \cdot 200 + 240 + 380 + \frac{1}{2} \cdot 480}{4-1} = 320 \text{ осіб,}\end{aligned}$$

де \bar{y} – середній рівень ряду динаміки;

y – конкретні рівні ряду динаміки;

n – число рівнів ряду динаміки, або число календарних дат.

Якщо інтервали (проміжки часу) між датами моментного ряду динаміки неоднакові, потрібно використати формулу середньої арифметичної зваженої, де як ваги (частоти) беруться протяжність відрізків часу між датами, а як варіанти – парні середні суміжних значень рівнів ряду динаміки.

Більш детально застосування формул для визначення середнього значення в динамічних рядах буде розглянуто в темі «ряди динаміки».

5.5. Поняття моди, медіани та їх застосування

В якості статистичних характеристик варіаційних рядів розподілу поряд з розглянутими вище середніми величинами розраховуються особливі структурні (порядкові) середні – мода і медіана, які кількісно описують структуру ряду.

Мода (M_0 – значення ознаки, яка найчастіше зустрічається в сукупності, яка досліджується. У дискретному ряду мода не вираховується, а визначається безпосередньо – ознака з найбільшою частотою є мода. На практиці частіше зустрічаються ряди з одним модальним значенням ознаки. Коли значення ознаки в ряду з однаковими частотами зустрічається два або більше разів він вважається, відповідно, бімодальним або мультимодальним. Це свідчить про неоднорідність сукупності.

В інтервальному ряду розподілу моду визначають за формулою:

$$M_0 = x_{m_0} + i_{m_0} \frac{f_{m_0} - f_{m_0-1}}{(f_{m_0} - f_{m_0-1}) + (f_{m_0} - f_{m_0+1})}, \quad (5.11)$$

де M_0 – мода;

x_{m_0} – нижня межа модального інтервалу;

i_{m_0} – ширина модального інтервалу;

$f_{m_0}, f_{m_0-1}, f_{m_0+1}$ – відповідно частоти модального інтервалу, передмодального і післямодального інтервалу.

Модальний інтервал – це межі ознаки з найбільшою частотою. Мода для інтервального ряду за вищенаведеною формулою – це умовне значення ознаки, в межах інтервалу в якій зустрічається найбільша частота.

Приклад 5.8. Визначимо моду за даними ряду розподілу (табл. 5.6).

Таблиця 5.6

**Розподіл продавців супермаркету
за рівнем заробітної плати**

Групи працівників за рівнем зарплати, (у.о.), X	Кількість працівників, f
1	2
До 260	4
260–280	7
280–300	10
300–320	18
320–340	35
340–360	10
360–380	7
380–400	6
>400	3
Усього	100

$$\begin{aligned}
 M_o &= x_{mo} + i_{mo} \frac{f_{mo} - f_{mo-1}}{(f_{mo} - f_{mo-1}) + (f_{mo} - f_{mo+1})} = \\
 &= 320 + 20 \frac{35 - 18}{(35 - 18) + (35 - 10)} = 320 + 20 \frac{17}{42} = 326,5 \text{ у.о.}
 \end{aligned}$$

Медіана – значення ознаки, яка знаходиться в середині ряду розподілу або, іншими словами, ділить ряд на дві рівні частини, в одній з яких ознаки менші медіани, а в другій – більші медіани.

Щоб знайти медіану, спочатку потрібно знайти її місцезнаходження, тобто порядковий номер ознаки, що виступає медіаною. Наприклад, у ряду знаходяться 7 студентів за величиною зросту (см)

169, 178, 178, 180, 182, 183, 183.

Місцезнаходження медіани (N_{me}) визначається за такою формулою:

$$N_{me} = \frac{n+1}{2} = \frac{7+1}{2} = 4.$$

У нашому прикладі це 4-е місце, тобто студент, зріст якого – 180 см і буде медіаною.

У випадку *парного* числа одиниць у ряду медіана буде дорівнювати середній із двох ознак, що знаходяться в середині ряду. Наприклад, якщо у нас 6 студентів, то $N_{me} = (6 + 1)/2 = 3,5$. Медіаною буде умовне середнє значення між зростом 3 і 4-го студента тобто $Me = (178 + 180)/2 = 179$.

В інтервальних рядах розподілу ряд на дві рівні частини ділить варіант (ознака), закладений у межах інтервалу. Щоб знайти її місцезнаходження, потрібно виконати додаткову операцію накопичення частот або перетворити ряд у кумулятивний. Тоді накопичена частота медіанного інтервалу буде рівною або більшою пів суми усіх частот ряду. І саме в ній знаходиться медіана, значення якої вираховується за такою інтерполяційною формулою:

$$Me = x_{me} + i_{me} \frac{\sum f/2 - S_{me-1}}{f_{me}}, \quad (5.12)$$

де Me – медіана;

x_{me} – нижня межа медіанного інтервалу;

$\sum f/2$ – половина накопичених частот;

S_{me-1} – сума накопичених частот до медіанного інтервалу;

i_{me} – частота медіанного інтервалу.

Приклад 5.9. Розрахуємо медіану за даними табл. 5.7.

Таблиця 5.7

Розподіл продавців за рівнем заробітної плати

Групи працівників за рівнем зарплати, x (у.о.)	Кількість працівників, f	Накопичена частота, Σf
До 260	4	4
260–280	7	11
280–300	10	21

**Розділ 5. СЕРЕДНІ ВЕЛИЧИНИ ТА ЗАГАЛЬНІ ПРИНЦИПИ
ЇХ ЗАСТОСУВАННЯ**

Закінчення табл. 5.7

Групи працівників за рівнем зарплати, x (у.о.)	Кількість працівників, f	Накопичена частота, Σf
300–320	18	39
320–340	35	74
340–360	10	84
360–380	7	91
380–400	6	97
>400	3	100
Усього	100	X

Визначимо у якій накопиченій частоті знаходиться медіанний інтервал? Ця частота буде більшою або рівною пів суми накопичених частот ряду. У нашому випадку пів сума частот ряду дорівнює 50 ($100/2$) і знаходиться в накопиченій частоті, що дорівнює 74. Таким чином, медіанним інтервалом у нашому прикладі є група працівників з рівнем зарплати 320–340 у.о. Підставимо числові значення з таблиці в формулу медіани для обраного інтервального ряду розподілу:

$$\begin{aligned} Me &= X_{me} + i_{me} \cdot \frac{\sum f / 2 - S_{me-1}}{f_{me}} = 320 + 20 \cdot \frac{100/2 - 39}{35} = \\ &= 320 + 20 \frac{11}{35} = 326,3 \text{ у.о} \end{aligned}$$

5.6. Задачі для самостійної роботи

Задача 5.1. Є такі дані про собівартість одиниці продукції (у.о.) та загальні грошові витрати на виробництво одного і того ж виду продукції на 4 підприємствах за місяць.

СТАТИСТИКА

Таблиця 5.8

Підприємства	1	2	3	4
Собівартість од. продукції	28	25	23	30
Витрати	100	125	158	110

Обчислити середню собівартість одиниці продукції по 4-х підприємствах, середні витрати на одне підприємство.

Задача 5.2. За наведеними даними обчислити середню місячну заробітну плату по чотирьох магазинах загалом за січень і за серпень.

Таблиця 5.9

№ магазину	Червень		Серпень	
	середня зарплата, у.о.	кількість продавців	середня зарплата, у.о.	фонд заробітної плати, у.о.
1	300	20	310	7130
2	310	18	312	6240
3	330	15	335	5695
4	315	22	320	6080

Задача 5.3. Залишки простроченої заборгованості банку по позичках становили (млн. грн):

на 1.04 – 40; на 1.05 – 39; на 1.06 – 42; на 1.07 – 43.

Обчислити середній залишок заборгованості по всіх позичках за квартал та за кожний місяць.

Задача 5.4. За нижченаведеними даними обчислити:

1) середній процент виконання плану виробництва продукції;

2) середній процент стандартної продукції.

**Розділ 5. СЕРЕДНІ ВЕЛИЧИНИ ТА ЗАГАЛЬНІ ПРИНЦИПИ
ЇХ ЗАСТОСУВАННЯ**

Таблиця 5.10

Підприємство	Вироб. продук. у звітному періоді, млн грн	% виконання плану	% стандартної продукції
1	630	110	85
2	970	98	88
3	750	110	95

Задача 5.5. Визначити середній розмір кредиту і середній термін кредиту банку за такими даними:

Таблиця 5.11

№ кредиту	Розмір кредиту, млн грн	Термін кредиту /днів
1	110	40
2	100	25
3	120	30
4	80	50
5	200	60

Задача 5.6. Маємо таку інформацію по 5 магазинах за рік:

Таблиця 5.12

№ магазину	Роздрібний товарооборот млн грн	Кількість робітників
1	260	5
2	860	16
3	350	9
4	520	12
5	450	18

СТАТИСТИКА

Обчислити:

- 1) середній товарооборот на 1 магазин;
- 2) середню чисельність робітників магазину;
- 3) роздрібний товарооборот, що припадає в середньому на одного робітника.

Задача 5.7. Маємо такі дані:

Таблиця 5.13

Однорічний прибуток тис. грн	30–35	35–40	40–45	45–50
Кількість підприємств	45	30	15	10

Обчислити:

1. Середній однорічний прибуток.
2. Моду і медіану.

Задача 5.8. Обчислити середній розмір продуктивності продавця, а також моду, медіану за такими даними:

Таблиця 5.14

Продуктивність праці, тис. грн	Число продавців
До 30	30
30–40	40
40–50	50
50 і більше	60

Задача 5.9. Є такі дані про частку експорту у вартості товарної продукції підприємств, що виробляють добрива:

**Розділ 5. СЕРЕДНІ ВЕЛИЧИНИ ТА ЗАГАЛЬНІ ПРИНЦИПИ
ЇХ ЗАСТОСУВАННЯ**

Таблиця 5.15

Добрива експорту у товарній продукції	Кількість підприємств	Товарна продукція підприємств групи, млн грн
0,15	5	200
0,2	7	460
0,3	4	600
Усього	16	1260

Обчислити середню частку експорту у товарній продукції підприємства.

Задача 5.10. Маємо дані про розподіл робітників магазину за обсягом заробітної плати.

Таблиця 5.16

Місячна заробітна плата, у.о.	Число робітників
220–250	5
250–280	10
280–310	20
310–340	25
340–370	20
370–400	15
400–430	5
Усього	

Обчислити:

- 1) по підприємству способом моментів середню заробітну плату;
- 2) моду і медіану.

СТАТИСТИКА

Задача 5.11. Є дані про кількість робочих місць у магазинах:

Таблиця 5.17

Кількість робочих місць	2	3	4	5	6	7	8
Кількість магазинів	13	18	23	26	11	7	3

Обчислити середню кількість робочих місць, а також моду і медіану.

Задача 5.12. Виробіток 24 продавців універмагу за місяць характеризуються такими даними, (у.о.): 4640, 4830, 5200, 4140, 5430, 6360, 4310, 5330, 5120, 3510, 3970, 4850, 4770, 5300, 4700, 5060, 5770, 4990, 6600, 3875, 5390, 4490, 3790, 4721.

Обчислити:

- 1) середній виробіток продавця;
- 2) середній прогресивний виробіток.

Задача 5.13. За наведеними даними обчислити середній одноденний прибуток тур фірми (тис. грн), моду та медіану.

Таблиця 5.18

Одноденний прибуток	10–20	20–30	30–40
Число турфірм	20	40	10

Задача 5.14. До водонагрівача підведена холодна вода через 3 труби діаметром 20, 25 та 30 міліметрів. Труби вирішено замінити на однакові. Якого діаметру труби потрібно встановити?

Задача 5.15. Три тур фірми продають путівки в одну країну за ціною 130, 135 і 150 у.о. Виручка від реалізації у них однакова. Визначити середню ціну путівки.

Задача 5.16. Щоб встановити середній процент браку товарів «М», було проведено перевірку 100 партій товару, отримано такі результати:

Таблиця 5.19

% браку	0–2	2–4	4–6	6–8	8–10
Кількість партій товару	8	24	42	17	9

Способом моментів обчислити середній процент браку, моду та медіану.

5.7. Питання для самоконтролю

1. Для чого використовуються середні величини в економіці?
2. Які є види середніх величин?
3. В яких випадках застосовується середня арифметична проста і зважена?
4. Які Ви знаєте математичні властивості середньої арифметичної.
5. Коли використовується середня гармонічна проста і зважена?
6. В яких випадках застосовується середня квадратична проста і зважена?
7. В яких випадках застосовується середня хронологічна і геометрична?
8. Що називається модою і медіаною в статистиці?
9. Як обчислюється мода і медіана для дискретного та інтервального рядів розподілу?

Розділ 6

АНАЛІЗ ВАРІАЦІЇ ТА ФОРМИ РОЗПОДІЛУ

План

- 6.1. Поняття варіаційного ряду розподілу. Характеристики центру розподілу.
- 6.2. Характеристики варіації.
- 6.3. Види та взаємозв'язок дисперсій.
- 6.4. Характеристики форми розподілу. Порівняльний аналіз структур.
- 6.5. Задачі для самостійної роботи.
- 6.6. Питання для самоконтролю.

6.1. Поняття варіаційного ряду розподілу. Характеристики центру розподілу

Одним із важливих завдань статистичного аналізу є дослідження характерних рис та особливостей структури статистичної сукупності, дослідження закономірностей розподілу значень ознаки. Інформаційним підґрунтям для аналізу закономірностей розподілу є варіаційний ряд розподілу.

Означення. Варіація – це властивість статистичної сукупності, що відображає здатність до зміни, обумовлену як зовнішніми, так і внутрішніми факторами, як пов'язаними з сутністю досліджуваного об'єкта, так і не пов'язаними з нею.

Варіаційний ряд розподілу – це статистичний ряд, згрупований за кількісною ознакою.

Позначення, що використовуються у варіаційному ряді:

- x_i – значення ознаки (варіанта);
- k – кількість значень ознаки x_i (у дискретному ряді) або число закритих інтервалів (в інтервальному ряді);
- x' – середина інтервалу;
- x_{\min} – нижня межа інтервалу;

**Розділ 6. АНАЛІЗ ВАРІАЦІЇ
ТА ФОРМИ РОЗПОДІЛУ**

- x_{\max} – верхня межа інтервалу;
 f_i – частота поширення ознаки;
 w_i – частка, тобто відносна частота ряду розподілу, показник, що характеризує частку окремих груп в загальній сумі частот.
 $\text{cum}f_i$ – кумулятивна частота (кумулятивний розподіл) – операція послідовного підсумовування частот, починаючи із першої та закінчуючи кожною наступною;
 $\text{cum}w_i$ – кумулятивна частка.

Варіаційний ряд складається із двох елементів: варіант x_i – значення групувальної ознаки та частот f_i – чисельності окремих варіант, показують, як часто повторюються варіанти. У співвідношенні варіант і частот виявляється закономірність розподілу.

Залежно від характеру варіації значень ознаки ряд розподілу може бути дискретним або інтервальним.

Дискретним рядом називається ряд розподілу у якому зазначаються конкретні значення ознаки. Приклад дискретного ряду розподілу наведено у табл. 6.1.

Таблиця 6.1

**Розподіл учнів старшої школи за рівнем навчальних
досягнень**

Рівень навчальних досягнень	1	2	3	4	5	6	7	8	9	10	11	12	Разом
Чисельність учнів	2	4	6	9	11	9	14	16	15	9	7	8	110

Інтервальний ряд – це ряд розподілу, у якому зазначаються інтервали значень ознаки «від – до».

Зауваження. При побудові варіаційного інтервального ряду розподілу існують деякі особливості, пов'язані з типом інтервалу.

Різновиди інтервальних рядів:

– інтервальний ряд з рівними інтервалами (наприклад, 25–30, 30–35, 35–40 ...);

– інтервальний ряд з нерівними інтервалами (20–25, 25–30, 30–40, 40–50 ...);

– інтервальний ряд з відкритими (неповними) межами (до 10, понад 50);

– інтервальний ряд із закритими (повними) межами (від 5 до 10).

Перш ніж виконати обчислення в інтервальному варіаційному ряду, що має відкриті межі, необхідно закрити межі відкритих інтервалів: інтервал закривається з орієнтацією на наступний (для першого інтервалу) або попередній (для останнього інтервалу) інтервал.

Приклад інтервального ряду розподілу наведено у табл. 6.2.

Таблиця 6.2

Розподіл населення за рівнем середньодушових еквівалентних загальних доходів за I квартал 2018 р.

Середньо-душовий загальний дохід у місяць, грн	до 1920	1920–2280	2280–2640	2640–3000	3000–3360	3360–3720	3720–4080	4080–4440	4440–4800	4800–5160	понад 5160
Кількість осіб, тис.	1272	2004	2120	3083	3816	3199	3854	2775	3006	2428	10986

**Розділ 6. АНАЛІЗ ВАРІАЦІЇ
ТА ФОРМИ РОЗПОДІЛУ**

Всебічний аналіз закономірностей розподілу передбачає визначення усіх аспектів і типових рис розподілу статистичної сукупності.

Виокремлюють такі основні групи аналітичних показників варіаційних рядів розподілу:

Показники, що характеризують структуру сукупності – частотний аналіз;

Показники, що визначають типовий рівень ознаки – характеристики центру розподілу;

Показники варіації – визначають ступінь згуртованості індивідуальних значень ознаки навколо центру розподілу, однорідність сукупності, яка є передумовою використання інших статистичних методів аналізу;

Показники форми розподілу – визначають ступінь відхилення від симетрії, пропорційність розподілу певного явища між окремими елементами сукупності;

Показники кореляційного зв'язку – показники, які виявляють наявність взаємозв'язку між явищами та ступінь його інтенсивності.

Кумулятивна частота (частка) – показник, який характеризує обсяг сукупності зі значенням варіант, які не перевищують значення x_i

Таблиця 6.3

Частотні характеристики ряду розподілу

Значення варіант x_i	Частоти f_i	Частки d_i	Кумулятивні	
			Частоти	Частки
x_1	f_1	d_1	f_1	d_1
x_2	f_2	d_2	$f_1 + f_2$	$d_1 + d_2$
x_3	f_3	d_3	$f_1 + f_2 + f_3$	$d_1 + d_2 + d_3$
...
x_n	f_n	d_n	Σf_i	1
Разом	Σf_i	1	x	x

Характеристики центру розподілу – це значення ознаки, які характерні для даного розподілу і є центром тяжіння статистичної сукупності. До них належать: середня величина, мода і медіана.

Середня величина – є найважливішою характеристикою центру розподілу, оскільки узагальнює усю різноманітність індивідуальних значень ознаки і відображає її типовий рівень.

Мода (M_o) – найбільш поширене, домінантне значення ознаки, тобто значення ознаки, яке найчастіше зустрічається у досліджуваній статистичній сукупності.

Медіана (M_e) – це значення ознаки, яке припадає на середину упорядкованого ряду розподілу, поділяє його на дві рівні за обсягом частини.

Формули та особливості розрахунку характеристик центру розподілу залежно від виду ряду розподілу наведені у попередньому розділі.

Для поглибленого аналізу структури сукупності виникає необхідність визначати значення ознаки у будь-якої одиниці, що посідає певне місце в упорядкованому варіаційному ряду.

Квартилі (Q) – визначають значення ознаки в одиниць, що поділяють сукупність на чотири однакові за обсягом частини.

Визначають нижній (Q_1) і верхній (Q_3) квартилі, середній квартиль – це медіана. Розрахунок квартилів у варіаційному ряду розподілу здійснюється аналогічно розрахунку медіани.

$$Q_1 = x_0 + h \frac{0,25 \sum_1^n f_i - cumf_{Q_{1-1}}}{f_{Q_1}} \quad (6.1)$$

$$Q_3 = x_0 + h \frac{0,75 \sum_1^n f_i - cumf_{Q_{3-1}}}{f_{Q_3}}, \quad (6.2)$$

де x_0 – нижня межа відповідного квартильного інтервалу;

h – ширина квартильного інтервалу;

f_{Q_1} – частота (частка) нижнього квартильного інтервалу;

f_{Q_3} – частота (частка) верхнього квартильного інтервалу;

$cumf_{Q_{1-1}}$, $cumf_{Q_{3-1}}$ – кумулятивні частоти (частки) передквартильних інтервалів.

**Розділ 6. АНАЛІЗ ВАРІАЦІЇ
ТА ФОРМИ РОЗПОДІЛУ**

Децилі поділяють сукупність на десять рівних частин. Перший дециль D_1 ділить сукупність у співвідношенні 1/10 до 9/10, другий D_2 – у співвідношенні 2/10 до 8/10 і т. д.

$$D_1 = x_0 + h \frac{0,10 \sum_1^n f_i - cumf_{D_{1-1}}}{f_{D_1}} \quad (6.3)$$

$$D_2 = x_0 + h \frac{0,20 \sum_1^n f_i - cumf_{D_{2-1}}}{f_{D_2}} \quad (6.4)$$

$$D_9 = x_0 + h \frac{0,90 \sum_1^n f_i - cumf_{D_{9-1}}}{f_{D_9}} \quad (6.5)$$

Приклад розрахунку характеристик центру розподілу, кватилів та децилів розподілу на прикладі варіаційного ряду розподілу 30-ти домогосподарств за середньомісячним доходом на одного члена домогосподарства.

Таблиця 6.4

Дані для розрахунку середньої, моди, медіани, кватилів та децилів у інтервальному ряду розподілу

№ п/п	Середньомісячний дохід на одного члена домогосподарства, грн	Кількість домогосподарств	x'_i	$x'_i f_i$	$cumf_i$
1	до 1452	6	1171	7026	6,00
2	1452–2014	5	1733	8665	11,00
3	2014–2576	2	2295	4590	13,00
4	2576–3138	13	2857	37141	26,00
5	3138 і більше	4	3419	13676	30,00
Разом		30	x	71098	x

Розрахунок середньої в інтервальному ряду розподілу передбачає попередній розрахунок серединних значення інтервалів (x') для кожної групи. Особливості розрахунку середини інтервалів та середньої наведені у попередньому розділі. За даними табл. 6.4. середня становить:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{71098}{30} = 2369,93 \text{ грн}$$

Порядок розрахунку моди та медіани в інтервальному ряду розподілу наведені у попередньому розділі. За даними табл. 6.4. мода та медіана становлять:

$$M_o = 2576 + 562 \frac{13-2}{(13-2)+(13-4)} = 2885,10 \text{ грн,}$$

$$M_e = 2576 + 562 \frac{15-13}{13} = 2662,46 \text{ грн.}$$

Розрахунок кватилів: визначаємо перший кватильний інтервал на основі значення кумулятивної частоти $\text{cum}f_i$, яке більше або дорівнює $1/4$ обсягу сукупності $0,25 \sum_1^n f_i = 7,5$. Зазначене значення $\text{cum}f_i$ припадає на другу групу з інтервалом 1452–2014.

Обчислюємо перший кватиль за формулою (6.1):

$$Q_1 = 1452 + 562 \frac{7,5-6}{5} = 1620,60 \text{ грн.}$$

Визначаємо третій кватильний інтервал на основі значення кумулятивної частоти $\text{cum}f_i$, яке більше або дорівнює $3/4$ обсягу сукупності $0,75 \sum_1^n f_i = 22,5$. Зазначене значення $\text{cum}f_i$ припадає на четверту групу з інтервалом 2576–3138.

Обчислюємо третій кватиль за формулою (6.2):

$$Q_3 = 2576 + 562 \frac{22,5-13}{13} = 2986,69 \text{ грн.}$$

Розрахунок децилів: визначаємо перший децильний інтервал на основі значення кумулятивної частоти $\text{cum}f_i$, яке більше або дорівнює $1/10$ обсягу сукупності $0,10 \sum_1^n f_i = 3$. Зазначене значення $\text{cum}f_i$ припадає на першу групу з інтервалом до 1452, тобто якщо закрити інтервал 890–1452.

Обчислюємо перший дециль за формулою (6.3):

$$D_1 = 890 + 562 \frac{3}{6} = 1171,00 \text{ грн.}$$

Визначаємо дев'ятий децильний інтервал на основі значення кумулятивної частоти $\text{cum}f_i$, яке більше або дорівнює $9/10$ обсягу сукупності $0,90 \sum_1^n f_i = 27$. Зазначене значення $\text{cum}f_i$ припадає на п'яту групу з інтервалом 3138 і більше, тобто якщо закрити інтервал 3138–3700.

Обчислюємо дев'ятий дециль за формулою (6.5):

$$D_9 = 3138 + 562 \frac{27 - 26}{13} = 3181,23 \text{ грн.}$$

Статистичний висновок. За результатами дослідження доходів 30-ти домогосподарств встановлено, що середній дохід на одного члена домогосподарства становить 2369,93 грн, у більшості домогосподарств середньодушовий дохід становить 2885,10 грн, у половини домогосподарств середньодушовий дохід не перевищує 2662,46 грн, у 25% домогосподарств середньодушовий дохід не перевищує 1620,60 грн, у 75% домогосподарств середньодушовий дохід не перевищує 2986,69 грн, 10% домогосподарств мають середньодушовий дохід не вищий ніж 1171,00 грн, а 90% домогосподарств мають дохід не вищий 3181,23 грн.

6.2. Характеристики варіації

Оцінювання ступеня коливання ознаки у статистичній сукупності, її варіації є невід'ємною складовою статистичного аналізу закономірностей розподілу. Оскільки, чим менші відхилення індивідуальних значень ознаки від центру розподілу, тим одноріднішою є досліджувана сукупність і більш надійними і типовими є характеристики центру розподілу, насамперед середня величина. Типовість середньої

величини є необхідною умовою використання інших методів статистичного аналізу.

З метою вимірювання і оцінювання варіації обчислюють абсолютні та відносні характеристики варіації.

До абсолютних належать: варіаційний розмах, середнє лінійне відхилення, стандартне (середнє квадратичне) відхилення, дисперсія.

До відносних належать коефіцієнти варіації, концентрації, локалізації, диференціації.

Варіаційний розмах (амплітуда коливань) – абсолютна різниця між максимальним та мінімальним значенням ознаки в сукупності.

$$R = x_{\max} - x_{\min} \quad (6.6)$$

У випадку, коли максимальне та/або мінімальне значення ознаки виявляються недостатньо типовим для досліджуваної сукупності, то доцільно використовувати квартильний (охоплює 50% сукупності) або децильний (охоплює 80% сукупності) розмахи:

$$\text{Квартильний розмах } R_Q = Q_3 - Q_1 \quad (6.7)$$

$$\text{Децильний розмах } R_D = D_9 - D_1 \quad (6.8)$$

Середнє лінійне відхилення (середня з модулів відхилень) характеризує середнє відхилення індивідуальних значень ознаки (x_i) від середнього значення (\bar{x}) у статистичній сукупності.

$$\bar{l} = \frac{\sum_1^n |x_i - \bar{x}| f_i}{\sum_1^n f_i} \quad (6.9)$$

За первинними, не згрупованими даними середнє лінійне відхилення розраховується за методом простої (незваженої) середньої:

$$\bar{l} = \frac{\sum_1^n |x_i - \bar{x}|}{\sum_1^n n} \quad (6.10)$$

Дисперсія – це середній квадрат відхилень індивідуальних значень ознаки від середнього значення у статистичній сукупності.

$$\sigma^2 = \frac{\sum_1^n (x_i - \bar{x})^2 f_i}{\sum_1^n f_i} \quad (6.11)$$

За первинними, не згрупованими даними за методом простої (незваженої) середньої:

$$\sigma^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{\sum_1^n n} \quad (6.12)$$

Стандартне (середнє квадратичне) відхилення є найбільш вживаною абсолютною мірою варіації, за змістом ідентичне середньому лінійному відхиленню.

$$\sigma = \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2 f_i}{\sum_1^n f_i}} \quad (6.13)$$

За первинними, не згрупованими даними за методом простої (незваженої) середньої:

$$\sigma = \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2}{\sum_1^n n}} \quad (6.14)$$

Середнє кuartильне відхилення – характеризує варіацію ознаки навколо центру розподілу, тобто тільки у центральній частині досліджуваної сукупності.

$$Q = \frac{Q_3 - Q_1}{2} \quad (6.15)$$

Усі абсолютні характеристики варіації, окрім дисперсії, є іменованими величинами, виражені в одиницях вимірювання ознаки.

Відносні характеристики варіації – коефіцієнт варіації – визначаються як відношення абсолютних характеристик варіації

до центру розподілу, що виражений середньою арифметичною або медіаною.

$$\text{Квадратичний коефіцієнт варіації } V_{\sigma} = \frac{\sigma}{\bar{x}} 100 \quad (6.16)$$

Квадратичний коефіцієнт варіації характеризує не лише відносний рівень варіації, а й однорідність сукупності.

Сукупність вважається однорідною за умови, що $V_{\sigma} \geq 33\%$

$$\text{Лінійний коефіцієнт варіації } V_{\bar{l}} = \frac{\bar{l}}{\bar{x}} 100 \quad (6.17)$$

$$\text{Квартильний коефіцієнт варіації } V_Q = \frac{Q_3 - Q_1}{2Me} \quad (6.18)$$

$$\text{Коефіцієнт осциляції } V_R = \frac{R}{\bar{x}} \quad (6.19)$$

$$\text{Коефіцієнт децильної диференціації } V_D = \frac{D_9}{D_1} \quad (6.20)$$

Приклад розрахунку характеристик варіації на прикладі варіаційного ряду розподілу 200 малих підприємств за розміром прибутку.

Таблиця 6.5

**Дані для розрахунку характеристик варіації
у інтервальному ряду розподілу**

№ п/п	Прибуток, тис. грн	Кількість підприємств, од.	x'_i	$x'_i f_i$	$(x'_i - \bar{x})^2$	$(x'_i - \bar{x})^2 f_i$	$ x'_i - \bar{x} f_i$
1	до 200	3	100	300	540225	1620675	2205
2	200–400	6	300	1800	286225	1717350	3210
3	400–600	24	500	12000	112225	2693400	8040
4	600–800	50	700	35000	18225	911250	6750
5	800–1000	65	900	58500	4225	274625	4225

**Розділ 6. АНАЛІЗ ВАРІАЦІЇ
ТА ФОРМИ РОЗПОДІЛУ**

Закінчення табл. 6.5

№ п/п	Прибуток, тис. грн	Кількість підприємс тв, од.	x'_i	$x'_i f_i$	$(x'_i - \bar{x})^2$	$(x'_i - \bar{x})^2 f_i$	$ x'_i - \bar{x} f_i$
6	1000–1200	43	1100	47300	70225	3019675	11395
7	1200–1400	7	1300	9100	216225	1513575	3255
8	1400 і більше	2	1500	3000	442225	884450	1330
Σ		200	x	167000	x	12365000	40410

Розрахунок стандартного відхилення:

Визначаємо серединне значення кожного інтервалу (x'_i).

Визначаємо середнє значення $\bar{x} = \frac{167000}{200} = 835$ грн.

Визначаємо в кожній групі квадрат відхилення значення ознаки від середнього значення $(x'_i - \bar{x})^2$, перемножуємо квадрат відхилення на частоту $(x'_i - \bar{x})^2 f_i$;

Обчислюємо стандартне відхилення за формулою (6.13):

$$\sigma = \sqrt{\frac{12365000}{200}} = 251,35 \text{ грн.}$$

Розрахунок середнього лінійного відхилення:

Визначаємо в кожній групі модуль відхилення значення ознаки від середнього значення та перемножуємо його на частоту $|x'_i - \bar{x}| f_i$;

Обчислюємо середнє лінійне відхилення за формулою (6.9):

$$\bar{l} = \frac{40410}{200} = 202,05 \text{ тис. грн.}$$

Розрахунок коефіцієнтів варіації:

Квадратичний коефіцієнт варіації

$$V_{\sigma} = \frac{251,35}{835} 100 = 30,10\%;$$

Лінійний коефіцієнт варіації

$$V_i = \frac{202,05}{835} 100 = 24,20\%$$

Статистичний висновок. За результатами дослідження розміру прибутку двохсот малих підприємств встановлено, що середній розмір прибутку становить 835 тис. грн. Стандартне відхилення дорівнює 251,35 тис. грн, це означає, що прибуток підприємств у досліджуваній сукупності відхиляється від середнього значення по сукупності в середньому на 251,35 тис. грн. Опираючись на значення квадратичного коефіцієнта варіації, яке дорівнює 30,10%, робимо висновок про те, що досліджувана сукупність є однорідною, а середнє значення прибутку є типовим для цієї сукупності.

6.3. Види та взаємозв'язок дисперсій

Дисперсія, завдяки своїм математичним властивостям, є невід'ємним і важливим елементом статистичних методів аналізу, входить до більшості теорем ймовірностей, використовується для вимірювання взаємозв'язків, для вибіркового методу дослідження тощо. Серед інших характеристик варіації, дисперсія найбільше використовується у статистичному аналізі економічних, соціальних та інших явищ.

Означення. *Дисперсія* (від лат.(латинський) *dispersio* — розсіяння) – це середній квадрат відхилень індивідуальних значень ознаки від середнього значення у статистичній сукупності.

$$\sigma^2 = \frac{\sum_1^n (x_i - \bar{x})^2 f_i}{\sum_1^n f_i} \quad (6.21)$$

Математичні властивості дисперсії:

➤ При зменшенні/збільшенні всіх значень варіанти x_i на сталу величину A дисперсія σ^2 не зміниться: $\sigma_{(x_i -/+A)}^2 = \sigma_{x_i}^2$

➤ При зміні всіх варіанти x_i в A разів – дисперсія зміниться в A^2 разів: $\sigma_{xA}^2 = \sigma_x^2 A^2$

➤ Дисперсія – це різниця квадратів: $\sigma^2 = \overline{x^2} - \bar{x}^2$, де $\overline{x^2}$ – середній квадрат значення ознаки; \bar{x}^2 – квадрат середньої величини.

➤ Дисперсія альтернативної ознаки обчислюється як добуток частки одиниць сукупності, яким властива ознака (d_1) та частки одиниць сукупності, яким ця ознака не властива (d_0): $\sigma^2 = d_1 d_0$

➤ Якщо частоти (f_i) замінити частками (d_i), дисперсія не зміниться.

Зауваження. Відхилення індивідуальних значень ознаки x від загального середнього значення \bar{x} у структурованих сукупностях розкладається на дві складові:

$$(x - \bar{x}) = (x - \bar{x}_l) + (\bar{x}_l - \bar{x})$$

де \bar{x}_l – групова середня.

Узагальнюючими характеристиками вищезазначених відхилень є дисперсії: загальна, групова та міжгрупова

Загальна дисперсія характеризує варіацію ознаки x навколо загальної середньої \bar{x} :

$$\sigma^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{\sum_1^n n} \quad (6.22)$$

Групова дисперсія характеризує варіацію ознаки x всередині кожної i -ї групи навколо групової середньої \bar{x}_l :

$$\sigma_i^2 = \frac{\sum_1^{f_i} (x - \bar{x}_l)^2}{f_i} \quad (6.23)$$

Середня із групових дисперсій є узагальнюючою мірою внутрішньо групової варіації:

$$\overline{\sigma^2} = \frac{\sum_1^n \sigma_i^2 f_i}{\sum_1^n f_i} \quad (6.24)$$

Міжгрупова дисперсія – характеризує варіацію групових середніх \bar{x}_i навколо загальної середньої \bar{x} :

$$\delta^2 = \frac{\sum_1^n (\bar{x}_i - \bar{x})^2 f_i}{\sum_1^n f_i} \quad (6.25)$$

Загальна дисперсія характеризує міру варіації, що спричинена впливом як факторної ознаки, яку покладено в основу групування, так і інших чинників, окрім факторної ознаки.

Міжгрупова дисперсія характеризує міру варіації, що спричинена тільки факторною ознакою.

Внутрішньогрупова дисперсія характеризує міру варіації, що спричинене іншими чинниками, окрім впливу факторної ознаки.

Загальна дисперсія складається із міжгрупової та внутрішньо групової дисперсій. Такий взаємозв'язок називається правилом декомпозиції (розкладання) варіації:

$$\sigma^2 = \overline{\sigma^2} + \delta^2 \quad (6.26)$$

Відношення міжгрупової дисперсії до загальної дисперсії характеризує частку вакації результативної ознаки, яка пов'язана із варіацією факторної (групувальної) ознаки. Таке відношення називається кореляційне відношення:

$$\eta = \frac{\delta^2}{\sigma^2} \quad (6.27)$$

Приклад розрахунку дисперсій наведений на прикладі варіації успішності студентів залежно від щоденно витраченого часу на самостійне вивчення навчальних дисциплін.

Таблиця 6.6

Дані для розрахунку загальної, групових дисперсій, міжгрупової та середньої з групових дисперсій

Витрати часу на самостійну підготовку, ГОДИН	Середній бал за 100 баловою шкалою x_i	Чисельність студентів, f_i	$\sum x_i$	Середній бал у групі \bar{x}_i	$\sum x_i^2$	\bar{x}_i^2	\bar{x}_i^2	Групова дисперсія σ_i^2	Для розрахунку дисперсій		
									Середня з групових $\sigma_i^2 f_i$	Міжгрупова	
										$(x_i - \bar{x})$	$(x_i - \bar{x})^2 f_i$
1	65, 55, 70, 45, 35	5	270	54,00	15400,00	3080,00	2916,00	164,00	820,00	-15,25	1162,81
2	70, 75, 65, 55, 80, 85, 70, 65	8	565	70,63	40525,00	5065,63	4987,89	77,74	621,88	1,38	15,13
3	75, 70, 80, 80, 90, 85, 70	7	550	78,57	43550,00	6221,43	6173,47	47,96	335,71	9,32	608,22
Загалом по сукупності		20	1385	69,25	99475,00	4973,75	4795,56	x	1777,59	x	1786,16

Розрахунок дисперсій

Обчислимо середній бал по сукупності:

$$\bar{x} = \frac{\sum_1^n x}{n} = \frac{1385}{20} = 69,25.$$

Розрахуємо загальну дисперсію як різницю квадратів

$$\sigma^2 = \overline{x^2} - \bar{x}^2 = 4973,75 - 4795,56 = 178,19.$$

Розрахуємо групові середні бали успішності та групові дисперсії

$$\bar{x}_1 = \frac{270}{5} = 54,00; \quad \bar{x}_2 = \frac{565}{8} = 70,63; \quad \bar{x}_3 = 7 = 78,57$$

$$\sigma_1^2 = \overline{x_1^2} - \bar{x}_1^2 = 3080,00 - 2916,00 = 164,00$$

$$\sigma_2^2 = \overline{x_2^2} - \bar{x}_2^2 = 5065,63 - 4987,89 = 77,74$$

$$\sigma_3^2 = \overline{x_3^2} - \bar{x}_3^2 = 6221,43 - 6173,47 = 47,96.$$

Розрахуємо середню із групових дисперсій

$$\overline{\sigma^2} = \frac{\sum_1^n \sigma_i^2 f_i}{\sum_1^n f_i} = \frac{1777,59}{20} = 88,88.$$

Розрахуємо міжгрупову дисперсію

$$\sigma^2 = \frac{\sum_1^n (x_i - \bar{x})^2 f_i}{\sum_1^n f_i} = \frac{1786,16}{20} = 89,31.$$

Сума середньої із групових дисперсій та міжгрупової дисперсії дорівнює загальній дисперсії

$$\sigma^2 = \overline{\sigma^2} + \sigma^2 = 88,88 - 89,31 = 178,19$$

Розрахуємо кореляційне відношення

$$\eta^2 = \frac{\delta^2}{\sigma^2} = \frac{89,31}{178,19} = 70,63$$

Статистичний висновок. У розглянутому прикладі кореляційне відношення становить 0,501. Це свідчить про те, що варіація успішності студентів на 50,1% пояснюється величиною витрат часу на самопідготовку та на 49,9% іншими факторами.

6.4. Характеристики форми розподілу. Порівняльний аналіз структур

Оцінювання ступеня однорідності сукупності є одним із найважливіших завдань аналізу закономірностей розподілу, оскільки однорідність є передумовою використання інших статистичних методів дослідження економічних, соціальних та інших явищ.

Однорідними вважаються сукупності, елементам яких властиві спільні риси і належать вони до одного типу. Але однорідність сукупності не передбачає цілковитої тотожності елементів, а тільки наявність у них спільних властивостей в істотному вираженні.

У однорідної сукупності розподіл одновершинний – одномодальний, тобто наявне лише одне модальне значення.

Одновершинний розподіл може бути симетричний та асиметричний (скошений), гостровершинний та плосковершинний.

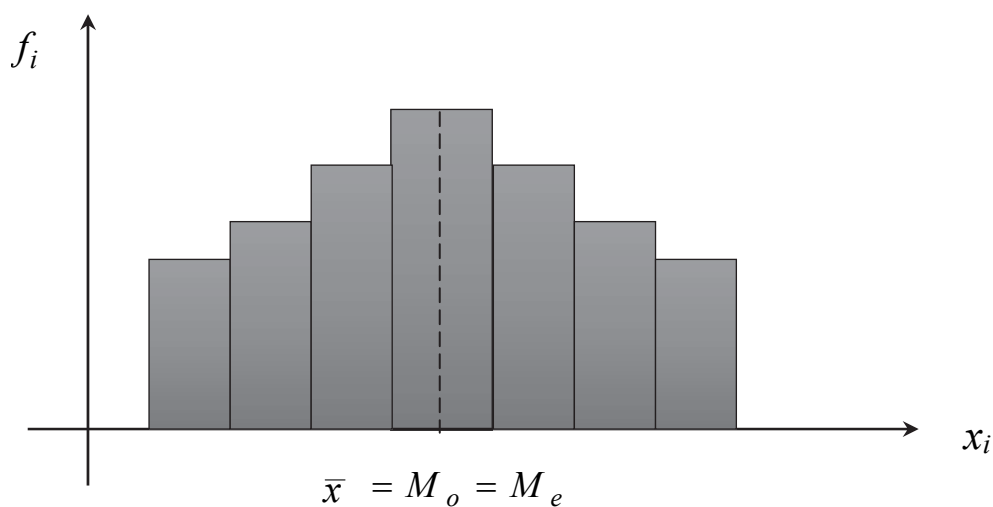


Рис. 6.1. Гістограма симетричного розподілу

Джерело: розроблено авторами.

Ступінь асиметрії розподілу вимірюється за допомогою коефіцієнтів, які характеризують напрям та міру скошеності:

$$A = \frac{\bar{x} - M_o}{\sigma} \quad (6.28)$$

$$\text{або } A = \frac{3(\bar{x} - M_e)}{\sigma} \quad (6.29)$$

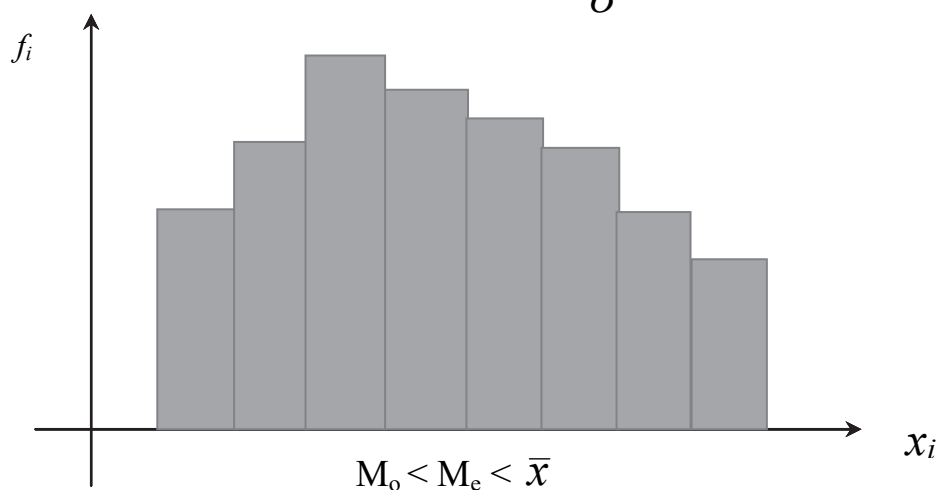
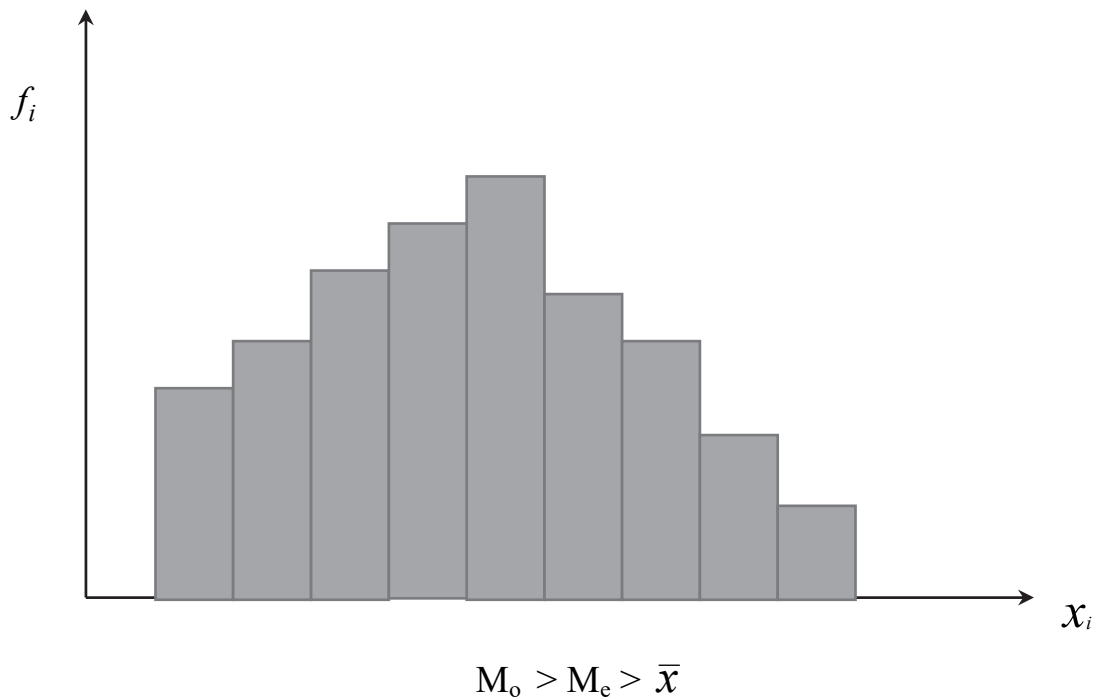


Рис. 6.2. Гістограма асиметричного розподілу із правобічною асиметрією

Джерело: розроблено авторами.



*Рис. 5.3. Гістограма асиметричного розподілу
із лівобічною асиметрією*

Джерело: розроблено авторами.

$A > 0$ – правобічна асиметрія;

$A < 0$ – лівобічна асиметрія;

$A < |0,25|$ – асиметрія низька;

$|0,25| < A < |0,50|$ – асиметрія помірною;

$A > |0,50|$ – асиметрія висока.

Екссес розподілу – це властивість одновершинного розподілу, яка характеризує ступінь зосередженості елементів сукупності навколо центру розподілу.

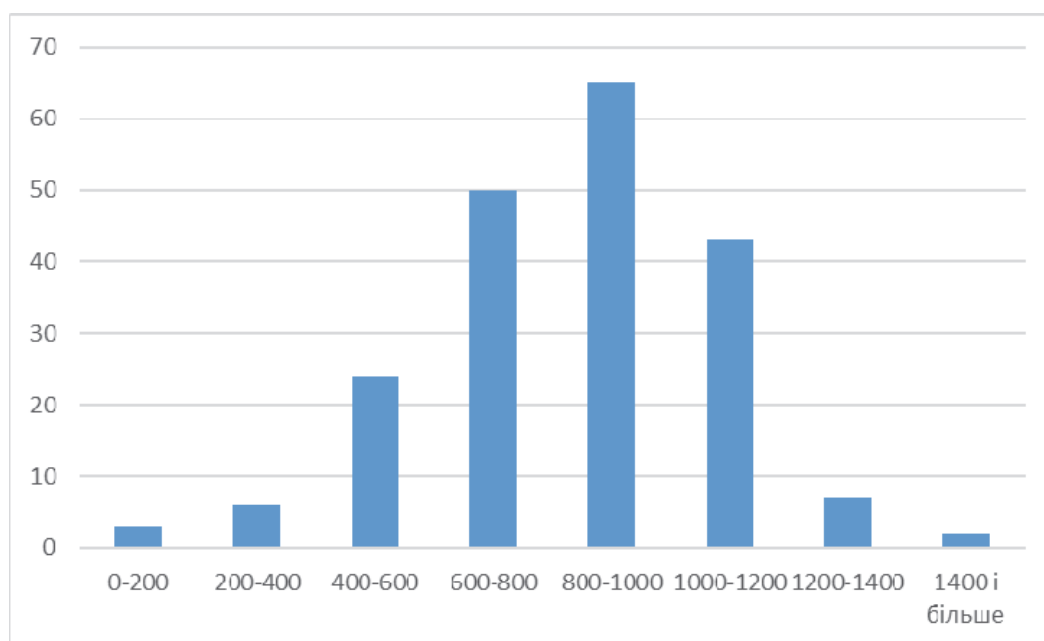


Рис. 6.4. Гістограма з гостровершинним розподілом

Джерело: розроблено авторами.

Чим менша варіація, тим гострішу вершину має розподіл і навпаки. Комплексне оцінювання асиметрії (скошеності) та ексцесу (крутизни) розподілу здійснюють на основі центральних моментів розподілу.

Алгебраїчно центральний момент розподілу – це середня арифметична k -го ступеня відхилення індивідуальних значень ознаки від середньої:

$$\mu_k = \frac{\sum_1^n (x_i - \bar{x})^k f_i}{\sum_1^n f_i} \quad (6.30)$$

Момент другого порядку (μ_2) – дисперсія

Момент третього порядку (μ_3) – характеризує асиметрію

$$\mu_3 = \frac{\sum_1^n (x_i - \bar{x})^3 f_i}{\sum_1^n f_i}. \quad (6.31)$$

У симетричному розподілі $\mu_3 = 0$. Чим більша скошеність ряду розподілу, тим більшим буде значення μ_3 .

Коефіцієнт асиметрії визначається за формулою:

$$A_s = \frac{\mu_3}{\sigma^3}, \quad (6.32)$$

де σ^3 – стандартне відхилення у третій степені.

Момент четвертого порядку (μ_4) – характеризує ексцес

$$\mu_4 = \frac{\sum_1^n (x_i - \bar{x})^4 f_i}{\sum_1^n f_i}. \quad (6.33)$$

Коефіцієнт ексцесу визначається за формулою:

$$E_k = \frac{\mu_4}{\sigma^4} - 3. \quad (6.34)$$

У симетричному розподілі $E_k = 0$, у гостровершинному розподілі $E_k > 0$, у плосковершинному – $E_k < 0$.

Для здійснення порівняльного аналізу структур розраховують низку коефіцієнтів, які дають змогу визначити подібність структури об'єктів однакового типу, оцінити концентрацію, локалізацію значень ознаки у складових сукупності, оцінити інтенсивність структурних зрушень тощо.

Коефіцієнт подібності структур – характеризує подібність (схожість) структур двох об'єктів або одного об'єкта за двома ознаками.

$$P = 1 - 0,5 \sum_1^m |d_i - d_k|, \quad (6.35)$$

де d_i – частка складових i -го об'єкта;

d_k – частка складових k -го об'єкта.

Якщо структури порівнюваних об'єктів однакові, то $P=1$, якщо абсолютно протилежні $P=0$. Отже, чим більше подібні структури, тим більше P наближається до 1.

Коефіцієнт локалізації – характеризує нерівномірність розподілу значень ознаки між окремими складовими сукупності.

$$L_i = \frac{D_i}{d_i} \cdot 100, \quad (6.36)$$

де D_i – частка розподілу за обсягом значень ознаки;
 d_i – частка розподілу за кількістю елементів сукупності.

Якщо розподіл рівномірний, то всі значення $L_i = 1$. Якщо відбувається концентрація значень ознаки в i -й складовій, то $L_i > 1$ і навпаки.

Коефіцієнт концентрації – характеризує концентрацію ознаки в окремих складових сукупності.

$$K = 0,5 \sum_1^m |d_i - D_i| \quad (6.37)$$

У рівномірному розподілі $K=0$. Чим більша концентрація, тим більшим буде значення K .

Важливо знати. Для визначення характеристик розподілу можна використовувати Excel, статистичні функції якого надають користувачам можливість автоматизації обчислення характеристик центру розподілу, характеристик варіації тощо (рис. 6.5).

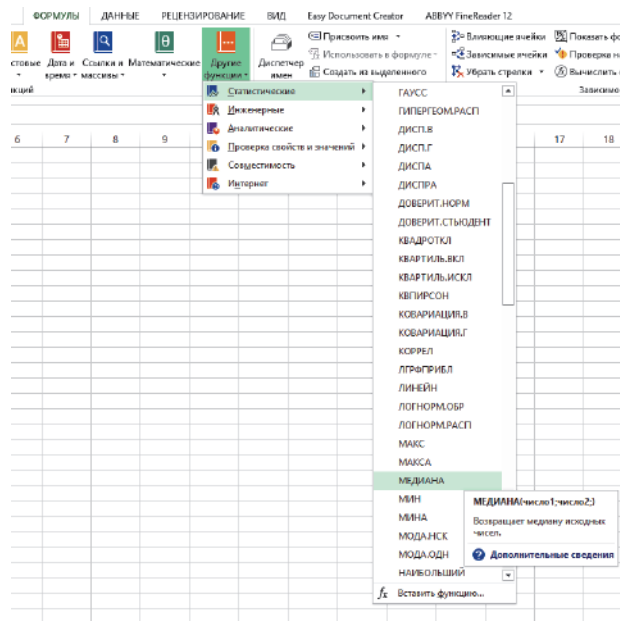


Рис. 6.5. Статистичні функції Excel

6.5. Задачі для самостійної роботи

Задача 6.1. Обчислити характеристики центру розподілу, абсолютні і відносні показники варіації та пояснити отримані результати, використовуючи дані табл. 6.7–6.18.

Таблиця 6.7

Середня заробітна плата спеціалістів сфери інформаційних технологій

Середня заробітна плата, тис. грн	Чисельність спеціалістів, осіб
До 15	12
15–20	13
20–25	15
25–30	19
30–35	19
35–40	13
Більше 40	14
Разом	105

Таблиця 6.8

Розподіл жінок віком від 18 до 23 років за бажаною кількістю дітей

Бажана кількість дітей	Чисельність жінок, осіб
0	3
1	9
2	58
3	15
Разом	85

СТАТИСТИКА

Таблиця 6.9

Розподіл туристів за тривалістю відпочинку на курорті

Тривалість відпочинку, днів	Чисельність туристів, осіб
До 7	22
7–14	33
14–21	25
21–28	19
Більше 28	6
Разом	105

Таблиця 6.10

Розподіл депозитів у банку за розміром

Розмір депозиту, тис. грн	Кількість угод, шт.
До 200	22
200–400	13
400–600	18
600–800	20
800–1000	35
1000–1200	42
Більше 1200	50
Разом	200

Таблиця 6.11

Розподіл студентів за розміром витрат на дозвілля

Витрати на дозвілля, грн/тиждень	Чисельність студентів, осіб
До 500	157
500-1000	162
1000-1500	137
1500-2000	29
Більше 2000	12
Разом	497

**Розділ 6. АНАЛІЗ ВАРІАЦІЇ
ТА ФОРМИ РОЗПОДІЛУ**

Таблиця 6.12

Розподіл відвідувачів ресторану за вартістю замовлення

Вартість замовлення, грн	Чисельність відвідувачів, осіб
До 150	2
150–300	7
300–450	10
450–600	20
600–750	25
750–900	4
Більше 900	2
Разом	70

Таблиця 6.13

**Розподіл кредитів, виданих банком під заставу майна за
величиною кредитної ставки**

Кредитна ставка, %	Сума кредитів, тис. грн
15	53 500
18	78 400
21	100 800
24	89 500
Разом	322 200

Таблиця 6.14

**Розподіл смартфонів, реалізованих торговельною
мережею протягом місяця за ціною категорією**

Ціна смартфона, USD	Кількість реалізованих смартфонів, шт.
До 100	55
100–200	78
200–300	110

СТАТИСТИКА*Закінчення табл. 6.14*

Ціна смартфона, USD	Кількість реалізованих смартфонів, шт.
300–400	85
400–500	62
Більше 500	35
Разом	425

*Таблиця 6.15***Розподіл платників податків за середньомісячним розміром податкових зобов'язань із податку на додану вартість**

Середньомісячний розмір податку, тис. грн	Кількість платників податку, шт.
До 20	125
20–40	185
40–60	260
60–80	480
80–100	360
Більше 100	165
Разом	1 575

*Таблиця 6.16***Розподіл аграрних підприємств регіону за розміром валового збору зернових культур**

Валовий збір, тис. т	Кількість аграрних підприємств, шт.
До 12	325
12–14	485
14–16	660
16–18	540
18–20	260
Більше 200	165
Разом	2435

**Розділ 6. АНАЛІЗ ВАРІАЦІЇ
ТА ФОРМИ РОЗПОДІЛУ**

Задача 6.2. Використовуючи дані табл. 6.17–6.18, визначити квартилі розподілу, перший та дев'яті дециль, децильний розмах, коефіцієнт децильної диференціації. Проаналізувати отримані результати, порівняти результати дослідження між двома регіонами.

Таблиця 6.17

Розподіл домогосподарств регіону А за розміром щомісячного середньодушового грошового доходу

Середньодушові грошові доходи на місяць, грн	Кількість домогосподарств, шт.
До 1500	22
1500–2000	13
2000–2500	18
2500–3000	20
3000–3500	35
3500–4000	42
4000–4500	28
4500–5000	12
5000 і більше	10
Разом	200

Таблиця 6.18

Розподіл домогосподарств регіону В за розміром щомісячного середньодушового грошового доходу

Середньодушові грошові доходи на місяць, грн	Кількість домогосподарств, шт.
До 1500	4
1500–2000	10
2000–2500	12
2500–3000	20
3000–3500	30
3500–4000	40

СТАТИСТИКА

Закінчення табл. 6.18

Середньодушові грошові доходи на місяць, грн	Кількість домогосподарств, шт.
4000–4500	48
4500–5000	22
5000 і більше	14
Разом	200

Задача 6.3. На підставі даних табл. 6.19–6.21 обчислити усі види дисперсій та кореляційне відношення. Зробити висновки щодо отриманих результатів.

Таблиця 6.19

Результати продажу смартфонів 15-ма торговельними точками характеризується даними

Виробник смартфона	Денний обсяг продажу, шт.														
Samsung	10	15	8	13	10	17	12	9	14	15	12	14	9	11	16
Apple	10	9	8	11	6	7	9	2	15	8	6	10	2	8	6
Xiaomi	5	8	10	4	12	4	12	1	10	3	11	8	4	10	2

Таблиця 6.20

Розподіл депозитних вкладів у комерційному банку

Категорія вкладника	Відсоток до загальної чисельності вкладників, %	Середній розмір депозитного вкладу, тис. грн	Дисперсія депозитних вкладень
Юридична особа	12	250	4500
Фізична особа – підприємець	8	20	300
Фізична особа	80	150	3750

**Розділ 6. АНАЛІЗ ВАРІАЦІЇ
ТА ФОРМИ РОЗПОДІЛУ**

Таблиця 6.21

**Розподіл платників податку за розміром зниження
податкових зобов'язань із податку на додану вартість
характеризується даними:**

Вид економічної діяльності	Відсоток до загальної чисельності вкладників, %	Середній розмір заниженого податкового зобов'язання, тис. грн	Дисперсія занижених податкових зобов'язань
Сільське, лісове та рибне господарство	15	180	3800
Промисловість	12	120	1750
Операції з нерухомим майном	10	100	1000
Оптова та роздрібна торгівля	27	350	12100
Будівництво	8	280	8400
Професійна, наукова та технічна діяльність	8	50	225
Інформація та телекомунікації	4	90	1020
Інші види діяльності	6	105	1205

Задача 6.4. За даними табл. 6.22 визначити коефіцієнти галузевої локалізації працівників, зробити висновок.

Таблиця 6.22

**Розподіл суб'єктів господарювання та працівників
за видами економічної діяльності**

Вид економічної діяльності	Кількість суб'єктів господарювання, одиниць	Середньооблікова чисельність штатних працівників, тис. осіб
Сільське, лісове та рибне господарство	76593	430,5

СТАТИСТИКА

Закінчення табл. 6.22

Вид економічної діяльності	Кількість суб'єктів господарювання, одиниць	Середньооблікова чисельність штатних працівників, тис. осіб
Промисловість	123876	1872,6
будівництво	50261	166,2
Оптова та роздрібна торгівля	837797	704,8
Транспорт, складське господарство, поштова та кур'єрська діяльність	95815	643,9
Тимчасове розміщування й організація харчування	57578	67,8
Інформація та телекомунікації	146909	116,2
Фінансова та страхова діяльність	10290	167,8
Операції з нерухомим майном	90553	83,0
Професійна, наукова та технічна діяльність	121546	218,1
Діяльність у сфері адміністративного та допоміжного обслуговування	46553	631,8
Освіта	11656	1401,0
Охорона здоров'я та надання соціальної допомоги	22085	949,4
Мистецтво, спорт, розваги та відпочинок	13045	150,5
Надання інших видів послуг	100587	27,0

Задача 6.5. За даними табл. 6.23 здійснити оцінювання подібності готелів двох міст за категорією зірок. Зробити висновки.

Таблиця 6.23

Розподіл готелів міста за категорією зірок

Категорія готелю	% до підсумку	
	Місто А	Місто В
Одна зірка	7	9
Дві зірки	17	7
Три зірки	23	25
Чотири зірки	48	15
П'ять зірок	5	44

6.6. Питання для самоконтролю

1. У чому полягає сутність аналізу закономірностей розподілу статистичних сукупностей?
2. Визначити аналітичні функції та види варіаційних рядів розподілу.
3. Надати визначення характеристик центру розподілу. При якому співвідношенні зазначених характеристик розподіл вважається симетричним?
4. Що таке квантілі розподілу? З якою метою їх розраховують?
5. Обґрунтувати необхідність здійснення оцінювання варіації. Які показники характеризують ступінь варіації у статистичній сукупності?
6. У чому полягає сутність дисперсійного аналізу?
7. Сформулювати основні математичні властивості дисперсії.
8. Розкрити сутність взаємозв'язку дисперсій.
9. Які коефіцієнти характеризують ступінь асиметрії розподілу?
10. У чому полягає сутність порівняльного аналізу структур?

Розділ 7

ВИБІРКОВИЙ МЕТОД

План

- 7.1. Суть вибіркового спостереження.
- 7.2. Види і способи вибіркового спостереження.
- 7.3. Похибки вибірки.
- 7.4. Оцінювання параметрів генеральної сукупності.
- 7.5. Визначення мінімально необхідного обсягу вибіркової сукупності.
- 7.6. Задачі для самостійної роботи.
- 7.7. Питання для самоконтролю.

7.1. Суть вибіркового спостереження

Вибірковим спостереженням називається такий вид несучільного спостереження, за характеристикою відібраної частини одиниць якого мають уявлення про всю сукупність.

Розрізняють генеральну сукупність і вибіркoву сукупність.

Генеральна сукупність – це загальна сукупність одиниць, що являє собою досліджуване явище, процес і з якої здійснюється відбір.

Вибірковою називається частина генеральної сукупності, що підлягає безпосередньому обстеженню.

До вибіркового спостереження статистика вдається у випадках, коли недоцільно або неможливо проводити суцільне спостереження. Зокрема, це труднощі з отримання потрібної інформації щодо всіх одиниць сукупності; властивості об'єкта дослідження; витрати, пов'язані із дослідженням кожної одиниці сукупності тощо.

Обсяг генеральної сукупності позначають через літеру N . Обсяг вибіркової сукупності – n .

Відповідно генеральна середня позначається: \bar{x} , а середня вибіркова – \tilde{x} .

Для достовірного представлення генеральної сукупності вибірка має бути *репрезентативною*. Вибірка вважається репрезентативною, якщо структура вибірки відповідає структурі генеральної сукупності і кожна одиниця останньої має однакову з іншими ймовірність потрапити до вибіркової сукупності. Відбір одиниць при цьому здійснюється випадковим чином.

Отримання репрезентативної вибірки забезпечується правильним застосуванням видів, способів та методів проведення вибіркового спостереження.

7.2. Види і способи вибіркового спостереження

Розрізняють два методи відбору одиниць у вибірку сукупність: повторний і безповторний.

Повторним називається такий метод відбору, за якого кожна раніше відібрана одиниця повертається в генеральну сукупність і може знову потрапити до вибірки.

Приклад. Проводиться опитування споживачів у магазині. Існує певна ймовірність того, що наступного дня буде опитаний той же споживач, оскільки він знов прийшов до магазину, де він є постійним покупцем.

Безповторним називається такий метод відбору, за якого кожна раніше відібрана одиниця не повертається в генеральну сукупність і в подальшому потрапити до вибірки вже не може.

Приклад. Аудитор здійснює перевірку документів фірми. Кожний перевірений документ він перекладає в окрему папку, отже, потрапити знову до перевірки цей документ не може.

Оскільки безповторний відбір охоплює постійно нові одиниці сукупності, а повторний – одну і ту ж сукупність, тому безповторний відбір дає більш точні результати.

Повторний і безповторний методи відбору, залежно від характеру одиниці відбору, застосовуються в поєднанні з іншими видами відбору.

У практиці статистичного дослідження використовуються три види відбору:

- 1) індивідуальний – відбір окремих одиниць сукупності;
- 2) груповий (серійний) – відбір груп (серій) одиниць;
- 3) комбінований – комбінація індивідуального і групового.

Способом відбору називається система організації відбору одиниць з генеральної сукупності.

За способом відбору одиниць для обстеження розрізняють такі види спостереження:

- 1) власне-випадкова вибірка;
- 2) механічна вибірка;
- 3) типова (районована) вибірка;
- 4) серійна (гніздова) вибірка;
- 5) комбінована вибірка;
- 6) одноступінчаста і багатоступінчаста вибірка;
- 7) однофазна і багатозфазна вибірка;
- 8) інші види вибірки.

Власне випадковою називається така вибірка, за якої відбір одиниць з генеральної сукупності є випадковим. Часто для цього застосовують жеребкування або таблицю (генератор) випадкових чисел.

Механічна вибірка – це послідовний відбір одиниць через рівні проміжки у порядку визначеного розташування їх в генеральній сукупності, або в якому-небудь переліку. Інтервали відбору визначаються відповідно з часткою відбору одиниць (кожна п'ята, десята, сота і т.д.).

Відбір елементів здійснюється через однакові інтервали, Крок інтервалу залежить від частки вибірки.

Так, при $\frac{n}{N} = 0,05$ крок інтервалу становить

$$\frac{1}{0,05} = 20.$$

Недоліком механічної вибірки є те, що перед вибіркою необхідно мати повний облік одиниць сукупності, потім потрібно провести ранжування і лише після цього можна робити вибірку з певним інтервалом.

При типовому (районованому) відборі генеральну сукупність поділяють на однорідні групи, райони за певною ознакою. Потім з кожної групи випадковим або механічним способом відбирають певну кількість одиниць, пропорційно частці групи в загальній сукупності.

Під час серійної (гніздової) вибірки відбір одиниць здійснюють цілими групами (серіями, гніздами) сукупності, в межах яких обстежують усі одиниці без винятку. Серії для спостереження відбирають випадково, частіше неповторним способом механічної вибірки.

Комбінованою називається така вибірка, коли комбінують два або кілька видів вибірок. Перш за все, комбінують суцільне і вибіркоче спостереження. У цьому випадку, за основною програмою обстежується генеральна сукупність, а за додатковою – вибіркова.

Одноступінчастою називається вибірка, коли із досліджуваної сукупності зразу відбираються одиниці або серії одиниць для безпосереднього обстеження.

Багатоступінчаста вибірка передбачає поступове вилучення із генеральної сукупності спочатку укрупнення груп одиниць, потім груп менших за обсягом, і так до тих пір, поки не відберуть відповідні групи або одиниці, які будуть досліджуватись. Вибірка може бути двоступінчастою, триступінчастою і більше.

Якщо необхідні дані можна отримати на основі вивчення всіх первинно відібраних одиниць, застосовують **однофазну вибірку**, а якщо тільки на основі деякої її частини, відібраної так, що вона складає підвибірку із початково проведеної вибірки – **багатофазну**.

Багатофазною називається така вибірка, коли одні відомості збираються від усіх одиниць відбору, потім відбираються ще деякі одиниці і обстежуються за більш широкою

програмою. При багатофазній вибірці на кожній фазі зберігається одна і та ж одиниця відбору.

Розрахунок помилок репрезентативності багатоступінчастої і багатофазної вибірок проводиться для кожного ступеня і фази окремо.

Бувають випадки, коли необхідно застосувати інші види відбору, такі як взаємопроникаючі і квантильні вибірки, направлений відбір, моментні спостереження, або скористатись малою вибіркою.

Взаємопроникненні – такі вибірки, коли із однієї генеральної сукупності здійснюють одним і тим же способом декілька незалежних вибірок. Взаємопроникненні вибірки завжди проводять різні, незалежні один від одного дослідники, що дозволяє порівнювати підсумки по всіх частинах і забезпечити взаємну перевірку їх роботи. Взаємопроникненні вибірки дають незалежні одна від одної оцінки значень досліджуваної сукупності, і, якщо результати різних вибірок близькі між собою, то такі оцінки дуже переконливі.

7.3. Похибки вибірки

Статистичні характеристики вибіркової сукупності розглядаються як оцінки відповідних характеристик генеральної сукупності. Оскільки вибіркова сукупність не точно відтворює структуру генеральної, то вибіркові оцінки відрізнятимуться від значень параметрів генеральної сукупності.

Розбіжності між ними називають *похибками репрезентативності*. За причин виникнення ці похибки поділяють на систематичні (тенденційні) та випадкові.

Систематичні похибки виникають за умови, що під час формування вибіркової сукупності порушується принцип випадковості відбору (упереджений відбір елементів, недосконала основа вибірки тощо).

Випадкові похибки – це наслідок випадковості відбору елементів сукупності для обстеження.

При організації вибіркового обстеження важливо запобігти виникненню систематичних похибок.

Що стосується випадкових похибок, то уникнути їх неможливо, проте на основі теорії вибіркового методу можна визначити розмір і по можливості регулювати.

Фактори, що впливають на помилки репрезентативності:

- показники варіації певної ознаки (наприклад дисперсії): чим більший показник варіації – тим більший розмір помилки;
- чисельність вибірки: чим більша вибірка – тим менша вірогідність помилки (розмір помилки);
- спосіб відбору (повторний чи безповторний).

Для узагальнюючої характеристики помилок вибірки обчислюють середню помилку репрезентативності (інша назва – стандартна).

Позначається грецькою літерою μ (мю).

Стандартна похибка вибірки μ – це середнє квадратичне відхилення вибіркових оцінок від значення характеристики в генеральній сукупності.

Для повторної вибірки середня похибка для кількісного показника обчислюється як:

$$\mu = \sqrt{\frac{\sigma^2}{n-1}} \quad (7.1)$$

Оскільки дисперсія частки є добутком часток одиниць сукупності з наявністю і без наявності значення ознаки:

$$\sigma^2 = w(1-w) \quad (7.2)$$

то середня помилка для частки обчислюється як:

$$\mu_w = \sqrt{\frac{w(1-w)}{n}}, \quad (7.3)$$

де w – вибіркова частка одиниць із певною властивістю;
 n – обсяг вибіркової сукупності.

Для безповторного відбору розрахунок середньої похибки для кількісного показника і частки здійснюється за такими формулами:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma_x^2}{n} \left(1 - \frac{n}{N}\right)}, \quad (7.4)$$

$$\mu_w = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}. \quad (7.5)$$

Застосування безповторного відбору диктується його перевагами перед повторним у плані підвищення репрезентативності вибірки, особливо при незначних обсягах генеральної сукупності.

Зіставлення формул середньої похибки для повторного і безповторного відбору свідчить про те, що застосування безповторного відбору призводить до зменшення стандартної похибки вибірки.

Похибка механічної вибірки обчислюється за формулою безповторної вибірки.

Для визначення середньої помилки вибірки власне випадкової і механічної вибірки застосовують такі формули для повторного і безповторного відбору (табл. 7.1).

Таблиця 7.1

Формули для визначення середньої помилки вибірки

Способи відбору	Помилка вибірки	
	при визначенні середньої	при визначенні частки
Повторний	$\mu = \sqrt{\frac{\sigma^2}{n}}$	$\mu = \sqrt{\frac{w \cdot (1-w)}{n}}$
Безповторний	$\mu = \sqrt{\frac{\sigma^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$	$\mu = \sqrt{\frac{w \cdot (1-w)}{n} \cdot \left(1 - \frac{n}{N}\right)}$

При порівнянні точності вибірових оцінок використовують відносну похибку вибірки V_μ , яка показує, на скільки процентів вибірова оцінка відхиляється від параметра генеральної сукупності:

$$V_\mu = \frac{\mu_x}{x} \cdot 100 \quad (7.6)$$

Відносну похибку вибірки можна розрахувати на основі коефіцієнта варіації ознаки V_x :

для повторної вибірки:

$$V_\mu = 100 \cdot \frac{V_x}{\sqrt{n-1}} \quad (7.7)$$

для безповторної вибірки:

$$V_\mu = 100 \cdot \frac{V_x}{\sqrt{n-1}} \sqrt{1 - \frac{n}{N}} \quad (7.8)$$

При практичному використанні наведених формул потрібно враховувати, що:

1) у великих за обсягом сукупностях (30 і більше одиниць) поправка $\frac{n}{n-1}$ при розрахунку дисперсії не вносить істотних змін у розрахунки, а тому враховується лише у малих вибірках;

2) коригуючий множник для безповторної вибірки:

$$\sqrt{\frac{N-n}{N-1}} \approx \sqrt{1 - \frac{n}{N}} \quad (7.9)$$

при малих (менше ніж 5%) значеннях $\frac{n}{N}$ наближається до 1, а тому обчислення можна виконувати за формулою для повторної вибірки.

Особливо слід акцентувати увагу на так звану *малу вибірку*, зміст якої дійсно відповідає назві.

Така вибірка складається з порівняно невеликої кількості одиниць (20–30) і поширення їх результатів має певні особливості.

7.4. Оцінювання параметрів генеральної сукупності

Ймовірність відхилень вибіркової середньої від генеральної середньої при достатньо великому обсязі вибірки і обмеженій дисперсії генеральної сукупності підпорядковується закону нормального розподілу (детальніше про цей та інші закони розподілу ймовірностей дод. 2).

Величина відхилення вибіркової середньої від генеральної середньої при заданому рівні ймовірності має назву *граничної помилки вибірки*.

Величину граничної помилки вибірки обчислюють з певною ймовірністю « P », якій відповідає t -разове значення середньої помилки « μ ». З введенням показника кратності помилки « t », формула граничної помилки вибірки матиме вигляд:

$$\Delta = t \cdot \mu, \quad (7.10)$$

де t – коефіцієнт довіри, який залежить від ймовірності визначення граничної помилки.

При малих вибірках величину t визначають за розподілом ймовірностей Стьюдента. У дод. 4 наведені значення t для різних значень ймовірності.

У практиці вибірових досліджень застосовують два типи оцінок характеристик генеральної сукупності – *точкові* та *інтервальні*.

Точкова оцінка – це значення параметра досліджуваної сукупності (середня \tilde{x} , частка p) обчислене за даними вибірки.

Інтервальні оцінки подаються у вигляді *довірчих інтервалів*.

Довірчий інтервал – це інтервал, що із ймовірністю P покриває можливе значення параметра досліджуваної генеральної сукупності.

Для середньої межі довірчого інтервалу визначають на основі точкової оцінки середньої та граничної похибки вибірки

$$\tilde{x} - \Delta \leq \bar{x} \leq \tilde{x} + \Delta \quad (7.11)$$

або

$$\bar{x} - t \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \bar{x} + t \frac{\sigma}{\sqrt{n}}.$$

Найпоширеніші значення ймовірностей і відповідні коефіцієнти t наведені у табл. 7.2:

Таблиця 7.2

Величина ймовірності	0,95	0,954	0,997
Значення коефіцієнта t	1,96	2,0	3,0

Довірчий інтервал для генеральної частки p визначається як:

$$w - \Delta \leq p \leq w + \Delta \quad (7.12)$$

або

$$w - t \sqrt{\frac{w(1-w)}{n}} \leq p \leq w + t \sqrt{\frac{w(1-w)}{n}},$$

де p – частка генеральної сукупності.

Приклад 7.1. З метою визначення середньої ціни певного товару методом простої випадкової вибірки було обстежено 25 магазинів. У результаті було встановлено, що середня ціна становить 30 грн, при середньому квадратичному відхиленні 4 грн.

З ймовірністю 0,954 потрібно визначити межі, в яких знаходиться середня ціна в генеральній сукупності.

Розв'язання. За умовою задачі:

Вибіркова середня $\tilde{x} = 30$ грн та $\sigma = 4$ грн, обсяг вибірки $n = 25$, $P = 0,954 \Rightarrow t = 2,0$.

Для визначення меж довірчого інтервалу спочатку потрібно розрахувати середню помилку:

$$\mu_{\tilde{x}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{4^2}{25}} = \sqrt{\frac{16}{25}} = 0,8 \text{ (грн)}.$$

Для заданої ймовірності гранична помилка вибірки $\Delta\tilde{x}$:

$$\Delta_{\tilde{x}} = t\mu_{\tilde{x}} = 2 \cdot 0,8 = 1,6 \text{ (грн)}.$$

Згідно з формулою 7.14 середня ціна в генеральній сукупності буде знаходитися в межах:

$$\tilde{x} - \Delta\tilde{x} \leq \bar{x} \leq \tilde{x} + \Delta\tilde{x}.$$

Звідси довірчий інтервал визначатиметься як:

$$\begin{aligned} 30 - 1,6 &\leq \bar{x} \leq 30 + 1,6 \\ 28,4 &\leq \bar{x} \leq 31,6, \end{aligned}$$

тобто з ймовірністю 0,954 можна стверджувати, що середня ціна в генеральній сукупності (у всіх магазинах, де продається цей товар) буде в межах від 28,4 до 30,6 грн.

Приклад 7.2. Під час дослідження 100 зразків виробів, відібраних із партії у випадковому порядку, 20 виявились невідповідними. З ймовірністю 0,954 визначити межі, в яких знаходиться частка невідповідної продукції в генеральній сукупності.

Розв'язання. За умовою задачі: обсяг вибірки $n = 100$, кількість одиниць, що відповідають певній умові (невідповідні) $m = 20$, для прийнятої ймовірності

$$P = 0,954 \Rightarrow t = 2,0$$

Розділ 7. ВИБІРКОВИЙ МЕТОД

Частка нестандартної продукції у вибірці становить:

$$w = \frac{m}{n} = \frac{20}{200} = 0,2.$$

Гранична помилка частки дорівнює:

$$w = t \sqrt{\frac{w(1-w)}{n}} = 2 \sqrt{\frac{0,2(1-0,2)}{100}} = 0,08.$$

Згідно з формулою (7.3) частка нестандартної продукції у генеральній сукупності буде знаходитися в межах:

$$w - \Delta \leq p \leq w + \Delta.$$

Підставивши отримані значення, знайдемо межі довірчого інтервалу:

$$0,2 - 0,08 \leq p \leq 0,2 + 0,08$$

$$0,12 \leq p \leq 0,28,$$

тобто з ймовірністю 0,954 частка невідповідної продукції в генеральній сукупності буде знаходитися в межах 12–28%.

Приклад 7.3. При проведенні 5% механічної вибірки дістали дані про відсоток залучених кредитів підприємствами:

Таблиця 7.3

Ставка по кредиту, %	20–22	22–24	24–26	26–28	28–30	Всього
Кількість підприємств	18	22	32	16	12	100

З ймовірністю 0,997 визначити:

- 1) межі середньої ставки по кредиту;
- 2) межі питомої ваги підприємств, що залучили кредити із ставкою понад 26%.

Розв'язання.

1. За даними вибірки середня ставка по кредиту $\tilde{x} = 24,64\%$, при дисперсії $\sigma^2 = 6,19$.

За умовою $n = 100$. Оскільки маємо 5% вибірку, то можна визначити кількість одиниць генеральної сукупності

$$N = \frac{100}{0,05} = 2000.$$

Тоді середня помилка за схемою безповторної вибірки

$$\mu = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{6,19}{100} \left(1 - \frac{100}{2000}\right)} = 0,056(\%).$$

Гранична помилка для середньої дорівнює:

$$\Delta = t \cdot \mu = 3 \cdot 0,056 = 0,168(\%).$$

Отже,

$$\bar{x} = \tilde{x} \pm \Delta = 24,64 \pm 0,168 \Rightarrow 24,472 \leq \bar{x} \leq 24,808.$$

2. Питома вага залучених кредитів із ставкою понад 26%:

$$w = \frac{16+12}{100} \cdot 100\% = 28\%.$$

Тоді середня помилка вибірки становить:

$$\mu = \sqrt{w \cdot \frac{1-w}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{0,28 \cdot \frac{1-0,28}{100} \left(1 - \frac{100}{2000}\right)} = 4\%$$

Гранична помилка частки дорівнює:

$$\Delta = t \cdot \mu = 3 \cdot 4 = 12\%$$

Довірчий інтервал визначатиметься як:

$$\begin{aligned} 0,28 - 0,12 &\leq \bar{p} \leq 0,28 + 0,12 \\ 0,16 &\leq \bar{p} \leq 0,40 \end{aligned}$$

Отже, частка підприємств, що залучили кредити із ставкою понад 26%, становитиме від 16 до 40%.

7.5. Визначення мінімально необхідного обсягу вибіркової сукупності

При організації проведення вибіркового спостереження важливе значення має правильне визначення необхідної чисельності вибірки, яка з відповідною ймовірністю забезпечить встановлену точність результатів спостереження.

Відповідно до теоретичних основ вибіркового спостереження, чисельність вибірки залежить від таких чинників:

1) від варіації досліджуваної ознаки. Чим більша варіація, тим більшою повинна бути чисельність вибірки, і навпаки;

2) від розміру можливої граничної помилки вибірки. Чим менший розмір можливої помилки, тим більша повинна бути чисельність вибірки. Так, якщо помилку потрібно зменшити в три рази, то чисельність вибірки збільшують у дев'ять разів;

3) від розміру ймовірності, з якою гарантуватимуть результати вибірки. Чим більша ймовірність, тим більша повинна бути чисельність вибірки;

4) від способу відбору одиниць у вибірку сукупність для обстеження.

Визначення необхідного обсягу вибірки будується на алгебраїчному перетворенні формул граничних похибок вибірок при різних способах відбору.

Мінімально необхідний обсяг вибірки для середньої при повторному відборі визначається з формули граничної похибки

$\Delta_x = t \sqrt{\frac{\sigma^2}{n}}$, в якій обидві частини цього рівняння підносять до квадрату $\Delta^2 = \frac{t^2 \sigma^2}{n}$, звідси **мінімально** необхідна чисельність вибірки дорівнює:

$$n = \frac{t^2 \sigma^2}{\Delta_x^2} \quad (7.13)$$

Аналогічно виводяться формули чисельності вибірки при визначенні часток ознаки (табл. 7.4).

Таблиця 7.4

Мінімально необхідна чисельність вибірки для різних способів формування вибіркової сукупності

Вид вибіркового спостереження	Повторний відбір	Безповторний відбір
Проста випадкова вибірка: при визначенні середньої ознаки	$n = \frac{t^2 \cdot \sigma^2}{\Delta^2}$	$n = \frac{t^2 \cdot \sigma^2 \cdot N}{\Delta^2 \cdot N + t^2 \cdot \sigma^2}$
при визначенні частки ознаки	$n = \frac{t^2 \cdot w \cdot (1-w)}{\Delta^2}$	$n = \frac{t^2 \cdot w \cdot (1-w) \cdot N}{\Delta^2 \cdot N + t^2 \cdot w \cdot (1-w)}$
Механічна вибірка	Та ж як і для простої вибірки	Та ж як і для простої вибірки

Для визначення обсягу вибірки n використовують оцінки дисперсій σ^2 аналогічних пробних обстежень.

Якщо такі обстеження відсутні, можна скористатися співвідношенням $\sigma = \frac{1}{6}(x_{\max} - x_{\min})$, а для частки взяти найбільше значення дисперсії:

$$\sigma^2 = 0,25.$$

Якщо в основу розрахунку n покласти відносну похибку вибірки V_{Δ} формули відповідно модифікуються:
для середньої:

$$n = \frac{t^2 V_x^2}{V_{\Delta}^2} \quad (7.14)$$

для частки:

$$n = \frac{t^2 (1-w)}{V_{\Delta}^2 \cdot w} \quad (7.15)$$

Приклад 7.4. З метою визначення середнього рівня задоволеності споживачів потрібно провести їх опитування. Скільки споживачів потрібно опитати, щоб з ймовірністю 0,954 гранична похибка не перевищувала 0,2 бали (за 10 бальною шкалою), якщо відомо, що дисперсія оцінок становить 0,5.

Розв'язання. З умови задачі маємо $t = 2$, гранична похибка $\Delta = 0,2$ бали, $\sigma^2 = 2,5$. Оскільки планується провести відбір як просту випадкову вибірку, скористаємося формулою 7.15:

$$n = \frac{t^2 \cdot \sigma^2}{\Delta^2} = \frac{2^2 \cdot 2,5}{0,2^2} = \frac{10}{0,04} = 250 \text{ (осіб)}$$

Отже, для отримання оцінок з потрібною точністю за обраного рівня достовірності потрібно опитати щонайменше 250 споживачів.

Приклад 7.5. На факультеті навчається 1000 студентів. З метою виявлення мотивації обраної спеціальності проводиться соціологічне дослідження. Якою повинна бути чисельність простої неповторної вибірки, щоб з ймовірністю 0,954 гранична похибка не перевищувала 10%?

Розв'язання. З умови задачі маємо $t = 2$, $N = 1000$, $\Delta = 0,1$. Оскільки дисперсія частки є невідомою, то приймаємо її рівною 0,25. Обсяг вибірки для неповторного відбору при визначенні частки становитиме:

$$n = \frac{t^2 W (1 - W) \times N}{\Delta^2 \times N + t^2 W (1 - W)} = \frac{4 \times 0,25 \times 1000}{0,01 \times 1000 + 4 \times 0,25} =$$
$$= \frac{1000}{11} = 91 \text{ (студент)}$$

7.6. Задачі для самостійної роботи

Задача 7.1. У результаті вибіркового дослідження встановлено, що стандартна похибка вибіркової середньої дорівнює $\mu_x = 2$. Якщо вибрано коефіцієнт довіри $t = 1,96$, то гранична похибка вибіркової середньої становить:

- а) 2; б) 1,96; в) 4; г) 3,92; д) 3,8416.

Задача 7.2. За даними вибіркового опитування 46% респондентів вважають рекламу основним джерелом інформації про товарний ринок. Стандартна похибка вибірки цього показника – 2,5%. З ймовірністю 0,954 можна стверджувати, що рекламою користуються:

- а) не менше 43,5% споживачів;
б) не більше 48,5%;
в) не менше 41 і не більше 51%;
г) не менше 51%.

Задача 7.3. За даними вибіркового обстеження 64 підприємств малого бізнесу в галузі інформаційно-обчислювального обслуговування, середня окупність витрат становить 37% при середньоквадратичному відхиленні 7,2%. Визначте граничну похибку вибірки середнього рівня окупності з імовірністю 0,954.

Задача 7.4. За даними вибіркового обстеження 25 фірм (19%-ний відбір) середня тривалість обороту дебіторської заборгованості 72 дні при середньоквадратичному відхиленні

10 днів. Визначити граничну похибку вибірки для середньої тривалості обороту з ймовірністю 0,954.

Задача 7.5. За результатами опитування домогосподарок (у плані маркетингового дослідження) 135 із 225 віддають перевагу споживанню чаю у разових упаковках. З ймовірністю 0,954 визначити похибку вибірки для частини домогосподарок, які є потенційними покупцями чаю в разових упаковках, %.

Задача 7.6. За результатами опитування 100 підприємців (5%-ний відбір) 20 оцінюють умови господарювання як несприятливі. З ймовірністю 0,954 визначити похибку вибірки для частини респондентів, не задоволених умовами господарювання.

Задача 7.7. Якою має бути чисельність вибірки при визначенні середнього вкладу в Ощадний банк, щоб з ймовірністю 0,954 гранична похибка вибірки не перевищила 5 гр. од.? Орієнтовна дисперсія вкладів 1200.

Задача 7.8. За даними попередніх вибірових обстежень частка нестандартної плодоовочевої продукції становить 0,10. Скільки треба провести перевірок, щоб похибка вибірки з ймовірністю 0,954 не перевищила 4%?

Задача 7.9. Скільки треба опитати респондентів, оцінюючи якість готельного обслуговування (задовольняє, не задовольняє), щоб гранична похибка вибірки часток з ймовірністю 0,954 при цьому не перевищила 5%?

Задача 7.10. За результатами контрольної перевірки податковими службами 400 бізнесових структур 140 з числа перевірених у податкових деклараціях за минулий рік вказали не всі доходи, які підлягають оподаткуванню.

СТАТИСТИКА

Визначити частку бізнесових структур, які приховують частину доходів від сплати податків, та довірчі межі частки з імовірністю 0,954.

Чи погоджуються вибірккові дані з твердженням, що 40% бізнесових структур не сплачують податків у повному обсязі?

Задача 7.11. За даними 5%-ного вибіркового обстеження, працівники за терміном роботи розподіляються так:

Таблиця 7.5

Термін роботи, років	До 4	4–8	8–12	12 і більше	Разом
Кількість працівників	25	40	20	15	100

Визначити:

а) середній термін служби верстатів та довірчий інтервал для середньої з імовірністю 0,954;

б) з тією ж ймовірністю визначити граничну похибку та довірчий інтервал частки працівників, що мають термін служби понад 12 років.

Задача 7.12. За даними опитування із 250 респондентів основними джерелами інформації про ринок цінних паперів вважають:

радіо та телебачення	170
газети та журнали	90

Для кожного джерела інформації визначити його частку та відносну похибку вибірки з імовірністю 0,954. Порівняти похибки вибірки.

Задача 7.13. З метою визначення потенціалу споживчого ринку планується анкетування населення (одна квартира \Leftrightarrow одна анкета). Визначити мінімально необхідний обсяг вибірки, щоб гранична похибка вибірки (з ймовірністю 0,954) для

середньомісячного розміру покупки не перевищувала 5 гр. од. За даними пробних обстежень дисперсія середньомісячного розміру покупок становить 1875.

7.7. Питання для самоконтролю

1. Яке спостереження називається вибірковим?
2. У чому полягає сутність вибіркового спостереження? Які його переваги порівняно з іншими видами спостереження?
3. Що означає репрезентативність вибірки? За яких умов вибірка репрезентативна?
4. Чому принцип випадковості добору є визначальним у процесі формування вибіркової сукупності? Які способи добору забезпечують додержання цього принципу?
5. Чому при вибіркового спостереженні є неминучими помилки та як вони класифікуються?
6. Чим випадкова похибка репрезентативності відрізняється від систематичної? Чи можна її уникнути?
7. Як визначити розмір похибки вибірки? Чим гранична похибка вибірки відрізняється від стандартної (середньої)?
8. У чому полягає різниця між повторною та безповторною вибіркою?
9. Що показує коефіцієнт довіри?
10. Якими способами здійснюється розповсюдження результатів вибіркового спостереження на всю сукупність?
11. Навіщо та як обчислюються граничні статистичні помилки вибірки?
12. Як визначити мінімально необхідний обсяг вибірки?

Розділ 8

СТАТИСТИЧНА ПЕРЕВІРКА ГІПОТЕЗ

План

- 8.1. Статистичні гіпотези. Принципи їх перевірки та формулювання статистичного висновку.
- 8.2. Перевірка гіпотез про дисперсії.
- 8.3. Перевірка гіпотез про середні значення.
- 8.4. Перевірка гіпотез про частки одиниць сукупності.
- 8.5. Задачі для самостійної роботи.
- 8.6. Питання для самоконтролю.

8.1. Статистичні гіпотези. Принципи їх перевірки та формулювання статистичного висновку

Вибіркове оцінювання дозволяє зробити висновок про генеральну сукупність тільки із певною ймовірністю. Внаслідок цього висуваються деякі припущення щодо параметрів генеральної сукупності.

Статистичними гіпотезами називаються припущення щодо розподілів досліджуваних ознак або їх параметри, що перевіряються за вибірковими даними.

Статистичні гіпотези позначають літерами H : (*Hypothesis*).

Статистичні гіпотези формулюються парами.

Гіпотезу, яку перевіряють, називають нульовою, або основною гіпотезою і позначають H_0 .

Основна гіпотеза (нульова) – висувається про:

- відповідність розподілу ознаки певному теоретичному розподілу (наприклад, нормальному закону розподілу),
- відсутність істотних відмінностей між статистичними сукупностями,
- про відсутність залежності ознак тощо.

Вона може бути сформульована наприклад так:

H_0 : параметр сукупності = певному значенню,

що формально записується як

$$H_0: \bar{x} = a \quad (8.1)$$

Кожній нульовій гіпотезі H_0 висувається конкуруюча (альтернативна) гіпотеза $H_1(H_a)$, яка вступає у протиріччя з нею і може мати вид односторонньої чи двосторонньої нерівності

$$H_1: \bar{x} \neq a, H_1: \bar{x} > a, H_1: \bar{x} < a \quad (8.2)$$

Для перевірки гіпотез застосовують *статистичні критерії*.

Статистичний критерій (тест) – це правило, що забезпечує статистично обґрунтоване прийняття істинної і відхилення помилкової гіпотези. Статистичні критерії будуються на основі деякої величини – функції від результатів вибірових спостережень.

Статистичний критерій, який умовно позначають через K , є випадковою величиною, точний або наближений закон розподілу ймовірностей якої заздалегідь відомий.

Емпіричним (спостережуваним) значенням $K_{\text{спост}}$ називають значення критерію, обчислене за вибіркою.

У розподілі статистичного критерію виокремлюють спеціальні області, ймовірність потрапляння в які $K_{\text{спост}}$ за правильної нульової гіпотези є малою. Такі області називають критичними областями критерію.

Областю прийняття гіпотези (допустимих значень) називають сукупність значень критерію, за яких приймають нульову гіпотезу H_0 .

Точку або кілька точок, що поділяють множину всіх можливих значень критерію на підмножини прийняття і відхилення основної гіпотези, називають критичними і позначають через $K_{\text{кр}}$. (рис. 8.1).

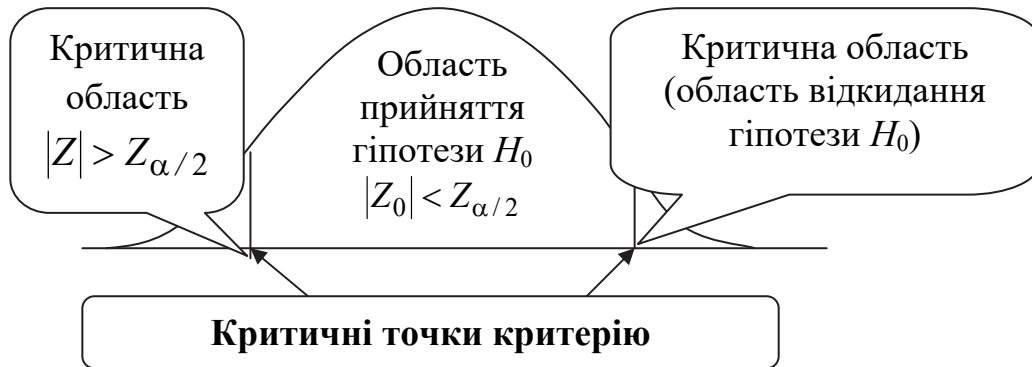


Рис. 8.1. Схема критерію перевірки статистичної гіпотези

Залежно від виду висунутої гіпотези критичні області бувають одно- та двосторонніми. Існують три види критичних областей: двосторонні критичні області виникають за умови гіпотези про будь-яку розбіжність, а односторонні – якщо висувається гіпотеза, що значення будуть тільки більше або менше від заданої величини.

Загальний алгоритм перевірки статистичних гіпотез проілюстровано на рис. 8.2.

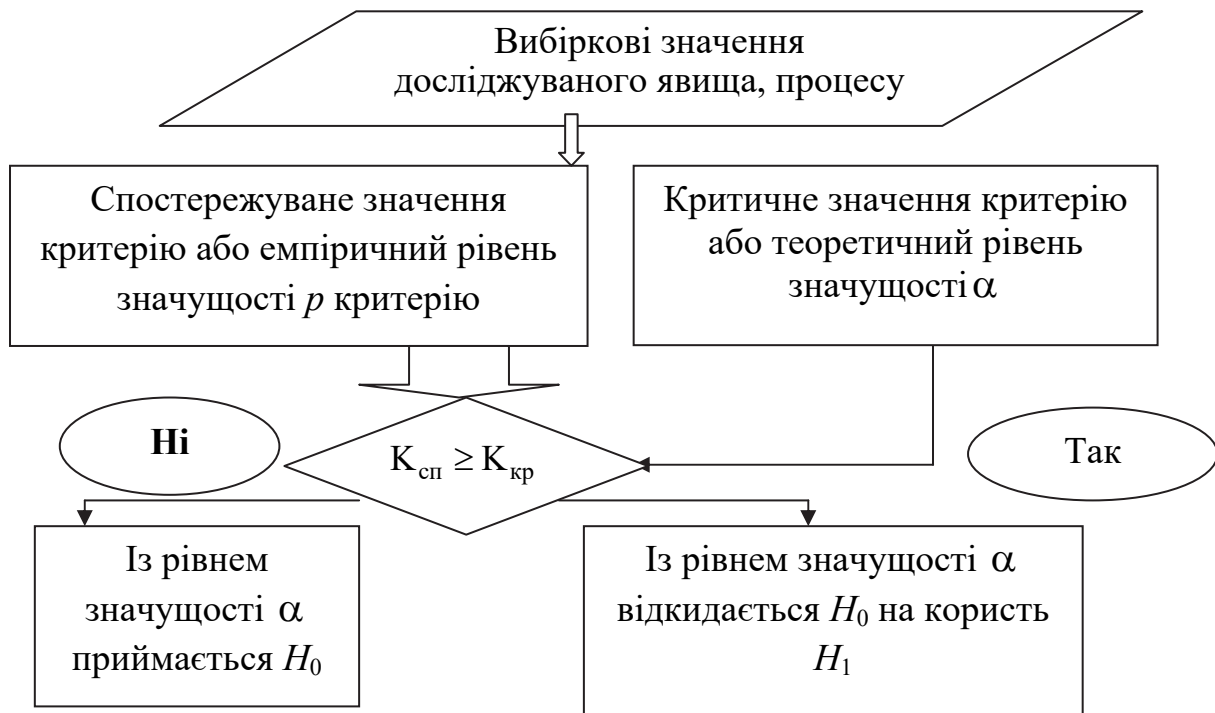


Рис. 8.2. Загальний алгоритм перевірки статистичних гіпотез

Загальний принцип перевірки статистичних гіпотез такий: якщо спостережуване значення критерію належить критичній області, то нульову гіпотезу H_0 відкидають і приймають альтернативну гіпотезу H_1 .

Якщо значення критерію, отримане на основі вибіркового даних, належить області прийняття гіпотези, то приймають нульову гіпотезу H_0 .

Унаслідок випадковості відбору одиниць у вибірковій сукупності, за якими перевіряються гіпотези, під час перевірки гіпотез можуть виникати помилки.

Помилка, яка полягає в тому, що гіпотеза H_0 відхиляється, коли насправді вона справедлива, називається *помилкою першого роду*.

Ймовірність помилки першого роду називають *рівнем значущості (significance level)* і позначають α .

Помилка, яка полягає в тому, що гіпотеза H_0 приймається, коли насправді вона несправедлива, називається *помилкою другого роду*. Її ймовірність позначається як β . Тобто прийнято гіпотезу H_0 , хоча насправді правильною є альтернативна гіпотеза H_1 .

Потужність критерію – ймовірність того, що не допускається помилка другого роду, тобто $1 - \beta$. Очевидно, що чим вища потужність критерію, що використовується, при заданому значенні α , тим краще він розрізняє гіпотези H_0 та H_1 . Особливо важливо, щоб критерій, що використовується добре розрізняв близькі альтернативи.

Найчастіше обираються рівні значущості, що дорівнюють значенням 0,1; 0,05; 0,01 і 0,001. Вибір рівня значущості визначається тими наслідками, до яких може призвести помилкове відхилення гіпотези. Якщо, наприклад, $\alpha = 0,01$, то це означає, що в одному випадку зі 100 є ризик припуститися помилки першого роду (відхилити гіпотезу H_0 , коли вона справедлива).

У соціально-економічних дослідженнях зазвичай обирають рівень значущості 0,05, у технологічних – рівень значущості може бути меншим, наприклад 0,001.

У результаті перевірки гіпотез можуть бути отримані висновки, як показано в табл. 8.1:

Таблиця 8.1

Матриця результатів в оцінці статистичних гіпотез

Результат	Гіпотези	
	H_0 приймається	H_0 відхиляється
Підтверджується H_0	Правильне рішення	Помилка з рівнем α
Підтверджується H_1	Помилка з рівнем β	Правильне рішення

Критичні значення критеріїв, табульовані для заданих рівнів значущості і степенів вільності, визначаються з таблиць відповідного розподілу, тому ще мають назву табличних значень критерію.

Степінь вільності – статистичний термін, що означає різницю між кількістю спостережень і числом параметрів ознаки, що оцінюються за цими спостереженнями.

За **одностороннього** критерію береться рівень помилки α , а за **двостороннього** – $\alpha / 2$.

Види статистичних гіпотез щодо параметрів сукупностей та основні критерії їх перевірки наведені в табл. 8.2.

Таблиця 8.2

Види критеріїв для перевірки статистичних гіпотез

Вид критеріїв	Вид гіпотези	1 вибірка	2 вибірки	Більше 2-х вибірок
Параметричні	Про рівність середніх	z -критерій t -критерій	t -критерій	Дисперсійний аналіз (ANOVA)

Закінчення табл. 8.2

Вид критеріїв	Вид гіпотези	1 вибірка	2 вибірки	Більше 2-х вибірок
	Про рівність дисперсій	<i>Хі-квадрат</i> критерій	<i>F</i> -критерій	Критерій Кохрена Критерій Бартлета
Непараметричні	Про рівність центрів розподілу (медіан)	<i>G</i> -критерій знаків <i>W</i> -критерій Уилкоксона	<i>U</i> -статистика Манна - Уїтні	<i>H</i> -критерій Краскела - Уоліса Критерій медіан Критерій Фрідмана
	Про рівність дисперсій	-	Критерій Зігеля - Тьюкі	Критерій Левіна

8.2. Перевірка гіпотез про дисперсії

Порівняння дисперсії з очікуваним значенням

На практиці часто виникають ситуації, коли потрібно перевірити стабільність фінансово-економічних процесів. Для цього на основі вибіркової дисперсії перевіряється гіпотеза про рівність дисперсії досліджуваної сукупності певному значенню.

Припустімо, потрібно перевірити гіпотезу про рівність дисперсії певному значенню, тобто

$$H_0: \sigma^2 = \sigma_0^2, \quad (8.3)$$

$$H_1: \sigma^2 \neq \sigma_0^2, \quad (8.4)$$

де σ^2 – невідома генеральна дисперсія;

σ_0^2 – очікуване значення генеральної дисперсії.

Припустімо, що з генеральної сукупності взята вибірка, де n – обсяг вибірки і s^2 – незміщена вибіркова дисперсія для цього обсягу.

Тоді для перевірки гіпотези обчислюється значення критерію хі-квадрат (χ^2) Пірсона:

$$\chi_{\text{спост}}^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad (8.5)$$

Для визначення критичного значення цього критерію з таблиці критичних точок χ^2 -розподілу для обраного рівня значущості α і числа степенів вільності $k = n - 1$ потрібно вибрати критичну точку $\lambda_{кр}^2(\alpha, k)$.

Альтернативно $\lambda_{кр}^2$ критерій можна знайти за допомогою статистичної функції Excel ХИ2.ОБР (дод. 3).

Якщо $\chi_{\text{спост}}^2 > \chi_{\alpha/2, k}^2$, то нульова гіпотеза H_0 відкидається на користь альтернативної.

У разі $\chi_{\text{спост}}^2 < \chi_{1-\alpha/2, k}^2$ – приймають нульову гіпотезу із заданим рівнем значущості.

Приклад 8.1. Перевірити гіпотезу про рівність дисперсії досліджуваної сукупності заданому значенню.

На основі хронометражу часу, потрібного на встановлення певного елемента у виріб різними працівниками, визначено, що дисперсія дорівнює $\sigma_0^2 = 2$ (хв). Результати хронометражу часу роботи нового працівника за 20 спостереженнями є такими:

$$\text{Середній час дорівнює: } \tilde{x} = \frac{1202}{20} = 60,1.$$

$$\text{Дисперсія становить: } s^2 = 3,98$$

Чи можна вважати, що новий працівник працює ритмічно?

Розв’язання. Сформулюємо завдання у термінах статистичних гіпотез. Метою є визначення постійності роботи робітника, тобто необхідно порівняти варіацію часу його роботи із заданим значенням. Потрібно перевірити гіпотезу, що дисперсія часу, який новий працівник витрачає на встановлення елемента, істотно не відрізняється від дисперсії часу інших робітників (очікуваного значення дисперсії).

Обравши за рівень значущості $\alpha = 0,05$, визначаємо дисперсію витрат часу нового робітника.

Розрахункове значення критерію становитиме:

$$\chi_{\text{спост}}^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(20-1)3,98}{2} = 37,81.$$

Оскільки альтернативна гіпотеза виражається рівністю, то потрібно встановити ліву і праву критичні точки.

У даному випадку для 19 степенів вільності і обраного рівня значущості ліва критична точка $\chi_{0,925;19}^2 = 8,9$, а права критична точка $\chi_{0,025;19}^2 = 32,85$.

Спостережуване значення перевищує праву критичну точку, тому є підставою відкинути гіпотезу про рівність дисперсії очікуваному значенню і зробити висновок, що новий працівник працює неритмічно.

Порівняння дисперсій двох сукупностей

Коли виникає потреба порівняти однорідність двох сукупностей (варіацію двох економічних процесів, два методи вимірювання), доцільно скористатися перевіркою гіпотези про рівність дисперсій двох генеральних сукупностей:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad (8.6)$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \quad (8.7)$$

де σ_1^2 – дисперсія першої, а

σ_2^2 – другої генеральної сукупності.

Нехай з двох генеральних сукупностей отримано вибірки обсягом n_1 і n_2 , для кожної з яких визначено скориговану (незміщену) вибірккову дисперсію S_1^2 та S_2^2 , відповідно.

Для перевірки гіпотези розраховують значення F -критерію як відношення більшої дисперсії до меншої:

$$F = \frac{S_1^2}{S_2^2}, \quad (8.8)$$

де S_1^2 – більша, а

S_2^2 – менша з двох вибіркових дисперсій.

За таблицею критичних точок розподілу Фішера – Снедекора (дод. 5) для рівня значущості α і числа степенів вільності $k_1 = n_1 - 1$, $k_2 = n_2 - 1$ (де значення k_1 береться для більшої дисперсії) потрібно знайти критичну точку $F_{кр}(\alpha, k_1, k_2)$ і порівняти спостережуване і критичне значення критерію:

якщо $F > F_{кр}$ – відкидаємо гіпотезу H_0 ;

якщо $F < F_{кр}$ – приймаємо гіпотезу H_0 .

Для користування таблицями F -розподілу спочатку обирається таблиця для рівня значущості, потім на перетині стовпця (число степенів вільності $K_1 = n_1 - 1$) і рядочка (число степенів вільності $K_2 = n_2 - 1$) визначається критичне значення критерію.

Альтернативно $F_{кр}$ можна знайти за допомогою статистичної функції Excel ФРАСПОБР.

Приклад 8.2. Порівняння дисперсій двох генеральних сукупностей.

Інвестор бажає порівняти ризик, пов'язаний з акціями двох компаній. Одним з методів оцінювання ризику певної акції є вимірювання варіації щоденних змін цін. Інвестор отримав випадкову вибірку із змін цін між днями акції 1 та акції 2. Інвестор може порівняти ризик, пов'язаний з акціями

цих компаній за допомогою перевірки гіпотези про рівність їх варіацій. Доцільним є прийняти рівень значущості 0.10.

Таблиця 8.4

Зміна цін на акції відносно попереднього дня

День	Зміна цін 1	Зміна цін 2
1	1,86	0,87
2	1,80	1,33
3	1,03	-0,27
4	0,16	-0,20
5	-0,73	0,25
6	0,90	0,00
7	0,09	0,09
8	0,19	-0,71
9	-0,42	-0,33
10	0,56	0,12
11	1,24	0,43
12	-1,16	-0,23
13	0,37	0,70
14	-0,52	-0,24
15	-0,09	-0,59
16	1,07	0,24
17	-0,88	0,66
18	0,44	-0,54
19	-0,21	0,55
20	0,84	0,08

Розв'язання. За умовою обсяги вибірок $n_1 = n_2 = 20$.

Знайдемо дисперсії для щоденної зміни цін першої та другої акції

$$S_1^2 = 0,8487, S_2^2 = 0,5291.$$

Для перевірки статистичної гіпотези про рівність дисперсій двох сукупностей, F -критерій

Знаходимо спостережуване значення F -критерію як відношення більшої дисперсії до меншої:

$$F = \frac{S_1^2}{S_2^2} = \frac{0,8487}{0,5291} = 1,604.$$

За умовою конкуруюча гіпотеза має вид нерівності, тому критична область є двосторонньою, для визначення критичного значення критерію потрібно брати рівень значущості удвічі менший, ніж заданий.

Аналіз результатів: Критичне значення розподілу Фішера для рівня значущості $\alpha/2 = 0,05$ і числа степенів вільності $k_1 = 20 - 1 = 19$ і $k_2 = 20 - 1 = 19$ становить

$$F(0,05; 19; 19) = 2,1555.$$

Оскільки $F < F_{кр}$, то немає підстав відкидати гіпотезу про рівність дисперсій.

Отже, за прийнятого рівня значущості можна вважати, що обидві акції мають однакову міру ризику.

8.3. Перевірка гіпотез щодо середніх значень

Порівняння середньої з очікуваним значенням

Часто постає питання перевірки того, чи відповідає середнє значення сукупності певному значенню, наприклад, виготовлена продукція має відповідати вимогам стандартів, виробничим специфікаціям.

Кількісною характеристикою міри такої відповідності є результати перевірки гіпотези про рівність середнього значення досліджуваної сукупності певному значенню.

Загальна ідея полягає у порівнянні різниці між значеннями відносно середньої помилки вибірки:

$$\left(\frac{\text{вибіркова середня} - \text{середня сукупності}}{\text{стандартна помилка}} \right).$$

У тому разі, якщо дисперсія генеральної сукупності відома, для перевірки гіпотези про рівність середньої величини очікуваному генеральному середньому:

$$H_1 : \mu \neq \mu_0 \quad (8.9)$$

$$H_0 : \mu = \mu_0$$

де μ – невідоме середнє із сукупності з відомою дисперсією;

μ_0 – очікуване значення середнього цієї сукупності, потрібно обчислити спостережуване значення Z -критерію за формулою:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad (8.10)$$

Z -критерій є випадковою величиною, що підпорядковується стандартному нормальному розподілу.

Якщо конкуруючою гіпотезою є нерівність, то праву критичну точку $z_{кр}$ можна визначити з умови:

$$\Phi(z) = \frac{1 - \alpha}{2}, \quad (8.11)$$

де $\Phi(z)$ є функцією Лапласа (інтегралом ймовірностей), що пов'язана з функцією стандартного нормального розподілу $F(z)$ (дод. 2) співвідношенням $\Phi(z) = F(z) - 1/2$.

Ліва й права критичні точки для одного й того ж рівня значущості пов'язані умовою $z_{кр} = -z_{кр}$.

Альтернативно критичне значення Z -критерію можна знайти за допомогою статистичної функції Excel НОРМ.СТ.ОБР.

Якщо $|Z| > u_{кр}$ – відкидаємо гіпотезу H_0 ,
в разі $|Z| < u_{кр}$ – приймаємо H_0 за визначеного рівня значущості α .

Приклад 8.3. Перевірити гіпотезу про відповідність середнього значення досліджуваної величини заданому значенню.

Відомо, що для безпечної дії ліків вміст певної речовини в одній таблетці має дорівнювати 0,5 мг. Вибіркове дослідження 120 виготовлених на новому обладнанні таблеток показало, що середній вміст цієї речовини в одній таблетці дорівнює 0,51 мг. При рівні значущості $\alpha = 0,01$ потрібно визначити, чи відповідає нове обладнання встановленим вимогам. Відомо, що вміст речовини має нормальний розподіл із середнім квадратичним відхиленням $\sigma = 0,11$ мг.

Розв'язання. Перевіримо гіпотезу про рівність середньої ваги речовини середньому $\mu_0 = 0,5$ мг.

За умовою $\bar{x} = 0,51$, $\mu_0 = 0,5$, $\sigma = 0,11$.

За відомими даними знаходимо спостережуване значення критерію:

$$Z = \frac{(\bar{x} - \mu_0) \sqrt{n}}{\sigma} = \frac{(0,51 - 0,50) \sqrt{120}}{0,11} = 0,995,$$

$$\Phi_{(z)} = \frac{1 - \alpha}{2} = \frac{1 - 0,01}{2} = 0,495.$$

Для визначення величини $z_{\text{кр}}$ можна використати функцію НОРМСТОБР електронних таблиць MS Excel або таблицю стандартного нормального розподілу, узявши як значення аргументу (функції) величину $(0,5 + 0,495 = 0,995)$.

Після цього знаходимо, що критичне значення $z_{\text{кр}} = 2,58$.

Оскільки спостережуване значення критерію не перевищує критичне, то за такого рівня значущості гіпотеза приймається, і можна зробити висновок, що вміст речовини відповідає встановленим вимогам.

Порівняння середньої з очікуваним значенням за невідомої дисперсії

У тому разі, якщо дисперсія генеральної сукупності не відома, для перевірки гіпотези про рівність генеральної середньої очікуваному значенню, користуються t -критерієм виду:

$$t_0 = \frac{(\bar{x} - \mu)\sqrt{n}}{S} \quad (8.12)$$

Розподіл критерію відповідає t -розподілу Стюдента з $n-1$ степенями вільності.

Основна гіпотеза H_0 відкидається якщо:

$$|t_0| > t_{\alpha/2, n-1}, \quad (8.13)$$

де $t_{\alpha/2, n-1}$ – критичне значення t -розподілу Стюдента з $n-1$ степенями вільності.

Якщо гіпотеза має вид односторонньої нерівності, тобто середнє очікується більшим за певне значення $H_1: \mu \geq \mu_0$, то береться рівень значущості α .

У разі альтернативна гіпотеза має вигляд нерівності, то критична область є двосторонньою і береться рівень значущості $\alpha/2$.

Критичне значення визначається з таблиць t -розподілу Стюдента (дод. 4) або за функцією EXCEL СТЮДРАСПОБР.

Приклад 8.4. Менеджер з туристичного напрямку бажає перевірити гіпотезу, що середня ціна за номер в певному регіоні дорівнює 120 \$. На основі запиту 12 готелів було встановлено, що середня ціна за номер становить $\bar{x} = 112,85$ \$, при середньому квадратичному відхиленні $S = 20,80$ \$. Рівень значущості обрано на рівні 5 %.

Розв'язання. За умовою відомо, що $\bar{x} = 112,85$ \$, $S = 20,80$ \$ і $n=12$.

Оскільки дисперсія цін невідома, було використано t -критерій і розраховано його значення:

$$t_0 = \frac{(\bar{x} - \mu)\sqrt{n}}{S} = \frac{(112,85 - 120)\sqrt{12}}{20,8} = -1,189$$

Визначене критичне значення $t_{кр} = -2,2010$.

Оскільки розраховане значення за абсолютною величиною є меншим за критичне, немає підстав відхиляти основну гіпотезу.

Це означає, що середня ціна за номер істотно не відрізняється від очікуваного рівня у 120 \$.

Порівняння середніх значень двох сукупностей

Порівняння середніх двох сукупностей, дисперсії яких є відомими

Нерідко виникають ситуації, коли потрібно порівняти між собою середні показники двох сукупностей. Для здійснення такого порівняння доцільно перевірити гіпотезу про рівність середніх двох сукупностей. Порівняння здійснюється на основі визначених з вибірки значень середніх. При цьому вибіркові середні порівнюються, одне з одним.

Для перевірки гіпотези про рівність середніх двох сукупностей:

$$H_0 : \mu_1 = \mu_2 \quad (8.14)$$

$$H_1 : \mu_1 \neq \mu_2 \quad (8.15)$$

у разі, дисперсії сукупностей відомі, користуються Z -критерієм:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \quad (8.16)$$

де \bar{x}_1 – вибіркове середнє першої сукупності;

\bar{x}_2 – середнє другої сукупності,

σ_1^2 і σ_2^2 – відомі дисперсії сукупностей.

Нульова гіпотеза H_0 відкидається, якщо абсолютне значення критерію буде більшим за табличне значення $Z_{кр}$ функції стандартного нормального розподілу:

$$|Z| > Z_{кр} \quad (8.17)$$

Приклад 8.5. За результатами спостереження упродовж 25 і 28 днів було встановлено, що середній щоденний обсяг продажу товару марки А становить 50 тис. грн, а марки Б – 60 тис. грн при дисперсіях 900 і 1600 відповідно. За рівня значущості 5 % менеджер бажає визначити, чи істотно відрізняється середній обсяг продажу цих марок.

Розв'язання. Товари не є взаємозалежними, тому вибірки розглядаються як незалежні. Дисперсії можуть вважатися відомими, тому можливо скористатися Z -критерієм.:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{50 - 60}{\sqrt{\frac{900}{25} + \frac{1600}{28}}} = -1.04$$

Беручи рівень значущості, що дорівнює 0,05 (5%), отримаємо для визначення критичної точки умову:

$$\Phi_{(z)} = \frac{1 - \alpha}{2} = \frac{1 - 0,05}{2} = 0,475.$$

Для цієї ймовірності значення функції стандартного нормального розподілу становить $z_{кр} = 1,96$.

Оскільки розрахункове значення критерію не перевищує критичне значення, то за прийнятого рівня значущості немає підстав відкидати нульову гіпотезу на користь альтернативної.

Робимо висновок, що середні істотно не відмінні, отже, середні обсяги продажу товарів різних марок істотно не відрізняються.

Порівняння середніх двох сукупностей за невідомих дисперсій

На практиці варіація сукупностей найчастіше невідома, тому користуються критеріями на основі розподілу Стьюдента.

При цьому порівняння середніх відрізняється залежно від того, однакові вибіркові дисперсії досліджуваних сукупностей чи ні.

Тому перед тим, як порівняти дві середні величини, потрібно перевірити гіпотезу про рівність дисперсій. Коли гіпотеза про рівність дисперсій вибіркових сукупностей підтверджується, використовують один критерій перевірки, якщо ні – інший.

Порівняння середніх двох сукупностей, дисперсії яких рівні

У разі дисперсії вважаються рівними, для перевірки гіпотези про рівність середніх двох сукупностей користуються критерієм виду:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_y^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \quad (8.18)$$

де n_1, n_2 – обсяг першої та другої вибіркової сукупності відповідно;

S_y^2 – узагальнена дисперсія.

Узагальнену дисперсію S_y^2 визначають як середню зважену з вибіркових дисперсій S_1^2 першої і S_2^2 другої вибіркових сукупностей:

$$S_y^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (8.19)$$

Розподіл критерію відповідає розподілу Стьюдента з двома степенями вільності $\nu = n_1 + n_2 - 2$ за рівня значущості α .

Нульову гіпотезу про рівність вибірових середніх H_0 відкидаємо, якщо абсолютне спостережуване значення критерію $|t|$ за абсолютним значенням перевищує критичне значення $t_{\alpha/2, \nu}$:

$$|t| > t_{\alpha/2, \nu}.$$

Приклад 8.6. На підприємстві встановили нову очисну систему. Перед встановленням на основі випадкової вибірки у 10 спостережень одержали таку інформацію про відсоток забруднення:

$$\bar{x}_1 = 9,85, S_1^2 = 1,73.$$

Після встановлення системи випадкова вибірка, яка становила 8 спостережень, дала результати:

$$\bar{x}_2 = 8,08, S_2^2 = 1,71.$$

Постає запитання стосовно того, чи справді використання нової очисної системи зменшує відсоток забруднення хімічними речовинами.

Розв'язання. Прийmemo рівень помилки першого виду, який нас влаштовує. Нехай $\alpha = 0,05$.

Спочатку потрібно перевірити гіпотезу про рівність вибірових дисперсій:

$$H_0: \sigma_1^2 = \sigma_2^2,$$

$$H_1: \sigma_1^2 \neq \sigma_2^2.$$

Скориставшись F -критерієм, визначаємо його розрахункове значення:

$$F = \frac{S_6^2}{S_M^2} = \frac{1,73}{1,71} = 1,011.$$

Знаходимо критичне значення критерію для числа степенів вільності $k_1 = 10 - 1 = 9$ у чисельнику та $k_2 = 8 - 1 = 7$ у знаменнику і обраного рівня значущості:

$$F_{0,05; 9; 7} = 3,68.$$

Оскільки розрахункове значення критерію 1,011 менше, ніж критичне, немає підстав відхилити гіпотезу про рівність дисперсій. Отже, вважаємо, що вони рівні.

Далі скористаємося критерієм для перевірки гіпотези про рівність середніх значень за умови, що дисперсії сукупностей рівні.

Якщо не буде підстав прийняти нульову гіпотезу, тобто середні істотно відмінні, підтвердиться ефективність використання нового очисного обладнання.

Узагальнена дисперсія дорівнює:

$$S_y^2 = \frac{1,73(10-1) + 1,71(8-1)}{10+8-2} = \frac{27,54}{16} = 1,721.$$

Отже, оскільки $S_y = \sqrt{1,721} = 1,311$, розрахункове значення критерію становитиме:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} S_y} = \frac{9,85 - 8,08}{\sqrt{\frac{1}{10} + \frac{1}{8}} 1,311} = \frac{1,77}{0,621} = 2,850.$$

Критичне значення t -критерію становить $t_{0,025;6} = 2,120$. Оскільки розрахункове значення критерію 2,85 перевищує критичне значення 2,12, то за прийнятого рівня значущості відкидаємо нульову гіпотезу на користь альтернативної.

Робимо висновок, що середні є значущо відмінні, отже, використання нового обладнання справді зменшує відсоток забруднювальних речовин.

*Порівняння середніх двох сукупностей,
дисперсії яких є відмінними*

Якщо дисперсії або їх відношення є невідомими і припущення про рівність дисперсій є необґрунтованим, то для перевірки гіпотези про рівність середніх двох сукупностей потрібно використати t -критерій виду:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (8.20)$$

що має t -розподіл Стьюдента із ν степенями вільності (округленим до найближчого цілочисельного значення):

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \quad (8.21)$$

Основна гіпотеза H_0 відкидається, якщо

$$|t| > t_{\alpha/2, \nu},$$

де $t_{\alpha/2, \nu}$ – табличне значення t -розподілу Стьюдента.

Приклад 8.7. Менеджер підприємства бажає перевірити, чи однаковий рівень цін його підприємства і конкурента. Випадковим чином було вибрано по 8 товарів однієї товарної групи і обрано припустимий рівень значущості 10%.

Результати обстеження, оцінки середніх та дисперсії наведені в табл. 8.5.

Таблиця 8.5

Результати спостереження цін

№ товару	Ціни фірми	Ціни конкурента
1	91	89
2	92	88
3	92	90
4	95	93
5	91	97
6	90	97
7	94	91
8	89	92
Середнє	91,750	92,125
Дисперсія	3,929	11,554

Розв’язання. Рівень ціни може вважатися однаковим, якщо однаковими будуть їх середні. Для визначення цього скористаємося перевіркою гіпотези про рівність середніх значень двох сукупностей. Першим етапом є порівняння дисперсій цін.

$$F = \frac{S_{\sigma}^2}{S_{\mu}^2} = \frac{11,554}{3,929} = 2,94.$$

Для прийнятого рівня значущості і числа степенів вільності значення $k_1 = 8 - 1 = 7$ у чисельнику та $k_2 = 8 - 1 = 7$ у знаменнику критичне значення критерію $F_{0,05; 7; 7} = 2.784$.

Оскільки розрахункове значення критерію 2,94 більше, ніж критичне, є підстави відхилити гіпотезу про рівність дисперсій. Отже, не вважаємо, що вони рівні.

Для перевірки гіпотези про рівність середніх значень двох сукупностей у разі різних дисперсій, застосовується критерій t -критерій Уелча. Знаходимо розрахункове значення t -критерію:

$$t_v = \frac{91,75 - 92,125}{\sqrt{\frac{3,929}{8} + \frac{11,554}{8}}} = \frac{-0,375}{1,391} = -0,269.$$

Визначаємо число степенів вільності для встановлення критичного значення критерію:

$$V = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1+1} + \frac{(S_2^2/n_2)^2}{n_2+1}} - 2 = \frac{\left(\frac{3,929}{8} + \frac{11,554}{8}\right)^2}{\frac{(3,929/8)^2}{8+1} + \frac{(11,554/8)^2}{8+1}} - 2 = 16,6$$

Порівнюємо розрахункове значення критерію з критичним.

У цьому випадку для рівня ймовірності 0,05 та числа степенів вільності 17 (після округлення) критичне значення критерію становить 2,12.

Таким чином, із заданою ймовірністю немає підстав відкидати гіпотезу про рівність середніх.

Отже, ціни фірми і конкурента можуть вважатися однаковими.

Двовибірковий t -критерій для залежних вибірок

Нехай дві генеральні сукупності X_1 та X_2 розподілені нормально, до того ж їх дисперсії рівні $\sigma^2(X_1) = \sigma^2(X_2)$, але невідомі. З них узято дві *пов'язані (залежні)* вибірки $x_{11}, \dots, x_{1n}, x_{21}, \dots, x_{2n}$ відповідно з першої та другої генеральної сукупності.

Пов'язаними (залежними) називаються такі сукупності, в яких варіанти пов'язані попарно, коли кожному елементу однієї з сукупностей відповідає певний елемент іншої сукуп-

ності. Це можуть бути результати спостережень над одними і тими ж об'єктами в різні моменти часу (наприклад, двома пробами одного і того ж тесту студента з певним інтервалом).

Нульова та альтернативна гіпотези формулюються аналогічно гіпотезі про рівність середніх.

Для обчислення емпіричного значення t -критерію в ситуації перевірки гіпотези про відмінності між двома залежними вибірками введемо величину

$$d_i = x_{1i} - x_{2i}, \quad i \sim \overline{1, n}.$$

Тоді середня різниця значень

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i,$$

а стандартне відхилення значень різниць d_i

$$\sigma_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}}$$

Таким чином, спостережуване значення статистичного критерію обчислюється за формулою

$$t_{cn} = \frac{\bar{d}}{\sigma_d / \sqrt{n}}.$$

Розподіл критерію відповідає наближеному розподілу Стьюдента, для якого кількість ступенів свободи розраховують як $k = n - 1$.

За таблицею критичних точок розподілу Стьюдента знаходимо двостороннє критичне значення $t(\alpha; k)$.

Якщо $|t_{cn}| < t_{кр}$ – немає підстав відхилити H_0 .

Якщо $|t_{cn}| > t_{кр}$ – H_0 відхиляється.

Приклад 8.8. З метою збільшення обсягів продажу було вирішено провести у 13 випадково визначених магазинах торгової мережі рекламну акцію. Середні щоденні обсяги продажу за період акції наведені у табл. 8.6. Менеджер бажає визначити чи мала рекламна акція ефект на обсяг продажу.

Таблиця 8.6

Середні щоденні обсяги продажу

№ магазину	До рекламної кампанії x_{1i}	Під час рекламної кампанії x_{2i}	Різниця $d_i = x_{1i} - x_{2i}$
1	65,3	67,2	1,9
2	54,7	59,4	4,7
3	81,3	80,1	-1,2
4	39,8	47,6	7,8
5	92,5	97,8	5,3
6	37,9	38,4	0,5
7	52,4	57,3	4,9
8	69,9	75,2	5,3
9	89	94,7	5,7
10	58,4	64,3	5,9
11	33	31,7	-1,3
12	41,7	49,3	7,6
13	53,6	54	0,4

Розв'язання. Позначимо дійсну різницю (для всієї мережі) між середніми μ_D . Тоді відсутність ефекту від рекламної кампанії полягатиме у відсутності різниці у середніх виторгах, або

$$H_0: \mu_D = 0$$

$$H_A: \mu_D > 0$$

Для цього числа спостережень і рівня значущості критичне значення $t_{кр} = 1.782$. У разі отримане розрахункове

значення критерію буде більшим за це значення, основна гіпотеза має бути відхилена.

Середнє значення різниць 3,6538, а їх стандартне відхилення 3,1855.

Розрахункове значення критерію становить:

$$t = \frac{\bar{d}}{\sigma_d / \sqrt{n}} = \frac{3,6538 - 0}{\frac{3,1855}{\sqrt{13}}} = 4,14$$

Оскільки розрахункове значення більше критичного, основна гіпотеза відхиляється за прийнятого рівня значущості.

Отже, рекламна кампанія мала істотний позитивний ефект на обсяг продажів.

8.4. Перевірка гіпотез про частки одиниць сукупності

Якщо сукупність характеризується за якісною ознакою, виникає питання про частки одиниць у сукупності з певною властивістю.

Припустимо, є n незалежних дослідів, ймовірність появи події p в яких є постійною, але невідомою. З цієї сукупності знайдена відносна частота кількості певних подій m у загальній кількості n одиниць.

Потрібно перевірити гіпотезу, що частота $p=m/n$ дорівнює очікуваній частоті p_0 , тобто

$$H_0: p = p_0,$$

$$H_1: p \neq p_0,$$

де p_0 – очікувана частота.

Для перевірки гіпотези, що частка одиниць у сукупності дорівнює певному рівню може бути здійснена на основі z -критерію, оскільки із зростанням обсягу вибірок біноміальний закон розподілу може бути наближено описаний за допомогою нормального закону розподілу.

Тестова статистика має вигляд:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}},$$

де \hat{p} – вибіркова оцінка частоти одиниць у сукупності.

Розподіл критерію відповідає стандартному нормальному закону розподілу $N(0,1)$ за умови, що $np_0 \geq 5$ та $n(1-p_0) \geq 5$.

У разі вибірок малого обсягу або необхідності отримання точного критичного значення доцільно скористатися біноміальним законом розподілу.

Приклад 8.9. Перевірка гіпотези про рівність частки одиниць у сукупності очікуваній величині.

Компанія з виробництва безалкогольних напоїв бажає вивести на ринок нову модифікацію напою. Компанія бажає бути впевненою в тому, що не менше ніж 70% її споживачів оберуть нову модифікацію напою. Новий напій було запропоновано 2 тис. осіб, і 1422 з них сказали, що він якісніший за попередній.

Чи може компанія зробити припущення про те, що принаймні 70 % всіх її споживачів оберуть новий продукт? Прийняти рівень значущості $\alpha = 0,05$.

Розв'язання. Сформулюємо задачу у термінах статистичної гіпотези. Потрібно перевірити гіпотезу, що частка одиниць (споживачів) дорівнює або більше 70%:

$$\begin{aligned}H_0 &: p \geq 0,7, \\H_1 &: p \leq 0,7.\end{aligned}$$

Для перевірки гіпотези про відповідність значення частки сукупності певному значенню, застосовується z -критерій

$$z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} = \frac{0,711 - 0,7}{\sqrt{\frac{0,7(1-0,7)}{2000}}} = \frac{0,011}{0,010} = 1,07.$$

Для прийнятого рівня значущості значення функції стандартного нормального розподілу становить $z_{кр} = 1,96$.

Отримане значення менше критичного, отже немає підстав для відхилення основної гіпотези.

Це дозволяє стверджувати із заданим рівнем ймовірності, що принаймні 70 % споживачів оберуть новий продукт.

Перевірка гіпотез про рівність часток одиниць у двох сукупностях

Для перевірки гіпотези про рівність часток у двох сукупностях:

$$\begin{aligned}H_0 &: p_1 = p_2, \\H_1 &: p_1 \neq p_2,\end{aligned}$$

з них отримують вибірки обсягами n_1 і n_2 відповідно і обчислюють вибіркові частки \hat{p}_1 і \hat{p}_2 , як відношення числа одиниць що відповідають умові m до обсягу одиниць у вибірці n .

На основі отриманих значень розраховують z -критерій, що має вигляд:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

де
$$\hat{p} = \frac{m_1 + m_2}{n_1 + n_2}.$$

якщо обсяги вибірок розрізняються не більше ніж на 25% або, якщо розміри вибірок значно відмінні

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)}}.$$

Приклад 8.10. Перевірити гіпотезу про рівність часток одиниць у двох сукупностях.

За результатами опитування 250 споживачів 120 були обізнані з торговою маркою. Після рекламної кампанії за результатами опитування 300 споживачів 150 були обізнані з досліджуваною торговою маркою. Приймаючи рівень помилки у 5%, чи можна припустити що рекламна кампанія підвищила рівень відомості марки?

Розв'язання. Результати опитування до рекламної кампанії розглядаються як перша сукупність, а після – як друга.

Для перевірки статистичної гіпотези про рівність часток двох сукупностей, застосовується z-критерій.

У результаті перевірки гіпотези отримано:

$$\hat{p} = \frac{120 + 150}{250 + 300} = \frac{270}{500} = 0,490.$$

$$z = \frac{|\hat{p}_1 - \hat{p}_2| - \frac{1}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{|0,5 - 0,48| - \frac{1}{2} \left(\frac{1}{250} + \frac{1}{300} \right)}{\sqrt{0,49(1-0,49) \left(\frac{1}{250} + \frac{1}{300} \right)}} = -0,467.$$

Для прийнятого рівня значущості значення функції стандартного нормального розподілу становить $z_{кр} = 1,96$.

Розраховане значення значно менше $z_{кр}$, таким чином, немає підстав відхилити гіпотезу про рівність часток сукупностей.

Це означає, що рекламна кампанія *не* підвищила рівень відомості марки.

Розглянуті процедури статистичної перевірки гіпотез ґрунтуються на підпорядкуванні розподілу значень ознак якості нормальному закону розподілу. За відсутності такого розподілу доцільно використовувати критерії перевірки гіпотез за *непараметричними* критеріями.

8.5. Задачі для самостійної роботи

Задача 8.1. Було перевірено 11 підприємств. Потрібно перевірити припущення дослідника, що середній рівень прибутковості становить 4,0 %. При перевірці отримані такі дані (y %):

4,21; 4.03; 3.99; 4.05; 3.89; 3.98; 4.01; 3.92; 4.23; 3.85; 4.20.

Зробити статистичний висновок за рівня значущості 10%.

Задача 8.2. На підприємстві вивчався вплив нової технології виготовлення ковбасних виробів на рівень їхньої якості. Результати дослідження такі: на 4 сортах, оброблених за традиційною технологією, рівень невідповідностей був 18, на 6 сортах, оброблених за новою технологією, – 16 випадків на 1000 виробів.

З ймовірністю 0,95 зробити висновок про ефективність нової технології.

Задача 8.3. З метою перевірки ефективності нового пристрою для вдосконалення технології обробки деталей працівників бригади розділили на 2 групи: на верстатах з

новим пристроєм працювало 4 робітники, за традиційною технологією обробки деталей – 6. Результати експерименту такі: у I групі середній виробіток за годину становить 84 деталі при $\sigma^2 = 10$, у II – 77 деталей при $\sigma^2 = 8$.

За допомогою t -критерію Стьюдента з рівнем істотності $\alpha = 0,05$ зробити висновок, чи дійсно нова технологія впливає на продуктивність праці.

Задача 8.4. Виробник стверджує, що строк придатності продукту становить 180 днів. За результатами обстеження 25 одиниць встановлено, що середній термін придатності становив 175 днів при дисперсії 36. Приймаючи рівень значущості у 5 % визначити, чи існують підстави сумніватися у даних виробника?

Задача 8.5. За результатами опитування 40 респондентів віком до 30 років 28 вважають, що держава має створити лише умови для плідної роботи, а людина повинна сама забезпечити свою сім'ю; із 60 респондентів старшого віку (понад 30 років) таку думку поділяють 24.

З достовірністю 0,954 перевірити, чи є випадковою розбіжність думок респондентів молодшого і старшого віку.

Задача 8.6. З кожної із 8 партій однотипних заготовок навмання бралися заготовки, які оброблялися на трьох верстатах різної модифікації. Кількість деталей, виготовлена верстатами, досліджувалася на стандартність. Результати досліджень подано в таблиці:

Фактор A (тип верстатів)	Кількість деталей, що відповідають стандарту
A_1	100; 86; 90; 89; 95; 22; 80; 79
A_2	99; 82; 98; 88; 100; 96; 98; 100

При рівні значущості $\alpha = 0,01$ перевірити вплив модифікації верстатів-автоматів на якість виготовлених деталей.

8.6. Питання для самоконтролю

1. Дати визначення нульової та альтернативної гіпотез.
2. Як формулюються та перевіряються статистичні гіпотези?
3. Які гіпотези називають параметричними?
4. Які гіпотези називають непараметричними?
5. Що називають простою та складною статистичними гіпотезами?
6. Що називається статистичним критерієм?
7. Що називається емпіричним значенням критерію?
8. Що таке область прийняття нульової гіпотези, критична область, критична точка.
9. Які Ви знаєте критичні області?
10. Загальна методика перевірки правильності нульової гіпотези.
11. Що таке рівень значущості ?
12. Що таке потужність критерію?
13. Назвати основні види статистичних гіпотез.
14. За яким критерієм перевіряється гіпотеза про рівність середніх за невідомих дисперсій?
15. За яких умов доцільно застосовувати непараметричні критерії перевірки гіпотез?

Розділ 9

КОРЕЛЯЦІЙНО-РЕГРЕСІЙНИЙ АНАЛІЗ

План

- 9.1. Загальне поняття кореляційно-регресійного методу аналізу.
- 9.2. Рівняння регресії і визначення його параметрів.
- 9.3. Множинна кореляційно-регресійна модель.
- 9.4. Задачі для самостійної роботи.
- 9.5. Питання для самоконтролю.

9.1. Загальне поняття кореляційно-регресійного методу аналізу

Будь-яке явище природи і суспільства не може бути усвідомленим і зрозумілим без обґрунтування його зв'язків з іншими явищами. Щоб пізнати сутність явищ, необхідно розкрити їх взаємовідносини, кількісно визначити вплив тих або інших об'єктивних і суб'єктивних факторів.

Вплив різноманітних факторів на рівень економічних показників визначається в основному за допомогою встановлення співвідношень, які мають місце при функціональних зв'язках, коли кожному значенню аргумента відповідає визначене значення функції. Параметри функціональних рівнянь оцінюються на основі методів кореляційно-регресійного аналізу.

Термін «кореляція» вперше застосував Ж. Кюв'є у праці «Лекции по сравнительной анатомии» (1800–1805 pp.). Початкове математичне визначення побудови методу кореляції дав О. Браве в 1846 р. («кореляція» – від лат. «correlation» відношення, що означає співвідношення, відповідність предметів або понять).

Означення. *Кореляцією* називається неповний зв'язок між досліджуваними явищами. Це така залежність, коли будь-якому значенню однієї змінної величини може відповідати декілька різноманітних значень іншої змінної. Вона

відображає закон множини причин і наслідків і є вільною неповною залежністю.

У дослідженнях важливо вивчити не стільки міру кореляції, скільки форму її й характер зміни однієї ознаки залежно від зміни іншої. Ці задачі розв'язуються методами регресійного аналізу. Перші спроби застосування цього методу в економіці були зроблені в кінці XIX і на початку XX ст. (праці Е.Е. Слуцького, А.А. Чупрова, В. Парето, Гукера та ін.).

Кореляційний аналіз допомагає вирішити цілий ряд завдань в економічному аналізі. Він дає можливість оцінити окремий вплив на залежну змінну багатьох факторів, які діють взаємопов'язано і взаємозумовлено.

Використання кореляційно-регресійних методів дозволяє вирішити такі основні завдання: 1) встановити характер і тісноту зв'язку між досліджуваними явищами; 2) визначити і кількісно виміряти ступінь впливу окремих факторів і їх комплексу на рівень досліджуваного явища; 3) на підставі отриманої моделі обчислити прогностні значення показника і дати об'єктивну оцінку розвитку досліджуваного явища.

Для визначення тісноти зв'язку між досліджуваними явищами необхідно обчислити коефіцієнт кореляції, перевірити його значущість (істотність) та знати його властивості.

Означення. Лінійний коефіцієнт кореляції – це середня величина з добутків нормованих відхилень для x та y :

$$R = \frac{1}{n} \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \quad (9.1)$$

або

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}, \quad (9.2)$$

або

$$R = \frac{n \sum xy - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}} \quad (9.3)$$

Розділ 9. КОРЕЛЯЦІЙНО-РЕГРЕСІЙНИЙ АНАЛІЗ

Для оцінювання значущості (істотності) коефіцієнта кореляції:

1) обчислимо значення t – критерію Стьюдента:

$$t_p = |R| \cdot \sqrt{\frac{n-2}{1-R^2}} \quad (9.4)$$

2) та порівняємо його з критичним (табличним) значенням (дод. 4):

$$t_{кр} = t_{\left(\frac{\alpha}{2}, n-2\right)},$$

де α – заданий рівень значущості,
 $n - 2$ – кількість степенів вільності;

3) якщо $t_p > t_{кр}$, отриманий коефіцієнт кореляції вважаємо значущим (істотним) з довірчою ймовірністю 0,95.

Властивості коефіцієнта кореляції:

- 1) $|R| \leq 1$;
- 2) якщо $R \rightarrow \pm 1$, то зв'язок між x та y тісний, якщо $R \rightarrow 0$, то немає зв'язку між x та y ;
- 3) якщо $R > 0$ – між x та y прямий зв'язок, якщо $R < 0$ – зв'язок між x та y обернений.

Силу зв'язку між x та y може бути визначено з таблиці:

Таблиця 8.1

Ступінь зв'язку	R
Слабкий	0,1–0,3
Помірний	0,3–0,5
Помітний	0,5–0,7
Щільний	0,7–0,9
Дуже щільний	0,9–0,99

Для вибраного рівня значущості α і відповідного степеня вільності $k = n - 1$ межі надійності для множинного коефіцієнта кореляції R :

$$(R - \Delta R; R + \Delta R),$$

де
$$\Delta R = t_{\alpha/2, k} \cdot \frac{1 - R}{\sqrt{n}}.$$

При визначенні форми зв'язку між досліджуваними явищами розрізняють два типи залежності явищ: функціональний і кореляційний.

При функціональному зв'язку зміна однієї ознаки чи показника на певну величину призведе до зміни другої ознаки чи показника на чітко визначену величину. Такого роду залежність у її чистому вигляді трапляється в математиці, фізиці, хімії.

При кореляційній залежності будь-якому значенню однієї змінної величини може відповідати декілька чи навіть безліч різноманітних, тобто варіюючих значень іншої змінної величини.

Головна відмінність кореляційної залежності від функціональної полягає в тому, що функціональний зв'язок має місце в кожному окремому випадку спостереження, а кореляційний проявляється так само лише в середньому або в цілому для всієї даної сукупності спостережень і є неточним у відношенні окремих спостережень.

Вивчення взаємозв'язків кореляційного типу має істотне значення особливо під час аналізу явищ, які складаються під впливом великої кількості певних умов.

За своїми математичними особливостями кореляційні залежності можуть бути додатними і від'ємними, прямолінійними і криволінійними, простими і множинними.

Якщо визначається зв'язок між двома ознаками, кореляція називається простою; якщо ж явище розглядається як

результат впливу декількох факторів – множинною. За формою кореляційна залежність буває прямолінійною і криво-лінійною, за напрямом – прямою (додатною) і оберненою (від’ємною).

Необхідно підкреслити дві особливості, властиві кореляційному аналізу: при використанні кореляційного методу вирішальне значення має всебічний, економічно усвідомлений попередній аналіз статистичних даних. Слід пам’ятати, що зв’язок між ознаками і властивостями не є результатом математичних розрахунків, а лежить у природі самих економічних явищ і за допомогою методів математичної статистики можна лише об’єктивно виразити існуючі закономірності економічних процесів; кореляцію можна виявити лише досліджуючи достатньо велику сукупність спостережень, оскільки кореляційні зв’язки виявляються у формі одночасної зміни двох або кількох зіставлених ознак.

Статистичне дослідження кореляційної залежності включає завдання визначення форми зв’язку і знаходження кількісної характеристики цієї форми. Процес встановлення форми зв’язку і вибору математичного рівняння, яке могло б найбільш повно відобразити характер взаємозв’язку між ознаками досліджуваного явища, має вирішальне значення в кореляційному аналізі.

Питання вибору форми зв’язку та математичного рівняння можна вирішити на основі кількісного соціально-економічного аналізу явищ, що вивчаються, використовуючи при цьому такі методи статистичного аналізу, як графічний, статистичні групування, дисперсійний аналіз та ін.

Основним завданням кореляційно-регресійного аналізу є встановлення виду функції, яка найкраще б описувала усереднене масове протікання економічного процесу. До *етапів* його вирішення відносяться:

I етап: проведення ґрунтового якісного економічного аналізу досліджуваного процесу;

II етап: на основі проведеного аналізу формулювання гіпотези про вид функції, яка описує цей процес;

III етап: прийняття гіпотези або відмова від неї шляхом статистичної перевірки за емпіричними даними;

IV етап: визначення кінцевої функції регресії, оцінення невідомих значень залежної змінної, одержання теоретично обґрунтованих і статистично надійних різноманітних прогнозів щодо поведінки досліджуваного процесу.

9.2. Рівняння регресії і визначення його параметрів

Одним з найпростіших типів кореляційного зв'язку є модель парної регресії. Хоча у чистому вигляді парна залежність лише двох показників трапляється рідко, однак використання цієї моделі дає змогу зрозуміти сутність багатьох процесів і дослідити їх.

Зауваження. Парну модель застосовують у тому випадку, коли з усіх факторів, які впливають на результат, можна виділити один такий, що має найсуттєвіший вплив (факторна змінна).

Якщо точки спостережень гуртуються навколо деякої прямої лінії, то залежність між змінними можна описати лінійним рівнянням

$$y = \beta_0 + \beta_1 x + u, \quad (9.5)$$

де β_0, β_1 – параметри регресії, які потрібно оцінити;

u – стохастична складова моделі.

Якщо прямої лінії не існує, то відповідну форму залежності потрібно шукати серед нелінійних функцій. Найчастіше використовують такі нелінійні функції:

- гіперболічна:
$$Y = \beta_0 + \beta_1 \frac{1}{X}; \quad (9.6)$$

• степенева: $Y = \beta_0 X^{\beta_1};$ (9.7)

• логарифмічна: $Y = \beta_0 + \beta_1 \log X;$ (9.8)

• поліноміальна: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_k X^k$ (9.9)

Перелічені залежності нескладними перетвореннями можуть бути зведені до лінійного відносно структурних параметрів моделі вигляду.

У всіх інших випадках заміна змінних дає змогу отримати лінійну функцію не лише за структурними параметрами моделі, але й за змінними.

Для оцінювання параметрів парної лінійної моделі застосовується метод найменших квадратів (МНК), введений в практику Гаусом.

Ідея методу найменших квадратів полягає в тому, щоб підібрати такі значення параметрів моделі, за яких сума квадратів відхилень між заданими y_i і розрахунковими \hat{y}_i значеннями результативної змінної буде найменшою, тобто

$$u = \sum (\hat{y}_i - y_i)^2 \rightarrow \min \quad (9.10)$$

МНК-оцінки параметрів парної лінійної моделі:

$$b_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (9.11)$$

$$b_0 = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (9.12)$$

або

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - (\bar{x})^2} = r_{xy} \cdot \frac{S_y}{S_x} \quad (9.13)$$

$$b_0 = \bar{y} - \hat{a}_1 \bar{x} \quad (9.14)$$

де $\bar{x} = \frac{1}{n} \sum x_i$;

$$\overline{x^2} = \frac{1}{n} \sum x_i^2$$
;

$$\bar{y} = \frac{1}{n} \sum y_i$$
;

$$\overline{y^2} = \frac{1}{n} \sum y_i^2$$

$$\overline{xy} = \frac{1}{n} \sum x_i y_i \quad \text{– середні арифметичні значення;}$$

$$S_x^2 = \overline{x^2} - (\bar{x})^2; \quad S_y^2 = \overline{y^2} - (\bar{y})^2 \quad \text{– вибіркові дисперсії.}$$

Зауваження (інтерпретація МНК-оцінок параметрів):

Оцінка b_0 є відрізком, який регресійна пряма відсікає на осі ординат; оцінка b_1 є тангенсом кута нахилу регресійної прямої до осі абсцис.

Алгоритм дослідження парної лінійної регресійної моделі включає такі кроки:

1. Побудувати хмару розсіяння (діаграму розсіяння, поле кореляції) по експериментальних даних і по ній вибрати форму криволінійної моделі (припустити, що форма криволінійної моделі має лінійний вигляд).

2. Оцінити параметри моделі, яка Вами вибрана.

3. Обчислити відносне і середнє значення відносно похибки. Визначити ступінь узгодженості розрахункових значень регресії та значень змінних y (для перевірки регресійної моделі на адекватність), що спостерігаються (залишкову дисперсію, коефіцієнт детермінації, F -критерій. Проаналізувати питання значущості оцінок параметрів моделі – t -критерій для перевірки гіпотез відносно істотності кожного з параметрів регресійної моделі).

4. Дослідити питання, які пов'язані з визначенням значущості вибіркового коефіцієнта кореляції.

5. Обчислити коефіцієнти еластичності.

6. Визначити зони надійності регресії при рівні значущості α .

7. Перевірити на значущість і побудова інтервалів довіри для параметрів регресії b_0, b_1 .

8 Обчислити прогнозні значення і знайти межі надійних інтервалів індивідуальних прогнозованих значень (знайти точковий та інтервальний прогнози).

Для визначення ступеня узгодженості розрахункових значень регресії та значень змінних y (для перевірки регресійної моделі на адекватність) необхідно застосувати F -критерій Фішера.

Обчислити F -статистику за формулою

$$F = \frac{(n-2)R^2}{1-R^2} \quad (9.15)$$

Обчислене значення F -критерію порівняти з табличним (дод. 5) $F_{\text{табл.}}(k_1, k_2, \alpha)$, при степенях вільності $k_1=1$ і $k_2=n-2$ і вибраному рівні значущості α (рівень помилки), або $p=1-\alpha$ (рівень довіри): якщо $F(1, n-2) > F_{\text{табл.}}(1, n-2, \alpha)$, то гіпотезу $H_0: b_1 = 0$ про неістотність зв'язку між залежною і незалежною змінною моделі не приймаємо з ризиком α , тобто коефіцієнт b_1 – значущий; інакше – приймаємо, тобто коефіцієнт $b_1 = 0$ – незначущий.

Коефіцієнти еластичності парної лінійної регресійної моделі обчислюється за формулою

$$E = \alpha = b_1 \frac{\bar{x}}{\bar{y}}, \quad (9.16)$$

$$\text{де } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \bar{y} = \frac{\sum_{i=1}^n y_i}{n}.$$

Коефіцієнт еластичності є показником впливу зміни питомої ваги x на y : показує, що показник моделі y зміниться на $\alpha\%$, якщо фактор x зміниться на 1%.

Для обчислення прогнозних значень y_{pi} у рівняння $\hat{y} = b_0 + b_1x$ необхідно підставити задані значення x_{pi} , які знаходяться за межами базового періоду (точковий прогноз);

Для обчислення меж надійних інтервалів індивідуальних прогнозованих значень необхідно знайти точність інтервальної оцінки за формулою ($k = n - 2$ – степінь вільності, α – вибраний рівень значущості):

$$\Delta \hat{y}_{pi} = t_{\alpha/2, k} S \sqrt{1 + \frac{1}{n} + \frac{(x_{pi} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (9.17)$$

де $t_{\alpha/2, k}$ – критичне значення t -статистики (дод. 4);

S – вибіркове середньоквадратичне відхилення.

Тоді межі надійних інтервалів індивідуальних прогнозованих значень – $(\hat{y}_{pi} - \Delta \hat{y}_{pi}; \hat{y}_{pi} + \Delta \hat{y}_{pi})$.

Приклад 9.1. У таблиці наведено дані про середньорічну кількість опадів та урожайність льону:

Розділ 9. КОРЕЛЯЦІЙНО-РЕГРЕСІЙНИЙ АНАЛІЗ

Таблиця 9.2

Середньорічна кількість опадів, мм	Урожайність льону, ц з 1 га зібраної площі
9,06	4,10
7,70	4,90
8,73	4,70
7,63	5,20
8,29	4,90

Розрахувати лінійний коефіцієнт кореляції та оцінити його істотність.

Розв'язання. Для визначення лінійного коефіцієнта кореляції застосуємо формулу (9.3). Наведемо розрахунки у вигляді таблиці:

Таблиця 9.3

x	y	xy	x ²	y ²
9,06	4,10	37,14	82,05	16,81
7,70	4,90	37,73	59,29	24,01
8,73	4,70	41,01	76,13	22,09
7,63	5,20	39,65	58,14	27,04
8,29	4,90	40,63	68,75	24,01
$\sum x = 41,4$	$\sum y = 23,8$	$\sum xy = 196,16$	$\sum x^2 = 344,36$	$\sum y^2 = 113,96$

Лінійний коефіцієнт кореляції:

$$R = \frac{n \sum xy - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}} =$$

$$= \frac{5 \cdot 196,16 - 41,4 \cdot 23,8}{\sqrt{(5 \cdot 344,36 - 41,4^2)(5 \cdot 113,96 - 23,8^2)}} = -0,88$$

Зв'язок між кількістю опадів та урожайністю льону щільний, обернений.

Оцінимо істотність лінійного коефіцієнта кореляції:

1) розрахунковий критерій:

$$t_p = \frac{|R|}{\sigma^2} = |R| \cdot \sqrt{\frac{n-2}{1-R^2}} = 0,88 \cdot \sqrt{\frac{5-2}{1-0,88^2}} = 3,21;$$

2) критична точка (дод. 4):

$$df = 13 - 2 = 11;$$

$$\alpha = 0,05;$$

$$t_{кр} = t_{\left(\frac{\alpha}{2}, n-2\right)} = 2,201;$$

3) оскільки $t_p > t_{кр}$, то коефіцієнт кореляції вважаємо істотним з ймовірністю 0,95.

Приклад 9.2. За даними прикладу 9.1 визначити параметри лінійного рівняння регресії та оцінити його адекватність.

Розв'язання. Для визначення параметрів лінійного рівняння скористаємося обчисленнями попередньої задачі:

$$b_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{5 \cdot 196,16 - 41,4 \cdot 23,8}{5 \cdot 344,36 - 41,4^2} = -0,58$$

$$b_0 = \frac{\sum y - b \sum x}{n} = \frac{23,8 - (-0,58) \cdot 41,4}{5} = 9,56$$

Отже, рівняння парної регресійної моделі має вигляд:

$$y = 9,56 - 0,58x$$

Оцінимо адекватність моделі на основі F -критерію Фішера.

Розділ 9. КОРЕЛЯЦІЙНО-РЕГРЕСІЙНИЙ АНАЛІЗ

Наведемо необхідні розрахунки у вигляді таблиці:

Таблиця 9.4

x	y_i	$y(x) = 9,56 - 0,58x$	$(y(x) - \bar{y})^2$	$(y(x) - y_i)^2$	$(y_i - \bar{y})^2$
9,06	4,10	4,31	0,204	0,044	0,436
7,70	4,90	5,10	0,113	0,04	0,020
8,73	4,70	4,50	0,066	0,04	0,004
7,63	5,20	5,14	0,144	0,004	0,194
8,29	4,90	4,75	0,000	0,021	0,020
$\sum x = 41,4$	$\sum y = 23,8$ $\bar{y} = 4,76$	$\sum y(x) = 23,80$	$\sum = 0,527$	$\sum = 0,149$	$\sum = 0,674$

Коефіцієнт детермінації:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y(x) - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{0,149}{0,674} = 1 - 0,22107 = 0,7789$$

Коефіцієнт кореляції:

$$R = \sqrt{R^2} = 0,883$$

Розрахунковий F -критерій Фішера:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - 2}{1} = \frac{0,7789}{1 - 0,7789} \cdot 3 = 10,57$$

Критична точка:

$$\alpha = 0,05$$

$$df_1 = k - 1 = 2 - 1 = 1$$

$$df_2 = n - k = 5 - 2 = 3$$

$$F_{кр} = 10,13$$

Розрахункове значення критерію більше за табличне, що дає підставу вважати знайдене рівняння адекватним статистичним даним з імовірністю 0,95.

9.3. Множинна кореляційно-регресійна модель

Під час економічного дослідження часто доводиться вивчати явища, які змінюються під впливом не одного, а багатьох різних факторів, кожний з яких окремо може не справляти вирішального впливу. Сукупний же вплив факторів іноді виявляється достатньо сильним, щоб за їх змінами можна було робити висновки про величини показника досліджуваного явища. Методи вимірювання кореляційного зв'язку одночасно між двома, трьома і більше кореляційними ознаками вивчається в теорії множинної кореляції (питання множинної кореляції вперше досліджував англійський вчений Ф.А. Еджворт у кінці ХІХ ст.).

Множинне кореляційне рівняння встановлює зв'язок між досліджуваними ознаками і дозволяє вирахувати очікувані значення результативної ознаки під впливом включених в аналіз факторів, пов'язаних таким рівнянням.

Припустимо, що між економічним показником Y та факторами X_1, X_2, \dots, X_n існує лінійний зв'язок.

Рівняння множинної або багатofакторної лінійної регресійної моделі має вигляд:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + u, \quad (9.18)$$

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n, \quad (9.19)$$

де y, \hat{y} – відповідно фактичні та розрахункові значення показника;

x_1, x_2, \dots, x_n – відповідно фактичні значення факторів;

$\beta_0, \beta_1, \beta_2, \dots, \beta_n$ та $b_0, b_1, b_2, \dots, b_n$ – параметри моделі, які потрібно оцінити та їх оцінки;

u – стохастична складова.

МНК-оцінки параметрів багатofакторної лінійної регресійної моделі обчислюються за формулою:

$$B = (X'X)^{-1} X'Y, \quad (9.20)$$

де $X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{T1} & \dots & x_{Tm} \end{pmatrix}$ – матриця спостережень за факторами;

X' – транспонована матриця матриці X ;

$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$ – вектор спостережень за показником;

$B = \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_m \end{pmatrix}$ – вектор оцінок параметрів моделі;

$U' = (u_1, u_2, \dots, u_n)$ – стохастична складова моделі

Алгоритм дослідження багатofакторної лінійної регресійної моделі включає:

1. Визначення ступеня узгодженості розрахункових значень регресії і значень змінних y (для перевірки регресійної моделі на адекватність), що спостерігаються:

- відносне і середнє значення відносної похибки;
- залишкову дисперсію;
- коефіцієнт детермінації;
- вибірковий коефіцієнт множинної кореляції та перевірити його на значущість;
- F -критерій (для нульової гіпотези):

$$H_0 : b_1 = b_2 = \dots = b_m = 0;$$

– t -критерій для перевірки гіпотез відносно істотності кожного з параметрів регресійної моделі (аналіз щодо значущості оцінок параметрів моделі);

2. Обчислення коефіцієнта еластичності.

3. Визначення зони надійності регресії при рівні значущості α .

4. Перевірка на значущість і побудову інтервалів довіри для параметрів регресії.

5. Тестування наявності зв'язків між факторами моделі.

6. Вивчення особливих випадків у поведінці стохастичної складової.

7. Обчислення прогнозних значень y_{pi} за x_{pi} , які лежать за межами базового періоду, і знаходження меж надійних інтервалів індивідуальних прогнозованих значень і меж надійних інтервалів для математичного сподівання (точковий та інтервальний прогнози).

Коефіцієнти еластичності множинної лінійної регресійної моделі обчислюються за формулою

$$\alpha_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}}. \quad (9.21)$$

Коефіцієнт еластичності є показником впливу зміни питомої ваги x_i на y : показує, що показник y зміниться на $\alpha_i\%$, якщо фактор x_i зміниться на 1%; загальна еластичність y від усіх факторів x_i дорівнює:

$$\alpha = \sum_{i=1}^m \alpha_i \quad (9.22)$$

Загальна еластичність засвідчує, що показник y зміниться на $\alpha\%$, якщо всі фактори x_i зміняться на 1%.

Зауваження. Існують можливості оцінити параметри та обчислити основні характеристики регресійних функцій засобами MS Excel за допомогою вбудованих функцій:

ЛИНЕЙН – МНК-оцінки параметрів лінійних моделей та їх основні характеристики;

ОТРЕЗОК – МНК-оцінка параметра b_0 парної лінійних моделей;

НАКЛОН – МНК-оцінка параметра b_1 парної лінійних моделей;

КОРРЕЛ – коефіцієнт кореляції для заданих x та y ;

ФРАСПОБР – критична точка F-критерію Фішера;

СТЮДРАСПОБР – критична точка t -критерію Стьюдента;

СРЗНАЧ – середнє (арифметичне) значення аргументів;

ДИСПР – дисперсія генеральної сукупності;

СТАНДОТКЛОН – стандартне відхилення за вибіркою.

9.4. Задачі для самостійної роботи

Задача 9.1. Результат оцінювання групи товарів за ступенем етнічного різноманіття та соціальної напруженості відображено в таблиці:

Таблиця 9.5

Країна	Етнічне різноманіття	Ступінь соціальної напруженості
А	101	135
Б	97	134
В	95	490
Г	90	515
Д	86	118
Е	71	13
Є	57	85
Ж	55	83
З	51	79
И	49	138
І	38	11
Ї	35	65
Й	26	59
К	10	59
Л	1	44

СТАТИСТИКА

Обчислити коефіцієнт кореляції та перевірити його на істотність. Проаналізувати отримані результати.

Задача 9.2. Заповнити порожні клітинки таблиці та перевірити коефіцієнти кореляції на істотність:

Таблиця 9.6

Показник	А	Б	В	Г	Ґ	Д
N	10	7	29	20	10	30
R	-0,8	0,8	0,3	-0,4	-0,9	0,6
t_p						
$t_{кр} (95\%)$						
$t_{кр} (99\%)$						
$t_{кр} (99,9\%)$						

Пояснити значення коефіцієнтів кореляції.

Задача 9.3. Оцінити істотність коефіцієнтів кореляції, поданих у таблиці, та визначити їх довірчі межі:

Таблиця 9.7

Показник	А	Б	В	Г	Ґ
N	60	160	90	60	110
R	0,4	0,4	-0,72	0,72	-0,9

Пояснити значення коефіцієнтів кореляції.

Задачі 9.4–9.8:

- 1) побудувати діаграми розсіювання та вибрати форму зв'язку між змінними;
- 2) обчислити оцінки параметрів лінійного рівняння регресії;
- 3) перевірити адекватність моделі статистичним даним за F -критерієм Фішера;
- 4) обчислити коефіцієнти кореляції, детермінації та еластичності;
- 5) перевірити значущість коефіцієнта кореляції.
Проаналізувати отримані результати.

Розділ 9. КОРЕЛЯЦІЙНО-РЕГРЕСІЙНИЙ АНАЛІЗ

Задача 9.4. У таблиці наведено демографічні дані країн Європи:

Таблиця 9.8

Країна	Середній вік жінок, які взяли другий шлюб, років	Коефіцієнт розлучень
Австрія	35,1	1,5
Білорусь	35,8	1,8
Франція	36,0	1,3
Німеччина	36,4	1,7
Бельгія	36,7	1,1
Норвегія	36,7	1,4
Швеція	38,3	1,3
Нідерланди	38,5	1,9
Іспанія	39,0	1,1

Задача 9.5. У таблиці наведено дані про територію та кількість населення країн Європи:

Таблиця 9.9

Країна	Територія, тис. км ²	Кількість населення, млн осіб
Чехія	79	10,1
Австрія	84	8,5
Швеція	450	9,7
Угорщина	93	9,8
Німеччина	357	82,8
Польща	312	38,4
Франція	551	66,8
Естонія	45	1,3
Литва	65	2,8
Португалія	92	10,3
Румунія	238	20,1
Фінляндія	338	5,5
Ісландія	103	0,3

СТАТИСТИКА

Задача 9.6. У таблиці наведено дані про взаємозв'язок між вмістом фтору в питній воді, ураженістю зубів карієсом та флюорозом:

Таблиця 9.10

Вміст фтору у воді, мг/л	Ураженість зубів флюорозом, %	Ураженість зубів карієсом, %
0,9	0	69
0,1	20	31
1,8	31	27
1,3	38	18
2,8	49	15
5,1	99	8

Задача 9.7. Дані про витрати пального кормозбирального комбайна Ягуар-589SL та урожайність культур наведено в таблиці:

Таблиця 9.11

Культура	Урожайність, ц/га	Витрати пального, л/га
1	2	3
А	110	11,9
Б	130	16,7
В	180	19,3
Г	260	20,4
Г	310	38,5
Д	440	39,8
Е	460	40,3
Є	570	41,8
Ж	610	45,6
З	720	46,5
И	780	49,6
І	790	50,1

Задача 9.8. У таблиці наведено дані про технічні характеристики машин та їх роздрібну ціну:

Таблиця 9.12

Марка машини	Об'єм двигуна, л	Потужність двигуна, к. с.	Роздрібна ціна, тис. EUR
Honda (Jazz)	1,8	91	17,0
Audi (A8)	3,7	230	70,3
BMW (X5 4.4)	5,1	300	78,6
Ford (Connect)	2,3	100	18,9
Mazda (Premacy)	3,3	153	24,2
Audi (A4 Avant)	2,2	178	35,3
Volvo (V50 N5 AWD)	3,4	230	45,7
Mersedes (C203)	2,8	198	40,1
Mazda (6)	3,5	145	23,2
Mersedes (C200 kompressor)	2,9	178	35,3
Saab (9-3 Sport Combi)	3,1	186	26,5
BMW (318 i)	3,2	178	29,6
BMW (318 i)	3,3	159	27,8
Peugeot (407)	4,0	167	36,5
Hyundai (Terracan)	3,1	151	36,5
BMW (3 Touring)			

9.5. Питання для самоконтролю

1. Що таке парна регресія?
2. Які види нелінійних функцій найчастіше застосовують в економічних дослідженнях? Як їх перетворити до лінійного вигляду?
3. У чому суть методу найменших квадратів?
4. Навести формули розрахунку оцінок параметрів парного лінійного рівняння регресії за методом найменших квадратів.
5. Які висновки можна зробити про оцінювання параметрів парної лінійної регресії за методом найменших квадратів?

6. Проінтерпретуйте оцінки параметрів парної лінійної регресії.
7. Навести критерій перевірки адекватності моделі та проаналізувати результат.
8. Навести критерій значущості коефіцієнта кореляції та проаналізувати результат.
9. Сформулювати властивості коефіцієнта кореляції.
10. Як обчислити коефіцієнт еластичності парної лінійної регресійної моделі?
11. Який висновок можна зробити за коефіцієнтом еластичності парної лінійної регресійної моделі?
12. Як дослідити парну лінійну регресію засобами MS Excel?
13. Дати визначення лінійної багатофакторної регресійної моделі.
14. Навести матричну форму класичної лінійної багатофакторної моделі.
15. Прокоментувати основні етапи побудови лінійної багатофакторної регресійної моделі.
16. Як розраховуються невідомі параметри лінійної моделі багатофакторної регресії за методом найменших квадратів?
17. Як обчислити коефіцієнт еластичності множинної лінійної регресійної моделі?
18. Який висновок можна зробити за коефіцієнтом еластичності множинної лінійної регресійної моделі?
19. Як дослідити багатофакторну лінійну регресію засобами MS Excel?

РОЗДІЛ 10

РЯДИ ДИНАМІКИ ТА ЇХ АНАЛІЗ

План

- 10.1. Поняття ряду динаміки. Зіставленість і змикання рядів динаміки.
- 10.2. Показники аналізу рядів динаміки.
- 10.3. Середні показники аналізу динаміки.
- 10.4. Дослідження динамічних та кореляційних властивостей статистичних показників. Прогнози за моделями динаміки.
- 10.5. Задачі для самостійної роботи.
- 10.6. Питання для самоконтролю.

10.1. Поняття ряду динаміки. Зіставленість і змикання рядів динаміки

Одним з основних завдань статистичного аналізу є дослідження розвитку, поведінки та зміни явищ у часі. Статистичні дані, які характеризують такі процеси і явища, утворюють ряди динаміки або часові чи хронологічні ряди.

Означення. *Ряд динаміки* – це числові значення визначеного статистичного показника виміряні в послідовні моменти або періоду часу.

Ряд динаміки складається із двох елементів: *часу або періоду* (t_i) та *рівня ряду* (y_i) – конкретного значення статистичного показника. Обидва елементи називаються членами ряду динаміки.

Залежно від способів вимірювання даних ряди динаміки класифікують як дискретні і неперервні.

Дискретним рядом називається ряд даних виміряних через певні проміжки часу (місяці, квартали, роки тощо).

Дискретні ряди бувають трьох видів: моментні, інтервальні (періодичні) та ряди середніх величин.

Моментні ряди динаміки характеризують розмір явищ на певний момент часу (на початок місяця, кварталу, року тощо). Наприклад, чисельність працівників, кількість об'єктів роздрібної торгівлі тощо.

Інтервальні ряди динаміки характеризують розмір явищ за певні проміжки (інтервали) часу (за добу, місяць, квартал, рік тощо). Наприклад, фонд заробітної плати, обсяг виробленої продукції тощо.

Часові інтервали між рівнями можуть бути однаковими (рівновіддалені ряди динаміки) і неоднаковими (нерівновіддалені).

Особливістю інтервальних рядів є те, що їх рівні можна ділити і додавати. Такі дії з рівнями моментного ряду не мають сенсу, оскільки кожен з наступних рівнів ряду охоплює попередні.

Ряди середніх характеризують зміну середніх рівнів явищ у часі. Наприклад, місячний рівень заробітної плати, денна продуктивність праці тощо.

Неперервним рядом називається ряд даних вимірених за певні проміжки часу (місяці, квартали, роки тощо). Такі ряди одержують при вимірюванні даних за допомогою відповідних приладів.

Статистичні дані, що входять до складу рядів динаміки, повинні бути порівняними між собою. Одна з вимог до рівнів динамічного ряду це зіставленість їх рівнів. Відсутність зіставленості може виникнути через зміну: меж території, які характеризують рівні динамічного ряду; методології обліку і розрахунку показників; дати обліку; одиниць виміру.

Зауваження. Для забезпечення зіставленості рівнів динамічного ряду застосовують такі методи:

1) **змикання рядів динаміки** – об'єднання в один динамічний ряд двох і більше рядів, що обчислені за різними методологіями. Для його застосування необхідно, щоб дані одного з періодів були обчислені за двома методологіями;

2) *зведення рядів до однієї основи* – перехід від абсолютних величин до відносних. Застосовується до декількох рядів динаміки як розрахунок рівнів ряду у відсотках або коефіцієнтах відносно рівня одного й того ж періоду.

10.2. Показники аналізу рядів динаміки

Для оцінки інтенсивності динаміки рівні динамічного ряду порівнюють між собою. Залежно від бази порівняння їх поділяють на *базисні* (стосовно базисного періоду) та *ланцюгові* (стосовно попереднього періоду).

Показники оцінювання інтенсивності ряду динаміки

Абсолютний приріст:

• базисний $\Delta = y_i - y_1$ (10.1)

• ланцюговий $\Delta_i = y_i - y_{i-1}$, (10.2)

де i – порядковий номер часового періоду ($i = 1, \dots, n$);

y_i – рівень ряду i -го періоду;

y_1 – початковий (базисний) рівень ряду (першого періоду);

y_{i-1} – рівень ряду $i - 1$ -го періоду.

Зауваження. Абсолютний приріст показує на скільки одиниць змінився рівень динамічного ряду порівняно з початковим (Δ) або попереднім (Δ_i).

Коефіцієнт зростання:

• базисний $K = \frac{y_i}{y_1}$ (10.3)

- ланцюговий
$$K_i = \frac{y_i}{y_{i-1}} \quad (10.4)$$

Зауваження. Коефіцієнт зростання показує у скільки разів змінився рівень динамічного ряду порівняно з початковим (K) або попереднім (K_i).

Темп зростання:

- базисний
$$T = K \cdot 100 = \frac{y_i}{y_1} \cdot 100 \quad (10.5)$$

- ланцюговий
$$T_i = K_i \cdot 100 = \frac{y_i}{y_{i-1}} \cdot 100 \quad (10.6)$$

Темп зростання показує відсоткове значення зміни рівня динамічного ряду порівняно з початковим (T) або попереднім (T_i).

Темп приросту:

- базисний
$$T_{\text{пр}} = T - 100 = (K - 1) \cdot 100 = \frac{\Delta}{y_1} \cdot 100 \quad (10.7)$$

- ланцюговий
$$T_{\text{пр}i} = T_i - 100 = \frac{\Delta_i}{y_{i-1}} \cdot 100 \quad (10.8)$$

Темп приросту показує на скільки відсотків змінився рівень динамічного ряду порівняно з початковим ($T_{\text{пр}}$) або попереднім ($T_{\text{пр}i}$).

Абсолютне значення 1-го % приросту

- ланцюгове
$$A_{1\%i} = \frac{\Delta_i}{T_{\text{пр}i}} = \frac{y_{i-1}}{100} = 0,01y_{i-1} \quad (10.9)$$

Зауваження. Абсолютне значення 1-го % приросту показує скільки абсолютних одиниць припадає на 1% приросту (зменшення).

Абсолютне прискорення (уповільнення)

• ланцюгове $\Delta'_i = \Delta_i - \Delta_{i-1}$ (10.10)

Зауваження. $\Delta'_i > 1$ свідчить про прискорення динамічного ряду; $\Delta'_i < 1$ свідчить про його уповільнення.

Відносне прискорення (уповільнення)

• ланцюгове $k_\Delta = \frac{k_i}{k_{i-1}}$ (10.11)

Зауваження. $k_\Delta > 1$ свідчить про прискорення динамічного ряду; $k_\Delta < 1$ свідчить про його уповільнення.

Властивості базисних та ланцюгових показників:

1) сума ланцюгових абсолютних приростів дорівнює кінцевому базисному:

$$\sum_{i=1}^m \Delta_i = \sum_{i=1}^m (y_i - y_{i-1}) = y_n - y_1, \quad (10.12)$$

де m – кількість абсолютних приростів ($m = n - 1$);

y_n – кінцевий рівень ряду;

y_1 – початковий рівень ряду;

2) добуток ланцюгових коефіцієнтів зростання дорівнює кінцевому базисному

$$\prod_{i=1}^m K_i = K_1 \cdot K_2 \cdot \dots \cdot K_m = \frac{y_n}{y_1}, \quad (10.13)$$

де m – кількість коефіцієнтів зростання ($m = n - 1$).

СТАТИСТИКА

Приклад 10.1. Задана статистична інформація про кількість об'єктів Київської області, які змінили форму власності протягом 2013–2018 рр.: 2013 – 5162 одиниць; 2014 – 6017; 2015 – 6137; 2016 – 6527; 2017 – 6149; 2018 – 5664 одиниць. Проаналізуйте статистичний ряд за допомогою базових і ланцюгових показників динаміки.

Розв'язання. Для розрахунку базових показників динаміки використаємо формули (10.1)– 10.11). Розрахунки оформлено в табл. 10.1 і 10.2.

Таблиця 10.1

Базові показники динаміки

Рік	Кількість об'єктів, y_i (одиниць)	Базові показники динаміки			
		Абсолютний приріст, Δ (одиниць)	Коефіцієнт зростання, k	Темпи зростання, T (%)	Темпи приросту, Tn (%)
2013	5162	0	1	0,0	0,0
2014	6017	855	1,166	116,6	16,6
2015	6137	975	1,189	118,9	18,9
2016	6527	1365	1,264	126,4	26,4
2017	6149	987	1,191	119,1	19,1
2018	5664	502	1,097	109,7	9,7

Таблиця 10.2

Ланцюгові показники динаміки

Рік	Кількість об'єктів, y_i (одиниць)	Ланцюгові показники динаміки				
		Абсолютний приріст, Δ (одиниць)	Коефіцієнт зростання, k	Темпи зростання, T (%)	Темпи приросту, Tn (%)	Абсолютне значення 1% приросту, $A_1\%$ (одиниць)
2013	5162	X	X	X	X	X
2014	6017	855	1,166	116,6	16,6	52
2015	6137	120	1,020	102,0	2,0	60

Закінчення табл. 10.2

Рік	Кількість об'єктів, y_i (одиниць)	Ланцюгові показники динаміки				
		Абсолютний приріст, Δ (одиниць)	Коефіцієнт зростання, k	Темпи зростання, T (%)	Темпи приросту, Tn (%)	Абсолютне значення 1% приросту, $A_1\%$ (одиниць)
2016	6527	390	1,064	106,4	6,4	61
2017	6149	-378	0,942	94,2	-5,8	65
2018	5664	-485	0,921	92,1	-7,9	61

Висновки. Потягом 2013–2016 рр. спостерігалось збільшення кількості об'єктів, що змінили форму власності, а з 2017 р. – різкий спад. Проте кількість об'єктів, що змінили форму власності 2014 – 2018 рр. перевищувала її рівень у 2013 р. До 2016 р. кількість підприємств, що змінили форму власності, збільшувалась: у 2014 р. – на 855 підприємств (16,6%) порівняно з попереднім роком, у 2015 р. – на 120 (2%) і у 2016 р. – на 390 підприємств (6,4%). З 2017 р. кількість таких підприємств почала зменшуватись: у 2017 р. – на 378 одиниць (5,8%), у 2018 р. – на 485 (7,9%). За весь період кількість таких підприємств збільшилася на 502 одиниці, або 9,7% (2018 р. – порівняно з 2013 р.)

10.3. Середні показники аналізу динаміки

Для узагальненої характеристики динамічних рядів розраховують середні величини. Однією з таких характеристик динамічного ряду є середній рівень ряду. Для різних видів рядів динаміки його розраховують по-різному.

Означення: *середній рівень інтервального ряду* з рівними періодами (інтервалами) розраховують як просту середню арифметичну:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}, \quad (10.14)$$

де y_i – рівень ряду i -го періоду;
 n – кількість рівнів ряду.

Означення: середній рівень моментного ряду з однаковими проміжками між датами розраховують як середню хронологічну:

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + \dots + y_{n-1} + \frac{1}{2}y_n}{n-1} = \frac{\frac{1}{2}(y_1 + y_n) + \sum_{i=2}^{n-1} y_i}{n-1}. \quad (10.15)$$

Якщо проміжки між датами моментного ряду не однакові й відомо, що між датами явище не змінювалося, то його середній рівень вираховують за формулою середньої арифметичної зваженої:

$$\bar{y} = \frac{\sum_{i=1}^n y_i t_i}{\sum_{i=1}^n t_i}, \quad (10.16)$$

де t_i – період часу, протягом якого рівень ряду не змінювався.

Якщо ж явище між періодами моментного ряду з неоднаковими проміжками змінювалося, то його середній рівень вираховують за формулою:

$$\bar{y} = \frac{\sum_{i=1}^{n-1} (y_i + y_{i+1}) t_i}{2 \sum_{i=1}^{n-1} t_i}, \quad (10.17)$$

де t_i – період часу між i -ю та $i+1$ -ю датами.

Означення: середній абсолютний приріст динамічного ряду визначають як середню арифметичну просту з ланцюгових абсолютних приростів:

$$\bar{\Delta} = \frac{\sum_{i=1}^m \Delta_i}{m}, \quad (10.18)$$

або на основі базового абсолютного приросту останнього рівня ряду (на основі властивості (5.12)):

$$\bar{\Delta} = \frac{y_n - y_1}{n-1} \quad (10.19)$$

Означення: середній коефіцієнт зростання визначають як середню геометричну з ланцюгових коефіцієнтів зростання:

$$\bar{K} = \sqrt[m]{\prod_{i=1}^m K_i} \quad (10.20)$$

або на основі базового абсолютного приросту останнього рівня ряду (на основі властивості (9.13)):

$$\bar{K} = \sqrt[n-1]{\frac{y_n}{y_1}} \quad (10.21)$$

У деяких випадках для визначення середнього коефіцієнта зростання важливо оцінити не досягнутий кінцевий рівень явища, а й його загальну суму за період. Тоді використовують формулу **середньої параболічної**:

$$\frac{\sum_{i=1}^n y_i}{y_1} = \bar{K} + (\bar{K})^2 + (\bar{K})^3 + \dots + (\bar{K})^n \quad (10.22)$$

Для визначення середньої параболічної розроблено спеціальну таблицю. У стовпчику з відповідною кількістю рівнів ряду треба знайти число, яке найбільш близьке до

розрахованого $\sum_{i=1}^n y_i / y_1$. Значення \bar{K} , яке розташоване в першій колонці відповідного рядка, і є середньою параболічною.

Зауваження. Середні темпи зростання та приросту і середнє абсолютне значення 1% приросту розраховують за їх означеннями.

Приклад 10.2. За даними прикладу 5.1 визначити середні показники динаміки.

Розв'язання. Дані представлено за інтервалом часу (роки), тому для обчислення середнього рівня ряду скористаємося формулою (10.14):

$$\bar{y} = \frac{\sum_{i=1}^6 y_i}{n} = \frac{5162 + 6017 + 6137 + 6527 + 6149 + 5664}{6} = 5943$$

Кількість підприємств Київської області, що змінили форму власності, становила в середньому 5943 за 2013–2018 рр.

Середній абсолютний приріст – це, з одного боку, середня арифметична ланцюгових абсолютних приростів (10.18):

$$\bar{\Delta} = \frac{\sum_{i=1}^5 \Delta_i}{m} = \frac{855 + 120 + 390 - 378 - 485}{5} = 100,4$$

А з іншого – різниця кінцевого та першого рівня ряду, поділена на кількість абсолютних приростів (10.19):

$$\bar{\Delta} = \frac{y_{n-y_1}}{n-1} = \frac{5664 - 5162}{6-1} = 100,4$$

За 2013–2018 рр. кількість підприємств Київської області, що змінили форму власності, щороку збільшувалася в середньому на 100 підприємств.

Середній коефіцієнт зростання також можна отримати двома способами (10.20) та (10.21):

$$\bar{K} = \sqrt[m]{\prod_{i=1}^m \kappa_i} = \sqrt[5]{1,166 \cdot 1,020 \cdot 1,064 \cdot 0,942 \cdot 0,921} = 1,019$$

або

$$K = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[5]{\frac{5664}{5162}} = 1,019$$

Середній коефіцієнт зростання за середньою параболічною (10.22):

$$\frac{\sum_{i=1}^n y_i}{y_1} = \frac{5162 + 6017 + 6137 + 6527 + 6149 + 5664}{5162} = 6,9$$
$$\bar{K} = 1,04$$

Середній коефіцієнт зростання за середньою параболічною має більше значення, оскільки тут ураховано значення проміжних рівнів ряду. Середній темп зростання та приросту:

$$\bar{T} = \bar{K} \cdot 100 = 1,019 \cdot 100 = 101,9\%;$$

$$\bar{T}_{\text{пр}} = \bar{T} - 100 = 1,019 - 100 = 1,9\%;$$

За 2013–2018 рр. кількість підприємств Київської області, що змінили форму власності, щороку збільшувалась у середньому на 1,9%.

Приклад 10.3. Відомими є показники попиту на робочу силу в Україні за 2017 р. – початок 2018 р. (станом на початок кожного місяця), подані в таблиці:

СТАТИСТИКА

Таблиця 10.3

Місяць	Попит на робочу силу, тис. осіб	Місяць	Попит на робочу силу, тис. осіб
Січень 2017	166,5	Серпень 2017	210,4
Лютий 2017	170,2	Вересень 2017	210,6
Березень 2017	167,1	Жовтень 2017	218,1
Квітень 2017	188,9	Листопад 2017	214,6
Травень 2017	197,5	Грудень 2017	200,9
Червень 2017	207,2	Січень 2018	204,4

Визначте середню величину показника попиту на робочу силу за: квартали; півріччя; рік. Обґрунтуйте вибір середньої.

Розв'язання. Показники попиту на робочу силу представлено моментними показниками (інформацію подано станом на початок кожного місяця). Інтервали часу між датами однакові, тому для розрахунку середніх величин застосуємо формулу (10.15):

1) середньомісячні показники попиту на робочу силу за кварталами:

$$\bar{x} = \frac{\frac{1}{2}(166,5 + 188,9) + 170,2 + 167,1}{3} = 171,7 \text{ тис. осіб} \text{ – для}$$

I кварталу;

$$\bar{x} = \frac{\frac{1}{2}(188,9 + 205,7) + 197,5 + 207,2}{3} = 200,7 \text{ тис. осіб} \text{ – для}$$

II кварталу;

$$\bar{x} = \frac{\frac{1}{2}(205,7 + 218,1) + 210,4 + 210,6}{3} = 211,0 \text{ тис. осіб} \text{ – для}$$

III кварталу;

Розділ 10. РЯДИ ДИНАМІКИ ТА ЇХ АНАЛІЗ

$$\bar{x} = \frac{\frac{1}{2}(218,1 + 204,4) + 214,6 + 200,9}{3} = 208,9 \text{ тис. осіб} - \text{ для}$$

IV кварталу;

2) середньомісячні показники попиту на робочу силу в першому та другому півріччях 2017 р.:

$$\begin{aligned}\bar{x} &= \frac{\frac{1}{2}(166,5 + 205,7) + 170,2 + 167,1 + 188,9 + 197,5 + 207,2}{6} = \\ &= \frac{1}{2}(171,7 + 200,7) = 186,2 \text{ тис. осіб};\end{aligned}$$

$$\begin{aligned}\bar{x} &= \frac{\frac{1}{2}(205,7 + 204,4) + 210,4 + 10,6 + 218,1 + 214,6 + 200,9}{6} = \\ &= \frac{1}{2}(211 + 208,9) = 209,9 \text{ тис. осіб};\end{aligned}$$

3) середньорічний обсяг попиту на робочу силу:

$$\begin{aligned}\bar{x} &= \frac{\frac{1}{2}(166,5 + 204,4) + 170,2 + 167,1 + 188,9 + 197,5 + 207,2 + 210,4}{12} + \\ &+ \frac{210,6 + 218,1 + 214,6 + 200,9}{12} = \frac{1}{4}(171,7 + 200,7 + 211 + 208,9) = \\ &= \frac{1}{2}(186,2 + 209,9) = 198,1 \text{ тис. осіб}.\end{aligned}$$

Приклад 10.4. У таблиці наведено дані про оптові ціни на цитрини в першому півріччі 2017 та 2018 рр.:

СТАТИСТИКА

Таблиця 10.4

Дата	Ціна 2017 р., грн. за 1 кг	Ціна 2018 р., грн. за 1 кг
01.01	25,0	25,2
07.02	25,1	24,7
05.03	25,2	24,4
03.04	25,1	24,9
05.05	24,6	25,4
06.06	24,0	25,3

Ціна цитрин між указаними датами не змінювалась. Визначте абсолютну та відносну зміни середнього рівня оптових цін на цитрини.

Розв'язання. Динамічний ряд, наведений в умові, є моментним (інформацію подано станом на дати часу) з неоднаковими проміжками між датами. Для розрахунку середніх значень треба визначити кількість днів між датами (дані наведено в таблиці):

Таблиця 10.5

Період	Кількість днів (<i>t</i>)	Ціна 2017 р., грн. за 1 кг	Ціна 2018 р., грн. за 1 кг
з 01.01 до 06.02	37	25,0	25,2
з 07.02 до 04.03	26	25,1	24,7
з 05.03 до 02.04	29	25,2	24,4
з 03.04 до 04.05	32	25,1	24,9
з 05.05 до 05.06	32	24,6	25,4
з 06.06 до кінця півріччя	25	24,0	25,3
Усього	181	X	X

Середня ціна цитрин у першому півріччі 2017 р.:

$$\bar{y} = \frac{\sum_{i=1}^6 y_i t_i}{\sum_{i=1}^6 t_i} = \frac{25,0 \cdot 37 + 25,1 \cdot 26 + 25,2 \cdot 29 + 25,1 \cdot 32 + 24,6 \cdot 32 + 24,0 \cdot 25}{181} =$$

= 24,86 грн.

Середня ціна цитрин у першому півріччі 2018 р.:

$$\bar{y} = \frac{\sum_{i=1}^6 y_i t_i}{\sum_{i=1}^6 t_i} = \frac{25,2 \cdot 37 + 24,7 \cdot 26 + 24,4 \cdot 29 + 24,9 \cdot 32 + 25,4 \cdot 32 + 25,3 \cdot 25}{181} =$$

= 25,00 грн.

Абсолютна зміна середньої ціни на цитрини:

$$\Delta = 25 - 24,86 = 0,14 \text{ грн};$$

Відносна зміна середньої ціни на цитрини:

$$T_{\text{пр}_1} = \frac{\Delta 1}{\bar{y}} \cdot 100 = \frac{0,14}{24,86} \cdot 100 = 0,56\%.$$

Отже, середня ціна на цитрини зросла на 14 коп., що становить 0,56%.

10.4. Дослідження динамічних та кореляційних властивостей статистичних показників. Прогнози за моделями динаміки

Проаналізуємо динаміку статистичного показника або його залежність від дискретних моментів часу за допомогою графічних засобів Excel. За статистичними даними будемо графіки залежності досліджуваного об'єкта від дискретних моментів часу (статистична ламана) та апроксимуємо її різними типами трендів. Для кожного тренду обчислюємо

коефіцієнт достовірності апроксимації (коефіцієнт детермінації) R^2 . За найбільшим коефіцієнтом детермінації робимо висновок про динаміку зміни досліджуваного статистичного показника як відповідної неперервної функції та обчислюємо його прогнозні значення.

Приклад 10.5. На основі статистичної інформації про кількість об'єктів роздрібної торгівлі за період з 2007 р. по 2018 р. дослідити динаміку кількості цих об'єктів та обчислити їх прогнозне значення.

Таблиця 10.6

Дата	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Кількість об'єктів, тис. одиниць	54,9	54,2	53,7	53,1	52,3	52,8	53,7	53,9	53,4	52,8	51,7	51,5

Розв'язання. За допомогою лінійного тренду апроксимуємо статистичну ламану кількості об'єктів роздрібної торгівлі лінійною функцією і визначаємо коефіцієнт детермінації $R^2 = 0,544$.

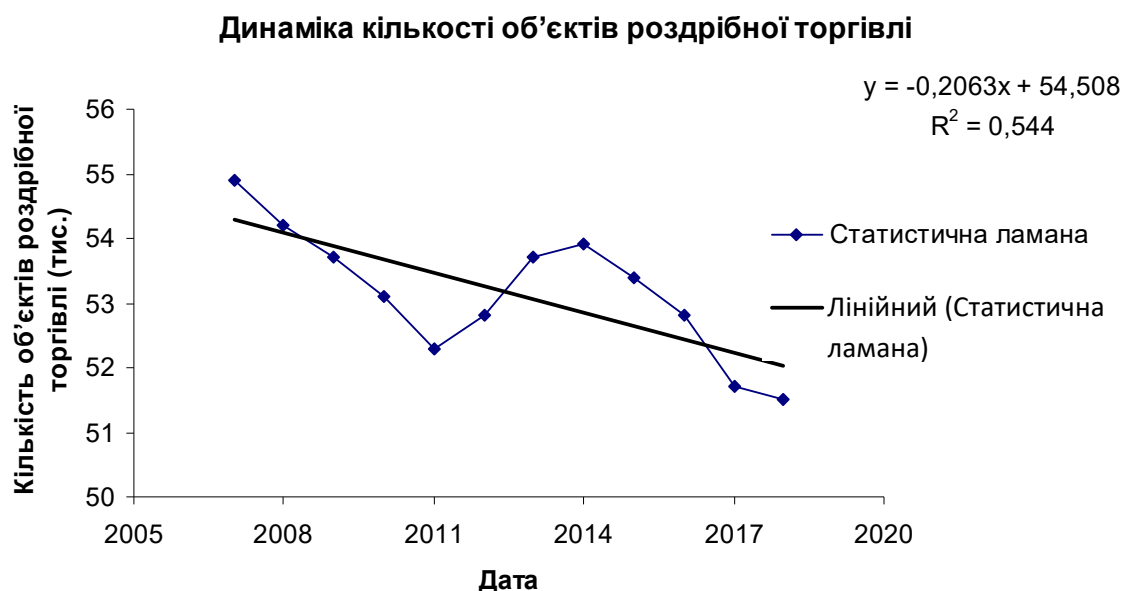


Рис. 10.1. Апроксимація кількості об'єктів роздрібної торгівлі лінійним трендом

Розділ 10. РЯДИ ДИНАМІКИ ТА ЇХ АНАЛІЗ

За допомогою логарифмічного тренду апроксимуємо статистичну ламану кількості об'єктів роздрібної торгівлі логарифмічною функцією і визначаємо коефіцієнт детермінації $R^2 = 0,544$.

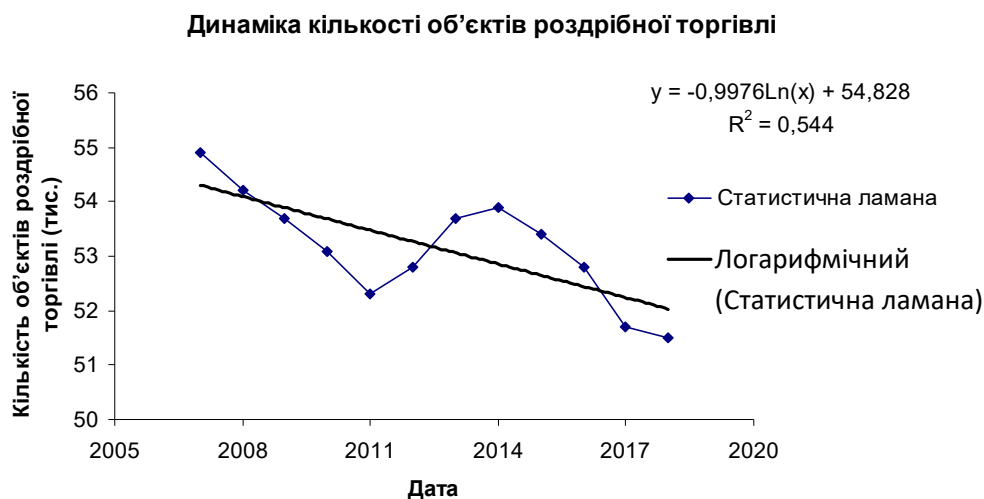


Рис. 10.2. Апроксимація кількості об'єктів роздрібної торгівлі логарифмічним трендом

За допомогою степеневого тренду апроксимуємо статистичну ламану кількості об'єктів роздрібної торгівлі степеневою функцією і визначаємо коефіцієнт детермінації $R^2 = 0,5436$.



Рис. 10.3. Апроксимація кількості об'єктів роздрібної торгівлі степеневим трендом

За допомогою експоненціального тренду апроксимуємо статистичну ламану кількості об'єктів роздрібної торгівлі експоненціальною функцією і визначаємо коефіцієнт детермінації $R^2 = 0,5437$.



Рис. 10.4. Апроксимація кількості об'єктів роздрібної торгівлі експоненційним трендом

За допомогою поліноміального тренду 2-го степеня апроксимуємо статистичну ламану кількості об'єктів роздрібної торгівлі поліномом 2-го степеня і визначаємо коефіцієнт детермінації $R^2 = 0,545$.

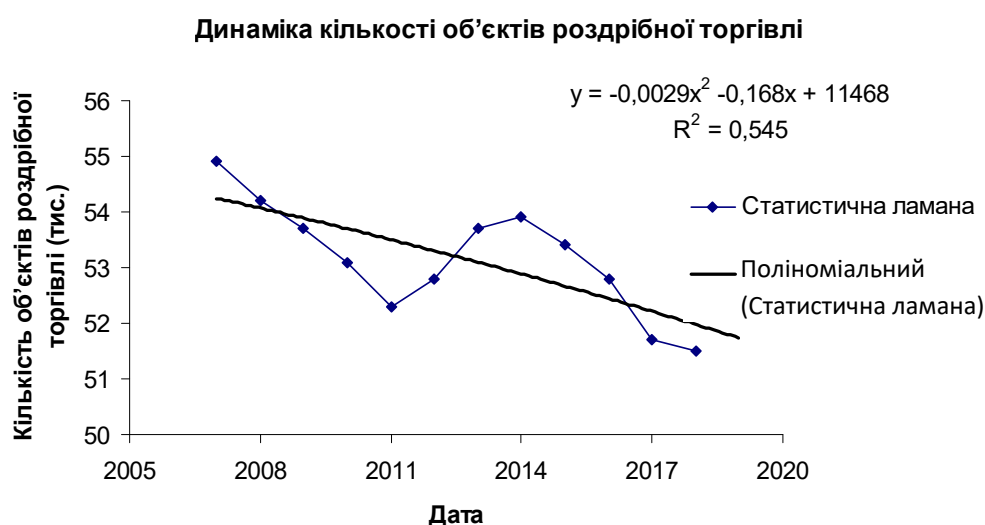


Рис. 10.5. Апроксимація кількості об'єктів роздрібної торгівлі поліноміальним трендом 2-го степеня

За допомогою поліноміального тренду 3-го степеня апроксимуємо статистичну ламану кількості об'єктів роздрібно-ї торгівлі поліномом 3-го степеня і визначаємо коефіцієнт детермінації $R^2 = 0,8385$.

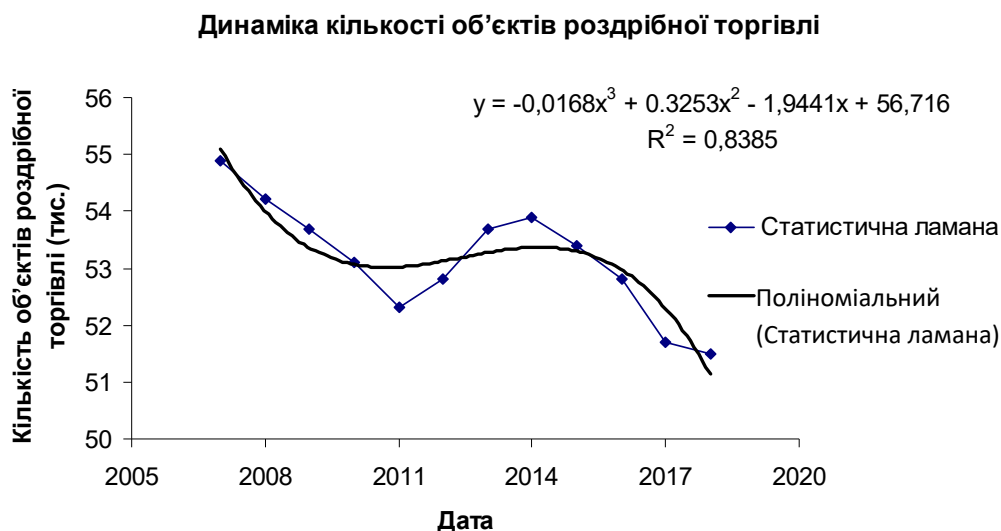


Рис. 10.6. Апроксимація кількості об'єктів роздрібно-ї торгівлі поліноміальним трендом 3-го степеня

За допомогою поліноміального тренду 4-го степеня апроксимуємо статистичну ламану кількості об'єктів роздрібно-ї торгівлі поліномом 4-го степеня і визначаємо коефіцієнт детермінації $R^2 = 0,8395$.

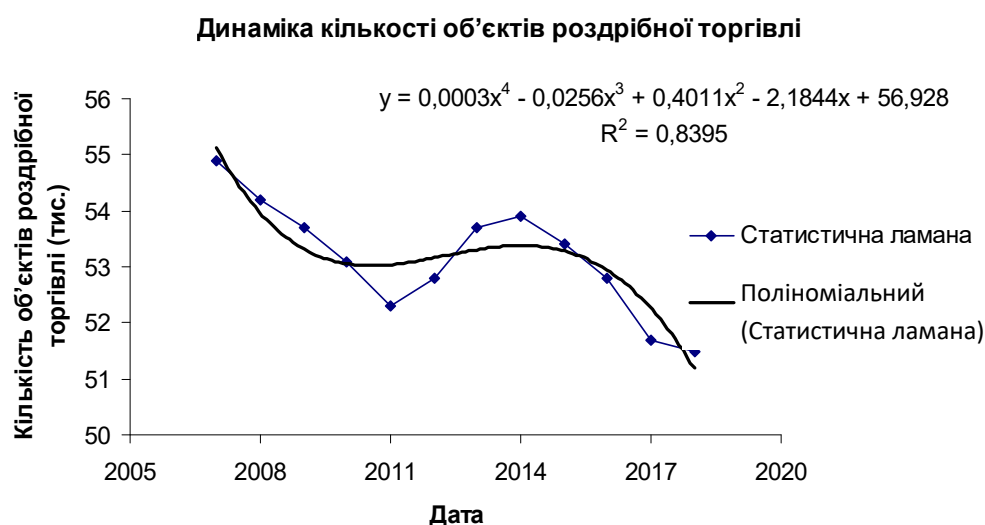


Рис. 10.7. Апроксимація кількості об'єктів роздрібно-ї торгівлі поліноміальним трендом 4-го степеня

За допомогою поліноміального тренду 5-го степеня апроксимуємо статистичну ламану кількості об'єктів роздрібно́ї торгівлі поліномом 5-го степеня і визначаємо коефіцієнт детермінації $R^2 = 0,9597$.

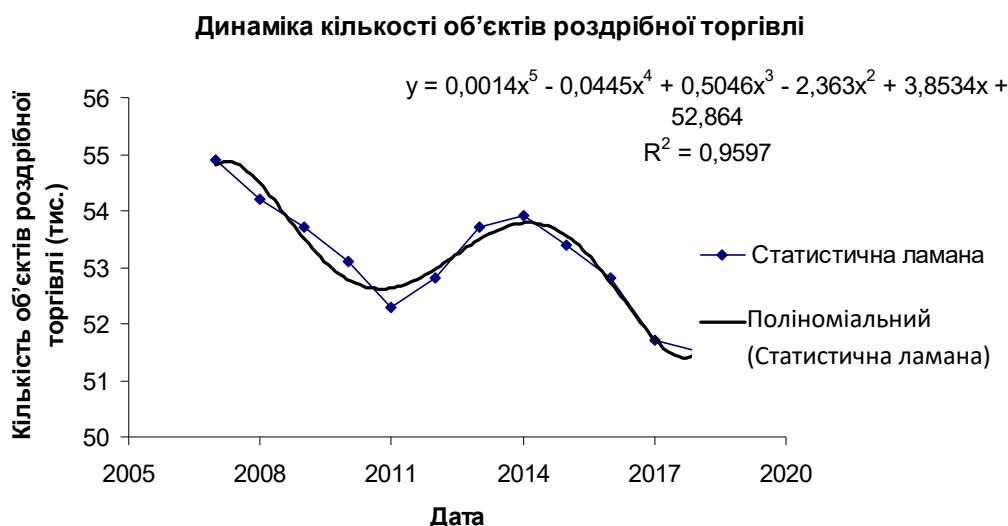


Рис. 10.8. Апроксимація кількості об'єктів роздрібно́ї торгівлі поліноміальним трендом 5-го степеня

За допомогою поліноміального тренду 6-го степеня апроксимуємо статистичну ламану кількості об'єктів роздрібно́ї торгівлі поліномом 6-го степеня і визначаємо коефіцієнт детермінації $R^2 = 0,965$.

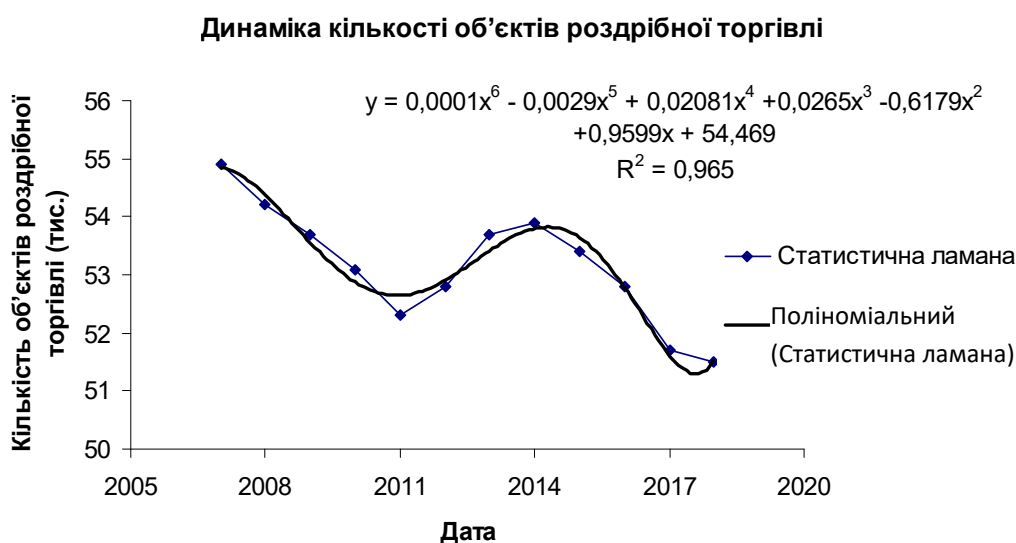


Рис. 10.9. Апроксимація кількості об'єктів роздрібно́ї торгівлі поліноміальним трендом 6-го степеня

Проаналізувавши дані, робимо висновок, що величина достовірності апроксимації досить висока для всіх типів трендів, а найбільша для трендів поліноміального типу. Але прогнози за моделями динаміки для поліномів 3-го та вищих степенів не є достовірними, виходячи з економічної інтерпретації статистичного показника. Таким чином, для отримання достовірного прогнозу кількості об'єктів роздрібної торгівлі за моделлю динаміки виберемо апроксимацію поліномом 2-го ступеня з $R^2 = 0,545$ та зробимо висновок, що досліджуваний статистичний показник змінюється у часі як поліномом 2-го ступеня. Для обчислення прогнозних значень y_{pi} у рівняння полінома 2-го ступеня необхідно підставити наступний дискретний момент часу $x_{pi} = 13$, який знаходяться за межами базового періоду (точковий прогноз):

$$y_{pi} = -0,0029 \times 13^2 - 0,168 \times 13 + 11468 = 51,744.$$

Отже, прогнозне значення кількості об'єктів роздрібної торгівлі у наступному 2019 році приблизно становить **51,744** тис. одиниць.

Обчислимо межі надійних інтервалів індивідуальних прогнозованих значень – $(\hat{y}_{pi} - \Delta\hat{y}_{pi}; \hat{y}_{pi} + \Delta\hat{y}_{pi})$.

Для цього необхідно знайти точність інтервальної оцінки за формулою (9.17):

$$\Delta\hat{y}_{pi} = t_{\alpha/2,k} S \sqrt{1 + \frac{1}{n} + \frac{(x_{pi} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}},$$

де $t_{\alpha/2,k}$ – критичне значення t -статистики (дод. 4);

$k = n - 2$ – степінь вільності;

α – вибраний рівень значущості;

S – вибіркове середньоквадратичне відхилення

Для заданого рівень значущості $\alpha = 0,05$ та степіня вільності $k = 10$ критичне значення t -критерія Стьюдента $t_{0,025, 10} = 2,634$ (дод. 4).

Обчислимо
$$\bar{x} = \sum_{i=1}^{12} x_i = 6,5$$

та
$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^{12} (x_i - 6,5)^2 = 143$$

Стандартна помилка або залишкова дисперсія обчислюється за формулою

$$S = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}} = 0,714$$

Тоді
$$\Delta \hat{y}_{pi} = 2,634 \cdot 0,714 \cdot \sqrt{1 + \frac{1}{12} + \frac{(13-6,5)^2}{143}} = 1,964$$

Отже, межі надійних інтервалів індивідуальних прогнозованих значень $(\hat{y}_{pi} - \Delta \hat{y}_{pi}; \hat{y}_{pi} + \Delta \hat{y}_{pi}) = (49,78; 53,71)$

Зауваження Були використані можливості обчислити основні характеристики регресіних функції засобами MS Excel за допомогою вбудованих функцій:

ЛИНЕЙН – МНК-оцінки параметрів лінійних моделей та їх основні характеристики;

СТЬЮДРАСПОБР – критична точка t -критерія Стьюдента;

СРЗНАЧ – середнє (арифметичне) значення аргументів;

ДИСПР – дисперсія генеральної сукупності;

СТАНДОТКЛОН – стандартне відхилення за вибіркою.

10.5 Задачі для самостійної роботи

Задача 10.1. У таблиці наведено дані про кількість диких звірів, що водяться у мисливських угіддях України, тис. голів:

Таблиця 10.7

Звірі	2013	2014	2016	2017	2018
Білки	58,2	59,5	61,5	61,5	60,7
Бобри	13,2	15,0	14,6	17,8	19,5
Дикі кролики	0,2	0,2	0,1	0,3	0,3
Коти лісові	0,5	0,6	0,6	0,9	0,9
Лисиці	126,0	116,3	106,4	104,3	88,0

Обчислити базові та ланцюгові показники динаміки. Проаналізувати отримані результати.

Задачі 10.2–10.13: заповніть порожні клітинки; проаналізуйте отримані результати.

Задача 10.2. У таблиці вміщено дані про лісову площу, що зазнала пожеж, у Миколаївській області, та ланцюгові показники динаміки:

Таблиця 10.8

Рік	Лісова площа, що зазнала пожеж, га	Ланцюгові показники динаміки			
		Абсолютний приріст, га	Коефіцієнт зростання	Темп приросту, %	Абсолютне значення 1% приросту, га
2012	310	X	X	X	X
2013	225				
2014	152				
2015	204				
2016	272				
2017	185				

Задача 10.3. У таблиці наведено дані про загальну кількість муфлонів у розплідниках та фермах України і базові показники динаміки:

СТАТИСТИКА

Таблиця 10.9

Рік	Кількість муфлонів, голів	Базові показники динаміки		
		Абсолютний приріст, тис. одиниць	Темп зростання, %	Темп приросту, %
2013	79	0	0,0	0,0
2014	71			
2015	64			
2016	57			
2017	71			
2018	79			

Задача 10.4. У таблиці наведено дані про знешкоджені й вилучені тверді шкідливі речовини у Житомирській області (2013–2018 рр.) та ланцюгові показники динаміки:

Таблиця 10.10

Рік	Шкідливі речовини, тис. т	Ланцюгові показники динаміки			
		Абсолютний приріст, тис. т	Темп зростання, %	Темп приросту, %	Абсолютне значення 1 % приросту, тис. т
2013		X	X	X	X
2014					0,245
2015	8,0	-0,3			
2016					
2017			100,0		0,054
2018				-5,2	

Задача 10.5. У таблиці наведено дані про кількість готелів у Львівській області та базові показники динаміки:

Таблиця 10.11

Рік	Готелі	Базові показники динаміки		
		Абсолютний приріст, тис. одиниць	Коефіцієнт зростання	Темп приросту, %
2013	97	0	0,000	0,0
2014		5		
2015				
2016		10		-5,7
2017			1,578	
2018	115			

Розділ 10. РЯДИ ДИНАМІКИ ТА ЇХ АНАЛІЗ

Задача 10.6. У таблиці наведено дані про динаміку експорту товарів з України до Індії та Австралії (2013–2018 рр.) та ланцюгові показники динаміки:

Таблиця 10.12

Рік	Експорт, млн дол.	Ланцюгові показники динаміки			
		Абсолютний приріст, млн дол.	Коефіцієнт зростання	Темп приросту, %	Абсолютне значення 1 % приросту, млн дол.
2013		X	X	X	X
2014		17,5			0,115
2015	5,1				
2016			1,034		
2017					
2018				11,3	0,025

Задача 10.7. У таблиці наведено дані про дорожньо-транспортні пригоди на дорогах і вулицях упродовж 2013–2018 рр. (тис. одиниць):

Таблиця 10.13

Рік	Кількість ДТП, тис. одиниць	Базові показники динаміки		
		Абсолютний приріст, тис. одиниць	Темп зростання, %	Темп приросту, %
2013		0	0,0	0,0
2014				-2,8
2015			87,8	
2016	24,1			-3,8
2017		-1,04		
2018		3,11		

Задача 10.8. У таблиці подано деякі показники роботи пасажирського транспорту Миколаївської області:

Таблиця 10.14

Показник	2012	2013	2014	2015	2016	2017
Тролейбуси: перевезено пасажирів, млн осіб	50,1	51,1	50,5	47,3	35,0	287,7
з них оплатили за проїзд	43,6	36,8	33,0	27,6	24,7	25,0

СТАТИСТИКА

Закінчення табл. 10.14

Показник	2012	2013	2014	2015	2016	2017
Трамваї: перевезено пасажирів, млн осіб	89,0	78,2	68,0	58,8	55,8	61,1
з них оплатили за проїзд	72,5	49,8	33,7	37,1	38,4	41,2

Проаналізувати динаміку частки пасажирів, що не оплачували проїзд у тролейбусах та трамваях протягом 2012–2017 рр.

Задача 10.9. Динаміку виробництва товарів легкої промисловості в регіоні характеризують дані, наведені в таблиці:

Таблиця 10.15

Вид товару	2010	% до 2010	
		2014	2018
Тканини, млн м ²	685	131,3	125,4
Взуття, млн пар	58	105,8	102,7

Для кожного виду товарів обчислити середні абсолютні прирости виробництва за 2001–2005, 2006–2015 та 2001–2015 рр.

Задача 10.10. Залишок строкових депозитів фізичних осіб банку на початок 2010 р. становив 56 млн грн, юридичних осіб – 74 млн грн. Протягом першого місяця зміна обсягів депозитів цього банку подана в таблиці (млн грн):

Таблиця 10.16

Депозити	05.01	7.01	11.01	14.01	21.01	22.01	28.01
Фізичних осіб	+8	-11	-	+20	-	-	+15
Юридичних осіб	-	-	+18	+35	-24	-10	+21

Обчисліть середні залишки строкових депозитів фізичних та юридичних осіб банку в січні 2010 р. Поясніть вибір середньої.

Розділ 10. РЯДИ ДИНАМІКИ ТА ЇХ АНАЛІЗ

Задача 10.11. У таблиці наведено дані про поголів'я худоби в Україні (млн голів):

Таблиця 10.17

Дата	Велика рогата худоба	Свині	Вівці та кози
На 1 січня 2015 р.	6,16	5,44	1,28
На 1 лютого 2015 р.	5,87	5,33	1,29
На 1 березня 2015 р.	5,50	5,55	1,57
На 1 квітня 2015 р.	5,66	5,15	2,01
На 1 травня 2015 р.	5,73	5,28	2,06
На 1 червня 2015 р.	5,89	5,58	2,03
На 1 липня 2015 р.	5,51	5,73	2,11
На 1 серпня 2015 р.	5,53	6,08	1,99
На 1 вересня 2015 р.	5,23	6,27	1,94
На 1 жовтня 2015 р.	5,22	6,40	1,92
На 1 листопада 2015 р.	5,19	6,57	1,77
На 1 грудня 2015 р.	4,97	6,80	1,73
На 1 січня 2016 р.	4,50	5,94	1,54

Визначити поголів'я ВРХ, свиней, овець та кіз у середньому за квартали; півріччя; рік. Результати подати у вигляді таблиці. Пояснити вибір середньої.

Задача 10.12. У таблиці вміщено дані про вантажні перевезення в Україні:

Таблиця 10.18

Показник	Перевезено вантажів, млн т		Вантажообіг, млрд т/км	
	2015 р.	У % до 2010 р.	2015 р.	У % до 2010 р.
Усіма видами транспорту	809,3	512,9	420,6	116,9
У тому числі:				
залізничним	248,7	134,7	123,4	115,3
автомобільним	126,5	185	59,7	101,1
водним	34,4	114,6	15,9	107,8
трубопровідним	112,6	76,6	187,3	103,5

Для кожного виду транспорту визначте середні абсолютні прирости обсягу перевезених вантажів та вантажообігу за 2011–2015 рр. Результати розрахунків подайте у вигляді таблиці та проаналізуйте.

Задача 10.13. У таблиці показано, як один і той самий динамічний ряд проіндексували з різними базами порівняння:

Таблиця 10.19

База порівняння	2013	2014	2015	2016	2017
2012 р.	87,3	134,9	1275,8		
2018 р.			267,0	201,7	157,3

На підставі цієї інформації заповніть порожні клітинки.

10.6. Питання для самоконтролю

1. Сформулюйте означення ряду динаміки.
2. З яких елементів складається динамічний ряд?
3. Що означає змикання рядів динаміки?
4. Що означає зведення рядів динаміки до однієї основи?
5. Назвіть показники оцінки інтенсивності ряду динаміки.
6. Сформулюйте означення абсолютного приросту.
7. Сформулюйте означення коефіцієнту зростання.
8. Сформулюйте означення темпу зростання.
9. Сформулюйте означення темпу приросту.
10. Сформулюйте означення абсолютного значення 1-го % приросту.
11. Що таке абсолютне прискорення (уповільнення)?
12. Що таке відносне прискорення (уповільнення)?
13. Що показує абсолютний приріст?

14. Що показує коефіцієнт зростання?
15. Що показує темп зростання?
16. Що показує темп приросту?
17. Що показує абсолютне значення 1-го % приросту?
18. Про що свідчить абсолютне прискорення (уповільнення)?
19. Про що свідчить відносне прискорення (уповільнення)?
20. Сформулюйте означення середнього рівні інтервального ряду.
21. Сформулюйте означення середнього рівні моментного ряду.
22. Як визначити середній абсолютний приріст?
23. Як визначити середній коефіцієнт зростання?
24. Як визначити середній темп зростання?
25. Як визначити середній темп приросту?
26. Як визначити середнє абсолютне значення 1-го % приросту?

Розділ 11

ІНДЕКСИ

План

- 11.1. Суть і функції індексів.
- 11.2. Методологічні основи побудови зведених індексів.
- 11.3. Агрегатна форма індексів.
- 11.4. Середньозважені індекси.
- 11.5. Системи взаємозв'язаних індексів.
- 11.6. Індекси середніх величин.
- 11.7. Індексні ряди.
- 11.8. Територіальні індекси.
- 11.9. Задачі для самостійної роботи.
- 11.10. Питання для самоконтролю.

11.1. Суть і функції індексів

Термін *індекс* (*index*) є синонімом певної узагальнюючої характеристики. Наприклад, індекс реальних доходів населення за рік, індекс курсової вартості цінних паперів за місяць, регіональний індекс злочинності тощо. Кожний індекс являє собою співвідношення двох значень показника, який індексується: оціночного (поточного) і взятого за базу порівняння. Тобто за статистичною природою індекс – це відносна величина, яка характеризує зміну соціально-економічного явища в часі чи просторі або ступінь відхилення значення показника від певного стандарту (нормативу, середньої). Форми вираження індексу: коефіцієнти, проценти, проміле.

Історично індекси створювались як інструмент вивчення динаміки споживчих цін. Ще на початку XVII ст. англійський купець Т. Ман доводив переваги торгівлі з Індією, порівнюючи вартість товарів, які ввозились в Англію з Індії і Туреччини (шовк-сирець, прянощі тощо), за цінами індійськими та

турецькими. Індекс цін становив 0,33, тобто закупівля зазначених товарів в Індії утричі дешевша порівняно з Туреччиною. Різницю вартостей Т. Ман визначив як суму економії від заміни торгового партнера. Такого роду розрахунки й досі вважаються найбільш логічним вираженням індексів.

Поступово коло показників, що піддавались індексному аналізу, розширювалось, а методи аналізу вдосконалювались. За допомогою індексів характеризують зміну (співвідношення) обсягів виробництва, чисельності зайнятих, заробітної плати, собівартості продукції тощо.

Означення. Індекс – це відносна величина, яка характеризує зміну соціально-економічного явища в часі чи просторі або ступінь відхилення значення показника від певного стандарту (нормативу, середньої). Індекс, як і будь-який статистичний показник, поєднує в собі якісний та кількісний аспекти. Назва індексу відбиває соціально-економічний зміст показника, числове його значення – інтенсивність змін або ступінь відхилення (індекс урожайності – 1,07, індекс доходів населення – 1,15 тощо).

За характером порівнянь (у часі, просторі, з певним стандартом) індекси поділяють на динамічні, територіальні, міжгрупові. **Динамічний індекс** характеризує інтенсивність динаміки; при його розрахунку базою порівняння є одне з попередніх значень показника. База порівняння ідентифікується підрядковою позначкою «0», поточне значення показника – «1».

При *просторових порівняннях* визначають ступінь відхилення значень показника у просторі – між об'єктами, країнами, регіонами, які ідентифікують певними літерами; вибір бази порівняння довільний.

Міжгруповий індекс характеризує відхилення від певного стандарту (еталонного, максимального чи мінімального значення) або від середнього рівня по сукупності загалом.

За ступенем охоплення елементів сукупності вирізняють індекси *індивідуальні та зведені*. Їх позначають символами *i*

та I відповідно. **Індивідуальні індекси** характеризують співвідношення рівнів показника окремих елементів сукупності, *зведені* – певної множини елементів. У структурованій сукупності зведений індекс може бути груповим (субіндексом) або загальним (тотальним). Наприклад, в ієрархії динамічних індексів промислового виробництва динаміку обсягів окремих видів продукції (чавун, електроенергія, тепловози) характеризують індивідуальні індекси, окремих галузей промисловості (металургія, енергетика, машинобудування) – субіндекси, промисловості загалом – загальний індекс.

Показник, динаміку чи співвідношення якого характеризує індекс, називають *індексованою величиною*, йому надається певний символ. Наприклад, індивідуальний індекс фізичного обсягу продукції позначають i_q , зведений індекс цін – I_p . Символи p та q не випадкові, вони відповідають початковим літерам англійських слів *price* (ціна) та *quantity* (кількість).

Індивідуальний індекс – це відносна величина динаміки

$$i_p = \frac{p_1}{p_0} \quad (11.1)$$

або порівняння

$$i_q = \frac{q_A}{q_B} \quad (11.2)$$

Неодмінною умовою їх обчислення є порівнянність методики вимірювання чисельника та знаменника відношення, що являє собою індекс.

Щодо зведених індексів, то методика їх розрахунку (модель) залежить від мети дослідження, змісту і статистичної природи показників, наявної інформації. Мета

дослідження визначає, зокрема, функцію, яку виконує індекс у конкретному аналізі, та характер порівнянь. Вирізняють дві функції індексів:

- синтетичну, пов'язану з побудовою узагальнюючих характеристик динаміки чи просторових порівнянь;
- аналітичну, спрямовану на вивчення закономірностей динаміки, взаємозв'язків між показниками, структурних зрушень.

Як узагальнюючий показник індекс виражає відносну зміну складного соціально-економічного явища, елементи якого безпосередньо не підсумовуються (фізичні обсяги товарів, що мають різні одиниці вимірювання, собівартість різних видів продукції тощо). Тож для побудови зведених індексів виникає необхідність зведення несумірних елементів до одного виду. Аналітична функція індексів полягає в оцінюванні впливу факторів чи структурних зрушень на динаміку індексованого показника, вона реалізується також у порівняльному аналізі стану і розвитку об'єктів дослідження.

Очевидно, що синтетична та аналітична функції індексів взаємозв'язані. Часто один і той самий індекс виконує обидві функції. Наприклад, індекс споживчих цін за рік становив 1,025. З одного боку, він характеризує середній приріст цін на 2,5%, а з іншого – свідчить про те, що за рахунок зростання цін вартість споживчого кошика зросла на 2,5%.

Зведені індекси широко використовують в практиці статистичних досліджень. За допомогою індексів характеризують розвиток національної економіки в цілому, окремих регіонів і видів економічної діяльності, аналізують результати господарської діяльності підприємств і організацій, досліджують роль окремих факторів у формуванні найважливіших економічних показників, індекси використовують у міжнародних порівняннях, в оцінюванні рівня життя населення, в моніторингу ділової активності, кон'юнктури ринку і т. ін.

11.2. Методологічні основи побудови зведених індексів

Зведений індекс свідчить як у середньому змінився показник по сукупності елементів. Методологічною основою побудови зведених індексів є агрегування, узагальнення інформації. Нехай за даними про ціни на товари ($j = 1, 2, 3, \dots$) за два періоди ($t = 0, 1$) необхідно визначити зведений індекс цін: $(p_{i0}, p_{20}, p_{30}, \dots, p_{n0})$ та $(p_{11}, p_{21}, p_{31}, \dots, p_{n1})$.

Узагальнення інформації для зведеного індексу цін I_p можна здійснити різними способами. Історія розвитку індексного методу свідчить про пошук таких способів побудови зведеного індексу, який би характеризував зміну цін певного набору товарів у середньому. До 20-х років ХХ ст. у світовій статистиці домінували незважені індекси, зокрема:

1. У вигляді відношення сум цін (індекс Дюто, 1735 р.):

$$I_p = \frac{\sum_1^n p_{j1}}{\sum_1^n p_{j0}} = \frac{\bar{p}_{j1}}{\bar{p}_{j0}} \quad (11.3)$$

Поділивши чисельник і знаменник на n , цей індекс можна подати як відношення середніх цін.

2. Як середня арифметична з індивідуальних індексів цін

$$i_p = \frac{p_{j1}}{p_{j0}} \quad (\text{індекс Карлі, 1751 р.):}$$

$$I_p = \frac{\sum_1^n \frac{p_{j1}}{p_{j0}}}{n} = \frac{\sum_1^n i_p}{n} \quad (11.4)$$

3. Як середня геометрична з індивідуальних індексів цін (Джевон, 1863 р.):

$$I_p = \sqrt[n]{\frac{p_{11} p_{21} \dots p_{n1}}{p_{10} p_{20} \dots p_{n0}}} \quad (11.5)$$

У незважених індексах усім елементам, незалежно від обсягу проданих однойменних товарів, надаються однакові ваги, в результаті зведений індекс неадекватно характеризує сукупну зміну цін. Як не можна визначити середню ціну, наприклад, 1 л вина та 1 кг хліба, так само без урахування вагомості окремих товарів неможливо усереднити індивідуальні індекси цін: якщо ціни на хліб зросли на 10%, а на вино – на 60%, то це не означає, що в середньому ціни зросли на 35%.

Отже, при розрахунку зведеного індексу кожному j -му елементові необхідно надати певну «вагу», яка б урахувувала його відносну значущість. Очевидно, що у ваг має бути однакова розмірність. Кількості товару q_j у натуральних одиницях вимірювання не можуть виконувати вагову функцію, вона покладається на вартість товарів $q_j p_j$.

Реалізація ідеї «зважування» індексів стала важливим етапом удосконалення індексної методології. Цей етап пов'язують з іменами німецьких статистиків Ласпейреса і Пааше (кінець XIX ст.), які вперше надали економічний зміст своїм індексам цін: замість простого підсумовування цін використали загальну вартість проданих товарів базисного і поточного періодів. Різні форми зважених індексів широко використовують донині. Припустимо, у базисному періоді вартості товарів становили $q_{i0} p_{i0}$, тоді зважений індекс цін набирає вигляду:

$$I_p = \frac{\sum_1^n \frac{p_{j1}}{p_{j0}} q_{j0} p_{j0}}{\sum_1^n q_{j0} p_{j0}} = \frac{\sum_1^n p_{j1} q_{j0}}{\sum_1^n p_{j0} q_{j0}} \quad (11.6)$$

Це дві форми одного індексу цін – середньозважена та агрегатна. Основою середньозваженої форми є індекс Карлі, агрегатної – індекс Дюто.

Аналогічна ситуація виникає при агрегуванні фізичного обсягу продукції (товарної маси, кількості), якщо елементи

сукупності не сумірні між собою, тобто не мають спільної міри: $(q_{10}, q_{20}, q_{30}, \dots, q_{n0})$ та $(q_{11}, q_{21}, q_{31}, \dots, q_{n1})$.

У такому разі їх агрегування можливе лише за умови зведення різнорідних елементів до одного порівнянного виду за допомогою певних сумірників. Найчастіше сумірниками є ціна або собівартість одиниці продукції. Наприклад, при визначенні обсягів проданих на біржі товарів – металу, лісоматеріалів, цементу – сумірниками виступають ціни.

Обсяги продажу записують як $\sum_1^n q_j p_j$. Якщо за сумірник узяти собівартість (c), агрегат інтерпретується як грошові витрати і набирає вигляду $\sum_1^n q_j c_j$.

Іноді зручніший трудовий сумірник. Приміром, для визначення обсягів виконаних у сільському господарстві робіт (оранка, боронування, сівба тощо) як сумірник використовують трудомісткість, тобто затрати праці на одиницю виконаної роботи (позначимо t). Тоді загальний обсяг трудових затрат буде $\sum_1^n q_j t_j$.

Зведений індекс фізичного обсягу I_q також має дві форми – середньозважену та агрегатну:

$$I_q = \frac{\sum_1^n \frac{q_{j1}}{q_{j0}} q_{j0} p_{j0}}{\sum_1^n q_{j0} p_{j0}} = \frac{\sum_1^n q_{j1} p_{j0}}{\sum_1^n q_{j0} p_{j0}} \quad (11.7)$$

Як бачимо з формул наведених індексів I_p та I_q , їх значення залежить від структури агрегату і динаміки окремих його елементів. Тому для визначення зведеного індексу необхідно передусім обґрунтувати коло елементів, які підлягають агрегуванню в одне ціле. Наприклад, обґрунтовується споживчий кошик (набір споживчих товарів) при обчисленні

індексу «вартості життя», перелік товарів і товарних груп при обчисленні індексу оптових цін тощо.

Підсумовування здійснюють по всьому визначеному колу елементів як у чисельнику, так і в знаменнику, а тому надалі для простоти викладу навчального матеріалу у формулах зведених індексів ідентифікаційні позначки елементів j не вказуються.

11.3. Агрегатна форма індексів

Означення. *Агрегатний індекс* – це співвідношення двох агрегатів, конкретних щодо змісту й часу. *Агрегат* являє собою добуток спряжених величин. Одна з цих величин індексована – у чисельнику і знаменнику вона в різних періодах, інша є вагою чи сумірником індексованої величини і фіксується на одному й тому самому рівні.

Так, в індексі цін індексується ціна p , а кількість q становить вагу ціни і фіксується на одному й тому ж рівні; в індексі фізичного обсягу продукції індексується кількість q , а сумірник – ціна p – фіксується:

<p>• індекс цін</p> $I_p = \frac{\sum p_1 q}{\sum p_0 q} \quad (11.8)$	<p>• індекс фізичного обсягу</p> $I_q = \frac{\sum q_1 p}{\sum q_0 p} \quad (11.9)$
--	---

Ваги в індексі цін і сумірники в індексі фізичного обсягу можна фіксувати на рівні як базисного, так і поточного періоду. Для ілюстрації варіантів зважування використаємо матрицю агрегатів (рис. 11.1).

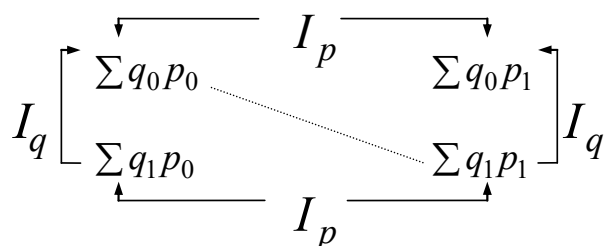


Рис. 11.1. Схема співвідношення агрегатів

На головній діагоналі матриці розміщено фактичні вартості товарів, на побічній – перехресні (умовні). По горизонталі розміщені агрегати з фіксованими вагами: у першому – на рівні базисного періоду, у другому – на рівні поточного. По вертикалі – агрегати з фіксованими сумірниками: у першому – на рівні базисного періоду, у другому – на рівні поточного. Порівняння агрегатів дає дві системи індексів – базисно-зважену (Ласпейреса) та поточно-зважену (Пааше).

У базисно-зваженій системі перехресні агрегати побічної діагоналі порівнюються з базисним агрегатом головної діагоналі $\sum q_0 p_0$, у поточно-зваженій системі агрегат головної діагоналі $\sum q_1 p_1$ порівнюється з перехресними агрегатами побічної діагоналі. Схематично системи зважування показані на рис. 11.1, а формули індексів наведені у табл. 11.1.

Таблиця 11.1

**Формули індексів цін і фізичного обсягу
за різних систем зважування**

Базисно-зважена система (Ласпейреса)	Поточно-зважена система (Пааше)
$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}$	$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$
$I_q = \frac{\sum p_0 q_1}{\sum q_0 p_0}$	$I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1}$

Джерело: опрацьовано авторами на основі [2, 297].

Обидві системи індексів є рівноправними. Реальний економічний зміст мають не лише чисельник і знаменник індексу, а й різниця між ними. Вибір форми індексу залежить від мети дослідження та наявної інформації. Так, у зарубіжній статистиці індекс цін розраховують за Ласпейресом, оскільки він ґрунтується на даних про обсяги, отримані з переписів, вибіркового обстежень домогосподарств, інших джерел за минулий період. У вітчизняній статистиці при розрахунках

Розділ 11. ІНДЕКСИ

індексів цін перевага надавалася формулі Пааше, оскільки визначальним показником була вартість поточного періоду. Індекс фізичного обсягу товарної маси, навпаки, обчислюють за формулою Ласпейреса з фіксованими сумірниками на рівні базисного періоду. У такому разі динаміка цінового фактора не впливає на величину індексу. Зауважимо, що в разі незначної кореляції між цінами та товарною масою індекси, розраховані за Ласпейресом і Пааше, практично однакові.

Приклад 11.1. Розглянемо порядок розрахунку агрегатних індексів за даними про ціни та обсяги продажу агропродукції через біржу (табл. 11.2). У цьому прикладі агрегатами виступають фактичні за кожний місяць та умовні обсяги торговельних оборотів біржі.

Таблиця 11.2

До розрахунку агрегатних індексів цін і фізичного обсягу

Продукція	Реалізовано, тис. т		Ціна за 1 т, тис. грн		Агрегати (торговельні обороты, млн грн)			
	серпень	вересень	серпень	вересень	$q_0 p_0$	$q_1 p_0$	$q_1 p_1$	$q_0 p_1$
	q_0	q_1	p_0	p_1				
Крупа гречана	22	28	30	25	660	840	700	550
Крупа пшенична	18	20	23	24	414	460	480	432
Олія	7	8	28	26	196	224	208	182
Разом	×	×	×	×	1270	1524	1388	1164

Розв'язання. За даними таблиці зведені індекси цін I_p та фізичного обсягу I_q реалізованої через біржу агропродукції становлять:

За Ласпейресом:	За Пааше:
$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{1164}{1270} = 0,91;$	$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{1388}{1524} = 0,91$
$I_q = \frac{\sum p_0 q_1}{\sum q_0 p_0} = \frac{1524}{1270} = 1,2;$	$I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{1388}{1164} = 1,2.$

Висновок. Отже, біржові ціни на агропродукцію у вересні порівняно із серпнем зменшились у середньому на 9%, реалізована товарна маса зросла в середньому на 20%.

Оскільки в структурі торговельного обороту не відбулося значних змін, то розбіжності між індексами, обчисленими за різними системами зважування, неістотні. Будь-який з розрахованих індексів має певний ступінь умовності.

За наявності структурних зрушень у торговельних оборотах використовують індекси із середніми вагами або усереднення різнозважених індексів за допомогою середньої геометричної, наприклад, індекс цін

$$I_p = \sqrt{\frac{\sum p_1 q_1}{\sum p_0 q_1} \times \frac{\sum p_1 q_0}{\sum p_0 q_0}} \quad (11.10)$$

Спираючись на формально-математичні критерії, яким відповідає усереднений індекс, І. Фішер назвав його «ідеальним», проте через відсутність конкретного економічного змісту цей індекс не набув широкого практичного використання.

11.4. Середньозважені індекси

Другою формою зведеного індексу є *середньозважений з індивідуальних індексів*. Використовують два види середніх – арифметичну та гармонічну. Вибір виду середньої ґрунтується на загальних засадах: середньозважений індекс має бути тотожним відповідному індексу агрегатної форми.

Розділ 11. ІНДЕКСИ

Подамо інформацію про біржові торги агропродукцією обсягами торговельного обороту (у серпні – $\sum q_0 p_0$, у вересні – $\sum q_1 p_1$) та індивідуальними індексами цін i_p і фізичного обсягу продажу i_q (табл. 11.3).

Таблиця 11.3

Розрахунок середньозважених індексів цін і фізичного обсягу

Товар	Торговельний оборот, млн. грн.		Індивідуальні індекси		Умовний агрегат	
	серпень	вересень	цін	фізичного обсягу	$i_q q_0 p_0$	$\frac{q_1 p_1}{i_p}$
	$q_0 p_0$	$q_1 p_1$	i_p	i_q		
Крупа гречана	660	700	0,8333	1,2727	839,982	840,034
Крупа пшенична	414	480	1,0435	1,1111	459,995	459,990
Олія	196	208	0,9286	1,1429	224,008	223,993
Разом	1270	1388	×	×	1523,985	1524,017

Умовний торговельний оборот $\sum q_1 p_0$ можна визначити, скоригувавши фактичні обороти індивідуальними індексами цін або фізичного обсягу продажів:

$$i_q q_0 p_0 = q_1 p_0 = \frac{1}{i_p} q_1 p_1 \quad (11.11)$$

У такому разі зведені індекси за Ласпейресом обчислюють як середньоарифметичну з вагами $q_0 p_0$, а індекси за Пааше – як середньогармонічну з вагами $q_1 p_1$:

$$I_p = \frac{\sum i_p p_0 q_0}{\sum p_0 q_0} \quad (11.12)$$

$$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} \quad (11.14)$$

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{1}{i_p} q_1 p_1} \quad (11.13)$$

$$I_q = \frac{\sum q_1 p_1}{\sum \frac{1}{i_q} q_1 p_1} \quad (11.15)$$

Обчислимо середньозважені індекси цін I_p та фізичного обсягу продажів I_q за даними табл. 11.3, використовуючи різні варіанти зважування (I_p – за Пааше, I_q – за Ласпейресом):

- середньозважений індекс цін:

$$I_p = \frac{700 + 480 + 208}{\frac{700}{0,8333} + \frac{480}{1,0435} + \frac{280}{0,9286}} = \frac{1388}{1524,017} = 0,91075 \approx 0,91;$$

- середньозважений індекс фізичного обсягу продажів:

$$I_q = \frac{1,2727 \cdot 660 + 1,1111 \cdot 414 + 1,1429 \cdot 196}{660 + 414 + 196} = \frac{1523,985}{1270} = 1,19998 \approx 1,2.$$

Як бачимо, значення середньозважених індексів такі ж, як і відповідних їм агрегатних індексів (див. табл. 11.2).

При побудові середньозважених індексів вартісні ваги можна замінити відносними величинами структури d , сума яких $\sum d = 1$. У цьому разі середньозважені індекси набирають вигляду:

$$I_q = \sum i_p d_0 \quad (11.16) \quad I_p = \frac{1}{\sum \frac{1}{i_p} d_1} \quad (11.17)$$

Ці формули підтверджують залежність значення зведеного індексу від динаміки окремих складових і структури сукупності.

Приклад 11.2. У регіоні виробництво споживчих товарів зменшилося: продовольчих – на 5, непродовольчих – на 10%, а ціни зросли на 6% і 8% відповідно. Внаслідок нерівномірності динаміки виробництва по групах споживчих товарів змінилась їх структура: на 2 відсоткові пункти зросла частка продовольчих товарів і на стільки ж зменшилася частка непродовольчих. Розрахунки зведених індексів фізичного обсягу та цін наведено у табл. 11.4.

Таблиця 11.4

**Розрахунок середньозважених індексів
з відносними вагами**

Товарні групи	Структура виробництва		Індивідуальні індекси		Розрахункові величини	
	d_0	d_1	i_q	i_p	$i_q d_0$	$\frac{d_1}{i_p}$
Продовольчі	0,62	0,64	0,95	1,06	0,589	0,604
Непродовольчі	0,38	0,36	0,90	1,08	0,342	0,333
Разом	1,00	1,00	×	×	0,931	0,937

Зведений індекс фізичного обсягу виробництва

$$I_q = 0,95 \cdot 0,62 + 0,90 \cdot 0,38 = 0,931,$$

тобто в середньому обсяги виробництва зменшилися на 6,9%.

Зведений індекс цін $I_p = 1,067$ показує, що ціни в середньому зросли на 6,7 %:

$$I_p = \frac{1}{\frac{0,64}{1,06} + \frac{0,36}{1,08}} = 1,067.$$

Зауваження. Середньозважені індекси мають перевагу над агрегатними, адже за їхньою допомогою можна виши-

кувати ієрархію індексів від індивідуальних на окремі товари через групові (субіндекси) до загального по всій сукупності елементів. Проте їм властиві й недоліки. Якщо динаміка окремих складових сукупності є протилежною, то зведений індекс не в змозі адекватно відобразити закономірність динаміки. Окрім того, середньозважений індекс визначається лише за порівнянним колом елементів. Коли ж окремі елементи сукупності відсутні в базисному чи поточному періоді, то розрахунок індивідуальних індексів неможливий. У цьому разі перевага надається індексу агрегатної форми.

Отже, за кожним індексом стоїть певне економічне явище, що зумовлює методику його розрахунку та змістовність.

11.5. Системи взаємозв'язаних індексів

Розглянуті зведені індекси узагальнюють динаміку складних сукупностей. Не менш важлива в статистичному аналізі друга функція індексів – аналітична, яка спирається на взаємозв'язок індексів. Практично кожний індекс є складовою певної індексної системи, а його зв'язки з іншими індексами цієї системи відбивають зв'язки між відповідними показниками. Так, товарооборот залежить від фізичного обсягу проданого товару q і цін p , відповідно індекс товарообороту I_{qp} являє собою добуток індексів фізичного обсягу і цін:

$$I_{qp} = I_q I_p \quad (11.18)$$

Аналогічно грошові витрати на виробництво можна подати як функцію фізичного обсягу виробництва q і собівартості c , тобто, $I_{qc} = I_q \cdot I_c$; обсяг виробництва – як функцію трудових затрат t та продуктивності праці w , тобто $I_q = I_t \cdot I_w$ і т. ін.

Таким чином, у будь-якій системі індекс добутку спряжених величин дорівнює добутку індексів цих величин.

У рамках такої індексної системи на основі двох індексів можна визначити третій.

Приклад 11.3. Витрати на виробництво зросли на 8,2%, а фізичний обсяг виробленої продукції – на 5%, то собівартість одиниці продукції зросла в середньому на 3%:

$$I_c = I_{qc} : I_q = 1,082 : 1,05 = 1,03.$$

Взаємозв'язані також індекси прямих і обернених показників, наприклад, споживчих цін і купівельної спроможності грошової одиниці, продуктивності праці й трудомісткості продукції тощо.

Приклад 11.4. Споживчі ціни зросли на 5,6%, то купівельна спроможність грошової одиниці зменшилася на 5,3%:

$$I_{1/p} = 1 : I_p = 1 : 1,056 = 0,947.$$

Показники-співмножники індексної системи є факторами показника-результату, а динаміка їх детермінує динаміку останнього. Отже, у межах індексної системи можна визначити роль кожного окремого фактора, оцінити його вплив на динаміку результату. Така оцінка ґрунтується на методі абстракції. Для того, щоб виявити вплив одного фактора, необхідно абстрагуватися від впливу іншого, зафіксувати його на постійному рівні. Проте постає питання: на рівні якого періоду – базисного чи поточного? Теоретично можливі два варіанти.

Перший варіант: коли обидва індекси-співмножники базисно-зважені, кожний з них оцінює окремий вплив, оцінки впливу порівнянні. Проте цей варіант не забезпечує пов'язування індексів у систему:

$$\frac{\sum q_1 p_0}{\sum q_0 p_0} \cdot \frac{\sum q_0 p_1}{\sum q_0 p_0} \neq \frac{\sum q_1 p_1}{\sum q_0 p_0}. \quad (11.19)$$

У другому варіанті індекси-співмножники різнозважені: ваги одного з них фіксуються на рівні базисного періоду, іншого — на рівні поточного. Через різнозваженість індексів оцінки впливу факторів непорівнянні, але саме такий порядок абстрагування впливу факторів забезпечує взаємозв'язок індексної системи:

$$\frac{\sum q_1 p_0}{\sum q_0 p_0} \cdot \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{\sum q_1 p_1}{\sum q_0 p_0}$$

або

$$\frac{\sum q_1 p_1}{\sum q_0 p_1} \cdot \frac{\sum q_0 p_1}{\sum q_0 p_0} = \frac{\sum q_1 p_1}{\sum q_0 p_0} \quad (11.20)$$

У розглянутому прикладі (див. табл. 11.2) результативний показник індексної системи – біржовий оборот. Його індекс обчислюється як відношення фактичних вартостей поточного періоду $\sum q_1 p_1 = 1\,388$ і базисного $\sum q_0 p_0 = 1\,270$:

$$I_{qp} = \frac{1\,388}{1\,270} = 1,092,$$

тобто біржовий оборот зріс на 9,3%. Цей індекс можна записати як добуток індексів фізичного обсягу продажів і цін:

$$1,092 = 1,2 \cdot 0,91$$

Висновок. Отже, зведені індекси цін I_p і товарної маси I_q , маючи самостійне значення, водночас виконують аналітичну функцію – оцінюють вплив відповідного фактора на динаміку біржового обороту. Ступінь впливу факторів на результат характеризують темпи приросту факторів. У розглянутому прикладі за рахунок збільшення товарної маси біржовий оборот зріс на 20%, зниження цін призвело до зменшення біржового обороту на 9%.

Розділ 11. ІНДЕКСИ

У межах індексної системи можна визначити також абсолютний вплив факторів на приріст результату. Абсолютний приріст біржового обороту

$$\Delta qp = \sum q_1 p_1 - \sum q_0 p_0 \quad (11.21)$$

У прикладі 11.1:

$$\Delta qp = 1\,388 - 1\,270 = 118 \text{ млн грн.}$$

Він спричинений обома факторами, тобто товарною масою Δq і цінами Δp :

$$\Delta qp = \Delta q + \Delta p \quad (11.22)$$

Абсолютний вплив кожного фактора окремо визначають як різницю між чисельником і знаменником відповідного індексу:

$$\Delta q = \sum (q_1 - q_0) p_0 = \sum q_1 p_0 - \sum q_0 p_0 \quad (11.23)$$

$$\Delta p = \sum (p_1 - p_0) q_1 = \sum p_1 q_1 - \sum p_0 q_1 \quad (11.24)$$

Згідно з даними табл. 11.2, у млн грн:

$$\Delta q = 1524 - 1270 = 254;$$

$$\Delta p = 1388 - 1524 = -136;$$

$$\Delta qp = 254 - 136 = 118.$$

Якщо абсолютний вплив факторів односпрямований, можна визначити питому вагу кожного фактора. За різноспрямованих впливів такі розрахунки не мають сенсу.

Коли факторів три і більше, передусім необхідно визначити їх послідовність, враховуючи суть кожного з них, порядок розрахунку, взаємозв'язок у системі. Наприклад,

$y = abc$. Припустимо, що результативний показник y – відносна величина. Тоді першим фактором-співмножником буде той, чисельник розрахункової формули якого є чисельником результативного показника; у наступного фактора-співмножника чисельник розрахункової формули – знаменник першого фактора і т.д.

Наприклад, y – прибутковість власного капіталу фірми, a – прибутковість поточних активів, b – коефіцієнт поточної ліквідності, c – частка поточних пасивів у власному капіталі. Згідно з розрахунковими формулами послідовність факторів у системі така:

$$\frac{\text{Чистий прибуток}}{\text{Власний капітал}} = \frac{\text{Чистий прибуток}}{\text{Поточні активи}} \times \frac{\text{Поточні активи}}{\text{Поточні пасиви}} \times \frac{\text{Поточні пасиви}}{\text{Власний капітал}} .$$

Отже,

$$I_y = I_a \cdot I_b \cdot I_c. \tag{11.25}$$

Ваги в індексах-співмножниках фіксуються за схемою: в індексі першого фактора – на рівні базисного періоду, в індексі другого фактора – ті, що праворуч від індексованої величини, на рівні базисного періоду, ті, що ліворуч, – на рівні поточного періоду, в індексі третього фактора – всі ваги фіксуються на рівні поточного періоду (вони розміщені ліворуч від індексованої величини).

У символах система зважування факторів має такий вигляд:

$$\frac{a_1 b_1 c_1}{a_0 b_0 c_0} = \frac{a_1 b_0 c_0}{a_0 b_0 c_0} \cdot \frac{a_1 b_1 c_0}{a_1 b_0 c_0} \cdot \frac{a_1 b_1 c_1}{a_1 b_1 c_0} \quad (11.26)$$

Як і в двофакторній індексній системі, абсолютний вплив зміни будь-якого фактора на динаміку результату визначається як різниця між чисельником і знаменником відповідного індексу. Тотожні оцінки абсолютного впливу факторів дає ланцюговий метод, який ґрунтується на умовних значеннях результативного показника. Замінімо $a_0 b_0 c_0 = y_0$. Тоді $i_a y_0$ – значення результативного показника за умови, що на динаміку останнього впливає лише фактор a . Різниця $i_a y_0 - y_0$ характеризує абсолютний приріст Y за рахунок фактора a . Аналогічно визначають абсолютний вплив інших факторів:

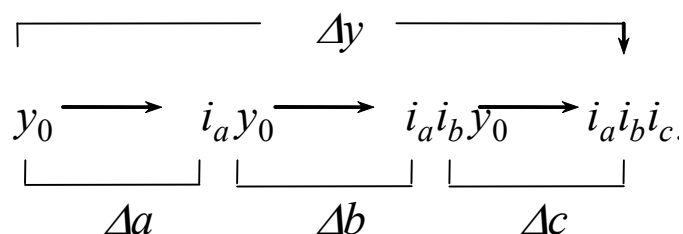


Рис. 11.2. Схема визначення абсолютного впливу інших факторів

Джерело: опрацьовано авторами на основі [2, 308].

Очевидно, що:

$$\Delta y = \Delta a + \Delta b + \Delta c \quad (11.27)$$

Приклад 11.5. Прибутковість власного капіталу зменшилась з 36% у базисному періоді до 27,37% у поточному, тобто на 8,63 відсоткові пункти. Індекс прибутковості становить: $I_y = 27,37 : 36 = 0,76$. За той же період прибутковість активів зменшилась на 10%, поточна ліквідність – на 12%, частка поточних пасивів у власному капіталі – на 4%. Розра-

хунок абсолютного впливу факторів на динаміку прибутковості капіталу подано у табл. 11.5.

Таблиця 11.5

**Розрахунок абсолютного впливу факторів на динаміку
результативного показника**

Фактор	Індекс	Розрахун- кова величина	Абсолютний вплив фактора, в. п.
<i>a</i>	0,90	$36,0 \cdot 0,90 =$ 32,4	$32,4 - 36,0 = -3,6$
<i>b</i>	0,88	$32,4 \cdot 0,88 =$ 28,51	$28,51 - 32,4 = -3,89$
<i>c</i>	0,96	$28,51 \cdot 0,96$ $= 27,37$	$27,37 - 28,51 = -$ 1,14
Разом	×	×	-8,63

Індексна система має вигляд: $0,90 \cdot 0,88 \cdot 0,96 = 0,76$.

За даними таблиці найвагомійший вплив на зменшення прибутковості капіталу виявив фактор *b* – поточна ліквідність, дещо менший вплив фактора *a* (прибутковість активів) і відносно невеликий вплив чинив фактор *c* (частка поточних пасивів у власному капіталі).

11.6. Індекси середніх величин

Поряд зі зведеними, агрегатними індексами в статистичній практиці широко використовують *індекси середніх величин* (індекси середньої заробітної плати, середньої уро-

жайності тощо). Як відомо, рівень середньої залежить від значень ознаки x_j і структури сукупності:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \sum_1^m x_j d_1 \quad (11.28)$$

де f_j – частота;

d_j – частка j -ї складової сукупності.

Виходячи з цього, динаміка середньої визначається цими факторами:

а) зміною значень ознаки x_j і б) структурними зрушеннями. Вплив кожного з них на динаміку середньої оцінюють за допомогою системи індексів середніх величин: змінного й фіксованого складу, а також структурних зрушень. У наведених формулах індексів ідентифікація складових сукупності відсутня.

Індексом змінного складу $I_{\bar{x}}$ називають *індекс середньої величини*, він відбиває не лише зміни значень ознаки x , а й зміни в структурі сукупності:

$$I_{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} = \sum x_1 d_1 : \sum x_0 d_0 \quad (11.29)$$

В **індексі фіксованого складу** I_x ваги постійні, тобто усувається вплив на динаміку середньої структурних зрушень. Величина I_x показує, як у *середньому* змінилися значення ознаки за незмінної, фіксованої структури:

$$I_x = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1} = \sum x_1 d_1 : \sum x_0 d_1 \quad (11.30)$$

Індекс структурних зрушень I_d , навпаки, показує, як змінилася середня за рахунок структурних зрушень; значення ознаки x фіксуються на постійному рівні:

$$I_d = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} = \sum x_0 d_1 : \sum x_0 d_0 \quad (11.31)$$

У кожній конкретній індексній системі I_d оцінює вплив на динаміку середньої того структурного фактора, який є основою поділу сукупності на складові.

Формули індексів фіксованого складу і структурних зрушень різнозважені: в I_x ваги фіксуються на рівні поточного періоду, в I_d – значення ознаки x – на рівні базисного періоду. Саме такий варіант зважування забезпечує пов'язування цих індексів у систему:

$$I_x = I_x \cdot I_d \quad (11.32)$$

Розглянемо систему індексів середніх величин з табл. 11.6. Є дані про середню заробітну плату двох категорій найманих працівників компанії: менеджери з продажу та обслуговуючий персонал:

Таблиця 11.6

Категорії працівників	Базисний період		Поточний період		Індекс заробітної плати, i_x
	середня заробітна плата, грн x_0	кількість працівників, у % до підсумку f_0	середня заробітна плата, грн x_1	кількість працівників, у % до підсумку f_1	
Менеджери з продажу	10 000	60	12 000	80	1,2
Обслуговуючий персонал	5 000	40	7 200	20	1,44
Разом	×	100	×	100	-

Розділ 11. ІНДЕКСИ

За поточний період заробітна плата одного менеджера зросла на 2% ($i_x = 1,2$), працівника обслуговуючого персоналу – на 44% ($i_x = 1,44$). Водночас змінилася структура працівників: на 20 в.п., зменшилася частка працівників обслуговуючого персоналу, на стільки ж пунктів зросла частка більш оплачуваних менеджерів. Середня заробітна плата у базисному періоді:

$$\bar{x} = \frac{10000 \cdot 60 + 5000 \cdot 40}{100} = 8000 \text{ грн,}$$

у поточному періоді:

$$\bar{x} = \frac{12000 \cdot 80 + 7200 \cdot 20}{100} = 11040 \text{ грн}$$

Індекс змінного складу показує як змінилася середня заробітна плата найманих працівників:

$$I_{\bar{x}} = \frac{\bar{x}_1}{x_0} = \frac{11040}{8000} = 1,38,$$

тобто середня заробітна плата зросла на 38%, що значно перевищує зростання заробітної плати по кожній категорії.

Такий парадоксальний результат пояснюється тим, що на динаміку середньої заробітної плати вплинула не лише динаміка заробітної плати окремих категорій, а й зміни в структурі працівників. Зафіксувавши структуру працівників на одному й тому ж (поточному) рівні, визначимо, якою б була середня заробітна плата:

$$\bar{x}_0^1 = \frac{10000 \cdot 80 + 5000 \cdot 20}{100} = 9000 \text{ грн}$$

Порівнявши середню заробітну плату поточного періоду (11040) з умовною (9000), знайдемо індекс фіксованого складу:

$$I_{\bar{x}} = \frac{\bar{x}_1}{x_0^1} = \frac{11040}{9000} = 1,227 \approx 1,23.$$

Тобто в середньому заробітна плата зросла на 23%.

Індекс фіксованого складу I_x – це середня з індивідуальних індексів значень ознаки (у нашому прикладі –

заробітної плати). Він тотожний середньозваженому гармонічному індексу з індивідуальних індексів заробітної плати:

$$I_x = \frac{12000 \cdot 0,8 + 7200 \cdot 0,2}{\frac{12000 \cdot 0,8}{1,2} + \frac{7200 \cdot 0,2}{1,44}} = \frac{11040}{9000} = 1,227 \approx 1,23.$$

Щоб оцінити, як вплинули на середню заробітну плату структурні зрушення (збільшення частки менеджерів з більшою заробітною платою), визначимо індекс структурних зрушень:

$$I_d = 9000 : 8000 = 1,125.,$$

тобто за рахунок структурних зрушень середня заробітна плата зросла на 12,5 %.

Усі розраховані індекси взаємозв'язані:

$$I_x^- = I_x \cdot I_d \quad (11.33)$$

$$I_x^- = 1,227 \times 1,125 = 1,38.$$

У рамках індексної системи можна визначити абсолютні прирости середньої $\Delta_{\bar{x}} = x_1 - x_0$ за рахунок кожного фактора:

$$\Delta_{\bar{x}} = 11040 - 8000 = 3040 \text{ грн};$$

у тому числі:

- за рахунок зростання заробітної плати в окремих групах:

$$\Delta_x = 11040 - 9000 = 2040 \text{ грн};$$

- за рахунок структурних зрушень:

$$\Delta_d = 9000 - 8000 = 1000 \text{ грн}.$$

Зауваження. Методологічною особливістю побудови системи індексів середніх величин є порівнянність складових

сукупності в часі. Проте більшість реальних сукупностей за своїм складом динамічні: одні частини сукупності зникають, інші (нові) – з’являються. Так, оновлюється асортимент продукції, на ринку цінних паперів з’являються нові емітенти, у видобувній промисловості вводяться в експлуатацію нові родовища і т. ін.

Щоб оцінити вплив на динаміку середньої такого роду змін, в індексну систему вводять три індекси структурних зрушень: I_d^0 – для оцінювання впливу змін у структурі порівнянного кола складових сукупності; I_d^n – для оцінювання впливу новоутворених складових; I_d^b – для оцінювання впливу вибулих складових. Індексна система має вигляд

$$I_{\bar{x}} = I_x^0 \cdot I_d^0 \cdot I_d^n \cdot I_d^b \quad (11.34)$$

Індекс фіксованого складу I_x^0 обчислюється для порівнянного кола складових. Вагами усіх індексів є відносні величини структури – частки d .

Отже,

$$I_{\bar{x}} = \frac{\sum x_1 d_1}{\sum x_0 d_0} \quad (11.35) \quad I_x^0 = \frac{\sum x_1 d_1^0}{\sum x_0 d_1^0} \quad (11.36)$$

$$I_d^0 = \frac{\sum x_0 d_1^0}{\sum x_0 d_0^0} \quad (11.37) \quad I_d^n = \frac{\sum x_1 d_1}{\sum x_1 d_1^0}; \quad (11.38)$$

$$I_d^b = \frac{\sum x_0 d_0^0}{\sum x_0 d_0} \quad (11.39)$$

Приклад. На ринку пального чотири постачальники бензину для автомобілів: А, Б, В, Г. У січні постачальниками А, Б, В поставлено 205 тис. т бензину, у лютому постачальниками А, Б, Г – 255 тис. т. Ціни на бензин для автомобілів у постачальників – різні (табл. 11.7).

Таблиця 11.7

Розрахунок системи індексів структурних зрушень

Постачальник	Ціна 1 т, тис. грн		Обсяг поставок, тис. т		Частка поставки			
	січень	лютий	січень	лютий	у загальному обсязі		по порівнянному колу	
	x_0	x_1	f_0	f_1	d_0	d_1	d_0^0	d_1^0
А	30	25	105	125	0,51	0,49	0,62	0,74
Б	34	28	65	45	0,32	0,18	0,38	0,26
В	28	–	35	–	0,17	–	–	–
Г	–	22	–	85	–	0,33	–	–
Разом	×	×	205	255	1,0	1,0	1,0	1,0

Розв’язання. Знайдемо частки поставок у загальному обсязі та у порівнянному колі.

У порівнянному колі постійними на ринку були постачальники А та Б:

$$d_0^0 = 105 / (105 + 65) = 0,62;$$

$$d_0^0 = 65 / (105 + 65) = 0,38;$$

$$d_1^0 = 125 / (125 + 45) = 0,74;$$

$$d_1^0 = 45 / (125 + 45) = 0,26.$$

У лютому вони знизили ціну на бензин у середньому на 16,9%, індекс фіксованого складу:

$$I_x^0 = \frac{25 \cdot 0,74 + 28 \cdot 0,26}{30 \cdot 0,74 + 34 \cdot 0,26} = \frac{25,78}{31,04} = 0,831.$$

За рахунок структурних зрушень в обсягах поставки постійних постачальників середня ціна бензину зменшилася на 1,5%:

$$I_d^0 = \frac{30 \cdot 0,74 + 34 \cdot 0,26}{30 \cdot 0,62 + 34 \cdot 0,38} = \frac{31,04}{31,52} = 0,985.$$

Вихід з ринку бензину постачальника В з відносно низькою ціною призвів до збільшення середньої ціни на 1,9%:

$$I_d^b = \frac{30 \cdot 0,62 + 34 \cdot 0,38}{30 \cdot 0,51 + 34 \cdot 0,32 + 28 \cdot 0,17} = \frac{31,52}{30,94} = 1,019.$$

Поява на ринку нового постачальника Г із найнижчою ціною сприяла зниженню середньої ціни на 4,8 %:

$$I_d^n = \frac{25 \cdot 0,49 + 28 \cdot 0,18 + 22 \cdot 0,33}{25 \cdot 0,74 + 28 \cdot 0,26} = \frac{24,55}{25,78} = 0,952.$$

Очевидний взаємозв'язок індексів:

$$I_{\bar{x}} = 0,831 \cdot 0,985 \cdot 1,019 \cdot 0,952 = 0,794.$$

Отже, динаміка середньої ціни на бензин формувалася за рахунок як динаміки цін в окремих постачальників, так і різноспрямованої дії структурних факторів.

11.7. Індексні ряди

Послідовність індексів, які характеризують зміну явища протягом певного інтервалу часу, утворюють *індексний ряд*. Кількість індексів у ряду на одиницю менше кількості включених в аналіз періодів часу.

База порівняння може бути *постійною* чи *змінною*, відповідно, індексні ряди – *базисними* або *ланцюговими*. Вибір бази порівняння залежить від мети дослідження. Ряди

базисних індексів дають уявлення про загальну тенденцію явища, ряди ланцюгових індексів – докладнішу картину послідовних змін явища в часі.

У табл. 11.8 наведено індекси тарифів на послуги зв'язку з різною базою порівняння.

Приклад. Маємо чотири індекси. Усі вони відображують об'єктивно існуючу динаміку ціни за 2014–2018 рр. Це чотири версії однієї динаміки з різним масштабом співвідношення.

Таблиця 11.8

Індекси ряди з різним масштабом співвідношення

Рік	Індекси (база порівняння становить 1)			
2014	1,00	.	.	.
2015	1,06	1,00	.	.
2016	1,10	1,04	1,00	.
2017	1,20	1,09	1,05	1,00
2018	1,18	1,11	1,07	1,02

Перехід від одного масштабу співвідношення до іншого здійснюють перебазуванням індексів:

$$I_{18\setminus 17} = 1,05 \cdot 1,02 = 1,07;$$

$$I_{18\setminus 16} = 1,04 \cdot 1,05 \cdot 1,02 = 1,11;$$

$$I_{18\setminus 15} = 1,06 \cdot 1,04 \cdot 1,05 \cdot 1,02 = 1,18.$$

При побудові індексних рядів зведених індексів постійною або змінною може бути не лише база порівняння, а й ваги (сумірники) індексів. Як приклад наведемо *ряди індексів цін*:

- базисних із постійними вагами:

$$\frac{\sum p_1 q_1}{\sum p_0 q_1}; \frac{\sum p_2 q_1}{\sum p_1 q_1}; \dots; \frac{\sum p_n q_1}{\sum p_{n-1} q_1} \quad (11.40)$$

- базисних зі змінними вагами:

$$\frac{\sum p_1 q_1}{\sum p_0 q_1}; \frac{\sum p_2 q_2}{\sum p_0 q_2}; \dots; \frac{\sum p_n q_n}{\sum p_0 q_n} \quad (11.41)$$

- ланцюгових зі змінними вагами:

$$\frac{\sum p_1 q_1}{\sum p_0 q_1}; \frac{\sum p_2 q_2}{\sum p_1 q_2}; \dots; \frac{\sum p_n q_n}{\sum p_{n-1} q_n} \quad (11.42)$$

- ланцюгових із постійними вагами:

$$\frac{\sum p_1 q_1}{\sum p_0 q_1}; \frac{\sum p_2 q_1}{\sum p_1 q_1}; \dots; \frac{\sum p_n q_1}{\sum p_{n-1} q_1} \quad (11.43)$$

Теоретично будь-який ряд можна використовувати для аналізу динаміки. На практиці поширеніші індекси з постійними вагами, які елімінують (усувають) вплив структурних зрушень на величину індексу. Окрім того, в індексних рядах із постійними вагами можна здійснити перебазування – перейти від ланцюгових індексів до базисних, і навпаки. *В індексних рядах зі змінними вагами перебазування індексів неможливе.*

На основі індексних рядів провадиться моніторинг динаміки соціально-економічних явищ, наприклад, кон'юнктури ринку цінних паперів. Спеціальні фірми-спостерігачі вивчають тенденції біржового ринку, оцінюють активність біржової діяльності та ступінь ризику інвесторів.

У моніторингу кон'юнктури біржового ринку використовують різні індекси: середньоарифметичні (зважені та незважені), середньогармонічні зважені, середньгеометричні. За допомогою індексних рядів здійснюються прогнози. Усі ці дані публікуються для широкого загалу.

11.8. Територіальні індекси

Означення. *Територіальний індекс* – показник, який використовують для порівняння однойменних ознак різних територій або об’єктів. Територіальний індекс використовується як інструмент порівняння соціально-економічних показників у просторі: за окремими країнами, територіями, регіонами, об’єктами. *Особливістю* цих індексів є *рівноправність* порівнюваних об’єктів А і Б. Жоден із них не може претендувати на роль бази порівняння, а отже рівноправними потрібно вважати індекси як з базою порівняння А, так і з базою порівняння Б:

$$I_{A/B} = \frac{\sum x_A f}{\sum x_B f} \quad (11.44)$$

$$I_{B/A} = \frac{\sum x_B f}{\sum x_A f} \quad (11.45)$$

де x – індексована величина;

f – вага (сумірник) індексованої величини.

За фіксованих значень ваг (сумірників) індекси I_A і I_B обернено пропорційні. Так, використовуючи світові ціни, можна звести до порівняльного вигляду обсяги експорту окремих країн, регіонів, спільних підприємств.

Приклад. Експорт об’єкта А перевищує експорт об’єкта Б на 18% ($I_{A/B} = 1,18$), то експорт об’єкта Б менший за експорт об’єкта А на 15 % ($I_{B/A} = 1 : 1,18 = 0,85$). Вибір бази порівняння підпорядковується меті дослідження.

Значно складнішим є вибір варіанта зважування. Якщо товарна структура експорту за об’єктами різниться, то

Розділ 11. ІНДЕКСИ

результати порівняння будуть неоднозначні, про що свідчать індекси, розраховані за умовними даними табл. 11.9.

Таблиця 11.9

Розрахунок територіальних індексів

Товар	Об'єкт А		Об'єкт Б	
	Ціна, дол. США	Кількість, шт.	Ціна, дол. США	Кількість, шт.
Бездротові наушники	20	150	16	380
USB-флеш- накопичувач	12	380	9	200

$$I_{A/B} = \frac{\sum x_A f_A}{\sum x_B f_A} = \frac{20 \cdot 150 + 12 \cdot 380}{16 \cdot 150 + 9 \cdot 380} = \frac{7560}{5820} = 1,299;$$

$$I_{B/A} = \frac{\sum x_B f_B}{\sum x_A f_B} = \frac{16 \cdot 380 + 9 \cdot 200}{20 \cdot 380 + 12 \cdot 200} = \frac{7880}{10000} = 0,788.$$

Індекс цін $I_{A/B}$ з вагами на рівні об'єкта А становить 1,299. Аналогічно розрахований індекс цін $I_{B/A}$ з вагами на рівні об'єкта Б – 0,788. Тобто, якщо ваги зафіксувати на рівні того об'єкта, що порівнюється, то ціни кожного з порівнюваних об'єктів виявляться вищими за ціни бази порівняння.

Щоб забезпечити однозначність висновку, застосовують спільні для обох об'єктів ваги. У нашому прикладі такою спільною вагою може бути сумарний обсяг продажу ($f_A + f_B$). За цим варіантом зважування індекс цін $I_{A/B}$ становить

$$I_{A/B} = \frac{\sum x_A f_{(A+B)}}{\sum x_B f_{(A+B)}} = \frac{20 \cdot 530 + 12 \cdot 580}{16 \cdot 530 + 9 \cdot 580} = \frac{17560}{13700} = 1,282;$$

$$I_{B/A} = \frac{\sum x_B f_{(A+B)}}{\sum x_A f_{(A+B)}} = \frac{16 \cdot 530 + 9 \cdot 580}{20 \cdot 530 + 12 \cdot 580} = \frac{13700}{17560} = 0,780.$$

Тобто, ціни об'єкта А вищі за ціни об'єкта Б у середньому на 28,2%. Відповідно, ціни об'єкта Б нижчі за ціни об'єкта А на 22%.

Спільною для обох об'єктів може бути середня або стандартна структура. Наприклад, при порівнянні смертності населення окремих країн за стандарт береться європейська структура населення за віком.

Зауваження. Розрахунок індексів можливий за допомогою вбудованих функцій MS Excel:

СУММ; Автосума – для підсумовування окремих значень показників;

ОКРУГЛ – для округлення чисел до необхідної кількості знаків;

СРЗНАЧ – середнє (арифметичне) значення показників;

ИНДЕКС – для пошуку необхідних значень;

ПОИСКПОЗ – для пошуку даних на основі введених значень;

СУММПРОИЗВ – для підсумкових розрахунків добутків.

11.9. Задачі для самостійної роботи

Задача 11.1. Визначити, які з наведених індексів індивідуальні, а які – зведені:

а) ціни на послуги зв'язку зросли в 1,12 разів а на кабельне телебачення – в 1,8 рази;

Розділ 11. ІНДЕКСИ

б) поголів'я великої рогатої худоби за рік зменшилося на 1,6%, а поголів'я корів – на 2,8%;

в) вантажообіг наземного транспорту за рік збільшився на 25%, зокрема автомобільного транспорту – на 32%.

Задача 11.2. Оптовий товарооборот на ринку борошна характеризують такі дані:

Таблиця 11.10

Складові оптового товарообороту	Товарооборот у фактичних цінах за період, тис. грн		Товарооборот поточного періоду за цінами базисного, тис. грн
	базисний	поточний	
Продано на внутрішньому ринку	315	381	363
Поставлено на експорт	176	331	172
Разом	491	712	535

Визначте зведені індекси: фізичного обсягу оптового продажу цукру, цін та товарообороту у фактичних цінах.

Показати взаємозв'язок індексів, проаналізувати отримані результати.

Задача 11.3. За наведеними даними про посівні площі та урожайність зернових культур у господарстві визначити:

– індивідуальні індекси посівних площ, урожайності та валового збору;

– загальні індекси посівних площ, урожайності та валового збору;

– абсолютну зміну валового збору, в тому числі за рахунок зміни посівних площ та урожайності.

Показати взаємозв'язок між отриманими індексами, пояснити їх економічний зміст.

СТАТИСТИКА

Таблиця 11.11

Зернові культури	Посівна площа, га		Урожайність, ц/га	
	Базовий період	Звітний період	Базовий період	Звітний період
Пшениця	125	130	30,5	32,8
Ячмінь	44	40	20,0	19,0
Овес	12	13	24,0	26,0

Задача 11.4. Використовуючи взаємозв'язок індексів, за наведеними даними визначити: індекс фізичного обсягу експорту виробів з чорного металу, індекс експортних цін на добрива та індекс вартості експорту полімерних матеріалів і пластмас.

Таблиця 11.12

Групи товарів	Індекс		
	вартості експорту	фізичного обсягу експорту	експортних цін
Вироби з чорних металів	1,26	?	1,08
Добрива	1,40	1,1	?
Полімерні матеріали і пластмаси	?	1,23	1,17

Задача 11.5. За наведеними даними визначити зведений індекс фізичного обсягу виробництва продукції на шинному заводі:

Таблиця 11.13

Шини для автомобілів	Вартість виробленої продукції у базисному періоді, млн грн	Індивідуальний індекс фізичного обсягу продукції
Вантажних	340	1,15
Легкових	290	1,06

Розділ 11. ІНДЕКСИ

Задача 11.6. За наведеними даними по заводу будівельних матеріалів визначити зведений індекс собівартості продукції та суму економії чи перевитрат:

Таблиця 11.14

Вид продукції	Грошові витрати на виробництво у поточному періоді, млн грн	Індивідуальний індекс собівартості продукції
Цегла	788	0,98
Блоки для фундаменту	576	0,94

Задача 11.7. Урожайність кукурудзи на площах з різною агротехнічною обробкою становила:

Таблиця 11.15

Площі посіву	Урожайність, ц/га		Посівна площа, у % до підсумку	
	базисний період	поточний період	базисний період	поточний період
Зрошувані	42	49	60	52
Незрошувані	37	41	40	48
У середньому	40	45	100	100

Визначити індекси середніх рівнів врожайності кукурудзи, пояснити їх економічний зміст.

Задача 11.8. Обсяг випуску продукції залежить від наявних виробничих потужностей і ступеня їх використання. Завод будівельних пластмас не виконав річного плану по випуску продукції. Через простої виробничого обладнання фактична потужність виявилася менше планової. Сформувати індексну модель, у рамках якої можна визначити вплив на невиконання плану з випуску продукції факторів, пов'язаних з використанням виробничих потужностей.

СТАТИСТИКА

Таблиця 11.16

№ п/п	Показник	Значення показника	
		за бізнес-планом	фактично
1	Обсяг продукції	Q_0	Q_1
2	Виробничі потужності	M_0	M_1
3	Випуск продукції на одиницю потужності (гр. 1 : гр. 2)	Q_0 / M_0	Q_1 / M_1
4	Коефіцієнт забезпечення річного випуску продукції виробничими потужностями (гр. 2: гр. 1)	M_0 / Q_0	M_1 / Q_1

Задача 11.9. Оптовий продаж цитрусових на внутрішньому ринку зазнав змін:

Таблиця 11.17

Товар	Продано, т		Ціна за 1 т, грн	
	базисний період	поточний період	базисний період	поточний період
Апельсини	9	12	20000	22000
Мандарини	3	7	36500	38000
Грейпфрути	2	3	30000	32250

Визначити:

- індивідуальні індекси цін та фізичного обсягу продажу;
- зведений індекс цін та фізичного обсягу продажу за Ласпейресом і Пааше.

Показати взаємозв'язок індексів, пояснити їх економічний зміст.

Розділ 11. ІНДЕКСИ

Задача 11.10. Виробництво та собівартість продукції харчової промисловості характеризуються такими даними:

Таблиця 11.18

Продукція	Базисний період		Поточний період	
	вироблено, кг	собівартість 1 кг, грн	вироблено, кг	собівартість 1 кг, грн
А	500	52	600	56
Б	420	100	480	97
В	510	46	570	45
Г	360	74	410	73

Визначити:

- зведені індекси фізичного обсягу виробництва та собівартості продукції;
- зведений індекс грошових витрат;
- економію (перевитрати) коштів за рахунок зміни собівартості продукції.

Показати взаємозв'язок індексів, пояснити їх економічний зміст.

Задача 11.11. За наведеними даними про витрати на телерекламу визначити:

- індивідуальні індекси собівартості, кількості та загальних витрат на виробництво рекламних роликів;
- загальні індекси собівартості, кількості та загальних витрат на виробництво рекламних роликів;
- абсолютну зміну загальних витрат на виробництво загалом і, у тому числі за рахунок зміни собівартості та кількості рекламних роликів.

Показати взаємозв'язок індексів, пояснити їх економічний зміст.

СТАТИСТИКА

Таблиця 11.19

Категорії товару	Собівартість одного рекламного ролика, тис. у.о.		Кількість рекламних роликів, одиниць.	
	базисний період	звітний період	базисний період	звітний період
Ліки	1,3	1.5	150	180
Засоби для прибирання	1,0	1,2	140	135

Задача 11.12. За наведеними даними визначити зведені індекси товарообороту та фізичного обсягу реалізації кави, а також абсолютний приріст (зменшення) товарообороту за рахунок динаміки фізичного обсягу реалізації і за рахунок динаміки цін:

Таблиця 11.20

Кава	Товарооборот у фактичних цінах за період, млн грн		Зміна фізичного обсягу реалізації, %
	базисний період	поточний період	
Молота	11	12	+15,2
Розчинна	22	23	+23,8
Кавові мікси	9	10	+12,3

Задача 11.13. За наведеними даними визначити зведений індекс біржових цін акцій та абсолютний приріст (зменшення) обсягу біржового обороту за рахунок динаміки цін:

Таблиця 11.21

Емітент	Обсяги продажу, млн грн		Індекс цін
	лютий	березень	
А	780	810	1,2
В	580	510	0,8
С	630	750	1,6

Розділ 11. ІНДЕКСИ

Задача 11.14. У таблиці наведено дані, що характеризують динаміку експорту товарів і послуг країни за останні два роки:

Таблиця 11.22

Складові експорту	Структура експорту		Індивідуальні індекси	
	базисний рік, d_0	поточний рік, d_1	фізичного обсягу, i_q	цін і тарифів, i_p
Товари	0,58	0,65	0,94	1,040
Послуги	0,42	0,35	0,98	1,054

Визначити зведені індекси:

- фізичного обсягу експорту товарів і послуг;
- цін і тарифів.

Пояснити їх економічний зміст.

Задача 11.15. За наведеними даними про реалізацію молочних продуктів на міському ринку визначити:

- зведені індекси товарообороту, цін та фізичного обсягу реалізації;
- абсолютну зміну товарообороту і в тому числі за рахунок зміни цін та кількості проданих товарів.

Показати взаємозв'язок індексів, проаналізувати отримані результати

Таблиця 11.23

Товар	Товарооборот, тис. грн		Зміна цін у грудні, у % до попереднього місяця
	листопад	грудень	
Молоко	10,8	7,4	+2,9
Масло	12,6	11,4	+3,8
Сметана	8,3	7,0	+4,2

Задача 11.16. За наведеними даними про реалізацію сільськогосподарських продуктів визначити:

СТАТИСТИКА

– загальні індекси товарообороту, фізичного обсягу реалізації та цін;

– абсолютну зміну товарообороту загалом і, в тому числі за рахунок зміни кількості проданих товарів та цін.

Показати взаємозв'язок індексів, проаналізувати отримані результати.

Таблиця 11.24

Товар	Товарооборот, тис. грн		Зміна кількості реалізованої продукції у травні, у % до попереднього місяця
	квітень	травень	
Картопля	30,1	37,4	+9,8
Буряк	15,4	23,6	-12,2
Цибуля	70,2	103,4	+22,1

Задача 11.17. За наведеними даними про витрати на виробництво продукції визначити:

– загальні індекси витрат на виробництво, собівартості одиниці продукції та фізичного обсягу виробництва продукції;

– абсолютну зміну витрат на виробництво всієї продукції загалом і, в тому числі за рахунок зміни собівартості одиниці продукції та фізичного обсягу виробництва.

Показати взаємозв'язок індексів, проаналізувати отримані результати.

Таблиця 11.25

Товар	Витрати на виробництво продукції, тис. грн		Зміна собівартості одиниці продукції, %
	базисний період	звітний період	
Тканини	840	809	+12
Одяг	760	772	+19
Взуття	394	357	+27

Розділ 11. ІНДЕКСИ

Задача 11.18. За наведеними даними визначити індекси середніх закупівельних цін на худобу змінного складу, фіксованого складу та структурних зрушень;

Пояснити економічний зміст обчислених індексів, показати їх взаємозв'язок.

Таблиця 11.26

Категорія вгодованості худоби	Базисний період		Поточний період	
	Обсяги закупівлі, т живої ваги	Середня ціна за 1 т, тис. грн	Обсяги закупівлі, т живої ваги	Середня ціна за 1 т, тис. грн
I	40	26	44	28
II	33	22	31	25
III	21	19	19	21

Задача 11.19. За наведеними даними визначити індекси середньої процентної ставки за користування кредитом: змінного складу, фіксованого складу і структурних зрушень. Пояснити їх зміст.

Таблиця 11.27

Філії банку	Базисний період		Поточний період	
	процентна ставка, %	розмір кредиту, млн грн	процентна ставка, %	розмір кредиту, млн грн
1	17	120	19	75
2	12	230	15	230
3	16	260	17	185

Задача 11.20. За підсумками торгів пшеницею на аграрних біржах середня ціна 1 т пшениці в поточному році порівняно з минулим зросла на 16%. Обсяги продажу пшениці за класами і динаміку цін характеризують такі дані:

СТАТИСТИКА

Таблиця 11.28

Клас пшениці	Обсяги продажу пшениці за рік, млн грн		Індекс цін
	минулий	поточний	
3-й	14,8	17,2	1,31
4-й	13,2	12,2	1,24

Визначити індекси біржових цін на пшеницю фіксованого складу і структурних зрушень. Пояснити їх зміст.

Задача 11.21. За наведеними даними визначити територіальні індекси забезпеченості населення житлом, взявши за базу порівняння регіон А. Використати три варіанти зважування. Проаналізувати отримані результати.

Таблиця 11.29

Тип поселення	Розмір загальної площі житла на одну особу в регіоні, м ²		Структура населення за типом поселення в регіоні, у % до підсумку		
	А	Б	А	Б	по країні загалом
Місто	20,0	21,2	60	70	67
Село	23,0	25,2	40	30	33
По країні загалом	21,2	22,6	100	100	100

Задача 11.21. За наведеними даними про прибуток і вартість активів умовного банку (тис. грн) для кожного періоду визначити показники, що характеризують ефективність банківської діяльності: прибутковість активів ($NP : A$); платоспроможність банку ($K : A$); ефективність використання власних коштів ($VP : K$), та ступінь податкового тиску ($NP : VP$):

Розділ 11. ІНДЕКСИ

Таблиця 11.30

Показники	Валовий прибуток, VP	Чистий прибуток, NP	Вартість активів, A	Власний капітал, K
Базисний період	9512	6900	431250	55200
Поточний період	10478	7440	495938	61824

На основі розрахованих показників сформулюйте 3-факторну індексну модель прибутковості активів і в рамках моделі оцінити абсолютний вплив факторів на динаміку прибутковості активів. Зробити висновки.

Задача 11.22. За даними таблиці про динаміку реалізації товарів здійснити необхідні розрахунки значень показників та заповнити порожні клітинки.

Таблиця 11.31

Товар	Ціна за 1 кг		Кількість реалізованих товарів, т		Індивідуальні індекси		
	Базовий період	Звітний період	Базовий період	Звітний період	Цін	Фізичного обсягу	Вартості реалізованих товарів
А	0,6	0,8	50			1,000	
Б	3,2			8,2	1,052	0,956	
В		2,7	10	11,4	1,087		

11.10. Питання для самоконтролю

1. Що називають індексами і яка їх особливість?
2. Що характеризує динамічний індекс?
3. Що характеризує міжгруповий індекс?
4. Які показники в розрахунку індексів належать до кількісних, якісних, змішаних?

5. Розкрити зміст функцій індексів: синтетичну, аналітичну.
6. Яку величину в розрахунку індексів називають індексованою, а яку сумірником (вагою)?
7. За якими ознаками характеризують індекси?
8. Що характеризують індивідуальні та загальні індекси? Навести приклади.
9. Які індекси називають ланцюговими та базисними?
10. У чому полягає методологічна суть побудови загальних індексів агрегатної форми?
11. Розкрити зміст двох існуючих індексних систем.
12. Навести приклади загальних індексів агрегатної форми: фізичного обсягу продукції, цін, собівартості, продуктивності праці, товарообороту.
13. Як розраховується загальний індекс ціни по Пааше та Ласпейресу?
14. У яких випадках використовуються середньозважені індекси? Які їх види?
15. Який загальний індекс називають середньоарифметичним і як він розраховується?
16. Який загальний індекс називається середньогармонічним і як він розраховується?
17. У яких випадках розраховуються індекси середніх величин?
18. Пояснити суть індексу змінного складу на прикладі індексу цін.
19. Пояснити суть індексу фіксованого складу на прикладі індексу цін.
20. Пояснити суть індексу структурних зрушень на прикладі індексу цін.
21. Якою залежністю пов'язані середні індекси?
22. Для чого будують індексні ряди? Які вони бувають стосовно бази порівняння?
23. Які задачі вирішуються за допомогою територіальних індексів?

ДОДАТКИ

Додаток 1

НАБЛИЖЕНІ ФОРМУЛИ ДЛЯ ОБЧИСЛЕННЯ ЗНАЧЕНЬ ЗАКОНІВ РОЗПОДІЛУ [17]

1. Нормальний розподіл

1.1. Інтеграл імовірностей $\Phi(x)$ визначається як

$$\Phi(x) = \sqrt{\frac{2}{\pi}} \int_0^x e^{-\frac{t^2}{2}} dt;$$

– подання $\tilde{\Phi}(x)$ у вигляді суми степеневого ряду для $x \leq 4$:

$$\tilde{\Phi}(x) = \sqrt{\frac{2}{\pi}} x \sum_{i=0}^N \frac{\left(-\frac{x^2}{2}\right)^i}{i!(2i+1)};$$

– подання $\tilde{\Phi}(x)$ у вигляді суми степеневого ряду для $x \geq 4$:

$$\tilde{\Phi}(x) = 1 - \sqrt{\frac{2}{\pi}} \frac{e^{-\frac{x^2}{2}}}{x} \sum_{i=0}^N \frac{(2i-1)!!}{(-x^2)^i};$$

– обчислення $\tilde{\Phi}(x)$ за формулою:

$$\tilde{\Phi}(x) = \sqrt{1 - e^{-\frac{2}{\pi}x^2}};$$

Продовження дод. 1

– обчислення $\tilde{\Phi}(x)$ за формулою:

$$\tilde{\Phi}(x) = \sqrt{1 - \frac{e^{-x^2\left(\frac{2-k}{\pi}\right)}}{1+kx^2}}, \quad k = 0,1523;$$

1.2. Функція, обернена інтегралу імовірностей.

Позначимо $x(z) = z^{-1}$, де $z = \Phi(x) = \sqrt{\frac{2}{\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$;

– обчислення $\tilde{x}(z)$ за формулою

$$\tilde{x}(z) = \sqrt{-\pi \ln \sqrt{1-z^2}};$$

– обчислення $\tilde{x}(z)$ за формулою

$$\tilde{x}(z) = \sqrt{-\frac{\pi}{2} \left[-\ln(1-z^k) \right]^{\frac{1}{k}}} \quad k = 1,898;$$

– обчислення $\tilde{x}(z)$ за формулою

$$\tilde{x}(z) = \sqrt{\pi \left\{ \frac{1}{41} \left[-\ln(1-z^4) \right]^{\frac{3}{2}} - \ln \sqrt{1-z^2} \right\}};$$

– обчислення $\tilde{x}(z)$ для $z \cong 1$:

$$\tilde{x}(1-q) = \sqrt{\pi \left\{ \frac{1}{23} \left[-\ln(5q) \right]^{\frac{5}{4}} - \ln \sqrt{2q} \right\}},$$

де $q = 1-z$ – доповнення z до 1.

2. χ^2 -розподіл

2.1. Щільність імовірності χ^2 -розподілу визначається із співвідношення:

$$\varphi_b(x) = \frac{x^{\frac{m}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} \quad (x > 0).$$

2.2. Інтегральний закон χ^2 -розподілу визначається співвідношенням:

$$F_m(x) = \int_0^x \varphi_m(t) dt = \frac{1}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} \int_0^x t^{\frac{m}{2}-1} e^{-\frac{t}{2}} dt;$$

– обчислення $F_m(x)$ для парних m ($m \geq 2$):

$$F_m(x) = 1 - e^{-\frac{x}{2}} \sum_{i=0}^{\frac{m}{2}-1} \frac{\left(\frac{x}{2}\right)^i}{i!};$$

– обчислення $F_m(x)$ для непарних m ($m \geq 3$):

$$F_m(x) = \Phi(\sqrt{x}) - \sqrt{\frac{2x}{\pi}} e^{-\frac{x}{2}} \sum_{i=0}^{\frac{m-3}{2}} \frac{x^i}{(2i+1)!!},$$

де $\Phi(\sqrt{x})$ – інтеграл імовірності, який наближено може бути поданий у вигляді:

$$\Phi(\sqrt{x}) \cong \sqrt{\frac{2}{\pi}} \int_0^{\sqrt{x}} e^{-\frac{t^2}{2}} dt \cong \sqrt{1 - \frac{e^{-x(2/\pi-k)}}{1+kx}}, \quad k=0,1,45.$$

3. Розподіл Стьюдента

Розподіл Стьюдента з n степенями вільності визначається як розподіл випадкової величини вигляду

$$t_n = \frac{z_0}{\sqrt{\frac{1}{n} \sum_{i=1}^n z_i^2}}, \text{ де } z_i \ (i=0,1,\dots,n) \text{ розподілені нормально з}$$

$M(z_i) = 0$ і однаковими дисперсіями σ_z^2 . При цьому t_n не залежить від величини дисперсії σ_z^2 .

3.1. Щільність імовірності розподілу Стьюдента з n степенями вільності для $n \geq 1$ визначається як

$$p_n(t) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

3.2. Інтегральний закон розподілу Стьюдента визначається співвідношенням:

$$S_n(t) = \int_{-t}^t p_n(x) dx = \frac{2}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \int_0^t \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx.$$

Позначивши $\varphi = \operatorname{arctg} \frac{t}{\sqrt{n}}$, запишемо $S_n(t)$ у вигляді:

$$S_n(t) = \sin \varphi \sum_{i=0}^{\frac{n}{2}-1} \cos^{2i} \varphi \prod_{j=0}^{i-1} \frac{j + \frac{1}{2}}{j + 1},$$

якщо n – парне,

$$S_n(t) = \frac{2}{\pi} \left\{ \phi + \operatorname{tg} \phi \sum_{i=1}^{\frac{n-1}{2}} \cos^{2i} \phi \prod_{j=0}^{i-1} \frac{j}{j + \frac{1}{2}} \right\},$$

якщо n – непарне.

4. F -розподіл

F -розподіл визначається як розподіл відношення двох випадкових величин, що мають χ^2 -розподіл відповідно з m і n степенями вільності.

Оскільки сума квадратів нормально розподілених випадкових величин має χ^2 -розподіл, то випадкова величина z , що має F -розподіл, визначається через нормально розподілені випадкові величини x та y у вигляді:

$$z = \frac{\frac{1}{m} \sum_{i=1}^m x_i^2}{\frac{1}{n} \sum_{j=1}^n y_j^2}.$$

4.1. Щільність імовірності F -розподілу запишемо із застосуванням бета-функції:

$$p_F(m, n, z) = \left(\frac{m}{2}\right)^{\frac{m}{2}} B^{-1}\left(\frac{m}{2}, \frac{n}{2}\right) z^{\frac{m}{2}-1} \left(1 + \frac{m}{n} z\right)^{-\frac{m+n}{2}}.$$

4.2. Інтегральний закон F -розподілу подається у вигляді:

$$F(m, n, z) = \int_0^z p_F(m, n, x) dx = \left(\frac{m}{n}\right)^{\frac{m}{2}} B^{-1}\left(\frac{m}{2}, \frac{n}{2}\right) \int_0^z x^{\frac{m}{2}-1} \left(1 + \frac{m}{n} x\right)^{-\frac{m+n}{2}} dx.$$

Продовження дод. 1

Наведений інтеграл обчислюється в елементарних функціях, але одержаний у загальному випадку вираз складний.

Нижче наведемо наближені формули обчислення $F(m, n, z)$ – для парного n , для парного m та для випадку, коли m і n одночасно непарні.

n – парне:

$$F(m, n, z) = \frac{n}{\left(1 + \frac{n}{mz}\right)^{n/2}} \left(\prod_{i=1}^{n/2} \frac{m + 2i - 2}{2i} \right) X_{n/2},$$

де $X_{n/2}$ визначається за допомогою рекурентного співвідношення:

$$X_{n/2} = \frac{2(i-1)X_{i-1} + \left(1 + \frac{mz}{n}\right)^{1-i}}{2i + m - 2}, \quad i = 1, 2, \dots, \frac{n}{2};$$

m – парне:

$$F(m, n, z) = m \left\{ \prod_{i=1}^{m/2} \frac{n + 2i - 2}{2i} \right\} Y_{m/2},$$

де $Y_{m/2}$ визначається за допомогою рекурентного співвідношення:

$$Y_j = \frac{2(j-1)Y_{j-1}}{2j-m-2} + \frac{1-\cos^{m+n-2}\phi}{m+n-2j}, \quad j = 2, 3, \dots, \frac{m}{2};$$

$$Y_1 = \frac{1-\cos^{m+n-1}\phi}{m+n-2}, \quad \phi = \arccos \left(\frac{1}{\sqrt{1+\frac{mz}{n}}} \right);$$

m і n – непарні:

$$F(m, n, z) = A_{m-1, n-1} \frac{2}{\pi} \left[\prod_{i=1}^{\frac{m-1}{2}} \frac{m+n-2i}{m-2i} \right] \left[\prod_{j=1}^{\frac{n-1}{2}} \frac{n-2j+1}{n-2j} \right],$$

де $A_{m-1, n-1}$ визначається за допомогою рекурентного співвідношення:

$$A_{i, n-1} = -\frac{\sin^{i-1}\phi \cos^n\phi}{n+i-1} + \frac{i-1}{n+i-1} A_{i-2, n-1}, \quad i = 2, 4, \dots, m-1,$$

а $A_{0, n-1}$ – за допомогою рекурентного співвідношення:

$$A_{0, j} = \frac{\sin\phi \cos^{j-1}\phi}{j} + \frac{j-1}{j} A_{0, j-2}, \quad j = 2, 4, \dots, n-1,$$

для $A_{0, 0} = \phi = \arctg \sqrt{\frac{mz}{n}}$.

**ТАБЛИЦЯ стандартизованого НОРМАЛЬНОГО
РОЗПОДІЛУ
 $A(z)$**

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5703	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9533
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9700	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987									

Джерело: Доугерти К. Введение в эконометрику: пер. с англ. – М.: ИНФРА-М, 1997. – XIV. – С. 367.

$A(z)$ – це інтеграл щільності імовірності стандартизованого нормального розподілу від $-\infty$ до z (площа під кривою зліва від z) і є імовірність того, що величина нормально розподіленої випадкової змінної не перевищує середнє значення більше, ніж на z стандартних відхилень.

Приклад: $p(z \leq 2,15) = 0,9842$; $p(z > 2,15) = 1 - (z \leq 2,15) = 1 - 0,9842 = 0,0158$.

ТАБЛИЦЯ χ^2 -РОЗПОДІЛУ
[критичні значення χ^2 для рівня значущості
 α та k степенів вільності]

k	Рівень значущості α , %)						
	0,1%	1%	2,5%	5%	9,5%	97,5%	99%
1	10,8	6.6	5.0	3.8	0.0039	0.001	0.0002
2	13,86	9.2	7.4	6.0	0.103	0.051	0.02
3	16,2	11.3	9.4	7.8	0.352	0.216	0.115
4	18,5	13.3	11.1	9.5	0.711	0.484	0.297
5	20,5	15.1	12.8	11.1	1.15	0.831	0.554
6	22,5	16.8	14.4	12.6	1.64	1.24	0.872
7	24,3	18.5	16.0	14.1	2.17	1.69	1.2
8	26,1	20.1	17.5	15.5	2.73	2.18	1.7
9	27,9	21.7	19.0	16.9	3.33	2.70	2.1
10	29,6	23.2	20.5	18.3	3.94	3.25	2.6
11	31,3	24.7	21.9	19.7	4.57	3.82	3.1
12	32,9	26.2	23.3	21.0	5.23	4.40	3.6
13	34,5	27.7	24.7	22.4	5.89	5.01	4.1
14	36,1	29.1	26.1	23.7	6.57	5.63	4.7
15	37,7	30.6	27.5	25.0	7.26	6.26	5.2
16	39,3	32.0	28.8	26.3	7.96	6.91	5.8
17	40,8	33.4	30.2	27.6	8.67	7.56	6.4
18	42,3	34.8	31.5	28.9	9.39	8.23	7.0
19	43,8	36.2	32.9	30.1	10.1	8.91	7.6
20	45,3	37.6	34.2	31.4	10.9	9.59	8.3
21	46,8	38.9	35.5	32.4	11.6	10.3	8.9
22	48,3	40.3	36.8	33.9	12.3	11.0	9.5
23	49,7	41.6	38.1	35.2	13.1	11.7	10.2
24	51,2	43.0	39.4	36.4	13.8	12.4	10.9
25	52,6	44.3	40.6	37.7	14.6	13.1	11.5
26	54,1	45.6	41.9	38.9	15.4	13.8	12.2
27	55,5	47.0	43.2	40.1	16.2	14.6	12.9
28	56,9	48.3	44.5	41.3	16.9	15.3	13.6
29	58,3	49.6	45.7	42.6	17.7	16.0	14.3
30	59,7	50.9	47.0	43.8	18.5	16.8	15.0

Джерело: Доугерти К. Введение в эконометрику: пер. с англ. – М.: ИНФРА-М, 1997. – XIV. – С. 371.

Толбатов Ю.А. Эконометрика: підруч. для студ. екон. спец. вищ. навч. закл. / Ю.А. Толбатов. – Київ: Четверта хвиля, 1997. – С.3 01.

ТАБЛИЦЯ t -РОЗПОДІЛУ СТЬЮДЕНТА
[критичні значення $t(\alpha, k)$]

Тести	Рівень значущості α , %							
	50%	20%	10%	5%	2%	1%	0,2%	0,1%
Двосторонній								
Односторонній	25%	10%	5%	2,5%	1%	0,5%	0,1%	0,05%
k								
1	1,000	3,078	6,314	12,706	31,821	63,657	318,31	636,62
2	0,861	1,886	2,920	4,303	6,965	9,925	22,327	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	10,214	12,924
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307	3,551
60	0,679	1,296	1,671	2,000	2,390	2,660	3,232	3,460
120	0,677	1,289	1,658	1,980	2,358	2,617	3,160	3,373
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Джерела: Доугерти К. Введение в эконометрику: пер. с англ. – М.: ИНФРА-М, 1997. – XIV. – С. 368.

Лук'яненко І.Г. Економетрика : підручник /І.Г. Лук'яненко, Л.І. Краснікова. – Київ: Знання, 1998. – С. 484.

ДОДАТКИ

Додаток 5

ТАБЛИЦЯ $F(k_1, k_2, \alpha)$ розподілу Фішера для рівня значущості $\alpha=0,05$ (5%)

k_2	k_1												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	161	200	216	225	230	234	237	239	241	242	243	244	245
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,20	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26
9	5,12	4,26	3,68	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20
23	4,48	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20	2,18
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18	2,15
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13	2,10
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	1,09
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,08
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,90	1,86	1,83
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,87	1,83	1,80
140	3,91	3,06	2,67	2,44	2,28	2,16	2,08	2,01	1,95	1,90	1,86	1,82	1,79

СТАТИСТИКА

Продовження дод. 5

k_2	k_1												
	1	2	3	4	5	6	7	8	9	10	11	12	13
160	3,90	3,05	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,81	1,78
180	3,89	3,05	2,65	2,42	2,26	2,15	2,06	1,99	1,93	1,88	1,84	1,81	1,77
200	3,88	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80	1,77
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75	1,72

Таблиця $F(k_1, k_2, \alpha)$ розподілу Фішера для рівня значущості $\alpha=0,05$ (5%)

k_2	14	15	16	17	18	19	20	30	40	50	100	200	∞
1	245	246	246	247	247	248	248	250	251	252	253	254	254
2	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
3	8,71	8,70	8,69	8,68	8,67	8,67	8,66	8,62	8,60	8,58	8,55	8,54	8,53
4	5,87	5,86	5,84	5,83	5,82	5,81	5,80	5,75	5,72	5,70	5,66	5,65	5,63
5	4,64	4,62	4,60	4,59	4,58	4,57	4,56	4,50	4,46	4,44	4,41	4,39	4,37
6	3,96	3,94	3,92	3,91	3,90	3,88	3,87	3,81	3,77	3,75	3,71	3,69	3,67
7	3,53	3,51	3,49	3,48	3,47	3,46	3,44	3,38	3,34	3,32	3,27	3,25	3,23
8	3,24	3,22	3,20	3,19	3,17	3,16	3,15	3,08	3,04	3,02	2,97	2,95	2,93
9	3,03	3,01	2,99	2,97	2,96	2,95	2,94	2,86	2,83	2,80	2,76	2,73	2,71
10	2,86	2,85	2,83	2,81	2,80	2,79	2,77	2,70	2,66	2,64	2,59	2,56	2,54
11	2,74	2,72	2,70	2,68	2,67	2,66	2,65	2,57	2,53	2,51	2,46	2,43	2,40
12	2,64	2,62	2,60	2,58	2,57	2,56	2,54	2,47	2,43	2,40	2,36	2,32	2,30
13	2,55	2,53	2,51	2,50	2,48	2,47	2,46	2,38	2,34	2,31	2,26	2,23	2,21
14	2,48	2,46	2,44	2,43	2,41	2,40	2,39	2,31	2,27	2,24	2,19	2,16	2,13
15	2,42	2,40	2,38	2,37	2,35	2,34	2,33	2,25	2,20	2,18	2,12	2,10	2,07
16	2,37	2,35	2,33	2,32	2,30	2,29	2,28	2,19	2,15	2,12	2,07	2,04	2,01
17	2,33	2,31	2,29	2,27	2,26	2,24	2,23	2,15	2,10	2,08	2,02	1,99	1,96
18	2,29	2,27	2,25	2,23	2,22	2,20	2,19	2,11	2,10	2,08	1,98	1,95	1,92
19	2,26	2,23	2,21	2,00	2,18	2,17	2,16	2,07	2,03	2,00	1,94	1,91	1,88
20	2,22	2,20	2,18	2,17	2,15	2,14	2,12	2,04	1,99	1,97	1,91	1,88	1,84
21	2,20	2,18	2,16	2,14	2,12	2,11	2,10	2,01	1,96	1,94	1,88	1,84	1,81
22	2,17	2,15	2,13	2,11	2,10	2,08	2,07	1,98	1,94	1,91	1,85	1,82	1,78
23	2,15	2,13	2,11	2,09	1,08	2,06	2,05	1,96	1,91	1,88	1,82	1,79	1,76
24	2,13	2,11	1,09	2,07	2,05	2,04	2,03	1,94	1,89	1,86	1,80	1,77	1,73
25	2,11	2,09	2,07	2,05	2,04	2,02	2,01	1,92	1,87	1,84	1,78	1,50	1,71
26	2,09	2,07	2,05	2,03	2,02	2,00	1,99	1,90	1,85	1,82	1,76	1,73	1,69
27	2,08	2,06	2,04	2,02	2,00	1,99	1,97	1,88	1,84	1,81	1,74	1,71	1,67
28	2,06	2,04	2,02	2,00	1,99	1,97	1,96	1,87	1,82	1,79	1,73	1,69	1,65
29	2,05	2,03	2,01	1,99	1,97	1,96	1,94	1,85	1,81	1,77	1,71	1,57	1,64
30	2,04	2,01	1,99	1,98	1,96	1,95	1,93	1,84	1,79	1,76	1,70	1,66	1,62
40	1,95	1,92	1,90	1,89	1,87	1,85	1,84	1,74	1,69	1,66	1,63	1,60	1,51
50	1,89	1,87	1,85	1,83	1,81	1,80	1,78	1,69	1,63	1,60	1,59	1,55	1,44
60	1,86	1,84	1,82	1,80	1,78	1,76	1,75	1,65	1,59	1,56	1,48	1,44	1,39

ДОДАТКИ

Продовження дод. 5

k_2	14	15	16	17	18	19	20	30	40	50	100	200	∞
70	1,84	1,81	1,79	1,77	1,75	1,74	1,72	1,62	1,57	1,53	1,45	1,40	1,35
80	1,82	1,79	1,77	1,75	1,73	1,72	1,70	1,60	1,54	1,51	1,43	1,38	1,32
90	1,80	1,78	1,76	1,74	1,72	1,70	1,69	1,59	1,53	1,49	1,41	1,36	1,30
100	1,79	1,77	1,75	1,73	1,71	1,69	1,68	1,57	1,52	1,48	1,39	1,34	1,28
120	1,78	1,75	1,73	1,71	1,69	1,67	1,66	1,55	1,50	1,46	1,37	1,32	1,25
140	1,76	1,74	1,72	1,70	1,68	1,66	1,65	1,54	1,48	1,44	1,35	1,30	1,23
160	1,75	1,73	1,71	1,69	1,67	1,65	1,64	1,53	1,47	1,43	1,34	1,28	1,21
180	1,75	1,72	1,70	1,68	1,66	1,64	1,63	1,52	1,46	1,42	1,33	1,27	1,20
200	1,74	1,72	1,69	1,67	1,66	1,64	1,62	1,52	1,46	1,41	1,32	1,26	1,19
∞	1,69	1,67	1,64	1,62	1,60	1,59	1,57	1,46	1,39	1,35	1,24	1,17	1,00

ТАБЛИЦЯ $F(k_1, k_2, \alpha)$ розподілу Фішера для рівня значущості $\alpha=0,01$ (1%)

k_2	k_1												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	4052	4999	5403	5625	5764	5859	5928	5981	6023	6056	6083	6106	6126
2	98,5	99,0	99,2	99,3	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4
3	34,1	30,8	29,4	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	27,1	27,0
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,5	14,4	14,3
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,96	9,89	9,82
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	5,61
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71	4,65
11	9,64	7,20	6,21	5,67	5,31	5,07	4,88	4,74	4,63	4,54	4,46	4,39	4,34
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,90
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37	3,32
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23	3,18
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17	3,12
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12	3,07
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07	3,02
24	7,82	5,64	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03	2,98
25	7,77	5,57	4,68	4,18	3,85	3,83	3,46	3,32	3,22	3,13	3,06	2,99	2,94
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96	2,90

СТАТИСТИКА

Продовження дод. 5

k_2	k_1												
	1	2	3	4	5	6	7	8	9	10	11	12	13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,99	2,93	2,87
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90	2,84
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93	2,87	2,81
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84	2,79
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66	2,61
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,62	2,56	2,51
60	7,07	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	2,44
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45	2,40
80	6,96	4,88	4,04	3,56	3,25	3,04	2,87	2,74	2,64	2,55	2,48	2,42	2,36
90	6,92	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39	2,33
100	6,88	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37	2,31
120	6,84	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,40	2,34	2,28
140	6,81	4,76	3,92	3,46	3,15	2,93	2,77	2,64	2,54	2,45	2,38	2,31	2,26
160	6,79	4,74	3,90	3,44	3,13	2,92	2,75	2,62	2,52	2,43	2,36	2,30	2,24
180	6,77	4,72	3,89	3,43	3,12	2,90	2,74	2,61	2,51	2,42	2,35	2,28	2,23
200	6,75	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27	2,22
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,25	2,18	2,13

ТАБЛИЦЯ $F(k_1, k_2, \alpha)$ розподілу Фішера для рівня значущості $\alpha=0,01$ (1%)

k_2	k_1												
	14	15	16	17	18	19	20	30	40	50	100	200	∞
1	6143	6157	6169	6182	6192	6201	6209	6261	6287	6303	6335	6350	6366
2	99,4	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5	99,5
3	26,9	26,9	26,8	26,8	26,8	26,7	26,7	26,5	26,4	26,4	26,2	26,2	26,1
4	14,2	14,2	14,2	14,1	14,1	14,0	14,0	13,8	13,7	13,7	13,6	13,5	13,5
5	9,77	9,72	9,68	9,64	9,61	9,58	9,55	9,38	9,29	9,24	9,13	9,08	9,02
6	7,60	7,56	7,52	7,48	7,45	7,42	7,40	7,23	7,14	7,09	6,99	6,93	6,88
7	6,36	6,31	6,28	6,24	6,21	6,18	6,16	5,99	5,91	5,86	5,75	5,70	5,65
8	5,56	5,52	5,48	5,44	5,41	5,38	5,36	5,20	5,12	5,07	4,96	4,91	4,86
9	5,01	4,96	4,92	4,89	4,86	4,83	4,81	4,65	4,57	4,52	4,41	4,36	4,31
10	4,60	4,56	4,52	4,49	4,46	4,43	4,41	4,25	4,17	4,12	4,01	3,96	3,91
11	4,29	4,25	4,21	4,18	4,15	4,12	4,10	3,94	3,86	3,81	3,70	3,65	3,60
12	4,05	4,01	3,97	3,94	3,91	3,88	3,86	3,70	3,62	3,57	3,47	3,41	3,36
13	3,86	3,82	3,78	3,74	3,72	3,69	3,66	3,51	3,42	3,37	3,27	3,22	3,17
14	3,70	3,66	3,62	3,59	3,56	3,53	3,51	3,35	3,27	3,22	3,11	3,06	3,00
15	3,56	3,52	3,49	3,45	3,42	3,40	3,37	3,21	3,13	3,08	2,98	2,92	2,87
16	3,45	3,41	3,37	3,34	3,31	3,28	3,26	3,10	3,02	2,97	2,86	2,81	2,75
17	3,35	3,31	3,27	3,24	3,21	3,19	3,16	3,00	2,92	2,87	2,76	2,71	2,65
18	3,27	3,23	3,19	3,16	3,13	3,10	3,08	2,92	2,84	2,78	2,68	2,62	2,57
19	3,19	3,15	3,12	3,08	3,05	3,03	3,00	2,84	2,76	2,71	2,60	2,55	2,49
20	3,13	3,09	3,05	3,02	2,99	2,96	2,94	2,78	2,69	2,64	2,54	2,48	2,42

ДОДАТКИ

Закінчення дод. 5

k_2	k_1												
	14	15	16	17	18	19	20	30	40	50	100	200	∞
21	3,07	3,03	2,99	2,96	2,93	2,90	2,88	2,72	2,64	2,58	2,48	2,42	2,36
22	3,02	2,98	2,94	2,91	2,88	2,85	2,83	2,67	2,58	2,53	2,42	2,36	2,31
23	2,97	2,93	2,89	2,86	2,83	2,80	2,78	2,62	2,54	2,48	2,37	2,32	2,26
24	2,93	2,89	2,85	2,82	2,79	2,76	2,74	2,58	2,49	2,44	2,33	2,27	2,21
25	2,89	2,85	2,81	2,78	2,75	2,72	2,70	2,54	2,45	2,40	2,29	2,23	2,17
26	2,86	2,81	2,73	2,75	2,72	2,69	2,66	2,60	2,42	2,36	2,25	2,19	2,13
27	2,82	2,78	2,75	2,71	2,68	2,66	2,63	2,47	2,38	2,33	2,22	2,16	2,10
28	2,79	2,75	2,72	2,68	2,65	2,63	2,60	2,44	2,35	2,30	2,19	2,13	2,06
29	2,77	2,73	2,69	2,66	2,63	2,60	2,57	2,41	2,33	2,27	2,16	2,10	2,03
30	2,74	2,70	2,66	2,63	2,60	2,57	2,55	2,39	2,30	2,25	2,13	2,07	2,01
40	2,56	2,52	2,48	2,45	2,42	2,39	2,37	2,20	2,11	2,06	2,02	1,96	1,80
50	2,46	2,42	2,38	2,35	2,32	2,29	2,27	2,10	2,01	1,95	1,94	1,87	1,68
60	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,03	1,94	1,88	1,75	1,68	1,60
70	2,35	2,31	2,27	2,23	2,20	2,18	2,15	1,98	1,89	1,83	1,70	1,62	1,54
80	2,31	2,27	2,23	2,20	2,17	2,14	2,12	1,94	1,85	1,79	1,65	1,58	1,49
90	2,29	2,24	2,21	2,17	2,14	2,11	2,09	1,92	1,82	1,76	1,62	1,55	1,46
100	2,27	2,22	2,19	2,15	2,12	2,08	2,07	1,89	1,80	1,74	1,60	1,52	1,43
120	2,23	2,19	2,15	2,12	2,09	2,06	2,03	1,86	1,76	1,70	1,56	1,48	1,38
140	2,21	2,17	2,13	2,10	2,07	2,04	2,01	1,84	1,74	1,67	1,53	1,45	1,35
160	2,20	2,15	2,11	2,08	2,05	2,02	1,99	1,82	1,72	1,66	1,51	1,42	1,32
180	2,18	2,14	2,10	2,07	2,04	2,01	1,98	1,81	1,71	1,64	1,49	1,41	1,30
200	2,17	2,13	2,09	2,06	2,03	2,00	1,97	1,79	1,69	1,63	1,48	1,39	1,28
∞	2,08	2,04	2,00	1,97	1,93	1,90	1,88	1,70	1,59	1,52	1,36	1,25	1,00

Джерело: Бронштейн И.Н. Справочник по математике: для инженеров и учащихся ВТУЗов/И.Н Бронштейн, К.А. Семенов. – М.: Наука, 1981. – С. 82–87.

СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

1. Про державну статистику: Закон України № 2614-ХІІ від 17.09.1992 [Електронний ресурс] // Верховна Рада України. – Режим доступу : <http://www.rada.gov.ua>.
2. Порядок проведення перевірок достовірності первинних і статистичних даних, вивчення стану первинного обліку та статистичної звітності органами статистики, затв. наказом Держкомстату України від 19.06.03 № 186. – Режим доступу : <http://www.ukrstat.gov.ua>.
3. Про заходи щодо розвитку державної статистики: Указ Президента України від 22 листопада 1997 року, № 1299/97 // Статистика України. – 1998. – № 1.
4. Програма реформування державної статистики на період до 2002 року // Постанова Кабінету Міністрів України № 971 від 27.06.1998.
5. Удосконалення законодавства України у галузі статистики. Совершенствование законодательства Украины в области статистики. Improvement of statistical legislation of Ukraine / за ред. Ю.М. Остапчука. – Київ : ІВЦ Держкомстату України, 2002. – 764 с.

Основний

1. Горкавий В.К. Статистика : навч. посіб. – Вид. 2-ге, перероб. і допов. / В.К. Горкавий. – Київ : Алерта, 2012. – 608 с.
2. Єріна А.М. Статистика : підручник / А.М. Єріна, З.О. Пальян. – Київ : КНЕУ, 2010. – 351 с.
3. Лугінін О.Є. Статистика національної та міжнародної економіки : навч. посіб. / О.Є. Лугінін, С.В. Фомішин. – Львів : Новий світ–2000, 2014. – 471 с.

СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

4. Мармоза А.Т. Теорія статистики : підручник / А.Т. Мармоза. – 2-ге вид., перероб. і допов. – Київ : Центр навч. літ., 2013. – 592 с.
5. Матковський С.О. Теорія статистики : навч. посіб. / С.О. Матковський, О.Р. Марець. – 2-ге вид. – Київ : Знання, 2010. – 534 с.
6. Стегней, М.І. Статистика: кредитно-модульний курс : навч. посіб. / М. І. Стегней, І. О. Іртищева. – Київ : Кондор, 2013. – 305 с.

Додатковий

7. Годин А. М. Статистика : учебник / А. М. Годин. – М. : Дашков и К, 2012. – 451 с.
8. Ковалевський Г.В. Статистика : підручник / Г.В. Ковалевський. – Харків : ХНАМГ, 2012. – 445 с.
9. Кулинич О.І Теорія статистики : підручник / О. І. Кулинич, Р.О. Кулинич. – 5-те вид., перероб. і допов. – Київ : Знання, 2010. – 239 с.
10. Ковтун Н. В. Теорія статистики : підручник / Н.В. Ковтун. – Київ : Знання, 2012. – 400 с.
11. Логунова Н.А. Статистика II : підручник / Н. А. Логунова. – Київ : Кондор, 2014. – 340 с.
12. Лугінін О.Є. Статистика : підручник / О.Є. Лугінін. – 2-ге вид., перероб. та допов. – Київ : Центр навч. літ., 2007. – 608 с.
13. Опря А.Т. Статистика (модульний варіант з програмованою формою контролю знань) : навч. посіб. / А.Т. Опря. – Київ : Цент навч. літ., 2012. – 448 с.
14. Горошанська О.О. Статистика в схемах, формулах і таблицях / О.О. Горошанська, О.В. Прокопова. – Харків : ХДУХТ, 2014. – 52 с.

15. Статистика для економістів : навч. посіб. / Р. М. Моторин, Е. В. Чекотовський. – 3-є вид., виправ. і допов. – Київ : Знання, 2013. – 381 с.
16. Статистика : підручник / С.С. Герасименко, А.В. Головач та ін. – Київ : КНЕУ, 2000.
17. Статистика: Структурно-логічні схеми та задачі: [навч. посіб.] / А. М. Єріна, В. Б. Захожай, І. Г. Манцуров та ін.; за наук. ред. А. М. Єріної. – Київ : КНЕУ, 2007. – С.304
18. Статистика : навч. посіб. / О.О. Горошанська, О.В. Прокопова. – Харків : ХДУХТ, 2014. – 320 с.
19. Статистичний словник / за ред. О.Г. Осауленка; Держ. служба статистики України, Наук.-техн. комплекс стат. досліджень. – Київ : Інформ.-аналіт. агентство, 2012. – 499 с.
20. Ткач Є.І. Загальна теорія статистики : підручник / Є.І. Ткач, В.П. Сторожук.– 3-тє вид., перероб. і допов. – Київ : Центр навч. літ., 2017. – 442 с.
21. Тринько Р.І. Основи теоретичної і практичної статистики : навч. посіб. / Р.І. Тринько, М.Є. Стадник. – Київ : Знання, 2011. – 397 с.
22. Чернелевський Л.М. Статистика : підручник / Л.М. Чернелевський, Л.М. Соломчук, М.В. Перетятко. – Київ : НУХТ, 2012. – 207 с.
23. Попов І.І. Статистика : практикум / І.І. Попов – Київ : Київ. нац. торг.-екон. ун-т, 2016. – 160 с.
24. Практикум по теории статистики : учеб. пособ. / под ред. Р.А. Шмойловой .– М. : Финансы и статистика, 2002. – 416 с.: ил.
25. Уманець Т.В. Статистика : навч. посіб. /Т.В. Уманець, Ю.Б. Пігарев.– Київ : Вікар, 2003.
26. Щурик М.В. Статистика : навч. посіб. / М.В. Щурик. – Львів : Магнолія, 2009. – 545 с.

СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

27. Бек В.Л. Теорія статистики : навч. посіб. /В.Л. Бек. – Київ : ЦУЛ, 2003.
28. Теорія статистики : навч. посіб. / П.Г. Вашків, П.І. Пастер, В.П. Сторожук, Є.І Ткач. – Київ : Либідь, 2001.

Інтернет-ресурси

1. Сайт Державної служби статистики України. – Режим доступу : www.ukrstat.gov.ua
2. Сайт Головного управління статистики у м. Києві. – Режим доступу : www.gorstat.kiev.ua
3. Сайт Кабінету Міністрів України. – Режим доступу : www.kmu.gov.ua
4. Сайт WorldometersReal time world statistics. – Режим доступу : www.worldometers.info
5. Сайт World Bank. – Режим доступу : www.worldbank.org
6. Сайт Організації об'єднаних націй. – Режим доступу : www.un.org
7. Сайт Національної бібліотеки України ім. В.І. Вернадського. – Режим доступу : archive.nbuv.gov.ua

Навчальне видання

ПИРОЖКОВ Сергій Іванович,
РЯЗАНЦЕВА Валентина Василівна,
МОТОРИН Руслан Миколайович,
ГОЛОВАЧ Наталія Анатоліївна,
ІВАНЧЕНКО Ніна Олексіївна,
МІЩЕНКО Яніна Олегівна,
ЧОРНИЙ Антон Юрійович,
АНТОНЮК Олександр Андрійович

СТАТИСТИКА

Підручник

Редактор К. Я. Савчук
Комп'ютерне верстання І. І. Віннік
Дизайн обкладинки Г. В. Поліщук

Формат 60x84/16. Ум. друк. арк. 16,95. Тираж 600 пр. Зам. 164.

Видавець і виготовлювач

Київський національний торговельно-економічний університет
вул. Кіото, 19, м. Київ-156, Україна, 02156

Свідоцтво суб'єкта видавничої справи серія ДК № 4620 від 03.10.2013 р.