ДЕРЖАВНА СЛУЖБА СТАТИСТИКИ УКРАЇНИ
НАЦІОНАЛЬНА АКАДЕМІЯ СТАТИСТИКИ, ОБЛІКУ ТА АУДИТУ

С. С. Герасименко, Г. В. Голубова, М. Ю. Потапова, С. П. Червона

СТАТИСТИКА

НАВЧАЛЬНИЙ ПОСІБНИК

За редакцією
доктора наук з державного управління, професора,
члена-кореспондента Національної академії наук України
О. Г. Осауленка

Київ – 2022

УДК 311(075.8)
С 78

Герасименко С. С., Голубова Г. В., Потапова М. Ю., Червона С. П. Статистика: навч. посіб. / За ред. О. Г. Осауленка. Київ: НАСОА, 2022. 265 с.

ISBN

Автори:

С. С. Герасименко, доктор економічних наук – розд. 1.4;
Г. В. Голубова, кандидат економічних наук – розд. 10;
М. Ю. Потапова, кандидат економічних наук – розд. 1 (крім 1.4) – 5;
С. П. Червона, кандидат економічних наук – розд. 6–9.

За редакцією доктора наук з державного управління, члена-кореспондента Національної академії наук України О. Г. Осауленка

Рецензенти:

Ковтун Н. В., доктор економічних наук, професор,
Київський національний університет імені Тараса Шевченка;
Моторин Р. М., доктор економічних наук, професор,
Київський національний торговельно-економічний університет

Рекомендовано до друку Вченою радою Національної академії статистики обліку та аудиту (протокол №3 від 04.11.2021 р.).

Розглянуто базові питання й положення організації та проведення статистичного дослідження соціально-економічних явищ і процесів. Зміст посібника узгоджено з типовою нормативною програмою дисципліни “Статистика” та міжнародними статистичними стандартами.

Розрахований на студентів, що навчаються за планом підготовки бакалаврів з економіки та менеджменту, слухачів навчальних закладів з надання другої економічної освіти, слухачів курсів перепідготовки та підвищення кваліфікації, державних службовців, працівників органів державної статистики. Буде корисний працівникам державних установ, банків.

© С. С. Герасименко, Г. В. Голубова, М. Ю. Потапова, С. П. Червона, 2021
© НАСОА, 2021

Зміст

РОЗДІЛ І.
МЕТОДОЛОГІЧНІ ЗАСАДИ СТАТИСТИКИ 6
1.1. Предмет статистики 6
1.2. Основні поняття і категорії статистики 10
1.3. Методи статистики 12
1.4. Система статистичних організацій світу 12

РОЗДІЛ 2.
СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ 21
2.1. Статистичне спостереження і статистична інформація 21
2.2. План статистичного спостереження 22
2.3. Помилки статистичного спостереження і контроль даних 25
2.4. Форми, види і способи статистичного спостереження 27

РОЗДІЛ 3.
ЗВЕДЕННЯ І ГРУПУВАННЯ СТАТИСТИЧНИХ ДАНИХ 35
3.1. Сутність статистичного зведення, класифікації і групування 35
3.2. Принципи формування груп 36
3.3. Види групувань 38
3.4. Статистичні таблиці 41

РОЗДІЛ 4.
СТАТИСТИЧНІ ПОКАЗНИКИ 54
4.1. Сутність статистичних показників 54
4.2. Абсолютні та відносні величини 55
4.3. Види відносних величин 56
4.4. Середні величини, їх види та форми 60

РОЗДІЛ 5.
РЯДИ РОЗПОДІЛУ.
АНАЛІЗ ВАРІАЦІЙ ТА ФОРМИ РОЗПОДІЛУ 78
5.1. Сутність і види рядів розподілу 78
5.2. Характеристики центру розподілу 80
5.3. Характеристики варіації 82
5.4. Оцінювання рівномірності розподілу, подібності структур, структурних зрушень 85
5.5. Правило розкладання дисперсій 87

РОЗДІЛ 6.
ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ 105
6.1. Завдання вибіркового спостереження 105
6.2. Методи і способи відбору одиниць до вибіркової сукупності ..108
6.3. Оцінювання похибок вибірки та необхідного обсягу вибірки 111
6.4. Способи поширення даних вибіркового спостереження на генеральну сукупність 114

РОЗДІЛ 7.
СТАТИСТИЧНІ МЕТОДИ
ВИМІРЮВАННЯ ВЗАЄМОЗВ'ЯЗКІВ 131
7.1. Види взаємозв'язків між явищами 131
7.2. Метод аналітичного групування 133
7.3. Основи кореляційно-регресійного аналізу 135
7.4. Непараметричні методи вимірювання взаємозв'язків 140

РОЗДІЛ 8.
АНАЛІЗ ІНТЕНСИВНОСТІ ДИНАМІКИ
І ТЕНДЕНЦІЙ РОЗВИТКУ 165
8.1. Поняття про ряди динаміки, їх елементи та правила побудови 165
8.2. Статистичні характеристики динамічних рядів 166

8.3. Середні характеристики динамічних рядів 169
8.4. Методи аналізу тенденції розвитку 170
8.5. Коливання та сталість динамічних рядів 173

РОЗДІЛ 9.
ІНДЕКСИ 199
9.1. Поняття і способи побудови індексів 199
9.2. Середньозважені індекси 203
9.3. Система взаємопов'язаних індексів
та визначення впливу окремих факторів 204
9.4. Індекси середніх величин, їх взаємозв'язок 206

РОЗДІЛ 10.
БАГАТОВИМІРНІ МЕТОДИ АНАЛІЗУ 224
10.1. Статистичні методи порівняння
соціально-економічних об'єктів 224
10.2. Формування сукупностей з подібними елементами 227
10.3. Оптимізація інформації у статистичному аналізі 233

ОСНОВНІ КАТЕГОРІЇ ТА ПОНЯТТЯ 257

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 263

Розділ І.
МЕТОДОЛОГІЧНІ ЗАСАДИ СТАТИСТИКИ

1.1. Предмет статистики

Протягом усієї історії становлення і розвитку статистики як науки навколо визначення її наукового змісту й насамперед предмета, методів та теоретичних засад виникали бурхливі дискусії. У 1839 році професор статистики Харківського університету О. Рославський почав свою вступну лекцію на тему, чи є статистика наукою, з оцінки стану тогочасної науки, підкреслюючи, що немає жодної науки, яка б так мало була визначена і характеризувалася б такою розмаїтістю суперечливих думок, як статистика. Так, одні науковці звеличують похвалами її сцієнтичну (з лат. scientia – наука) гідність, не підтверджуючи утім на досвіді свої блискучі теорії. На думку інших, це навіть не знання, а тільки запас більш-менш корисних відомостей [36, с. 90].

Термін "**статистика**" походить від латинського слова "статус" (status), яке має декілька значень: під ним розуміють "положення", "стан", воно ж означає і "держава". Саме слово "статистика" в науковий обіг увів у 1746 році німецький вчений, професор філософії і права Готфрід Ахенваль (1719–1772).

У XVII столітті в Західній Європі склалися умови, які призвели до становлення статистики. Величезних змін зазнала сфера соціально-економічних відносин, інтенсивно розвивалися товарно-грошові відносини, зовнішня торгівля займала все більше місце в господарстві європейських країн. Розширювалися зовнішні ринки, збільшувався торговий капітал. Усе це посилювало соціальне замовлення на те, що нині ми називаємо *інформацією*. Зародження статистики стало неминучим. Вона виникла в XVII ст. на європейському континенті майже одночасно у Німеччині й у Великій Британії. Однак форми

виникнення та її зміст були різними: у першому випадку – як *державоведення*, в другому – як *політична арифметика*. Останній термін підкреслював, що йдеться про прикладання математики до політики, про вимірювання фактів соціально-економічного життя (розуміючи під терміном "політика" науку про суспільство).

Державоведення визнавало державу єдиним об'єктом спостереження. Цьому напряму була притаманна недооцінка кількісного аспекту пізнання. Числа трактувались як окремий випадок опису, звідси й інша назва – описова статистика. Пізніше з державоведення виділилися статистика, економічна і політична географія, а з політичної арифметики – політична економіка, статистика, демографія.

На початковому етапі розвитку статистики виділяється фігура Вільяма Петті (1623–1687) – англійського економіста, винахідника (як він себе назвав) нового числового методу умовиводу.

Визначаючи зміст статистики та її завдання, всесвітньо відомий бельгійський учений-статистик Адольфа Кетле (1796–1874) зазначав, що статистика вивчає державу у певну епоху, збирає елементи, пов'язані з життям цієї держави, намагається зробити їх порівнянними та комбінує їх щонайкраще, щоб пізнати все те, що вони можуть висвітлити [36, с. 90].

Наприкінці XVII ст. був відкритий закон, пізніше названий **законом великих чисел**, який став математичним підґрунтям теорії статистики. Сутність цього закону полягає в тому, що при достатньо великому числі спостережень випадкові відхилення від середньої величини взаємно гасяться і в середніх числах відображаються порядок явищ та їх закономірності. Статистика стає **наукою, що вивчає закономірності масових явищ**.

Одну з перших проб періодизації історії статистичної науки містить праця відомого німецького вченого-статистика, професора Мюнхенського університету Георга Майєра "Статистика і суспільствознавство". Він запропонував розділити весь процес розвитку статистичної науки на дві частини, що розвиваються паралельно та пов'язані між собою різними зв'язками:

1) *статистика як мистецтво* або *статистичне мистецтво*, тобто елементарне спостереження соціальних мас (Г. Майєр

трактував статистику як науку про соціальні маси) та елементарне зведення результатів спостереження з їх використанням лише для практичних життєвих цілей, насамперед для державного управління, але без будь-якої подальшої їх наукової розробки (має тривалу тисячолітню історію);

2) *статистична наука*, що спирається на зібраний під час елементарного масового спостереження матеріал, використовуючи його для з'ясування сутності соціальних умов і явищ та виводячи закономірності, які існують у досліджуваних соціальних масах (статистичній науці передувало державоведення; для її розвитку важливе значення мали праці представників школи політичної арифметики) [35, с. 111].

У всіх розвинених країнах статистика є однією з фундаментальних складових економічної освіти, жоден економіст не може обійтися без знання статистики. Але статистична підготовка необхідна фахівцю в будь-якій сфері.

Надати визначення статистиці як науці намагалися філософи, математики, економісти, соціологи, державні діячі та, звісно, самі статистики. Під сучасним терміном "статистика" найчастіше розуміють таке:

- статистика – одна з форм практичної діяльності людей, мета якої – збирання, обробка й аналіз масових даних про ті чи інші явища;
- статистика – наука, яка має свій предмет і свою методологію, займається розробкою теоретичних положень та методів, що використовуються на практиці;
- статистика – статистичні дані, статистична інформація, сукупність підсумкових цифрових показників, що характеризують різні сторони життя держави: політичні відносини, культуру, населення, виробництво тощо;
- статистика – метод пізнання, інструментарій, який використовують у суспільних науках для встановлення закономірностей масових явищ і процесів;
- статистика – як еквівалент словосполучення "статистичні методи". Статистичні методи можна охарактеризувати як

такі, що застосовуються для збирання, подання, аналізу й інтерпретації даних.

Об'єктом статистики є суспільство, явища і процеси суспільного життя. Визначимо основні *риси та особливості предмета статистичної науки*:

- у статистиці як науці досліджуються не окремі факти, а масові соціально-економічні явища і процеси, що виступають як множина окремих фактів, які мають як індивідуальні, так і загальні ознаки;
- статистика вивчає насамперед кількісний бік суспільних явищ і процесів у конкретних умовах місця і часу;
- статистика характеризує структуру суспільних явищ. Структура – це внутрішня будова статистичної множини. Статистика повинна цю структуру виявити, виразити і відобразити за допомогою статистичних показників;
- у статистиці зміни рівня і структури явища досліджуються у часі, тобто в динаміці;
- явища суспільного життя взаємопов'язані та взаємозумовлені: зміна одних явищ визначає інші. Тому виявлення зв'язків є особливістю статистики як науки.

Отже, **предметом статистики** є розміри та кількісні співвідношення масових суспільних явищ, закономірності їх зв'язку, формування і розвитку.

Статистика – це наука, яка вивчає кількісний аспект масових суспільних явищ і процесів у нерозривному зв'язку з їх якісним аспектом, досліджує кількісне вираження закономірностей суспільного розвитку в конкретних умовах місця і часу.

Статистика – багатогалузева наука. Дотепер виділилися такі її складові частини:

1) *загальна теорія статистики*, яка розглядає категорії статистичної науки, а також методи і засоби аналізу, спільні для будь-яких масових явищ;
2) *економічна статистика*, яка вивчає явища і процеси, що мають місце в економіці, розробляє систему економічних показників;

3) *галузеві статистики*, які вивчають окремі галузі народного господарства чи суспільного життя, розробляють та аналізують статистичні показники розвитку відповідних галузей, зокрема промисловості, сільського господарства, транспорту, охорони довкілля тощо;
4) *соціально-демографічна статистика*, що формує та аналізує систему показників, які комплексно характеризують різні аспекти соціального життя та праці населення, споживання матеріальних благ і послуг, народонаселення тощо;
5) *статистика підприємств*, об'єктом вивчення якої є підприємства, а її інструментами – статистичний реєстр підприємств, міжнародні стандартні економічні класифікації як основа для відповідних національних класифікацій і методологічно узгоджені показники.

1.2. Основні поняття і категорії статистики

Статистика як наука має свій особливий понятійний апарат, що охоплює найбільш характерні категорії, поняття та визначення.

Так, **ознакою** у статистиці називають відмінну рису, властивість, якість, що є характерною для окремих одиниць, об'єктів (явищ). Ознаки можуть бути кількісними й атрибутивними (якісними). *Кількісні ознаки* мають числове вираження (наприклад вік, стаж роботи, рентабельність тощо), *атрибутивні* (стать людини, релігія, професія, колір очей тощо) – не мають.

Кількісні ознаки поділяються на дискретні (перервні) та неперервні. *Дискретні* ознаки мають тільки деякі певні значення, між якими не може бути проміжних значень, і виражені окремими цілими числами (наприклад, кількість дітей у родині, кількість кімнат у квартирі тощо). *Неперервні* ознаки можуть в окремих межах набувати будь-яких значень.

Кількісні зміни значень ознаки при переході від однієї одиниці сукупності до іншої називаються *варіацією*.

Ознаки мають різні рівні вимірювання, що відображається у відповідних типах шкал. Є *три основні типи шкал*:

1) *номінальна шкала* – це шкала найменувань, порядок розташування яких не має значення (наприклад, перелік видів економічної діяльності, форм власності тощо);
2) *порядкова (рангова) шкала* – це шкала, для якої порядок значень ознаки має значення; така шкала встановлює відношення послідовності (порядку) – ранги; наприклад, рівень інвестиційної привабливості підприємств може бути різним: низьким (ранг 1), середнім (ранг 2), високим (ранг 3);
3) *метрична шкала* – це звичайна шкала дійсних чисел, її використовують для вимірювання фізичних величин (ваги, часу і т. ін.) або результатів обчислення (наприклад, кількість зареєстрованих злочинів, продаж алкогольних напоїв на одну особу тощо);

Вивчаючи множину явищ або фактів, статистика в кожному випадку має справу з масою, сукупністю. **Статистична сукупність** складається з одиниць, яким притаманні загальні властивості, однак вони розрізняються між собою за однією чи декількома ознаками. Наприклад, є сукупність студентів одного навчального закладу. Оскільки усі вони навчаються в одному закладі, то являють собою одне ціле, що надає можливість визначити їх загальну чисельність, встановити загальну суму нарахованої їм стипендії тощо. Однак вони відрізняються за статтю, віком, успішністю та ін., що приводить до необхідності вивчати склад сукупності за цими ознаками, визначати середній бал успішності тощо.

Статистична закономірність – закономірність зміни у просторі й часі масових явищ і процесів суспільного життя, які складаються з множини елементів (одиниць сукупності). Під закономірністю взагалі розуміють повторюваність, послідовність і порядок зміни в явищах.

Усі соціально-економічні явища і процеси статистика вивчає за допомогою статистичних показників, з допомогою яких створюється, передається і зберігається статистична інформація. Під **статистичним показником** розуміють узагальнену кількісну характеристику соціально-економічних явищ і процесів у їх визначеності щодо конкретних умов місця і часу.

1.3. Методи статистики

Статистична методологія – це комплекс спеціальних, властивих лише статистиці методів і засобів дослідження. Будь-яке статистичне дослідження складається з чотирьох послідовних етапів:

1) збирання первинного статистичного матеріалу шляхом реєстрації фактів або опитування респондентів;
2) систематизація і групування зібраних даних;
3) перехід від характеристики окремих елементів до узагальнюючих показників у формі відносних або середніх величин;
4) аналіз варіації, динаміки, взаємозв'язків (результати аналізу подаються текстом, у формі таблиць і графіків).

До статистичних методів дослідження належать: метод статистичного спостереження; метод статистичних групувань; метод абсолютних та відносних величин; середніх величин; табличний; графічний; аналізу варіації; вибірковий; кореляційно-регресійного аналізу; аналізу динаміки; індексний та ін.

1.4. Система статистичних організацій світу

Основне завдання міжнародних організацій у сфері статистики – збирання порівнянних даних від національних статистичних служб, узагальнення та поширення. Якісне й ефективне виконання цього завдання можливе тільки за наявності єдиної методологічної бази та спроможності країн представляти статистику в повному обсязі відповідно до рекомендацій міжнародних організацій. Отже, розвиток статистичного потенціалу країн – ще один важливий напрям діяльності в рамках глобальної статистичної системи.

У цілому на сьогодні джерелами даних для міжнародної статистики є близько двох тисяч міжнародних організацій, які регулярно збирають та обробляють статистичні дані. Можна виділити такі важливі за масштабом і характером статистичних даних групи організацій:

- всесвітні – організації системи ООН, які вхідну інформацію окремих країн зводять у регіональні та світові

підсумки за певною системою і спільними методами розрахунку показників;

- регіональні – організації певних угруповань країн (таких як ЄС, країни Тихоокеанського регіону та ін.), статистичні служби яких збирають і узагальнюють дані та публікують інформацію щодо цих груп країн з огляду на завдання таких організацій;
- професійні – організації, що мають на меті розроблення статистичних засад аналізу соціально-економічних явищ, зокрема Міжнародний статистичний інститут, Панамериканський статистичний інститут та ін.;
- вузькопрофесійні – міжнародні організації, що узагальнюють інформацію, яка надходить від окремих країн, щодо окремих видів продукції та послуг, зокрема міжнародні організації виробників кави, цукру, олії та інших сировинних продуктів.

Отже, основними механізмами ефективного функціонування глобальної статистичної системи є координація, спеціалізація і розвиток робіт у галузі статистичних досліджень. Нині існує такий набір інструментів координації статистичної діяльності та об'єднання зусиль різних міжнародних організацій у сфері статистики:

- ✓ щорічні сесії СК ООН;
- ✓ регулярні наради Комітету з координації статистичної діяльності, що об'єднує всі міжнародні організації, які займаються статистичною діяльністю;
- ✓ діяльність Міжнародного статистичного інституту;
- ✓ Міжнародні статистичні стандарти.

Міжнародний статистичний інститут (MCI, International Statistical Institute, ISI) утворено у 1885 р. як наукову організацію, що об'єднала представників статистичної науки та практики 50 країн. Метою діяльності Міжнародного статистичного інституту є удосконалення статистичних методів збирання й оброблення даних та сприяння їх застосуванню в окремих країнах і міжнародних організаціях. Від 1916 р. Інститут видає статистичний щорічник. Членами MCI було і зараз є багато українських учених-статистиків.

У 1920 р. була заснована Ліга Націй. Створена у Секретаріаті Ліги секція економіки та фінансів з допомогою Міжнародного статистичного інституту розпочала розробку системи показників міжнародної статистики (торгівлі, рибальства, товарних запасів).

Сучасна система міжнародної статистики сформувалася після Другої світової війни. У центрі системи – статистичні органи ООН (Статистична комісія та Статистичний відділ Секретаріату ООН), з ними співпрацюють інші міжнародні організації, що мають власні статистичні служби, які вносять пропозиції щодо розробки й упровадження відповідних статистичних стандартів. Тим самим ООН заклала основи міжнародної статистики на однакових методологічних принципах і координує статистичні роботи для окремих країн та міжнародних організацій.

Статистична комісія ООН (СК ООН, United Nations Statistical Commission, UNSC) була утворена у 1946 р. і є найвищим органом глобальної статистичної системи. Вона об'єднує керівників статистичних відомств країн-членів ООН усього світу. Комісія – головний орган, який приймає рішення в галузі міжнародної статистичної діяльності, зокрема щодо статистичних стандартів, розвитку концепцій і методів та їх упровадження на національному й міжнародному рівнях.

Статистична комісія ООН є функціональною комісією Економічної і соціальної ради (ЕКОСОР) ООН і надає їй сприяння у розвитку національних статистичних систем, підвищенні порівнянності національних статистичних даних, координації статистичної роботи спеціалізованих агентств ООН й удосконаленні статистики та статистичної методології. Останнім часом зростає роль СК ООН як глобального координатора міжнародної статистичної діяльності; зусилля Комісії спрямовані, зокрема, на створення єдиної системи збирання, обробки та розповсюдження міжнародної статистичної інформації. Комісія складається з 24 країн-членів ООН, обирається ЕКОСОР ООН на основі рівноправного географічного представництва й оновлюється за принципом ротації. Членство обмежене чотирирічним терміном.

Однією з істотних функцій СК ООН є узагальнення досвіду країн щодо вдосконалення й уніфікації статистичної інформації у світі з метою внесення відповідних коректівів у систему міжнародних стандартів обліку та статистики. Для цього утворюються робочі групи фахівців, які вивчають усі нововведення в різних галузях статистики і готують свої проєкти пропозицій щодо вдосконалення діючих або вироблення нових міжнародних стандартів ООН. Стандарти розроблялись і розробляються фахівцями міжнародних організацій, що входять у систему установ ООН. Крім цього, діють також стандарти інших організацій, що не входять в систему ООН. Вони ґрунтуються на принципах стандартів ООН, а також рекомендаціях фахівців, які працюють у системі ООН.

Основне завдання усіх міжнародних стандартів полягає в забезпеченні спільних засад побудови національних систем статистичної інформації, порівнянності показників за їх змістом і методологією обчислення, щоб уможливити аналіз розвитку країн у зіставленні з регіональними та світовими явищами. Особливо важливо дотримуватися таких стандартів при публікації даних у статистичних щорічниках, довідниках, бюлетенях, аналітичних доповідях ООН та ін.

Статистична комісія ООН як головний механізм об'єднання зусиль статистиків усіх країн світу та її регіональні відділення, зокрема Конференція європейських статистиків, на своїх засіданнях і в поточній роботі відповідних секретаріатів постійно приділяють увагу проблемам оптимізації збирання статистичних відомостей від національних статистичних служб.

Статистичний відділ ООН (СВ ООН, United Nations Statistics Division, UNSD) є структурним підрозділом Департаменту економічних і соціальних питань ООН. **Статистичний відділ** був створений одночасно з СК ООН у 1947 р. як орган Секретаріату ООН для сприяння виконанню завдань, що стоять перед Статистичною комісією і глобальною статистичною системою. СВ ООН складається з професійних статистиків та економістів з усього світу. До його складу входять п'ять основних підрозділів: з економічної статистики; демографічної та соціальної статистики; статистики енергетики та навколишнього середовища; статистики торгівлі; поширення статистичної

інформації. Отже, у своїй діяльності СВ ООН охоплює практично всі основні напрями статистики. До спеціалізованих агенцій системи ООН належать ті, що пов'язані з ООН спеціальними угодами і мають підрозділи для узагальнення статистичної інформації у своїх сферах діяльності, зокрема:

- Продовольча та сільськогосподарська організація, ФАО;
- Міжнародна організація праці, МОП;
- Міжнародний валютний фонд, МВФ;
- Світовий банк, СБ;
- Всесвітня організація охорони здоров'я, ВООЗ;
- Організація з питань освіти, науки та культури, ЮНЕСКО та ін.

Міжнародна статистика поєднує також діяльність та інформацію інших організацій, які не входять до системи ООН. Комітет з координації статистичної діяльності (Committee for the Coordination of Statistical Activities, CCSA), створений у 2002 р, відіграє важливу роль в узгодженні статистичної діяльності міжнародних організацій. Він сприяє координації та інтеграції статистичних програм, здійснюваних міжнародними організаціями. Це дозволяє уникати дублювання в роботі, знизити навантаження на країни щодо подання статистичної інформації, уніфікувати підходи до вирішення статистичних питань. До складу Комітету входять СВ ООН, інші підрозділи Департаменту ООН з економічних і соціальних питань, регіональні комісії, а також більшість програм, фондів та спеціалізованих агентств ООН, міждержавні організації, регіональні банки розвитку в особі керівників статистичних відділів. Щорічно Комітет звітує про виконану роботу перед СК ООН. Найважливішим досягненням Комітету з координації статистичної діяльності стало прийняття в 2005 р. Принципів управління міжнародної статистичної діяльністюОрганізація Об'єднаних Націй має п'ять регіональних комісій, сформованих за географічним принципом: ЕКА (Економічна комісія для Африки); ЄЕК (Європейська економічна комісія); ЕСКАТО (Економічна і соціальна комісія для Азії і Тихого океану); ЕСКЗА (Економічна і соціальна комісія для Західної Азії); ЕКЛАК (Економічна комісія для Латинської Америки і Карибського басейну).

Практично всі спеціалізовані служби, органи та фонди системи ООН, інші наднаціональні й регіональні організації мають статистичні підрозділи і спеціалізуються на розвитку окремих напрямів статистики та збиранні статистичних даних відповідно до специфіки своєї діяльності. Статистичні дані від країн збирають за допомогою міжнародних статистичних запитальників або формуються міжнародними організаціями самостійно на основі власних обстежень, оцінок і розрахунків. На підставі даних, зібраних від країн з використанням запитальників, Статистичний відділ секретаріату ООН публікує інформацію щодо національних рахунків у статистичних довідниках: National Accounts Statistics: Main Aggregates and Detailed Tables; Statistical Yearbook; Trends in Europe and North America. МВФ аналогічні дані публікує в довіднику International Financial Statistics.

Статистичне управління європейського співтовариства (Євростат, Statistical office of the European Union, Eurostat) – статистична служба Європейського Союзу (ЄС), заснована у 1953 р. Євростат працює для задоволення попиту на статистичну інформацію і переважне місце в його роботі займає вирішення питань узгодження статистичної діяльності як країн-членів ЄС, так і інших країн, що потребують допомоги ЄС. Євростат отримує від національних статистичних служб різноманітну статистичну інформацію, узагальнює і оприлюднює її у формі понад 100 публікацій на рік. Серед них найбільше практичне значення мають методологічні пояснення до статистичних щорічників та місячних бюлетенів. Особливу роль відіграють вказівки щодо проведення спеціальних обстежень – переписів населення, промисловості, сільського господарства тощо, – в яких містяться роз'яснення щодо порядку і методології обчислення відповідних показників.

У публікаціях-довідниках наводяться міжнародні стандарти, які періодично уточнюються й оновлюються. Це, зокрема:

- ✓ номенклатура територіальних одиниць;
- ✓ країни світу та національності;
- ✓ митні зони світу;
- ✓ перелік статистичних показників, що їх розробляють міжнародні організації;

- ✓ коефіцієнти перерахунку з національних одиниць вимірювання у метричні;
- ✓ довідник методів реєстрації природного руху населення та ін.

Організація економічного співробітництва та розвитку (ОЕСР, Organization for Economic Cooperation and Development, OECD) – міжнародна міжурядова економічна організація, створена з ініціативи США в 1961 р. Основною метою ОЕСР є аналіз стану економіки держав-членів і вироблення рекомендацій щодо здійснення ними економічного регулювання на галузевому та макрорівнях.

У структурі ОЕСР функціонує Директорат статистики, який забезпечує формування на порівнянній основі економічної статистики для проведення аналітичної діяльності Організації. Крім того, в рамках ОЕСР створено Комітет зі статистики що забезпечує координацію статистичної діяльності країн-членів ОЕСР та інших найбільших країн, а також прийняття рекомендацій у галузі статистики.

ОЕСР активно поширює свою статистичну інформацію, забезпечуючи вільний доступ до потужної відкритої бази даних OECD.stat, що є сховищем всієї статистичної інформації ОЕСР. Основним друкованим статистичними виданням ОЕСР є OECD Factbook – щорічник, який містить глобальний огляд всесвітніх економічних, соціальних та екологічних тенденцій. Також ОЕСР випускає популярні статистичні бюлетені та довідники за статистикою країн-членів, тематичні збірники за напрямами статистики.

Світова організація торгівлі (СОТ, World Trade Organization, WTO) була заснована в 1994 р. і є спадкоємицею Генеральної угоди про тарифи і торгівлю (ГАТТ). СОТ покликана регулювати торгово-політичні відносини учасників на основі ряду обов'язкових багатосторонніх угод з торгівлі товарами, послугами, правами інтелектуальної власності. СОТ також є форумом для переговорів між її членами з питань міжнародної торгівлі. За всю історію ГАТТ/СОТ відбулося вісім раундів багатосторонніх торговельних переговорів.

Для виконання своїх функцій СОТ розпоряджається великим масивом статистичних даних. Статистична база даних СОТ містить дані з торгівлі (Trade Profiles), що представляють стандартні

інформаційні файли щодо торговельної ситуації у країнах-членах, країн-спостерігачів та інших окремих країн, дані за тарифами (Tariff Profiles), що забезпечують інформацію про ситуацію з доступом на ринки, а також секцію часових рядів з інтерактивним пошуком даних щодо статистики торгівлі. Крім того, СОТ випускає щорічники "Міжнародна статистика торгівлі" і "Статистика торгівлі", які містять статистику про розвиток всесвітньої торгівлі, а також основні економічні показники, показники торговельної політики, торгівлі товарами і послугами, промислової власності.

СОТ є активним учасником міжнародної статистичної діяльності, виступаючи ініціатором створення міжвідомчих цільових груп ООН з питань статистики міжнародної торгівлі послугами і товарами, а також основним автором "Керівництва зі статистики міжнародної торгівлі послугами" і "Навчального модуля з вимірювання торгівлі послугами".

Світовий банк (СБ, World Bank, WB) є одним із найважливіших джерел фінансової і технічної допомоги, яка надається країнам, що розвиваються світу. Банк складається з двох унікальних організацій розвитку, до яких належать 185 країн-членів, – Міжнародного банку реконструкції та розвитку і Міжнародної асоціації розвитку. Кожна з цих двох установ відіграє особливу роль у процесі досягнення мети, що стоїть перед банком і полягає у скороченні масштабів бідності в глобальному масштабі та підвищенні рівня життя населення планети. Світовий банк націлений на постійне підвищення якості та кількості статистичних даних про всі аспекти процесу розвитку.

Світовий банк – активний користувач і розповсюджувач статистичних даних. Основна статистична продукція СБ – щорічна публікація "Показники глобального розвитку", що містить більше ніж 900 показників, а також "Атлас Цілей розвитку тисячоліття", що представляє карти за ключовими індикаторами ЦСР.

Міжнародний валютний фонд (МВФ, International Monetary Fund, IMF) – міжурядова фінансово-кредитна організація, що має статус спеціалізованої установи ООН. Фонд заснований у 1945 р. відповідно до рішення Бреттон-Вудської конференції. Основними завданнями МВФ

є підтримка стабільності світової фінансової системи, сприяння стійкості валютних курсів і рівноваги платіжних балансів країн-членів, розвиток міжнародного співробітництва у валютно-фінансовій сфері.

Традиційно Департамент статистики МВФ є одним із ключових гравців у рамках міжнародного статистичного співробітництва. Сфера інтересів МВФ у галузі статистики – це насамперед національні рахунки, платіжний баланс, а також фінансова статистика. МВФ є автором міжнародних рекомендацій і стандартів за цими напрямами. Щорічно МВФ готує доповіді про дотримання стандартів та кодексів (ROSCs) на основі аналізу ситуації в окремих країнах.

Запитання для самоконтролю

1. Від якого слова походить термін “статистика”? Що під ним розуміють?
2. Чим було викликане зародження статистики?
3. Які Ви можете назвати статистичні школи?
4. Якою є сутність закону великих чисел?
5. Що Ви можете віднести до основних рис та особливостей предмету статистичної науки?
6. Які складові частини статистики як багатогалузевої науки Ви можете виділити?
8. Яке визначення статистичної ознаки Ви можете надати? Які види ознак знаєте? Наведіть їх приклади.
9. Що розуміють під статистичною сукупністю?
10. Що розуміють під статистичною закономірністю?
11. Що являє собою статистичний показник?
12. Що Ви розумієте під статистичною методологією?
13. За якими етапами здійснюється статистичне дослідження?
14. Які статистичні методи дослідження Ви знаєте?
15. Яким є основне завдання міжнародних організацій у сфері статистики?
16. Які групи організацій за масштабом і характером статистичних даних Ви можете назвати?

Розділ 2.
СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ

2.1. Статистичне спостереження і статистична інформація

Першим етапом будь-якого дослідження є збирання інформації, а саме, статистичне спостереження. Це складна і відповідальна робота. Помилки, допущені у процесі статистичного спостереження, дуже важко виправити на його наступних етапах. Тому при організації та проведенні статистичного спостереження необхідно точно дотримуватися відповідних правил і вимог.

Статистичне спостереження – це спланована, науково організована реєстрація масових даних про соціально-економічні явища та процеси. Під час статистичного спостереження дістають необхідну статистичну інформацію, яку потім систематизують, зводять, обробляють, аналізують і узагальнюють. Статистичне спостереження може бути первинним або вторинним. *Первинне статистичне спостереження* – це реєстрація даних, що надходять безпосередньо від об'єкта, який їх продукує. *Вторинне* – це збирання раніше зареєстрованих та оброблених даних.

Статистичні дані – масові системні характеристики соціально-економічних явищ і процесів. Від якості даних статистичного спостереження залежать результати подальшого дослідження. Статистичні дані повинні відповідати певним вимогам:

- ✓ вірогідність даних, тобто їх відповідність реальному стану;
- ✓ повнота даних (за обсягом і по суті);
- ✓ своєчасність даних (інформація має дійти до користувача раніше, ніж застаріє);

- ✓ порівнянність даних у часі або у просторі (дані мають бути порівнянними за складом сукупності, за одиницями вимірювання, за методикою збирання даних та обчислення статистичних показників, за територіальною належністю одиниць);
- ✓ доступність даних (доступ до статистичної інформації забезпечується через систематичну публікацію її у друкованих виданнях, поширення її засобами масової інформації, безпосереднє її надання органам державної влади й органам місцевого самоврядування, іншим юридичним і фізичним особам).

Статистична інформація – це система статистичних показників, що характеризують соціально-економічні явища і процеси й використовуються в управлінні.

2.2. План статистичного спостереження

Статистичне спостереження здійснюється за такими *етапами*: підготовка спостереження, реєстрація статистичних даних, формування бази даних. Найважливішим елементом підготовки спостереження є розробка плану його проведення. При складанні плану обов'язково необхідно враховувати кінцеву мету дослідження.

План статистичного спостереження – це сукупність програмно-методологічних і організаційних питань. До *програмно-методологічних питань* відносять такі:

- ✓ для чого воно проводиться (*мета спостереження*);
- ✓ що обстежується (*об'єкт спостереження*);
- ✓ складові частини об'єкта (*одиниця сукупності*);
- ✓ джерело інформації (*одиниця спостереження*);
- ✓ на які питання планується одержати відповіді (*програма спостереження*).

У мету статистичного спостереження закладається основний очікуваний результат статистичного дослідження. Мета має бути сформульована ясно, цілком чітко й розгорнуто, із зазначенням конкретних завдань, які необхідно вирішити для її досягнення. Кінцевою

метою є підготовка управлінських рішень та вжиття відповідних заходів.

Відповідно до завдань і поставленої мети визначаються об'єкт і одиниця спостереження.

Об'єкт статистичного спостереження – це сукупність явищ, що підлягають дослідженню. Визначити об'єкт спостереження означає встановити, яка саме сукупність підлягає спостереженню, а також її межі й напрями вивчення. Для цього потрібні глибокі знання про об'єкт. Інколи для відмежування об'єкта спостереження вдаються до *цензу*, під яким розуміють одну або декілька встановлених ознак, що визначають межі об'єкта. Для об'єкта статистичного спостереження характерне те, що його не можна вивчати безпосередньо у цілому, потрібне виділення в його складі окремих одиниць.

Одиниця сукупності – це первинний елемент об'єкта статистичного спостереження, який є носієм ознак, що підлягають спостереженню. Від одиниці статистичної сукупності слід відрізняти одиницю спостереження.

Одиниця спостереження – це первинна одиниця, від якої одержують інформацію. Інакше кажучи, одиниця сукупності – це те, що підлягає дослідженню, а одиниця спостереження – це джерело одержання відомостей.

Наприклад, якщо метою спостереження є визначення кількості наявного устаткування та його придатності на підприємствах певного виду діяльності, то при проведенні перепису устаткування об'єктом спостереження буде сукупність одиниць усього устаткування (встановленого та невстановленого, діючого та недіючого). Одиницею сукупності виступатиме окрема одиниця устаткування, а одиницею спостереження – підприємство. Одиниця сукупності та одиниця спостереження можуть збігатися. Наприклад, при проведенні опитування студентів навчального закладу щодо якості викладання курсу одиницею спостереження і одиницею сукупності виступатиме студент.

Після того, як визначені мета, об'єкт, одиниця сукупності й одиниця спостереження, переходять до складання програми спостереження.

Програма спостереження – це перелік запитань, на які необхідно одержати відповіді в результаті спостереження. При складанні програми велике значення має чітке формулювання запитань. Для більшості статистичних спостережень це складна і трудомістка робота, у виконанні якої бере участь багато людей. Поставлені запитання мають бути однаково зрозумілими для всіх. До програми спостереження також включаються: розробка статистичного інструментарію, визначення виду і способу обстеження.

Статистичний інструментарій – це набір статистичних формулярів, інструкцій та роз'яснень щодо проведення спостереження.

Статистичний формуляр – обліковий документ єдиного зразка, що містить адресну характеристику об'єкта спостереження та статистичні дані про нього. Для забезпечення однакового тлумачення програми спостереження у формулярі, в якому вона міститься, подаються пояснення, викладені в документі, що називається *інструкцією*. Обов'язковим елементом статистичного формуляра є його окрема частина, необхідна для перевірки зібраних даних та їх обробки. Остання забезпечується завдяки застосуванню системи шкал. Сам же формуляр може мати вигляд звітності, переписних та опитувальних листків, анкет і т. ін.

Другою складовою частиною плану статистичного спостереження є *організаційні питання*, які визначають:

- ✓ хто проводить спостереження (*органи і персонал*);
- ✓ де проводиться спостереження (*місце спостереження*);
- ✓ коли проводиться спостереження (*час і період спостереження*);
- ✓ за допомогою чого проводять спостереження (*матеріально-технічне забезпечення*);
- ✓ який спосіб забезпечення точності результатів (*система контролю та пробні обстеження*).

Органами спостереження можуть бути різні установи системи державної статистики, відомства, підприємства, спеціальні інститути, агентства, міжнародні організації, громадські організації тощо. Також визначаються рівень підготовки організаторів і виконавців, їх права й обов'язки. Перед статистичним спостереженням проводять

відповідні підготовчі заходи, зокрема визначається потреба в кадрах, здійснюється їх добір та інструктаж.

При плануванні спостереження необхідно встановити, в межах якої території воно буде проводитися, а також місце, де безпосередньо відбуватиметься реєстрація. *Місцем спостереження* вважають пункт, де безпосередньо реєструються ознаки окремих одиниць сукупності. Здебільшого воно збігається з місцем перебування одиниць спостереження.

Для одержання достовірних даних важливе значення має точне встановлення періоду або моменту часу, до якого належать відомості, що реєструються. Також необхідно визначити час, протягом якого проводитиметься спостереження. Розрізняють об'єктивний і суб'єктивний час спостереження.

Об'єктивний час (*час спостереження*) – це час, до якого належать відомості, що реєструються (або певний момент часу, або його період). Якщо об'єктом спостереження є процес, то обирається інтервал часу, за який нагромаджуються дані, якщо ж певний стан – то критичний момент. Термін спостереження має максимально наближатися до критичного моменту.

Критичний момент – це момент, на який фіксуються дані про явище. Наприклад, критичним моментом перепису населення 2001 року в Україні була 12 година ночі з 4 на 5 грудня.

Суб'єктивний час (період спостереження) – це період, протягом якого реєструються ознаки об'єкта спостереження. Так, при переписі населення він становив 10 днів – з 5 по 14 грудня 2001 р. *Сезон спостереження* слід обирати в найсприятливіший період, коли об'єкт перебуває у звичайному для нього стані. Наприклад, перепис здійснюється, як правило, взимку, коли має місце найменша міграція населення.

2.3. Помилки статистичного спостереження і контроль даних

Завдання статистичного спостереження – одержання даних, які б точно відображали дійсність. Проте навіть при досконалій організації статистичного спостереження можуть виникати різні неточності та помилки, яким необхідно намагатися запобігти.

Неточності, неправильності у статистичних даних, одержаних при спостереженні, прийнято називати ***помилками спостереження***. Залежно від причини виникнення виділяють помилки реєстрації і помилки репрезентативності.

Помилки реєстрації можуть виникнути при як суцільному, так і несуцільному спостереженні. За природою виникнення розрізняють випадкові та систематичні помилки реєстрації. *Випадкові помилки реєстрації* виникають унаслідок збігу випадкових обставин, через неуважність, недбалість, некваліфікованість реєстратора або незосередженість респондента. Їх дія в масі випадків урівноважується і на результати не впливає. *Систематичні помилки реєстрації* виникають унаслідок постійних викривлень в одному напрямі. Їх дія призводить до зміщення результатів у бік збільшення або зменшення. Так, відома тенденція до заокруглення доходів до цілих чисел у бік зменшення, а витрат – у бік збільшення. Систематичні помилки поділяють на ненавмисні та навмисні. *Ненавмисні помилки* виникають через необґрунтованість програми спостереження, некомпетентність реєстраторів, неосвіченість респондентів. *Навмисні помилки* – внаслідок свідомого викривлення фактів з певною метою.

Помилки репрезентативності властиві тільки вибірковому спостереженню та виникають унаслідок несуцільного характеру реєстрації або порушення правил випадковості відбору.

Контроль даних – це засіб попередження, виявлення та виправлення помилок спостереження, що полягає в перевірці даних щодо повноти і вірогідності. Повнота даних перевіряється візуально, а їх вірогідність – за допомогою логічного та арифметичного контролів.

Логічний контроль – це перевірка сумісності даних порівнянням взаємозалежних ознак, яка встановлює лише наявність помилки, а не її величину. Наприклад, порівняння віку респондентів з їх сімейним станом або доходів домогосподарства з його витратами.

Арифметичний контроль – це перевірка зареєстрованих даних за допомогою прямих або непрямих перерахунків. Наприклад, загальну кількість підприємств по країні можна визначити, якщо відома кількість підприємств за різними видами економічної діяльності або розмірами.

2.4. Форми, види і способи статистичного спостереження

Різноманітність сфер спостереження зумовлює застосування різних організаційних форм, видів і способів статистичного спостереження. З погляду організації розрізняють такі його ***форми***: звітність, спеціально організовані спостереження, реєстри.

Звітністю називають форму спостереження, за якої кожен суб'єкт діяльності регулярно та у відповідні терміни подає свої дані до державних органів статистики у вигляді документів спеціально затвердженої форми (звітів). Звітність має юридичну силу і є обов'язковою.

За порядком надходження звітність поділяється на централізовану (загальнодержавну) і децентралізовану (відомчу). Перша подається як у вищу організацію, так і в органи державної статистики, друга – лише у вищу організацію і може мати закритий характер.

За періодичністю подання звітність поділяється на поточну (подається протягом року, наприклад, місячна, квартальна) і річну.

За ознаками звітність поділяють на типову (єдину для всіх підприємств) і спеціалізовану (для тих підприємств або окремих виробництв, які мають специфічні особливості).

Спеціально організоване спостереження – це одержання статистичних даних на основі переписів, разових обліків, спеціальних обстежень, опитувань. За допомогою спеціально організованих статистичних спостережень вивчають явища і процеси, не охоплені статистичною звітністю. Водночас інколи вони використовуються для перевірки звітних даних. Прикладом спеціально організованих спостережень можуть бути різні соціологічні та маркетингові дослідження; переписи населення, устаткування.

Перепис – суцільне спостереження масових явищ з метою визначення їх розміру та складу станом на певну дату. Переписи проводяться одночасно для всієї сукупності (території) за єдиною для всіх одиниць програмою (як правило, періодично, з рівним інтервалом). Наприклад, перепис населення рекомендовано проводити кожні 10 років.

Обліки – суцільні спостереження масових явищ, що ґрунтуються на даних опитувань, огляду та документальних записів. Наприклад, облік готівки в касі, облік об'єктів незавершеного будівництва.

Спеціальні обстеження – несуцільні спостереження окремих масових явищ за певною тематикою – здійснюються періодично або одноразово. Наприклад, бюджетні обстеження домогосподарств, обстеження з питань неформальної зайнятості.

Опитування – частіше несуцільні спостереження думок, мотивів, оцінок, які реєструються зі слів респондентів. Винятком із цього є суцільне опитування всього населення – *референдум,* масове волевиявлення з принципових соціально-політичних або економічних питань.

Статистичний реєстр – список або перелік одиниць певного об'єкта спостереження з наведенням необхідних ознак, який складається та оновлюється під час постійного відстежування. Прикладом статистичного реєстру є реєстр підприємств та організацій – перелік суб'єктів усіх видів економічної діяльності із зазначенням їх реквізитів і основних показників.

Статистичне спостереження може поділятися на окремі ***види*** за такими ознаками, як повнота охоплення одиниць сукупності та час реєстрації фактів.

За часом реєстрації фактів спостереження може бути одноразовим, періодичним і поточним.

Одноразові спостереження проводяться епізодично, в міру виникнення потреби в дослідженні явища або процесу. Наприклад, маркетингові дослідження думки споживачів щодо якості послуг, адаптації товару до місцевого ринку.

Періодичні спостереження – це реєстрація фактів через певні (як правило, рівні) проміжки часу. Наприклад, перепис населення, виробничих площ, технологій.

Поточні (неперервні) спостереження – це спостереження, які здійснюються в часі неперервно, коли факти, події та явища реєструються в міру їх виникнення. Наприклад, реєстрація актів цивільного стану населення, звітність.

За повнотою охоплення одиниць досліджуваної сукупності статистичне спостереження поділяється на два види: суцільне і несуцільне.

Суцільне обстеження передбачає реєстрацію всіх без винятку одиниць сукупності. Наприклад, при переписі населення обліком має бути охоплена кожна людина.

Несуцільне (часткове) спостереження – це спостереження тільки певної частини одиниць сукупності, яку вивчає статистика. Несуцільні спостереження залежно від завдань дослідження і характеру об'єкта поділяються на такі різновиди: вибіркове, основного масиву, монографічне, анкетне, моніторинг.

Вибірковим спостереженням називається такий вид несуцільного спостереження, при якому характеристика всієї сукупності одиниць подається по їх частині, відібраній за відповідними правилами. Прикладом є опитування суб'єктів середнього бізнесу.

Обстеження основного масиву – це реєстрація даних щодо переважної частини одиниць сукупності, які є визначальними для характеристики об'єкта спостереження. Наприклад, обстеження комерційних банків, в яких працює тимчасова адміністрація, обстеження міст з найвищим рівнем забруднення.

Монографічне обстеження – це детальне обстеження окремих типових одиниць сукупності з метою їх досконалого вивчення. Прикладом є обстеження Фонду державного майна.

Анкетне спостереження ґрунтується на принципі добровільного заповнення адресатами надісланих (чи розданих) їм спеціальних анкет. Наприклад, опитування студентів щодо якості викладання курсів.

Моніторинг – спеціально організоване систематичне спостереження за станом певного середовища. Наприклад, моніторинг біржових цін.

Статистичне спостереження здійснюється трьома ***способами***: безпосередній облік фактів, документальний облік, опитування.

Безпосередній облік – це таке спостереження, в процесі якого реєстратори збирають потрібні відомості шляхом самостійних вимірювань, зважувань, особистого огляду тощо одиниць об'єкта і на цій основі роблять записи у формулярі. Наприклад, облік готівкової грошової маси в банках, реєстрація митних потоків, що перетинають митні кордони.

При *документальному обліку* реєстрація фактів відбувається за даними, наведеними у документах первинного обліку (звітності). Наприклад, визначення обсягів експорту та імпорту товарів, розміру інвестицій.

Опитування – отримання інформації на основі реєстрації усних чи письмових (анкетних) відповідей опитуваних на поставлені запитання. Опитування може здійснюватися експедиційним способом, самореєстрацією, кореспондентським або анкетним способами.

Експедиційне опитування – реєстрація фактів спеціально підготованими обліковцями з одночасною перевіркою точності реєстрації. *Самореєстрація* – реєстрація фактів самими респондентами після попереднього інструктажу. *Кореспондентське опитування* – реєстрація фактів на місцях виникнення явищ добровільно обраними особами, які надсилають результати у відповідні інстанції. *Анкетне опитування* – реєстрація думок, намірів і мотивів респондентів шляхом їх самостійного заповнення анкети.

Запитання для самоконтролю

1. Яке визначення статистичного спостереження Ви можете надати?
2. Які вимоги існують до статистичних даних? Надайте пояснення кожної з них.
3. Який найважливіший елемент підготовки статистичного спостереження Ви знаєте?
4. Які питання плану спостереження належать до програмно-методологічних?
5. Що являє собою мета спостереження? Які вимоги до неї Ви можете назвати?
6. Що таке об'єкт статистичного спостереження? Що Ви розумієте під цензом?
7. Як Ви розумієте, що являє собою одиниця сукупності й одиниця спостереження?
8. Що являє собою програма статистичного спостереження?
9. Що відносять до статистичного інструментарію?

10. Які питання плану спостереження належать до організаційних?

11. Що Ви розумієте під часом і періодом спостереження?

12. Хто проводить статистичні спостереження?

13. Що розуміють під місцем спостереження?

14. Які форми статистичного спостереження Ви знаєте?

15. Що таке звітність? На які види її поділяють?

16. Що являє собою спеціально організоване спостереження?

17. Що являє собою статистичний реєстр? Наведіть приклад.

18. Як поділяють статистичні спостереження за повнотою охоплення одиниць досліджуваної сукупності?

19. Що являє собою суцільне спостереження?

20. Які різновиди несуцільних спостережень Ви знаєте? Охарактеризуйте їх.

21. Якими трьома способами здійснюється статистичне спостереження?

22. Що являють собою помилки реєстрації?

23. Унаслідок чого виникають випадкові помилки?

24. Які види систематичних помилок Ви знаєте?

25. Яким спостереженням властиві помилки репрезентативності? Унаслідок чого вони виникають?

26. З якою метою проводиться контроль даних спостереження?

27. Яким чином проводиться логічний контроль даних?

28. Що являє собою арифметичний контроль?

Завдання для самостійної роботи

Тестові завдання

1. Статистичне спостереження здійснюється у такі етапи:
   а) підготовка спостереження,
   б) реєстрація статистичних даних;
   в) формування бази даних;
   г) розрахунок статистичних показників.

Відповіді: 1) а, б; 2) б; в 3) а, б, в; 4) а, б, в, г.

2. До програмно-методологічних питань плану спостереження належать такі:

а) для чого проводиться обстеження;
б) що обстежується;
в) коли проводиться спостереження;
г) джерело інформації.

Відповіді: 1) а, б; 2) б; в 3) а, б, г; 4) а, б, в.

3. Організаційні питання плану спостереження визначають:

а) хто проводить спостереження;
б) джерело інформації;
в) де проводиться спостереження;
г) коли проводиться спостереження.

Відповіді: 1) а, б; 2) б; в 3) а, б, г; 4) а, в, г.

4. Помилки репрезентативності властиві:

а) усім статистичним спостереженням;
б) тільки вибірковому спостереженню.

Вони виникають:

в) унаслідок несуцільного характеру реєстрації;
г) унаслідок неуважності реєстраторів.

Відповіді: 1) а, в; 2) б; в 3) а, г; 4) б, г.

5. Логічний контроль даних спостереження – це:

а) перевірка сумісності даних порівнянням взаємозалежних ознак, яка встановлює лише наявність помилки, а не її величину;
б) перевірка зареєстрованих даних за допомогою прямих або непрямих перерахунків.

Прикладом логічного контролю даних є:

в) порівняння доходів домогосподарства з його витратами;
г) визначення середнього розміру заробітної плати працівників підприємства, якщо відомі кількість працівників і розмір фонду оплати праці.

Відповіді: 1) а, в; 2) б; в 3) а, г; 4) б, г.

Практичні завдання (задачі)

6. Визначте вид статистичного спостереження за формою спостереження, часом реєстрації фактів, повнотою охоплення одиниць сукупності, способом одержання інформації :

- ✓ облік РАЦСами природного руху населення (народжуваності та смертності);
- ✓ обстеження житлових умов працівників банку;
- ✓ перепис навчальних закладів міста;
- ✓ анкетне опитування пасажирів авіакомпанії, яка обслуговує міжнародні рейси;
- ✓ щомісячний звіт про собівартість продукції;
- ✓ опитування молодих сімей регіону з питань планування сім'ї.

7. Складіть перелік запитань, що входитимуть до програм спостереження:

- опитування випускників навчального закладу щодо спрямованості навчального процесу на практичну діяльність;
- опитування жінок репродуктивного віку з питань планування сім'ї;
- опитування працездатного населення щодо стану його здоров'я.

8. Планується проведення статистичного спостереження, мета якого – поліпшення якості обслуговування в супермаркеті. Розробіть план статистичного спостереження.

9. Планується проведення опитування студентів групи щодо якості викладання навчальної дисципліни "статистика". Розробіть анкету, проведіть анкетне опитування у вашій групі. Результати спостереження подайте у вигляді статистичної таблиці.

10. Проведіть логічний контроль правильності заповнення анкети студентом другого курсу денної форми навчання одного з вищих навчальних закладів:

Стать	Чоловіча
Дата народження	15.04.2017 р.
Місце народження	м. Київ
Освіта	Вища
Місце навчання	Університет
Громадянство	США
Сімейний стан	Одружений

11. Проведіть арифметичний контроль наведених даних:

Вид поселення	Чисельність населення, тис. осіб	У тому числі		Кількість чоловіків на 1000 жінок
		чоловіки	жінки	
Міські поселення	560	256	320	800
Сільська місцевість	380	160	220	727
У цілому	980	416	524	794

12. Зазначте, які види помилок (випадкові, систематичні) допущені при проведенні спостережень:

а) під час опитування респондентів їх значна частина заокруглювала свої доходи до цілих чисел (у бік зменшення);

б) під час опитування жінок їх значна частина заокруглювала свій вік до цілих або кратних п’яти чисел;

в) під час опитування жінок репродуктивного віку не було враховано вікову групу 20–24 роки.

Розділ 3.
ЗВЕДЕННЯ І ГРУПУВАННЯ СТАТИСТИЧНИХ ДАНИХ

3.1. Сутність статистичного зведення, класифікації і групування

Статистичне зведення є другим етапом дослідження масових суспільних явищ. Зведення – це комплекс послідовних операцій з узагальнення конкретних поодиноких факторів, які утворюють сукупність, для виявлення типових рис і закономірностей, що характеризують досліджуване явище в цілому. Тобто у результаті зведення здійснюється перехід від даних, які характеризують окремі одиниці об'єкта спостереження, до даних, які характеризують сукупність загалом.

На етапі зведення елементи сукупності за принципом схожості та відмінності певних ознак об'єднуються у групи, для характеристики яких використовують систему показників. Результати зведення подаються у вигляді статистичних таблиць.

Існують загальноприйняті методологічні стандарти розподілу сукупностей на групи – *класифікації* (наприклад, класифікація видів економічної діяльності (КВЕД), українська класифікація товарів зовнішньоекономічної діяльності (УКТЗЕД)). Вони розробляються міжнародними та національними статистичними органами і рекомендуються як статистичний стандарт. Класифікація – це стійке групування за атрибутивною ознакою, яке, як правило, містить дуже докладну номенклатуру груп і підгруп.

Для розв'язання конкретних аналітичних завдань проводять нестандартні групування за певними ознаками, що легко розпізнаються. До *принципів групування* слід віднести такі:

- в одну групу об'єднуються елементи, подібні між собою;
- ступінь подібності елементів однієї групи значно вищий, ніж елементів інших груп.

При проведенні групування вирішуються три питання:

1) що взяти за основу групування?
2) скільки груп необхідно виділити?
3) як розподілити ці групи?

Залежно від виду ознаки, покладеної в основу, розрізняють групування за атрибутивними й кількісними ознаками.

3.2. Принципи формування груп

На практиці використовують три формальні *принципи формування груп*:

1) рівність інтервалів;
2) кратність інтервалів;
3) рівність частот.

Інтервал – це кількісне значення ознаки, яка варіює в певних межах, тобто інтервал окреслює кількісні межі (верхню і нижню) груп.

За величиною інтервали поділяються на рівні та нерівні. Рівні інтервали застосовують тоді, коли зміни кількісної ознаки всередині сукупності відбуваються рівномірно. Нерівні інтервали застосовують, якщо значення групувальної ознаки нерівномірно змінюються від мінімального до максимального.

Розрізняють інтервали відкриті та закриті. *Закриті* інтервали – це такі, в яких максимальні та мінімальні значення відомі (наприклад, 25–35, 35–45). Прикладами *відкритих* інтервалів можуть бути: до 500, 500 і вище, 60 років і старші).

Рівність інтервалів. Цей принцип використовується, як правило, в структурних і аналітичних групуваннях. Ширина рівного інтервалу h залежить від діапазону варіації ознаки x та обґрунтованого числа груп (інтервалів) m:

$$h = \frac{x_{\max} - x_{\min}}{m}.$$

При кількісних групуваннях слід додатково визначати, якураховувати верхні межі групи – "включно" або "виключно".

Приклад. У групу вивчення іноземної мови потрапили 10 осіб віком 15, 21, 19, 25, 30, 18, 18, 27, 18, 20 років. Згрупуйте студентів групи за віком, виділивши три рівні інтервали.

Визначимо ширину рівного інтервалу. Для цього спочатку знайдемо найбільше і найменше значення віку серед студентів групи. Так, наймолодшому студентові 15 років, а найстаршому – 30 років. Тоді

$$h = \frac{x_{\max} - x_{\min}}{m} = \frac{30 - 15}{3} = 5 \text{ (років)}.$$

Отже, матимемо три рівні інтервали: 15–20 або "молодше 20" (20 = 15 + 5); 20–25 (25 = 20 + 5); 25–30 (30 = 25 + 5) або "25 років і старші".

Слід відмітити, що верхня межа попереднього інтервалу і нижня межа наступного інтервалу збігаються. Так, у нашому випадку перший інтервал закінчується значенням віку 20 років, а наступний також починається з 20 років. У цьому випадку віднесення значення 20 років до певного інтервалу визначатиметься формулюванням останнього відкритого інтервалу: "25 років і старші". Так, нижня межа закритого інтервалу вважається "включно", а верхня – "виключно". У нашому випадку студенти віком 20 років належатимуть до другої вікової групи (20–25), а віком 25 років – до третьої вікової групи (25 років і старші). Тобто до першого інтервалу (15–20 років) входять 5 студентів (зі значеннями віку 15, 19, 18, 18, 18 років), до другого (20–25 років) – 2 студенти (віком 20 і 21 рік), до третього (25 років і старші) – 3 студенти (віком 25, 30 і 27 років).

Кратність інтервалів. Якщо діапазон варіації ознаки надто широкий і розподіл значень нерівномірний, використовують нерівні інтервали, зокрема сформовані за принципом кратності, коли ширина кожного наступного інтервалу в k разів більша (менша) за попередній.

Рівність частот (кількостей). Цей принцип використовується нечасто і переважно в аналітичних групуваннях.

Кількість груп пов'язана з обсягом сукупності. Не існує чітко визначених наукових прийомів, які дозволяють вирішувати питання вибору кількості груп за будь-яких умов, кожного разу слід враховувати конкретні обставини. Слід намагатися, щоб виділені групи були достатньо заповнені одиницями спостереження. Наявність незаповнених інтервалів або попадання в них лише окремих одиниць сукупності є результатом того, що інтервали обрані невдало, а їх кількість, можливо, є надмірною. Варто пам'ятати, що оптимальна

кількість інтервалів дозволяє виявити всі суттєві особливості досліджуваної сукупності.

У сукупностях з розподілом, близьким до нормального, кількість груп m за рівних інтервалів орієнтовно можна визначати емпірично або за формулою, запропонованою американським ученим Г. Стерджесом:

$$m = 1 + 3{,}322 \lg N,$$

де N – кількість одиниць сукупності.

Визначене за формулою Стерджеса число інтервалів, яке відповідає обсягу сукупності N, наведено в табл. 3.1.

Таблиця 3.1

Число інтервалів за формулою Стерджеса

Обсяг сукупності, N	15–24	25–44	45–89	90–179	180–359	360–719
Число інтервалів, m	5	6	7	8	9	10

Іноді доводиться перегруповувати дані, передусім для забезпечення порівнянності структур двох сукупностей за однією ознакою. Результат перегрупування називають *вторинним групуванням*, яке являє собою утворення нових груп на основі здійсненого раніше групування статистичних даних. Розрізняють такі способи вторинного групування:

1) укрупнення інтервалів первинного групування;
2) перегруповування за питомою вагою груп у загальному підсумку, тобто на основі закріплення за кожною групою певної частки одиниць сукупності.

Вторинне групування за другим способом є складнішим процесом, ніж за першим.

3.3. Види групувань

Групування як метод пізнання дозволяє в початковому статистичному матеріалі відокремити суттєві риси від несуттєвих, побачити перехід кількісних змін у якісні, у великій кількості випадковостей виявити необхідність у вигляді закономірностей досліджуваного явища.

За кількістю ознак, покладених в основу групування, розрізняють два *види групування* – просте і складне. *Перше* являє собою групування за будь-якою однією ознакою, *друге* – за двома і більше ознаками.

Складне групування може бути комбінаційним або багатовимірним. *Комбінаційне* – це групування, в якому групи, утворені за однією ознакою, потім поділяють на підгрупи за іншою ознакою або ознаками. *Багатовимірними* є такі групування, які проводять за певною множиною ознак одночасно.

На групування у статистичному аналізі покладають певні *функції*:

1) вивчення структури та структурних зрушень (*структурні групування*);
2) визначення типів соціально-економічних явищ, виокремлення однорідних груп і підгруп (*типологічні групування*);
3) виявлення взаємозв'язків між ознаками (*аналітичні групування*).

Структурне групування характеризує склад однорідної сукупності за певними ознаками, обсяги явища та вагомість окремих груп (табл. 3.2).

Таблиця 3.2

Розподіл кандидатів наук за віковими групами

(%)

Вікова група, років	Базисний рік	Поточний рік
До 30	3,1	6,2
31–40	15,8	20,5
41–50	28,4	22,2
51–55	15,9	12,9
56–60	12,6	12,9
61–70	20,8	18,9
71 і старші	3,4	6,4
Разом	100,0	100,0

Типологічне групування – це розподіл якісно неоднорідної сукупності на класи, соціально-економічні типи, однорідні групи (табл. 3.3).

Таблиця 3.3

Середньорічна кількість найманих працівників в Україні за розмірами підприємств

(тис. осіб)

Розмір підприємств	Базисний рік	Поточний рік
Малі	1834,2	1674,2
Середні	4397,5	4269,8
Великі	3138,2	3131,2
Разом	9369,9	9075,2

Завдання, вирішувані типологічними та структурними групуваннями, тісно пов'язані між собою, внаслідок чого ці групування доповнюють одне одного і застосовуються, як правило, комплексно.

За допомогою *аналітичного групування* виявляють наявність та напрям зв'язку між двома ознаками, з яких одна представляє результат (результативна ознака), а інша – фактор, що впливає на результат (факторна ознака), табл. 3.4.

Таблиця 3.4

Залежність вартості квартири на первинному ринку від її загальної площі

Загальна площа квартири, м^2	Кількість квартир, од.	Середня вартість квартири, тис. USD
42–48	7	99
48–54	5	133
54–60	3	161
У цілому	15	123

Характерна особливість аналітичних групувань полягає в тому, що одиниці сукупності групуються за факторною ознакою (у нашому випадку це загальна площа квартири), а розрахунок групових середніх проводять за значеннями результативної ознаки (в нашому прикладі – за середньою вартістю квартири). За зміною цих величин визначають наявність зв'язку та взаємозалежності між ознаками.

Міру впливу факторної ознаки на результативну вивчають за допомогою дисперсійного аналізу. Метод аналітичних групувань розглянуто у параграфі 7.2 навчального посібника.

Різновидом структурних групувань є ряди розподілу. *Рядом розподілу* у статистиці називають упорядкований розподіл одиниць сукупності на групи за будь-якою варіюючою ознакою, різновиди якої розташовані в певному порядку. Ряд розподілу складається з двох елементів: *варіант* (окремі значення групувальної ознаки) та *частот* (показують, як часто повторюються значення варіант). Частоти можуть виражатись як в абсолютних, так і у відносних величинах, у вигляді часток або відсотків до підсумку. Залежно від статистичної природи групувальної ознаки розрізняють ряди розподілу *атрибутивні* (в їх основу покладено якісні ознаки) і *варіаційні* (в їх основі – ознаки, що мають кількісне вираження).

3.4. Статистичні таблиці

Результати статистичного зведення та групування подаються у статистичних таблицях або графіках.

Таблиці є найбільш раціональною формою подання оброблених статистичних даних. *Статистичними таблицями* називають форму раціонального, наочного, систематизованого викладу й аналізу цифрових характеристик суспільних явищ і процесів.

У статистичній практиці використовують таблиці різного ступеня складності залежно від мети дослідження, особливостей об'єкта дослідження, обсягу наявної інформації.

Таблицю за своїм логічним змістом розглядають як *статистичне речення*, в якому є підмет і присудок. *Підмет таблиці* характеризує об'єкт дослідження, а *присудок* – це система показників, що відображає підмет як об'єкт. Основний зміст таблиці зазначають у її назві.

Залежно від побудови статистичного підмета таблиці поділяються на три види: *прості*, *групові* та *комбінаційні*. Простими називають таблиці, підмет яких являє собою перелік елементів сукупності, територіальний або хронологічний ряд. У групових таблицях підметом є групування за однією ознакою, в комбінаційній – за двома і більше ознаками.

Складання статистичної таблиці здійснюється у два етапи. На першому проєктується макет таблиці, на другому таблицю заповнюють статистичними даними.

Макет таблиці – це комбінація горизонтальних рядків та вертикальних граф (стовпчиків), на перетині яких утворюються клітини для запису відповідної інформації. Під час розроблення і заповнення макета таблиці слід дотримуватися певних правил:

- таблиця повинна мати порядковий номер; у назві таблиці вказується об'єкт, його часова і географічна ознаки;
- назва таблиці, заголовки рядків та граф таблиці повинні бути чіткими й лаконічними, без скорочень, зайвої та другорядної інформації;
- якщо назви окремих граф або рядків повторюються, мають однакові терміни або однаковий зміст, то їх доцільно об'єднувати загальним заголовком;
- у верхніх або бічних заголовках слід вказати одиниці вимірювання з використанням загальноприйнятих скорочень (т, кг, грн, м тощо); загальну для всієї таблиці одиницю вимірювання вказують над таблицею у дужках;
- у випадку побудови великих таблиць, які виникає необхідність продовжити на наступну сторінку, рядки і графи таблиці доцільно нумерувати: графу підмета – літерою алфавіту, інші графи – цифрами;
- інформація, що міститься в рядках (графах) таблиці, узагальнюється підсумковим рядком (графою) "Разом" або "У цілому"; якщо підсумковий рядок розташовано першим, то деталізацію подають як, наприклад, "у тому числі";
- значення показників у клітинах таблиці слід округлювати в межах одного рядка або графи з обов'язковим однаковим ступенем точності; якщо цифри надто малі, записують 0,0 або 0,00 відповідно до розрядності даних у рядках або графах даної таблиці;
- відсутність інформації позначають трьома крапками (…) або (н/д), якщо ж відсутнє саме явище, використовують тире (–); знак "х" свідчить про те, що клітина не підлягає заповненню.

Приклад. Побудуймо макет таблиці, яка б відображала зміну перевезення вантажів (млн т) в Україні протягом трьох останніх років за видами транспорту (залізничним, автомобільним, морським, річковим).

Оскільки в назві таблиці зазначають її основний зміст, об'єкт, його часову і географічну ознаки, то назва нашої таблиці буде такою: "Динаміка перевезення вантажів за видами транспорту в Україні за 2018–2020 рр.". Над таблицею вкажемо загальну одиницю вимірювання – млн т. Отже, макет матиме такий вигляд:

Динаміка перевезення вантажів за видами транспорту в Україні за 2018–2020 рр. *(макет таблиці)*

(млн т)

Вид транспорту	Рік		
	2018	2019	2020
А	1	2	3
Залізничний			
Автомобільний			
Морський			
Річковий			
Разом			

Назву підмета позначимо літерою алфавіту (А), інші графи пронумеруємо цифрами (1, 2 і 3). Інформацію в останньому рядку таблиці узагальнимо (Разом).

Розв'язання типових задач

Задача 1

Навесні у новобудовах центру міста пропонувалися 13 однокімнатних квартир такої вартості.

№	Загальна площа квартири, м²	Вартість квартири, тис. USD	№	Загальна площа квартири, м²	Вартість квартири, тис. USD
1	42	79,5	8	47	140,0
2	43	79,9	9	52	140,0
3	43	85,5	10	60	150,0
4	45	94,0	11	55	154,0
5	51	98,0	12	52	159,0
6	45	117,0	13	59	178,5
7	50	135,0			

Завдання:

1. Побудуйте ряд розподілу квартир за загальною площею, утворивши три групи з рівними інтервалами. Проаналізуйте структуру квартир, які пропонуються на первинному ринку житла у центрі міста, за загальною площею.

2. Для вивчення зв'язку між площею квартир та їх вартістю побудуйте аналітичне групування. Зробіть висновки щодо існування зв'язку та його напряму. Результати групувань подайте у вигляді таблиці.

Розв'язання задачі

1. Обидві ознаки є кількісними, оскільки вимірюються числами.

Визначимо ширину рівного інтервалу (h). Найбільше значення загальної площі квартири (x_{max}) становить 60 м², найменше (x_{min}) – 42 м², кількість інтервалів (m) за умовою задачі – 3.

Тоді ширина інтервалу $h = \frac{x_{max} - x_{min}}{m} = \frac{60 - 42}{3} = 6 м^2.$

Отже, сформовані такі три інтервали (групи):

1) 42–48 (48 = 42 + 6);

2) 48–54 (54 = 48 + 6);

3) 54–60 (60 = 54 + 6).

Наступним кроком є підрахунок кількості квартир, які за своєю площею потрапили до кожної вищезазначеної групи (групування здійснюється за принципом "нижня межа кожної групи "включно"). Наприклад, квартира № 1 площею 42 м^2 увійде до першої групи (42–48), а квартира № 7 площею 50 м^2 – до другої групи (48–54). Результати проведеного групування занесемо до табл. 3.5.

Таблиця 3.5

Розподіл квартир за розміром загальної площі

Група квартир за розміром площі, м^2	Кількість квартир, од.	Частка у підсумку, %
42–48	6	46,1
48–54	4	30,8
54–60	3	23,1
Разом	13	100,0

Отже, майже половина (46,1%) однокімнатних квартир, які були запропоновані на ринку новобудов міста, мали загальну площу 42–48 м^2.

2. У наведеному прикладі факторною ознакою є загальна площа квартир, яка впливає на вартість квартир (результативну ознаку).

З метою виявлення зв'язку між цими ознаками проведемо аналітичне групування. В його основу покладемо результати групування, проведеного в завданні 1. Побудуємо робочу таблицю (табл. 3.6).

Таблиця 3.6

Розподіл квартир за розміром загальної площі

№ групи	Група квартир за розміром загальної площі, м^2	Порядковий номер квартири	Загальна площа квартири, м^2	Вартість квартири, тис. USD
1	2	3	4	5
1	42–48	1, 2, 3, 4, 6, 8	42, 43, 43, 45, 45, 47	79,5; 79,9; 85,5; 94,0; 117,0; 140,0
	Разом	6	265	595,9

Продовження табл. 3.6

1	2	3	4	5
2	48–54	5, 7, 9, 12	51, 50, 52, 52	98,0; 135,0; 140,0; 159,0
	Разом	4	205	532,0
3	54–60	10, 11, 13	60, 55, 59	150,0; 154,0; 178,5
	Разом	3	174	482,5
Разом по таблиці		13	644	1610,4

За даними робочої таблиці визначимо середні значення ознак за кожною групою і в цілому по всій сукупності квартир:

№ групи	Середня загальна площа квартири, м^2	Середня вартість квартири, тис. USD
1	265 : 6 = 44,2	595,9 : 6 = 99,3
2	205 : 4 = 51,3	532,0 : 4 = 133,0
3	174 : 3 = 58,0	482,5 : 3 = 160,8
У цілому	644 : 13 = 49,5	1610,4 : 13 = 123,9

Використовуючи одержані значення, побудуємо табл. 3.7.

Таблиця 3.7

Залежність вартості квартир від їх загальної площі

Група квартир за розміром їх загальної площі, м^2	Кількість квартир, од.	Середній розмір загальної площі квартир, м^2	Вартість квартир, тис. USD	
			усього	однієї квартири
42–48	6	44,2	595,9	99,3
48–54	4	51,3	532,0	133,0
54–60	3	58,0	482,5	160,8
У цілому	13	49,5	1610,4	123,9

Отже, зі зростанням загальної площі квартири збільшується і її вартість, тобто між цими ознаками існує прямий зв'язок. Середня вартість однокімнатної квартири у новобудовах центру міста становила 123,9 тис. USD.

Задача 2

Перегрупуйте наведені в таблиці дані про розподіл будівельно-монтажних організацій за обсягом виконаних робіт, утворивши такі три групи: до 25 млн грн, 25–100, 100 млн грн. і більше. Результати вторинного групування подайте у формі статистичної таблиці.

Обсяг будівельно-монтажних робіт, млн грн	Частка у підсумку, %	
	кількість організацій	обсяг робіт
До 10	8	2
10–20	10	7
20–40	24	12
40–80	32	18
80–150	21	28
150 і більше	5	33
Разом	100	100

Розв'язання задачі

До першого інтервалу вторинного групування (до 25 млн грн) точно будуть входити: 8 + 10 = 18% організацій і 2 + 7 = 9% обсягу робіт первинного групування. Однак до цього інтервалу також має увійти певна частка організацій та обсягу робіт інтервалу 20–40 млн грн. первинного групування. Визначимо, яка саме:

$$25 - 20 = 5;\ \frac{5}{20} = \frac{1}{4},\ \text{отже},\ 24 \times \frac{1}{4} = 6,\ 12 \times \frac{1}{4} = 3.$$

Тобто до першого інтервалу вторинного групування (до 25 млн грн) увійдуть: 8 + 10 + 6 = 24% організацій і 2 + 7 + 3 = 12% обсягу робіт. Техніку перегрупування подамо в табл. 3.8.

Таблиця 3.8

Розподіл будівельно-монтажних організацій за обсягом виконаних робіт

Обсяг будівельно-монтажних робіт, млн грн	Частка у підсумку, %	
	кількість організацій	обсяг робіт
1	2	3
До 25	$8 + 10 + \frac{1}{4} \times 24 = 24$	$2 + 7 + \frac{1}{4} \times 12 = 12$

Продовження табл. 3.8

1	2	3
25–100	$\frac{3}{4}\times 24+32+21\times\frac{2}{7}=56$	$\frac{3}{4}\times 12+18+28\times\frac{2}{7}=35$
100 і більше	$\frac{5}{7}\times 21+5=20$	$\frac{5}{7}\times 28+33=53$
Разом	100	100

Запитання для самоконтролю

1. Яким чином проводиться зведення? Дайте його визначення.
2. Що являють собою класифікації? Наведіть їх приклади.
3. Які принципи групування Ви знаєте? Які питання вирішуються при його проведенні?
4. Які принципи формування груп Ви можете назвати?
5. Як визначають ширину рівного інтервалу?
6. Які види групувань Ви знаєте?
7. Що характеризує структурне групування?
8. Що являє собою типологічне групування?
9. З якою метою проводиться аналітичне групування? Яким чином?
10. Які визначення ряду розподілу Ви можете надати? Назвіть його елементи.
11. Для чого використовуються статистичні таблиці і графіки?
12. Що являють собою статистичні таблиці?
13. Що характеризують підмет і присудок статистичної таблиці?
14. Які види статистичних таблиць Ви знаєте?
15. Що являє собою макет статистичної таблиці?
16. Які правила, яких слід дотримуватися при розробці та заповненні макета таблиці, Ви можете назвати?

Завдання для самостійної роботи

Тестові завдання

1. Загальноприйняті методологічні стандарти розподілу сукупностей на групи – це:
   а) групування; б) зведення; в) класифікації.

Відповіді: 1) а; 2) б; 3) в.

2. Зазначте атрибутивні групувальні ознаки:
   а) стать людини; б) бал успішності; в) національність;
   г) вид економічної діяльності.

Відповіді: 1) а; 2) б; 3) а, в; 4) а, в, г.

3. Ширина рівного інтервалу h залежить:
   а) тільки від діапазону варіації ознаки x;
   б) тільки від обґрунтованого числа груп (інтервалів) m;
   в) від діапазону варіації ознаки x та від обґрунтованого числа груп (інтервалів) m.

Відповіді: 1) а; 2) б; 3) в.

4. Зазначте вид наведеного групування:
   а) типологічне; б) структурне; в) аналітичне.

Залежність якості роботи верстатів від їх технічного стану

Технічний стан верстата	Кількість верстатів	Середнє число обривів нитки на 100 м^2 тканини
Пройшов ремонт	8	68
Не пройшов ремонту	12	73
У цілому	20	71

Відповіді: 1) а; 2) б; 3) в.

5. До основних правил розробки і заповнення макета таблиці слід віднести такі:
   а) у назві таблиці, заголовках рядків і граф таблиці можливі скорочення;
   б) таблиця повинна мати порядковий номер; у назві таблиці вказується об’єкт, його часова і географічна ознаки;

в) значення показників у клітинах таблиці не слід округлювати в межах одного і того ж рядка або графи з обов'язковим однаковим ступенем точності;

г) відсутність інформації позначають "х".

Відповіді: 1) а; 2) б; 3) а, в; 4) б, г.

Практичні завдання (задачі)

6. Відомо, що загальна чисельність співробітників підприємства складає 30 осіб, мінімальна і максимальна заробітна плата по підприємству відповідно становить 4600 і 10600 грн. Визначте:
   а) число інтервалів групування співробітників за рівнем заробітної плати, користуючись формулою Стерджеса;
   б) величину інтервалу групування (за результатами завдання а)).

7. За даними про тип підприємств утворіть атрибутивний ряд розподілу (позначено: "м" – мале підприємство; "с" – середнє підприємство; "в" – велике підприємство). Результати подайте у вигляді таблиці. Проаналізуйте структуру підприємств.

м	м	с	в	м	с	в	с	с	м
с	м	м	с	м	в	м	с	м	м
с	в	с	с	м	м	м	с	м	м
м	м	м	с	м	м	с	с	с	м
в	с	с	м	м	с	в	м	м	м

8. За даними про денну виручку від реалізації торговельних підприємств (тис. грн) утворіть інтервальний ряд розподілу з такими нерівними інтервалами: 10–15, 15–25, 25–40.

12	11	31	13	23	22	12	21	23	25
23	25	36	29	40	19	23	30	19	16
34	24	28	32	35	20	34	14	16	27
15	19	14	22	34	31	38	18	22	31
18	32	17	21	29	14	17	23	24	19

9. У зимову сесію 20 студентів групи отримали такі бали успішності зі статистики за 5-бальною системою оцінювання знань:

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

Побудуйте ряд розподілу студентів за одержаними оцінками. Виділивши дві групи студентів – неуспішних (одержали 2 бали) і успішних (3 бали і вище) – проаналізуйте їх структуру. Вкажіть, якими є отримані ряди розподілу – варіаційними чи атрибутивними.

10. Розмір посівних площ картоплі, овочевих та баштанних продовольчих культур за регіонами України у минулому році становив (тис. га):

37	68	70	87	102
127	49	58	73	82
79	66	115	75	93
84	70	43	67	43
98	125	74	113	

Згрупуйте регіони України за розміром посівних площ, утворивши три групи з рівними інтервалами. Результати групування подайте у вигляді статистичної таблиці. Проаналізуйте структуру регіонів за посівними площами картоплі, овочевих та баштанних продовольчих культур.

11. Восени в одному з районів міста було 37 пропозицій для оренди однокімнатних квартир з такою щомісячною платнею:

№	Загальна площа квартири, м^2	Щомісячна орендна плата, USD	№	Загальна площа квартири, м^2	Щомісячна орендна плата, USD
1	2	3	4	5	6
1	35	350	20	36	225
2	37	375	21	34	250
3	35	375	22	30	250

Продовження табл.

1	2	3	4	5	6
1	35	350	20	36	225
2	37	375	21	34	250
3	35	375	22	30	250
4	32	375	23	30	280
5	36	400	24	30	295
6	33	400	25	32	300
7	32	400	26	30	300
8	35	415	27	33	300
9	42	450	28	40	325
10	34	450	29	40	325
11	42	450	30	35	325
12	34	500	31	30	325
13	30	500	32	34	325
14	35	500	33	35	350
15	35	500	34	35	350
16	32	500	35	37	350
17	30	600	36	34	350
18	35	700	37	30	350
19	32	225			

Завдання:

1) згрупуйте квартири за розміром загальної площі, утворивши три групи з рівними інтервалами. Результати групування подайте у вигляді статистичної таблиці. Проаналізуйте структуру квартир за розміром загальної площі;
2) побудуйте аналітичне групування залежності щомісячної орендної плати від загальної площі квартири. Зробіть висновки щодо напряму зв'язку.

12. Розподіл підприємств трьох регіонів за вартістю основного капіталу характеризується такими даними (% до підсумку):

Група підприємств за вартістю основного капіталу, млн грн	Регіони		
	A	B	C
100–300	14	13	1
300–600	27	30	15
600–1000	25	26	30
1000–3000	30	27	52
3000 і більше	4	4	2
Разом	100	100	100

Перегрупуйте наведені в таблиці дані, утворивши такі три групи підприємств за вартістю основного капіталу (млн грн): 100–400, 400–1000, 1000 і більше.

13. Побудуйте макети статистичних таблиць, які характеризуватимуть:
    1) зміну чисельності зайнятих за видами економічної діяльності за останні два роки;
    2) захворюваність населення України за класами хвороб за три останніх роки;
    3) зміну оплати праці найманих працівників у двох видах економічної діяльності (освіті та будівництві) за регіонами України за два останніх роки;
    4) структуру і динаміку перевезення вантажів (млн т) усіма видами транспорту (наземним, водним, авіаційним) за два останніх роки.

Розділ 4.
СТАТИСТИЧНІ ПОКАЗНИКИ

4.1. Сутність статистичних показників

Статистичне дослідження незалежно від його масштабів та мети завжди закінчується визначенням і аналізом різних статистичних показників.

Статистичний показник – це узагальнююча характеристика соціально-економічного явища чи процесу, в якій поєднуються зміст та кількісна визначеність останнього. *Зміст* показника знаходить своє відображення у назві (наприклад, народжуваність, урожайність, прибутковість тощо). *Кількісний бік* явища – це число та його вимірник (наприклад, 15‰, 38 ц/га, 85% тощо).

Сполучною ланкою між змістом і числовим вираженням є правило побудови, або *модель показника*, яка розкриває його статистичну структуру та встановлює, *що, де, коли* і *яким чином* підлягає вимірюванню. *Модель показника* забезпечує адекватність відображення явища (процесу) і точність його вимірювання, тобто обидва аспекти вірогідності статистичної інформації.

Статистичні показники виконують такі *функції*: пізнавальну, управлінську, контрольну та стимулюючу. *Пізнавальна* функція статистичних показників полягає в тому, що економіст аналізує, пізнає якісний бік явища, яке вивчається, заглиблюється в його суть. *Управлінська* функція статистичних показників є важливою складовою процесу управління на всіх його рівнях. З розвитком ринкових відносин зростає значення *контрольної* функції статистичних показників, зокрема щодо виконання договірних умов. Аналіз результатів роботи уможливлює реалізацію *стимулюючої* функції статистичних показників, що дозволяє знайти невикористані резерви, окреслити шляхи подолання недоліків, прийняти правильне рішення в умовах ринкової системи господарювання.

За способом обчислення розрізняють *первинні* (подаються у формі абсолютних величин) та *похідні* (обчислюються на основі первинних, мають форму середніх або відносних величин) показники.

За ознакою часу показники поділяються на *моментні* (надають кількісну характеристику явищ на певний момент часу) і *інтервальні* (характеризують явище за певний час).

Окрему групу становлять *взаємообернені показники* – пара характеристик одного й того самого явища. Якщо прямий показник змінюється в напрямі зміни явища, то обернений – у протилежному. Наприклад, якщо прямі показники – продуктивність праці, споживчі ціни, то обернені, відповідно, – трудомісткість праці, купівельна спроможність грошової одиниці.

Статистичні показники можуть виражатись у формі абсолютних, відносних і середніх величин.

4.2. Абсолютні та відносні величини

Абсолютні статистичні величини характеризують розміри соціально-економічних явищ – обсяги сукупності або обсяги значень певних ознак. Абсолютні статистичні величини завжди є іменованими числами, кожна з них має свою одиницю вимірювання. Іменовані числа містять вимірники ознак. Розрізняють такі одиниці вимірювання: *натуральні* (наприклад, метри, години, штуки, тонни), *вартісні* (наприклад, гривні, долари, євро) і *трудові* (наприклад, людино-години, людино-дні).

У свою чергу, натуральні одиниці вимірювання можуть бути *простими* та *комбінованими (складними)*. Останні являють собою комбінування декількох різнойменних одиниць вимірювання і застосовуються для всебічного визначення розмірів деяких явищ. Наприклад, виробництво і споживання електроенергії вимірюють у кіловат-годинах.

Якщо виникає потреба звести воєдино кілька різновидів однієї споживчої властивості, обсяги такого явища виражають в *умовно-натуральних* одиницях (умовне паливо, умовна банка).

Абсолютні статистичні величини поділяють на індивідуальні та сумарні (загальні, підсумкові). *Індивідуальні* абсолютні величи-

ни характеризують величину ознаки окремих одиниць сукупності. *Сумарні (загальні, підсумкові)* характеризують величину ознаки певної сукупності або окремих її груп.

Поряд з абсолютними показниками велику роль у статистиці відіграють відносні величини.

Відносні величини є результатом ділення абсолютних величин і характеризують співвідношення різнойменних та однойменних показників.

Відносна величина є результатом співвідношення величини, яку порівнюють (порівнюваної), з величиною, з якою порівнюють (базою порівняння):

$$\textit{Відносна величина} = \frac{\textit{порівнювана величина}}{\textit{база порівняння}}.$$

Відносна величина показує, у скільки разів чисельник більший за знаменник або яку частку чисельник становить від знаменника, або скільки одиниць чисельника припадає на 100, 1000, 10000 і т. д. одиниць знаменника.

4.3. Види відносних величин

Відносні величини можна поділити на *дві групи*:

1) відносні величини як співвідношення однойменних величин;
2) відносні величини як результат зіставлення різнойменних величин.

До *першої групи* відносять такі відносні величини: динаміки, структури, координації, просторового порівняння, планового завдання, виконання плану (договірних зобов'язань, державних замовлень).

До *другої групи* належать відносні величини інтенсивності.

Відносні величини динаміки (ВВД) характеризують напрям та інтенсивність зміни явища у часі й визначаються співвідношенням значень показника за два періоди чи моменти часу.

Приклад. Якщо кількість підприємств, які здійснювали діяльність у сфері послуг у місті, у 2020 році становила 4458, а у 2019 році – 3914, то:

$$ВВД = \frac{y_t}{y_{t-1}} = \frac{y_{2020}}{y_{2019}} = \frac{4458}{3914} = 1{,}139, \text{ або } 113{,}9\%.$$

Отже, кількість підприємств, які здійснювали діяльність у сфері послуг у місті, у 2020 році порівняно з 2019 роком збільшилася на 13,9%.

Відносні величини планового завдання (прогнозування) (ВВПЗ) характеризують очікуваний розмір збільшення або зменшення розмірів явища за планом (прогнозом) у наступному періоді порівняно з базисним (одним із періодів, що прийнято за базу порівняння).

Відносні величини виконання плану (договірних зобов'язань, державних замовлень) (ВВВП) характеризують рівень виконання прогнозних розрахунків і обчислюються як співвідношення фактично досягнутого рівня у звітному періоді до рівня, передбаченого планом (договором або державним замовленням).

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

$$ВВД = ВВПЗ \cdot ВВВП.$$

Відносні величини структури (*ВВС*) характеризують склад, структуру сукупності за тією чи іншою ознакою й обчислюються як відношення розміру складової частини до загального підсумку. Відносні величини структури є адитивними. Сума відносних величин структури по всій сукупності становить 1 або 100%. Відносні величини структури можна порівняти у часі в межах порівнянних складових. Так оцінюють структурні зрушення, тобто зміни у складі сукупності за певний період часу. Різницю між відповідними частками сукупностей вимірюють процентними (або відсотковими) пунктами (п. п. або в. п.).

Приклад. Маємо такі дані про чисельність наявного населення України на початок 2015 і 2020 рр. (табл.4.1).

Таблиця 4.1

Динаміка чисельності населення України

(на 1 січня, тис. осіб)

Роки	Усе населення	у тому числі	
		міське	сільське
2015	42 929,3	29 673,1	13 256,2
2020	41 902,4	29 139,3	12 763,1

У цьому випадку можна визначити такі відносні величини структури: частку міського населення та частку сільського населення в усьому населенні України. Так, частка міського населення на початок 2015 року становить:

$$ВВС^{м}_{2015} = \frac{y_м}{\sum y} = \frac{29\ 673,1}{42\ 929,3} = 0,691, \text{ або } 69,1\% .$$

Частка сільського населення на початок 2015 року, відповідно,

$$ВВС^{с}_{2015} = \frac{y_с}{\sum y} = \frac{13\ 256,2}{42\ 929,3} = 1 - 0,691 = 0,309, \text{ або } 30,9\%.$$

Ці самі відносні величини на початок 2020 року дорівнюють:

$$ВВС^{м}_{2020} = \frac{y_м}{\sum y} = \frac{29\ 139,3}{41\ 902,4} = 0,695, \text{ або } 69,5\%;$$

$$ВВС^{с}_{2020} = \frac{y_с}{\sum y} = \frac{12\ 763,1}{41\ 902,4} = 1 - 0,695 = 0,305, \text{ або } 30,5\%$$

Отже, у складі населення України на початок 2020 року порівняно з початком 2015 року відбулися певні структурні зрушення: частка міського населення за цей період зросла на 69,5 – 69,1 = 0,4 п. п. У свою чергу, частка сільського населення скоротилася на таку ж величину.

Відносні величини координації (*ВВК*) показують, скільки одиниць однієї частини сукупності припадає на одиницю (або на 100 одиниць) іншої, прийнятої за базу порівняння.

Приклад. Використовуючи дані попередньої таблиці, визначимо співвідношення міського і сільського населення України. Так,

$$ВВК_{2015} = \frac{y_м}{y_с} = \frac{29\ 673,1}{13\ 256,2} = 2,24;$$

$$ВВК_{2020} = \frac{y_м}{y_с} = \frac{29\ 139,3}{12\ 763,1} = 2,28$$

Отже, якщо на початок 2015 року на 100 осіб сільського населення припадало 224 особи міського населення, то на початок 2020 року на 100 осіб сільського населення припадало вже 228 осіб міського населення.

Відносні величини порівняння (*ВВПр*) – це співвідношення однойменних величин, які характеризують різні об'єкти (підприємства, види економічної діяльності), території (міста, регіони, країни), але в один і той же самий період чи момент часу. Важливим є те, щоб методика розрахунку зіставних показників була однаковою.

Відносні величини порівняння мають широке застосування в міжнародній статистичній практиці. Для їх обчислення застосовують як абсолютні, так і відносні величини.

Приклади. 1. Якщо площа Данії складає 43 тис. км^2, а площа Угорщини – 93 тис. км^2, то:

$$ВВПр = \frac{y_У}{y_Д} = \frac{93}{43} = 2{,}2 .$$

Отже, площа Угорщини більша за площу Данії у 2,2 раза.

2. Якщо щільність населення Латвії – 35 осіб/км^2, а Фінляндії – 17 осіб/км^2, то:

$$ВВПр = \frac{y_Л}{y_Ф} = \frac{35}{17} = 2{,}1 .$$

Щільність населення в Латвії у 2,1 раза перевищує щільність населення в Фінляндії, отже, Латвія є більш густонаселеною країною.

Базою порівняння може бути певне еталонне значення показника (стандарт, норматив тощо). У такому випадку визначають *відносну величину порівняння зі стандартом (еталоном)*. Її відхилення від 1 (або 100%) свідчить про порушення оптимального процесу.

Відносні величини інтенсивності (*ВВІ*) характеризують ступінь поширення явища у певному середовищі. ВВІ завжди є відношенням двох різнойменних абсолютних величин, пов'язаних між собою, але які не є додатками або частинами цілого. На відміну від інших видів, відносна величина інтенсивності завжди є іменованою, в ній поєднуються одиниці вимірювання чисельника і знаменника. Прикладами ВВІ є показники народжуваності, смертності населення на 1000 осіб, валового внутрішнього продукту на одну особу тощо.

Приклад. Якщо чисельність населення Кіпру 800,0 тис. осіб, а його площа – 9250 км^2, то щільність населення складає:

$$ВВІ = \frac{800{,}0}{9{,}250} = 86{,}5 \ осіб / км^2.$$

Важливою умовою правильного використання статистичних показників є вивчення явищ і процесів через застосування абсолютних та відносних величин у їх поєднанні.

4.4. Середні величини, їх види та форми

Серед узагальнюючих показників, якими статистика характеризує суспільні явища та властиві їм закономірності, важлива роль належить середнім величинам. Вони відображають характерну, типову величину ознаки одиниць статистичної сукупності, яка утворюється в конкретних умовах місця і часу під впливом усіх факторів.

Середня величина ($\overline{X}$) – це узагальнююча міра варіюючої ознаки, яка характеризує її рівень у розрахунку на одиницю сукупності. У середній нівелюються (розчиняються) відмінності й особливості індивідуальних значень ознак, вона є рівнодіючою значень цих ознак.

До *основних вимог* наукового застосування середніх величин належать:

- ✓ однорідність сукупності;
- ✓ визначення середньої на підставі масових даних;
- ✓ обґрунтований вибір форми середньої величини.

У статистичній практиці використовують такі *види середніх величин*: середня арифметична, середня гармонічна, середня геометрична, середня квадратична тощо. Кожна з цих середніх може набувати двох форм: простої і зваженої.

Обираючи вид середньої величини, слід дотримуватися певних правил.

1. Вид середньої величини визначає логічна формула показника, який осереднюється. Зрозуміло, що кожний показник має свій, притаманний лише йому економічний зміст. Наприклад,

$$\frac{\textit{Заробітна плата}}{\textit{одного працівника}} = \frac{\textit{фонд заробітної плати усіх працівників}}{\textit{кількість працівників}}.$$

2. Середня величина буде обчислена правильно у тому випадку, коли під час заміни нею усіх варіантів усередненого показника без зміни залишиться загальний обсяг усередненого показника, так званий *визначальний показник*. Так, у наведеному вище прикладі таким

визначальним показником є загальні витрати на харчування, а саме, чисельник формули.

3. Вибір виду середньої величини залежить від характеру взаємозв'язку індивідуальних значень усередненої ознаки з їх визначальним показником. Так, якщо визначальний показник є сумою індивідуальних значень усередненої ознаки, застосовують середню арифметичну. Якщо ж він є сумою обернених індивідуальних значень усередненої ознаки, то використовують середню гармонічну.

Якщо середня величина визначається за первинними (незгрупованими) даними, застосовують *просту (незважену) форму*, якщо за вторинними (згрупованими) – то *зважену*.

Середня арифметична використовується для осереднення прямих значень ознак шляхом їх підсумовування. Логічна формула середньої арифметичної має такий вигляд:

$$\overline{X} = \frac{\textit{загальний обсяг значень ознаки}}{\textit{обсяг сукупності}}.$$

Якщо дані незгруповані, то використовують *середню арифметичну просту*:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n},$$

де X_i – окремі значення ознаки; n – обсяг сукупності.

Якщо дані згруповані, то застосовують ***середню арифметичну зважену***:

$$\overline{X} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_n f_n}{f_1 + f_2 + f_3 + \ldots + f_n} = \frac{\sum_{j=1}^{m} x_j f_j}{\sum_{j=1}^{m} f_j},$$

де f_j – це частоти.

Частоти (або ваги) f_j – це числа, що показують, скільки разів зустрічаються індивідуальні значення ознаки або варіанти x_j (тобто їх кількість).

Середня арифметична зважена обчислюється шляхом ділення суми зважених варіант на суму частот. Середню арифметичну зважену застосовують тоді, коли варіанти, які складають досліджувану сукупність, зустрічаються неоднакову кількість разів чи мають різну вагу.

Вагами можуть бути також і частки (відносні величини структури d_j):

$$\overline{X} = \sum_1^m x_j \, d_j$$

де d_j – частка j-ї групи.

Величина d_j розраховується за формулою

$$d_j = \frac{f_j}{\sum_1^m f_j}, \; \sum_1^m d_j = 1.$$

Вагою може бути також абсолютна величина, логічно пов'язана з осереднюваним показником. При цьому вибір ваг ґрунтується на *логічній формулі показника*. Оскільки середня величина визначається у розрахунку на одиницю сукупності, то *вага завжди буде розташована у знаменнику логічної формули*.

Приклад. При визначенні середніх витрат на харчування на одного члена родини спочатку побудуємо логічну формулу цього показника. Для того щоб визначити витрати на одного члена родини, потрібно знати, по-перше, загальні витрати всіх членів родини на харчування, а по-друге – кількість членів родини:

$$\frac{\textit{Витрати на харчування}}{\textit{на одного члена родини}} = \frac{\textit{загальні витрати на харчування}}{\textit{кількість членів родини}}$$

Отже, в нашому випадку вагою буде кількість членів родини – знаменник логічної формули показника, який потрібно осереднити.

Середня арифметична величина має деякі *математичні властивості*, використання яких спрощує її обчислення:

- алгебраїчна сума відхилень усіх варіант від середньої величини дорівнює нулю:

$$\sum (x_j - \overline{x}) = 0;$$

- сума квадратів таких відхилень наближається до мінімуму:

$$\sum (x_j - \overline{x})^2 \Rightarrow \min;$$

- якщо кожну варіанту збільшити або зменшити на будь-яку постійну величину, то й середня зміниться відповідно:

$$\frac{\sum (x_j \pm A) \cdot f}{\sum f} = \overline{x} \pm A;$$

- якщо кожну варіанту збільшити (зменшити) у k разів, то і середня величина зміниться відповідно:

$$\frac{\sum (x_j \cdot k) \cdot f}{\sum f} = \overline{x} \cdot k;$$

- якщо частоти всіх варіант збільшити (зменшити) у k разів, то середня при цьому не зміниться:

$$\frac{\sum x_j \cdot (f \cdot k)}{\sum (f \cdot k)} = \overline{x}.$$

Для того, щоб визначити *середню величину в динамічному ряду* (коли відомі дані за декілька років, місяців тощо), використовують різні способи залежно від наявних даних:

1) якщо відомі дані на початок ($X_{поч}$) і на кінець ($X_{кін}$) періоду, тобто моменті показники, то середню величину визначають так:

$$\overline{X} = \frac{X_{поч} + X_{кін}}{2};$$

2) якщо відомі дані на початок кожного періоду, наприклад на початок кожного місяця або 01.01 кожного року, при цьому моментів більше ніж два й інтервали між ними рівні, то використовують ***середню хронологічну***:

$$\overline{X} = \frac{\frac{X_1 + X_n}{2} + X_2 + X_3 + ... + X_{n-1}}{n-1},$$

де n – кількість моментів.

3) якщо відомі дані за ряд місяців, років, то використовують звичайну середню арифметичну просту.

Середню гармонічну використовують у тих випадках, коли відомі на варіанти, а обернені до них числа.

Для незгрупованих даних це ***середня гармонічна проста***:

$$\overline{X} = \frac{n}{\sum \frac{1}{x}},$$

де $\sum \frac{1}{x}$ – це сума обернених значень варіант; n – число варіант.

Якщо дані згруповані, то використовують ***середню гармонічну зважену***:

$$\overline{X} = \frac{\sum_{1}^{m} z_j}{\sum_{1}^{m} \frac{z_j}{x_j}},$$

де z_j – обсяг значень ознаки, тобто .

Середню гармонічну доцільно використовувати в тих випадках, коли немає інформації про знаменник логічної формули показника, тобто *відсутні ваги*. Так, якщо потрібно визначити середню урожайність зернових за різними категоріями господарств, то спочатку побудуємо логічну формулу цього показника:

$$Урожайність = \frac{валовий\ збір}{фактично\ зібрана\ площа}.$$

Якщо в умові задачі інформації про розмір площі немає (відсутні ваги), а є інформація про валовий збір (чисельник логічної формули) і урожайність за різними категоріями господарств, то використаємо *середню гармонічну*.

Середня геометрична визначається як добуток відносних величин динаміки $ВВД_i$, які є співвідношенням *i*-го значення показника до попереднього (i–1):

$$\overline{X} = \sqrt[n]{ВВД_1 \cdot ВВД_2 \cdot ВВД_3 \cdot \ldots \cdot ВВД_n} = \sqrt[n]{\prod_{1}^{n} ВВД_i},$$

де **П** – символ добутку; n – число значень *ВВД*, які осереднюються.

Середня квадратична використовується для характеристики варіації (див. параграф 5.3 навчального посібника).

Метод середніх величин широко використовується в економічних дослідженнях. При цьому вимірюються самі середні величини й оцінюються коливання навколо них індивідуальних значень ознак досліджуваного явища.

Розв’язання типових задач

Задача 1

На підприємстві виготовляють різні зошити. Запланований та фактичний обсяги виробництва наведені нижче. Визначити відносні величини планового завдання, виконання плану і динаміки (у перерахунку на зошити 12 аркушів). Показати взаємозв’язок між обчисленими відносними величинами.

Зошити	Випуск, тис. шт.		
	фактично за минулий рік	заплановано на звітний рік	фактично у звітному році
12 аркушів	2500	2600	2600
18 аркушів	2300	2200	2250
24 аркуші	2000	2100	2050
48 аркушів	1500	1500	1500
96 аркушів	1000	800	850

Розв'язання задачі

З метою зіставлення випуску зошитів різної кількості аркушів для перерахунку випуску в умовно-натуральні одиниці обчислимо спочатку необхідні для цього коефіцієнти перерахунку:

$$K_{12\ арк} = \frac{12}{12} = 1{,}0; \qquad K_{18\ арк} = \frac{18}{12} = 1{,}5; \qquad K_{24\ арк} = \frac{24}{12} = 2{,}0;$$

$$K_{96\ арк} = \frac{96}{12} = 8{,}0. \qquad K_{48\ арк} = \frac{48}{12} = 4{,}0;$$

Визначені коефіцієнти й обчислені з їх допомогою значення випуску, перераховані в умовно-натуральні одиниці, наведено нижче.

Зошити	Випуск, тис. шт.			Коефіцієнти перерахунку, *K*	Випуск, тис. ум. од.		
	фактично за минулий рік	заплановано на звітний рік	фактично у звітному році		фактично за минулий рік	заплановано на звітний рік	фактично у звітному році
12 аркушів	2500	2600	2600	1,0	2500	2600	2600
18 аркушів	2300	2200	2250	1,5	3450	3300	3375
24 аркуші	2000	2100	2050	2,0	4000	4200	4100
48 аркушів	1500	1500	1500	4,0	6000	6000	6000
96 аркушів	1000	800	850	8,0	8000	6400	6800
Разом	x	x	x	x	23950	22500	22875

Відносна величина планового завдання:

$$ВВПЗ = \frac{22500}{23950} = 0{,}939, \text{ або } 93{,}9\%.$$

Зокрема, за зошитами 96 аркушів:

$$ВВПЗ_1 = \frac{6400}{8000} = 0{,}800, \text{ або } 80{,}0\%.$$

Відносна величина виконання плану:

$$ВВВП = \frac{22875}{22500} = 1{,}017, \text{ або } 101{,}7\%.$$

За зошитами 96 аркушів:

$$ВВВП_1 = \frac{6800}{6400} = 1{,}063, \text{ або } 106{,}3\%.$$

Відносна величина динаміки:

$$ВВД = \frac{22875}{23950} = 0{,}955, \text{ або } 95{,}5\%; \; ВВД_1 = \frac{6800}{8000} = 0{,}850, \text{ або } 85{,}0\%.$$

Покажемо взаємозв'язок між обчисленими відносними величинами:

$$ВВД = ВВПЗ \cdot ВВВП,$$

$$0{,}955 = 0{,}939 \cdot 1{,}017 \; ; \; 0{,}850 = 0{,}800 \cdot 1{,}063 \; .$$

Отже, планом за звітний рік передбачалося зменшити випуск зошитів на 6,1% (зокрема, зошитів 96 арк. – на 20%). Фактично випуск зошитів зменшився на 4,5% (зошитів 96 арк. – на 15%). План був перевиконаний на 1,7% (за зошитами 96 арк. – на 6,3%).

Задача 2

Розмір фінансування установи становив (млн грн): на 01.01 – 52; на 01.02 – 45; на 01.03 – 48; на 01.04 – 60. Визначити розмір середньомісячного фінансування установи в I кварталі.

Розв'язання задачі

Маємо моментні показники, рівні часові інтервали, а отже, застосуємо середню хронологічну:

$$\overline{X} = \frac{\frac{X_1 + X_n}{2} + X_2 + X_3 + \ldots + X_{n-1}}{n-1}.$$

$$\overline{X} = \frac{\frac{52+60}{2} + 45 + 48}{4-1} = \frac{149}{3} = 49{,}7 \text{ млн грн.}$$

Тобто середньомісячний розмір фінансування установи в I кварталі становив 49,7 млн грн.

Задача 3

Відомо, що в одну зміну на підприємстві працювало 5 одиниць устаткування, у дві зміни – 20 одиниць, у три зміни – 25. Визначити середню змінність роботи (кількість змін) устаткування підприємства. Обґрунтувати вибір форми середньої величини і пояснити одержаний результат.

Розв'язання задачі

Маємо згруповані дані (за змінами роботи) по підприємству, на якому працювало всього 50 одиниць устаткування. Обчислимо відносні величини структури:

в одну зміну працювало $BBC_1 = \frac{5}{50} \cdot 100 = 10\%$,

у дві зміни – $BBC_2 = \frac{20}{50} \cdot 100 = 40\%$,

у три зміни – $BBC_3 = \frac{25}{50} \cdot 100 = 50\%$ усього устаткування.

Для визначення середньої змінності роботи устаткування застосуємо середню арифметичну зважену:

$$\overline{X} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j}; \quad \overline{X} = \frac{1 \cdot 5 + 2 \cdot 20 + 3 \cdot 25}{50} = 2{,}4 \text{ зміни,}$$

або $\overline{X} = \sum_1^m x_j d_j$; $\overline{X} = 1 \cdot 0{,}1 + 2 \cdot 0{,}4 + 3 \cdot 0{,}5 = 2{,}4$ зміни.

Отже, середня змінність роботи устаткування становила 2,4 зміни.

Задача 4

Використовуючи наведені нижче дані, визначити середню місячну заробітну плату одного робітника трьох підприємств.

Підприємства	Середня місячна заробітна плата робітника, грн	Фонд заробітної плати, тис. грн
1	7 200	2 560
2	6 700	1 755
3	6 950	3 245
Разом	х	7 560

Розв'язання задачі

Логічна формула показника середньої місячної заробітної плати робітника виглядатиме так:

$$\textit{Заробітна плата одного робітника} = \frac{\textit{фонд заробітної плати}}{\textit{кількість робітників}}.$$

Згідно з умовою задачі, інформація про знаменник логічної формули показника, тобто про ваги, відсутня (кількість робітників). Однак відома інформація про чисельник формули – фонд заробітної плати. Отже, застосовуємо середню гармонічну зважену.

$$\overline{X} = \frac{\sum_{1}^{m} z_j}{\sum_{1}^{m} \frac{z_j}{x_j}}; \quad \overline{X} = \frac{7560}{\frac{2560}{7{,}20} + \frac{1755}{6{,}70} + \frac{3245}{6{,}95}} = \frac{7560}{1084{,}4} = 6{,}971 \text{ тис. грн.}$$

Тобто середня місячна заробітна плата одного робітника трьох підприємств становить 6971 грн.

Задача 5

Кількість зареєстрованих розлучень в області змінювалася так (% до попереднього року):

2017	2018	2019	2020
–5,3	+9,4	+ 4,9	–4,0

Визначити середній темп зміни кількості зареєстрованих розлучень протягом всього вищезазначеного періоду.

Розв'язання задачі

Оскільки маємо динаміку зареєстрованих розлучень за ряд років у відсотках до попереднього року, то визначення середньої величини проводитимемо за формулою середньої геометричної:

$$\overline{X} = \sqrt[n]{ВВД_1 \cdot ВВД_2 \cdot ВВД_3 \cdot ... \cdot ВВД_n}.$$

Переведемо подані у відсотках зміни у відносні величини динаміки в коефіцієнтах:

$$ВВД_{2017} = \frac{-5,3 + 100}{100} = 0,947; \quad ВВД_{2018} = 1,094;$$

$$ВВД_{2019} = 1,049; \qquad ВВД_{2020} = 0,960.$$

Отже, середньорічний темп зміни кількості зареєстрованих розлучень становив:

$$\overline{X} = \sqrt[4]{0,947 \cdot 1,094 \cdot 1,049 \cdot 0,960} = \sqrt[4]{1,043} = 1,011 \text{ або } 101,1\%.$$

Тобто кількість зареєстрованих розлучень в області за цей період зростала у середньому щорічно на 1,1%.

Запитання для самоконтролю

1. Що являють собою статистичні показники? Наведіть їх приклади.
2. Які функції виконують статистичні показники?
3. Які види статистичних показників Ви знаєте?
4. Яке визначення абсолютних статистичних величин Ви можете назвати? Які одиниці вимірювання вони мають? Наведіть їх приклади.
5. Що являють собою відносні величини? Що вони показують?
6. Які види відносних величин Ви можете назвати?
7. Що характеризують відносні величини динаміки? Наведіть їх приклади.
8. Що являють собою відносні величини структури? Наведіть їх приклади.
9. Що показують відносні величини координації? Наведіть їх приклади.

10. Яке визначення відносної величини порівняння Ви можете надати? Що може бути базою порівняння? Наведіть власні приклади застосування цієї відносної величини.

11. Що характеризують відносні величини інтенсивності? Наведіть їх приклади.

12. Що таке середня величина?

13. Якими є основні вимоги до наукового застосування середніх величин?

14. Які види середніх величин Ви знаєте?

15. В яких випадках застосовують середню хронологічну?

16. Назвіть форми середніх величин, які Ви знаєте. В яких випадках застосовують кожну з них?

17. Що являє собою логічна формула показника? Наведіть власні приклади. В яких випадках доцільно використовувати середню гармонічну?

19. Яким чином визначають середню геометричну?

Завдання для самостійної роботи

Тестові завдання

1. Серед нижченаведених величин оберіть відносну величину інтенсивності:

а) щільність населення України становить 74 особи на 1 км^2;
б) на 100 чоловіків в Україні припадає 116 жінок;
в) ВВП у розрахунку на одну особу в Україні становить 3660 USD;
г) частка готівки в загальному обсязі грошової маси становить 37%.

Відповіді: 1) а; 2) б; 3) а, в; 4) б, г.

2. Зазначте відносну величину динаміки:

а) частка дітей, народжених матерями, які не перебували у зареєстрованому шлюбі, становила 20,1%;
б) чисельність населення області скоротилася за рік на 10 тис. осіб;

в) виручка підприємства від реалізації алкогольних напоїв у грудні місяці була удвічі більшою, ніж у листопаді;
г) чисельність населення країни скоротилася за рік на 2,7%.

Відповіді: 1) а; 2) б, в; 3) в, г; 4) б.

3. Зазначте відносну величину координації:
а) щільність населення України становить 74 особи на 1 км^2;
б) на 100 чоловіків в Україні припадає 116 жінок;
в) ціна вершкового масла у місті N – 225 грн/кг, у місті М – 250 грн/кг;
г) виробництво м'яса на одну особу населення в регіоні становить 50 кг.

Відповіді: 1) а, г; 2) б; 3) в; 4) б, в.

4. Зазначте відносну величину структури:
а) частка готівки в загальному обсязі грошової маси становить 37%;
б) площа під посівами ярових зернових у господарстві утричі менша, ніж під посівами озимих;
в) частка дітей, народжених матерями, які не перебували у зареєстрованому шлюбі, становила 20,1%;
г) доходи підприємства за рік зменшилися порівняно з попереднім роком на 10%.

Відповіді: 1) а; 2) б; 3) а, в; 4) г.

5. Заробітну плату працюючих на підприємстві скоригували на рівень інфляції (становить 1,3). Середня заробітна плата працюючих:
а) залишилася такою самою, як і була;
б) збільшилася відповідно до рівня інфляції (в 1,3 раза);
в) збільшилася на 30%;
г) зросла в два рази.

Відповіді: 1) а; 2) б; 3) б, в; 4) г.

Практичні завдання (задачі)

6. Відомі такі дані про виробництво двигунів на підприємстві:

Потужність двигуна, кВт	Виробництво двигунів, одиниць	
	базисний рік	поточний рік
30	850	900
60	1500	2000
120	1000	850

Проаналізуйте динаміку виробництва двигунів у цілому на підприємстві в перерахунку на двигун потужністю 30 кВт.

7. Нижче наведено кількість працівників, задіяних в Україні у виконанні наукових досліджень і розробок, за категоріями персоналу у 2015 і 2019 рр., осіб.

Кількість працівників	2015 р.	2019 р.
Усього, у т. ч.	122 504	79 262
– дослідники	90 249	51 121
– техніки	11 178	7 470
– допоміжний персонал	21 077	20 671

Визначте відносні величини структури і динаміки. Які структурні зрушення відбулись у складі різних категорій персоналу за цей період?

8. Маємо такі дані про динаміку зареєстрованих безробітних в Україні за статтю у 2019 р. (на кінець звітного періоду, тис. осіб):

Кількість зареєстрованих безробітних	жовтень	листопад	грудень
Усього, у т. ч.	259,3	288,9	338,2
– жінки	155,8	164,5	177,3
– чоловіки	103,5	124,4	160,9

Визначте відносні величини динаміки, структури та координації. Прокоментуйте одержані результати.

9. Дві молоді родини вирішили порівняти свої витрати на харчування. Перше подружжя витратило в листопаді 6000 грн, у грудні воно планувало витратити на харчування вже на 800 грн більше, а фактично витратило на 2000 грн більше. Друге подружжя витратило в листопаді 8000 грн на харчування, запланувало в грудні витратити 10000 грн, а фактично витратило 9600 грн. Визначте відносні величини динаміки, планового завдання і виконання планів двох родин. Зробіть висновки щодо "апетитів" кожної родини.

10. За наведеними нижче даними визначте валовий внутрішній продукт (ВВП) у розрахунку на одну особу в Україні та динаміку цього показника. Зазначте види використаних відносних величин.

Показники	2018 р.	2019 р.
ВВП у фактичних цінах, млрд грн	3 560,6	3 974,6
Чисельність населення, млн осіб	42,4	42,2

11. Маємо такі дані про динаміку чисельності населення України та окремих країн ЄС (станом на 1 січня, млн осіб):

Країни	2010 р.	2015 р.	2020 р.
Україна	46,0	42,9	41,9
Польща	38,0	38,0	38,0
Франція	64,7	66,5	67,1

Визначте усі можливі види відносних величин. Зробіть висновки.

12. Група студентів-заочників, яка складалася з семи осіб, здала іспит зі статистики так (балів за 100-бальною системою оцінювання): 50, 75, 85, 55, 50, 60, 65. Який середній бал успішності в групі? Обґрунтуйте вибір форми середньої величини.

13. П'ять перукарень в одному з районів міста пропонують надання послуг за такими цінами (грн):

Стрижки	Перукарні				
	"Краса"	"Диво"	"Симона"	"Ганна"	"Чарівниця"
Жіноча	145	145	150	145	165
Чоловіча	130	140	135	135	150

Визначте середню, найнижчу (мінімальну) та найвищу (максимальну) ціни жіночої та чоловічої стрижок.

14. Маємо такі дані про динаміку заборгованості з виплати заробітної плати (млн грн): на 01.01.2016 р. – 2232; на 01.01.2017 р. – 1111; на 01.01.2018 р. – 960; на 01.01.2019 р. – 806; на 01.01.2020 р. – 669. Якою є середньорічна сума заборгованості з виплати заробітної плати за 2016–2020 рр.? Обґрунтуйте вибір виду середньої величини.

15. Студенти факультету банківської справи за результатами екзаменаційної сесії одержали такі оцінки з банківської статистики (за 5-бальною системою оцінювання знань):

Оцінка знань, балів	5	4	3	2	Разом
Кількість оцінок	20	68	48	14	150

Визначте середній бал успішності студентів факультету з банківської статистики.

16. На підприємстві працюють два цехи з 80 одиницями устаткування з різною змінністю роботи:

Режим роботи	Кількість одиниць устаткування	
	Цех № 1	Цех № 2
Одна зміна	10	11
Дві зміни	16	16
Три зміни	20	7

Визначте середню змінність роботи устаткування кожного цеху. Поясніть, чому середні величини відрізняються.

17. Три перукарі-комбіновані майстри (Тетяна, Світлана та Наталя), які працюють у перукарні "Світло", протягом тижня виконали такі стрижки:

Вид стрижки	Тетяна	Світлана	Наталя
Дитяча	5	3	2
Жіноча звичайна	15	20	17
Жіноча модельна	10	8	6
Чоловіча звичайна	20	25	30
Чоловіча модельна	7	8	10

У перукарні встановлено такий прейскурант (грн): стрижка дитяча – 120, жіноча звичайна – 140, жіноча модельна – 160, чоловіча звичайна – 130, чоловіча модельна – 145.

Визначте виручку від наданих послуг кожного перукаря за тиждень. Якою була середня ціна виконаних стрижок кожного перукаря? Поясніть, чому середні ціни відрізняються.

18. Обсяг реалізації платних послуг за трьома районами міста був таким:

Райони	Чисельність населення, тис. осіб	Обсяг реалізації платних послуг, млн грн
1	207,1	884
2	339,6	1345
3	306,4	1186

Яким був обсяг реалізації платних послуг на одну особу в цілому за трьома районами міста?

19. За наведеними нижче даними визначте, скільки квадратних метрів загальної площі припадає в середньому на одного мешканця України у кожному році. Порівняйте одержані результати.

Житловий фонд	Усього житловий фонд, млн $м^2$ загальної площі		Припадає житлового фонду на одного мешканця, $м^2$	
	базисний рік	поточний рік	базисний рік	поточний рік
Міський	664,5	677,6	20,5	21,2
Сільський	375,5	380,0	24,4	25,4
У цілому	1040,0	1057,6	x	x

20. У результаті перевірки було виявлено, що частина виробів є бракованою:

Партія товару	Кількість перевірених виробів, штук	Частка забракованих виробів, %
№ 1	2500	3,5
№ 2	1650	2,7
№ 3	560	0,9

Визначте середню частку забракованих виробів за всіма партіями товару. Скільки було забраковано виробів (шт.) у кожній партії?

21. Маємо такі дані про частку міського населення за окремими регіонами країни.

Регіони	Частка міського населення, %	Загальна чисельність населення, млн осіб
K	61	4,1
M	45	5,2
S	52	3,8

Визначте частку сільського населення в цілому по країні.

22. Загальні витрати виробництва і собівартість одиниці продукції по кожному з трьох нижченаведених підприємств характеризуються такими даними:

Номер підприємства	Витрати виробництва, тис. грн	Собівартість одиниці продукції, грн
1	500	250
2	720	240
3	630	252

Визначте середню собівартість одиниці продукції в цілому на трьох підприємствах.

23. Відомі такі дані про частку осіб старше працездатного віку та щільність населення за окремими регіонами країни.

Регіони	Щільність населення, осіб/км^2	Частка осіб старше працездатного віку в загальній чисельності населення, %	Чисельність осіб старше працездатного віку, тис. осіб
C	76	42	570
F	58	49	820
N	49	40	790

Визначте частку осіб старше працездатного віку і щільність населення в цілому за трьома регіонами. У кожному випадку побудуйте логічну формулу показника.

24. Споживання картоплі й олії на одну особу за рік (кг) у країні змінювалося так (у % до попереднього року):

Продукт	2017 р.	2018 р.	2019 р.	2020 р.
Картопля	+2,2	–3,5	–1,5	–3,0
Олія	+15,0	+3,8	+0,7	+5,1

Як у середньому щороку змінювалося споживання картоплі та олії на одну особу (у %) у країні за цей період часу?

Розділ 5.
РЯДИ РОЗПОДІЛУ. АНАЛІЗ ВАРІАЦІЙ ТА ФОРМИ РОЗПОДІЛУ

5.1. Сутність і види рядів розподілу

Після обробки та систематизації первинних статистичних матеріалів одержують ряди показників, які характеризують окремі аспекти досліджуваних явищ і процесів. Такі статистичні ряди за своїм змістом поділяють на ряди розподілу та ряди динаміки.

Ряд розподілу характеризує склад, структуру сукупності за певною ознакою. *Елементами* ряду розподілу є варіанти x_j та частоти f_j. *Варіантами* називають окремі значення ознаки групування. *Частоти* показують, як часто зустрічаються в ряду розподілу ті чи інші варіанти.

Наприклад, маємо ряд розподілу студентів-заочників однієї групи за віком (табл. 5.1):

Таблиця 5.1

Розподіл студентів за віком

Групи студентів за віком, років	До 20	20–25	25–30	30 і старші
Кількість студентів, осіб	8	10	4	2

У цьому випадку варіантами виступають значення віку студентів, а частотами – кількість студентів різних вікових груп.

Залежно від статистичної природи варіант ряди розподілу поділяються на *атрибутивні* (в їх основу покладено якісні ознаки) та *варіаційні* (за ознаками, які мають кількісний вираз).

У співвідношенні варіант і частот проявляється *закономірність розподілу*. Вона описується такими статистичними характеристиками як:

– частотні характеристики;
– характеристики центру розподілу;
– характеристики варіації;
– характеристики нерівномірності розподілу, концентрації, асиметрії.

Частотні характеристики – це абсолютна чисельність *j-ї* групи (*частота* f_j) та відносна частота (*частка* d_j).

Додатковою характеристикою варіаційних рядів є *кумулятивні (накопичені) частоти* S_{f_j} (частки S_{d_j}), які утворюються послідовним підсумовуванням абсолютних (f_j) або відносних (d_j) частот. Наведемо приклад визначення кумулятивних частот (табл. 5.2).

Таблиця 5.2

Допоміжна таблиця для розрахунку кумулятивних частот

Групи студентів за віком, років	x_j	До 20	20–25	25–30	30 і старші	Разом
Кількість студентів, осіб	f_j	8	10	4	2	24
Кумулятивні частоти	S_{f_j}	8	8+10=18	18+4=22	22+2=24	x

Отже, $S_{f_1} = f_1 = 8$; $S_{f_2} = f_1 + f_2 = 8 + 10 = 18$;

$S_{f_3} = f_1 + f_2 + f_3 = S_{f_2} + f_3 = 18 + 4 = 22$;

$S_{f_4} = f_1 + f_2 + f_3 + f_4 = S_{f_3} + f_4 = 22 + 2 = 24$.

Якщо інтервали варіаційного ряду нерівні, то для зіставлення потрібно частоти або частки привести до одиниці інтервалу. Це відношення має назву *щільності розподілу* (абсолютної або відносної):

$$g_j = \frac{f_j}{h_j} \quad \text{або} \quad g_j = \frac{d_j}{h_j},$$

де g_j – щільність розподілу в *j*-му інтервалі; h_j – ширина *j*-го інтервалу.

Щільність розподілу використовується як для розрахунку узагальнюючих показників, так і для графічного зображення варіаційних рядів з нерівними інтервалами.

5.2. Характеристики центру розподілу

До характеристик центру розподілу відносять середню величину, моду та медіану. *Середня величина* характеризує типовий рівень ознаки у сукупності. За даними ряду розподілу середня визначається як середня арифметична зважена:

– на основі частот $\bar{x} = \frac{\sum_{1}^{m} x_j \cdot f_j}{\sum_{1}^{m} f_j}$;

– на основі часток $\bar{x} = \sum_{1}^{m} x_j \cdot d_j$,

де m – число груп; x_j – середини інтервалів.

В інтервальних рядах розподілу, припускаючи рівномірний розподіл у межах j-го інтервалу, як варіанту x_j використовують *середину інтервалу*. При цьому ширина відкритого інтервалу умовно вважається такою самою, як і найближчого закритого інтервалу. Так, перший інтервал (відкритий) умовно прирівнюється за шириною до другого (закритого), а останній (відкритий) – до передостаннього (закритого).

Найпоширеніше значення ознаки або значення ознаки, яке зустрічається в ряду розподілу найчастіше, називається *модою* (*Mo*).

У *дискретних рядах розподілу* моду визначають візуально, без додаткових розрахунків, як *значення ознаки з найбільшою частотою (часткою)*. Найчастіше зустрічаються ряди розподілу з одним модальним значенням ознаки. Ряди, що мають дві моди, називають бімодальними; ряди, що мають декілька мод, – полімодальними.

В *інтервальних рядах розподілу* за найбільшою частотою (часткою) визначають спочатку *модальний інтервал*. У межах цього інтервалу знаходять значення так званої *точкової моди*:

$$Mo = x_0 + h \cdot \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})},$$

де x_0 – нижня межа модального інтервалу; h – його ширина; f – частоти (замість частот можуть бути частки): $f_{Mo\text{-}1}$ – передмодального інтервалу, f_{Mo} – модального інтервалу, f_{Mo+1} – післямодального інтервалу.

Мода являє собою значення, яке найчастіше зустрічається, або типове. Її широко використовують, наприклад, при вивченні купівельного попиту, при визначенні розмірів одягу та взуття, які мають більший попит.

Медіана (Me) – це варіанта, яка припадає на середину упорядкованого ряду розподілу і ділить його на дві рівні за обсягом частини: зі значеннями ознаки, меншими за медіану, і зі значеннями ознаки, більшими за медіану.

Медіана не залежить від крайніх значень ознаки. Через це вона застосовується для характеристики центру в ряду розподілу з невизначеними межами. *Для визначення медіани використовують кумулятивні частоти (частки)*. Їх також називають *накопиченими*.

Медіана ділить ряд розподілу навпіл, а отже, в дискретному ряду розподілу медіаною є значення ознаки, де накопичена частота складає половину або більше половини всієї суми частот, а попередня накопичена частота – менше половини обсягу сукупності.

В *інтервальному варіаційному ряду* таким чином визначається *медіанний інтервал*. А значення *медіани* обчислюється за формулою:

$$Me = x_0 + h \cdot \frac{0{,}5 \cdot \sum_1^m f_j - S_{f_{Me-1}}}{f_{Me}},$$

де x_0 та h – відповідно, нижня межа та ширина медіанного інтервалу; f_{Me} – частота медіанного інтервалу; $S_{f_{Me-1}}$ – кумулятивна частота передмедіанного інтервалу.

Якщо значення середньої величини, моди і медіани збігаються ($\overline{x} = Me = Mo$), то такий ряд розподілу є *симетричним*. Для правосторонньої асиметрії $\overline{x} > Me > Mo$; для лівосторонньої – $\overline{x} < Me < Mo$. Для помірно асиметричних розподілів справедливою є рівність:

$|Mo - \overline{x}| = 3 \cdot |Me - \overline{x}|$.

В аналізі закономірностей розподілу використовуються також інші порядкові характеристики: квартилі й децилі. Квартилі Q – це значення варіант, які ділять упорядкований ряд за обсягом на чотири рівних частини, а децилі D – значення варіант, які ділять упорядкований ряд на десять рівних частин. Отже, в ряду розподілу визначають

три квартилі та дев'ять децилів. Медіана є другим квартилем і п'ятим децилем. Визначення квартилів і децилів, як і медіани, ґрунтується на кумулятивних частотах (частках).

Так, перший і третій квартилі визначатимуться за такими формулами:

$$Q_1 = x_0 + h\,\frac{0{,}25\sum_1^m f_j - S_{Q_1-1}}{f_{Q_1}}; \qquad Q_3 = x_0 + h\,\frac{0{,}75\sum_1^m f_j - S_{Q_3-1}}{f_{Q_3}},$$

де x_0 – нижня межа квартильного інтервалу; h – його ширина; f_Q – частота квартильного інтервалу; $S_{Q\text{-}1}$ – кумулятивна частота передквартильного інтервалу.

Перший та дев'ятий децилі визначатимуться так:

$$D_1 = x_0 + h\,\frac{0{,}1\sum_1^m f_j - S_{D_1-1}}{f_{D_1}}; \qquad D_9 = x_0 + h\,\frac{0{,}9\sum_1^m f_j - S_{D_9-1}}{f_{D_9}},$$

де x_0 – нижня межа децильного інтервалу; h – його ширина; f_D – частота децильного інтервалу; $S_{D\text{-}1}$ – кумулятивна частота переддецильного інтервалу.

Кількісні зміни значень ознаки при переході від однієї одиниці сукупності до іншої називаються *варіацією*. Чим меншою є варіація, тим більш надійними й типовими є характеристики центру розподілу, насамперед середня величина.

5.3. Характеристики варіації

Для вимірювання та оцінки варіації використовують *абсолютні та відносні характеристики*. До *абсолютних* належать: розмах варіації (варіаційний розмах), середнє лінійне відхилення, середнє квадратичне відхилення, дисперсії; до *відносних* – коефіцієнти варіації, нерівномірності, локалізації, концентрації.

Розмах варіації (R) характеризує діапазон варіації ознаки та є різницею між максимальним (x_{max}) і мінімальним (x_{min}) значеннями ознаки: $R = x_{max} - x_{min}$. Оскільки розмах варіації визначається на основі тільки двох крайніх значень ознак сукупності, він дає уявлення про загальний розмір варіації і не дозволяє встановити рівень варіації

всередині сукупності. З цієї причини його використовують для попередньої наближеної оцінки варіації.

Більш точними показниками варіації є *середнє лінійне відхилення* ($\bar{l}$) та *середнє квадратичне відхилення* (σ), які ґрунтуються на відхиленні всіх індивідуальних значень ознаки від середньої величини ($x_j - \bar{x}$). Оскільки алгебраїчна сума відхилень $\sum_1^m (x_j - \bar{x}) \cdot f_j = 0$, то використовують або модулі $|x_j - \bar{x}|$, або квадрати $(x_j - \bar{x})^2$ відхилень. Середнє лінійне відхилення і середнє квадратичне відхилення є іменованими величинами та визначаються в одиницях вимірювання ознаки.

Для згрупованих даних $\bar{l}$ і σ розраховуються за принципом зваженої середньої:

$$\bar{l} = \frac{\sum_1^m \left| x_j - \bar{x} \right| \cdot f_j}{\sum_1^m f_j}; \qquad \sigma = \sqrt{\frac{\sum_1^m (x_j - \bar{x})^2 \cdot f_j}{\sum_1^m f_j}}.$$

За первинними (незгрупованими) даними зазначені характеристики варіації визначаються як проста середня:

$$\bar{l} = \frac{\sum_1^n \left| x_j - \bar{x} \right|}{n}; \qquad \sigma = \sqrt{\frac{\sum_1^n (x_j - \bar{x})^2}{n}}.$$

Середнє лінійне і середнє квадратичне відхилення за змістом ідентичні, через математичні властивості $\sigma > \bar{l}$. У симетричному, близькому до нормального, розподілі $\sigma = 1{,}25 \cdot \bar{l}$, $P = 6 \cdot \sigma = \bar{x} \pm 3 \cdot \sigma$.

Дисперсія – це середній квадрат відхилень σ^2. Застосування дисперсії ширше, ніж тільки для оцінки варіації. Її використовують, зокрема, при вимірюванні взаємозв'язків. Для ознак метричної шкали визначення дисперсії проводиться за такими формулами:

$$\sigma^2 = \frac{\sum_1^m (x_j - \bar{x})^2 \cdot f_j}{\sum_1^m f_j} \quad \text{або} \quad \sigma^2 = \overline{x^2} - \bar{x}^2.$$

Для альтернативної ознаки, варіація якої має два взаємовиключні значення – "1" (наявність) та "0" (відсутність), а розподіл характеризується, відповідно, двома частками – p і q ($p + q = 1$, звідки $p = 1 - q$,

$q = 1 - p$), дисперсія обчислюється як добуток часток $\sigma^2 = p \cdot q = p \cdot (1 - p)$.

При статистичному аналізі варіацій різних ознак або порівнянні варіації однієї і тієї самої ознаки в різних сукупностях використовують відносні характеристики варіації.

Коефіцієнти варіації обчислюються як відношення абсолютних характеристик варіації (R, $\bar{l}$, σ) до центру розподілу $\bar{x}$ і найчастіше виражаються у відсотках:

– *коефіцієнт осциляції* $V_R = \frac{R}{\bar{x}} \cdot 100$;

– *лінійний коефіцієнт* $V_l = \frac{\bar{l}}{\bar{x}} \cdot 100$;

– *квадратичний коефіцієнт* $V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100$.

Чим більшою є величина коефіцієнта варіації, тим більшим є розсіювання значень ознаки навколо середньої величини, тим більша неоднорідність сукупності.

Квадратичний коефіцієнт варіації V_σ використовують як *критерій однорідності сукупності.* Вважається, що при значенні $V_\sigma \leq 33\%$ сукупність є однорідною, а $\bar{x}$ – це типова і надійна характеристика цієї сукупності.

Аналіз закономірностей розподілу передбачає оцінку рівня однорідності сукупності та симетричності розподілу.

Найпростішою мірою асиметричності розподілу є відхилення між характеристиками центру розподілу. Оскільки у симетричному розподілі $\bar{x} = Me = Mo$, то чим помітніша асиметрія, тим більше відхилення ($\bar{x} - Mo$). Стандартне відхилення називають *коефіцієнтом асиметрії*:

$$As = \frac{\bar{x} - Mo}{\sigma}.$$

У випадку правосторонньої асиметрії коефіцієнт $As > 0$, лівосторонньої – $As < 0$.

5.4. Оцінювання рівномірності розподілу, подібності структур, структурних зрушень

Оцінка рівномірності розподілу значень ознаки між окремими складовими сукупностей ґрунтується на порівнянні часток двох розподілів за кількістю елементів сукупності d_j та за обсягом значень ознаки D_j. У випадку *рівномірного розподілу* значень ознаки $d_j = D_j$. Відхилення часток свідчить про певну нерівномірність розподілу, яка вимірюється коефіцієнтами:

– *локалізації:* $L_j = \frac{D_j}{d_j}$;

– *концентрації:* $K = \frac{1}{2}\sum_{1}^{m}\left|d_j - D_j\right|$.

Коефіцієнти локалізації обчислюються для кожної *j*-ї складової сукупності. У випадку рівномірного розподілу всі значення $L_j = 1$. Якщо значення ознаки концентруються в *j*-й складовій, то $L_j > 1$.

Коефіцієнт концентрації є узагальнюючою характеристикою відхилення розподілу від рівномірного. Його значення коливаються в межах від 0 до 1. У випадку рівномірного розподілу $K = 0$. Чим помітніша концентрація, тим більше значення K відхиляється від 0.

Коефіцієнти концентрації та локалізації є ефективним засобом вимірювання диференціації сукупності за даними інтервальних рядів розподілу з нерівними інтервалами та за даними атрибутивних рядів.

Для порівняння структур двох об'єктів або одного об'єкта за двома ознаками визначають *коефіцієнт подібності (схожості) структур P*:

$$P = 1 - \frac{1}{2}\sum\left|d_j - d_k\right|.$$

Чим більшими є відхилення структур, тим менше значення коефіцієнта P. Якщо структури однакові, то $P = 1$.

Структура того чи іншого явища не залишається постійною ані у часі, ані у просторі. Структури однойменних явищ, що стосуються різних просторових пунктів, підприємств тощо, як правило, не збігаються. Відмінності в структурі однойменних явищ (порівнюваних явищ) визначаються різницею часток окремих частин сукупностей.

Зміна часток окремих частин явища у часі свідчить про зміну структури, тобто *структурні зрушення*. Динамічний аналіз показників структури є дуже важливим для вивчення закономірностей розвитку економічних явищ у часі.

Для окремих складових сукупності (її частин) розглядають такі показники структурних змін у часі:

- ✓ *абсолютні показники* – абсолютні прирости часток (Δd_j):

$$\Delta d_j = d_{j1} - d_{j0},$$

де d_{j1} та d_{j0} – частки розподілу за два періоди (поточний і базисний);

- ✓ *відносні показники* – коефіцієнти зростання питомої ваги (k_{d_j}):

$$k_{d_j} = \frac{d_{j1}}{d_{j0}}.$$

Темпи зростання структурних зрушень виражаються формулою:

$$T_{d_j} = k_{d_j} \cdot 100\,\%.$$

Абсолютні показники структурних зрушень показують, на скільки процентних пунктів змінилася – збільшилася (“+”) чи зменшилася (“–”) – питома вага кожної частини явища. *Відносні показники* структурних зрушень відображають інтенсивність зміни часток окремих частин явища, тобто відповідають на запитання, у скільки разів або на скільки відсотків збільшилася (зменшилася) питома вага окремих частин явища.

Всебічна оцінка зміни структури явища може бути надана тільки при паралельному використанні обох показників. Абсолютний приріст і коефіцієнт зростання структурних зрушень дозволяють розглянути зміну питомої ваги окремих частин явища, що вивчається. Однак ці показники не можна використати для оцінки структурних змін у цілому за всім явищем та для характеристики ступеня зміни його структури за певний період часу, для порівняння інтенсивності структурних зрушень за різні періоди часу й у просторі.

Інтенсивність структурних змін залежить від ступеня коливання показників абсолютних приростів і коефіцієнтів зростання часток окремих частин досліджуваного явища. В основу побудови узагальнюючих показників структурних зрушень покладено відповідні показники варіації.

Мірою інтенсивності структурних зрушень є середнє лінійне відхилення $\overline{l}_d$ або середнє квадратичне відхилення σ_d часток. Ці характеристики називають *коефіцієнтами структурних зрушень*:

– *лінійний:* $\overline{l}_d = \dfrac{\sum\limits_1^m \left| d_{j1} - d_{j0} \right|}{m}$;

– *квадратичний:* $\sigma_d = \sqrt{\dfrac{\sum\limits_1^m (d_{j1} - d_{j0})^2}{m}}$,

де m – кількість складових сукупності.

5.5. Правило розкладання дисперсій

При групуванні виникає варіація, зумовлена дією різних факторів, які впливають на варіацію ознаки в межах групи і між групами.

У структурованій сукупності, які поділена на групи за факторною ознакою x, дисперсія результативної ознаки y може бути розкладена на:

– дисперсію в кожній групі (*внутрішньогрупову*);
– дисперсію між групами (*міжгрупову*).

Загальна дисперсія характеризує варіацію ознаки y за рахунок впливу всіх причин (факторів).

Міжгрупова дисперсія характеризує результат впливу фактора x (групувального). Її обчислюють за формулою:

$$\delta^2 = \frac{\sum\limits_1^m (\overline{y}_j - \overline{y})^2 \cdot f_j}{\sum\limits_1^m f_j},$$

де $\overline{y}_j$ та – відповідно, середня j-ї групи та загальна середня варіюючої ознаки y; f_j – частота j-ї групи.

Внутрішньогрупові дисперсії розраховуються для кожної j-ї групи:

$$\sigma_j^2 = \frac{\sum\limits_1^{f_j} (y - \overline{y}_j)^2}{f_j},$$

де y – значення ознаки окремих елементів сукупності.

Загалом для всіх груп визначають *середню з внутрішньогрупових дисперсій*, яка показує результат впливу інших факторів, що не враховані у групуванні (тобто всіх факторів, окрім фактора x). Для її розрахунку внутрішньогрупові дисперсії (σ_j^2) зважують на частоти відповідних груп (f_j):

$$\overline{\sigma_j^2} = \frac{\sum_1^m \sigma_j^2 \cdot f_j}{\sum_1^m f_j} \cdot$$

Взаємозв'язок між трьома дисперсіями називається *правилом розкладання дисперсій*:

$$\sigma^2 = \delta^2 + \overline{\sigma_j^2},$$

тобто *загальна дисперсія дорівнює сумі міжгрупової дисперсії і середньої з внутрішньогрупових дисперсій*. Отже, знаючи дві будь-які дисперсії, завжди можна визначити третю на основі наведеної рівності.

Відношення міжгрупової дисперсії до загальної називається *кореляційним відношенням*:

$$\eta^2 = \frac{\delta^2}{\sigma^2} \cdot$$

Розв'язання типових задач

Задача 1

Визначити середній, модальний і медіанний вік студентів за даними вищенаведеної таблиці про розподіл студентів-заочників однієї групи (табл. 5.1).

Розв'язання задачі

Для розрахунку зазначених характеристик центру розподілу побудуємо табл. 5.3.

Таблиця 5.3

Допоміжна таблиця
для розрахунку характеристик центру розподілу

Групи студентів за віком, років, x	Кількість студентів, осіб, f_j	Кумулятивні частоти, осіб, S_{f_j}	Середини інтервалів вікових груп, років, x_j	$x_j \cdot f_j$
До 20	8	8	17,5	140
20–25	10	18	22,5	225
25–30	4	22	27,5	110
30 і старші	2	24	32,5	65
Разом	24	х	х	540

Перший інтервал (відкритий) умовно прирівнюємо за шириною до другого (закритого), а останній (відкритий) – до передостаннього (закритого). Так, якщо ширина другого інтервалу дорівнює п'ять років (20–25), то вважаємо, що ширина першого інтервалу буде також п'ять років (15–20), відповідно його середина – 17,5.

Середнє значення віку визначимо за формулою середньої арифметичної зваженої:

$$\overline{x} = \frac{\sum_{1}^{m} x_j \cdot f_j}{\sum_{1}^{m} f_j} = \frac{540}{24} = 22{,}5 \text{ р.}$$

Серед студентів найбільш вагомою віковою групою є 20–25 років, представників цієї групи налічується 10 осіб. Отже, модальний інтервал визначено. За формулою розрахуємо значення моди в межах цього інтервалу:

$$Mo = x_0 + h \cdot \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})}.$$

Нижня межа визначеного модального інтервалу $x_0 = 20$, його ширина $h = 5$, частота модального інтервалу $f_{Mo} = 10$, частота передмодального інтервалу $f_{Mo\text{-}1} = 8$, частота післямодального інтервалу $f_{Mo+1} = 4$. Тоді мода дорівнюватиме:

$$Mo = 20 + 5 \cdot \frac{10 - 8}{(10 - 8) + (10 - 4)} = 20 + 5 \cdot \frac{2}{8} = 21{,}3 \text{ року.}$$

Отже, в групі найчастіше зустрічаються студенти віком приблизно 21 рік.

Для визначення медіани спочатку з'ясуємо, якою є половина обсягу сукупності. У нашому випадку ця величина обчислюється як $0{,}5 \cdot \sum_1^m f_j = 0{,}5 \cdot 24 = 12$. Перша кумулятивна частота, що перевищує половину обсягу сукупності, становить $S_{f_{Me}} = 18$. Ця накопичена частота належить до другого інтервалу – 20–25 років. У межах визначеного медіанного інтервалу визначимо значення медіани:

$$Me = x_0 + h \cdot \frac{0{,}5 \cdot \sum_1^m f_j - S_{f_{Me-1}}}{f_{Me}}$$

Нижня межа медіанного інтервалу $x_0 = 20$, його ширина $h = 5$. Частота медіанного інтервалу $f_{Me} = 10$, кумулятивна частота передмедіанного інтервалу $S_{f_{Me-1}} = 8$. Тоді

$$Me = 20 + 5 \cdot \frac{0{,}5 \cdot 24 - 8}{10} = 20 + 5 \cdot \frac{4}{10} = 22 \text{ роки.}$$

Отже, половина студентів групи мають вік до 22 років, а інша половина – більше 22 років.

Оскільки $(\bar{x} > Me > Mo) \cdot (22{,}5 > 22 > 21{,}5)$, то маємо правосторонню асиметрію розподілу. Перевіримо рівність : $|Mo - \bar{x}| = 3 \cdot |Me - \bar{x}|$

$$|21{,}3 - 22{,}5| \approx 3 \cdot |22 - 22{,}5|,$$

отже, цей ряд розподілу є помірно асиметричним.

Задача 2

Використовуючи дані попередньої задачі, зробити висновки щодо однорідності досліджуваної сукупності й асиметричності розподілу.

Розв'язання задачі

Для цього побудуємо табл. 5.4.

Таблиця 5.4

Розрахунок показників варіації віку студентів

Вікові групи студентів, років	Кількість студентів, осіб, f_j	Середини інтервалів вікових груп, років, x_j	$(x_j - x)^2 \cdot f_j$
До 20	8	17,5	200
20–25	10	22,5	0
25–30	4	27,5	100
30 і старші	2	32,5	200
Разом	24	x	500

Використовуючи визначене в задачі 1 значення $\bar{x} - 2{,}5$ року, заповнимо останню графу табл. 5.4. Так, у першому інтервалі (до 20 років):

$$(x_j - \bar{x})^2 \cdot f_j = (17{,}5 - 22{,}5)^2 \cdot 8 = 200.$$

Для висновку щодо однорідності досліджуваної сукупності визначимо квадратичний коефіцієнт варіації: $V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100$. Середнє квадратичне відхилення для згрупованих даних розрахуємо за формулою

$$\sigma = \sqrt{\frac{\sum_1^m (x_j - \bar{x})^2 \cdot f_j}{\sum_1^m f_j}}.$$

Використовуючи дані табл. 5.4:

$$\sigma = \sqrt{\frac{500}{24}} = 4{,}6 \text{ року}.$$

Отже, значення віку студентів групи відрізняються від середнього значення ($\bar{x} = 22{,}5$ року) у середньому на 4,6 року.

Квадратичний коефіцієнт варіації становитиме:

$$V_\sigma = \frac{4,6}{22,5} \cdot 100 = 20,4\%,$$

що свідчить про однорідність досліджуваної сукупності.

Визначимо коефіцієнт асиметрії:

$$As = \frac{\bar{x} - Mo}{\sigma} = \frac{22,5 - 21,3}{4,6} = 0,26.$$

Оскільки $As > 0$, робимо висновок про правосторонню асиметрію.

Задача 3

Склад засуджених у регіоні за віковими групами характеризується такими даними:

Вікові групи, років	14–18	18–25	25–30	30–50	50–80	Разом
Кількість засуджених, осіб	120	840	450	1344	60	2814

Визначити модальний вік засуджених.

Розв'язання задачі

Оскільки маємо варіаційний ряд розподілу з нерівними інтервалами, то для зіставлення потрібно частоти (кількість засуджених) привести до одиниці інтервалу. Визначимо абсолютні щільності розподілу за формулою: $g_j = \frac{f_j}{h_j}$, де g_j – щільність розподілу в j-му інтервалі; h_j – ширина j-го інтервалу.

У результаті розрахунків одержимо такі значення: в першому інтервалі $g_1 = \frac{120}{4} = 30$; у другому – $g_2 = \frac{840}{7} = 120$; у третьому – $g_3 = \frac{450}{5} = 90$; у четвертому – $g_4 = \frac{1344}{20} = 67,2$; у п'ятому – $g_5 = \frac{60}{30} = 2$.

За максимальним значенням щільності розподілу визначимо модальний інтервал. Так, в нашому випадку модальним інтервалом віку засуджених буде 18-25 років (другий інтервал).

За відповідною формулою визначимо значення моди:

$$Mo = 18 + 7 \cdot \frac{120 - 30}{(120 - 30) + (120 - 90)} = 18 + 7 \cdot \frac{90}{120} = 23,3 \text{ (року)}.$$

Отже, найпоширеніший вік засуджених у регіоні – 23 роки.

Задача 4

За наведеними в табл. 5.5 даними визначте коефіцієнти локалізації і концентрації, оцінити нерівномірність розподілу оптового товарообороту суб'єктів оптової торгівлі.

Таблиця 5.5

Групування підприємств області
за обсягом оптового товарообороту

Класифікаційні групи підприємств за обсягом обороту, тис. грн	Частка у підсумку, %	
	кількість підприємств	обсяг оптового обороту
До 500	49,9	0,9
500,1–10000,0	41,7	17,3
10000,1–50000,0	6,7	19,2
50000,1 і більше	1,7	62,6
Разом	100,0	100,0

Розв'язання задачі

Оцінка нерівномірності розподілу значень ознаки між окремими складовими сукупностей ґрунтується на порівнянні часток двох розподілів – за кількістю елементів сукупності d_j та за обсягом значень ознаки D_j. У нашому випадку d_j – це частка j-ї класифікаційної групи підприємств у загальній їх кількості, D_j – частка оптового обороту j-ї групи підприємств у загальному обсязі обороту. Для обчислення коефіцієнтів локалізації (L_j) і концентрації (K) побудуємо табл. 5.6.

Розрахуємо значення коефіцієнта концентрації:

$$K = \frac{1}{2} \cdot 1{,}468 = 0{,}734.$$

Таблиця 5.6

Розрахунок коефіцієнтів нерівномірності розподілу
оптового товарообороту суб'єктів оптової торгівлі

Класифікаційні групи підприємств за обсягом обороту, тис. грн	% до підсумку		$L_j = \frac{D_j}{d_j}$	$\frac{\sum_1^m \lvert d_j - D_j \rvert}{100}$
	кількість підприємств, d_j	обсяг оптового обороту, D_j		
1	2	3	4	5
До 500	49,9	0,9	0,018	0,490
500,1–10000,0	41,7	17,3	0,415	0,244

Продовження табл.5.6

1	2	3	4	5
10000,1–50000,0	6,7	19,2	2,866	0,125
50000,1 і більше	1,7	62,6	36,823	0,609
Разом	100,0	100,0	x	1,468

Оскільки значення K можуть коливатися в межах від 0 до 1, то $K = 0{,}734$ свідчить про достатньо високий ступінь концентрації обсягу оптового товарообороту за підприємствами, що займаються оптовою торгівлею.

Задача 5

Відомі такі дані про структуру перевезення пасажирів за видами транспорту в одній з областей (табл. 5.7):

Таблиця 5.7

Структура перевезення пасажирів за видами транспорту

Види транспорту	Частка у підсумку, %	
	базисний рік, d_{j0}	поточний рік, d_{j1}
Залізничний	7,9	8,6
Автомобільний	53,1	61,4
Тролейбусний	39,0	30,0
Разом	100,0	100,0

Проаналізувати інтенсивність структурних зрушень за допомогою квадратичного коефіцієнта структурних зрушень.

Розв'язання задачі

Проведемо необхідні розрахунки зміни часток, результати занесемо в табл. 5.8.

Таблиця 5.8

Структура і структурні зрушення перевезення пасажирів за видами транспорту

Види транспорту	Базисний рік, d_{j0}	Поточний рік, d_{j1}	$\Delta_{d_j} = d_{j1} - d_{j0}$, п. п.	$\Delta^2_{d_j} = (d_{j1} - d_{j0})^2$.
1	2	3	4	5
Залізничний	7,9	8,6	+0,7	0,49
Автомобільний	53,1	61,4	+8,3	68,89

Продовження табл.5.7

1	2	3	4	5
Тролейбусний	39,0	30,0	–9,0	81,00
Разом	100,0	100,0	0,0	150,38

Отже, найбільші зміни за цей період часу відбулися за часткою перевезень тролейбусним транспортом, вона зменшилася на 9,0 п. п.

Визначимо квадратичний коефіцієнт структурних зрушень:

$$\sigma_d = \sqrt{\frac{\sum_{1}^{m} (d_{j1} - d_{j0})^2}{m}}; \; m = 3,$$

оскільки маємо три елементи структури (за видами транспорту).

Тоді $\sigma_d = \sqrt{\frac{150,38}{3}} = 7,1.$

Отже, структура перевезення пасажирів за видами транспорту (частки окремих видів транспорту) в області змінилася за досліджуваний період часу в середньому на 7,1 п. п.

Задача 6

Вартість однокімнатних квартир різної загальної площі на первинному ринку центру міста характеризується такими даними:

Групи квартир за розміром загальної площі, м^2	Кількість квартир, f_j, од.	Вартість квартир, y, тис. USD	$\sum y$	$\overline{y}_j = \frac{\sum y}{f_j}$
42–48	6	79,5; 79,9; 85,5; 94,0; 117,0; 140,0	595,9	$99,3 = \frac{595,9}{6}$
48–54	4	98,0; 135,0; 140,0; 159,0	532,0	$133,0 = \frac{532,0}{4}$
54–60	3	150,0; 154,0; 178,5	482,5	$160,8 = \frac{482,5}{3}$
У цілому	13	x	1610,4	$123,9 = \frac{1610,4}{13}$

Визначити внутрішньогрупові, міжгрупову та загальну дисперсії вартості однокімнатних квартир, а також питому вагу міжгрупової дисперсії у загальній. Зробити відповідні висновки.

Розв'язання задачі

Варіація вартості квартир для кожної їх групи визначається трьома внутрішньогруповими дисперсіями (для трьох груп квартир). Для першої групи квартир внутрішньогрупова дисперсія становитиме:

$$\sigma_1^2 = \frac{(79{,}5-99{,}3)^2+(79{,}9-99{,}3)^2+(85{,}5-99{,}3)^2+(94{,}0-99{,}3)^2}{6} +$$

$$+\frac{(117{,}0-99{,}3)^2+(140{,}0-99{,}3)^2}{6} = 492{,}785.$$

Для другої групи квартир:

$$\sigma_2^2 = \frac{(98-133)^2+(135-133)^2+(140-133)^2+(159-133)^2}{4} = 488{,}5.$$

Для третьої групи квартир:

$$\sigma_3^2 = \frac{(150{,}0-160{,}8)^2+(154{,}0-160{,}8)^2+(178{,}5-160{,}8)^2}{3} = 158{,}7.$$

Середня з групових (внутрішньогрупових) дисперсій становитиме:

$$\overline{\sigma_j^2} = \frac{492{,}785\cdot 6+488{,}5\cdot 4+158{,}7\cdot 3}{13} = 414{,}37.$$

Визначимо міжгрупову дисперсію:

$$\delta^2 = \frac{(99{,}3-123{,}9)^2\cdot 6+(133{,}0-123{,}9)^2\cdot 4+(160{,}8-123{,}9)^2\cdot 3}{13} = 619{,}0\,.$$

Загальну дисперсію можемо визначити за правилом складання дисперсій як суму міжгрупової дисперсії і середньої з внутрішньогрупових дисперсій:

$$\sigma^2 = 414{,}37+619{,}0 = 1033{,}37.$$

Також можемо обчислити загальну дисперсію за формулою . У цьому випадку:

$$\sigma^2 = \frac{79{,}5^2+79{,}9^2+85{,}5^2+94{,}0^2+\ldots..+154{,}0^2+178{,}5^2}{13} - 123{,}9^2 = 1033{,}37.$$

Питома вага міжгрупової дисперсії у загальній становить:

$$\eta^2 = \frac{\delta^2}{\sigma^2} = \frac{619{,}0}{1033{,}37} = 0{,}599, \text{ або } 59{,}9\%.$$

Отже, варіація вартості однокімнатних квартир у новобудовах майже на 60% пояснюється (обумовлена) варіацією їх загальної площі та на 40% (100% – 60%) – варіацією інших факторів (які не були враховані у групуванні).

Запитання для самоконтролю

1. Що характеризують ряди розподілу? Які елементи рядів розподілу Ви знаєте?
2. Який приклад ряду розподілу Ви можете навести? Назвіть його варіанти і частоти.
3. У чому проявляється закономірність розподілу? Якими статистичними характеристиками вона описується?
4. Що являють собою частотні характеристики? Яким чином одержують кумулятивні частоти?
5. Які характеристики центру розподілу Ви знаєте?
6. Як обчислюють середню величину в інтервальному ряду розподілу?
7. Яке визначення Ви можете надати моді?
8. Як визначають моду в дискретних та інтервальних рядах розподілу?
9. Як медіана ділить упорядкований ряд розподілу? Які частоти використовують для визначення медіани?
10. Як визначають медіану в дискретних та інтервальних рядах розподілу?
11. Що являє собою варіація? Які характеристики варіації Ви знаєте?
12. Які характеристики варіації відносять до абсолютних, а які – до відносних?
13. Що характеризує розмах варіації?
14. На чому ґрунтується визначення середнього лінійного та середнього квадратичного відхилення? Які одиниці вимірювання мають ці характеристики?

15. Як визначають коефіцієнти варіації? Для чого їх використовують?

16. Критерієм чого є квадратичний коефіцієнт варіації?

17. Як визначають коефіцієнт асиметрії?

18. За допомогою визначення яких коефіцієнтів роблять висновки про нерівномірність розподілу?

19. Який коефіцієнт допомагає порівнювати структури двох об'єктів або одного об'єкта за двома ознаками?

20. Що являють собою структурні зрушення? З якою метою їх визначають?

21. Які показники структурних зрушень Ви знаєте?

22. Поясніть правило складання (розкладання) дисперсій.

Завдання для самостійної роботи

Тестові завдання

1. При визначенні середньої величини в інтервальних рядах розподілу, припускаючи рівномірний розподіл у межах *j*-го інтервалу, як варіанту x_j використовують:

а) нижню межу інтервалу;
б) верхню межу інтервалу;
в) середину інтервалу.

Відповіді: 1) а; 2) б; 3) в.

2. Мода – це:

а) найпоширеніше значення ознаки;
б) варіанта, яка припадає на середину упорядкованого ряду розподілу і ділить його на дві рівні за обсягом частини;
в) характеристика варіації;
г) значення ознаки, яке зустрічається в ряду розподілу найчастіше.

Відповіді: 1) а, в; 2) б; г; 3) а, г; 4) б, в.

3. Медіана – це:

а) характеристика варіації;
б) найпоширеніше значення ознаки;

в) варіанта, яка припадає на середину упорядкованого ряду розподілу і ділить його на дві рівні за обсягом частини;

г) характеристика центру розподілу.

Відповіді: 1) а, в; 2) б; г; 3) б, в; 4) в, г.

4. Середнє квадратичне відхилення для згрупованих даних обчислюється за такою формулою:

а) $\sigma = \sqrt{\frac{\sum_{1}^{m}(x_j - \bar{x})^2 \cdot f_j}{\sum_{1}^{m} f_j}}$; б) $\sigma = \sqrt{\frac{\sum_{1}^{n}(x_j - \bar{x})^2}{n}}$; в) $\sigma = \frac{\sum_{1}^{m}\left|x_j - \bar{x}\right| \cdot f_j}{\sum_{1}^{m} f_j}$;

через математичні властивості

г) $\sigma > \bar{l}$; д) $\sigma < \bar{l}$.

Відповіді: 1) а, д; 2) б; г; 3) в, д; 4) а, г.

5. Згідно з правилом складання (розкладання) дисперсій:

а) $\delta^2 = \overline{\sigma_j^2} + \sigma^2$; б) $\sigma^2 = \delta^2 + \overline{\sigma_j^2}$; в) $\overline{\sigma_j^2} = \delta^2 + \sigma^2$; г) $\overline{\sigma_j^2} = \sigma^2 - \delta^2$.

Відповіді: 1) а, г; 2) б; 3) в; 4) б, г.

Практичні завдання (задачі)

6. У результаті опитування одержано такі дані щодо бажаної кількості дітей (якщо у респондентів різних вікових груп склалися б для цього ідеальні умови), %:

Бажана кількість дітей	Вікова група, років		
	18–25	26–35	36–45
Не бажав(-ла) б мати дітей	4	3	2
Одного	16	10	10
Двох	50	53	48
Трьох	20	26	24
Чотирьох і більше	10	8	16
Разом	100	100	100

Визначте модальну і медіанну бажану кількість дітей респондентів різних вікових груп. Зробіть висновки.

7. Відомі такі дані про розподіл працівників фірми за величиною заробітної плати:

Заробітна плата, грн	До 7600	7600–7700	7700–7800	7800–7900	7900–8000	8000 і більше
Кількість працівників, осіб	4	8	18	30	13	7

Завдання:

а) вкажіть модальний інтервал заробітної плати працівників, обґрунтувавши його вибір;

б) визначте моду поданого ряду розподілу. Зробіть висновки.

8. У результаті опитування 300 школярів про час, який вони витрачають щоденно на ігри у смартфоні, були одержані такі результати:

Час ігри, хв.	До 20	20–40	40–60	60–80	80–100	100 і більше
Кількість дітей, осіб	40	40	60	90	50	20

Завдання: а) визначте модальний час щоденних витрат часу школярами на ігри; б) на основі розрахованих кумулятивних (накопичених) частот визначте, який інтервал буде медіанним; в) обчисліть медіанне значення часу щоденних витрат часу на ігри. Зробіть висновки.

9. Маємо такі дані про розподіл робітників, зайнятих одним із видів економічної діяльності регіону, за віком:

Вікові групи робітників, років	15–20	20–25	25–30	30–50	50–55	55–60	60–75	Разом
Кількість робітників, %	2,0	10,0	8,5	62,0	7,0	7,5	3,0	100,0

Визначте модальне значення віку робітників. Зробіть висновки.

10. Ціни на товари в магазинах міста за результатами вибіркового спостереження становили (грн):

Товар	Магазин						
	1	2	3	4	5	6	7
A	10	12	9	12	9	10	13
B	25	20	27	26	28	24	25
C	40	42	37	41	38	42	40

Проведіть порівняльний аналіз варіації цін товарів. Зробіть висновки.

11. Маємо такі дані про зріст дівчат певної вікової групи:

Групи дівчат за зростом, см	152–156	156–160	160–164	164–168	168–172
Кількість дівчат, осіб	4	8	9	16	3

Обчисліть такі характеристики розподілу: середню величину, моду, медіану, середнє квадратичне відхилення, коефіцієнти варіації та асиметрії. Зробіть висновки про характер розподілу дівчат за зростом.

12. На дослідних ділянках урожайність винограду "аліготе" становила 8 кг з одного куща. Визначте середнє квадратичне відхилення урожайності за умови, що її середній квадрат $\overline{x^2} = 73$.

13. Відомі такі статистичні характеристики розподілу мігрантів за віком (років):

Характеристика розподілу	Прибулі	Вибулі
$\overline{x}$	29,9	30,3
Mo	19,0	20,2
σ	18,8	18,3

Порівняйте варіацію та асиметрію розподілу мігрантів за віком. Зробіть висновки.

14. За наведеними даними визначте: а) коефіцієнти регіональної локалізації водо- та теплопостачання міст; б) коефіцієнти концентрації водо- та теплопостачання. Зробіть висновки.

Регіони	Частка у підсумку, %		
	міського населення	постачання питної води	постачання теплової енергії
A	38,5	37,8	38,9
B	17,3	18,2	15,1
C	20,8	22,4	22,4
D	12,1	8,0	11,8
E	11,3	13,6	11,8
Разом	100,0	100,0	100,0

15. За наведеними даними визначте коефіцієнт концентрації зайнятих на підприємствах одного з видів економічної діяльності:

Групи підприємств за кількістю зайнятих, осіб	Частка у підсумку, %	
	підприємств	зайнятих
До 50	30	4
50–100	35	7
100–200	18	13
200–500	12	32
500 і більше	5	44
Разом	100	100

16. Відомі такі дані про структуру населення за віковими групами в регіонах країни (тис. осіб):

Вікові групи	Регіони		
	A	B	C
Молодше працездатного віку	272,2	256,5	375,6
Працездатного віку	957,3	827,0	1463,6
Віку, старшого за працездатний	435,7	296,3	544,4
Разом	1665,2	1379,8	2383,6

Використовуючи відповідний коефіцієнт, оцініть ступінь подібності структур населення різних регіонів. Зробіть висновки.

17. Структура витрат організацій на виконання власними силами наукових та науково-технічних робіт за їх видами характеризується такими даними (у фактичних цінах, млн грн):

Види робіт	Базисний період	Поточний період
Фундаментальні дослідження	886,1	1462,3
Прикладні дослідження	655,8	1058,5
Науково-технічні розробки	2195,8	2793,5
Науково-технічні послуги	648,6	594,1
Разом	4386,3	5908,4

Оцініть інтенсивність структурних зрушень за допомогою лінійного коефіцієнта. Зробіть висновки.

18. Аналітичне групування підприємств характеризує залежність прибутку від обсягу реалізованої продукції (робіт, послуг). Виділено 4 групи, міжгрупова дисперсія прибутку становить 324, середня з групових дисперсій – 51. Визначте кореляційне відношення. Зробіть висновки.

19. За результатами опитування кількість очікуваних дітей серед жінок з різним рівнем освіти виявилася такою:

Освіта	Опитано жінок, тис. осіб	Середня очікувана кількість дітей, осіб
Середня	9	2,0
Середня спеціальна	5	1,8
Вища	6	1,5
У цілому	20	x

Визначте міжгрупову дисперсію кількості очікуваних дітей.

20. Визначте кореляційне відношення, що характеризує щільність зв'язку між прибутком і обсягом продажу на підприємствах одного з видів економічної діяльності. Прокоментуйте одержаний результат.

Групи підприємств за обсягом реалізованої продукції (робіт, послуг), тис. грн	Кількість підприємств, од.	Чистий прибуток підприємства у середньому по групі, тис. грн
До 400	40	28
400–600	50	58
Понад 600	30	79

Загальна дисперсія чистого прибутку становить 545.

21. У таблиці наведено депозитні процентні ставки 12 відділень банків.

Відділення комерційних банків	Депозитні процентні ставки в окремих відділеннях, %						
Філії	16,0	16,5	16,0	17,0	16,5	17,5	16,5
Центральні	19,5	19,0	20,5	21,0	20,0		

Визначте групові, міжгрупову та загальну дисперсії депозитної процентної ставки. Покажіть їх взаємозв'язок.

22. Відомі такі дані про підприємства трьох районів:

Райони	Кількість підприємств, од.	Чистий прибуток підприємств, млн грн
A	7	2, 4, 6, 3, 7, 8, 5
B	3	7, 9, 10
C	4	11, 19, 21, 16

Визначте групові дисперсії чистого прибутку, середню з групових дисперсій, міжгрупову дисперсію. Покажіть взаємозв'язок

Розділ 6.
ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ

6.1. Завдання вибіркового спостереження

Вибіркове спостереження є найбільш поширеним видом несуцільного спостереження. Воно широко використовується як у практичній роботі органів державної статистики (зокрема, в обстеженнях умов життя домогосподарств, робочої сили, державних статистичних спостереженнях, що складають статистику підприємств), так і в роботі установ, які здійснюють аудиторську, соціологічну, маркетингову, медичну, фармакологічну та інші види діяльності.

Вибірковим спостереженням називається науково обґрунтований вид спостереження, за яким обстежуються не всі елементи сукупності, а лише певним чином відібрана їх частина. Це вид несуцільного спостереження, при якому поширення даних, отриманих за вибірковою сукупністю (вибіркою), на всю досліджувану сукупність має статистико-математичне обґрунтування.

При цьому розрізняють генеральну сукупність та вибіркову сукупність. Та сукупність, з якої проводиться відбір, називається *генеральною сукупністю*; відібрані одиниці складають *вибіркову сукупність*. Саме основні характеристики вибіркової сукупності дають підставу для висновків про параметри та властивості генеральної сукупності.

Для позначення узагальнюючих показників генеральної та вибіркової сукупностей використовують таку систему умовних позначень (табл. 6.1).

Таблиця 6.1

Умовні позначення параметрів генеральної та вибіркової сукупностей

Параметри	Генеральна сукупність	Вибіркова сукупність
Обсяг сукупності	N	n
Середня величина	$\overline{X}$	$\overline{x}$
Відносна величина	D	d
Дисперсія	σ^2	$\sigma_в^2$

Від інших видів несуцільного спостереження (основного масиву, монографічного, анкетного) вибірковий метод відрізняється тим, що попередньо визначається:

1) яка частина одиниць генеральної сукупності буде обстежуватися;
2) порядок відбору одиниць, який би надійно репрезентував середні та відносні параметри генеральної сукупності.

Причин використання вибіркового методу декілька:

1) підвищення точності даних, адже зменшення числа одиниць спостереження у вибірці різко знижує помилки реєстрації, а також дозволяє залучити більше кваліфікованих виконавців (інтерв'юєрів, реєстраторів), що позитивно позначається на якості даних вибіркового обстеження;
2) забезпечує економію матеріальних, трудових і фінансових ресурсів, а також часу;
3) без вибірки не обійтися, коли спостереження пов'язане з псуванням об'єктів, що спостерігаються. На вибірках засновані маркетингові дослідження, оцінки якості поставки.

Практика застосування вибіркового методу дуже різноманітна. Іноді після проведення суцільного спостереження при розробці даних застосовують вибірковий метод: відбирають частину даних для більш докладної розробки за розширеною програмою.

Отже, вибірковий метод має такі *переваги*:

✓ відносно невеликі (порівняно з суцільним спостереженням)

матеріальні, трудові та вартісні витрати на збирання даних (включаючи витрати на планування і формування вибірки);

- ✓ оперативність отримання результатів;
- ✓ широка область застосування;
- ✓ висока достовірність результатів.

Усі ці переваги проявляються, однак, лише за умови правильного розв'язання *проблем вибіркового обстеження*. До них належать:

1) визначення генеральної сукупності;
2) розробка програми спостереження та інструкцій;
3) визначення основи для формування вибірки – списку одиниць генеральної сукупності, відомостей про їх розміщення і т. д.;
4) встановлення допустимого розміру помилки і визначення обсягу вибірки;
5) визначення виду вибіркового спостереження;
6) окреслення термінів проведення спостереження;
7) визначення потреби в кадрах для проведення вибіркового спостереження, їх підготовка;
8) оцінка точності та достовірність результатів спостереження, визначення порядку їх поширення на генеральну сукупність.

Проведення вибіркового спостереження складається з таких етапів:

а) обґрунтування мети вибіркового спостереження;
б) складання програми спостереження і розробка відповідних даних;
в) вирішення організаційних питань проведення спостереження;
г) визначення частки і способу відбору одиниць у вибіркову сукупність;
ґ) проведення відбору;
д) реєстрація ознак у відібраних для дослідження одиниць;
е) узагальнення даних спостереження та визначення їх вибіркових характеристик;
є) обчислення похибок вибірки;

ж) поширення кількісних характеристик вибіркового спостереження на генеральну сукупність.

Вибіркове спостереження проводиться для вирішення таких завдань:

1) визначення середнього розміру досліджуваної ознаки;
2) визначення частки досліджуваної ознаки в певній сукупності;
3) знаходження меж для середнього розміру та частки при повторному і безповторному відборі;
4) поширення даних вибіркового спостереження на генеральну сукупність.

6.2. Методи і способи відбору одиниць до вибіркової сукупності

Відбір одиниць з генеральної сукупності до вибіркової може бути повторним і безповторним. При *повторному відборі* кожна раніше відібрана одиниця повертається у генеральну сукупність і може повторно брати участь у побудові вибірки. При *безповторному відборі* кожна раніше відібрана одиниця не повертається у генеральну сукупність і не бере участь у подальшому відборі. У зв'язку з цим безповторний відбір порівняно з повторним дозволяє отримати точніші результати спостереження: при однаковому обсязі вибірки дослідженням охоплюється більше одиниць за рахунок того, що одна й та сама одиниця не може потрапити до вибірки декілька разів.

Повторний і безповторний методи залежно від характеру відбору одиниць поділяють на такі *види*:

1) індивідуальний (проводиться відбір окремих одиниць сукупності);
2) кластерний (проводиться відбір груп або серій одиниць);
3) комбінований (передбачає комбінацію використання індивідуального та групового видів відбору).

Для формування вибіркової сукупності використовують різні види вибірки за способом добору одиниць сукупності для обстеження та за кількістю ступенів добору (рис. 6.1).

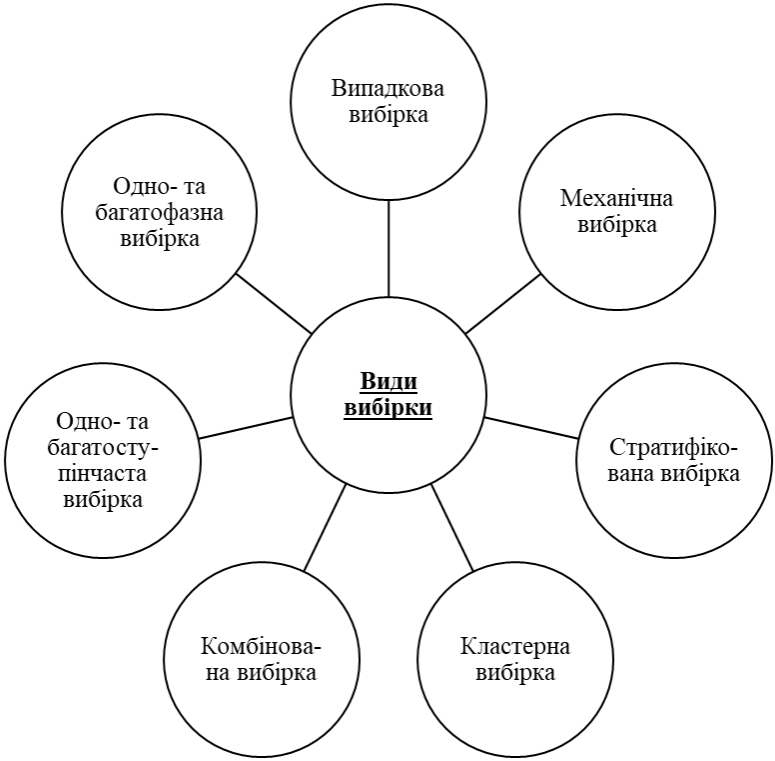


Рис. 6.1. Види вибірки за способом добору одиниць сукупності для обстеження та за кількістю ступенів добору

Випадковою називається вибірка, для якої відбір одиниць із генеральної сукупності є випадковим і проводиться за допомогою жеребкування або таблиці випадкових чисел. Цей спосіб відбору є доцільним за умови однорідності генеральної сукупності, тобто коли між її одиницями немає значних відхилень, а отже, він рідко використовується на практиці.

Найбільш поширеним способом відбору одиниць генеральної сукупності у вибіркову є механічний. ***Механічна вибірка*** являє собою результат послідовного відбору одиниць через рівні проміжки з упорядкованої певним чином сукупності. Інтервал відбору визначається

як $h = \frac{N}{n}$, а перша одиниця вибірки визначається як випадкове число всередині першого інтервалу. Механічний відбір завжди безповторний.

Найчастіше в практиці роботи органів державної статистики використовують стратифіковану вибірку. *Стратифікований відбір* передбачає стратифікацію, тобто поділ генеральної сукупності за певною ознакою на однорідні групи (страти) з подальшим відбором (випадковим або механічним) з кожної виділеної групи визначеної кількості одиниць. Використання стратифікованого відбору доцільно у випадку неоднорідності одиниць досліджуваної сукупності. Крім того, цей спосіб відбору дозволяє отримати більш достовірні дані, адже до вибірки потрапляють представники всіх типових груп.

Простішим в організаційному відношенні й дешевшим є *кластерний відбір*, який передбачає відбір одиниць цілими кластерами – групами або серіями, в межах яких обстежують всі одиниці без винятку. Для відбору серій використовують випадковий або механічний спосіб відбору. Недоліком кластерного способу є те, що похибка вибірки, яка виникає через те, що обстежено не всю генеральну сукупність, а тільки її частину, може бути більшою, ніж за інших способів відбору, а для зменшення цієї похибки необхідно збільшити обсяг вибірки.

Вибірка, при якій із досліджуваної сукупності одразу відбираються одиниці або серії одиниць для безпосереднього обстеження, називається *одноступінчастою*. *Багатоступінчаста вибірка* передбачає спочатку вилучення з генеральної сукупності укрупнених груп одиниць, потім виділення із останніх менших за обсягом груп і так до тих пір, поки не відберуть відповідні групи або окремі одиниці, які й будуть обстежуватися. При багатоступінчастому відборі поєднуються різні способи відбору.

У випадках, коли необхідні дані можна отримати на основі вивчення всіх первинно відібраних одиниць, використовують *однофазну вибірку*. *Багатофазною* називається вибірка, на основі якої проводиться декілька стадій (фаз) вибіркових обстежень, що відрізняються між собою широтою програми обстеження та обсягом вибірки для кожної обстежуваної ознаки; при цьому для всіх фаз зберігається одна й та сама одиниця відбору.

6.3. Оцінювання похибок вибірки та необхідного обсягу вибірки

Вибіркова сукупність не точно відтворює склад генеральної сукупності, тому вибіркові оцінки не збігаються з відповідними характеристиками генеральної сукупності. Різниця між ними визначається похибкою спостереження і похибкою вибірки.

Усі похибки вибіркового спостереження поділяються на похибки вибірки (випадкові); похибки, викликані відхиленням від схеми відбору (систематичні); похибки спостереження (випадкові та невипадкові).

Систематичні похибки можуть виникнути через методологічні аспекти – формулювання запитань, час опитування тощо. Випадкові похибки – ті, які змінюються за законами ймовірності, до них належить похибка вибірки.

Похибка вибірки або, інакше кажучи, *похибка репрезентативності* – це різниця між значеннями показника, отриманого за вибіркою та у генеральній сукупності. Так, похибка репрезентативності вибіркової середньої дорівнює $\varepsilon_{\bar{x}} = \bar{x} - \overline{X}$, вибіркової відносної величини $\varepsilon_d = d - D$, дисперсії $\varepsilon_{\sigma_в^2} = \sigma_в^2 - \sigma^2$.

Характер розподілу похибок вибіркових середніх є нормальним або наближається до нього навіть у випадках, коли генеральна сукупність має іншу форму розподілу.

Ймовірність, яка приймається при розрахунку похибки вибіркової характеристики, називається довірчою. Частіше за все приймають за довірчу ймовірність, що дорівнює 0,95; 0,954; 0,997 або навіть 0,999. Довірчий рівень ймовірності 0,95 означає, що тільки у п'яти випадках зі 100 похибка може вийти за встановлені межі; ймовірності 0,954 – у 46 випадках із 1000; при 0,997 – у трьох випадках, а при 0,999 – в одному випадку з 1000.

Щоб обчислити похибку вибірки за прийнятою довірчою ймовірністю, треба розрахувати величину *середньої* (або *стандартної*) *похибки* μ. Середня похибка вибірки показує розбіжність між вибірковими та генеральними середніми (частками) і визначається для різних методів та способів відбору за формулами, наведеними у табл. 6.2.

Таблиця 6.2

Формули середньої похибки вибіркової середньої і вибіркової відносної величини

Способи відбору	Характеристики сукупності	Повторна вибірка	Безповторна вибірка
Випадкова і механічна вибірки	Середня величина	$\mu_{\bar{x}} = \sqrt{\frac{\sigma_в^2}{n}}$	$\mu_{\bar{x}} = \sqrt{\frac{\sigma_в^2}{n}\left(1 - \frac{n}{N}\right)}$
	Частка	$\mu_d = \sqrt{\frac{d\,(1-d)}{n}}$	$\mu_d = \sqrt{\frac{d\,(1-d)}{n}\left(1 - \frac{n}{N}\right)}$
Стратифікована вибірка	Середня величина	$\mu_{\bar{x}} = \sqrt{\frac{\overline{\sigma}_в^2}{n}}$	$\mu_{\bar{x}} = \sqrt{\frac{\overline{\sigma}_в^2}{n}\left(1 - \frac{n}{N}\right)}$
	Частка	$\mu_d = \sqrt{\frac{\overline{d\,(1-d)}}{n}}$	$\mu_d = \sqrt{\frac{\overline{d\,(1-d)}}{n}\,(1 - \frac{n}{N})}$
Кластерна вибірка	Середня величина	$\mu_{\bar{x}} = \sqrt{\frac{\delta_s^2}{s}}$	$\mu_{\bar{x}} = \sqrt{\frac{\delta_s^2}{s}\left(1 - \frac{s}{S}\right)}$
	Частка	$\mu_d = \sqrt{\frac{d_s(1-d_s)}{s}}$	$\mu_d = \sqrt{\frac{d_s(1-d_s)}{s}\left(1 - \frac{s}{S}\right)}$

Тут $\sigma_в^2 = \frac{\sum (x - \bar{x})^2 \cdot f}{\sum f}$ – вибіркова дисперсія; d – частка одиниць сукупності, які мають певні значення ознаки, визначається за формулою: $d = \frac{m}{n}$, m – число одиниць, які мають певні значення ознаки; $\overline{\sigma}_в^2$ – середня із внутрішньогрупових дисперсій; $\overline{d\,(1-d)}$ – середня з часткових дисперсій за групами; δ_s^2 – межгрупова дисперсія; S – загальне число рівних кластерів (серій, груп) у генеральній сукупності; s – число кластерів (серій, груп), відібраних для обстеження; d_s – середня частка за всіма обстежуваними кластерами (серіями, групами), яка визначається за формулою: $d_s = \frac{\sum d_i}{s}$, d_i – частка в кожному кластері.

Гранична похибка (Δ_x або Δ_p) – це максимально можлива похибка для прийнятої ймовірності, розмір якої для середньої величини і частки відповідно визначається за формулами:

$$\Delta_x = t \cdot \mu_{\bar{x}};$$
$$\Delta_d = t \cdot \mu_d,$$

де t – коефіцієнт довіри, величина якого залежить від рівня ймовірності P.

Величини значень t і P, що відповідають одне одному, наведено в спеціальних таблицях. У практиці статистичної роботи використовуються такі їх значення: $t = 1$ відповідає ймовірність $P = 0{,}683$; $t = 2$ – ймовірність $P = 0{,}954$; $t = 3$ – ймовірність $P = 0{,}997$.

Гранична похибка вибірки є абсолютною величиною. Для порівняння похибок вибірки двох і більше ознак або однієї й тієї самої ознаки у двох і більше вибірках використовують *відносну похибку* V_μ, яка показує, на скільки відсотків вибіркова оцінка відхиляється від параметра генеральної сукупності:

$$V_\mu = \frac{\mu_{\bar{x}}}{\bar{x}} \cdot 100.$$

Відносну похибку вибірки можна розрахувати на основі коефіцієнта варіації ознаки V_x:

– для повторного відбору: $V_\mu = \frac{V_x}{\sqrt{n-1}} \cdot 100$;

– для безповторного відбору: $V_\mu = \frac{V_x}{\sqrt{n-1}} \cdot \sqrt{1 - \frac{n}{N}} \cdot 100$.

Відносну похибку вибірки для частки розраховують за формулою:

$$V_\mu = \frac{\mu_d}{d} \cdot 100.$$

Необхідно зазначити, що за будь-якого способу відбору одиниць розрахунок обсягу вибірки, що проєктується, починають за формулою:

$$n = \frac{t^2 \sigma^2}{\Delta^2},$$

де Δ – похибка, що задається дослідником з огляду на задану точність результатів, для вибірки; t – табличне значення, що відповідає довірчій імовірності, з якою будуть гарантовані оцінки генеральної сукупності за даними вибіркового обстеження; σ^2 – генеральна дисперсія.

Якщо в результаті розрахунку n частка відбору перевищить 5%, то використовують формулу:

$$n = \frac{N t^2 \sigma^2}{N\Delta^2 + t^2 \sigma^2}.$$

Необхідно пам'ятати, що для альтернативних ознак часто використовується максимальне значення дисперсії:

$$\sigma^2 = d(1-d) = 0{,}5 \cdot 0{,}5 = 0{,}25.$$

Якщо в основу розрахунку *n* покласти відносну похибку вибірки, яка враховує ймовірність статистичного висновку, $V_\Delta = t \cdot V_\mu$, то формули набувають такого вигляду:

– для середньої величини: $n = \dfrac{t^2 V_x^2}{V_\Delta^2}$;

– для частки: $n = \dfrac{t^2(1-d)}{V_\Delta^2 d}$.

6.4. Способи поширення даних вибіркового спостереження на генеральну сукупність

Кінцевою метою вибіркового спостереження є характеристика генеральної сукупності на основі даних, отриманих по вибірці. При цьому виходять з того, що всі середні та відносні показники, отримані для вибіркової сукупності, є незміщеними й ефективними характеристиками генеральної сукупності.

Вибіркові середні та відносні величини розповсюджуються на генеральну сукупність обов'язково з урахуванням межі їх можливої похибки. Вибірковий показник приводять із зазначенням межі похибки (довірчої ймовірності): $\overline{x} \pm \Delta_{\overline{x}}$, $d \pm \Delta_d$, або ж вказують межі значень характеристики генеральної сукупності з певною ймовірністю:

$$\overline{x} - \Delta_{\overline{x}} \le \overline{X} \le \overline{x} + \Delta_{\overline{x}};$$
$$d - \Delta_d \le D \le d + \Delta_d.$$

Остання форма запису є основною.

На основі вибірки також можуть бути отримані підсумкові (або загальні) значення показників для генеральної сукупності. Такий розрахунок здійснюється двома способами: шляхом прямого розрахунку і способом коефіцієнтів. Прямий розрахунок використовується у випадках, коли метою вибіркового спостереження є визначення обсягу ознаки в генеральній сукупності, і полягає в тому, що вибіркова середня або частка множиться на об'єм генеральної сукупності:

$$\overline{x}N = \sum_{1}^{N} x_i .$$

Оскільки середня величина містить похибку репрезентативності $\pm \Delta_{\overline{x}}$, то можна вважати, що підсумковий результат у генеральній сукупності варіює в межах:

$$(\overline{x} - \Delta_{\overline{x}})N \le \sum_{i=1}^{N} x_i \le (\overline{x} + \Delta_{\overline{x}})N.$$

Перш ніж провести розрахунок підсумкових показників для генеральної сукупності треба пересвідчитись, що структура вибірки відповідає структурі генеральної сукупності. За наявності значних зміщень у структурі вибірки в частках окремих груп (0,03 і вище) потрібно застосувати метод перезважування, тобто розраховувати генеральну середню на основі вибіркових середніх $\overline{x}_i$ по групах і питомої ваги цих груп у генеральній сукупності w_i:

$$\overline{X} = \overline{x}_1 w_1 + ... + \overline{x}_m w_m ,$$

де $w_i = \dfrac{N_i}{N}$.

Спосіб коефіцієнтів використовується для уточнення результатів суцільного або вибіркового спостереження щодо кількості одиниць сукупності. При застосуванні способу коефіцієнтів також використовуються не тільки вибіркові дані, а й відомості про генеральну сукупність:

$$N_1 = N_0 \frac{n_1}{n_0},$$

де N_0 та N_1 – обсяг сукупності без поправки та з поправкою на недооблік; n_0 та n_1 – обсяг вибірки за первинними й уточненими даними.

При прямому розрахунку береться частка відбору одиниць сукупності, при способі коефіцієнтів – частка відбору за значенням якоїсь ознаки.

Ефективність останнього способу коефіцієнтів порівняно з методом прямого розрахунку залежить від того, наскільки тісно пов'язані між собою ознаки, які лежать в основі розрахунку коефіцієнта, тобто ознака, за якою підраховується підсумок, і ознака, за якою визначається частка відбору. Ефект виявляється, якщо коефіцієнт кореляції між цими ознаками більше за 0,8.

Розв'язання типових задач

Задача 1

За даними про пропозицію до продажу двокімнатних квартир, наведеними в табл. 6.3 (10-відсоткова механічна вибірка), з імовірністю 0,954 визначити похибку вибірки та межі, в яких буде перебувати середнє значення ціни пропозиції для генеральної сукупності.

Таблиця 6.3

Вибіркова сукупність
пропозицій до продажу двокімнатних квартир

Ціна пропозиції, тис. євро	Кількість пропозицій
До 44	3
44–48	5
48–52	8
52 і більше	4
Усього	20

Розв'язання задачі

Варіант 1. Середня ціна пропозиції до продажу двокімнатних квартир перебуває в межах

$$\bar{x} - \Delta_{\bar{x}} \le \overline{X} \le \bar{x} + \Delta_{\bar{x}}.$$

Гранична похибка вибірки при механічному способі відбору визначається за формулою:

$$\Delta_{\bar{x}} = t\sqrt{\frac{\sigma_в^2}{n}\left(1 - \frac{n}{N}\right)},$$

де вибіркова дисперсія $\sigma_в^2$ визначається за формулою:

$$\sigma_в^2 = \frac{\sum (x - \bar{x})^2 \cdot f}{\sum f}.$$

Результати допоміжних розрахунків наведені в табл. 6.4.

Таблиця 6.4

Допоміжна таблиця для розрахунку $\overline{x}$ та $\sigma_в^2$

Ціна пропозиції, тис. євро	Кількість пропозицій, f	Середина інтервалу, x	$x \cdot f$	$(x-\overline{x})^2 \cdot f$
До 44	3	42	42·3=126	$(42-48{,}6)^2$·3=130,68
44–48	5	46	46·5=230	$(46-48{,}6)^2$·5=33,80
48–52	8	50	400	15,68
52 і більше	4	54	216	116,64
Усього	20	–	972	296,80

Середня ціна пропозиції до продажу двокімнатних квартир $\overline{x} = \frac{972}{20} = 48{,}6$ (тис. євро). Тоді вибіркова дисперсія $\sigma_в^2 = \frac{296{,}80}{20} = 14{,}84$, а гранична похибка вибірки $\Delta_{\overline{x}} = 2\sqrt{\frac{14{,}84}{20}(1-0{,}10)} = 1{,}55$ (тис. євро).

Отже, середня ціна пропозиції варіює в межах:

$$48{,}6 - 1{,}55 \le \overline{X} \le 48{,}6 + 1{,}55,\ 47{,}05 \le \overline{X} \le 50{,}15.$$

У 954 випадках з 1000 середня ціна пропозиції до продажу двокімнатних квартир буде не менше 47,05 тис. євро і не перевищить 50,15 тис. євро, лише у 46 випадках з 1000 досліджуваний показник може вийти за встановлені межі.

Варіант 2. Вибіркова дисперсія розраховується за формулою

$$\sigma_в^2 = \overline{x^2} - (\overline{x})^2 = \frac{\sum x^2 f}{\sum f} - \left(\frac{\sum x f}{\sum f}\right)^2.$$

Результати допоміжних розрахунків наведені в табл. 6.5.

Таблиця 6.5

Допоміжна таблиця для розрахунку $\overline{x}$ та $\sigma_в^2$

Ціна пропозиції, тис. євро	Кількість пропозицій, f	Середина інтервалу, x	$x \cdot f$	$x^2 f$
1	2	3	4	5
До 44	3	42	42·3=126	42^2·3=5292
44–48	5	46	46·5=230	46^2·5=10580

Продовження табл.6.5

1	2	3	4	5
48–52	8	50	400	20000
52 і більше	4	54	216	11664
Усього	20	–	972	47536

Середня ціна пропозиції до продажу двокімнатних квартир $\bar{x} = \frac{972}{20} = 48{,}6$ (тис. євро).

Тоді вибіркова дисперсія $\sigma_в^2 = \frac{47536}{20} - \left(\frac{972}{20}\right)^2 = 2376{,}80 - 2361{,}96 = 14{,}84$,

а гранична похибка вибірки $\Delta_{\bar{x}} = 2\sqrt{\frac{14{,}84}{20}(1-0{,}10)} = 1{,}55$ (тис. у. о).

Отже, середня ціна пропозиції перебуває в межах:

$$48{,}6 - 1{,}55 \le \overline{X} \le 48{,}6 + 1{,}55,\ \ 47{,}05 \le \overline{X} \le 50{,}15.$$

Тобто, як і в попередньому варіанті розрахунку, у 954 випадках з 1000 середня ціна пропозиції до продажу двокімнатних квартир буде не менше 47,05 тис. євро і не перевищить 50,15 тис. євро, лише у 46 випадках із 1000 досліджуваний показник може вийти за встановлені межі.

Задача 2

За даними про пропозицію до продажу двокімнатних квартир, наведеними в табл. 6.6 (10-відсоткова механічна вибірка), визначити:

- дисперсію частки пропозиції до продажу двокімнатних квартир у будинках з цегляними стінами;
- похибку цієї частки з ймовірністю 0,997, а також межі, в яких вона перебуває у генеральній сукупності.

Таблиця 6.6

Розподіл пропозицій до продажу двокімнатних квартир за матеріалом стін

Матеріал стіни	Кількість пропозицій
Панельні	13
Цегляні	7
Усього	20

Розв'язання задачі

Частка пропозицій до продажу двокімнатних квартир у будинках з цегляними стінами становить $d_1 = \frac{7}{20} = 0{,}35$, або 35%. Ця частка у генеральній сукупності варіює в межах:

$$d - \Delta_d \le D \le d + \Delta_d .$$

Гранична похибка вибірки при механічному способі відбору визначається за формулою:

$$\Delta_d = t\sqrt{\frac{\sigma_в^2}{n}\left(1 - \frac{n}{N}\right)},$$

де – в\ибіркова дисперсія частки, що визначається за формулою $\sigma_в^2 = d\,(1-d)$.

У нашому випадку дисперсія частки пропозиції до продажу двокімнатних квартир у будинках з цегляними стінами $\sigma_d^2 = 0{,}35\,(1-0{,}35) = 0{,}35 \cdot 0{,}65 = 0{,}228$. Тоді гранична похибка вибірки для частки пропозицій до продажу двокімнатних квартир у будинках з цегляними стінами

$$\Delta_d = 3\sqrt{\frac{0{,}228}{20}\,(1-0{,}10)} = 0{,}304.$$

Отже, частка пропозиції до продажу двокімнатних квартир у будинках з цегляними стінами перебуває в межах

$$0{,}35 - 0{,}304 \le D \le 0{,}35 + 0{,}304,\ \ 0{,}046 \le D \le 0{,}654.$$

У 997 випадках з 1000 аналізована частка буде не менше 4,6% і не перевищить 65,4%, лише у 3 випадках з 1000 досліджуваний показник може вийти за встановлені межі.

Задача 3

Для визначення середньомісячної заробітної плати працівників малих підприємств трьох видів економічної діяльності в регіоні була побудована 17-відсоткова стратифікована (типологічна) вибірка з відбором одиниць пропорційно кількості малих підприємств у групах за видами економічної діяльності. Всередині груп застосовувався механічний відбір. Результати формування вибірки наведені в таблиці 6.7.

Таблиця 6.7

Розподіл малих підприємств
регіону за видом економічної діяльності

Вид економічної діяльності	Кількість підприємств, од.	Середньомісячна заробітна плата, тис. грн			
Освіта	12	3,6	5,1	3,7	4,7
		4,8	4,8	4,5	4,9
		3,7	4,9	4,6	5,7
Охорона здоров'я та надання соціальної допомоги	11	2,4	2,4	4,0	5,7
		4,5	3,7	6,1	3,8
		3,5	5,4	5,2	–
Професійна, наукова та технічна діяльність	25	3,9	9,0	5,9	5,2
		7,0	3,2	3,0	3,6
		5,2	5,1	3,1	4,1
		7,7	3,9	3,6	7,6
		5,7	3,8	3,1	–
		4,2	7,3	4,9	–
		8,6	4,6	3,9	–

З імовірністю 0,954 визначити межі, в яких варіює середньомісячна заробітна плата працівників малих підприємств зазначених видів економічної діяльності.

Розв'язання задачі

Середньомісячна заробітна плата працівників малих підприємств розглянутих видів економічної діяльності перебуває в межах:

$$\bar{x} - \Delta_{\bar{x}} \le \overline{X} \le \bar{x} + \Delta_{\bar{x}}.$$

Гранична похибка вибірки при безповторному випадковому і механічному способах відбору всередині страт (типологічних груп) визначається за формулою:

$$\Delta_{\bar{x}} = t\sqrt{\frac{\overline{\sigma_в^2}}{n}\left(1-\frac{n}{N}\right)},$$

де – вибіркова дисперсія, що визначається за формулою $\overline{\sigma_в^2} = \frac{\sum \sigma_i^2 \cdot n_i}{\sum n_i}$.

Визначимо дисперсії середньої зарплати за кожним видом економічної діяльності за формулою $\sigma_i^2 = \overline{y_i^2} - (\overline{y_i})^2 = \frac{\sum y_i^2}{n_i} - \left(\frac{\sum y_i}{n_i}\right)^2$. Результати допоміжних розрахунків наведені в табл. 6.8.

Таблиця 6.8

Допоміжна таблиця для розрахунку σ_i^2

Вид економічної діяльності	Кількість підприємств, од.	Середньомісячна заробітна плата, тис. грн, y_i	y_i^2
Освіта $\overline{y_i^2} = \frac{\sum y_i^2}{n_i} =$ 256,44 : 12 = 21,37; $\overline{y_i} = \frac{\sum y_i}{n_i} =$ 55,00 : 12 = 4,58	12	3,60	12,96
		4,80	23,04
		3,70	13,69
		5,10	26,01
		4,80	23,04
		4,90	24,01
		3,70	13,69
		4,50	20,25
		4,60	21,16
		4,70	22,09
		4,90	24,01
		5,70	32,49
Разом (1)		55,00	256,44
Охорона здоров'я та надання соціальної допомоги $\overline{y_i^2} = \frac{\sum y_i^2}{n_i} =$ 4,05 : 11 = 19,46 $\overline{y_i} = \frac{\sum y_i}{n_i} =$ 46,70 : 11 = 4,25	11	2,40	5,76
		4,50	20,25
		3,50	12,25
		2,40	5,76
		3,70	13,69
		5,40	29,16
		4,00	16,00
		6,10	37,21
		5,20	27,04
		5,70	32,49
		3,80	14,44

Продовження табл. 6.8

Вид економічної діяльності	Кількість підприємств, од.	Середньомісячна заробітна плата, тис. грн, y_i	y_i^2
Разом (2)		46,70	214,05
Професійна, наукова та технічна діяльність $\overline{y_i^2} = \frac{\sum y_i^2}{n_i} =$ $725{,}80 : 25 = 29{,}03$ $\overline{y_i} = \frac{\sum y_i}{n_i} =$ $127{,}20 : 25 = 5{,}09$	25	3,90	15,21
		7,00	49,00
		5,20	27,04
		7,70	59,29
		5,70	32,49
		4,20	17,64
		8,60	73,96
		9,00	81,00
		3,20	10,24
		5,10	26,01
		3,90	15,21
		3,80	14,44
		7,30	53,29
		4,60	21,16
		5,90	34,81
		3,00	9,00
		3,10	9,61
		3,60	12,96
		3,10	9,61
		4,90	24,01
		3,90	15,21
		5,20	27,04
		3,60	12,96
		4,10	16,81
		7,60	57,76
Разом (3)		127,20	725,80

Дисперсія за групами σ_i^2:

1) освіта – $\sigma_1^2 = 21{,}37 - (4{,}58)^2 = 0{,}39$;

2) охорона здоров'я та соціальна допомога –

$\sigma_2^2 = 19{,}46 - (4{,}25)^2 = 1{,}40$;

3) колективні, громадські та особисті послуги –

$\sigma_3^2 = 29{,}03 - (5{,}09)^2 = 3{,}12$.

Тоді вибіркова дисперсія

$\overline{\sigma_в^2} = \frac{0{,}39 \cdot 12 + 1{,}40 \cdot 11 + 3{,}12 \cdot 25}{48} = \frac{98{,}084}{48} = 2{,}04$, а гранична похибка вибірки $\Delta_{\overline{x}} = 2\sqrt{\frac{2{,}04}{48}(1 - 0{,}17)} = 0{,}38$.

Середньомісячна заробітна плата працівників обстежених малих підприємств:

$$\overline{x} = \frac{4{,}58 \cdot 12 + 4{,}25 \cdot 11 + 5{,}09 \cdot 25}{48} = 4{,}77 \text{ (тис. грн).}$$

Отже, середньомісячна заробітна плата працівників малих підприємств розглянутих видів економічної діяльності перебуває в межах:

$$4{,}77 - 0{,}38 \le \overline{X} \le 4{,}77 + 0{,}38,\ 4{,}39 \le \overline{X} \le 5{,}15.$$

У 954 випадках з 1000 середньомісячна заробітна плата працівників малих підприємств розглянутих видів економічної діяльності буде не менше 4,39 тис. грн і не перевищить 5,15 тис. грн, лише у 46 випадках з 1000 досліджуваний показник може вийти за встановлені межі.

Задача 4

У районі мешкає 2000 сімей. Планується провести їх вибіркове обстеження методом випадкового безповторного відбору з метою визначення середнього розміру сім'ї.

Визначити необхідний обсяг вибірки за умови, що з імовірністю 0,954 похибка вибірки не перевищить однієї особи при середньому квадратичному відхиленні три особи.

Розв'язання задачі

При безповторному випадковому відборі обсяг вибірки визначається за формулою $n = \frac{Nt^2\sigma^2}{N\Delta^2 + t^2\sigma^2}$. Тоді кількість сімей, які необхідно обстежити, становитиме $n = \frac{2000 \cdot 2^2 \cdot 3^2}{2000 \cdot 1^2 + 2^2 \cdot 3^2} = 35$ (сімей).

Запитання для самоконтролю

1. Що називають вибірковим спостереженням?
2. Які переваги має вибіркове спостереження порівняно з іншими видами спостереження?
3. Які приклади практичного використання вибіркового спостереження ви знаєте?
4. Які існують методи відбору одиниць із генеральної сукупності та в чому їх особливість?
5. Які способи формування вибіркової сукупності ви знаєте?
6. Чим відрізняється багатофазна вибірка від багатоступінчастої?
7. Як здійснюють випадковий, механічний та стратифікований відбір?
8. Як визначають середню похибку вибірки для середньої та частки?
9. Як розраховується і для чого використовується гранична похибка вибірки?
10. Як визначається довірчий інтервал для середньої та частки?
11. Як визначають необхідний обсяг вибірки у випадках повторного та безповторного відбору?
12. Як поширюються дані вибіркового спостереження на генеральну сукупність?

Завдання для самостійної роботи

Тестові завдання

1. Вибіркове спостереження – це

а) вид несуцільного спостереження, при якому поширення даних, отриманих за вибірковою сукупністю (вибіркою), на всю досліджувану сукупність має статистико-математичне обґрунтування;

б) вид спостереження, при якому поширення даних, отриманих за вибірковою сукупністю (вибіркою), на всю досліджувану сукупність має статистико-математичне обґрунтування;

в) вид несуцільного спостереження, при якому поширення даних здійснюється на всю досліджувану сукупність.

Відповіді: 1) а; 2) б; 3) в.

2. Вибірка, що складається з випадково відібраних одиниць сукупності, у якій кожна одиниця має однакову ймовірність бути включеною, називається:

а) проста випадкова вибірка;
б) механічна вибірка;
в) стратифікована вибірка;
г) кластерна вибірка.

Відповіді: 1) а; 2) б; 3) в; 4) г.

3. Вибірка, що базується на систематичному відборі з генеральної сукупності (через певний інтервал), називається:

а) проста випадкова вибірка;
б) механічна вибірка;
в) стратифікована вибірка;
г) кластерна вибірка.

Відповіді: 1) а; 2) б; 3) в; 4) г.

4. Вибірка, що передбачає попереднє розшарування генеральної сукупності на однорідні групи, з яких буде проводитись відбір одиниць, називається:

а) проста випадкова вибірка;
б) механічна вибірка;
в) стратифікована вибірка;
г) кластерна вибірка.

Відповіді: 1) а; 2) б; 3) в; 4) г.

5. Вибірка, формування якої передбачає попередній поділ сукупності на групи, які суттєво не відрізняються одна від одної, та їх включення до вибірки з допомогою випадкового або систематичного добору, називається:

а) проста випадкова вибірка;
б) механічна вибірка;
в) стратифікована вибірка;
г) кластерна вибірка.

Відповіді: 1) а; 2) б; 3) в; 4) г.

6. Які з наведених формул використовують для обчислення середньої похибки вибірки за умови повторного відбору:

а) $\sqrt{\frac{\sigma_в^2}{n}\left(1-\frac{n}{N}\right)}$;

б) $\sqrt{\frac{d\,(1-d)}{n}}$;

в) $\sqrt{\frac{\sigma_в^2}{n}}$;

г) $\sqrt{\frac{d\,(1-d)}{n}\left(1-\frac{n}{N}\right)}$.

Відповіді: 1) а, б; 2) б, в; 3) в, г; 4) а, г.

7. Які з наведених формул використовують для обчислення середньої похибки вибірки за умови безповторного відбору:

а) $\sqrt{\frac{\sigma_в^2}{n}\left(1-\frac{n}{N}\right)}$;

б) $\sqrt{\frac{d\,(1-d)}{n}}$;

в) $\sqrt{\frac{\sigma_в^2}{n}}$;

г) $\sqrt{\frac{d\,(1-d)}{n}\left(1-\frac{n}{N}\right)}$.

Відповіді: 1) а, б; 2) б, в; 3) в, г; 4) а, г.

8. Гранична похибка вибірки – це максимально можлива похибка при заданій ймовірності, яка визначається за формулою:

а) $t\cdot\mu$;

б) $\sqrt{\frac{\sigma_в^2}{n}\left(1-\frac{n}{N}\right)}$;

в) $\sqrt{\frac{\sigma_в^2}{n}}$.

Відповіді: 1) а; 2) б; 3) в.

9. Розмір граничної похибки залежить від:

а) тільки варіації ознаки в генеральній сукупності σ^2 та обсягу вибірки n;

б) прийнятого рівня довіри (позначається t);

в) варіації ознаки в генеральній сукупності σ^2, обсягу вибірки n та частки вибірки в генеральній сукупності $\frac{n}{N}$;

г) варіації ознаки в генеральній сукупності σ^2, обсягу вибірки n, частки вибірки в генеральній сукупності $\frac{n}{N}$ та прийнятого рівня довіри t.

Відповіді: 1) а; 2) б; 3) в; 4) г.

10. Межі довірчого інтервалу для середньої визначаються за формулою:

а) $\bar{x} - t\mu_{\bar{x}} \le \overline{X} \le \bar{x} + t\mu_{\bar{x}}$;

б) $d - t\mu_d \le D \le d + t\mu_d$;

в) $\bar{x} - \mu_{\bar{x}} \le \overline{X} \le \bar{x} + \mu_{\bar{x}}$;

г) $d - \mu_d \le D \le d + \mu_d$.

Відповіді: 1) а; 2) б; 3) в; 4) г.

11. Межі довірчого інтервалу для частки визначаються за формулою:

а) $\bar{x} - t\mu_{\bar{x}} \le \overline{X} \le \bar{x} + t\mu_{\bar{x}}$;

б) $d - t\mu_d \le D \le d + t\mu_d$;

в) $\bar{x} - \mu_{\bar{x}} \le \overline{X} \le \bar{x} + \mu_{\bar{x}}$;

г) $d - \mu_d \le D \le d + \mu_d$.

Відповіді: 1) а; 2) б; 3) в; 4) г.

12. Для порівняння похибки вибірки різних ознак або однієї і тієї самої ознаки в різних сукупностях використовують:

а) середню похибку вибірки;

б) відносну похибку вибірки;

в) граничну похибку вибірки.

Відповіді: 1) а; 2) б; 3) в.

13. Відносна похибка частки визначається за формулою:

а) $V_{\mu} = \frac{\mu_{\bar{x}}}{\bar{x}} \cdot 100$;

б) $V_{\mu} = \frac{V_x}{\sqrt{n}} \cdot 100$;

в) $V_{\mu} = \frac{V_x}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}} \cdot 100$;

г) $V_{\mu} = \frac{\mu_d}{d} \cdot 100 = \frac{\sqrt{\frac{d(1-d)}{n}}}{d} = \sqrt{\frac{(1-d)}{nd}}$.

Відповіді: 1) а; 2) б; 3) в; 4) г.

14. Мінімально достатній обсяг вибірки для повторного відбору визначається за формулою:

а) $n = \frac{t^2\sigma^2}{\Delta^2}$.

б) $n = \frac{Nt^2\sigma^2}{N\Delta^2 + t^2\sigma^2}$.

в) $n' = \frac{n}{1 - \frac{n}{N}}$.

Відповіді: 1) а; 2) б; 3) в.

Практичні завдання (задачі)

15. Для вивчення суспільної думки населення регіону щодо ставлення до дій влади було опитано 900 осіб. З числа опитаних 400 осіб позитивно оцінили дії влади.

З імовірністю 0,997 визначте межі, в яких знаходиться частка осіб регіону, що позитивно оцінили дії влади.

16. Для визначення середнього віку робітників підприємства була побудована 20-відсоткова безповторна вибірка. У результаті обстеження отримані такі дані:

Вік робітників, років	до 30	30–40	40–50	50 і старше
Кількість робітників, осіб	20	60	15	5

З імовірністю 0,954 визначте:

1) межі, в яких перебуває середній вік робітників підприємства;
2) межі, в яких перебуває частка робітників підприємства у віці 50 років і старше.

17. У місті з метою визначення середньої тривалості поїздки на роботу передбачається проведення вибіркового обстеження методом випадкового відбору.

Яким має бути обсяг вибірки, щоб з ймовірністю 0,997 похибка вибіркової середньої не перевищувала 10 хвилин при середньому квадратичному відхиленні 40 хвилин?

18. З метою визначення забезпеченості населення міста загальною житловою площею було проведено 5-відсоткове вибіркове обстеження. В результаті отримано такі дані:

Розмір загальної житлової площі на одного члена домогосподарства, м^2	Кількість домогосподарств, % до підсумку
До 10	9
10–12	32
12–14	24
14–16	20
16–18	8
18–20	5
20 і більше	2
Усього	100

З імовірністю 0,954 визначте:

1) межі, в яких перебуває середній розмір загальної житлової площі на одного члена домогосподарства;
2) межі, в яких перебуває частка домогосподарств із розміром загальної житлової площі на одного члена домогосподарства 16 кв. м і більше.

19. Для визначення середнього віку чоловіків, що одружуються, була побудована 5-відсоткова стратифікована вибірка з пропорційним відбором одиниць у кожній страті. Всередині страт застосовувався механічний відбір:

Соціальна група	Кількість чоловіків, осіб	Середній вік, років	Середнє квадратичне відхилення, років	Частка чоловіків, які одружуються вдруге, %
Робітники	60	27	5	10
Службовці	40	31	7	15

З імовірністю 0,954 визначте межі, в яких перебуватиме середній вік чоловіків, що одружуються, та частка чоловіків, які одружуються вдруге.

20. З метою прогнозування врожаю пшениці в господарстві була побудована 10-відсоткова серійна вибірка, до якої потрапили 4 ділянки. В результаті обстеження встановлено, що врожайність пшениці на цих ділянках становила 20, 21, 23 та 25 ц/га.

З імовірністю 0,954 визначте межі, в яких буде варіювати середня врожайність пшениці в господарстві.

21. У місті, де мешкає 10 тис. сімей, передбачається проведення вибіркового обстеження методом механічного відбору з метою визначення частки сімей із дітьми дошкільного віку.

Яким має бути обсяг вибірки, щоб з ймовірністю 0,954 похибка вибірки не перевищувала 0,02, якщо дисперсія дорівнює 0,2?

22. Для виявлення витрат часу на оброблення деталей робітниками різної кваліфікації на підприємстві була сформована 15-відсоткова стратифікована вибірка з пропорційним відбором одиниць у кожній страті із застосуванням всередині страт механічного відбору:

Розряд робітників	Кількість робітників, осіб	Середні витрати часу на оброблення деталей, хвилин	Середнє квадратичне відхилення, хвилин
I	30	10	1
II	50	14	4
III	20	20	2

З імовірністю 0,954 визначте межі, в яких перебуватимуть середні витрати часу на оброблення деталей робітниками.

Розділ 7.
СТАТИСТИЧНІ МЕТОДИ ВИМІРЮВАННЯ ВЗАЄМОЗВ'ЯЗКІВ

7.1. Види взаємозв'язків між явищами

Явища суспільного життя існують неізольовано, вони органічно пов'язані між собою, залежать одні від інших, обумовлюють одне одного і перебувають у постійному русі та розвитку. Розкриваючи взаємозв'язки та взаємозалежності між явищами, можна пізнати їх суть і закони розвитку. Тому вивчення взаємозв'язків є одним з основних завдань статистичного аналізу.

Статистичне дослідження взаємозв'язків не тільки дає можливість виявити наявність і напрям зв'язку, але й дозволяє оцінити його і виразити аналітично. Зв'язки між явищами, їх окремими ознаками різноманітні, при цьому одні ознаки виступають як фактори, що впливають на інші та зумовлюють їх зміну, – факторні ознаки, а інші – як результати дії цих факторів, або результативні ознаки.

Різноманітність зв'язків між явищами та процесами суспільного життя зумовлює необхідність їх класифікації.

1. За характером залежності зв'язки поділяють на функціональні, стохастичні та кореляційні (табл. 7.1).

Таблиця 7.1

Види взаємозв'язків

Факторна ознака, x_i	Результативна ознака y_i за наявністю зв'язку:		
	функціонального	стохастичного	кореляційного
x_1	y_1	$y_1,\ y_2$	$\overline{y_1}$
x_2	y_2	y_1, y_2, y_3	$\overline{y_2}$
x_3	y_3	$y_2,\ y_3,\ y_4$	$\overline{y_3}$
…	…	…	…
x_n	y_n	$y_n - 1,\ y_n$	$\overline{y_n}$

Функціональний зв'язок передбачає, що кожному можливому значенню факторної ознаки (x) відповідає визначене значення результативної ознаки (y), тобто існує повна відповідність між причиною і наслідком, яку можна описати за допомогою математичної формули. *Стохастичні зв'язки* проявляються як узгодженість варіації двох чи більше ознак: кожному значенню факторної ознаки відповідає кілька значень результативної, що утворюють так званий умовний розподіл. Тобто між причиною і наслідком немає повної відповідності. У випадку, коли умовні розподіли замінюються одним параметром – середнім значенням результативної ознаки, то зв'язок називають *кореляційним.* Тобто кореляційний зв'язок проявляється у зміні середніх умовних розподілів і є різновидом стохастичного зв'язку;

2. За напрямом дії розрізняють зв'язки прямі й обернені. *Прямим* називають зв'язок, при якому факторна та результативна ознаки змінюються в одному напряму: при збільшенні значень факторної ознаки збільшується і значення результативної ознаки і навпаки. *Оберненим* називають зв'язок, при якому значення результативної ознаки змінюється в протилежному напрямі до зміни значення факторної ознаки, тобто при збільшенні значень факторної ознаки значення результативної ознаки зменшуються і навпаки.

3. За аналітичним виразом (за формою) зв'язки поділяються на прямолінійні та криволінійні. *Прямолінійним* називають зв'язок, що описується відповідністю однакових змін середніх значень факторної ознаки та приблизно однакових змін середніх значень результативної ознаки. При *криволінійному* зв'язку однаковим змінам середніх значень факторної ознаки відповідають нерівні зміни середніх значень результативної ознаки.

При дослідженні кореляційних зв'язків вирішуються такі завдання:

1) визначається форма зв'язку;
2) вимірюється щільність зв'язку;
3) виявляється вплив окремих факторів на результативну ознаку.

7.2. Метод аналітичних групувань

Суть *методу аналітичних групувань* полягає у групуванні всіх елементів сукупності за факторною ознакою *x* й обчисленні для кожної групи середнього значення результативної ознаки *y*. Порівнюючи паралельні ряди *x* і *y*, можна встановити наявність зв'язку між досліджуваними ознаками.

Вивчення взаємозв'язку між явищами за допомогою методу аналітичного групування здійснюється за такими етапами:

1. Обґрунтування моделі кореляційного зв'язку. На цьому етапі:
   - здійснюється вибір факторної ознаки;
   - визначається число груп;
   - визначаються межі інтервалів;
   - проводиться комбінаційне групування одиниць сукупності за досліджуваними ознаками;
   - будується кореляційна таблиця.
2. Оцінка лінії регресії, для чого:
   - у кожній групі за факторною ознакою обчислюються середні значення результативної та факторної ознак;
   - будується аналітичне групування.

Якщо має місце зростання групових середніх значень результативної ознаки від групи до групи, то це свідчить про наявність прямого кореляційного зв'язку між факторною та результативною ознаками. За оберненого зв'язку, навпаки, має місце зменшення групових середніх значень результативної ознаки від групи до групи.

3. Вимірювання щільності кореляційного зв'язку зважаючи на правило розкладання дисперсій $\sigma^2 = \delta^2 + \overline{\sigma^2}$.

Загальна дисперсія, яка характеризує варіацію результативної ознаки під впливом усіх причин чи умов, може бути обчислена за формулами:

$$\sigma^2 = \overline{y^2} - (\overline{y})^2 \text{ або } \sigma^2 = \frac{\sum (y_i - \overline{y})^2 f}{\sum f}.$$

Міжгрупова дисперсія, що характеризує варіацію результативної ознаки, пов'язану з варіацією групувальної ознаки, обчислюється за формулою:

$$\delta^2 = \frac{\sum (\overline{y}_i - \overline{y})^2 f_i}{\sum f_i},$$

де $\overline{y}_i$ – групові середні значення результативної ознаки; $\overline{y}$ – середнє значення результативної ознаки; f_i – частоти за окремими групами.

Середню з групових дисперсій, яка вимірює варіацію результативної ознаки, пов'язану з впливом усіх факторних ознак, крім покладеної в основу групування, можна обчислити за такою формулою:

$$\overline{\sigma^2} = \sigma^2 - \delta^2 \text{ або } \overline{\sigma^2} = \frac{\sum \sigma_i^2 f_i}{\sum f_i},$$

де σ_i^2 – внутрішньогрупові дисперсії, які обчислюються за формулою:

$$\sigma_i^2 = \frac{\sum (y_i - \overline{y}_i)^2 f_i}{\sum f_i},$$

де y_i – індивідуальні значення результативної ознаки в *i*-й групі; f_i – частоти.

Щоб оцінити щільність зв'язку, слід обчислити співвідношення міжгрупової та загальної дисперсій, тобто *кореляційне відношення* η^2:

$$\eta^2 = \frac{\delta^2}{\sigma^2}.$$

Кореляційне відношення характеризує частку варіації результативної ознаки (y), пов'язану з варіацією групувальної ознаки. Для оцінки щільності зв'язку між факторною та результативною ознаками використовують також η, яке коливається в межах від 0 до 1: $\eta = \sqrt{\frac{\delta^2}{\sigma^2}}$.

Якщо зв'язок відсутній, то $\eta = 0$. Якщо зв'язок функціональний, то $\eta - 1$. Слід відмітити, що якщо $\eta > 0$, то це не може бути доказом наявності кореляційного зв'язку між ознаками. Відмінне від нуля кореляційне відношення може бути наслідком впливу інших чинників.

Для якісної оцінки щільності зв'язку між досліджуваними ознаками y та x на основі η використовують шкалу Чеддока, наведену у табл. 7.2.

Таблиця 7.2

Шкала оцінки щільності зв'язку на основі η

Значення η	0,1–0,3	0,3–0,5	0,5–0,7	0,7–0,9	0,9–0,99
Зв'язок	Слабкий	Помірний	Помітний	Сильний	Дуже сильний

Перевірка істотності відхилень групових середніх здійснюється з допомогою критеріїв, розроблених математичною статистикою. Слід застосувати критичні значення η^2 або критичні значення F-критерію (критерію Фішера).

Фактичні значення F-критерію обчислюють за формулами:

$$F = \frac{\eta^2}{1-\eta^2} \cdot \frac{k_2}{k_1} \text{ або } F = \frac{\delta^2}{\sigma^2} \cdot \frac{k_2}{k_1},$$

де k_1, k_2 – число ступенів вільності; $k_1 = m - 1$, m – число груп; $k_2 = n - m$, n – число одиниць сукупності.

Фактичні значення η^2 і F-критерію необхідно порівняти з критичними для рівнів істотності $\alpha = 0{,}05$ або $\alpha = 0{,}01$. Якщо фактичні значення η^2 і F-критерію перевищують відповідні критичні, то зв'язок між ознаками визнається істотним. Якщо фактичні значення η^2 і F-критерію менше відповідних критичних, то висновок є невизначеним, а наявність або відсутність зв'язку – не доведеною.

7.3. Основи кореляційно-регресійного аналізу

Основне завдання кореляційно-регресійного методу полягає в аналізі статистичних даних з метою виявлення залежності між досліджуваними ознаками у вигляді певної математичної формули та встановлення за допомогою коефіцієнта кореляції порівняльної оцінки щільності зв'язку.

Кореляційно-регресійний метод аналізу використовується для розв'язання двох основних завдань:

1) визначає за допомогою рівнянь регресії аналітичну форму зв'язку між варіацією ознак x і y;
2) встановлює ступінь щільності зв'язку між ознаками.

В основі кореляційно-регресійного аналізу лежить припущення, що залежність між факторною і результативною ознаками може бути виражена функцією $Y = f(x)$, яка називається рівнянням регресії. За аналітичним виразом залежність може бути лінійною і нелінійною. Найбільш поширені такі рівняння регресії:

- $Y - \alpha + bx$ – лінійне;
- $Y = ab^x$ – показникове;
- $Y = ax^b$ – степеневе;
- $Y = a + bx + cx^2$ – параболічне;
- $Y = a + \frac{b}{x}$ – гіперболічне,

де Y – теоретичні значення результативної ознаки; a, b і c – параметри рівняння регресії (b і c називають коефіцієнтами регресії).

На першому етапі кореляційно-регресійного аналізу при обґрунтуванні моделі, як і в аналітичному групуванні, вирішують два питання: вибір факторної і результативної ознаки та вибір виду рівняння регресії.

Правильний вибір ознак і виду рівняння регресії потребує теоретичного аналізу взаємозв'язку. Для вибору виду рівняння регресії часто застосовується графічне зображення зв'язку у вигляді кореляційного поля (рис. 7.1).

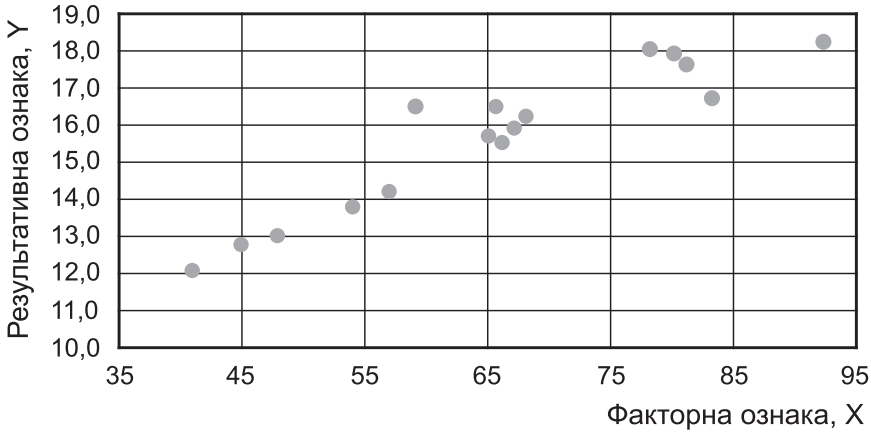


Рис. 7.1. Кореляційне поле залежності результативної ознаки від факторної

При його побудові на осі абсцис відкладають значення факторної ознаки *x*, а на осі ординат – результативної ознаки *y*. Кожній одиниці сукупності на графіку відповідає окрема точка. За формою розміщення точок на кореляційному полі робиться висновок щодо виду регресійного рівняння.

При великому обсязі сукупності доцільно на графіку зображати групові середні попередньо побудованого аналітичного групування. Лінію групових середніх називають емпіричною лінією регресії.

Для визначення виду рівняння регресії застосовується також спосіб перебору функцій, коли будують рівняння регресії різних видів і з них на основі статистико-математичних критеріїв вибирають найкраще.

У практиці економіко-статистичних досліджень часто доводиться мати справу з прямолінійною формою зв'язку, яка виражається *рівнянням регресії*:

$$Y = \alpha + bx,$$

де *b* – коефіцієнт регресії, який показує, на скільки одиниць власного виміру змінюється середнє значення результативної ознаки зі збільшенням факторної ознаки на одиницю її власного виміру; *a* – теоретичне значення *Y* для $x = 0$, якщо 0 перебуває в межах фактичної варіації ознаки *x* (інакше параметр *a* не має реального змісту).

Параметри *a* та *b* визначають методом найменших квадратів щодо складеної і розв'язаної системи нормальних рівнянь:

$$\begin{cases} \sum y = na + b\sum x \\ \sum yx = a\sum x + b\sum x^2 \end{cases},$$

де *n* – число членів за кожною ознакою.

Розв'язавши зазначену систему рівнянь, визначимо параметри *a* та *b*:

$$a = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - (\sum x)^2},$$

$$b = \frac{\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}.$$

Використовуючи регресійну модель, можна визначити відносний ефект впливу факторної ознаки *x* на результативну ознаку *y* за допо-

могою коефіцієнта еластичності. *Коефіцієнт еластичності* показує, на скільки відсотків зміниться в середньому результативна ознака *y* при зміні факторної ознаки *x* на 1%. Цей коефіцієнт може бути розрахований за кожним досліджуваним об'єктом сукупності (ε_i), а також у середньому по сукупності (ε) за формулами:

$$\varepsilon_i = b\frac{x_i}{y_i} \text{ та } \varepsilon = b\frac{\overline{x}}{\overline{y}}.$$

Визначення щільності зв'язку в кореляційно-регресійному аналізі, як і в методі аналітичного групування, базується на правилі розкладання дисперсій, але в цьому випадку для оцінки лінії регресії використовують не групові середні значення результативної ознаки, а теоретичні значення *Y*.

Щільність кореляційного зв'язку вимірюється шляхом обчислення коефіцієнта детермінації, індексу кореляції та лінійного коефіцієнта кореляції.

Мірою щільності зв'язку – як лінійного, так і нелінійного – є *коефіцієнт детермінації* R^2 :

$$R^2 = \frac{\delta_Y^2}{\sigma^2},$$

де δ_Y^2 – факторна дисперсія, що характеризує варіацію результативної ознаки, спричинену дією факторної ознаки *x*, визначається за формулами:

$$\delta_Y^2 = \frac{\sum (Y-\overline{y})^2}{n} \text{ або } \delta_Y^2 = \frac{a\sum x + b\sum xy}{n} - (\overline{y})^2;$$

– загальна дисперсія, що розраховується за формулами:

$$\sigma^2 = \frac{\sum (y-\overline{y})^2}{n} \text{ або } \sigma^2 = \delta_Y^2 + \sigma_e^2;$$

σ_e^2 – залишкова дисперсія, обчислюється за формулою:

$$\sigma_e^2 = \frac{\sum (y-Y)^2}{n}.$$

Коефіцієнт детермінації приймає значення від 0 до 1 і характеризує частку варіацій результативної ознаки, пов'язану з факторною ознакою при відповідній формі зв'язку.

Корінь квадратний з коефіцієнта детермінації називають *індексом кореляції R:*

$$R = \sqrt{R^2}.$$

Індекс кореляції приймає значення в межах від 0 до 1. Коли R= 0, то зв'язку між варіацією ознак *y* та *x* немає. При R = 1 зв'язок між досліджуваними ознаками лінійно-функціональний. Індекс кореляції оцінює щільність зв'язку, але не вказує його напрям.

Щільність та напрям зв'язку можна виміряти з допомогою *лінійного коефіцієнта кореляції Пірсона r*:

$$\text{або } r = b\frac{\sigma_x}{\sigma_y},$$

де $\overline{xy} = \frac{\sum xy}{n}$; $\overline{x}$ і $\overline{y}$ – середні значення факторної і результативної ознак; σ_x і σ_y – середні квадратичні відхилення відповідних ознак.

Цей показник коливається в межах від –1 до +1: додатне значення свідчить про прямий зв'язок між досліджуваними ознаками, від'ємне – про обернений.

Необхідно знати, що при лінійній формі зв'язку абсолютна величина лінійного коефіцієнта кореляції дорівнює індексу кореляції, тобто $|r| = R$.

Оцінити ступінь щільності зв'язку на основі лінійного коефіцієнта кореляції можна за допомогою градації, наведеної у табл. 7.3, а також за допомогою шкали Чеддока (табл. 7.2).

Таблиця 7.3

Оцінка щільності зв'язку на основі лінійного коефіцієнта кореляції

Зв'язок	Значення лінійного коефіцієнта кореляції *r*	
	Прямий зв'язок	Обернений зв'язок
Слабкий	0,10–0,30	(–0,10)–(–0,30)
Середній	0,30–0,70	(–0,30)– (–0,70)
Щільний	0,70–0,99	(–0,70)–(–0,99)

У кореляційно-регресійному аналізі істотність зв'язку перевіряється за допомогою R^2 чи *F*-критерію, фактичне значення якого в цьому випадку обчислюється за формулою:

$$F = \frac{R^2}{1-R^2} \cdot \frac{k_2}{k_1},$$

де k_1, k_2 – число ступенів вільності.

При визначенні числа ступенів вільності дійсні ті самі формули, що і в аналітичному групуванні, але *m* позначає число параметрів у регресійному рівнянні.

7.4. Непараметричні методи вимірювання взаємозв'язків

У статистичній практиці широко використовуються також непараметричні методи вимірювання взаємозв'язків, які не потребують використання кількісних значень ознак і обчислення параметрів їх розподілів. Непараметричні методи вимірювання взаємозв'язків можуть використовуватись як для атрибутивних, так і для кількісних ознак.

Аналіз взаємозв'язків може проводитись на основі таблиць взаємної спряженості (співзалежності) – комбінаційних розподілів сукупностей за двома ознаками: факторною *x* та результативною *y*. Оцінка щільності зв'язку ґрунтується на відхиленнях фактичних частот f_{ij} від теоретичних F_{ij}, пропорційних до підсумкових:

$$F_{ij} = \frac{f_{i0}\, f_{0j}}{n},$$

де f_{i0} – підсумкові частоти за факторною ознакою; f_{0j} – підсумкові частоти за результативною ознакою; *n* – обсяг сукупності.

Абсолютну величину відхилень ($f_{ij} - F_{ij}$) характеризує квадратична спряженість Пірсона χ^2 (хі-квадрат):

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - F_{ij})^2}{F_{ij}} = n\left[\sum_i \sum_j \frac{f_{ij}^2}{f_{i0}\, f_{0j}} - 1\right].$$

Якщо $\chi^2 = 0$, то зв'язок між ознаками відсутній. Для перевірки істотності зв'язку фактичне значення χ^2 порівнюється з критичним для заданої ймовірності $1 - \alpha$ та числа ступенів свободи $k = (m_x - 1)(m_y - 1)$, де 1- m_x та m_y – кількість груп за факторною та результативною ознаками відповідно. Критичні значення χ^2 наведені у табл. 7.4.

Таблиця 7.4

Критичні значення χ^2 при $\alpha = 0{,}5$

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1-α	3,84	5,99	7,82	9,49	11,07	12,59	14,07	15,51	16,92	18,31	19,68	21,03	22,36	23,69	25,00

Відносною мірою щільності зв'язку є коефіцієнти взаємної спряженості *C*, які за змістом ідентичні коефіцієнтам кореляції і змінюються в межах від 0 до 1.

Коефіцієнт взаємної спряженості Пірсона розраховується за формулою:

$$C_{П}=\sqrt{\frac{\varphi^2}{1+\varphi}}=\sqrt{\frac{\chi^2}{n+\chi^2}},$$

де $\varphi^2=\frac{\chi^2}{n}$.

Але результат, отриманий за цією формулою, не враховує залежність коефіцієнта спряженості від кількості груп за кожною ознакою.

У випадках, коли $m_x=m_y$, використовують коефіцієнт спряженості Чупрова:

$$C_{ч}=\sqrt{\frac{\chi^2}{n\sqrt{(m_x-1)\,(m_y-1)}}},$$

а якщо $m_x\neq m_y$, то коефіцієнт спряженості Крамера:

$$C_K=\sqrt{\frac{\chi^2}{n\sqrt{(m_{\min}-1)}}},$$

де $m_{\min}$ – мінімальна кількість груп за факторною чи результативною ознакою.

Для оцінки щільності зв'язку між двома ознаками, які мають альтернативне вираження, розраховують коефіцієнт асоціації та колігації Юла, коефіцієнт контингенції Пірсона з допомогою чотириклітинної таблиці (табл. 7.5), в якій кожна клітина відповідає певній альтернативі тієї чи іншої ознаки. Альтернативною називається ознака, яка має тільки два варіанти значень.

Таблиця 7.5

Макет чотириклітинної таблиці для оцінки взаємозв'язків між альтернативними ознаками

Ознака	*A*	*не A*	Разом
B	*a*	*b*	$(a+b)$
не B	*c*	*d*	$(c+d)$
Разом	$(a+c)$	$(b+d)$	$a+b+c+d$

де *A* та *B* – значення ознак, зв'язок між якими вивчають; *не A* та *не B* – відповідні альтернативні значення ознак; *b, c, d* – частоти відповідних комбінацій ознак для досліджуваних одиниць сукупності

Коефіцієнт асоціації, запропонований Е. Дж. Юлом, розраховується за формулою:

$$K_{ac} = \frac{ad - bc}{ad + bc},$$

де *a, b, c, d* – частоти таблиці.

Коефіцієнт колігації, також запропонований Е. Дж. Юлом, використовується для дослідження кореляції альтернативних ознак і визначається за формулою

$$K_{кол} = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}.$$

Однак якщо у однієї з клітинок таблиці не буде значення, то показник зв'язку буде помилково збільшеним. Тому використовують *коефіцієнт контингенції:*

$$K_{конт} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}.$$

Зв'язок вважається істотним, якщо :

$$|K_{ac}| > 0{,}5 \text{ або } |K_{kont}| > 0{,}3.$$

Значення коефіцієнтів колігації та контингенції як показників щільності зв'язку інтерпретуються так само, як і коефіцієнта кореляції. Причому завжди $K_{ac} > K_{кол} > K_{конт}$.

Серед поширених методів аналізу взаємозв'язків виділяють *метод порівняння паралельних рядів*, який передбачає розташування матеріалів статистичного спостереження ранжованими паралельними рядами за факторною та результативною ознаками, що дає можливість, порівнюючи їх, простежити співвідношення, виявити зв'язок та його напрям. На основі порівняння паралельних рядів напрям та щільність зв'язку між досліджуваними ознаками оцінюють за допомогою коефіцієнтів Фехнера та рангової кореляції Спірмена.

Коефіцієнт Фехнера оцінює щільність зв'язку на підставі порівняння знаків відхилень окремих значень від середнього значення за кожною ознакою: знак «–» свідчить, що значення ознаки менше за середнє, а знак «+» – що значення ознаки більше за середнє. Якщо

знаки за обома ознаками збігаються (*З*), то варіація ознак узгоджена, незбігання знаків (*Н*) вказує на порушення узгодженості варіації досліджуваних ознак. Коефіцієнт Фехнера розраховується за формулою

$$K_{\Phi} = \frac{\sum З - \sum H}{\sum З + \sum H},$$

де $\sum З$ – сума знаків відхилень, що збіглися в обох рядах; $\sum H$ – сума знаків відхилень, що не збіглися в обох рядах.

Коефіцієнт Фехнера набуває значення в межах $-1 \le K_{\Phi} \le 1$. Наближення значення K_{Φ} до +1 свідчить про сильну пряму узгодженість, а наближення до (–1) – про сильну обернену узгодженість між досліджуваними ознаками. Якщо $K_{\Phi} = 0$, узгодженість між досліджуваними ознаками відсутня.

Для оцінки щільності зв'язку між ознаками порядкової (рангової) шкали використовують *коефіцієнт рангової кореляції Спірмена* ρ, який змінюється в межах від –1 до +1 і за змістом відповідає лінійному коефіцієнту кореляції, тобто одночасно оцінює щільність зв'язку та вказує його напрям:

$$\rho = 1 - \frac{6\sum_{1}^{n} d^2}{n(n^2-1)},$$

де $d = R_x - R_y$ – відхилення рангів факторної (R_x) та результативної (R_y) ознак; n – кількість рангів.

Ранги – це порядкові номери одиниць сукупності у ранжованому ряді. Так, якщо проранжувати сукупність за двома ознаками, зв'язок між якими вивчається, то повне збігання рангів означає максимально щільний прямий зв'язок, а повна протилежність рангів – максимально щільний обернений зв'язок. Ранжувати обидва показника, зв'язок між якими досліджується, необхідно в одному порядку – від менших значень ознаки до більших (або навпаки). У разі збігання значення ознаки у різних одиниць сукупності їм присвоюється об'єднаний середній порядковий номер.

Коефіцієнт рангової кореляції Спірмена можна використовувати у випадках, коли ознаки не мають кількісного вираження, але одиниці сукупності можна впорядкувати за значенням ознаки.

Розв'язання типових задач

Задача 1

Маємо дані про вартість оборотних коштів і прибуток ***де***сяти підприємств (табл. 7.6).

Таблиця 7.6

Показники діяльності підприємств

№ підприємства	Вартість оборотних коштів, млн грн	Прибуток, млн грн
1	7	3
2	11	4
3	12	4
4	16	6
5	22	8
6	27	7
7	30	11
8	38	14
9	45	18
10	46	17

Визначити міжгрупову, середню з групових та загальну дисперсії прибутку. Розрахувати показники щільності зв'язку між вартістю оборотних коштів та прибутком підприємств. Перевірити істотність зв'язку з ймовірністю 0,95. Зробіть висновки.

Розв'язання задачі

На першому етапі проводиться комбінаційне групування підприємств за двома ознаками: за вартістю оборотних коштів і розміром одержаного прибутку, утворивши по три групи з рівними інтервалами за даними табл. 7.6. Для цього:

1) визначаємо ширину інтервалів.

Вартість оборотних коштів (млн грн):

$$h_x = \frac{x_{\max} - x_{\min}}{n} = \frac{46 - 7}{3} = 13.$$

Прибуток (млн грн):

$$h_y = \frac{y_{\max} - y_{\min}}{n} = \frac{18 - 3}{3} = 5.$$

2) утворюємо групи підприємств за кожною ознакою, використовуючи ширину інтервалу:

За вартістю оборотних коштів (млн грн):	За прибутком (млн грн):
– від 7 до 20 (20=7+13)	– від 3 до 8 (8=3+5)
– від 20 до 33 (33=20+13)	– від 8 до 13 (13=8+5)
– від 33 до 46 (46=33+13)	– від 13 до 18 (18=13+5)

Таблиця 7.7

Розподіл підприємств за вартістю оборотних коштів та розміром прибутку

Групи підприємств за вартістю оборотних коштів, млн грн	Групи підприємств за розміром прибутку, млн грн			Разом
	3–8	8–13	13–18	
7–20	4			4
20–33	2	1		3
33–46			3	3
Разом	6	1	3	10

Побудовану у такий спосіб таблицю називають кореляційною. Якщо частоти у кореляційній таблиці розташовані по діагоналі з лівого верхнього кута у правий нижній кут, тобто більшим значенням фактора відповідають більші значення функції, то передбачається наявність прямого кореляційного зв'язку між ознаками. Якщо ж частоти розташовані по діагоналі з правого кута у лівий, то передбачають наявність оберненого зв'язку між ознаками. Табл. 7.7 показує, що частоти розподілу підприємств концентруються по діагоналі, яка йде з верхнього лівого кута у правий нижній. Це свідчить про те, що існує прямий зв'язок між фактором (вартістю оборотних коштів) і розміром прибутку підприємств, який збільшується зі зростанням вартості оборотних коштів.

На другому етапі проводиться оцінка лінії регресії, для чого у кожній групі за факторною ознакою обчислюють середні значення результативної та факторної ознак. Групові середні обчислюють за вихідними незгрупованими даними. У нашому прикладі:

Середній прибуток, млн грн

$$\bar{y}_1 = \frac{3+4+4+6}{4} = 4{,}25;$$

$$\bar{y}_2 = \frac{8+7+11}{3} = 8{,}67;$$

$$\bar{y}_3 = \frac{14+18+17}{3} = 16{,}33.$$

Середня вартість оборотних коштів, млн грн

$$\bar{x}_1 = \frac{7+11+12+16}{4} = 11{,}50;$$

$$\bar{x}_2 = \frac{22+27+30}{3} = 26{,}33;$$

$$\bar{x}_3 = \frac{38+45+46}{3} = 43{,}00.$$

У цілому по досліджуваній сукупності:

Середній прибуток, млн грн

$$\bar{y} = \frac{4{,}25 \cdot 4 + 8{,}67 \cdot 3 + 16{,}33 \cdot 3}{4+3+3} = 9{,}20.$$

Середня вартість оборотних коштів, млн грн

$$\bar{x} = \frac{11{,}50 \cdot 4 + 26{,}33 \cdot 3 + 43{,}00 \cdot 3}{4+3+3} = 25{,}40.$$

Результати розрахунків оформляємо у вигляді табл. 7.8.

Таблиця 7.8

Залежність прибутку від вартості оборотних коштів

Групи підприємств за вартістю оборотних коштів, млн грн	Кількість підприємств, од., f	Середній прибуток, млн грн, у $\bar{y}_i$	Середня вартість оборотних коштів, млн грн, $\bar{x}_i$
7–20	4	4,25	11,50
20–33	3	8,67	26,33
33–46	3	16,33	43,00
Всього	10	9,20	25,40

Зростання групових середніх прибутку від групи до групи свідчить про наявність кореляційного зв'язку між вартістю оборотних коштів і розміром прибутку.

Третій етап аналітичного групування – вимірювання щільності зв'язку з використанням дисперсійного аналізу. Розрахунки, необхідні для визначення міжгрупової дисперсії, наведені в табл. 7.9.

Таблиця 7.9

Допоміжні розрахунки для міжгрупової дисперсії

Групи підприємств за вартістю оборотних коштів, млн грн	Кількість підприємств, од., f	Середній прибуток, млн грн, $\overline{y_i}$	$(\overline{y_i} - \overline{y})^2 \cdot f$
7–20	4	4,25	$(4,25 - 9,20)^2 \cdot 4 = 98,01$
20–33	3	8,67	$(8,67 - 9,20)^2 \cdot 3 = 0,84$
33–46	3	16,33	$(16,33 - 9,20)^2 \cdot 3 = 152,51$
Разом	10	9,20	251,36

Тоді міжгрупова дисперсія:

$$\delta^2 = \frac{\sum (\overline{y_i} - \overline{y})^2 f}{\sum f} = \frac{251,36}{10} = 25,136 .$$

Для обчислення середньої з групових (часткових) дисперсій необхідно спочатку визначити дисперсії прибутку для кожної групи підприємств за вартістю оборотних коштів за формулою . Результати допоміжних розрахунків наведені в табл. 7.10.

Таблиця 7.10

Допоміжна таблиця для розрахунку σ_i^2

Групи підприємств за вартістю оборотних коштів, млн грн	Кількість підприємств, од., f	Прибуток, млн грн, y_i	y_i^2	$\overline{y_i^2} = \frac{\sum y_i^2}{n_i}$
1	2	3	4	5
7–20	4	3	9	77:4 = 19,250
		4	16	
		4	16	
		6	36	
Усього (1)		17	77	–
20–33	3	8	64	234:3 = 78,000
		7	49	
		11	121	

Продовження табл.7.10

1	2	3	4	5
Усього (2)		26	234	–
33–46	3	14	196	809:3 = 269,667
		18	324	
		17	289	
Усього (3)		49	809	–

Дисперсія σ_i^2 :

$\sigma_i^2 = 19{,}250 - (4{,}25)^2 = 1{,}188;$

$\sigma_2^2 = 7\,8{,}000 - (8{,}67)^2 = 2{,}889;$

$\sigma_3^2 = 269{,}667 - (16{,}33)^2 = 2{,}889.$

Тоді середня з групових дисперсія становить:

$$\overline{\sigma^2} = \frac{1{,}188 \cdot 4 + 2{,}889 \cdot 3 + 2{,}889 \cdot 3}{10} = 2{,}208,$$

а загальна дисперсія:

$$\sigma^2 = \delta^2 + \overline{\sigma^2} = 25{,}136 + 2{,}208 = 27{,}344.$$

Кореляційне відношення η^2 показує ступінь варіації ознаки під впливом фактора, покладеного в основу групування і визначається як відношення міжгрупової дисперсії до загальної.

$$\eta^2 = \frac{\delta^2}{\sigma^2} = \frac{25{,}136}{27{,}344} = 0{,}919.$$

Кореляційне відношення показує, що варіація прибутку підприємств на 91,9% залежить від варіації вартості оборотних коштів і на 9,1% – від інших факторів.

Мірою щільності зв'язку між факторною і результативною ознаками є η, яке визначається за формулою:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{0{,}919} = 0{,}959.$$

Отже, зв'язок між прибутком підприємств і вартістю оборотних коштів дуже сильний (див. шкалу оцінки щільності зв'язку на основі η, наведену у табл. 7.2).

Перевірка істотності відхилень групових середніх здійснюється з допомогою критерію Фішера (*F*-критерій) за формулою:

$$F = \frac{\delta^2}{\overline{\sigma_i^2}} \cdot \frac{k_2}{k_1} \quad \text{або} \quad F = \frac{\eta^2}{1-\eta^2} \cdot \frac{k_2}{k_1}.$$

За даними нашого прикладу, число ступенів вільності для міжгрупової дисперсії $k_1 = m - 1 = 3 - 1 = 2$, для середньої з групових – $k_2 = n - m = 10 - 3 = 7$.

Тоді

$$F = \frac{\delta^2}{\overline{\sigma_i^2}} \cdot \frac{k_2}{k_1} = \frac{25{,}136}{2{,}208} \cdot \frac{7}{2} = 39{,}84 \quad \text{або} \quad F = \frac{\eta^2}{1-\eta^2} \cdot \frac{k_2}{k_1} = \frac{0{,}919}{1-0{,}919} \cdot \frac{7}{2} = 39{,}71.$$

Для оцінки отриманого відношення його порівнюють з критичним (табличним). Для ймовірності 0,95 і числа ступенів вільності $k_1 = 2$, $k_2 = 7$ значення $F_{0,95}(2; 7) = 4{,}74$ (табл. 7.11). Отже, $F > F_{кр}$, що свідчить про суттєвий вплив вартості оборотних коштів підприємств на їх прибуток.

Таблиця 7.11

Критичні значення *F*-критерію (рівень істотності $\alpha = 0{,}050$

k_2	k_1								
	1	2	3	4	5	6	8	10	20
1	161,4	199,5	215,7	224,6	230,2	243,0	238,9	242,0	248,0
2	18,54	19,00	19,16	19,25	19,30	19,33	19,37	19,39	19,44
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,78	8,66
4	7,71	6,04	6,59	6,36	6,26	6,16	6,04	5,96	5,80
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,63	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,34	3,15
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,13	2,93
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,97	2,77
11	4,82	3,98	3,59	3,36	3,20	3,09	2,95	2,86	2,65
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,76	2,54
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,60	2,39
16	4,49	3,63	3,24	3,01	2,85	5,74	2,59	2,49	2,28

Продовження табл.7.11

k_2	k_1								
	1	2	3	4	5	6	8	10	20
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,41	2,19
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35	2,12
30	4,17	3,32	,92	2,69	2,53	2,42	2,27	2,16	1,93
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,12	1,84
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	2,04	1,75
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,90	1,65
¥	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,83	1,57

Задача 2

Використовуючи дані про вартість оборотних коштів і прибуток десяти підприємств (табл. 7.6) побудувати однофакторну регресійну модель залежності між вартістю оборотних коштів та прибутком підприємств. Побудувати графік кореляційного поля. Визначити коефіцієнти еластичності; показники щільності кореляційного зв'язку. Перевірити істотність з імовірністю 0,95.

Розв'язання задачі

Для вибору виду форми рівняння регресії використаємо кореляційне поле (рис. 7.2).

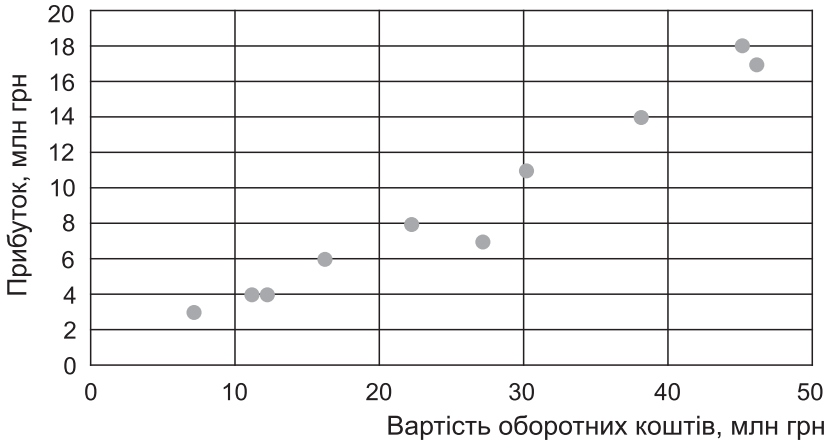


Рис. 7.2. Кореляційне поле зв'язку між вартістю оборотних коштів і прибутком підприємств

При його побудові на осі абсцис відкладаємо значення факторної ознаки x (вартість оборотних коштів), а на осі ординат – результативної ознаки y (прибуток).

За формою розміщення точок на кореляційному полі обираємо лінійне рівняння регресії. Розрахунки показників, необхідних для обчислення параметрів рівняння регресії, представлено у табл. 7.12.

Таблиця 7.12

Розрахункові показники
для складання системи нормальних рівнянь

№ підприємства	Вартість оборотних коштів, млн грн, x	Прибуток, млн грн, y	x^2	y^2	xy	Y
А	1	2	3	4	5	6
1	7	3	7·7=49	3·3=9	7·3=21	–0,4655+ 0,3805·7=2,20
2	11	4	11·11=121	4·4=16	11·4=44	3,72
3	12	4	144	16	48	4,10
4	16	6	256	36	96	5,62
5	22	8	484	64	176	7,91
6	27	7	729	49	189	9,81
7	30	11	900	121	330	10,95
8	38	14	1444	196	532	13,99
9	45	18	2025	324	810	16,66
10	46	17	2116	289	782	17,04
Усього	254	92	8268	1120	3028	92

Параметри a та b визначимо за формулами:

$$a = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - (\sum x)^2} = \frac{8268 \cdot 92 - 254 \cdot 3028}{10 \cdot 8268 - 254 \cdot 254} = -0{,}4655;$$

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{10 \cdot 3028 - 254 \cdot 92}{10 \cdot 8268 - 254 \cdot 254} = 0{,}3805.$$

Рівняння кореляційного зв'язку має вигляд $Y = -0{,}4655 + 0{,}3805x$, тобто зв'язок між досліджуваними показниками є прямим і при збільшенні вартості оборотних коштів на 1 млн грн прибуток підприємств у середньому збільшується на 0,3805 млн грн (або 380,5 тис. грн).

Після визначення параметрів рівняння регресії визначаємо теоретичну лінію регресії шляхом підстановки у рівняння значень факторної ознаки (див. графу 6 табл. 7.12).

Розрахуємо коефіцієнт еластичності:

$$\varepsilon = b\frac{\overline{x}}{\overline{y}} = 0{,}3805\,\frac{25{,}4}{9{,}2} = 1{,}051\ ;$$

$$\overline{x} = 254/10 = 25{,}4 \text{ (млн грн)}, \quad \overline{y} = 92/10 = 9{,}2 \text{ (млн грн)}.$$

Коефіцієнт еластичності показує, що при зростанні вартості оборотних коштів на 1% у середньому прибуток підприємств збільшиться на 1,051%.

Щільність кореляційного зв'язку вимірюється шляхом обчислення коефіцієнта детермінації, індексу кореляції та лінійного коефіцієнта кореляції. Для визначення цих показників здійснимо додаткові розрахунки (табл. 7.13).

Таблиця 7.13

Розрахункові показники
для розрахунку показників щільності зв'язку

№ підприємства	Вартість оборотних коштів, млн грн, x	Прибуток, млн грн, y	Y	$(Y-\overline{y})^2$	$(y-\overline{y})^2$	$(y-Y)^2$	$(x-\overline{x})^2$
1	2	3	4	5	6	7	8
1	7	3	2,20	$(2{,}20-9{,}20)^2=$ 49,000	$(3-9{,}20)^2=$ 38,440	$(3-2{,}20)^2=$ 0,640	$(7-25{,}4)^2=$ 338,560
2	11	4	3,72	30,030	27,040	0,078	207,360
3	12	4	4,10	26,010	27,040	0,010	179,560
4	16	6	5,62	12,816	10,240	0,144	88,360
5	22	8	7,91	1,664	1,440	0,008	11,560
6	27	7	9,81	0,372	4,840	7,896	2,560

Продовження табл. 7.13

1	2	3	4	5	6	7	8
7	30	11	10,95	3,063	3,240	0,003	21,160
8	38	14	13,99	22,944	23,040	0,000	158,760
9	45	18	16,66	55,652	77,440	1,796	384,160
10	46	17	17,04	61,466	60,840	0,002	424,360
Усього	254	92	92,00	263,017	273,600	10,577	1816,400

Мірою щільності зв'язку є коефіцієнт детермінації R^2:

$$R^2 = \frac{\delta_Y^2}{\sigma^2} = \frac{26,302}{27,360} = 0,961,$$

де δ_Y^2 – факторна дисперсія:

$$\delta_Y^2 = \frac{\sum (Y - \overline{y})^2}{n} = \frac{263,017}{10} = 26,302,$$

– загальна дисперсія:

$$\sigma^2 = \frac{\sum (y - \overline{y})^2}{n} = \frac{273,600}{10} = 27,360$$

$$\text{або } \sigma^2 = \delta_Y^2 + \sigma_e^2 = 26,302 + 1,058 = 27,360,$$

σ_e^2 – залишкова дисперсія:

$$\sigma_e^2 = \frac{\sum (y - Y)^2}{n} = \frac{10,577}{10} = 1,058.$$

Коефіцієнт детермінації показує, що на 96,1% варіація прибутку підприємств пов'язана з варіацією вартості оборотних коштів.

Індекс кореляції R вказує на щільний зв'язок між вартістю оборотних коштів та прибутком підприємств, але не вказує його напрям:

$$R = \sqrt{R^2} = \sqrt{0,961} = 0,980.$$

Щільність та напрям зв'язку можна оцінити з допомогою лінійного коефіцієнта кореляції r

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y} = \frac{302,8 - 25,4 \cdot 9,2}{13,477 \cdot 5,231} = 0,980 \text{ або } r = b \frac{\sigma_x}{\sigma_y} = 0,3805 \frac{13,477}{5,231} = 0,980,$$

де $\overline{xy} = \frac{\sum xy}{n} = \frac{3028}{10} = 302,8;\ \overline{x} = 25,4;\ \overline{y} = 9,2;$

$$\sigma_x = \sqrt{\frac{\sum(x-\overline{x})^2}{n}} = \sqrt{\frac{1816{,}400}{10}} = 13{,}477,\ \sigma_y = \sqrt{\frac{\sum(y-\overline{y})^2}{n}} = \sqrt{\frac{273{,}600}{10}} = 5{,}231.$$

Цей показник свідчить про прямий ($r > 0$) і щільний зв'язок між досліджуваними ознаками.

При лінійній формі зв'язку абсолютна величина лінійного коефіцієнта кореляції дорівнює індексу кореляції, тобто $|r| = 0{,}980 = R$.

Істотність зв'язку перевіримо з допомогою F-критерію, фактичне значення якого в цьому випадку становить:

$$F = \frac{R^2}{1-R^2} \cdot \frac{k_2}{k_1} = \frac{0{,}961}{1-0{,}961} \cdot \frac{8}{1} = 197{,}13,$$

де $k_1 = 2–1=1$, $k_2 = 10–2=8$.

Для оцінки отриманого відношення його порівнюють з критичним (табличним). Знаходимо $F_{кр}$ при ймовірності 0,95 і ступенях вільності k_1, k_2 за таблицею 7.11. Воно становить $F_{0,95}(1; 8) = 5{,}12$. Отже, $F > F_{кр}$, що свідчить про суттєвий вплив вартості оборотних коштів підприємств на їх прибуток.

Задача 3

Розподіл респондентів за віком та схильністю до ризику наведений у табл. 7.14. До групи ризикових віднесено респондентів, які мають намір придбати криптовалюту, незважаючи на ризик, обережні не уявляють ризику без певних гарантій, не ризикові уникають ризику взагалі.

Таблиця 7.14

Розподіл респондентів за віком
та схильністю до ризику при придбанні криптовалюти

Вік, років	Тип інвестора:			Разом
	агресивний	помірний	консервативний	
До 30	24	12	4	40
30–50	20	50	30	100
50 і старше	6	18	36	60
Разом	50	80	70	200

Визначити щільність зв'язку між досліджуваними ознаками: віком респондентів та їх схильністю до ризику при придбанні криптовалюти.

Розв'язання задачі

Концентрація частот навколо діагоналі з верхнього лівого кута у правий нижній свідчить про наявність стохастичного зв'язку: чим більше вік, тим менша схильність до ризику. Фактичне значення χ^2 становить

$$\chi^2 = n\left[\sum_i \sum_j \frac{f_{ij}^2}{f_{i0} f_{0j}} - 1\right] =$$

$$= 200\left[\frac{24^2}{40\cdot 50} + \frac{12^2}{40\cdot 80} + \frac{4^2}{40\cdot 70} + \frac{20^2}{100\cdot 50} + \frac{50^2}{100\cdot 80} + \right.$$

$$\left. \frac{30^2}{100\cdot 70} + \frac{6^2}{60\cdot 50} + \frac{18^2}{60\cdot 80} + \frac{36^2}{60\cdot 70} - 1\right] = 49{,}6\,.$$

Критичне значення χ^2 для заданої імовірності $1 - \alpha$ та числа ступенів вільності $k = (m_x - 1)(m_y - 1) = (3 - 1)(3 - 1) = 4$, становить $\chi^2_{0,95}(4) = 9{,}49$ (табл. 7.4). Фактичне значення значно перевищує критичне ($\chi^2 = 49{,}6 > \chi^2_{0,95}(4) = 9{,}49$), що доводить істотність зв'язку між віком респондентів та їх схильністю до ризику при придбанні криптовалюти з імовірністю 0,95.

Оскільки $m_x = m_y = 3$, то для оцінки щільності зв'язку використаємо коефіцієнт спряженості Чупрова:

$$C_ч = \sqrt{\frac{\chi^2}{n\sqrt{(m_x - 1)(m_y - 1)}}} = \sqrt{\frac{49{,}6}{200\sqrt{(3-1)(3-1)}}} = 0{,}352\,,$$

що свідчить про помірний зв'язок між досліджуваними ознаками: віком респондентів та їх схильністю до ризику при придбанні криптовалюти.

Задача 4

Використовуючи дані про вартість оборотних коштів і прибуток десяти підприємств (табл. 7.6) за допомогою коефіцієнта Фехнера оцінити зв'язок між досліджуваними ознаками. Зробити висновки.

Розв'язання задачі

Для оцінювання щільність зв'язку за допомогою коефіцієнта Фехнера порівнюватимемо знаки відхилень окремих значень від середнього значення за кожною ознакою: $\bar{x} = \frac{254}{10} = 25{,}4$ млн грн, $\bar{y} = \frac{92}{10} = 9{,}2$ млн грн.

Результати цих порівнянь представлено у табл. 7.15. Якщо знаки відхилень по обох ознаках збігаються, то варіація ознак узгоджена і у графі 5 ставимо «З», незбігання знаків вказує на порушення узгодженості варіації досліджуваних ознак і у графі 6 ставимо «Н».

Таблиця 7.15

Розрахункові показники для визначення коефіцієнта Фехнера

№ підприємства	Вартість оборотних коштів, млн грн, x	Прибуток, млн грн, y	Відхилення окремих значень ознак від їх середнього значення		Знаки відхилень окремих значень ознак від їх середнього значення:	
			$x - \bar{x}$	$y - \bar{y}$	збігаються (З)	не збігаються (Н)
А	1	2	3	4	5	6
1	7	3	7,0–25,4= =–18,40	3,0–9,20= =–6,20	З	
2	11	4	–14,40	–5,20	З	
3	12	4	–13,40	–5,20	З	
4	16	6	–9,40	–3,20	З	
5	22	8	–3,40	–1,20	З	
6	27	7	1,60	–2,20		Н
7	30	11	4,60	1,80	З	
8	38	14	12,60	4,80	З	
9	45	18	19,60	8,80	З	
10	46	17	20,60	7,80	*З*	
Усього	254	92	–	–	9	1

Підрахувавши суму знаків відхилень, що збіглися і не збіглися, визначаємо коефіцієнт Фехнера

$$K_{\Phi} = \frac{\sum З - \sum Н}{\sum З + \sum Н} = \frac{9-1}{9+1} = 0{,}8.$$

Отже, коефіцієнт Фехнера вказує на те, що між прибутком підприємств і вартістю оборотних коштів існує прямий і дуже щільний зв'язок.

Задача 5

Використовуючи дані про вартість оборотних коштів і прибуток десяти підприємств (табл. 7.6) за допомогою коефіцієнта рангової кореляції Спірмена оцінити залежності між досліджуваними ознаками. Зробити висновки.

Розв'язання задачі

Кожному значенню ознаки присвоюємо відповідний ранг від менших значень ознаки до більших. Для різних одиниць сукупності з однаковими значеннями ознаки ранг визначають як середню арифметичну відповідних рангів: для другого і третього підприємств, у яких прибуток становить по 4 млн грн, середній ранг становитиме 2,5 ($\frac{2+3}{2}$).

Порядок розрахунку коефіцієнта кореляції рангів наведено у табл. 7.16.

Таблиця 7.16

Розрахункові показники
для визначення коефіцієнта рангової кореляції

№ підприємства	Вартість оборотних коштів, млн грн, x	Прибуток, млн грн, y	R_x	R_y	$d^2 = (R_x - R_y)^2$
1	2	3	4	5	6
1	7	3	1	1	$(1-1)^2 = 0{,}00$
2	11	4	2	2,5	0,25
3	12	4	3	2,5	0,25
4	16	6	4	4	0,00
5	22	8	5	6	1,00

Продовження табл. 7.16

1	2	3	4	5	6
6	27	7	6	5	1,00
7	30	11	7	7	0,00
8	38	14	8	8	0,00
9	45	18	9	10	1,00
10	46	17	10	9	1,00
Усього	254	92	–	–	4,50

Знаходимо коефіцієнт рангової кореляції Спірмена:

$$=1-\frac{6\sum_{1}^{n} d_j^2}{n(n^2-1)}=1-\frac{6\cdot 4{,}50}{10\cdot(10^2-1)}=1-0{,}027=0{,}973.$$

Коефіцієнт кореляції рангів Спірмена вказує на дуже щільний прямий зв'язок між вартістю оборотних коштів та прибутком підприємств.

Запитання для самоконтролю

1. Які форми зв'язку ви знаєте?
2. Як за допомогою аналітичного групування оцінити силу зв'язку між результативною та факторною ознаками?
3. Які показники щільності зв'язку вам відомі?
4. Що характеризують коефіцієнти регресії та еластичності? У чому відмінність між ними?
5. Як здійснюється перевірка істотності зв'язку?
6. Для чого використовуються та як розраховуються коефіцієнти Спірмена та Фехнера?
7. Як розраховується та інтерпретується лінійний коефіцієнт кореляції?
8. У чому особливість використання коефіцієнтів взаємної спряженості Крамера, Пірсона та Чупрова?
9. Що розуміють під прямим і оберненим зв'язком між результативною та факторною ознаками? Наведіть приклади.
10. Для чого використовується та як розраховується критерій Фішера?

Завдання для самостійної роботи

Тестові завдання

1. Як називають зв'язок у випадку, коли умовні розподіли замінюються одним параметром – середнім значенням результативної ознаки:

а) стохастичний;

б) функціональний;

в) кореляційний.

Відповіді: 1) а; 2) б; 3) в.

2. Зв'язок вважають істотним, якщо значення коефіцієнта асоціації за модулем

а) більше 0,5;

б) більше 0,3;

в) більше 0,7.

Відповіді: 1) а; 2) б; 3) в.

3. Для таблиць взаємної спряженості кількість ступенів свободи розраховують за формулою:

а) $(m_x - 1)+(m_y - 1)$;

б) $(m_x - 1)(m_y - 1)$;

в) $m_x - 1$;

г) $m_x = m_y$.

Відповіді: 1) а; 2) б; 3) в; 4) г.

4. В яких межах змінюються значення коефіцієнтів Чупрова, Крамера та Пірсона?

а) від -1 до 1;

б) від 0 до 1;

в) від -1 до ∞;

г) від $-\infty$ до $+\infty$.

Відповіді: 1) а; 2) б; 3) в; 4) г.

5. Який зв'язок передбачає, що кожному можливому значенню факторної ознаки x відповідає чітко визначене значення результатив-

ної ознаки *y*, тобто існує повна відповідність між причиною і наслідком, яку можна описати за допомогою точної математичної формули?

а) стохастичний;

б) функціональний;

в) кореляційний.

Відповіді: 1) а; 2) б; 3) в.

6. Який з параметрів рівняння регресії $Y=a+bx$ показує, на скільки одиниць власного виміру змінюється середнє значення результативної ознаки зі збільшенням факторної ознаки на одиницю власного виміру?

а) a ;

б) b;

в) x.

Відповіді: 1) а; 2) б; 3) в.

7. В яких межах змінюються значення лінійного коефіцієнта кореляції?

а) від –1 до 1;

б) від 0 до 1;

в) від –1 до ∞;

г) від –∞ до +∞.

Відповіді: 1) а; 2) б; 3) в; 4) г.

8. Який зв'язок проявляється як узгодженість варіації двох чи більше ознак? Кожному значенню факторної ознаки відповідає кілька значень результативної, що утворюють так званий умовний розподіл:

а) стохастичний;

б) функціональний;

в) кореляційний.

Відповіді: 1) а; 2) б; 3) в.

9. Який коефіцієнт показує, на скільки відсотків зміниться в середньому результативна ознака *y* при зміні факторної ознаки *x* на 1%:

а) варіації;

б) детермінації;

в) еластичності;

г) кореляції.

Відповіді: 1) а; 2) б; 3) в; 4) г.

10. Коефіцієнти асоціації, колігації та контингенції згідно з правилом мажорантності співвідносяться між собою так:

а) $K_{кол} > K_{конт} > K_{ас}$;

б) $K_{ас} > K_{кол} > K_{конт}$;

в) $K_{конт} > K_{кол} > K_{ас}$;

г) $K_{ас} = K_{кол} = K_{конт}$.

Відповіді: 1) а; 2) б; 3) в; 4) г.

Практичні завдання (задачі)

11. За наведеними нижче даними визначте міжгрупову, середню з групових та загальну дисперсії врожайності озимої пшениці. Розрахуйте показники щільності зв'язку. Перевірте істотність зв'язку з ймовірністю 0,95. Зробіть висновки.

Залежність втрат від браку продукції від процента порушення технологічної дисципліни

Процент порушень технологічної дисципліни	Кількість господарств, од.	Середні втрати від браку продукції, тис. грн	Втрати від браку продукції, тис. грн
1,2–1,7	9	1,3	1,0; 1,2; 1,4; 1,2; 0,9; 1,5; 1,4; 1,4; 1,3
1,7–2,2	7	1,6	1,6; 1,5; 1,4; 1,7; 1,7; 1,8; 1,6
2,2–2,7	6	2,0	1,9; 2,1; 1,6; 2,0; 2,1; 2,0
Усього	22	–	–

12. Маємо дані про обстеження споживання електроенергії у житловому секторі:

Категорія житла	Кількість споживачів, тис.	Середньомісячне споживання електроенергії в розрахунку на одного споживача, кВт·год
Квартира з газовою плитою	40	300
Квартира з електричною плитою	85	600

Визначте міжгрупову та середню з групових дисперсій споживання електроенергії у житловому секторі, якщо відомо, що загальна дисперсія дорівнює 26112. Розрахуйте показники щільності зв'язку. Перевірте істотність зв'язку з ймовірністю 0,95. Зробіть висновки.

13. Маємо дані щодо трьох груп робітників підприємства з різним стажем роботи:

Стаж роботи, років	Число робітників, осіб	Середня зарплата, грн	Середнє квадратичне відхилення зарплати, грн
До 3	10	5000	120
3–10	15	6000	100
Понад 10	25	7000	200

Визначте міжгрупову, середню з групових та загальну дисперсії середньої зарплати. Розрахуйте показники щільності зв'язку. Перевірте істотність зв'язку з ймовірністю 0,95. Зробіть висновки.

14. Маємо такі дані про показники діяльності підприємств:

№ з/п	Випуск продукції, млн грн	Прибуток, млн грн
1	65	15,7
2	78	18,0
3	41	12,1
4	54	13,8
5	66	15,5
6	80	17,9

7	45	12,8
8	57	14,2
9	67	15,9
10	81	17,6
11	92	18,2
12	48	13,0
13	59	16,5
14	68	16,2
15	83	16,7

Побудуйте однофакторну регресійну модель залежності між випуском продукції та прибутком. Побудуйте графік кореляційного поля. Визначте показники щільності кореляційного зв'язку. Перевірте зв'язок на істотність з імовірністю 0,95.

15. Маємо такі дані по 10 підприємствах:

№ підприємства	1	2	3	4	5	6	7	8	9	10
Кількість працівників з професійною підготовкою, %	10	12	14	17	24	28	30	35	40	50
Кількість бракованої продукції, %	18	17	14	12	10	10	8	9	6	6

За допомогою коефіцієнта рангової кореляції Спірмена оцініть залежності між кількістю бракованої продукції кількістю працівників з професійною підготовкою. Зробіть висновки

16. Маємо такі дані про показники діяльності підприємств:

№ з/п	Втрати робочого часу, тис. людино-днів	Прибуток, млн грн
1	66,0	15,7
2	44,0	18,0
3	91,0	12,1
4	78,0	13,8

5	57,4	15,5
6	42,0	17,9
7	100,0	12,8
8	79,8	14,2
9	57,0	15,9
10	38,0	17,6
11	23,1	18,2
12	112,0	13,0
13	72,0	16,5
14	55,7	16,2
15	36,0	16,7

За допомогою коефіцієнта Фехнера оцініть зв'язок між прибутком та втратами робочого часу. Зробіть висновки.

17. Молоді робітники за ступенем задоволеності умовами праці та професійною мобільністю розподілені так:

Ступінь задоволеності умовами праці	Чи маєте намір змінити професію?			Усього
	Так, найближчим часом	Так, у перспективі	Ні	
Задоволений	–	20	26	46
Ставлюся байдуже	7	18	9	34
Не задоволений	15	5	–	20
Усього	22	43	35	100

Проаналізуйте комбінаційний розподіл працівників, оцініть щільність зв'язку між задоволеністю умовами праці та професійною мобільністю з допомогою коефіцієнта взаємної спряженості. Перевірте істотність зв'язку з імовірністю 0,95, зробіть висновки.

Розділ 8.
АНАЛІЗ ІНТЕНСИВНОСТІ ДИНАМІКИ І ТЕНДЕНЦІЙ РОЗВИТКУ

8.1. Поняття про ряди динаміки, їх елементи та правила побудови

Суспільні явища безперервно змінюються в часі під впливом різних соціальних, економічних, технічних та інших чинників. Для вивчення цієї зміни складаються й аналізуються ряди динаміки. *Динамічний ряд* – впорядковані у часі значення статистичних показників для вивчення розвитку та зміни соціально-економічних явищ у динаміці. Будь-якому динамічному ряду властиві два елементи: перелік хронологічних дат (моментів) або інтервалів часу t і конкретні відповідні значення показників, що відповідають цим моментам (інтервалам) часу, які називаються рівнями ряду y.

За часовою ознакою динамічні ряди поділяють на *інтервальні* та *моментні*. В інтервальних рядах рівні ряду відображають розміри досліджуваного явища за певний проміжок (інтервал) часу (день, місяць, квартал, півріччя, рік тощо). Рівень моментного ряду фіксує стан досліджуваного явища на певний момент часу t (на перше число місяця, на початок чи кінець досліджуваного періоду).

За способом вираження рівнів динамічні ряди поділяються на ряди абсолютних, відносних і середніх величин.

При побудові динамічних рядів потрібно дотримуватися порівнянності всіх рівнів ряду між собою за періодами часу, територією, методологією обчислення показників, одиницями вимірювання, колом охоплюваних об'єктів.

Порівнянність за колом охоплюваних об'єктів забезпечується зімкненням динамічний рядів шляхом заміни абсолютних рівнів ряду відносними (коли за базу порівняння для кожного ряду беруть один

(перехідний) рівень) або абсолютними величинами, що коригуються на коефіцієнт зміни, визначений як відношення рівня нового ряду до рівня старого ряду за певний перехідний період.

8.2. Статистичні характеристики динамічних рядів

При вивченні швидкості та інтенсивності динаміки соціально-економічних явищ і процесів статистика використовує низку взаємопов'язаних характеристик, розрахунок яких базується на абсолютному чи відносному порівнянні рівнів динамічного ряду. До таких характеристик належать абсолютний приріст або зменшення, коефіцієнт (темп) зростання або зменшення, темп приросту або скорочення, абсолютне значення 1% приросту або зменшення. Рівень ряду, що порівнюється, називається поточним (y_t), а рівень ряду, з яким проводиться порівняння, – базисним. За базисний рівень приймається або попередній (y_{t-1}), або початковий (y_0) рівень ряду. Якщо кожний рівень ряду порівнюється з попереднім (змінна база порівняння), характеристики динаміки називаються ланцюговими, якщо з початковим (постійна база порівняння) – то базисними (рис. 8.1).

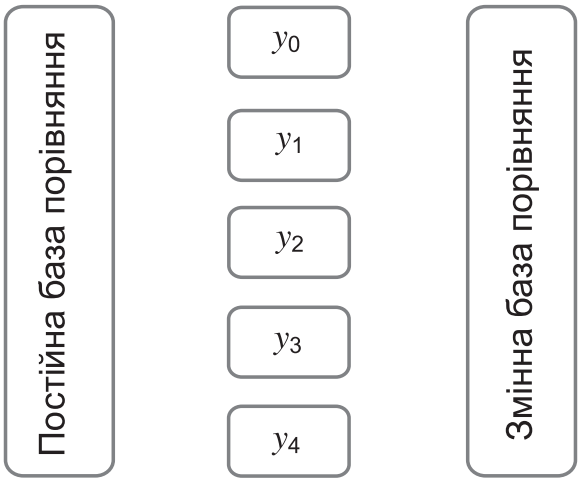


Рис. 8.1. Схеми порівняння при розрахунку характеристик динаміки

Абсолютний приріст Δ показує, на скільки одиниць власного виміру підвищився або знизився рівень за певний проміжок часу, тобто характеризує абсолютну швидкість зміни рівнів ряду динаміки. Він може бути додатнім або від'ємним і обчислюється як різниця рівнів ряду динаміки.

Коефіцієнт зростання k є відносною характеристикою інтенсивності зміни рівнів ряду динаміки, він характеризує відносну швидкість їх зміни, тобто показує, у скільки разів рівень y_t більший (менший) за рівень, взятий за базу порівняння. Коефіцієнт обчислюють, зіставляючи два рівні ряду динаміки. При збільшенні рівня $k > 1$, при зменшенні $k < 1$. Якщо коефіцієнт зростання k помножити на 100%, то отримаємо темп зростання.

Темп приросту T показує, на скільки відсотків рівень y_t підвищився або знизився порівняно з базою порівняння. Він може бути додатним або від'ємним.

Абсолютне значення 1% приросту A показує, яка абсолютна величина прихована за відносним показником – 1% приросту. Цей показник також може бути додатним або від'ємним.

Формули для розрахунку характеристик динаміки за ланцюговою та базисною системами наведені у табл. 8.1.

Таблиця 8.1

Розрахунок характеристик динаміки
за ланцюговою та базисною системами

Назва показника	Ланцюгова система	Базисна система
Абсолютний приріст, D	$\Delta_л = y_t - y_{t-1}$	$\Delta_б = y_t - y_0$
Коефіцієнт зростання, k	$k_л = \frac{y_t}{y_{t-1}}$	$k_б = \frac{y_t}{y_0}$
Темп зростання, k, %	$k_л = \frac{y_t}{y_{t-1}} \cdot 100$	$k_б = \frac{y_t}{y_0} \cdot 100$
Темп приросту, T, %	$T = (k_л - 1) \cdot 100$	$T = (k_б - 1) \cdot 100$
Абсолютне значення 1% приросту, A	$A = \frac{\Delta_л}{T_л} = \frac{y_{t-1}}{100}$	$A = \frac{\Delta_б}{T_б} = \frac{y_0}{100}$

Ланцюгові та базисні характеристики динаміки взаємопов'язані:

1) сума ланцюгових абсолютних приростів дорівнює кінцевому базисному:

$$\sum_{1}^{n}\Delta_{л}=\sum_{1}^{n}(y_t-y_{t-1})=y_n-y_0;$$

2) добуток ланцюгових коефіцієнтів зростання дорівнює кінцевому базисному:

$$k_1\cdot k_2\cdot\ldots\cdot k_n=\prod_{1}^{n}k_{л}=\frac{y_n}{y_0};$$

3) відношення наступного базисного коефіцієнта зростання до попереднього дорівнює відповідному ланцюговому коефіцієнту зростання:

$$\frac{y_t}{y_0}\div\frac{y_{t-1}}{y_0}=\frac{y_t}{y_{t-1}}.$$

Якщо швидкість розвитку в межах періоду, що вивчається, неоднакова, то порівнянням однойменних характеристик швидкості вимірюється прискорення чи уповільнення динаміки:

- абсолютне прискорення чи уповільнення δ_t – різниця між абсолютними приростами: $\delta_t=\Delta_t-\Delta_{t-1}$. Якщо $\delta_t>0$, має місце прискорення, якщо $\delta_t<0$ – уповільнення;
- коефіцієнт прискорення чи уповільнення абсолютної швидкості φ_{Δ} розраховується за формулою $\varphi_{\Delta}=\frac{\Delta_t}{\Delta_{t-1}}$;
- коефіцієнт прискорення або уповільнення відносної швидкості розвитку φ_t визначається за формулою $\varphi_t=\frac{k_t}{k_{t-1}}$.

Інтенсивність динаміки в різних рядах досліджується за допомогою коефіцієнта випередження, який розраховується як відношення коефіцієнтів зростання за однакові проміжки часу за двома динамічними рядами k' та k'': $K_{вип}=\frac{k'}{k''}$.

8.3. Середні характеристики динамічних рядів

Важливими узагальнюючими показниками динамічного ряду є його середні характеристики: середній рівень динамічного ряду, середній абсолютний приріст (зменшення), середній коефіцієнт зростання, середній темп приросту (зменшення).

Розрахунок середнього рівня динамічного ряду залежить від виду цього ряду та величини інтервалу, що відповідає кожному рівню. Для інтервальних рядів з однаковими періодами часу середній рівень $\overline{y}$ визначається за формулою середньої арифметичної простої:

$$\overline{y} = \frac{\sum_{1}^{n} y_t}{n},$$

де n – число рівнів ряду.

Якщо в інтервальному ряду періоди часу різні за тривалістю, то середній рівень визначається за формулою середньої арифметичної зваженої:

$$\overline{y} = \frac{\sum y_t \cdot D_t}{\sum D_t},$$

де D_t – тривалість періодів часу.

У моментному ряді за припущення про рівномірну зміну показника між датами середній рівень ряду визначається як напівсума значень на початок і кінець досліджуваного періоду:

$$\overline{y} = \frac{y_0 + y_n}{2}.$$

Для моментного ряду з однаковими інтервалами часу середній рівень ряду визначається за формулою середньої хронологічної:

$$\overline{y} = \frac{\frac{y_1 + y_n}{2} + \sum_{2}^{n-1} y_t}{n-1}.$$

Середній абсолютний приріст показує, на скільки у середньому за одиницю часу (у середньому щорічно, щомісячно і т. ін.) у досліджуваний період змінювалися рівні ряду динаміки.

Середній абсолютний приріст обчислюють за формулою:

$$\overline{\Delta} = \frac{\sum \Delta_n}{n-1} = \frac{y_n - y_0}{n-1}$$

де $\overline{\Delta}$ – середній абсолютний приріст; n – число рівнів ряду динаміки.

Середній коефіцієнт зростання обчислюють за формулою середньої геометричної:

$$\overline{k} = \sqrt[n-1]{k_1 \cdot k_2 \cdot \ldots \cdot k_{n-1} \cdot k_n} = \sqrt[n-1]{\frac{y_n}{y_0}},$$

де n – число рівнів ряду динаміки.

Середній темп приросту $\overline{T}$ показує, на скільки відсотків щорічно підвищується або знижується рівень, і розраховується за формулою:

$$\overline{T} = (\overline{k} - 1) \cdot 100.$$

8.4. Методи аналізу тенденції розвитку

При аналізі рядів динаміки важливо виявити загальну тенденцію розвитку (тренд) соціально-економічного явища, тобто встановити, в якому напрямі (зростає, зменшується) і за якою залежністю (лінійна чи нелінійна) вона змінюється. У статистичній практиці виявлення основної тенденції розвитку суспільних явищ у часі проводиться за допомогою методів:

1) укрупнення інтервалів – дані динамічного ряду об'єднуються в групи за періодами (як правило три роки, п'ять років тощо) та розраховується середній показник за період;
2) плинної (ковзної) середньої – укрупнення інтервалів проводиться шляхом послідовних зміщень на одну дату при збереженні постійного інтервалу періоду;
3) аналітичне вирівнювання – рівні ряду динаміки розглядаються як функція часу $Y = f(t)$, задача вирівнювання зводиться до знаходження того виду функції, значення якої були б найближчими до фактичних величин динамічного ряду. Цей метод є найбільш ефективним для визначення основної тенденції розвитку суспільних явищ і процесів.

Найбільш поширеними формулами, які виражають тенденцію розвитку явищ, є:

- ❑ лінійна функція $Y = a + bt$; використовується, якщо ланцюгові абсолютні прирости відносно стабільні, не мають чіткої тенденції до зростання чи зменшення;

- ❑ парабола $Y = a + bt + ct^2$; використовується, якщо абсолютні ланцюгові прирости мають деяку тенденцію, але абсолютні ланцюгові прирости абсолютних ланцюгових приростів (різниці другого порядку) ніякої тенденції розвитку не демонструють;
- ❑ експонента $Y = ab^t$; використовується, якщо відносно стабільними є ланцюгові темпи приросту.

У розглянутих функціях Y – вирівняні значення динамічного ряду, t – порядковий (умовний) номер періоду, a – рівень ряду при t = 0, b – швидкість динаміки (середня абсолютна в лінійній функції та параболі, середня відносна – в експоненті), c – прискорення розвитку. Параметри трендових рівнянь визначаються методом найменших квадратів.

При застосуванні аналітичного вирівнювання найчастіше використовується лінійна функція $Y = a + bt$, параметри якої визначаються розв'язанням системи нормальних рівнянь

$$\begin{cases} na + b\sum t = \sum y \,, \\ a\sum t + b\sum t^2 = \sum yt \,, \end{cases}$$

де n – кількість рівнів ряду динаміки.

Якщо відлік значень t перенести у середину ряду динаміки, що вивчається, то $\sum t = 0$, система рівнянь спрощується і параметри a і b обчислюються за формулами:

$$a = \frac{\sum y}{n}; \; b = \frac{\sum yt}{\sum t^2} \,.$$

Визначені параметри трендового рівняння і відповідні значення t підставляються у формулу рівняння і розраховуються вирівняні (теоретичні) рівні динамічного ряду Y .

У теорії і практиці на основі вивчення динаміки та тенденції зміни у часі статистичних показників досить часто використовують прийом інтерполяції або екстраполяції рядів динаміки. Інтерполяція – це знаходження відсутніх проміжних рівнів ряду. Знаючи рівняння тренду для обчислення теоретичних рівнів і підставляючи в нього проміжне значення, можна визначити відповідний йому теоретичний рівень Y.

Екстраполяція використовується при прогнозуванні суспільних явищ у майбутньому за припущення, що виявлена тенденція зберігатиметься і надалі за межами досліджуваного ряду динаміки. При цьому значення *t* за межами наявного динамічного ряду підставляють у трендове рівняння й отримують точкове прогнозне значення рівня тренду Y_{t+v} у майбутньому.

На практиці спрогнозовані у результаті екстраполяції рівні соціально-економічних явищ зазвичай доповнюють інтервальними оцінками, які з певною ймовірністю визначають довірчі межі прогнозного рівня $Y_{t+v} \pm t_{1-a}(k) \cdot S_p$. При цьому $t_{1-a}(k)$ – довірче число для прийнятого рівня ймовірності (*t*-критерій Стьюдента – коефіцієнти або числові характеристики, що широко використовуються для побудови довірчих інтервалів), критичні значення *t*-критерію для рівня істотності α=0,05 та числа ступенів свободи *k* наведені у табл. 8.2; S_p – стандартна похибка прогнозу, яка залежить від адекватності трендової моделі (оцінки залишкової дисперсії σ_e^2), довжини динамічного ряду *n* та періоду упередження *v*:

$$S_p = \sigma_e \cdot \sqrt{\frac{n+1}{n} + \frac{3(n+2v-1)^2}{n(n^2-1)}},$$

$$\sigma_e = \sqrt{\sigma_e^2} = \sqrt{\frac{1}{n-m} \cdot \sum (y-Y)^2},$$

де *m* – кількість параметрів трендового рівняння.

Таблиця 8.2

Критичні значення *t*-критерію Стьюдента
для рівня істотності α=0,05 та числа ступенів свободи *k*

$k=(n-m)$	3	4	5	6	7	8	9	10	11	12	14	16	18	20
$t_{0,95}(k)$	2,35	2,13	2,02	1,94	1,89	1,86	1,83	1,81	1,80	1,78	1,76	1,75	1,73	1,72

Оцінку якості обраного рівняння тренда проводять з допомогою абсолютної похибки апроксимації:

$$\overline{A} = \frac{\sum |y-Y| \div y}{n} 100\%.$$

Якщо модель підібрана з високою точністю, $\overline{A} < 10\%$, доброю – $10\% < \overline{A} < 20\%$, задовільною – $20\% < \overline{A} < 50\%$, незадовільною – $\overline{A} > 50\%$.

8.5. Коливання та сталість динамічних рядів

Під впливом різних чинників фактичні рівні динамічного ряду відхиляються від теоретично визначеної основної тенденції розвитку. Ці відхилення можуть мати систематичний (або закономірний) характер, тобто повторюватися через певні інтервали часу, чи випадковий характер. Для конкретного ряду характерним може бути поєднання систематичних і випадкових коливань.

Випадкові коливання ряду динаміки можна оцінити з допомогою таких показників:

- середнє квадратичне відхилення σ_e як абсолютна міра випадкових коливань є;
- відносною мірою випадкових коливань є коефіцієнт варіації V_e:

$$V_e = \frac{\sigma_e}{\overline{y}} \cdot 100.$$

Для оцінки сталості динамічного ряду використовують формулу $100 - V_e$. У випадку, якщо $(100 - V_e) \to 100$, можна стверджувати, що має місце сталий характер тенденції розвитку.

Багато суспільних явищ і процесів характеризуються систематичними змінами, що повторюються з року в рік помісячно чи поквартально. Ці зміни називаються сезонними коливаннями. Вони пов'язані з нерівномірним розвитком явищ і процесів протягом року, а отже, потребують статистичного вивчення з метою регулювання їх зміни. Сезонні коливання описуються за допомогою моделі "сезонна хвиля", яку утворює сукупність індексів сезонності I_c. Індекси сезонності визначаються як відношення фактичних місячних чи квартальних рівнів y_t до середньомісячного рівня $\overline{y}$ або теоретичного рівня Y, якщо ряд динаміки демонструє тенденцію ($Y = f(t)$):

$$I_c = \frac{y_t}{\overline{y}} \cdot 100 \text{ або } I_c = \frac{y_t}{Y} \cdot 100.$$

Узагальнюючими характеристиками сезонних коливань, що дозволяють порівнювати інтенсивність сезонності різних явищ або одного явища за різні роки, є:

1) амплітуда коливань $R_c = I_c^{\max} - I_c^{\min}$, де $I_c^{\max}$, $I_c^{\min}$ – максимальне і мінімальне значення індексів сезонності;

2) середнє лінійне відхилення $\bar{l}_c = \frac{1}{n}\sum_{1}^{n}|I_c - 100|$, де n – кількість періодів: $n = 12$ для місячних даних, $n = 4$ – для квартальних;

3) середнє квадратичне відхилення $\sigma_c = \sqrt{\frac{1}{n}\sum_{1}^{n}(I_c - 100)^2}$.

У рядах динаміки з невеликими інтервалами часу з метою характеристики закономірних коливань використовують коефіцієнти нерівномірності K_{max} та K_{min}:

$$K_{max} = \frac{y_{max}}{\bar{y}},$$

$$K_{min} = \frac{y_{min}}{\bar{y}},$$

де y_{max}, y_{min}, та $\bar{y}$ – максимальний, мінімальний та середній рівні динамічного ряду відповідно.

На основі коефіцієнтів нерівномірності визначається амплітуда коливань $R_K = (K_{max} - K_{min}) \cdot 100$, яка використовується для оцінки істотності нерівномірності. Чим більша різниця між K_{max} та K_{min}, тим більшою є нерівномірність досліджуваного процесу.

Розв'язання типових задач

Задача 1

Використовуючи наведену у табл. 8.3 інформацію про обсяг виготовленої продукції зімкнути динамічний ряд.

Таблиця 8.3

Обсяг виготовленої продукції певного виду

(млн грн)

Рік	2014	2015	2016	2017	2018	2019
Продукція чотирьох підприємств	35	45	50	–	–	–
Продукція семи підприємств	–	–	70	85	90	98

Розв'язання задачі

Для побудови зімкненого ряду виготовлення продукції нинішнім складом підприємств на основі даних 2016 року обчислюємо коефіцієнт зміни обсягу реалізованої продукції за рахунок збільшен-

ня кількості підприємств: 70/50=1,4. Потім коригуємо дані за 2014–2015 рр. (табл. 8.4).

Таблиця 8.4

Обсяг реалізованої продукції певного виду

(млн грн)

Рік	2014	2015	2016	2017	2018	2019
Продукція семи підприємств	35·1,4=49	45·1,4=63	70	85	90	98

Або обсяг продукції 2016 року приймається за 100% та визначаються відносні показники динаміки для попередніх і наступних років (табл. 8.5).

Таблиця 8.5

Обсяг реалізованої продукції певного виду

(%)

Рік	2014	2015	2016	2017	2018	2019
Продукція восьми підприємств	$\frac{35}{50}\cdot 100=70$	$\frac{45}{50}\cdot 100=90$	100	–	–	–
Продукція десяти підприємств	–	–	100	$\frac{85}{70}\cdot 100=121$	$\frac{90}{70}\cdot 100=129$	$\frac{98}{70}\cdot 100=140$
Зімкнений ряд	70	90	100	121	129	140

Задача 2

Динаміка середньої кількості безробітних в Україні за методологією МОП у 2010–2018 роках представлена у табл. 8.6.

Таблиця 8.6

Середня кількість безробітних за методологією МОП, Україна, 2010–2018 рр.

Рік	2010	2011	2012	2013	2014	2015	2016	2017	2018
Середня кількість безробітних, тис. осіб	1713,9	1661,9	1589,8	1510,4	1847,6	1654,7	1678,2	1698,0	1578,6

Визначити:

1) абсолютні та відносні характеристики динаміки за ланцюговою та базисною системами;
2) середній рівень динамічного ряду;
3) середні значення показників динаміки: абсолютного приросту; коефіцієнта зростання; темпу приросту.

Розв'язання задачі

Формули та розрахунок характеристик динаміки середньої кількості безробітних в Україні за методологією МОП (далі – кількість безробітних) за ланцюговою системою наведений у графах 2, 4, 6 та 8 табл. 8.7, а за базисною системою – у графах 3, 5 та 7 табл. 8.7.

Таблиця 8.7

Характеристики динаміки середньої кількість безробітних за методологією МОП, Україна, 2010–2018 рр.

Рік	Середня кількість безробітних, тис. осіб	Абсолютний приріст, тис. осіб		Коефіцієнт зростання		Темп приросту, %		Абсолютне значення 1% приросту, тис. осіб
		ланцюговий	базисний	ланцюговий	базисний	ланцюговий	базисний	
А	1	2	3	4	5	6	7	8
	y	$\Delta_л$	$\Delta_б$	$k_л$	$k_б$	$T_л$	$T_б$	A
2010	1713,9	–	–	–	1,000	–	–	–
2011	1661,9	$1661,9 - 1713,9 = -52,0$	$1661,9 - 1713,9 = -52,0$	$\frac{1661,9}{1713,9} = 0,970$	$\frac{1661,9}{1713,9} = 0,970$	$(0,970 - 1) \cdot 100 = -3,0$	$(0,970 - 1) \cdot 100 = -3,0$	17,139

Продовження табл.8.7

A	1	2	3	4	5	6	7	8
2012	1589,8	$1589,8 - 1661,9 = 72,1$	$1589,8 - 1713,9 = -124,1$	$\frac{1589,8}{1661,9} = 0,957$	$\frac{1589,8}{1713,9} = 0,928$	$(0,957 - 1) \cdot = -4,3$	–7,2	16,619
2013	1510,4	–79,4	–203,5	0,950	0,881	–5,0	–11,9	15,898
2014	1847,6	337,2	133,7	1,223	1,078	22,3	7,8	15,104
2015	1654,7	–192,9	–59,2	0,896	0,966	–10,4	–3,4	18,476
2016	1678,2	23,5	–35,7	1,014	0,979	1,4	–2,1	16,547
2017	1698,0	19,8	–15,9	1,012	0,991	1,2	–0,9	16,782
2018	1578,6	–119,4	–135,3	0,930	0,921	–7,0	–7,9	16,980

Як показали обчислені показники інтенсивності динаміки, кількість безробітних порівняно з попередніми періодами зменшувалася у 2011–2013 роках, а у 2014 році стрімко зросла: порівняно з 2010 роком вона збільшилася на 133,7 тис. осіб, або на 7,80%, а порівняно з 2013 – на 337,2 тис. осіб, або на 22,33% (див. табл. 2.1). У 2015 році середня кількість безробітних скоротилася порівняно як із попереднім роком (на 10,44%), так і з 2010 роком (на 3,45%). Починаючи з 2016 року досліджуваний показник почав поступово зростати, але у 2017 році не було досягнуто рівня 2010 року. У 2018 році мало місце зменшення середньої чисельності безробітних порівняно як з попереднім, так і з базовим роком. У цілому по Україні за дев'ять досліджуваних років зазначений показник зменшився на 135,3 тис. осіб, або на 7,9%.

Динаміка кількості безробітних в Україні за методологією МОП – це інтервальний динамічний ряд з однаковими періодами часу, тому середній рівень $\overline{y}$ визначаємо за формулою середньої арифметичної простої:

$$\overline{y} = \frac{\sum_{1}^{n} y_t}{n} = \frac{1713{,}9 + 1661{,}9 + 1589{,}8 + 1510{,}4 + 1847{,}6 + 1654{,}7 + 1678{,}2 + 1698{,}0 + 1578{,}6}{9} = 1659{,}2 \text{ (тис. осіб).}$$

Тобто середня кількість безробітних в Україні за методологією МОП у 2010-2018 роках становить 1659,2 тис. осіб

Середній абсолютний приріст можна обчисли двома способами:

1) використовуючи ланцюгові абсолютні прирости (графа 2 табл. 8.7):

$$\overline{\Delta} = \frac{\sum \Delta_л}{n-1} = \frac{-52{,}0 + (-72{,}1) + (-79{,}4) + 337{,}2 + (-192{,}9) + 23{,}5 + 19{,}8 + (-119{,}4)}{9-1} = \frac{-135{,}3}{8} = -16{,}9 \text{ (тис. осіб),}$$

2) використовуючи рівні динамічного ряду (графа 1 табл. 8.7):

$$\overline{\Delta} = \frac{y_n - y_0}{n-1} = \frac{1578{,}6 - 1713{,}9}{9-1} = \frac{-135{,}3}{8} = -16{,}9 \text{ (тис. осіб).}$$

Отже, кількість безробітних в Україні за методологією МОП у 2010–2018 роках щорічно зменшувалась у середньому на 16,9 тис. осіб.

Середній коефіцієнт зростання також можна визначити двома способами:

1) використовуючи ланцюгові коефіцієнти зростання (графа 4 табл. 8.7):

$$\overline{k} = \sqrt[n-1]{k_1 \cdot k_2 \cdot \ldots \cdot k_{n-1} \cdot k_n} = \sqrt[9-1]{0{,}970 \cdot 0{,}957 \cdot 0{,}950 \cdot 1{,}223 \cdot 0{,}896 \cdot 1{,}014 \cdot 1{,}012 \cdot 0{,}930} = \sqrt[8]{0{,}921};$$

2) використовуючи рівні динамічного ряду (графа 1 табл. 8.7):

$$\overline{k} = \sqrt[n-1]{\frac{y_n}{y_0}} = \sqrt[9-1]{\frac{1578{,}6}{1713{,}9}} = \sqrt[8]{0{,}921} = 0{,}990.$$

Середній темп приросту $\overline{T}$ становить:

$$\overline{T} = (\overline{k} - 1) \cdot 100 = (0{,}990 - 1) \cdot 100 = -10{,}0 \text{ (\%).}$$

Тобто кількість безробітних в Україні за методологією МОП у 2010–2018 роках у середньому щорічно зменшувалася на 10%.

Задача 3

Кількість зареєстрованих безробітних в Україні за даними Державної служби зайнятості характеризується такими даними (табл. 8.8)

Таблиця 8.8

Кількість зареєстрованих безробітних у 2018 році

(на кінець звітного періоду; тис. осіб)

Період	Грудень 2017	2018											
		Січень	Лютий	Березень	Квітень	Травень	Червень	Липень	Серпень	Вересень	Жовтень	Листопад	Грудень
Кількість безробітних	354,4	378,9	383,7	366,9	326,8	316,0	303,9	298,0	292,8	287,1	271,4	301,0	341,7

Визначте середню кількість безробітних, зареєстрованих у 2018 році.

Розв'язання задачі

Наведений у табл. 8.8 динамічний ряд характеризує кількість зареєстрованих безробітних на кінець кожного місяця, а отже, є моментним динамічним рядом з однаковими інтервалами часу. Середній рівень такого динамічного ряду визначається за формулою:

$$\bar{y} = \frac{\frac{354,4 + 341,7}{2} + (378,9 + 383,7 + 366,9 + 326,8 + 316,0 + 303,9 + 298,0 + 298,0 + 292,8 + 287,1 + 271,4 + 301,0)}{13 - 1} = 322,9 \text{ (тис. осіб).}$$

Отже, середня кількість зареєстрованих безробітних у 2018 році за даними Державної служби зайнятості становить 322,9 тис. осіб.

Задача 4

Маємо дані про динаміку середньомісячних сукупних ресурсів у розрахунку на одне домогосподарство області у 2014–2019 роках (табл. 8.9):

Таблиця 8.9

Середньомісячні сукупні ресурси у розрахунку на одне домогосподарство області у фактичних цінах у 2014–2019 рр.

Роки	2014	2015	2016	2017	2018	2019
Середньомісячні сукупні ресурси, грн	3820,7	4441,7	5361,1	7091,1	8721,0	9893,9

Охарактеризувати тенденцію розвитку середньомісячних сукупних ресурсів у розрахунку на одне домогосподарство області з допомогою трендового рівняння.

Визначити розмір середньомісячних сукупних ресурсів у розрахунку на одне домогосподарство області через 2 роки.

Розв'язання задачі

Для виявлення основної тенденції розвитку середньомісячних сукупних ресурсів у розрахунку на одне домогосподарство регіону застосуємо методи аналітичного вирівнювання й екстраполяції. Результати зроблених необхідних розрахунків занесемо у графи 2–6 табл. 4.1. Зазначимо, що у графі 2 табл. 8.10 умовні показники часу t обираємо так, щоб їхня сума $\sum t = 0$, а оскільки число членів рівнів ряду є парним, то для 2016 року $t = -1$, а для 2017 року $t = 1$. Використовуючи рівняння прямої $Y_t = a + bt$ і виконуючи розрахунки за даними, наведеними у графах 1, 3 і 4 табл. 8.10, визначимо параметри трендового рівняння:

$$a = \frac{\sum y}{n} = \frac{39329,5}{6} = 6554,95.$$

Лінійний тренд має вигляд $Y = 6554,95 + 641,91 \cdot t$. Тобто середній розмір середньомісячних сукупних ресурсів у розрахунку на одне домогосподарство області у 2014–2019 рр. становив 6554,95 грн і щорічно за цей період досліджуваний показник зростав у середньому на 641,91 грн. Знаючи рівняння тренда, обчислимо теоретичні рівні

середньомісячних сукупних ресурсів у розрахунку на одне домогосподарство Y за 2014–2019 рр. (графа 5 табл. 8.10). Як видно з табл. 8.10, суми фактичних рівнів $\sum y$ і розрахованих за лінійним трендом теоретичних рівнів однакові: $\sum y = \sum Y = 39329{,}5$ грн.

Таблиця 8.10

Динаміка середньомісячних сукупних ресурсів у розрахунку на одне домогосподарство області у фактичних цінах у 2014–2019 рр.

Роки	Середньомісячні сукупні ресурси, грн, y	t	t^2	$y \cdot t$	Y, грн	$(y - Y)^2$
А	1	2	3	4	5	6
2014	3820,7	–5	25	$3820{,}7 \cdot (-5) = -19103{,}5$	$6554{,}92 + 641{,}91 \cdot 3345{,}4$	$(3820{,}7 - 3345{,}4)^2 = 225910{,}09$
2015	4441,7	–3	9	$4441{,}7 \cdot (-3) = -13325{,}1$	$6554{,}92 + 641{,}91 \cdot 4629{,}2$	$(4441{,}7 - 4629{,}2)^2 = 35156{,}25$
2016	5361,1	–1	1	5361,1	5913,0	304593,61
2017	7091,1	1	1	7091,1	7196,8	11172,49
2018	8721,0	3	9	26163,0	8480,6	57792,16
2019	9893,9	5	25	49469,5	9764,5	16744,36
Усього	39329,5	0	70	44933,9	39329,5	651368,96

Припускаючи, що умови, в яких формувалася тенденція середньомісячних сукупних ресурсів у розрахунку на одне домогосподарство області у 2014–2019 рр. найближчим часом не зміняться і враховуючи тривалість періоду, що прогнозується (для 2020 року $t = 7$, для 2021 року $t = 9$), визначимо прогнозний рівень середньомісячних сукупних ресурсів у розрахунку на одне домогосподарство на 2021 рік:

$Y_{2021} = 6554{,}95 + 641{,}91 \cdot 9 = 12332{,}14$ грн.

Визначимо прогнозні межі, скориставшись формулою $Y_{t+v} \pm t_{1-\alpha}(k) \cdot S_p$. Для визначення похибки прогнозу за формулою $S_p = \sigma_e \cdot \sqrt{\frac{n+1}{n} + \frac{3(n+2v-1)^2}{n(n^2-1)}}$ необхідно спочатку обчислити залишкове середнє квадратичне відхилення:

$$\sigma_e = \sqrt{\sigma_e^2} = \sqrt{\frac{1}{n-m} \cdot \sum (y - Y)^2} = \sqrt{\frac{651368{,}96}{6-2}} = 403{,}54$$

Похибка прогнозу для 2021 року дорівнюватиме:

$$S_p = 403{,}54 \cdot \sqrt{\frac{6+1}{6} + \frac{3(6+2\cdot 2-1)^2}{6(6^2-1)}} = 615{,}16$$

Критичне значення t-критерію для рівня істотності $\alpha = 0{,}05$ та числа ступенів вільності $6 - 2 = 4$ становить $t_{0{,}95}(4) = 2{,}13$ (табл. 8.2). Отже, $t_{1-\alpha}(k) \cdot S_{p\,2021} = 2{,}13 \cdot 615{,}16 = 1310{,}29$.

Довірчі межі прогнозного рівня для 2021 року становлять:

$$12332{,}14 \pm 1310{,}29 \text{ або } [11021{,}85;\ 13642{,}43].$$

Отже, за умови збереження комплексу причин, що формують тенденцію, з рівнем імовірності 0,95 розмір середньомісячні сукупні ресурси у розрахунку на одне домогосподарство області у 2021 році становитимуть не менше 11021,85 грн. і не перевищать 13642,43 грн.

Оцінимо якість обраного трендового рівняння з допомогою абсолютної похибки апроксимації $\overline{A} = \frac{\sum |y - Y| \div y}{n} 100\%$ (необхідні проміжні розрахунки наведено у табл. 8.11).

Таблиця 8.11

Динаміка середньомісячних сукупних ресурсів у розрахунку на одне домогосподарство області у фактичних цінах у 2014–2019 рр.

Роки	Середньомісячні сукупні ресурси, грн		$\lvert y - Y \rvert \div y$
	фактичні, y	теоретичні, Y	
2014	3820,7	3345,4	$\lvert 3820{,}7 - 3345{,}4 \rvert \div 3820 = 0{,}124$
2015	4441,7	4629,2	$\lvert 4441{,}7 - 4629{,}2 \rvert \div 4441 = 0{,}042$
2016	5361,1	5913,0	$\lvert 5361{,}1 - 5913{,}0 \rvert \div 5361 = 0{,}1029$
2017	7091,1	7196,8	0,0149
2018	8721,0	8480,6	0,0276
2019	9893,9	9764,5	0,0131
Усього	39329,5	39329,5	0,3251

Абсолютна похибка апроксимації становить:

$$\overline{A} = \frac{0{,}3251}{6} \cdot 100 = 5{,}42 .$$

Оскільки $\overline{A} < 10\%$, то обране трендове рівняння добре підібране і може використовуватися для опису тенденції зміни досліджуваного показника.

Задача 5

Використовуючи інформацію про помісячне споживання населенням електроенергії у 2016–2018 рр. (табл. 8.12), визначити індекси сезонності та середні квадратичні відхилення й описати сезонну хвилю графічно.

Таблиця 8.12

Помісячне споживання населенням електроенергії у 2016–2018 рр.

Місяць	Споживання населенням електроенергії, тис кВт		
	2016	2017	2018
А	1	2	3
1	30	31	32
2	29	30	30
3	27	28	29
4	26	27	27
5	25	26	25
6	23	22	23
7	22	21	22
8	24	23	23
9	26	26	25
10	28	29	27
11	30	30	29
12	31	32	30

Розв'язання задачі

Сезонні коливання описуються за допомогою моделі сезонної хвилі, яку утворює сукупність індексів сезонності I_c.

Індекси сезонності можуть бути розраховані для кожного року як відношення фактичних місячних рівнів y_t до середньомісячного рівня $\overline{y}$: $I_c = \frac{y_t}{\overline{y}} \cdot 100$ (графи 4–6 табл. 8.13).

Таблиця 8.13

Споживання населенням електроенергії та індекси сезонності

Місяць	Споживання електроенергії, тис. кВт			Індекс сезонності, %		
	2016	2017	2018	2016	2017	2018
А	1	2	3	4	5	6
1	30	31	32	$\frac{30}{26,8}\cdot 100 = 111,9$	$\frac{31}{27,1}\cdot 100 = 114,4$	$\frac{32}{26,8}\cdot 100 = 119,4$
2	29	30	30	$\frac{29}{26,8}\cdot 100 = 108,2$	$\frac{30}{27,1}\cdot 100 = 110,7$	$\frac{30}{26,8}\cdot 100 = 111,9$
3	27	28	29	100,7	103,3	108,2
4	26	27	27	97,0	99,6	100,7
5	25	26	25	93,3	95,9	93,3
6	23	22	23	85,8	81,2	85,8
7	22	21	22	82,1	77,5	82,1
8	24	23	23	89,6	84,9	85,8
9	26	26	25	97,0	95,9	93,3
10	28	29	27	104,5	107,0	100,7
11	30	30	29	111,9	110,7	108,2
12	31	32	30	115,7	118,1	111,9
Усього	321	325	322	–	–	–
Середнє значення	$\frac{321}{12} = 26,8$	$\frac{325}{12} = 27,1$	$\frac{322}{12} = 26,8$	–	–	–

Індекси сезонності показують, що пик споживання електроенергії у 2016 та 2017 роках припадав на грудень, а у 2018 році – на січень, а найменше споживання електроенергії щороку спостерігалось у липні. Наочно сезонні хвилі споживання електроенергії за кожний рік представлені на рис. 8.2.

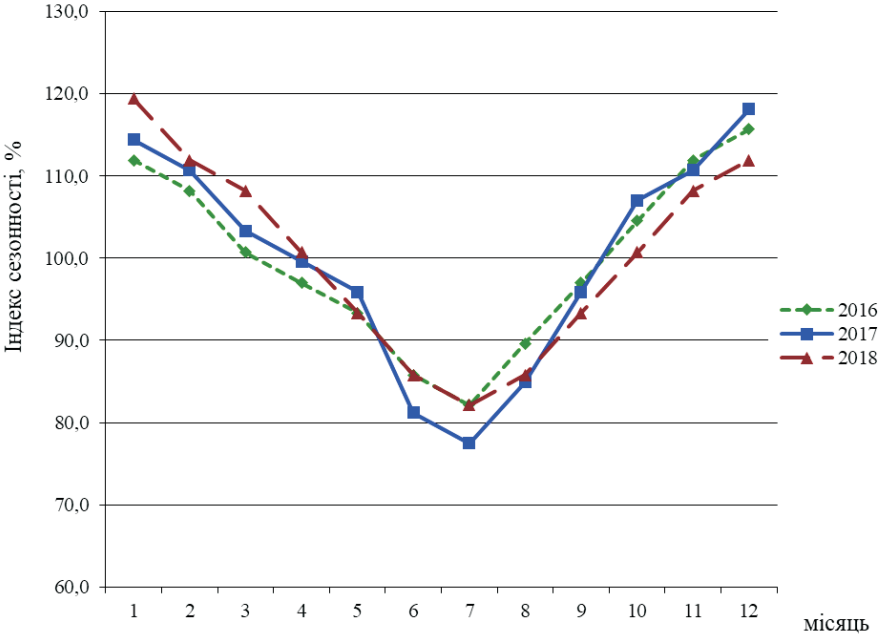


Рис. 8.2. Сезонні хвилі споживання населенням електроенергії у 2016–2018 рр.

Порівняємо інтенсивність сезонності споживання електроенергії за різні роки з допомогою середнього квадратичного відхилення $\sigma_c = \sqrt{\frac{1}{n}\sum_{1}^{n}(I_c - 100)^2}$. Необхідні розрахунки наведені у графах 4–6 табл. 8.14.

Таблиця 8.14

Розрахунок середнього квадратичного відхилення

Місяць	Індекси сезонності, %			$(I_c - 100)^2$		
	2016	2017	2018	2016	2017	2018 р.
А	1	2	3	4	5	6
1	111,9	114,4	119,4	$(111,9-100)^2 = 141,61$	$(114,4-100)^2 = 207,36$	$(119,4-100)^2 = 376,36$
2	108,2	110,7	111,9	$(108,2-100)^2 = 67,24$	$(110,7-100)^2 = 114,49$	$(111,9-100)^2 = 141,61$
3	100,7	103,3	108,2	0,49	10,89	67,24
4	97,0	99,6	100,7	9,00	0,16	0,49
5	93,3	95,9	93,3	44,89	16,81	44,89
6	85,8	81,2	85,8	201,64	353,44	201,64
7	82,1	77,5	82,1	320,41	506,25	320,41
8	89,6	84,9	85,8	108,16	228,01	201,64
9	97,0	95,9	93,3	9,00	16,81	44,89
10	104,5	107,0	100,7	20,25	49,00	0,49
11	111,9	110,7	108,2	141,61	114,49	67,24
12	115,7	118,1	111,9	246,49	327,61	141,61
Усього	_	_	_	1310,79	1945,32	1608,51

Середнє квадратичне відхилення:

- 2016 рік: $\sigma_c = \sqrt{\frac{1}{12} \cdot 1310,79} = 10,45$;
- 2017 рік: $\sigma_c = \sqrt{\frac{1}{12} \cdot 1945,32} = 12,73$;
- 2018 рік: $\sigma_c = \sqrt{\frac{1}{12} \cdot 1608,51} = 11,58$.

Отже, найбільш інтенсивними сезонні коливання споживання електроенергії були у 2017 році.

Крім того, індекси сезонності можуть бути розраховані у середньому за досліджуваний період. При цьому для кожного місяця визначається середній рівень $\overline{y_i}$ (графи 4, 5 табл. 8.15), а також середньомісячний рівень для всього періоду, що аналізується, $\overline{y}$ (останній рядок графи 5 табл. 8.15). За цими даними визначається індекс сезонності як відношення місячних середніх рівнів $\overline{y_i}$ до загального середньомісячного рівня $\overline{y}$: $I_c = \frac{\overline{y_i}}{\overline{y}} \cdot 100$ (графа 6 табл. 8.15).

Таблиця 8.15

Споживання електроенергії та індекси сезонності

Місяць	Споживання електроенергії, тис кВт			Загальне споживання електроенергії, тис. кВт	Середнє споживання електроенергії, тис. кВт	Індекс сезонності, %
	2016	2017	2018			
А	1	2	3	4	5	6
1	30	31	32	$30+31+32=93$	$\frac{93}{3}=31{,}0$	$\frac{31}{26{,}9}\cdot 100=115{,}2$
2	29	30	30	$29+30+30=89$	$\frac{89}{3}=29{,}7$	$\frac{29{,}7}{26{,}9}\cdot 100=110{,}4$
3	27	28	29	$27+28+29=84$	$\frac{84}{3}=28{,}0$	$\frac{28}{26{,}9}\cdot 100=104{,}1$
4	26	27	27	80	26,7	99,3
5	25	26	25	76	25,3	94,1
6	23	22	23	68	22,7	84,4
7	22	21	22	65	21,7	80,7
8	24	23	23	70	23,3	86,6
9	26	26	25	77	25,7	95,5
10	28	29	27	84	28,0	104,1
11	30	30	29	89	29,7	110,4
12	31	32	30	93	31,0	115,2
Усього	321	325	322	968	–	–
Середнє значення	–	–	–	–	$\frac{968}{3\cdot 12}=26{,}9$	–

Індекси сезонності показують, що пік споживання електроенергії у 2016–2018 роках припадав на січень та грудень, а найменше споживання електроенергії було у липні. Наочно сезонну хвилю споживання електроенергії за досліджуваний період представлено на рис. 8.3.

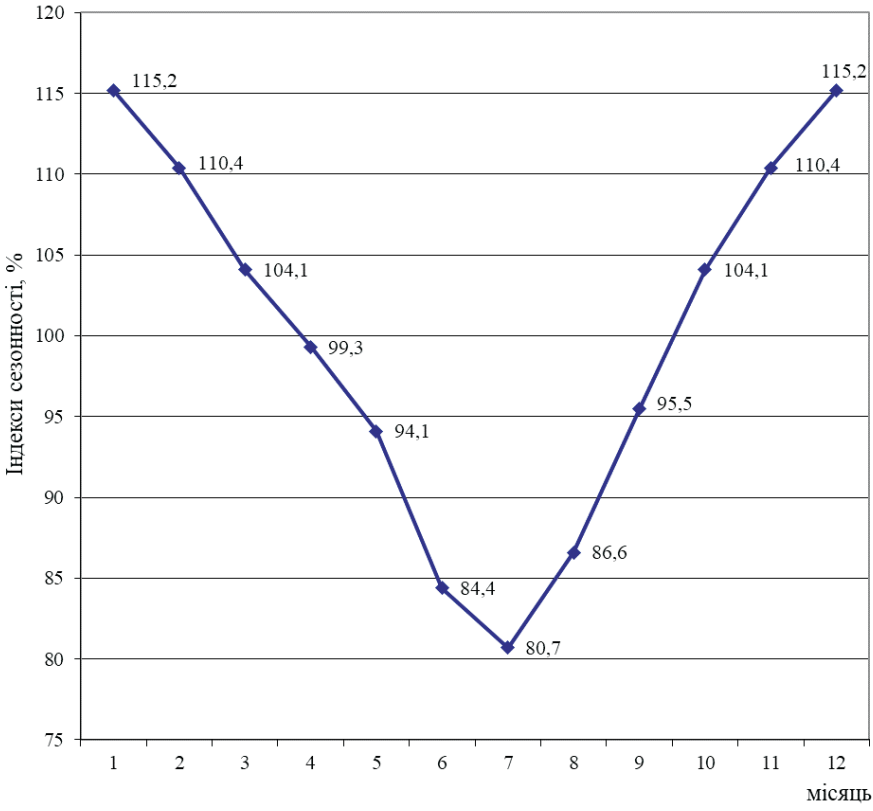


Рис. 8.3. Сезонна хвиля споживання населенням електроенергії в середньому за 2016—2018 рр.

Запитання для самоконтролю

1. Що називають рядом динаміки? З яких елементів він складається?

2. Яких умов необхідно дотримуватися при побудові динамічних рядів?

3. Які види рядів динаміки ви знаєте та наведіть їх приклади?

4. Які аналітичні показники ряду динаміки ви знаєте?

5. Який взаємозв'язок існує між ланцюговими та базисними показниками динаміки?

6. Як розраховують середній рівень моментного та інтервального рядів динаміки?

7. Як розраховують середній абсолютний приріст, середній коефіцієнт зростання та середній темп приросту?

8. Які методи виявлення основної тенденції розвитку ви знаєте? У чому їх особливість?

9. У чому полягає суть аналітичного вирівнювання динамічного ряду?

10. За допомогою яких показників досліджують сезонні коливання?

Завдання для самостійної роботи

Тестові завдання

1. Ланцюгові показники ряду динаміки розраховуються при порівнянні:

а) кожного наступного рівня ряду з попереднім;
б) першого рівня ряду з кожним наступним рівнем;
в) кожного рівня ряду з одним і тим самим рівнем, прийнятим за базу порівняння;
г) останнього рівня ряду з попередніми.

Відповіді: 1) а; 2) б; 3) в; 4) г.

2. Ряди динаміки відображають:

а) структуру сукупності за будь-якою ознакою;
б) кількісну послідовність показників;
в) хронологічну послідовність показників у сукупності;

г) підсумок значень показників сукупності за певний проміжок часу.

Відповіді: 1) а; 2) б; 3) в; 4) г.

3. Значення коефіцієнта зростання може:

а) бути величиною від'ємною;
б) бути величиною додатною;
в) дорівнювати нулю.

Відповіді: 1) а; 2) б; 3) в.

4. Показник, що характеризує величину зміни рівня ряду за певний проміжок часу і розраховується як різниця між двома рівнями динамічного ряду називається:

а) абсолютним приростом;
б) темпом зростання;
в) коефіцієнтом зростання;
г) темпом приросту.

Відповіді: 1) а; 2) б; 3) в; 4) г.

5. Базисний абсолютний приріст дорівнює:

а) сумі ланцюгових абсолютних приростів;
б) добутку ланцюгових абсолютних приростів;
в) сумі ланцюгових коефіцієнтів зростання;
г) добутку ланцюгових коефіцієнтів зростання.

Відповіді: 1) а; 2) б; 3) в; 4) г.

6. Базисний коефіцієнт зростання дорівнює:

а) сумі ланцюгових абсолютних приростів;
б) добутку ланцюгових абсолютних приростів;
в) сумі ланцюгових коефіцієнтів зростання;
г) добутку ланцюгових коефіцієнтів зростання.

Відповіді: 1) а; 2) б; 3) в; 4) г.

7. У лінійному рівнянні тренду $Y = a + bt$ параметр a характеризує:

а) рівень ряду динаміки при $t = 0$;
б) середній річний абсолютний приріст;

в) середній річний темп приросту;

г) середній річний рівень ряду динаміки.

Відповіді: 1) а; 2) б; 3) в; 4) г.

8. За формулою $\bar{y} = \frac{\frac{y_1 + y_n}{2}}{n-1} + \sum_{2}^{n-1} y_t$ визначається середній рівень динамічного ряду:

а) для інтервальних рядів з однаковими за тривалістю періодами часу;

б) для моментних рядів з однаковими за тривалістю періодами часу;

в) для інтервальних рядів з неоднаковими за тривалістю періодами часу;

г) для моментних рядів з неоднаковими за тривалістю періодами часу.

Відповіді: 1) а; 2) б; 3) в; 4) г.

9. Індекс сезонності розраховується як:

а) ковзна середня з інтервалом згладжування $m < n$;

б) різниця між мінімальним та максимальним значеннями досліджуваного явища;

в) відношення y_t до середнього рівня за цикл;

г) відношення мінімального значення y_t до середнього рівня за цикл.

Відповіді: 1) а; 2) б; 3) в; 4) г.

10. Числа послідовності $y_1, y_2, y_3, \ldots, y_n$, що характеризують зміну явища або рівень явища на певний момент часу, називаються:

а) переліком хронологічних дат;

б) інтервалами часу;

в) рівнями ряду.

Відповіді: 1) а; 2) б; 3) в.

11. Параметри лінійної функції $Y = a + bt$ визначаються за формулами:

а) $b = \frac{\sum y}{\sum t^2}$;

б) $a = \frac{\sum y}{n}$; $b = \frac{\sum yt}{\sum t^2}$;

в) $b = \frac{\sum y}{n}$.

Відповіді: 1) а; 2) б; 3) в.

12. Похибка прогнозу розраховується за формулою:

а) $\sigma_e \cdot \sqrt{\frac{n+1}{n} + \frac{3(n+2v-1)^2}{n(n^2-1)}}$;

б) $\sqrt{\frac{n+1}{n} + \frac{3(n+2v-1)^2}{n(n^2-1)}}$;

в) $\sqrt{\frac{n+1}{n} + \frac{3(n+2v-1)^2}{n(n^2-1)}}$;

г) $\sqrt{\frac{1}{n-m} \cdot \sum (y-Y)^2}$.

Відповіді: 1) а; 2) б; 3) в; 4) г.

Практичні завдання (задачі)

13. Маємо такі дані про динаміку середньомісячної номінальної заробітної плати штатних працівників, грн:

Роки	2016	2017	2018	2019
Зарплата, грн	5183	7104	8865	10497

Визначте: а) ланцюгові абсолютні прирости; б) середні значення показників динаміки: абсолютного приросту, темпу приросту. Зробіть висновки.

14. Кількість зареєстрованих безробітних у регіоні на 1 січня 2012–2019 рр. характеризується даними:

Роки	2012	2013	2014	2015	2016	2017	2018	2019
Кількість зареєстрованих безробітних, тис. осіб	15,6	15,8	16,1	19,6	20,1	15,6	13,5	12,3

Визначте: 1) абсолютні та відносні характеристики динаміки кількості зареєстрованих безробітних; 2) середні характеристики динаміки кількості зареєстрованих безробітних. Зробіть висновки.

15. Є такі дані про роздрібний товарооборот району, млн грн:

Роздрібний товарооборот	Рік					
	2014	2015	2016	2017	2018	2019
У старих межах	520	540	600	–	–	–
У нових межах	–	–	750	792	810	835

Для аналізу динаміки товарообігу: 1) приведіть ряди динаміки до порівнянного вигляду (зімкніть ряди); 2) визначте: середньорічний товарообіг; середньорічні показники абсолютного приросту, темпів зростання та темпів приросту. Зробіть висновки.

16. Дані щодо перевезення вантажів автомобільним транспортом в області у 2012–2018 рр. наведені в таблиці:

Роки	2012	2013	2014	2015	2016	2017	2018
Перевезення вантажів автотранспортом, млн тонн	10,1	9,5	8,9	9,1	9,2	8,7	8,8

Охарактеризуйте тенденцію розвитку з допомогою трендового рівняння. Визначте обсяг перевезення вантажів автомобільним транспортом через 3 роки.

17. Динаміка середньорічної чисельності наявного населення області у 2013–2019 рр. характеризується даними:

Роки	Середньорічна чисельність наявного населення, тис. осіб
2013	3 300,1
2014	3 284,5
2015	3 265,8
2016	3 242,6
2017	3 230,8
2018	3 218,8
2019	3 191,6

Охарактеризуйте тенденцію розвитку з допомогою трендового рівняння. Визначте середньорічну чисельність наявного населення області через 3 роки.

18. Маємо таку інформацію про кількість шлюбів у 2019 році в Україні:

Місяць	Кількість шлюбів		
	Міська та сільська місцевість	Міська місцевість	Сільська місцевість
Січень	10928	8172	2756
Лютий	13710	10473	3237
Березень	13882	10984	2898
Квітень	10706	8664	2042

Травень	16659	12322	4337
Червень	25813	20761	5052
Липень	23927	19014	4913
Серпень	29095	22918	6177
Вересень	29176	23041	6135
Жовтень	23334	17697	5637
Листопад	21233	16249	4984
Грудень	19395	15252	4143

Для кожного ряду визначте індекси сезонності, амплітуду коливань та середнє квадратичне відхилення. Опишіть сезонні хвилі графічно. Зробіть висновки.

19. Маємо таку інформацію про кількість народжених і померлих у 2019 році в Україні:

Місяць	Кількість народжених, осіб	Кількість померлих, осіб
Січень	27947	62196
Лютий	23839	47891
Березень	24782	49177
Квітень	23108	47879
Травень	27777	53791
Червень	22991	43396
Липень	29916	47930
Серпень	28417	42359
Вересень	25768	41914
Жовтень	29248	50245
Листопад	24624	46986
Грудень	22188	47350

Для кожного ряду визначте індекси сезонності й опишіть сезонні хвилі графічно. За допомогою узагальнюючих показників сезонних коливань порівняйте сезонність. Зробіть висновки.

20. Використовуючи взаємозв'язок характеристик динаміки:
а) заповніть таблицю;
б) визначте середній абсолютний приріст перевезення вантажів за цей період.

Зробіть висновки.

Рік	Перевезено вантажів, млн т	Ланцюгові характеристики динаміки:			
		абсолютний приріст, млн т	коефіцієнт зростання	темп приросту, %	абсолютне значення 1% приросту, млн т
2017	300	–	–	–	–
2018			1,1		
2019		20			

21. Використовуючи взаємозв'язок характеристик динаміки визначте середньомісячну заробітну плату штатного працівника у промисловості за всі роки. Заповніть таблицю, дорахувавши характеристики динаміки. Розрахуйте середні рівні характеристик динаміки середньомісячної заробітної плати штатного працівника за 2012–2016 рр. Зробіть висновки щодо тенденції зміни середньомісячної заробітної плати.

Рік	Середньомісячна заробітна плата, грн	Абсолютний приріст, грн		Коефіцієнт зростання		Темп приросту, %	
		ланцюговий	базисний	ланцюговий	базисний	ланцюговий	базисний
2014		–	–	–	–	–	–
2015	4789						
2016			1914			23,24	
2017		1729					
2018					2,4155		
2019							195,59
2020				1,0824			

22. Використовуючи наведену нижче інформацію про кількість вакансій та середній запропонований розмір заробітної плати станом на початок місяця, визначте середні значення показників за 1 квартал та І півріччя.

Місяць	Січень	Лютий	Березень	Квітень	Травень	Червень	Липень
Кількість вакансій, од.	59018	68081	76514	64057	53099	58591	57927
Розмір запропонованої зарплати, грн	6662	6724	6735	6645	6889	6839	6927

Розділ 9.
ІНДЕКСИ

9.1. Поняття і способи побудови індексів

Індекс – відносний статистичний показник, що характеризує відношення в часі (динамічний індекс) або в просторі (територіальний індекс) рівнів суспільних явищ і процесів. При цьому зіставляються значення однойменних показників, що мають однаковий економічний зміст. Отже, індекс – це співвідношення двох однойменних показників, де в чисельнику наводиться показник, зміну якого досліджуємо, а у знаменнику – показник, з яким здійснюється порівняння (база порівняння або базисний показник). В індексах динаміки базисним може бути показник попереднього періоду або періоду, взятого за базу порівняння; в індексах порівняння у просторі базою може бути показник, що прийнятий за еталон або стосується якоїсь території.

Усі індекси, що використовуються в статистиці, за ступенем охоплення елементів сукупності поділяються на індивідуальні (елементарні) та зведені (складні). Індивідуальний індекс (i) – це відносна величина, що характеризує співвідношення рівнів показника для окремих елементів сукупності. Зведений індекс (I) – співвідношення рівнів показника, до складу якого входять різнорідні елементи. Якщо сукупність, що вивчається, складається з кількох груп, то можуть бути визначені зведені індекси групові (субіндекси) та зведений індекс щодо всієї сукупності (загальний або тотальний).

За формою індекси поділяються на агрегатні та середньозважені. Вибір форми індексу залежить від мети дослідження і наявної інформації.

Для порівняння рівнів середніх величин використовують систему індексів середніх величин: індекс змінного складу, індекс фіксованого складу, індекс структурних зрушень.

Методика розрахунку індексу залежить від мети дослідження, статистичної природи показника, ступеня агрегування інформації. З допомогою індексів вивчається динаміка явищ, здійснюються різні регіональні та пооб'єктні порівняння. Застосування індексів дозволяє оцінити вплив окремих факторів на динаміку показників, що вивчаються, як у відносних, так і в абсолютних вимірниках.

Показник, динаміку чи співвідношення якого характеризує індекс, називають індексованою величиною. При побудові індексів необхідно дотримуватися таких загальноприйнятих у теорії та практиці статистики умовних позначень:

1. Показники, зміна яких вивчається за допомогою індексів, мають такі (найбільш вживані) умовні позначення (табл. 9.1):

Таблиця 9.1

Основні умовні позначення показників

Назва показника	Умовне позначення
Кількість (або фізичний обсяг) продукції, товару чи послуги в натуральному вираженні	q
Ціна одиниці продукції, товару чи послуги	p
Собівартість одиниці продукції, товару чи послуги	z
Трудомісткість	t
Продуктивність праці (виробіток)	w
Кількість працівників	T
Урожайність	$в$
Розмір посівної площі	s
Вартість продукції, товарів чи послуг	pq
Загальна собівартість продукції, товарів чи послуг	zq
Загальні витрати робочого часу на виробництво продукції, товарів чи послуг	tq
Валовий збір	$вs$

2. Показники базисного періоду мають підрядковий знак «0», а звітного – «1». Наприклад, собівартість продукції базового періоду – z_0, звітного – z_1, вартість продукції базисного періоду – $P_0 q_0$, звітного – $P_1 q_1$. Якщо зміна явища досліджується більш ніж за два періоди, то вони позначаються відповідно «0», «1», «2», «3» і т. д.

3. Індекси (як індивідуальні, так і зведені) мають підрядкове позначення, що вказує на індексовану величину. Наприклад, індивідуальний індекс собівартості – i_z, зведений індекс вартості продукції – I_{pq}.

Методологічні принципи побудови індивідуальних і зведених індексів розглянемо на прикладі індексів динаміки. Індивідуальний індекс динаміки характеризує зміну в часі якогось індивідуального явища, наприклад рівня собівартості певного виду продукції (товару чи послуги): $i_z = \frac{z_1}{z_0}$. Із використанням наведених вище умовних позначень може бути побудована низка індивідуальних індексів (табл. 9.2).

Зведений індекс характеризує зміну у часі рівнів показника, до складу якого входять різнорідні елементи (наприклад, вартість товарів, реалізованих у супермаркеті). Тому порівняно з індивідуальним індексом методика побудови та розрахунку зведеного індексу є складнішою.

Таблиця 9.2

Формули розрахунку індивідуальних індексів

Назва індивідуального індексу	Формула розрахунку
1	2
Індекс кількості продукції, товару чи послуги в натуральному вираженні (індекс фізичного обсягу)	$i_q = \frac{q_1}{q_0}$
Індекс ціни одиниці продукції, товару чи послуги	$i_p = \frac{p_1}{p_0}$
Індекс собівартості одиниці продукції, товару чи послуги	$i_z = \frac{z_1}{z_0}$
Індекс трудомісткості	$i_t = \frac{t_1}{t_0}$
Індекс продуктивності праці (виробітку)	$i_w = \frac{w_1}{w_0}$
Індекс кількості працівників	$i_T = \frac{T_1}{T_0}$
Індекс урожайності	$i_y = \frac{y_1}{y_0}$

Продовження табл.9.2

1	2
Індекс розміру посівної площі	$i_s = \frac{s_1}{s_0}$
Індекс вартості продукції, товару чи послуги	$i_{pq} = \frac{p_1 q_1}{p_0 q_0}$
Індекс загальної собівартості продукції, товару чи послуги	$i_{zq} = \frac{z_1 q_1}{z_0 q_0}$
Індекс загальних витрати робочого часу на виробництво продукції, товару чи послуги	$i_{tq} = \frac{t_1 q_1}{t_0 q_0}$
Індекс валового збору	$i_{ys} = \frac{в_1 s_1}{в_0 s_0}$

Слід виділяти екстенсивний (кількісний) та інтенсивний (якісний) фактори. Зокрема серед розглянутих у табл. 9.1 виділимо такі фактори (табл. 9.3).

Таблиця 9.3

Класифікація факторів

Інтенсивні	Екстенсивні
Ціна одиниці продукції, товару чи послуги, *p*	Кількість (або фізичний обсяг) продукції, товару чи послуги в натуральному вираженні, *q*
Собівартість одиниці продукції, товару чи послуги, *z*	Кількість працівників, *T*
Трудомісткість, *t*	Розмір посівної площі, *s*
Продуктивність праці (виробіток), *w*	
Урожайність, *в*	

Зведений індекс, що характеризує зміну складного явища в цілому, визначається за формулою:

$$I_{xv} = \frac{\sum x_1 v_1}{\sum x_0 v_0},$$

де x_0 та x_1 – значення інтенсивного показника у базисному та звітному періодах; v_0 та v_1 – значення екстенсивного показника у базисному та звітному періодах.

Наведена формула розрахунку індексу називається агрегатною.

Також може бути окремо оцінений вплив на зміну складного явища змін екстенсивного та інтенсивного факторів. При цьому слід дотримуватися правила, прийнятого у статистичній практиці: інтенсивні фактори-співмножники фіксуються на рівні базисного періоду (базисно-зважена система (Ласпейреса)), а екстенсивні – на рівні звітного періоду (поточно-зважена система (Пааше)). Тоді зведений індекс інтенсивного (якісного) показника матиме вигляд (формула Пааше):

$$I_x = \frac{\sum x_1 v_1}{\sum x_0 v_1},$$

а зведений індекс екстенсивного (кількісного) показника (формула Ласпейреса):

$$I_v = \frac{\sum x_0 v_1}{\sum x_0 v_0}.$$

При побудові як індивідуальних, так і зведених індексів необхідно пам'ятати, що:

- індекси, що характеризують зміну явищ, поєднаних між собою як співмножники, мають такий зв'язок: добуток індексів співмножників дорівнює індексу добутку (наприклад, $i_{pq} = i_p \cdot i_q$, $I_{xv} = I_x \cdot I_v$);
- величина, обернена індексу прямого показника дорівнює індексу оберненого йому показника (наприклад, $i_w = \frac{1}{i_t}$, якщо досліджується зміна щодо одного об'єкта, або $I_w = \frac{1}{I_t}$, якщо досліджується зміна за групою об'єктів).

9.2. Середньозважені індекси

Крім агрегатної форми зведеного індексу, в окремих випадках (коли зміну складного явища неможливо вивчити з її допомогою через відсутність інформації про рівні окремо екстенсивного та інтенсивного показників у базисному та звітному періодах) використовують іншу форму – середньозважені індекси: за формулою середньої арифметичної зваженої і за формулою середньої гармонічної зваженої.

Вибір форми індексу залежить від мети дослідження і наявних даних. Зокрема, за формулою середньої арифметичної зваженої визначається середньозважений індекс у випадках, коли необхідно охарактеризувати зміну екстенсивного (кількісного) показника в середньому за рядом різнорідних елементів, маючи інформацію про базисний рівень досліджуваного показника $x_0 v_0$ за окремими елементаи сукупності (наприклад, вартість виготовленої продукції, витрати на виробництво продукції) та індивідуальні індекси екстенсивного показника i_v (наприклад, зміну фізичного обсягу (кількості) продукції):

$$I_v = \frac{\sum i_v \cdot x_0 v_0}{\sum x_0 v_0}.$$

Формула середньої гармонічної зваженої використовується для обчислення загального індексу інтенсивного (якісного) показника у випадках, коли наявна інформація про звітний рівень показника $x_1 v_1$ за окремими елементами сукупності (наприклад, вартість виготовленої продукції, витрати на виробництво продукції) та індивідуальні індекси інтенсивного показника i_x (наприклад, зміну ціни одиниці виготовленої продукції або зміну собівартості одиниці продукції):

$$I_x = \frac{\sum x_1 v_1}{\sum \frac{x_1 v_1}{i_x}}.$$

9.3. Система взаємопов'язаних індексів та визначення впливу окремих факторів

У процесі аналізу динаміки суспільних явищ і процесів виникає потреба у визначенні впливу окремих чинників на зміну результативної ознаки, наприклад впливу зміни рівня собівартості та фізичного обсягу продукції на зміну загальних витрат. Одним зі способів вирішення цієї проблеми є застосування системи взаємопов'язаних індексів. При цьому оцінити вплив окремих чинників можна як у відносному, так і в абсолютному вираженні, обчисливши індекси факторних показників відповідної системи взаємозв'язаних індексів. Як зазначалось у п. 9.2, добуток індексів співмножників дорівнює індексу добутку, тобто двофакторні зведені індекси поєднані між собою так:

– індекси, що характеризують динаміку індивідуального явища: $i_{xv} = i_x \cdot i_v$;
– індекси, що характеризують зміну складного явища, тобто є співвідношенням рівнів показника, до складу якого входять різнорідні елементи: $I_{xv} = I_x \cdot I_v$.

Отже, $\frac{\sum x_1 v_1}{\sum x_0 v_0} = \frac{\sum x_1 v_1}{\sum x_0 v_1} \cdot \frac{\sum x_0 v_1}{\sum x_0 v_0}$. У цьому випадку відносний вплив зміни інтенсивного (якісного) показника визначається як:

$$I_x = \frac{\sum x_1 v_1}{\sum x_0 v_1},$$

а відносний вплив зміни екстенсивного (кількісного) показника:

$$I_v = \frac{\sum x_0 v_1}{\sum x_0 v_0}.$$

Абсолютна зміна індивідуального явища за рахунок зміни окремих факторів визначається як різниця між чисельником і знаменником відповідного факторного індексу. Зокрема, загальна абсолютна зміна визначається як:

$$\Delta_{xv} = x_1 v_1 - x_0 v_0.$$

Відповідно, абсолютна зміна за рахунок зміни інтенсивного фактора:

$$\Delta_x = x_1 v_1 - x_0 v_1,$$

а за рахунок екстенсивного фактора:

$$\Delta_v = x_0 v_1 - x_0 v_0.$$

Отже, $\Delta_{xv} = \Delta_x + \Delta_v$.

Для дослідження абсолютної зміни складного явища, а також впливу зміни інтенсивного (x) та екстенсивного (v) факторів використовують такі формули:

– абсолютна загальна зміна:

$$\Delta_{xv} = \sum x_1 v_1 - \sum x_0 v_0;$$

– абсолютна зміна за рахунок зміни інтенсивного фактора:

$$\Delta_x = \sum x_1 v_1 - \sum x_0 v_1;$$

– абсолютна зміна за рахунок зміни екстенсивного фактора

$$\Delta_v = \sum x_0 v_1 - \sum x_0 v_0$$

Показник абсолютної зміни може набувати як додатних, так і від'ємних значень. Якщо абсолютна зміна має знак «+», то це означає, що відбулось абсолютне збільшення (приріст) досліджуваного показника; знак «–» вказує на абсолютне зменшення (скорочення) досліджуваного показника.

9.4. Індекси середніх величин, їх взаємозв'язок

У статистичній практиці при дослідженні сукупностей, що складаються з об'єктів одного й того самого типу, поряд зі зведеними індексами використовуються індекси середніх величин, що дозволяють не лише дослідити динаміку середніх величин, а й виявити вплив факторів на неї.

Як відомо, середня величина визначається за формулою

$$\bar{x} = \frac{\sum x_j \nu_j}{\sum \nu_j},$$

де x_j – значення досліджуваної ознаки (інтенсивний фактор); ν_j – частота (екстенсивний фактор).

Виходячи із запропонованих нами позначень інтенсивного й екстенсивного факторів формула середньої величини може бути представлена так:

$$\bar{x} = \frac{\sum xv}{\sum v} \text{ або } \bar{x} = \sum xd,$$

де d – частка j-ї складової досліджуваної сукупності.

Отже, на зміну середнього рівня досліджуваного показника впливають:

- ❑ зміна інтенсивного фактора – значень ознаки x,
- ❑ зміна екстенсивного фактора – структури сукупності.

Оцінити вплив кожного з цих факторів на динаміку середнього рівня досліджуваного (інтенсивного) показника можна за допомогою системи індексів середнього рівня інтенсивного показника: змінного складу, фіксованого складу, структурних зрушень.

Індекс змінного складу ($I_{\bar{x}}$) характеризує зміну середнього рівня досліджуваного показника у звітному періоді порівняно з базисним

за рахунок впливу двох факторів: осереднюваного показника для окремих одиниць сукупності та структури досліджуваної сукупності.

Для будь-яких інтенсивних показників індекс змінного складу має вигляд:

$$I_{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 v_1}{\sum v_1} \div \frac{\sum x_0 v_0}{\sum v_0}, \text{ або } I_{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 d_1}{\sum x_0 d_0}.$$

Індекс фіксованого складу (I_x) враховує зміну тільки індексованої величини і показує, як у звітному періоді порівняно з базисним змінився середній рівень досліджуваного показника під впливом зміни лише осереднюваного показника для окремих одиниць сукупності:

$$I_x = \frac{\sum x_1 v_1}{\sum v_1} \div \frac{\sum x_0 v_1}{\sum v_1}, \text{ або } I_x = \frac{\sum x_1 d_1}{\sum x_0 d_1}.$$

Індекс структурних зрушень (I_d) характеризує вплив зміни структури досліджуваного явища на динаміку середнього рівня індексованого показника і розраховується за формулою:

$$I_d = \frac{\sum x_0 v_1}{\sum v_1} \div \frac{\sum x_0 v_0}{\sum v_0}, \text{ або } I_d = \frac{\sum x_0 d_1}{\sum x_0 d_0}.$$

Індекс структурних зрушень показує, як у звітному періоді порівняно з базисним змінився середній рівень досліджуваного показника під впливом зміни лише структури досліджуваної сукупності.

Між розглянутими індексами існує такий зв'язок:

$$\left(\frac{\sum x_1 v_1}{\sum v_1} \div \frac{\sum x_0 v_0}{\sum v_0}\right) = \left(\frac{\sum x_1 v_1}{\sum v_1} \div \frac{\sum x_0 v_1}{\sum v_1}\right) \cdot \left(\frac{\sum x_0 v_1}{\sum v_1} \div \frac{\sum x_0 v_0}{\sum v_0}\right),$$

тобто $I_{\bar{x}} = I_x \cdot I_d$.

Крім відносної зміни середнього показника та впливу на нього окремих факторів, можна визначити й абсолютну зміну середнього рівня досліджуваного показника:

– загальну:

$$\Delta_{\bar{x}} = \bar{x}_1 - \bar{x}_0 = \frac{\sum x_1 v_1}{\sum v_1} - \frac{\sum x_0 v_0}{\sum v_0} = \sum x_1 d_1 - \sum x_0 d_0;$$

– за рахунок зміни лише осереднюваного показника для окремих одиниць сукупності:

$$\Delta_x = \frac{\sum x_1 v_1}{\sum v_1} - \frac{\sum x_0 v_1}{\sum v_1} = \sum x_1 d_1 - \sum x_0 d_1 ;$$

– за рахунок зміни лише структури досліджуваної сукупності:

$$\Delta_d = \frac{\sum x_0 v_1}{\sum v_1} - \frac{\sum x_0 v_0}{\sum v_0} = \sum x_0 d_1 - \sum x_0 d_0 .$$

Загальну зміну середнього рівня досліджуваного показника можна також визначити як

$$\Delta_{\bar{x}} = \Delta_x + \Delta_d .$$

Показники абсолютної зміни середнього рівня досліджуваного показника можуть набувати як додатних, так і від'ємних значень: абсолютна зміна має знак «+», якщо відбулось абсолютне збільшення (приріст) досліджуваного показника, а знак «–» вказує на абсолютне зменшення (зниження) досліджуваного показника.

Розв'язання типових задач

Задача 1

Маємо такі дані:

Вид продукції	Кількість виробленої продукції, шт.		Ціна одиниці продукції, грн/шт.	
	базисний період, q_0	звітний період, q_1	базисний період, P_0	звітний період, P_1
А	378	400	15,0	18,0
Б	448	500	12,0	16,5

Визначити:

1) індивідуальні індекси фізичного обсягу, ціни та вартості виробленої продукції;
2) зведені (загальні) індекси ціни, фізичного обсягу виробництва та вартості виробленої продукції.

Розв'язання задачі

1) Індивідуальний індекс характеризує зміну в часі певного індивідуального явища (фізичного обсягу, ціни чи вартості виробленої продукції) і розраховується за кожним видом продукції:

❑ за продукцією А:

– індекс фізичного обсягу: $i_q^A = \frac{q_1}{q_0} = \frac{400}{378} = 1{,}058$, або 105,8% (+5,8%);

– індекс ціни одиниці продукції: $i_p^A = \frac{p_1}{p_0} = \frac{18{,}0}{15{,}0} = 1{,}200$, або 120,0% (+20,0%);

– індекс вартості продукції: $i_{pq}^A = \frac{p_1 q_1}{p_0 q_0} = \frac{18{,}0 \cdot 400}{15{,}0 \cdot 378} = \frac{7200}{5670} = 1{,}270$, або 127,0% (+27,0%),

при цьому $i_{pq}^A = i_p^A \cdot i_q^A = 1{,}200 \cdot 1{,}058 = 1{,}270$.

❑ за продукцією Б:

– індекс фізичного обсягу: $i_q^Б = \frac{q_1}{q_0} = \frac{500}{448} = 1{,}116$, або 111,6% (+11,6%);

– індекс ціни одиниці продукції: $i_p^Б = \frac{p_1}{p_0} = \frac{16{,}5}{12{,}0} = 1{,}375$, або 137,5% (+37,5%);

– індекс вартості продукції: $i_{pq}^Б = \frac{p_1 q_1}{p_0 q_0} = \frac{16{,}5 \cdot 500}{12{,}0 \cdot 448} = \frac{8250}{5376} = 1{,}535$, або 153,5% (+53,5%),

при цьому $i_{pq}^Б = i_p^Б \cdot i_q^Б = 1{,}375 \cdot 1{,}116 = 1{,}535$.

2) Зведений індекс цін становитиме

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{18{,}0 \cdot 400 + 16{,}5 \cdot 500}{15{,}0 \cdot 400 + 12{,}0 \cdot 500} = \frac{15450}{12000} = 1{,}288 \ \ (+28{,}8\%).$$

У звітному періоді порівняно з базисним ціни за двома видами продукції у середньому збільшилися на 28,8% .

Загальний індекс фізичного обсягу

$$I_q = \frac{\sum p_0 q_1}{\sum p_0 q_0} = \frac{15{,}0 \cdot 400 + 12{,}0 \cdot 500}{15{,}0 \cdot 378 + 12{,}0 \cdot 448} = \frac{12000}{11046} = 1{,}086 \ \ (+8{,}6\%).$$

У звітному періоді порівняно з базисним фізичний обсяг виробленої продукції за двома її видами у середньому збільшився на 8,6% .

Зведений індекс вартості виготовленої продукції

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{18{,}0 \cdot 400 + 16{,}5 \cdot 500}{15{,}0 \cdot 378 + 12{,}0 \cdot 448} = \frac{15450}{11046} = 1{,}399 \ (+39{,}9\%)$$

або $I_{pq} = I_p \cdot I_q = 1{,}288 \cdot 1{,}086 = 1{,}399$.

Вартість виготовленої продукції у звітному періоді порівняно з базисним збільшилась на 39,9%. Це сталося за рахунок збільшення як ціни, так і кількості виробленої продукції.

Задача 2

Маємо такі дані про продажі продукції в магазині:

Вид продукції	Товарооборот у фактичних (поточних) цінах, тис. грн		Зміни фізичного обсягу продажів у звітному періоді порівняно з базисним, %
	базисний період, $p_0 q_0$	звітний період, $p_1 q_1$	
А	149,0	156,3	+1,5
Б	182,5	176,8	–7,2

Визначити загальний індекс фізичного обсягу.

Розв'язання задачі

Загальний індекс фізичного обсягу визначається за формулою $I_q = \frac{\sum p_0 q_1}{\sum p_0 q_0}$.

У таблиці відсутні дані, необхідні для визначення загального індексу безпосередньо в агрегатній формі, а саме: інформація за кожним видом продукції про ціну (p) та кількість (q) реалізованого товару у базисному та звітному періодах, однак відомі індивідуальні індекси фізичного обсягу i_q й відповідна їм вага (товарооборот у фактичних цінах за базисний період $p_0 q_0$**)**. Отже для розрахунку загального індексу фізичного обсягу слід скористатися формулою середньої арифметичної зваженої:

$$I_q = \frac{\sum i_q \cdot p_0 q_0}{\sum p_0 q_0},$$

де i_q– індивідуальні індекси фізичного обсягу, які у цьому прикладі визначаються на підставі інформації про зміну фізичного обсягу продажів у звітному періоді порівняно з базисним.

Вид продукції	Зміни фізичного обсягу продажів у звітному періоді порівняно з базисним, %	Індивідуальні індекси фізичного обсягу, i_q
А	+1,5	$\frac{100+1,5}{100}=1,015$
Б	–7,2	$\frac{100-7,2}{100}=0,928$

Тоді $I_q=\frac{1,015\cdot 149,0+0,928\cdot 182,5}{149\ ,0\ +182\ ,5}=\frac{320,60}{331,5}=0,967$ (–3,3%).

У звітному періоді порівняно з базисним фізичний обсяг товарообороту за двома видами продукції у середньому зменшилися на 3,3%.

Задача 3

Маємо такі дані щодо магазину:

Тканини	Товарооборот у фактичних цінах, тис. грн		Зміни ціни у звітному періоді порівняно з базисним, %
	базисний період, p_0q_0	звітний період, p_1q_1	
Бавовняні	14,9	15,3	+2
Шовкові	180,0	177,6	–4

Визначити загальний індекс цін. Скориставшись правилом взаємозв'язку залежних індексів, визначити загальний індекс фізичного обсягу реалізації тканин.

Розв'язання задачі

Індекс цін визначається за формулою $I_p=\frac{\sum p_1q_1}{\sum p_0q_1}$

Оскільки в таблиці відсутні дані, необхідні для визначення загального індексу безпосередньо в агрегатній формі (інформація щодо кожного виду тканин про ціну (p) і кількість (q) реалізованого товару у базисному та звітному періодах), але відомі індивідуальні індекси цін i_p та відповідна їм вага (товарооборот у фактичних цінах за звітний період p_1q_1), для розрахунку загального індексу цін необхідно скористатися формулою середньої гармонічної зваженої:

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}},$$

де i_p – індивідуальні індекси цін, які в цьому прикладі визначаються на підставі інформації про зміну ціни у звітному періоді порівняно з базисним.

Тканини	Зміни ціни у звітному періоді порівняно з базисним, %	Індивідуальні індекси цін, i_p
Бавовняні	+2	$\frac{100+2}{100} = 1{,}02$
Шовкові	–4	$\frac{100-4}{100} = 0{,}96$

Тоді $I_p = \frac{15{,}3 + 177{,}6}{\frac{15{,}3}{1{,}02} + \frac{177{,}6}{0{,}96}} = 0{,}965$ (– 3,5%).

У звітному періоді порівняно з базисним ціни за двома видами тканин у середньому зменшилися на 3,5% .

Індекс фізичного обсягу реалізації визначимо, скориставшись правилом взаємозв'язку залежних індексів: $I_{pq} = I_p \cdot I_q$. Індекс обсягу реалізації:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{15{,}3 + 177{,}6}{14{,}9 + 180{,}0} = \frac{192{,}9}{194{,}9} = 0{,}990 \quad (–1{,}0\%).$$

Тобто у звітному періоді порівняно з базисним обсяг реалізації зменшився на 1%. Тоді $I_q = I_{pq} \div I_p = 0{,}99 / 0{,}965 = 1{,}023$ (+2,3%).

У звітному періоді порівняно з базисним фізичний обсяг товарообороту збільшився на 2,3%.

Задача 4

Маємо такі дані:

Вид продукції	Кількість виробленої продукції, шт.		Собівартість одиниці продукції, грн/шт.	
	базисний період, q_0	звітний період, q_1	базисний період, z_0	звітний період, z_1
А	378	400	5,0	4,8
Б	448	500	12,0	10,0

Визначити загальні індекси собівартості одиниці продукції, фізичного обсягу виробництва та виробничих витрат на виготовлення всієї продукції. Обчислити абсолютну суму економії (перевитрат) грошових засобів на виробництво в цілому і в тому числі за рахунок зміни собівартості та кількості виробленої продукції.

Розв'язання задачі

Загальний індекс собівартості розраховується за формулою:

$$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1},$$

де z_1 та z_0 – собівартість одиниці продукції звітного і базисного періоду відповідно; q_1 та q_0 – кількість виробленої продукції звітного і базисного періоду відповідно.

Розрахуємо загальний індекс собівартості по двох видах продукції $I_z = \frac{4{,}8 \cdot 400 + 10{,}0 \cdot 500}{5{,}0 \cdot 400 + 12{,}0 \cdot 500} = \frac{6920}{8000} = 0{,}865$ (–13,5%), тобто собівартість у звітному періоді порівняно з базисним у середньому знизилася на 13,5%.

Загальний індекс фізичного обсягу виробництва розраховується за формулою:

$$I_q = \frac{\sum z_0 q_1}{\sum z_0 q_0} = \frac{5{,}0 \cdot 400 + 12{,}0 \cdot 500}{5{,}0 \cdot 378 + 12{,}0 \cdot 448} = \frac{8000}{7266} = 1{,}101 \ (+10{,}1\%).$$

У звітному періоді порівняно з базисним фізичний обсяг виробництва продукції збільшився на 10,1% .

Загальний індекс виробничих витрат на виготовлення всієї продукції розраховується за формулою:

$$I_{zq} = \frac{\sum z_1 q_1}{\sum z_0 q_0} = \frac{4{,}8 \cdot 400 + 10{,}0 \cdot 500}{5{,}0 \cdot 378 + 12{,}0 \cdot 448} = \frac{6920}{7266} = 0{,}952$$

або $I_{zq} = I_z \cdot I_q = 0{,}865 \cdot 1{,}101 = 0{,}952$ $(-4{,}8\%)$.

Виробничі витрат на виготовлення всієї продукції в звітному періоді порівняно з базисним зменшилися на 4,8%.

Абсолютна сума економії (перевитрат) грошових засобів на виробництво за рахунок зміни собівартості розраховується як:

$$\Delta_z = \sum z_1 q_1 - \sum z_0 q_1 = 6920 - 8000 = -1080 \text{ (грн)}$$

або $\Delta_z = \sum (z_1 - z_0) \cdot q_1 = (4{,}8 - 5{,}0) \cdot 400 + (10{,}0 - 12{,}0) \cdot 500 = -1080$ (грн).

За рахунок зниження собівартості у звітному періоді порівняно з базисним економія грошових коштів на виробництво становила 1080 грн.

Абсолютна сума економії (перевитрат) грошових засобів на виробництво за рахунок зміни кількості виробленої продукції розраховується як:

$$\Delta_q = \sum z_0 q_1 - \sum z_0 q_0 = 8000 - 7266 = 734 \text{ (грн)}$$

або $\Delta_q = \sum z_0 \cdot (q_1 - q_0) = 5{,}0 \cdot (400 - 378) + 12{,}0 \cdot (500 - 448) = 734$ (грн).

За рахунок збільшення кількості виробленої продукції у звітному періоді порівняно з базисним перевитрати грошових коштів на виробництво становили 734 грн.

Абсолютну суму економії (перевитрат) грошових засобів на виробництво в цілому обчислюємо як:

$$\Delta_{zq} = \sum z_1 q_1 - \sum z_0 q_0 = 6920 - 7266 = -346 \text{ (грн)}$$

або $\Delta_{zq} = \Delta_z + \Delta_q = -1080 + 734 = -346$ (грн).

У звітному періоді порівняно з базисним виробничі витрат на виготовлення всієї продукції зменшилися на 346 грн.

Задача 5

Маємо такі дані про виробництво двох видів продукції на підприємстві:

Вид продукції	Кількість виробленої продукції, шт.		Ціна одиниці продукції, грн/шт.	
	базисний період, q_0	звітний період, q_1	базисний період, p_0	звітний період, p_1
А	378	400	15,0	18,0
Б	448	500	12,0	16,5

Визначити:

1) середню ціну одиниці виробленої продукції;
2) індекси середньої ціни виробленої продукції змінного, фіксованого складу та структурних зрушень.

Розв'язання задачі

1) Середню ціну виробленої продукції визначаємо за формулою: $\overline{p} = \frac{\sum pq}{\sum q}$:

– для базисного періоду:

$$\overline{p}_0 = \frac{\sum p_0 q_0}{\sum q_0} = \frac{15{,}0 \cdot 378 + 12{,}0 \cdot 448}{378 + 448} = \frac{11046}{826} = 13{,}37 \text{ (грн);}$$

– для звітного періоду:

$$\overline{p}_1 = \frac{\sum p_1 q_1}{\sum q_1} = \frac{18{,}0 \cdot 400 + 16{,}5 \cdot 500}{400 + 500} = \frac{15450}{900} = 17{,}17 \text{ (грн).}$$

2) Індекс змінного складу ($I_{\overline{p}}$) характеризує зміну середнього рівня ціни виробленої продукції у звітному періоді порівняно з базисним за рахунок впливу двох факторів: ціни окремих видів продукції та структури виробленої продукції:

$$I_{\overline{p}} = \frac{\overline{p}_1}{\overline{p}_0} = \frac{\sum p_1 q_1}{\sum q_1} \div \frac{\sum p_0 q_0}{\sum q_0} = \frac{17{,}17}{13{,}37} = 1{,}284 \quad (+28{,}4\%).$$

Середня ціна одиниці виробленої продукції за двома видами у звітному періоді порівняно з базисним збільшилась на 28,4%.

Індекс фіксованого складу (I_p) враховує зміну тільки ціни окремих видів продукції і показує, як під впливом лише її зміни у звітному періоді порівняно з базисним змінився середній рівень ціни виробленої продукції:

$$I_p = \frac{\sum p_1 q_1}{\sum q_1} \div \frac{\sum p_0 q_1}{\sum q_1} = 17{,}17 \div \frac{15{,}0 \cdot 400 + 12{,}0 \cdot 500}{400 + 500}$$

$$= 17{,}17 \div \frac{12000}{900} = 17{,}17 \div 13{,}33 = 1{,}288 \quad (+28{,}8\%).$$

У середньому за двома видами продукції ціна одиниці виробленої продукції зросла на 28,8%.

Індекс структурних зрушень (I_d) характеризує вплив зміни структури виробленої продукції на динаміку середнього рівня ціни останньої:

$$I_d = \frac{\sum p_0 q_1}{\sum q_1} \div \frac{\sum p_0 q_0}{\sum q_0} = \frac{13{,}33}{13{,}37} = 0{,}997 \quad (-0{,}3\%).$$

Середня ціна одиниці виробленої продукції за двома видами у звітному періоді порівняно з базисним зменшилась на 0,3% за рахунок зміни структури окремих видів у загальному обсязі виробленої продукції.

Запитання для самоконтролю

1. Що таке індекси і для чого вони використовуються?
2. У чому полягає відмінність зведеного індексу від індивідуального?
3. В яких випадках використовуються середньозважені індекси?
4. За якими формулами розраховують зведені індекси інтенсивного та екстенсивного показників?
5. У чому полягають особливості побудови середньозважених індексів?
6. Які завдання виконуються з допомогою системи взаємозв'язаних індексів?
7. Як з допомогою індексного аналізу визначити абсолютну зміну досліджуваних явищ?
8. Для чого використовуються індекси середніх величин?
9. Як розраховуються індекси середніх величин?
10. Який взаємозв'язок існує між індексами середніх величин?

Завдання для самостійної роботи

Тестові завдання

1. Індекси використовуються в аналізі для:

а) вивчення динаміки соціально-економічних явищ та порівняння рівнів соціально-економічних явищ у просторі (за різними об'єктами та регіонами);
б) вивчення динаміки соціально-економічних явищ;
в) порівняння рівнів соціально-економічних явищ у просторі (за різними об'єктами та регіонами).

Відповіді: 1) а; 2) б; 3) в.

2. Вкажіть, які твердження є правильними:

а) якщо I_p=0,986, а I_{pq}=1,678, то I_q=1,702;
б) якщо I_c=1,322, а I_q=1,342, то I_{cq}=1,574;
в) якщо I_{pq}=0,974, а I_q=1,565, то I_p=0,566;
г) якщо I_q=1,224, а I_{tq}=0,998, то I_t=0,815.

Відповіді: 1) а; 2) б; 3) в; 4) г.

3. Визначте, як вплинула зміна регіональної структури обсягу реалізації на середню ціну товару, якщо у звітному періоді порівняно з базисним середня ціна товару збільшилася на 1,7%, у тому числі за рахунок зміни ціни товару за регіонами – на 5,7%.

а) за рахунок структурних зрушень в обсягах реалізації середня ціна зменшилася на 3,8%;
б) за рахунок структурних зрушень в обсягах реалізації середня ціна збільшилася на 0,962%;
в) середня ціна збільшилася на 3,8%;
г) якщо б структура не змінювалась, то його середня ціна зросла б у 0,962 раза.

Відповіді: 1) а; 2) б; 3) в; 4) г.

4. Зміну в часі індивідуального явища характеризують:

а) зведені індекси;
б) індивідуальні індекси;
в) індекси середніх величин.

Відповіді: 1) а; 2) б; 3) в.

5. Зведений індекс інтенсивного показника визначається за формулою:

а) $\frac{\sum x_1 v_1}{\sum x_0 v_0}$;

б) $\frac{\sum x_0 v_1}{\sum x_0 v_0}$;

в) $\frac{\sum x_1 v_1}{\sum x_0 v_1}$.

Відповіді: 1) а; 2) б; 3) в.

6. Індекси, що характеризують зміну явищ, поєднаних між собою як співмножники, мають такий зв'язок:

а) $I_{xv} = I_x \div I_v$;
б) $I_{xv} = I_x \cdot I_v$;
в) $I_{xv} = I_x + I_v$.

Відповіді: 1) а; 2) б; 3) в.

7. Зміну середнього рівня досліджуваного показника у звітному періоді порівняно з базисним під впливом зміни лише осереднюваного показника для окремих одиниць сукупності характеризує:

а) індекс змінного складу;
б) індекс фіксованого складу;
в) індекс структурних зрушень.

Відповіді: 1) а; 2) б; 3) в.

8. Індекс структурних зрушень розраховується за формулою:

а) $\frac{\sum x_0 v_1}{\sum v_1} \div \frac{\sum x_0 v_0}{\sum v_0}$;

б) $\frac{\sum x_1 v_1}{\sum v_1} \div \frac{\sum x_0 v_1}{\sum v_1}$;

в) $\frac{\sum x_1 v_1}{\sum v_1} \div \frac{\sum x_0 v_0}{\sum v_0}$.

Відповіді: 1) а; 2) б; 3) в.

9. Абсолютна зміна досліджуваного явища за рахунок екстенсивного фактора визначається за формулою:

а) $\sum x_1 v_1 - \sum x_0 v_0$;

б) $\sum x_1 v_1 - \sum x_0 v_1$;

в) $\sum x_0 v_1 - \sum x_0 v_0$.

Відповіді: 1) а; 2) б; 3) в.

10. Для обчислення загального індексу інтенсивного (якісного) показника у випадку, коли наявна інформація про звітний рівень показника $x_1 v_1$ за окремими елементами сукупності та індивідуальні індекси інтенсивного показника i_x використовують:

а) агрегатну форму індексу;
б) формулу середньої арифметичної зваженої;
в) формулу середньої гармонічної зваженої.

Відповіді: 1) а; 2) б; 3) в.

Практичні завдання (задачі)

11. Маємо дані про продаж товарів на ринку міста:

Товар	Продано товару, тис. кг		Ціна за 1 кг, грн/кг	
	Березень	Квітень	Березень	Квітень
Яблука	90	85	20	25
Морква	60	40	10	15

Визначте:

1) індивідуальні індекси цін і обсягу проданого товару;
2) загальний індекс товарообороту;
3) загальний індекс фізичного обсягу товарообороту;
4) загальний індекс цін;
5) абсолютну зміну товарообороту – всього і в тому числі за рахунок зміни цін і обсягу продаж товарів.

Покажіть взаємозв'язок між розрахованими індексами. Зробіть висновки.

12. Як у середньому змінилися споживча ціна на продовольчі товари в поточному періоді порівняно з базисним, якщо фізичний обсяг продажу зменшився на 2%, а товарообіг збільшився на 5,4%?

13. Динаміка ринкових цін на окремі види послуг характеризується такими даними:

Вид послуг	Вартість послуг, тис. грн		Темп зміни цін, %
	базисний період	поточний період	
Транспорт	45	73	+4,0
Зв'язок	29	50	+12,1

Визначити зведені індекси: 1) вартості послуг; 2) цін; 3) фізичного обсягу послуг, використовуючи взаємозв'язок співзалежних індексів. Зробити висновки.

14. Маємо такі дані про продані товари:

Товари	Одиниця вимірювання	Кількість, тис. од.		Ціна, грн/од.	
		базисний період	звітний період	базисний період	звітний період
А	кг	1000	750	15	20
Б	л	2000	1800	25	46

Визначте:

1) індивідуальні індекси обсягів продажів у натуральному вираженні, цін, товарообороту;
2) агрегатні індекси фізичного обсягу;
3) агрегатні індекси цін;
4) загальний індекс товарообороту;
5) абсолютну зміну товарообороту за рахунок зміни обсягів продажів, цін і за рахунок спільної дії обох факторів.

Покажіть взаємозв'язок між загальними індексами та між абсолютними приростами товарообороту. Зробіть висновки.

15. Динаміка експорту паперово-картонної продукції характеризується такими даними:

Вид продукції	Експорт продукції, млн дол. США		Темп скорочення обсягу експортованої продукції, %
	базисний період	поточний період	
Папір	25	35	–8
Картон	55	68	–4

Визначте:

а) зведений індекс фізичного обсягу експортованої продукції;
б) зведений індекс цін, використовуючи взаємозв'язок співзалежних індексів.

Зробіть висновки.

16. Маємо дані за одним видом продукції, що виробляється на кількох підприємствах:

Номер підприємства	Кількість випущених виробів, тис. шт.		Загальні витрати на весь випуск, млн грн	
	базисний період	звітний період	базисний період	звітний період
1	1000	2000	3000	4000
2	1500	1200	5250	8000

Обчисліть індекси середньої собівартості одиниці продукції змінного, фіксованого складу та впливу структурних зрушень. Поясніть економічний зміст кожного індексу. Зробіть висновки.

17. Продаж безалкогольних напоїв на оптовому ринку характеризується такими даними:

Напій	Кількість проданого товару, тис. шт.		Ціна за пляшку, грн	
	базисний період	поточний період	базисний період	поточний період
Мінеральна вода	8	10	12	15
Томатний сік	6	9	15	28

Визначте:

1) індивідуальні індекси товарообороту, цін та фізичного обсягу;
2) індекси середньої ціни змінного, фіксованого складу та впливу структурних зрушень.

Зробіть висновки.

18. Витрати на телерекламу окремих категорій товару характеризуються такими даними:

Категорія товару	Собівартість одного рекламного ролика, тис. дол. США		Кількість виготовлених роликів, од.	
	базисний період	поточний період	базисний період	поточний період
Солодощі	33	45	15	18
Напої	30	42	14	13

Визначте:

1) зведені індекси:
   а) собівартості та кількості виготовлених рекламних роликів;
   б) загальних витрат;
2) абсолютну зміну витрат на телерекламу в цілому і за рахунок зміни:
   а) собівартості одного рекламного ролика;
   б) кількості виготовлених роликів.

Зробіть висновки.

Розділ 10.
БАГАТОВИМІРНІ МЕТОДИ АНАЛІЗУ

10.1. Статистичні методи порівняння соціально-економічних об'єктів

Соціально-економічним об'єктам притаманна багатовимірність. Це означає, що для їх вичерпної характеристики слід застосовувати сукупність ознак. З їх допомогою можна визначити рівень розвитку країни, здійснити регіональну диференціацію, провести типологічне групування, сформувати рейтинг тощо. Узагальнення ознак дозволяє здійснити інтегральне оцінювання об'єктів, за результатами якого виконується їх ранжування або типологізація. Подібні порівняння можна проводити як у статиці, так і в динаміці.

Величина *інтегральної оцінки* G_j визначається за допомогою багатовимірної середньої:

$$G_j = \frac{1}{m}\sum_{1}^{m} Z_{ij}, \qquad (10.1)$$

де Z_{ij} – стандартизовані значення i-ї ознаки j-го об'єкта; m – кількість ознак.

Алгоритм розрахунку інтегральних оцінок такий:

1. Теоретичне обґрунтування складу ознакової множини показників, що характеризують величину ознак.
2. Вибір способу стандартизації первинних показників.
3. Реалізація процедури зважування (за необхідності).
4. Визначення процедури агрегування показників.

І етап. Формування ознакової множини показників визначає можливості та глибину подальшого дослідження. Наприклад, для оцінювання рівня розвитку зовнішньоекономічної діяльності (ЗЕД) можна обрати такі показники, як експорт товарів і послуг, імпорт товарів і послуг, прямі іноземні інвестиції, валова додана вартість (ВДВ) тощо.

За принципом інтегрального оцінювання розраховується індекс людського розвитку (Human Development Index), що характеризує людський потенціал країни. В основі його розрахунку лежить система показників очікуваної тривалості життя, рівня грамотності дорослого населення, тривалості навчання, добробуту (рівня життя), оціненого за розміром ВВП на одну особу населення, тощо.

II етап. Формуючи інформаційну базу аналізу, важливо забезпечити односпрямованість показників. Водночас багатьом соціально-економічним характеристикам властива різноспрямованість (стимулятори/дестимулятори). Зважаючи на відмінності одиниць вимірювання та напряму дії показників останні необхідно стандартизувати (нормувати), тобто привести до порівнянної основи. Способи стандартизації показників грунтуються на порівнянні поточних значень (x_{ij}) з їх мінімальним ($x_{\min}$), максимальним ($x_{\max}$), середнім ($\bar{x}$), еталонним (x_0) значеннями або на нормуванні, яке здійснюється з допомогою середньоквадратичного відхилення (σ) чи розмаху варіації ($x_{\max} - x_{\min}$), табл. 10.1.

Таблиця 10.1

Методи нормування показників

Показники-стимулятори	Показники-дестимулятори
$z_{ij} = \frac{x_{ij} - \bar{x}}{s}$	$z_{ij} = \frac{\bar{x} - x_{ij}}{s}$
$z_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}}$	$z_{ij} = \frac{x_{\max} - x_{ij}}{x_{\max} - x_{\min}}$
$z_{ij} = \frac{x_{ij}}{x_{\max}}$	$z_{ij} = 1 - \frac{x_{ij}}{x_{\max}}$
$z_{ij} = \frac{x_{ij}}{x_{\min}}$	$z_{ij} = \frac{x_{\min}}{x_{ij}}$
$z_{ij} = \frac{x_{ij}}{\bar{x}}$	$z_{ij} = \frac{x_{ij}}{\bar{x}}$
$z_{ij} = \frac{x_{ij}}{x_0}$	$z_{ij} = \frac{x_{ij}}{x_0}$

Слід зазначити, що віднесення показників до стимуляторів (рівень заробітної плати, рівень грамотності населення, коефіцієнти

народжуваності, очікувана тривалість життя) чи до дестимуляторів (рівень безробіття, рівень смертності, зовнішній борг, індекс інфляції) залежить від мети дослідження (окресленої на першому етапі), яка визначає подальший спосіб нормування показників та обчислення інтегральних оцінок.

Наприклад, показник імпорту можна віднести до стимуляторів при оцінюванні рівня розвитку ЗЕД, водночас цей показник можна трактувати як дестимулятор, перевищення якого над експортом породжує негативне сальдо торговельного балансу, зростання імпортної залежності країни тощо.

III етап. Застосування вагових коефіцієнтів в інтегральному оцінюванні не є обов'язковим. Однак за умови різновагомості показників багатовимірна середня G_j розраховується на основі арифметичної зваженої:

$$G_j = \frac{1}{m}\sum_{1}^{m} Z_{ij} W_i , \qquad (10.2)$$

де Z_{ij} – стандартизовані значення i-ї ознаки j-го об'єкта; m – кількість ознак; Wi – вагові коефіцієнти.

Наприклад, в основу оцінювання рівня розвитку інформаційно-комунікаційних технологій (ІКТ), який щороку розраховується й оприлюднюється Міжнародним союзом електрозв'язку, закладено три субіндекси з ваговими коефіцієнтами: доступ до ІКТ з часткою 0,4, використання ІКТ з часткою 0,4 та практичні навички у сфері ІКТ з часткою 0,2.

IV етап. Залежно від змісту та соціально-економічної природи показників зазвичай обирають той метод стандартизації, за яким інтегральна оцінка G_j може набувати значення від 0 до 1: чим вищий рівень розвитку явища, тим більшою мірою значення G_j наближається до одиниці. Якщо нормування показників здійснюється шляхом порівняння з середнім рівнем по сукупності, то при $G_j > 1$ рівень розвитку явища для j-го об'єкта вважається вищим за середній по сукупності, а при $G_j < 1$ – нижчим. Однак, такий висновок можна застосовувати лише за умови односпрямованості показників (наприклад, лише стимулятори або виняткова дестимулятори). Якщо нормування ви-

конується шляхом порівняння з еталонним показником, то висновки робляться відповідно до напряму відхилень від еталонних значень (вище норми чи нижче).

При використанні комп'ютерних пакетів обробки даних слід зважати на те, що вони не здатні розрізняти семантики показників, а тому здійснюють стандартизацію тільки шляхом порівняння з середньою, за таких умов інтегральна оцінка G_j може набувати і від'ємних значень, якщо не врахувати її по модулю:

$$z_{ij} = \frac{x_{ij} - \bar{x}}{s}. \qquad (10.3)$$

Основне завдання інтегральних оцінок – це групування, ранжування та типологія в багатовимірному просторі. У такий спосіб за низкою економічних показників оцінюють рівень економічного розвитку країни; на основі показників медичної статистики – стан здоров'я нації; за соціально-демографічними показниками аналізують рівень соціального добробуту та рівень грамотності населення тощо.

10.2. Формування сукупностей з подібними елементами

Завдання багатовимірної класифікації вирішує *кластерний аналіз (КА),* мета якого – формування однорідних груп у багатовимірному просторі. *Багатовимірна класифікація* – це віднесення об'єктів спостереження, подібних за певними критеріями чи ознаками, до окремих груп. Будь-яка класифікація є відносною та залежить від мети дослідження.

Кластер – це клас, група, об'єднання однорідних одиниць сукупності. Сукупності вважаються однорідними, якщо їм властивий симетричний нормальний розподіл. Як правило, економічним сукупностям притаманна неоднорідність, а для вирішення завдань економіко-статистичного моделювання важливим є забезпечення саме однорідності розподілу. У таких випадках зазвичай оперують поняттям "відносна однорідність", тобто "близькість", "схожість". Інколи не вдається приєднати об'єкт до жодного кластера, тоді такий об'єкт визначають, як нетиповий.

Кластерний аналіз набув практичного застосування в регіональній економіці, медицині, маркетингу, соціології та в багатьох інших сферах. КА дозволяє виділити однорідні групи (кластери) в багатовимірному інформаційному просторі. Результати кластеризації використовуються в науково-практичних дослідженнях, наприклад для дискримінантного аналізу, який передбачає вивчення нових об'єктів та віднесення їх до того чи іншого кластера.

Алгоритм кластерного аналізу такий:

1. Формування вибірки (ознакової множини) для кластеризації.
2. Обґрунтування вибору міри подібності між об'єктами.
3. Застосування методів КА для формування груп.
4. Перевірка достовірності результатів кластеризації.
5. Інтерпретація отриманих результатів.

Існує сім сімейств методів КА (ієрархічні методи, ітеративні, методи згущення, методи, на основі теорії графів та інші). Найпоширенішими є ієрархічні та ітераційні процедури.

І. Ієрархічні кластер-процедури. Ієрархічні процедури поділяють на дивізимні (ті, що розділяють) та агломеративні (ті, що збирають). Дивізимна процедура розглядає первинну сукупність елементів вибірки (n) як єдиний кластер (матрицю), поетапно розбиваючи її на n кластерів. Агломеративна навпаки – послідовно об'єднує елементи сукупності в окремі кластери, ніби нанизуючи один на одного.

Існують різні правила послідовного об'єднання кластерів:

- ❑ правило одиничного зв'язку (Single linkage) (рис. 10.1);
- ❑ правило повного зв'язку (Complete linkage);
- ❑ правило середнього зв'язку (декілька різновидів);
- ❑ центроїдний метод (Centroid method);
- ❑ метод Уорда (Ward's method) (рис. 10.2).

Отже, за результатами кластеризації регіонів України методом Уорда за рівнем інвестиційної привабливості візуально можна виділити три кластери: перший – Київ і Дніпропетровська область, другий – Одеська, Львівська, Київська, Харківська, Полтавська, Запорізька та Донецька області, третій – решта регіонів України.

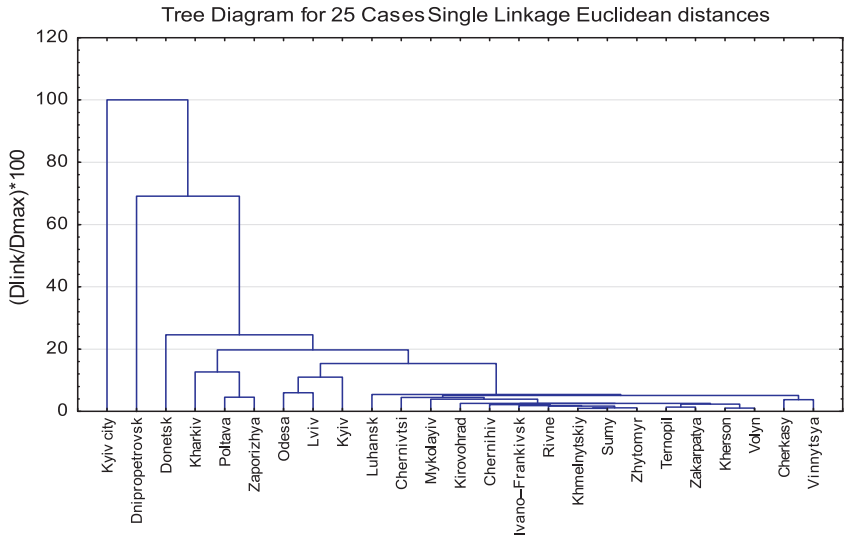


Рис. 10.1. Дендрограма результатів кластеризації регіонів України за рівнем інвестиційної привабливості, "Метод одиничного зв'язку"

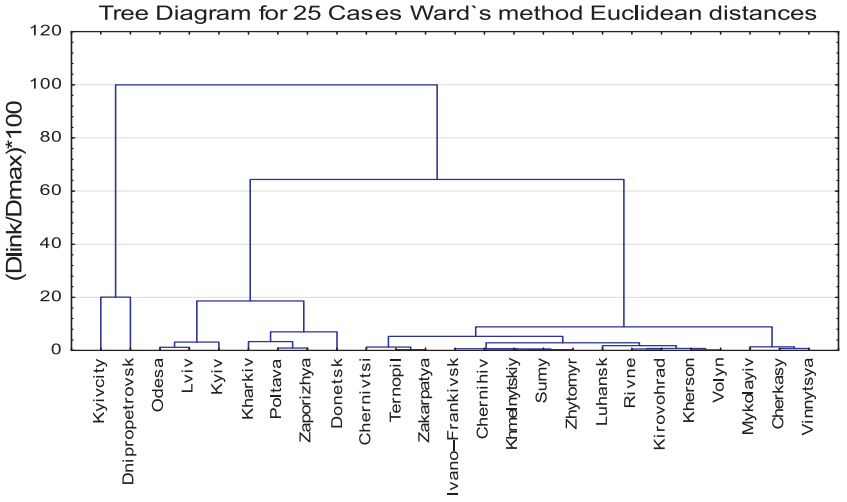


Рис. 10.2. Дендрограма результатів кластеризації регіонів України за рівнем інвестиційної привабливості, "Метод Уорда"

Кластери утворюються шляхом пошуку мінімальної відстані між ними. Використовують різні метрики відстаней: Евклідова відстань (Euclidean distance), Манхеттенська (Manhattan distance), відстань Чебишова (Chebychev distance), відстань Махаланобіса та ін.

Ієрархічна кластеризація є доволі простою, але непрактична для великих за обсягом сукупностей.

II. Ітераційні кластер-процедури. За принципом ітераційної кластеризації навпаки формуються кластери одного рангу, тобто ієрархічно не підпорядковані. Найпоширенішим є метод *k*-середніх (*k*-means clustering), який ґрунтується на принципі найближчого центру, тобто об'єкти відбираються в кластер з найближчим центром тяжіння. При цьому кількість кластерів визначається заздалегідь, тобто апріорі.

Перевагою цього методу є можливість здійснити перевірку статистичної значущості міжкласових розбіжностей на основі дисперсійного аналізу.

Слід зважати на те, що не існує єдиного для всіх випадків і найкращого способу кластеризації, оскільки визначення кількості кластерів у різний спосіб може призвести до різних результатів.

В основі кластерного аналізу лежать складні математичні обрахунки, тому для спрощення обчислень використовують комп'ютерні обчислювальні програми. Наразі є чимало статистичних програмних продуктів (R, SAS, STATISTICA, SPSS, NCSS, Prism та багато інших). Найпотужнішими вважаються SAS та R, однак вони вимагають певних навичок програмування, оскільки працюють за принципом прописування команд. Найпоширенішим продуктом є RStudio через свою доступність, безкоштовність, постійну оновлюваність версій тощо. STATISTICA (виробник StatSoft) – це статистичний пакет обробки й аналізу даних, який використовується як навчальний інструмент у багатьох закладах вищої освіти та загалом є зручним у використанні.

У пакеті STATISTICA кластерний аналіз здійснюється в модулі Cluster Analysis і включає три методи кластеризації (рис. 10.3):

- Joining (tree clustering) – належить до ієрархічної кластер-процедури, виконує деревоподібну кластеризацію (вертикальну та горизонтальну) (рис. 10.4);

- k-means clustering, або метод k-середніх – належить до ітераційної кластер-процедури (рис. 10.5);
- Two-way joining – передбачає двоетапну кластеризацію, тобто за об'єктами та за ознаками, зі створенням кольорової мапи.

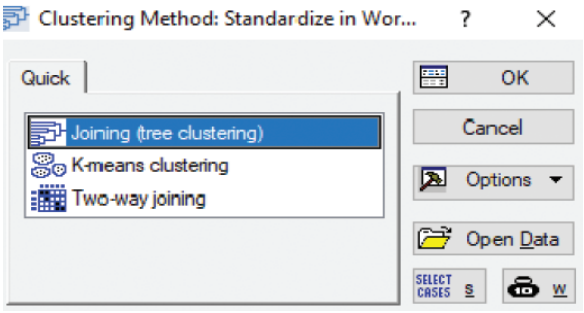


Рис. 10.3. Скриншот стартового вікна кластерного методу

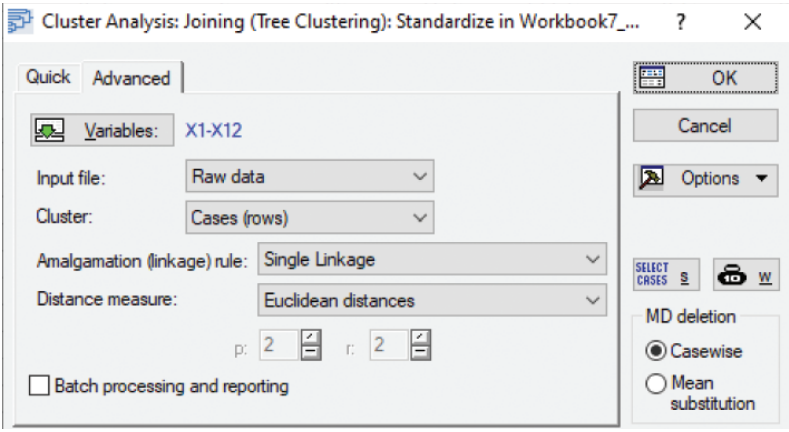


Рис. 10.4. Скриншот меню Joining (tree clustering)

Кластеризацію виконують за об'єктами (країнами, регіонами, підприємствами, респондентами тощо), які відповідно до організації даних подаються в рядках, а показники (ознаки) – у колонках, тому зазвичай обирають Cases (rows). У модулі Cluster Analysis за замовчуванням використовують одиничний зв'язок та Евклідову відстань (рис. 10.4).

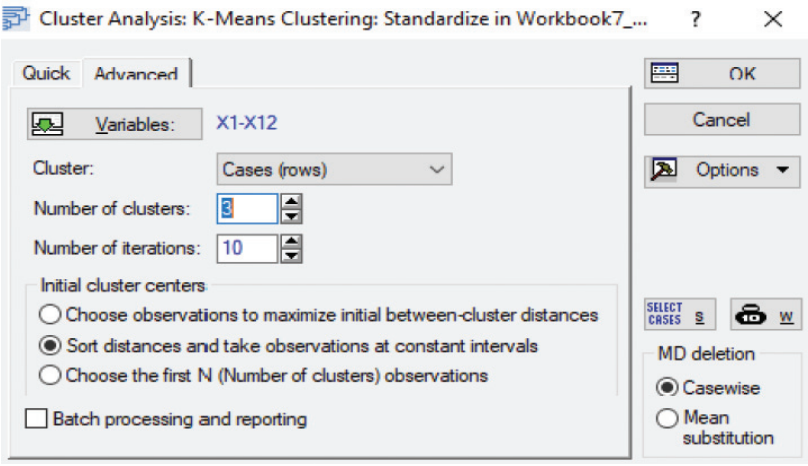


Рис. 10.5. Скриншот меню *k*-means clustering

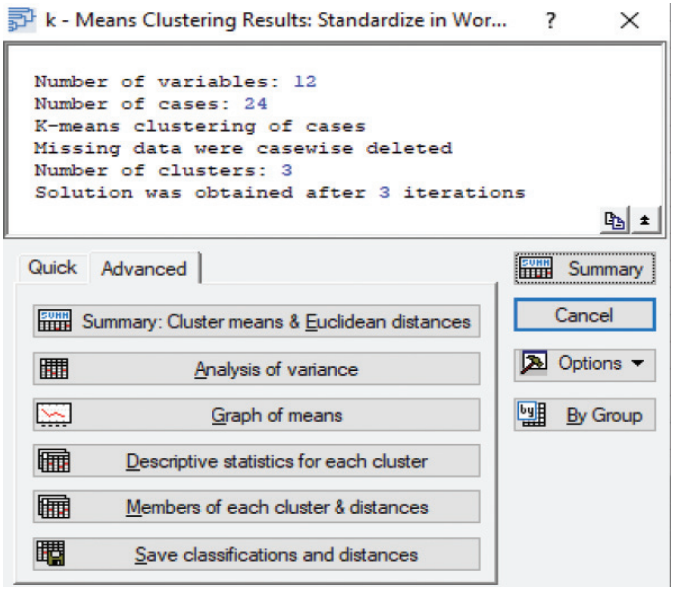


Рис. 10.6. Скриншот меню результатів k-means clustering

За командою Members of each cluster & distances отримуємо перелік об’єктів та їх відстані, які увійшли до кожного кластера (рис.

10.6). За командою Analysis of variance можна отримати результати дисперсійного аналізу.

10.3. Оптимізація інформації у статистичному аналізі

Багатовимірність соціально-економічних явищ породжує значні за розмірами бази даних, потоки інформації. З метою дослідження внутрішньої структури об'єкта слід угрупувати великий масив інформації, тобто "стиснути" розмірність початкової ознакової множини, замінивши її мінімальною кількістю компонент. Це дозволить позбутися мультиколінеарності між первинними показниками (ознаками) та уникнути автокореляції в часових рядах.

Мультиколінеарність – це наявність лінійної залежності між показниками, що оцінюється на основі коефіцієнта кореляції Пірсона (r) і коливається в межах від –1 до +1. Зазвичай значення $r > 0{,}85$ свідчить про наявність мультиколінеарності, якої слід позбутися шляхом вилучення з аналізу висококорельованих показників. У кореляційно-регресійному аналізі залишають ті показники, які більше корелюють з результативною ознакою.

Автокореляція – це залежність поточних рівнів динамічного ряду від попередніх рівнів. Щоб забезпечити побудову адекватної статистичної моделі, яка б охарактеризувала реальний процес розвитку явища, а не складову, пов'язану зі зміною часу, динамічні ряди слід перевіряти на наявність автокореляції.

Найпоширенішим методом оптимізації інформації є метод головних компонент (МГК). Головні компоненти (ГК) зберігають всі необхідні характеристики об'єкта дослідження. На основі цього методу можна ранжувати та класифікувати об'єкти (країни, регіони), вимірювати взаємозв'язки між первинними показниками та головними компонентами, виконувати регресійний аналіз тощо.

Не слід плутати компонентний аналіз із факторним. Основою факторного аналізу, популярного у суспільних науках, є виявлення інтерпретованих лінійних зв'язків між змінними, які називаються факторами. Натомість МГК – це метод зменшення кількості спостережуваних змінних до оптимального набору незалежних компонентів, а його основними цілями є:

1) візуалізація даних для дослідницького аналізу, що дає змогу розкрити латентні (приховані) характеристики даних та інтерпретувати компоненти;
2) зменшення кількості предикторів для майбутніх досліджень, наприклад регресійного аналізу основних компонент.

Найпростішим методом формування підмножини ГК є відбір першої кількості компонент (*k*) із найбільшими власними значеннями λ (лямбда), тобто значеннями дисперсій. Сумарна величина дисперсій компонент дорівнює *m* ознак.

Інший класичний метод вибору ГК ґрунтується на вивченні частки загальної дисперсії, що пояснюється кожною компонентою. Встановивши заздалегідь визначений поріг (як правило, 75% або 80% від загальної пояснюваної дисперсії), перші *k* головних компонент, які сукупно пояснюють принаймні цю велику частину дисперсії, можуть бути обрані як підмножина компонент.

Класичні методи відбору ГК не завжди враховують дисперсію даних, імовірно через випадкову помилку або шум. Тому найоптимальнішим методом відбору головних компонент є застосування паралельного аналізу (Parallel analysis, PA), який використовує багаторазове моделювання даних, щоб подолати цю проблему. Parallel analysis (його ще називають паралельний аналіз Горна, на честь його творця Дж. Горна) – це метод вибору ГК, що враховує дисперсію даних через випадкову помилку або шум. Процес виконання паралельного аналізу ґрунтується на методі Монте-Карло, тобто передбачає симуляцію великої кількості наборів даних (наприклад 1000 або більше), при цьому кожен імітований набір містить таку саму кількість змінних і спостережень, що і вихідні дані.

Отже, метод головних компонент дозволяє упорядкувати великий масив інформації, виділивши оптимальну кількість компонент (факторів) та зберегти основну інформацію про об'єкт дослідження.

Соціально-економічний розвиток країни можна оцінювати за різними критеріями: матеріальний добробут населення; рівень бідності; рівень надання медичних послуг; рівень цифровізації економіки та суспільства; доступ до освіти; гендерна рівність; рівень довіри суспільства; безпека та захист з боку держави тощо. Усі ці по-

казники можна узагальнити з використанням МГК, який дозволить серед низки різнонаправлених статистичних показників виділити ті головні компоненти, що найповніше характеризують досліджуваний процес, та ідентифікувати їх.

У пакеті STATISTICA модуля Principle Components & Classification Analysis (PC&CA) за замовчуванням відбір головних компонент здійснюється на основі кореляційної матриці стандартизованих значень показників. Коваріаційну матрицю, в основу якої закладено центрування значень, застосовують рідше (рис. 10.7).

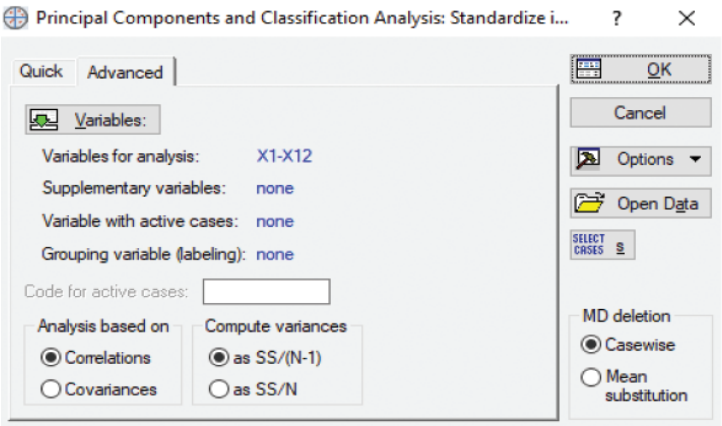


Рис. 10.7. Скриншот стартового вікна PC&CA

На закладці Variables за опцією Eigenvalues можна отримати власні значення λ кореляційної матриці для визначення кількості ГК; за опцією Factor & variables correlations – факторні навантаження, які характеризують щільність зв'язку між *i*-ю ознакою та *j*-ю компонентою й коливаються в межах від –1 до +1; за опцією Scree plot будується графік кам'янистого спаду (рис. 10.8). На закладці Cases за опцією Factor coordinates of cases можна отримати індивідуальні (для кожного окремого спостереження) значення ГК (рис. 10.9). Знаки (додатні або від'ємні), які присвоєні значенням кожної компоненти, свідчать про рівень головної компоненти для кожного спостереження, тобто показують, наскільки він вищий (+) або нижчий (–) за середній.

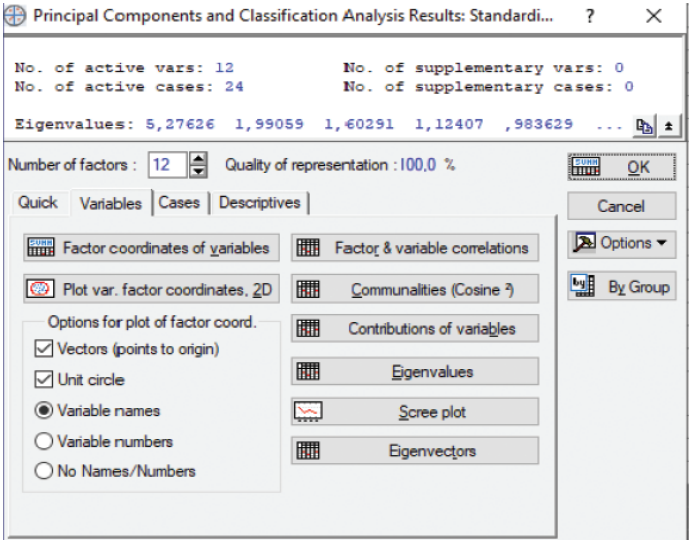


Рис. 10.8. Скриншот меню Variables PC&CA

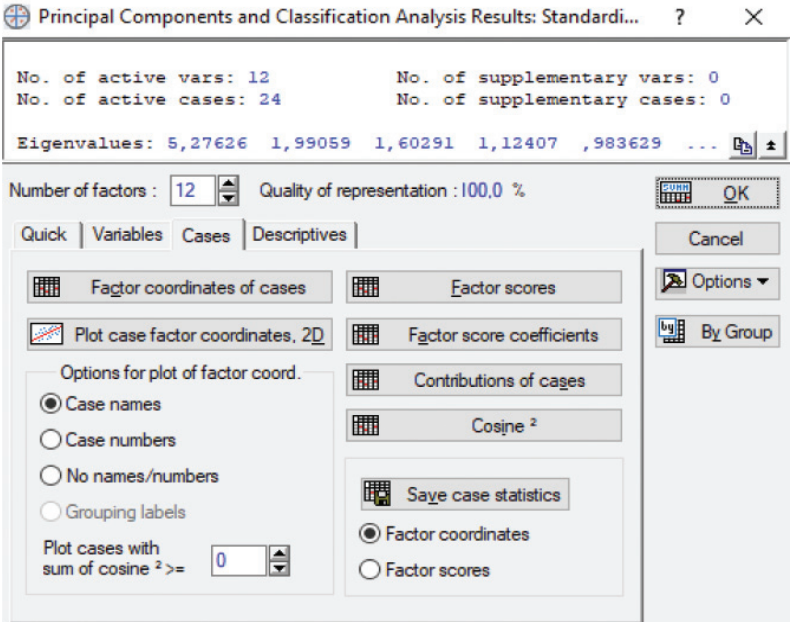


Рис. 10.9. Скриншот меню Cases PC&CA

Розв'язання типових задач

Задача 1

Розрахувати інтегральний показник інформаційно-комунікаційних технологій (ІКТ) України за період 2016–2019 рр. за показниками табл. 10.2, обґрунтувавши їх вибір.

Розв'язання задачі

Обґрунтуємо вибір показників, що характеризують розвиток ІКТ.

1. Кількість абонентів Інтернету щороку зростає і характеризує активність інтернетизації населення України. Так, у 2019 р. проти 2016 р. кількість інтернет-користувачів зросла майже на 56%.

Базисний темп зростання:

$$К2019/2016 = 26066,8/16723,0 = 1,558.$$

Обсяг реалізованих інтернет-послуг характеризує рівень розвитку електронної комерції, наприклад онлайн купівлі-продажі, отримання/надання послуг в онлайн-формі тощо. Середньорічний темп зростання послуг, реалізованих через Інтернет, становить 14,6%.

Ланцюгові коефіцієнти зростання:

$$К2017 = 10817,9/9101,8 = 1,188;$$

$$К2018 = 12273,0/10817,9 = 1,134;$$

$$К2019 = 13727,0/12273,0 = 1,118.$$

Середньорічний темп зростання: $\sqrt[3]{1,188 \cdot 1,134 \cdot 1,118} = 1,146$.

3. Пропускна спроможність Інтернету характеризує швидкість передачі даних, що є вагомим показником рівня розвитку ІКТ країни. Цей коефіцієнт щороку оприлюднюється у звіті Всесвітнього економічного форуму.

4. Наявність у домогосподарствах ПК / ноутбуків / планшетів / мобільних телефонів характеризує рівень забезпечення комунікаційними засобами домогосподарств України.

Таблиця 10.2

Вихідні дані для обчислення інтегрального показника розвитку ІКТ

Рік	Кількість абонентів Інтернету, тис. осіб	Обсяг реалізованих інтернет-послуг, млн грн	Коефіцієнт WEF (пропускна спроможність Інтернету)	Наявність у домогосподарствах ПК, у середньому на 100 домогосподарств, шт.	Наявність у домогосподарствах ноутбуків, у середньому на 100 домогосподарств, шт.	Наявність у домогосподарствах планшетів, у середньому на 100 домогосподарств, шт.	Наявність у домогосподарствах мобільних телефонів, у середньому на 100 домогосподарств, шт.
2016	16723,0	9101,8	40,7	38	27	15	201
2017	22957,5	10817,9	45,7	38	30	18	201
2018	23632,3	12273,0	50,1	37	35	19	203
2019	26066,8	13727,0	56,1	39	38	22	204

Наступним кроком виконуємо стандартизацію показників. Обрані показники за своєю природою не є однорідними та виражаються в різних одиницях вимірювання. Але всі вони є стимуляторами, оскільки їх зростання сприяє розвитку інформатизації суспільства України, тому їх нормування слід виконати шляхом порівняння з максимальним значенням.

Приклад нормування першого показника для 2016 року:

$$z_{1,2016} = \frac{x_{ij}}{x_{i,\max}} = \frac{16726,0}{26066,8} = 0,642 .$$

За формулою 10.1 обчислено інтегральний показник ІКТ 2016 року

$$G_{2016} = \frac{\sum_{1}^{7} z_{i,2016}}{7} = \frac{0.642 + 0.663 + 0.725 + 0.974 + 0.711 + 0.682 + 0.985}{7} = 0,769 .$$

Нормовані показники та розрахований інтегральний показник розвитку ІКТ України представлені в табл. 10.3.

Таблиця 10.3

Нормовані показник та інтегральний показник розвитку ІКТ

Роки	z_1	z_2	z_3	z_4	z_5	z_6	z_7	$\sum_i z_i$	Інтегральний показник ІКТ
2016	0,642	0,663	0,725	0,974	0,711	0,682	0,985	5,382	0,769
2017	0,881	0,788	0,815	0,974	0,789	0,818	0,985	6,051	0,864
2018	0,907	0,894	0,893	0,949	0,921	0,864	0,995	6,422	0,917
2019	1,000	1,000	1,000	1,000	1,000	1,000	1,000	7,000	1,000

Отримані результати характеризують позитивну тенденцію розвитку та впровадження ІКТ. Це обумовлено багатьма чинниками, в тому числі щорічними додатними темпами зростання показників, включених до розрахунку інтегрального показника ІКТ.

Задача 2

За даними Державної служби статистики України щодо екологічної ситуації в країні проаналізувати міжрегіональну диференціацію екологічних ризиків на основі інтегрального оцінювання за 12-ма показниками:

1. Викиди забруднюючих речовин в атмосферне повітря від стаціонарних джерел забруднення, тис. тонн.
2. Викиди забруднюючих речовин в атмосферне повітря на одиницю площі, тонн/км^2.
3. Скидання забруднених зворотних вод у поверхневі водні об'єкти, млн м^3.
4. Потужність очисних споруд за регіонами, млн м^3.
5. Обсяг унесених пестицидів, тонн.
6. Утворення відходів за регіонами, тис. тонн.
7. Утворення відходів I-III класів небезпеки за регіонами, тис. тонн.
8. Утилізація відходів за регіонами, тис. тонн.
9. Утилізація відходів I-III класів небезпеки за регіонами, тис. тонн.
10. Спалювання відходів за регіонами, тис. тонн.
11. Площа рубок лісу за видами рубок, га.
12. Площа загибелі лісових насаджень, га.

Розв'язання задачі

1. У пакеті STATISTICA за опцією Standardize, вікно Data на Toolbar, виконуємо нормування показників (рис. 10.10, 10.11).

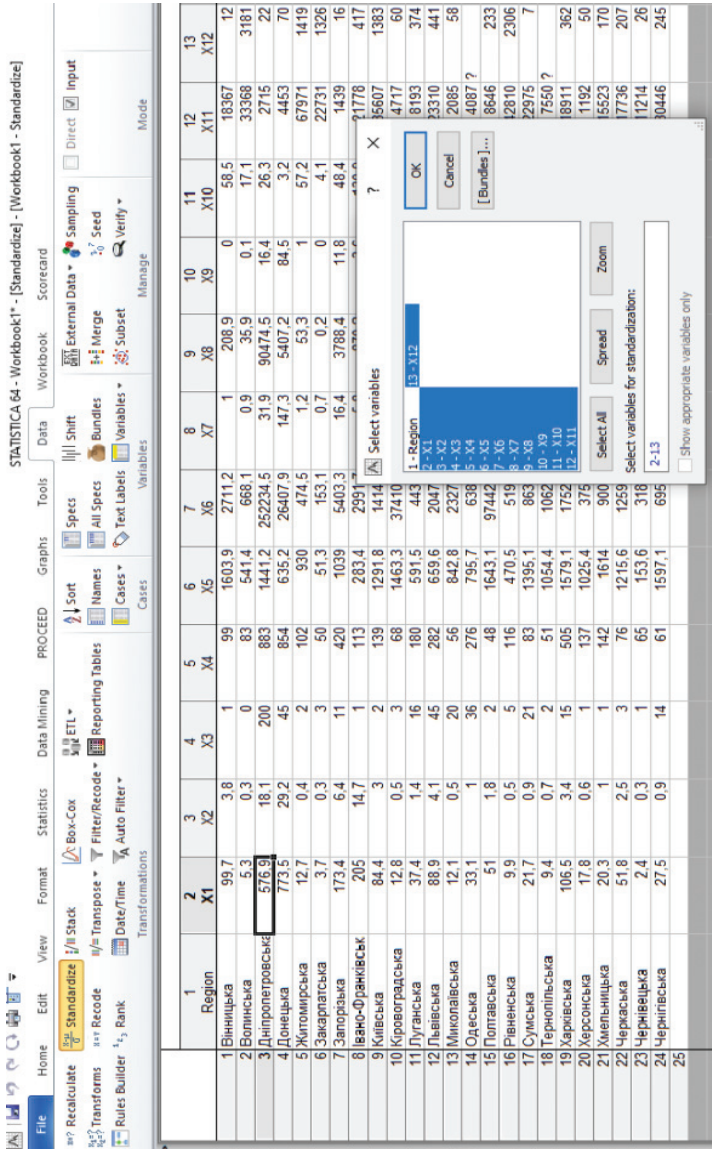
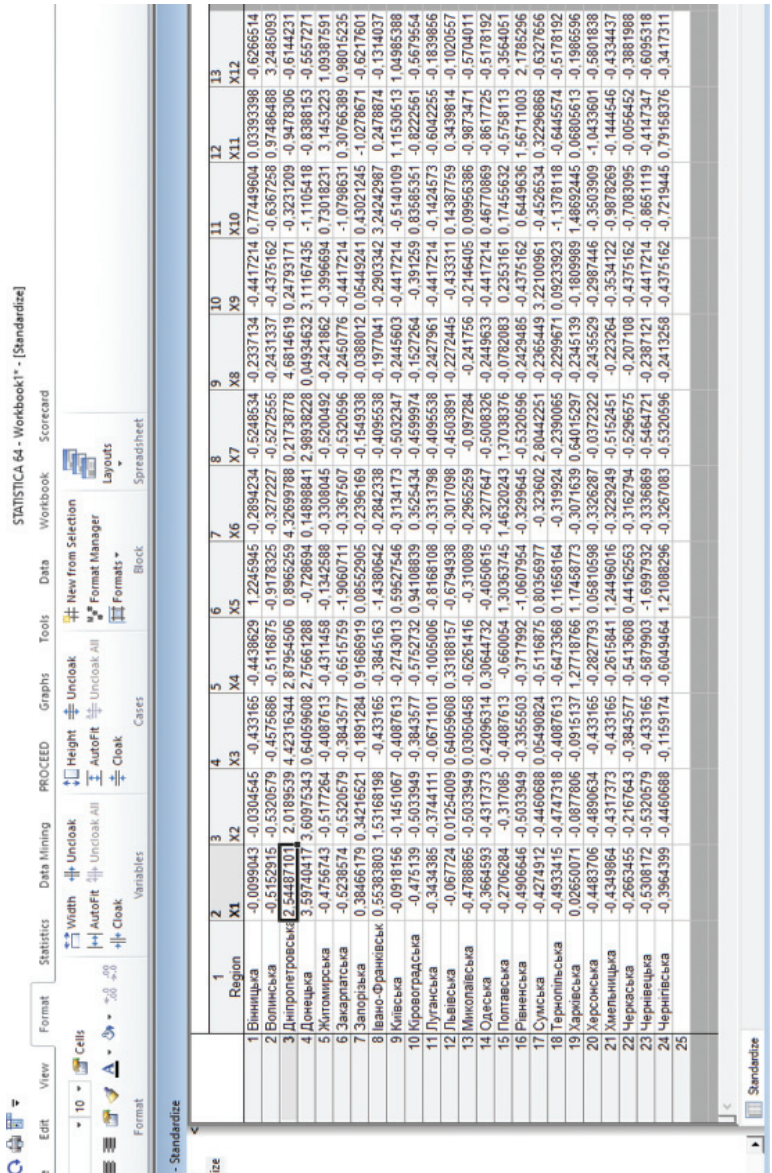


Рис. 10.10. Скриншот діалогового вікна Standardize



	1 Region	2 X1	3 X2	4 X3	5 X4	6 X5	7 X6	8 X7	9 X8	10 X9	11 X10	12 X11	13 X12
1	Вінницька	-0.0099043	-0.0304545	-0.433165	-0.4438629	1.2245945	-0.2894234	-0.5248534	-0.2337134	-0.4417214	0.7749604	0.03393398	-0.6266514
2	Волинська	-0.5152915	-0.5320579	-0.4575686	-0.5116875	-0.9178325	-0.3272227	-0.5272555	-0.2431337	-0.4375162	-0.6367258	0.97486488	3.2485093
3	Дніпропетровська	2.54487101	2.0189539	4.42316344	2.87954506	0.8965259	4.32699788	0.21738778	4.6814619	0.24793171	-0.3231209	-0.9478306	-0.6144231
4	Донецька	3.59704017	3.60975343	0.64059608	2.75661288	-0.728694	0.14898841	2.98938228	0.04934632	3.11167435	-1.1105418	-0.8388153	-0.5557271
5	Житомирська	-0.4756743	-0.5177264	-0.4087613	-0.4311458	-0.1342588	-0.3308045	-0.5200492	-0.2421862	-0.3996694	0.73018231	3.1453223	1.09387591
6	Закарпатська	-0.5238574	-0.5320579	-0.3843577	-0.6515759	-1.9060711	-0.3387507	-0.5320596	-0.2450776	-0.4417214	-1.0798631	0.30766389	0.98015235
7	Запорізька	0.38466179	0.34216521	-0.1891284	0.91686919	0.08552905	-0.2396169	-0.1549338	-0.0388012	0.05449241	0.43021245	-1.0278671	-0.6217601
8	Івано-Франківська	0.55383803	1.53168198	-0.433165	-0.3845163	-1.4380642	-0.2842338	-0.4095538	-0.1977041	-0.2903342	3.24242987	0.2478874	-0.1314037
9	Київська	-0.0918156	-0.1451067	-0.4087613	-0.2743013	0.59527546	-0.3134173	-0.5032347	-0.2445603	-0.4417214	-0.5140109	1.11530513	1.04985388
10	Кіровоградська	-0.475139	-0.5033949	-0.3843577	-0.5752732	0.94108839	0.3525434	-0.4599974	-0.1527264	-0.391259	0.83585361	-0.8222561	-0.5679554
11	Луганська	-0.3434385	-0.3744111	-0.0671101	-0.1005006	-0.8168108	-0.3313798	-0.4095538	-0.2427961	-0.4417214	-0.1424573	-0.6042255	-0.1839856
12	Львівська	-0.067724	0.01254009	0.64059608	0.33188157	-0.6794938	-0.3017098	-0.4503891	-0.2272445	-0.433311	0.14387759	0.3439814	-0.1020557
13	Миколаївська	-0.4788865	-0.5033949	0.03050458	-0.6261416	-0.310089	-0.2965259	-0.097284	-0.241756	-0.2146405	0.09956386	-0.9873471	-0.5704011
14	Одеська	-0.3664593	-0.4317373	0.42096314	0.30644732	-0.4050615	-0.3277647	-0.5008326	-0.2449633	-0.4417214	0.46770869	-0.8617725	-0.5178192
15	Полтавська	-0.2706284	-0.317085	-0.4087613	-0.660054	1.30363745	1.46320243	1.37038376	-0.0782083	0.2353161	0.17455632	-0.5758113	-0.3564051
16	Рівненська	-0.4906646	-0.5033949	-0.3355503	-0.3717992	-1.0607954	-0.3299645	-0.5320596	-0.2429485	-0.4375162	0.6449636	1.56711003	2.1785296
17	Сумська	-0.4274912	-0.4460688	0.05490824	-0.5116875	0.80356977	-0.323602	2.80442251	-0.2365449	3.22100961	-0.4526534	0.32296868	-0.6327656
18	Тернопільська	-0.4933415	-0.4747318	-0.4087613	-0.6473368	0.11658164	-0.319924	-0.2390065	-0.2299671	0.09233923	-1.1378118	-0.6445574	-0.5178192
19	Харківська	0.02650071	-0.0877806	-0.0915137	1.27718766	1.17458773	-0.3071639	0.64015297	-0.2345139	-0.1809989	1.48692445	0.06805613	-0.1986596
20	Херсонська	-0.4483706	-0.4890634	-0.433165	-0.2827793	0.05810598	-0.3326287	-0.0372322	-0.2435529	-0.2987446	-0.3503909	-1.0433601	-0.5801838
21	Хмельницька	-0.4349864	-0.4317373	-0.433165	-0.2615841	1.24496016	-0.3229249	-0.5152451	-0.223264	-0.3534122	-0.9878269	-0.1444546	-0.4334437
22	Черкаська	-0.2663455	-0.2167643	-0.3843577	-0.5413608	0.44162563	-0.3162794	-0.5296575	-0.207108	-0.4375162	-0.7083095	-0.0056452	-0.3881988
23	Чернівецька	-0.5308172	-0.5320579	-0.433165	-0.5879903	-1.6997932	-0.3336869	-0.5464721	-0.2387121	-0.4417214	-0.8651119	-0.4147347	-0.6095318
24	Чернігівська	-0.3964399	-0.4460688	-0.1159174	-0.6049464	1.21088296	-0.3267083	-0.5320596	-0.2413258	-0.4375162	-0.7219445	0.79158376	-0.3417311
25													

Рис. 10.11. Скриншот стандартизованих результатів

2. За формулою 10.1 обчислюємо інтегральний індекс екологічної небезпеки або, інакше кажучи, інтегральний індекс екологічних ризиків (табл. 10.4). Лідерами екологічної небезпеки передбачувано є промислові регіони України: Дніпропетровська (1,694) та Донецька (1,137) області. Найменші екологічні ризики – у Чернівецькій (–0,605), Закарпатській (–0,452), Херсонській (–0,376), Тернопільській (–0,366), Миколаївській (–0,342) областях України. Зазначимо, що через відсутність даних щодо ситуації у Києві інтегральний індекс екологічної небезпеки для столиці не обчислювався.

Таблиця 10.4

Ранжування областей України за рівнем екологічних ризиків

№ з/п	Області	Інтегральний індекс екологічної небезпеки/ризику
1	2	3
1	Дніпропетровська	1,694
2	Донецька	1,137
3	Сумська	0,346
4	Харківська	0,294
5	Івано-Франківська	0,164
6	Полтавська	0,154
7	Житомирська	0,119
8	Рівненська	–0,002
9	Запорізька	–0,007
10	Київська	–0,021
11	Львівська	–0,069
12	Вінницька	–0,086
13	Волинська	–0,086
14	Кіровоградська	–0,186
15	Одеська	–0,199
16	Чернігівська	–0,218
17	Хмельницька	–0,277
18	Черкаська	–0,299
19	Луганська	–0,342
20	Миколаївська	–0,352

Продовження табл.10.4

1	2	3
21	Тернопільська	–0,366
22	Херсонська	–0,376
23	Закарпатська	–0,452
24	Чернівецька	–0,605

3. На завершальному етапі побудуємо типологію регіонів України за рівнем екологічних ризиків (табл. 10.5)

Таблиця 10.5

Групування областей України за рівнем екологічних ризиків

Тип	Межі інтегрального індексу	Кількість областей у групі	Регіони України
Високий ступінь ризику	> 0,5	2	Дніпропетровська, Донецька
Середній ступінь ризику	0–0,5	5	Сумська, Харківська, Івано-Франківська, Полтавська, Житомирська
Помірний ступінь ризику	<0	17	Рівненська, Запорізька, Київська, Львівська, Вінницька, Волинська, Кіровоградська, Одеська, Чернігівська, Хмельницька, Черкаська, Луганська, Миколаївська, Тернопільська, Херсонська, Закарпатська, Чернівецька
Разом	–	24	–

Відповідно до запропонованої класифікації, більшість регіонів України (17) мають помірний ступінь екологічних ризиків. В основному до цієї групи увійшли області, що не мають великих промислових потужностей (за винятком Запорізької області), туристичні регіони України та західні області, для яких характерними є гірські ландшафти та лісистість.

До групи із середнім ступенем ризику віднесено 5 областей. У Сумській області налічується понад 120 великих, середніх та малих підприємств харчової та переробної промисловості. Харківщина характеризується машинобудуванням та підприємствами хімічної, нафтохімічної та фармацевтичної галузі. Івано-Франківщина є великим центром машинобудівної, металообробної, деревообробної, хімічної, харчової, легкої промисловості. Полтавщина відома видобутком залізної руди, переробкою нафти і газу (наприклад, ПАТ "Полтавський гірничо-збагачувальний комбінат"). До головних галузей промисловості Житомирщини належать машинобудування і металообробка, легка, харчова, деревообробна, хімічна промисловість (зокрема завод "Хімволокно").

Слід відмітити, що диференціація між групою з високим ступенем ризику (Донецька й Дніпропетровська області) та групами з середнім і помірним ступенями ризику є досить суттєвою.

Задача 3

На основі кластерного аналізу методом *k*-середніх з розбиттям на три кластери оцінити інвестиційну привабливість регіонів України за такими соціально-економічними показниками:

1. Прямі іноземні інвестиції у регіони України, млн. дол. США;
2. Капітальні інвестиції, млрд. грн.
3. Кількість зайнятих економічною діяльністю, тис. осіб;
4. Середня місячна заробітна плата, грн.;
5. Обсяг Валового регіонального продукту, млн грн.;
6. Обсяг реалізованої промислової продукції, млн грн.;
7. Експорт товарів, млн. дол. США;
8. Експорт послуг, млн. дол. США;
9. Імпорт товарів, млн. дол. США;
10. Імпорт послуг, млн. дол. США.

Розв'язання задачі

1. У пакеті STATISTICA виконаємо ітераційну процедуру кластеризації методом *k*-середніх з розбиттям на три кластери (рис. 10.12).

	Members of Cluster Number 1 (Invest.sta) and Distances from Respective Cluster Center Cluster contains 2 cases
	Distance
Dnipropetrovsk	38170,85
Kyiv city	38170,85

	Members of Cluster Number 2 (Invest.sta) and Distances from Respective Cluster Center Cluster contains 5 cases
	Distance
Donetsk	19226,59
Zaporizhya	5296,32
Kyiv	17346,23
Poltava	3998,25
Kharkiv	6068,56

	Members of Cluster Number 3 (Invest.sta) and Distances from Respective Cluster Center Cluster contains 18 cases
	Distance
Vinnytsya	8788,72
Volyn	5780,14
Zhytomyr	2040,48
Zakarpatya	7520,22
Ivano–Frankivsk	610,92
Kirovohrad	4093,39
Luhansk	6395,34
Lviv	19574,79
Mykolayiv	3308,58
Odesa	18963,02
Rivne	3311,67
Sumy	815,66
Ternopil	8467,10
Kherson	5136,78
Khmelnytskiy	1310,88
Cherkasy	7128,49
Chernivtsi	11849,49
Chernihiv	2318,22

Рис. 10.12. Результати кластеризації регіонів України за рівнем інвестиційної привабливості

2. За командою Graph of means візуалізуємо середні значення показників для кожного кластера. Найбільш суттєві міжкласові розбіжності характерні за такими показниками: обсяг валового регіонального продукту та обсяг реалізованої промислової продукції (рис. 10.13).

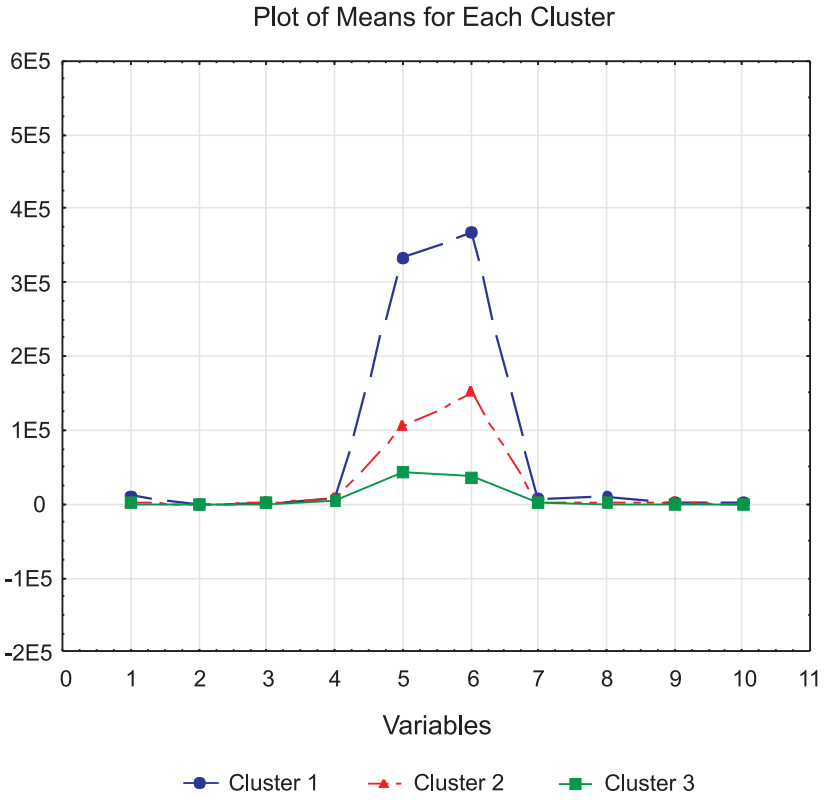


Рис.10.13. Середні значення показників для кожного кластера

3. За командою Analysis of variance виконуємо дисперсійний аналіз, який підтверджує статистичну значущість проведеної кластеризації (табл. 10.6).

Таблиця 10.6

Результати дисперсійного аналізу кластеризації регіонів України за рівнем інвестиційної привабливості

Показники	Міжгрупова дисперсія	Внутрішньо-групова дисперсія	*F*-критерій	Рівень значущості
Прямі іноземні інвестиції в Україну	230136400,0	119926200,0	21,109	0,00001
Капітальні інвестиції	7083,6	3302,0	23,598	0,00000
Кількість зайнятих економічною діяльністю	1508802,0	927103,3	17,902	0,00002
Середня місячна заробітна плата	14025730,0	9640060,0	16,004	0,00005
Обсяг валового регіонального продукту	155058800000,0	36868710000,0	46,263	0,00000
Обсяг реалізованої промислової продукції	223748500000,0	12463420000,0	197,477	0,00000
Експорт товарів	79051960,0	10752790,0	80,869	0,00000
Експорт послуг	153893000,0	86931620,0	19,473	0,00001
Імпорт товарів	3055504,0	3954769,0	8,499	0,00184
Імпорт послуг	2479873,0	1616270,0	16,878	0,00004

Чим менше значення внутрішньогрупової диспеpсії і чим більше значення міжгрупової дисперсії – тим об'єктивніша кластеризаія. Таким чином, вагомий вклад міжгрупових дисперсій, значення *F*-критерію та рівень значущості, що менший за 0,05, підтверджують істотність міжкласових розбіжностей у цілому і цим самим засвідчують високу якість проведеної кластеризації на основі обраних десяти показників із розбиттям на три кластери.

За результатами проведеного аналізу найбільш інвестиційно привабливі регіони України – це місто Київ, який лідирує за усіма показниками, окрім "Кількість зайнятих економічною діяльністю", та

Дніпропетровщина, що є потужним промисловим центром України. До другого кластера увійшло п'ять областей, що мають виробничі потужності, гарно розвинену торговельну мережу та транспортну інфраструктуру. Регіони, що ввійшли до третього кластера, є менш інвестиційно привабливими.

Задача 4

За даними задачі 2 щодо екологічної ситуації в Україні визначити оптимальну кількість головних компонент, що характеризують екологічну ситуацію в Україні. Відбір головних компонент здійснити за правилом Кайзера, обчислити факторні навантаження ГК, їх індивідуальні значення та інтерпретувати їх економічний зміст.

Розв'язання задачі

У пакеті STATISTICA модуля Principle Components & Classification Analysis за опцією Eigenvalues визначаємо кількість головних компонент, значення власних чисел (λ_j) кореляційної матриці та внесок їх дисперсій у сумарну варіацію, рис. 10.14.

	Eigenvalues of correlation matrix, and related statistics (Standardize.stw) Active variables only			
Value number	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	5,276256	43,96880	5,27626	43,9688
2	1,990592	16,58827	7,26685	60,5571
3	1,602908	13,35757	8,86976	73,9146
4	1,124074	9,36729	9,99383	83,2819
5	0,983629	8,19691	10,97746	91,4788
6	0,446876	3,72397	11,42434	95,2028
7	0,264534	2,20445	11,68887	97,4073
8	0,198247	1,65206	11,88712	99,0593
9	0,072988	0,60823	11,96011	99,6675
10	0,027080	0,22567	11,98719	99,8932
11	0,009495	0,07913	11,99668	99,9723
12	0,003319	0,02766	12,00000	100,0000

Рис. 10.14. Власні значення кореляційної матриці та частки дисперсій, що їх пояснює

За правилом Кайзера (власні числа $\lambda_j > 1$) виділено чотири головні компоненти, що визначають екологічний стан країни. Внесок першої компоненти в сумарну дисперсію ознакової множини становить

майже 44%, другої – 16,5%, третьої – 13,3% та четвертої – 9,4%, (рис. 10.15). У цілому ці компоненти пояснюють 83,3% сумарної варіації, що свідчить про високий ступінь факторизації.

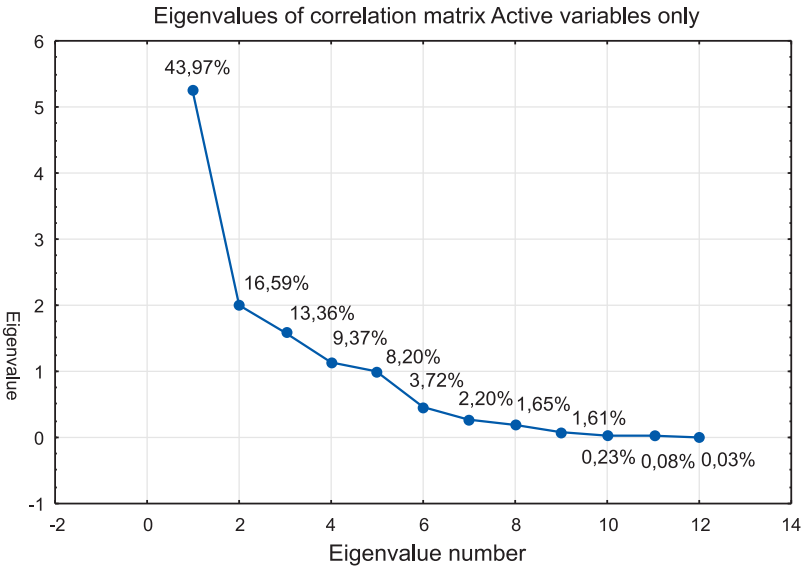


Рис. 10.15. Графік кам'янистого спаду (за правилом Кайзера)

2. Наступним кроком є визначення матриці факторних навантажень, що реалізується за опцією Factor & variables correlations і характеризує щільність зв'язку між *i-тою* ознакою та *j-тою* компонентою (рис. 10.16).

	Factor-variable correlations (factor loadings), based on correlations (Standardize.stw)											
Variat	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Fact.10	Fact.11	Fact.12
X1	0,908	-0,104	0,288	-0,135	0,040	-0,215	0,016	0,114	0,001	0,010	0,012	0,044
X2	0,830	-0,162	0,351	-0,288	0,022	-0,183	-0,007	0,202	0,047	0,025	-0,020	-0,033
X3	0,840	0,450	0,002	0,124	0,040	0,115	-0,136	-0,159	0,077	0,109	0,002	0,000
X4	0,890	0,042	0,193	-0,109	0,026	-0,243	0,064	-0,289	-0,088	-0,046	0,014	-0,012
X5	0,202	0,016	-0,638	0,264	-0,617	-0,290	0,115	0,011	0,057	0,011	-0,002	-0,001
X6	0,769	0,498	-0,128	0,176	-0,103	0,219	0,094	0,171	-0,121	-0,001	0,044	-0,008
X7	0,578	-0,716	0,051	0,177	-0,215	0,223	0,084	-0,035	-0,107	0,035	-0,049	0,004
X8	0,781	0,570	-0,034	0,163	-0,010	0,132	-0,059	0,020	0,064	-0,093	-0,050	0,005
X9	0,563	-0,736	0,131	0,216	-0,123	0,195	-0,073	-0,026	0,120	-0,047	0,044	-0,004
X10	-0,096	0,182	0,123	-0,774	-0,540	0,215	0,049	-0,051	0,025	-0,004	0,004	0,004
X11	-0,446	0,152	0,701	0,289	-0,480	-0,103	-0,297	0,023	-0,065	-0,004	0,001	-0,001
X12	-0,403	0,259	0,734	0,329	-0,068	0,025	0,336	-0,028	0,066	0,011	-0,001	-0,000

Рис. 10.16. Значення факторних навантажень

Значення факторного навантаження в групі повинно бути більше за 0,70, тоді вважається, що показник навантажує цю компоненту. До першої компоненти ввійшли такі показники: викиди забруднюючих речовин в атмосферне повітря від стаціонарних джерел забруднення (0,908); викиди забруднюючих речовин в атмосферне повітря на одиницю площі (0,830); скидання забруднених зворотних вод у поверхневі водні об'єкти (0,840); потужність очисних споруд (0,890); утворення відходів (0,769); утилізація відходів (0,781). Пропонуємо ідентифікувати цю компоненту як забруднення навколишнього середовища.

До другої компоненти ввійшло два показники: утворення відходів I-III класів небезпеки (–0,716) та утилізація відходів I-III класів небезпеки (–0,736). Відходи I-III класів небезпеки вважаються найнебезпечнішими для життя людини, рослинного та тваринного середовища. Кореляція цих показників з другою компонентою свідчить про щільний, але обернений зв'язок, цю компоненту назвемо "поводження з небезпечними відходами".

Третю компоненту навантажує два показники – площа вирубок лісу (0,701) та площа загибелі лісових насаджень (0,734), тому її можна ідентифікувати як знищення лісофонду України.

Четверту компоненту навантажує лише один показник – спалювання відходів (–0,774). Його від'ємне значення свідчить про зворотній зв'язок з компонентою, яку можна трактувати як екологічні ризики у зв'язку зі спалюванням відходів.

Отже, екологічну ситуацію в Україні можна описати чотирма факторами, або напрямами: забруднення навколишнього середовища, поводження з небезпечними відходами, знищення лісофонду країни та екологічні ризики.

3. На третьому етапі слід визначити індивідуальні значення для кожної ГК, що виконується за опцією Factor coordinates of cases (рис. 10.17).

Як свідчать індивідуальні значення першої компоненти, рівень забруднення навколишнього середовища в Дніпропетровській (8,01), Донецькій (5,85), Запорізькій (0,74), Полтавській (0,52), Сумській (0,93) та Харківській (0,44) областях був вищим за середній рівень

за регіонами України. Отримані результати є очікуваними, оскільки вищий рівень забруднення атмосферного повітря та водних ресурсів характерний для промислових центрів України.

	Factor coordinates of cases, based on correlations (Standardize.stw)									
Case	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Fact.1
Vinnitsa	-0,583	0,156	-1,000	-0,703	-0,986	-0,658	-0,070	0,355	0,2697	-0,002
Volyn	-2,007	0,828	2,377	1,450	0,363	0,126	1,290	-0,048	0,3558	0,041
Dnipro	8,010	4,008	-0,232	0,889	-0,052	0,461	-0,196	0,004	0,1038	-0,066
Donetsk	5,853	-3,804	2,205	-0,252	0,778	-1,008	0,153	0,278	-0,1192	0,080
Zhytomyr	-1,959	0,842	1,925	0,566	-1,770	-0,214	-1,072	0,004	-0,3967	-0,141
Zakrpaty	-1,585	0,292	1,177	0,662	1,699	0,490	-0,112	0,146	-0,2185	-0,067
Zapirizh	0,741	-0,319	-0,504	-1,084	0,345	-0,470	0,344	-0,421	0,0450	-0,360
Iv-Frank	-0,159	0,216	1,560	-3,401	-0,796	0,736	-0,235	0,783	0,2926	-0,009
Kyiv	-1,094	0,415	0,657	0,921	-0,534	-0,905	0,153	0,236	0,1000	0,038
Kirovograd	-0,571	0,379	-1,539	-0,686	-0,552	0,216	0,478	0,269	0,1358	-0,076
Lugansk	-0,656	0,140	-0,215	-0,424	1,043	0,211	-0,006	-0,271	-0,0803	0,035
Lviv	-0,172	0,503	0,474	-0,415	0,402	-0,128	-0,576	-0,530	-0,0933	0,420
Mykolaiv	-0,601	-0,159	-0,967	-0,477	0,690	0,582	0,031	-0,094	0,1840	0,232
Odesa	-0,253	0,357	-0,642	-0,922	0,650	0,185	-0,018	-0,787	0,0732	0,210
Poltava	0,521	-0,517	-1,436	0,579	-1,133	0,898	0,898	0,801	-0,7302	0,185
Rivne	-1,884	0,899	2,269	0,295	-0,447	0,421	0,322	-0,218	0,0374	-0,046
Sumy	0,936	-3,357	-0,476	1,690	-1,377	1,610	-0,577	-0,419	0,4077	-0,082
Ternopil	-0,716	-0,501	-1,080	0,624	0,906	-0,086	-0,114	0,239	0,1817	-0,116
Kharkiv	0,437	-0,240	-0,310	-0,974	-1,612	-0,500	0,522	-1,125	-0,3888	-0,087
Kherson	-0,575	-0,362	-1,169	-0,186	0,727	0,051	0,278	-0,092	-0,0560	-0,092
Khmelnytsky	-0,733	-0,030	-1,322	0,760	0,003	-0,998	-0,008	0,110	0,1225	-0,100
Cherkasy	-0,813	0,050	-0,752	0,346	0,299	-0,606	-0,248	0,377	0,0698	0,002
Chernivtsi	-1,153	-0,049	-0,164	-0,146	1,915	0,489	-0,651	0,165	-0,4001	-0,210
Chernihiv	-0,984	0,255	-0,837	0,888	-0,560	-0,902	-0,584	0,238	0,1041	0,211

Рис. 10.17. Значення головних компонент для кожного регіону

Другу компоненту ми визначили, як індикатор поводження з небезпечними відходами. Цей показник був нижчим за середній рівень по Україні у таких областях: Донеччина (–3,80), Запоріжжя (–0,32), Миколаївщина (–0,16), Полтавщина (–0,52), Сумщина (–3,35), Тернопільщина (–0,5), Харківщина (–0,24), Хмельниччина (–0,3) та Буковина (–0,049), що є негативним і свідчить про необхідність екологізації виробництва в цих регіонах. Найбільша частка відходів I-III класів небезпеки зосереджена в Донецькій області – 26,6% від загального обсягу по країні, в Сумській – 25,2% та в Полтавській – 14,5%.

Хоча Дніпропетровська область є потужним промисловим регіоном України, проте частка відходів саме I-III класів небезпеки тут

складає лише 5,7%. Тому за індивідуальним значенням другої головної компоненти (4,0) вважається, що на Дніпропетровщині рівень поводження з небезпечними відходами є вищим за середній, а це є позитивним напрямом екологізації регіону.

Рівень знищення лісофонду України вищий за середній на Волині (2,37), Донеччині (2,21), Житомирщині (1,93), Закарпатті (1,18), Івано-Франківщині (1,56), Київщині (0,65), Львівщині (0,47) та Рівненщині (2,27), що негативно впливає на збереження лісистості України та на відтворення рослинного та тваринного генофондів у цілому.

Четверта компонента засвідчує, що найвищі екологічні ризики, пов'язані зі спалюванням відходів, характерні для Волинської (1,45), Сумської (1,69), Київської (0,92) і Чернігівської (0,88) областей.

Запитання для самоконтролю

1. Які відомі вам пособи нормування показників?
2. Що таке показники стимулятори та дестимулятори? Наведіть приклади.
3. Як розраховується багатовимірна середня?
4. За яким принципом здійснюється стандартизація показників в пакеті Statistica?
5. Що таке кластер та кластеризація?
6. Яка мета кластерного аналізу?
7. Скільки існує сімейств кластерного аналізу? Назвіть їх.
8. Що являють собою ієрархічні кластер-процедури?
9. Що являють собою ітераційні кластер-процедури?
10. Які вам відомі правила об'єднання кластерів?
11. Які вам відомі метрики відстаней?
12. Які вам відомі способи зменшення розмірності інформації?
13. Що таке головні компоненти?
14. Як діє правило Кайзера при виборі головних компонент?
15. Що характеризує напрям (+/–) індивідуальних значень головних компонент?

Завдання для самостійної роботи

Тестові завдання

1. На практиці застосовують такі способи стандартизації:
   а) нормування середнім значенням;
   б) нормування мінімальним значенням;
   в) нормування максимальним значенням;
   г) нормуванням еталонним значенням;
   д) всі відповіді правильні.

Відповіді: 1) а; 2) б, в; 3) а, г; 4) д.

2. Узагальнююча інтегральна оцінка розраховується на основі:
   а) багатовимірної середньої;
   б) середньої геометричної;

в) середньої гармонійної.

Відповіді: 1) а; 2) б; 3) в; 4) а, б, в.

3. В основі агломеративних кластер-процедур лежить принцип:

а) послідовного об'єднання кластерів;

б) поетапного розбиття кластерів.

Відповіді: 1) а; 2) б.

4. Скільки існує сімейств кластерного аналізу:

а) два (ітераційні та ієрархічні);

б) сім;

в) безліч.

Відповіді: 1) а; 2) б; 3) в.

5. Найпоширенішою метрикою відстані між кластерами є:

а) відстань Чебишова;

б) Манхеттенська відстань;

в) Евклідова відстань.

Відповіді: 1) а; 2) б; 3) в; 4) а, б, в.

Практичні завдання (задачі)

Задача 1

За даними таблиці оцініть рівень розвитку зовнішньоекономічної діяльності окремих регіонів України за такими показниками:

X1 – Обсяг валового регіонального продукту (млн грн);

X2 – Експорт товарів (млн дол. США);

X3 – Імпорт товарів (млн дол. США);

X4 – Експорт послуг (млн дол. США);

X5 – Імпорт послуг (млн дол. США);

X6 – Прямі іноземні інвестиції в регіон (млн дол. США);

X7 – Прямі іноземні інвестиції із регіону (млн дол. США).

Вихідні дані для обчислення інтегрального показника ЗЕД

Області України	X1	X2	X3	X4	X5	X6	X7
Вінницька	36191,0	679,3	496,2	59,3	33,2	248,2	0,2
Дніпропетровська	152905,0	9794,9	5357,3	542,9	504,1	8432,3	191,1
Донецька	164926,0	12408,6	4054,6	639,6	621,2	3187,0	5226,9
Київська	68931,0	1973,7	4736,2	477,8	309,4	1835,2	7,0
Одеська	69760,0	1628,4	3438,7	1178,3	268,6	1629,1	53,9
Сумська	26765,0	924,6	645,6	55,4	77,9	386,5	2,7
Тернопільська	18085,0	365,1	349,7	58,3	12,6	63,9	–
Харківська	85315,0	1987,3	2280,0	345,0	181,4	2174,3	12,9
Чернівецька	13757,0	125,3	158,1	17,0	2,4	64,2	–
м. Київ	312552,0	12336,8	26677,5	4018,9	3132,7	27278,1	671,9

Задача 2

За даними таблиці здійсніть кластерний аналіз (за методом *k*-середніх) з розбиттям на три кластери. Визначте, які марки автомобілів ввійшли до кожного кластера. Обгрунтуйте отримані результати. Вихідні дані для кластеризації марок автомобілів

Марки	Ціна	Прискорення авто	Ефективність гальмування	Індекс спроможності тримати дорогу	Пробіг авто
Acura	–0,521	0,477	–0,007	0,382	2,079
Audi	0,866	0,208	0,319	–0,091	–0,677
BMW	0,496	–0,802	0,192	–0,091	–0,154
Buick	–0,614	1,689	0,933	–0,210	–0,154
Corvette	1,235	–1,811	–0,494	0,973	–0,677
Chrysler	–0,614	0,073	0,427	–0,210	–0,154
Dodge	–0,706	–0,196	0,481	0,145	–0,154
Eagle	–0,614	1,218	–4,199	–0,210	–0,677
Ford	–0,706	–1,542	0,987	0,145	–1,724
Honda	–0,429	0,410	–0,007	0,027	0,369
Isuzu	–0,798	0,410	–0,061	–4,230	1,067

Mazda	0,126	0,679	–0,133	0,500	–1,724
Mercedes	1,051	0,006	0,120	–0,091	–0,154
Mitsub.	–0,614	–1,003	0,084	0,382	0,718
Nissan	–0,429	0,073	–0,007	0,263	0,997
Olds	–0,614	–0,734	0,409	0,382	2,114
Pontiac	–0,614	0,679	0,536	0,145	0,195
Porsche	3,454	–2,215	–0,296	0,618	–1,026
Saab	0,588	0,679	0,246	0,263	0,021
Toyota	–0,059	1,218	0,228	0,736	–0,851
VW	–0,706	–0,128	0,102	0,382	0,195
Volvo	0,219	0,612	0,138	–0,210	0,369

** Файл знаходиться у пакеті STATISTICA (File-Open Examples-Datasets-Cars.sta)*

Задача 3

За результатами соціологічного опитування 100 респондентів щодо ступеня задоволеності їх життям визначте оптимальну кількість головних компонент за правилом Кайзера, факторні навантаження, значення ГК для кожного спостереження, побудуйте графіки. Інтерпретуйте головні компоненти та обґрунтуйте отримані результати.

Результати опитування наведені у файлі *Factor.sta* пакету STATISTICA *(File-Open Examples-Datasets)* і містять відповіді на такі запитання:

WORK_1 – самооцінка професійного статусу респондента;
WORK_2 – оцінка умов праці;
WORK_3 – оцінка рейтингу компанії;
HOBBY_1 – оцінка можливостей самореалізації поза роботою;
HOBBY_2 – ефективність відпочинку;
HOME_1 – оцінка матеріального добробуту сім'ї;
HOME_2 – задоволеність соціальним статусом сім'ї;
HOME_3 – психологічний комфорт в сім'ї;
MISCEL_1 – оцінка навколишнього середовища;
MISCEL_2 – оцінка своєї мобільності.

ОСНОВНІ КАТЕГОРІЇ ТА ПОНЯТТЯ

А

Абсолютна статистична величина
Абсолютне значення 1% приросту
Абсолютний приріст
Адекватність показника
Анкетне спостереження

Б

Багатовимірна середня
База порівняння

В

Ваги
Варіанта
Варіаційний розмах
Варіація
Вартісні показники споживання
Вибірка моментна
    багатоступенева
    багатофазна
Вибіркова оцінка
Вибіркове спостереження
Визначальна властивість середньої
Вимірники
Відносна величина динаміки
    структури
    координації
    порівняння зі стандартом
    просторового порівняння
    інтенсивності
Відношення шансів
Вірогідність інформації

Г

Групувальна ознака
Групування
- структурне
- типологічне
- аналітичне
- просте
- комбінаційне
- вторинне

Д

Дециль
Динамічний ряд)
- моментний
- інтервальний

Дисперсія
- загальна
- міжгрупова
- внутрішньогрупова
- середня з групових
- альтернативної ознаки

Добір
- випадковий
- механічний
- розшарований, районований
- серійний

Довірчі межі

Е

Екстраполяція тренду
Ефект
Ефект впливу
Ефективність

З

Закон великих чисел
Закономірність розподілу
Зв'язок функціональний

стохастичний
кореляційний
Згладжування ряду

І

Індекс агрегатний
динаміки
індивідуальний
зведений
міжгруповий
сезонності
середньозважений
територіальний
Індекси доходів населення
ефективності
матеріаломісткості
обороту позики
обсягу споживання
середніх величин
змінного складу
фіксованого складу
структурних зрушень
Індекси цін продукції окремих галузей
Індексна система
Індексні ваги
Індексована величина
Інтегральна, рейтингова оцінка
Інтервал групувань

К

Квартиль
Класифікація
Ковзна (плинна) середня
Код, кодування
Коефіцієнт асиметрії
варіації
взаємної спряженості

випередження
детермінації
ексцесу
еластичності
контингенції
концентрації
кореляції
локалізації
нерівномірності
прискорення (уповільнення)
рангової кореляції
структурних зрушень
регресії
децильний
квадратичний
квартильний
лінійний
осциляції
Кореляційне відношення
Критичний момент спостереження

Л

Лінія регресії
емпірична
теоретична
Логічна формула середньої

М

Макет статистичної таблиці
Медіана
Мода
Модель показника
Моніторинг
Монографічне обстеження

Н

Несуцільне спостереження

О

Об’єкт спостереження
Обстеження основного масиву
Обсяг споживання
Одиниця спостереження
Ознака
    атрибутивна
    дискретна
    кількісна
    неперервна
Опитування

П

Похибка репрезентативності
    випадкова
    відносна
    гранична
    систематична
    стандартна, середня
Правило розкладання варіації
Принципи кредитування і завдання статистики
Процент, процентний пункт

Р

Ряд розподілу

С

Середнє відхилення лінійне
    квадратичне
Середня арифметична
    гармонічна
    геометрична
    зважена
    проста
    розподілу
    хронологічна
Системи зважування індексів
Сталість динаміки

Статистична закономірність
Статистична методологія
Статистична сукупність
Статистична таблиця
Статистичне зведення
Статистичний критерій
Статистичний показник
    інтервальний
    моментний
    обернений
    первинний
    похідний
    прямий
Статистичний ряд
Статистичні реєстри
Ступені свободи
Сучасні статистичні національні класифікації

Т

Темп зростання
    приросту
Тенденція розвитку
Територіальні порівняння цін
Трендове рівняння

Ф

Фактор

Ц

Центр розподілу

Ч

Час спостереження
Частка
    кумулятивна
Частота
    кумулятивна

Ш

Шкала вимірювання

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Андрієнко В.Ю. Статистика: методологія та практика досліджень соціально-економічних явищ: Метод. рек. до вивч. курсу. – К.: Держкомстат України, 2004. – 117 с.
2. Акімова О.В., Дубинська О.С. Статистика в малюнках та схемах: Навчальний посібник. – К.: Центр учбової літератури, 2007. – 168 с.
3. Бек В.Л. Капленко Г.В. Теорія статистики: Практикум. – видання друге, доп.та випр. – Львів: Новий світ-2000, 2007. – 320 с.
4. Вайнберг Дж., Шумекер Дж. Статистика/ Пер. с англ. Л.А. Клименко и Б.И.Клименко; Под ред.. и с предисл. И.Ш.Амирова. Статистика, 1979. – 389 с.
5. Вашків П.Г. та ін. Статистика підприємництва: Навч.посібник/ П.Г.Вашків, П.І.Пастер, В.П.Сторожук, Є.І.Ткач; за ред. П.Г.Вашківа, В.П.Сторожука. – К.: "Слобожанщина", 1999. – 600 с.
6. Герасименко С.С. Фундаментальний характер статистики в ефективному управлінні економікою.// Статистика України. – 2007. – №3. – с.26-28.
7. Єріна А.М., Пал'ян З.О. Теорія статистики: Практикум. – 6-те вид., стер. – К.: Знання, 2008. – 267 с.
8. Кулинич О.І., Кулинич Р.О. Теорія статистики: Підручник. – 3-тє вид., перероб. і доп. – К.: Знання, 2006. – 294 с.
9. Лугінін О.Є. Статистика. Підручник. 2-ге вид., перероб. та доп. – К.: Центр учбової літератури, 2007. – 608 с.
10. Мазуренко В.П. Статистика: Навч.-метод.посіб. для самостійного вивчення дисципліни. – К.: Видавничо-поліграфічний центр "Київський університет", 2006. – 315 с.
11. Мармоза А.Т. Практикум із статистики. – Київ: Кондор, 2005. – 512 с.
12. Матковський С.О., Гальків Л.І., Гринькевич О.С., Сорочак О.З. Статистика: Навчальний посібник. – Львів: "Новий світ-2000",

2009. – 430 с.
13. Опря А.Т. Статистика: (з програмованою формою контролю знань). – К.: Урожай, 1996. – 448 с.
14. Положення про Державний комітет статистики України // Інформ. бюл. Держкомстату України. – 1997. – №11-12.
15. Практикум з навчальної дисципліни "Статистика": навч.посіб./ [Матковський С.О., Вдовин М.Л., Гринькевич О.С. та ін.]; за ред. С.О.Матковського. – Львів: Видавництво ЛНУ ім. Івана Франка, 2009. – 215 с.
16. Про державну статистику: Закон України // Голос України. – 1992. – 21 жовтня 1992.
17. Состояние коррупции в мире // Всеукраинская газета "ВВ". – 2008. – №184(2104). – с.7, 9.
18. Статистика: Навч.-метод.посібник для самост.вивч.дисц./ А.М. Єріна, Р.М. Моторин, А.В. Головач та ін.; За заг.ред. А.М.Єріної, Р.М.Моторина. – К.: КНЕУ, 2001. – 448 с.
19. Статистика: Підручник/ А.В.Головач, А.М.Єріна, О.В.Козирєв та ін.; За ред.. А.В.Головача, А.М.Єріної, О.В.Козирєва. – К.: Вища шк., 1993. 623 с.
20. 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. – К.: КНЕУ, 2007. – 304 с.
21. Статистика: теоретичні засади і прикладні аспекти. Навчальний посібник./ Р.В.Фещур, А.Ф.Барвінський, В.П.Кічор та інші; За наук.ред. Р.В.Фещура. – 2-е вид.оновлене і доповнене. – Львів: "Інтелект-Захід", 2003. – 576 с.
22. Чекотовський Е.В. Підходи і напрями вивчення історії статистичної науки// Статистика України. – 2006. – №1. – с.110-115.
23. Чекотовський Е.В. Підходи і напрями вивчення історії статистичної науки// Статистика України. – 2006. – №2. – с.90-93.
24. Штангрет А.М., Копилюк О.І. Статистика: Навчальний посібник. – Київ: Центр навчальної літератури, 2005. – 232 с.

СТАТИСТИКА

Навчальне видання

Комп'ютерна верстка – Грищенко І. О.

Підписано до друку 03.11.2022. Формат 60x84 1/16.
Папір офсетний. Гарнітура Times New Roman.
Ум.-друк. арк. 15,31. Обл.-вид.арк. 16,47.
Тираж 100 прим. Зам. № 00-00-0.