

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

СТАТИСТИКА

Навчальний посібник

За загальною редакцією
д-ра екон. наук, професора О. В. Раєвської

Харків
ХНЕУ ім. С. Кузнеця
2019

УДК 311(075.034)

C78

Авторський колектив: д-р екон. наук, професор *О. В. Раєвнева*; канд. екон. наук, доцент *І. В. Аксьонова*; канд. екон. наук, доцент *О. І. Бровко*; д-р екон. наук, доцент *Л. В. Гриневич*; канд. екон. наук, доцент *В. І. Дериховська*; канд. екон. наук, доцент *Г. І. Свидло*; канд. екон. наук, доцент *І. А. Серова*; канд. екон. наук, доцент *В. О. Шликова*.

Рецензенти: завідувач кафедри статистики, обліку та економічної інформатики Дніпровського національного університету імені Олеся Гончара, д-р екон. наук, професор, акад. Академії економічних наук України *О. К. Єлісєєва*; завідувач кафедри статистики Одеського національного економічного університету, канд. екон. наук, професор *А. З. Підгорний*; доцент кафедри статистики обліку та аудиту Харківського національного університету імені В. Н. Каразіна, канд. екон. наук *О. А. Чуприна*.

Рекомендовано до видання рішенням ученої ради Харківського національного економічного університету імені Семена Кузнеця.

Протокол № 9 від 27.05.2020 р.

Самостійне електронне текстове мережеве видання

Статистика [Електронний ресурс] : навчальний посібник C78 / *О. В. Раєвнева, І. В. Аксьонова, О. І. Бровко* ; за заг. ред. д-ра екон. наук, професора *О. В. Раєвневої*. – Харків : ХНЕУ ім. С. Кузнеця, 2019. – 389 с.

ISBN 978-966-676-770-0

Із метою поглиблення спеціальних знань зі статистики, практико-орієнтований матеріал подано у вигляді взаємозв'язку компонентів триєдиної системи: прикладних теоретичних знань зі статистики, економіко-математичних методів оброблення статистичної інформації та інформаційної підтримки оброблення й аналізу даних, що відповідає сучасним вимогам підготовки компетентного фахівця в галузях народного господарства та підвищує якість набуття студентами отриманих професійних компетентностей.

Рекомендовано для студентів усіх спеціальностей, аспірантів, викладачів, економістів, аналітиків, які провадять прикладні статистичні дослідження.

УДК 311(075.034)

© Раєвнева О. В., Аксьонова І. В.,
Бровко О. І. та ін., 2019

© Харківський національний економічний
університет імені Семена Кузнеця, 2019

ISBN 978-966-676-770-0

Передмова

Перехід до ринкових умов господарювання передбачає розумне втручання держави в економіку, враховуючи те, що ступінь і методи такого втручання варіюють залежно від обраної моделі розвитку ринку. Але будь-яка модель реформування економіки базується на наявності систематизованої, повної та своєчасної інформації щодо процесів та явищ, взаємозв'язок яких визначається практичними потребами суспільства.

Тільки знаючи реальне становище, можна правильно розподілити ресурси та прогнозувати наслідки явищ, що стають особливо актуальними в умовах політичних та економічних обмежень на сучасному етапі розвитку країни.

Джерелом формування суспільної свідомості за умов розширення гласності та доступності зведеної інформації зі збереженням принципу конфіденційності індивідуальних даних виступає статистика.

У сучасних умовах необхідно мати новий підхід до статистики, який складається з послідовного переходу від фрагментарного принципу до системного реформування. Цей процес передбачає компетентне, взаємозв'язане вдосконалення всіх елементів статистичного спостереження з урахуванням ринкового попиту на інформацію, нових вимог до якості інформації з боку органів державної влади, комерційних структур, приватних осіб та інших споживачів.

Головним завданням статистики в ринковій економіці є удосконалення методології розрахунку статистичних показників, які комплексно характеризують стан сучасної національної моделі економіки країни з урахуванням міжнародних стандартів, приведення їх до системного виду, що відповідають потребам відповідного етапу соціально-економічного розвитку країни.

Ключовим напрямом розвитку та реформування статистики є забезпечення взаємозв'язку статистичних показників, що відбивають господарські процеси як на мікро-, так і на макрорівні.

Необхідність інтеграції інформаційних масивів з безлічі форм звітності, наявність компетентностей щодо коректної роботи з великими масивами

даних, застосування знань законів статистики на практиці з використанням ІТ-технологій і програмного забезпечення висувають високі вимоги до спеціалістів з вищою освітою.

Майбутні фахівці в межах економічної освіти повинні опанувати знання з основ статистичної методології. Це надасть їм змогу бути професіоналами в будь-якій сфері господарювання чи управління, стати універсальними спеціалістами, що вільно володіють компетентностями в галузі статистики й економіки, фінансах і юриспруденції, менеджменту та підприємництва.

Однією з важливих функцій, що виконує статистична наука є забезпечення підготовки фахівців у галузі прикладного економіко-статистичного аналізу, формування навичок використання та опрацювання великих масивів статистичної інформації засобами інформаційних технологій для ухвалення рішень в економіці й інших видах діяльності. Уміння користуватися основами статистики дозволяє орієнтуватися в існуючих суперечностях, критично осмислювати різні статистичні дані, грамотно інтерпретувати їх.

У навчальному посібнику "Статистика" розглянуто базові питання статистичної методології, що дадуть змогу сформуванню у майбутніх фахівців уміння комплексно характеризувати стан сучасної національної моделі економіки країни з урахуванням міжнародних стандартів і тренду розвитку процесів та явищ, навчить визначати характерні відмінності перебігу економічних процесів, будувати адаптивні до змін ринкового оточення системи показників, які віддзеркалюють різні аспекти соціально-економічного розвитку країни.

Навчальний посібник "Статистика" підготовлено на основі курсу лекцій дисципліни "Статистика", що розроблено та викладається в Харківському національному економічному університеті імені Семена Кузнеця.

Автори вдячні за ретельне рецензування рукопису, цінні зауваження та за допомогу в підготовці навчального посібника до видання доктору економічних наук професору Єлісеєвій О. К., кандидату економічних наук професору Підгорному А. З. і кандидату економічних наук доценту Чуприні О. А.

Розділ 1

Введення до статистики

1. Методологічні засади статистики

Основні питання:

- 1.1. Поняття про статистику як суспільну науку, її виникнення та розвиток.
- 1.2. Предмет і методологічні основи статистики.
- 1.3. Етапи статистичного дослідження та специфічні прийоми статистичного аналізу.
- 1.4. Поняття та категорії в статистиці.
- 1.5. Організація та завдання статистики в сучасних умовах.

1.1. Поняття про статистику як суспільну науку, її виникнення та розвиток

Термін "статистика" походить від латинських слів "status" (певний стан явища, стан речей) і "stato" (держава). Він був введений у науковий обіг у 1749 р. німецьким вченим Г. Ахенвалем у трактаті під назвою "Статистика", в якому наводився опис політичного устрою держав Європи.

Історично розвиток статистики пов'язаний із розвитком держав, з потребами державного управління. Господарські та військові потреби вже в давню епоху історії людства вимагали даних про населення, його склад, майнове становище. Перші роботи такого роду відзначені навіть у священних книгах різних народів [2; 5].

У Давньому Єгипті склалися розвинені системи господарювання та адміністративного обліку. Контролювалося використання земель, їжі, робочої сили, складалася відомість доходів і збитків тощо.

У Стародавньому Китаї ще в XXIII ст. до н. е. збиралися зведення про чисельність населення, його склад за статтю та віком, прибутковість земель і зміни в торгівлі.

В античних Афінах був організований облік народжених; молоді люди, які досягли 18 років, вносилися в списки військовозобов'язаних, а після досягнення 20 років реєструвалися як повноправні громадяни.

Склалися земельні кадастри, до яких вносилися відомості про будівлі, рабів, худобу, реманент, отримувані доходи; з'явилися описи держав. Велика заслуга в цьому належить грецькому філософу Аристотелю (384 – 392 рр. до н. е.) – він склав опис 157 міст і держав свого часу.

У Стародавньому Римі Сервій Тулій створив ценз для проведення перепису вільних громадян. Чинovníки цензу – цензори встановлювали прізвище, стать і вік всіх членів сім'ї, а також перераховували їх майно. Засновник Римської імперії Октавіан Август, за словами Тацита, керував складанням спеціальної книги "Бреваріум", в якій вказувалися кошторисні асигнування, зведення про стан фінансів армії, чисельність громадян. Це свідчить про розвиток бюджетного обліку в масштабах всієї держави. Але особливого розвитку в Римі отримав облік приватних господарств. Голова кожної сім'ї був зобов'язаний вести книгу домашнього майна щодо доходів і витрат.

Середньовіччя залишило унікальну пам'ятку – "Книги страшного суду" (1061 р.) – зведення матеріалів загального перепису населення Англії та його майна (включаючи дані про 240 тис. подвір'їв). На Русі в X – XII ст. збиралися різноманітні відомості, пов'язані з оподаткуванням населення.

Зростання суспільного виробництва, розширення торговельних і міжнародних відносин у період становлення капіталізму послуговували стимулом розвитку обліку та статистики. Поряд із простим обліком в Італії (приблизно початок XIV ст.) з'являється система подвійного обліку, за якого операція фіксується двічі. Значно зростає потреба в аналізі економічної кон'юнктури, тому обсяг статистичної інформації особливо різко збільшується: потрібні відомості про розміри та розміщення промислового та сільськогосподарського виробництва, ринки збуту товарів, ринки праці, сировинні ресурси тощо. З часом збирання даних про масові суспільні явища набуло регулярного характеру. З середини XVIII ст. завдяки зусиллям бельгійця – математика, астронома і статистика А. Кетле (1796 – 1874 рр.) були розроблені правила переписів населення і запропоновано регулярність їх проведення в розвинених країнах. Для координації розвитку статистики за ініціативою А. Кетле проводилися міжнародні статистичні конгреси, а в 1885 р. був заснований Міжнародний статистичний інститут, який існує донині [68].

Розширення практики обліково-статистичних робіт у різних країнах сприяло формуванню статистичної науки.

Статистика як наука почала розвиватися з середини XVII ст., коли в Європі склалися **дві основні школи**: в Англії – *математична школа статистики (школа політичних арифметиків)*, у Німеччині – *описова школа статистики (державоведення)* [2].

Яскравим представником **математичної школи статистики** (школи політичних арифметиків) був В. Петті (1623 – 1687 рр.; англійський економіст і статистик), він заслужено вважається засновником статистичної науки, оскільки першим широко застосував математику для економічного аналізу. "Політична арифметика" В. Петті ґрунтувалася на точному спостереженні, кількісному описі економічних явищ на основі підрахунку числових даних. Політичні арифметики за мету ставили вивчення суспільних явищ за допомогою числових характеристик. Вони намагалися виявити закономірності розвитку та взаємозв'язку економічних явищ за допомогою математичних розрахунків, усвідомлюючи необхідність урахування в статистичних дослідженнях вимог закону великих чисел, оскільки закономірність може виявитися лише за великого обсягу статистичної сукупності.

Відомим представником **описової школи статистики** був німецький вчений Г. Конринг (1606 – 1681 рр.) – відомий німецький полігістор, історик і державознавець, який був професором медицини та політики в Гельмштадті. Г. Ахенвалль (1719 – 1772 рр.) – знаний статистик, професор у Геттінгені; є засновником теорії статистики, основні положення якої оприлюднені в 1749 р. Саме Г. Ахенвалль уперше в Геттінгенському університеті почав читати навчальну дисципліну "Статистика". Основною метою статистики представники німецької описової школи вважали опис території, населення, державного устрою, релігії, зовнішньої політики тощо. Отже, предмет статистики не обмежувався тими явищами, які мають числову характеристику. Більш того, ранні представники цього напрямку уникали користування числовими даними, і лише пізніше (в середині XVIII ст.) числові дані поступово завоювали право бути включеними до описової статистики. Друга особливість підходу полягала в тому, що у роботах представників школи був відсутній аналіз закономірностей і взаємозв'язків, притаманних суспільним процесам. Отже, те, що представники описової статистики називали статистикою, було досить далеким від дійсної статистики в сучасному її розумінні.

Отже, від німецької описової школи сучасна статистика взяла систему словесного опису соціально-економічних явищ без цифр; від англійської

школи політичної арифметики – вивчення суспільних явищ за допомогою числових характеристик, статистичне узагальнення отриманих характеристик з метою виявлення закономірностей розвитку досліджуваних явищ.

У першій половині XIX ст. виник **третій напрям статистики – статистико-математичний** (А. Кетле – один з родоначальників наукової статистики), представники якого вважали основою статистики теорію ймовірностей. Вчені, які розвивали розглядуваний напрям: А. Кетле – засновник вчення про середні величини; англійці Ф. Гальтон (1822 – 1911 рр.) і К. Пірсон (1857 – 1936 рр.) використовували математичні методи статистики в біології; американці Р. Фішер (1890 – 1962 рр.), М. Мітчелл (1874 – 1948 рр.) і В. Госсет (1876 – 1937 рр., псевдонім – Стьюдент) використовували в статистичних дослідженнях методи теорії ймовірностей [68].

Методологію статистичних досліджень розвивали та вдосконалювали російські та вітчизняні вчені: І. Ф. Герман (1755 – 1815 рр.), Д. М. Журавський (1810 – 1856 рр.), які вивчали питання взаємодії статистики та політекономії, постановку статистичного спостереження та аналізу статистичних даних, розробляли теорію групування; Ю. Є. Янсон (1835 – 1893 рр.) – основоположник статистичного аналізу; О. І. Чупров (1842 – 1908 рр.) – вивчав питання аналізу зв'язку та залежності суспільних явищ, досліджував проблеми стійкості динамічних рядів. Одним з основоположників вітчизняної демографічної статистики є М. В. Птуха (1884 – 1961 рр.). Як теоретик статистики населення, М. В. Птуха розвивав теорію А. Кетле про "середню людину". Як методист-практик – розробив схему вивчення демографічних процесів і запропонував методику її реалізації. Його висновки у вивченні шлюбності та смертності населення здобули світове визнання. Велике значення мають його методи побудови сумарних таблиць смертності [77]. Сучасна статистична методологія знайшла розвиток також у працях таких вчених-статистиків, як В. С. Немчінов, С. Г. Струмилін, Б. С. Ястремський, А. Я. Боярський, Т. В. Рябушкін, С. С. Сергеев та ін.

Розвиток статистичної науки, розширення сфери практичної статистичної роботи призвели до зміни змісту терміну "статистика", різні трактування якого наведені в табл. 1.1.

Зміст поняття "статистика"

Автори	Зміст терміна
Лугінін О. Є. [36]	Статистика – це галузь практичної діяльності, економічної науки та навчальна дисципліна з вивчення способів збирання, оброблення й аналізу даних про масові соціально-економічні явища та процеси
Мармоза А. Т. [42]	Під терміном "статистика" розуміють три пов'язані між собою значення: 1) цифри, які характеризують рівні, розміри й обсяги тих чи інших суспільних явищ; 2) специфічну галузь практичної діяльності, спрямовану на збирання, накопичення, оброблення й аналіз даних, які характеризують населення, економіку, культуру, освіту та інші явища суспільного життя; 3) самостійну суспільну науку, яка спрямована на розроблення методів збирання, зведення, оброблення, аналізу та теоретичного узагальнення цифрових даних про явища суспільного життя
Герасименко С. С., Головач А. В., Єріна А. М., Козирев О. В., Пальян З. О., Шустіков А. А. [71]	Слово "статистика" (від лат. status – стан речей) – це синонім сукупності фактів, певних відомостей про соціально-економічні явища та процеси. Визначальною рисою таких відомостей є кількісна характеристика. Статистикою називають також науку, яка об'єднує принципи та методи роботи з масовими числовими даними
Осауленко О. Г., Васечко О. О., Пугачова М. В. [51]	Статистика (statistics): 1) галузь знань – складна та розгалужена система наукових дисциплін, що мають певну специфіку та вивчають кількісну сторону масових явищ і процесів у нерозривному зв'язку з їх якісною стороною. Основними розділами статистики є: теорія статистики, де розглядаються найбільш загальні категорії, принципи та методи статистичної науки; статистика економічна – вивчає явища та процеси, які мають місце в економіці; статистика соціальна, що аналізує соціальні явища та процеси; статистика математична, основними завданнями якої є статистична перевірка гіпотез, оцінювання розподілу статистичних ймовірностей та його параметрів, вивчення статистичної залежності, визначення основних числових характеристик випадкових вибірок; 2) галузь практичної діяльності, що охоплює збирання, оброблення, аналіз і публікацію статистичної інформації про явища та процеси суспільного життя; 3) комплекс цифрових відомостей, що характеризують стан масових явищ і процесів суспільного життя (або їх сукупність)

У сучасному розумінні **статистика** – це:

сукупність підсумкових відомостей, що кількісно характеризують різні аспекти суспільного життя: виробництво, розподіл і обмін товарами, політику, культуру тощо;

практична діяльність щодо збирання, оброблення й аналізу кількісних даних про суспільне життя та їх публікація;

навчальна дисципліна, яка вивчає кількісну сторону масових явищ і процесів у нерозривному зв'язку з їх якісним вираженням з метою виявлення закономірностей їх розвитку.

У статистиці, як навчальній дисципліні, виділяють *три рівні* [11; 12]:

- **загальна теорія статистики**, яка передбачає розроблення понятійного апарату та системи категорій статистичної науки, загальних принципів і правил проведення статистичних досліджень, універсальних методів оброблення інформації – тобто створення загальної методології статистичного дослідження масових суспільних явищ;

- **економічна та соціальна статистики**. *Економічна статистика* провадить дослідження економічних явищ і процесів, їх кількісне оцінювання і розроблення синтетичних економічних показників – таких, як: національне багатство, валовий внутрішній продукт, національний дохід та ін. *Соціальна статистика* – призначена для дослідження і розроблення узагальнюючих показників у різних галузях суспільного життя: освіта, охорона здоров'я, культура, політика, наука тощо;

- **галузі економічної та соціальної статистики**. В економічній статистиці виділяються такі її галузі, як: статистика промисловості, сільськогосподарства, транспорту, будівництва, торгівлі, зв'язку і т. ін. У соціальній статистиці: статистика науки, права, охорони здоров'я, політична статистика, статистика населення, статистика рівня життя тощо.

Завданнями галузевих статистик є: вивчення масових явищ, що мають місце у відповідних галузях і сферах суспільно-економічного життя; розроблення узагальнюювальних показників цих галузей; виявлення тенденцій і закономірностей розвитку зазначених сфер.

Статистика розвивається як єдина наука, і розвиток кожної галузі сприяє її вдосконаленню в цілому [41].

Статистика має величезне *пізнавальне значення*, оскільки:

дає чисельне та змістовне висвітлення досліджуваних явищ і процесів, слугуючи надійним способом оцінювання дійсності;

дає доказову силу економічним висновкам, дозволяючи перевірити висунуті гіпотези й окремі теоретичні положення;

розкриває взаємозв'язки між явищами, показує їх конкретну форму та силу;

виявляє закономірності розвитку нових явищ, надаючи їм кількісну та якісну характеристику.

Отже, **статистика** як суспільна наука розробляє методологію статистичного дослідження, яка використовується іншими науками. Знання статистики необхідне сучасному фахівцю для ухвалення рішень в умовах, коли аналізовані явища підпадають під вплив випадковостей, для аналізу елементів ринкової економіки, прогнозування та розроблення сценаріїв поведінки економічних систем за зміною умов їх функціонування.

1.2. Предмет і методологічні основи статистики

Предмет історії статистики – це процес виникнення та розвитку статистичного обліку та теорії статистики в концепції статистичної науки. Розвиток статистики визначається перш за все розвитком суспільства та держави, їх соціально-економічними потребами [68].

На початку XIX ст. з'явилася необхідність у розробленні та вдосконаленні методів статистики, які забезпечували би збирання і узагальнення масових облікових даних. Це потребувало створення теорії статистики. Засновником теорії статистики є А. Кетле (1796 – 1874 рр.). З одного боку, він був засновником науки про суспільство – соціальної фізики, або соціології, з іншого – засновником теорії статистики. Основні роботи – "Розрахунок вірогідних злочинів", "Листи про теорію ймовірностей". Успіх цих робіт у сучасників пояснюється тим, що науковцю вдалося на масовому фактичному матеріалі показати наявність закономірностей суспільного життя, причому на прикладі найбільш гострих питань суспільства [86].

У роботах Кетле синтезувалися обидва попередні напрями розвитку статистичної думки. Як представник описової школи він визначав статистику як науку, що вивчає державу, проте стверджував, що статистика є наукою про суспільство – "соціальною фізикою", яка займається вивченням життя "соціального тіла". Кетле довів наявність стійких закономірностей, властивих соціальним явищам. Для пояснення причин формування цих закономірностей він запропонував учення про постійні та пертурбаційні (індивідуальні, такі, що діють нерівномірно й у різних напрямках)

причини, під впливом яких складається явище. Сумісну дію обох причин він вважав за найважливішу мету предмета статистичного пізнання. Погашення індивідуальних відмінностей, які викликані випадковими причинами, на думку Кетле, і могло забезпечуватися масовістю даних [41; 82; 83].

Як основний прийом статистичного аналізу Кетле висунув середні величини. Вчений багато в чому сприяв проведенню перепису населення і розробленню методології його проведення. Об'єктом перепису встановлювалося наявне населення. Одиницею спостереження – людина або сім'я. Була рекомендована періодичність проведення – 1 раз на 10 років.

Заслуга А. Кетле полягає також в обґрунтуванні таких положень статистики, як:

- предмет статистики – об'єктивні закономірності, що визначають розвиток суспільства;
- усі явища складаються під сумісною дією незалежних один від одного загальних та індивідуальних причин;
- основою методології статистичного пізнання є масове спостереження та узагальнювальні показники, що забезпечують погашення випадковостей;
- середні величини є найважливішим прийомом, що дозволяє визначити реально існуючі типи досліджуваних явищ;
- теорія ймовірностей є теоретичною основою статистики. Вона розкриває дію загальних та індивідуальних причин, дозволяючи дати оцінку надійності узагальнюючим статистичним показникам;
- можливе та необхідне збирання масових статистичних даних на прикладі перепису населення.

Упродовж півстоліття статистики різних країн розділялися за своїм ставленням до ідей Кетле. У Німеччині сильнішою була позиція антикетлеанців. Наприклад, відомий психолог М. Дробіш (1802 – 1896 рр.) вважав, що головним завданням статистики є не перерахунок, скільки здійснено злочинів, скільки вироблено продукції, а відповідь на питання: чому здійснено злочини, чому не випущена необхідна кількість продукції тощо [3].

Німецькому вченому, представникові описової школи Г. Рюмеліну належать твердження, що слово "статистика" має два різні сенси та визначає дві різні науки: статистика як наука, яка вивчає кількісні методи, що застосовуються в будь-якій сфері людської діяльності; статистика як державознавство, або суспільна наука.

Під впливом ідей Кетле формувався англійський статистик А. Боулі (1869 – 1957 рр.), який об'єднав ідеї політичних арифметик з новітніми тенденціями, пов'язаними з використанням математичних методів, особливо теорії ймовірностей. Статистика трактувалася ним як наука про середні [68].

Уникнути впливу ідей Кетле вдалося французькому соціологові Ф. Ле Пле (1806 – 1882 рр.). Для нього мета статистики – це точне спостереження або опис соціальних фактів, кращим джерелом для отримання яких він вважав експертів, "соціальних авторитетів". Ле Пле ввів монографічний метод соціально-економічної статистики, підкреслюючи, що глибина дослідження досягається тим, що розглядаються не всі явища, а тільки типові випадки.

Ідеї Ле Пле сприяли формуванню двох основних шкіл бюджетної статистики XIX ст.:

- бельгійської, яка приділяла основну увагу питанням пошуку раціональних норм споживання їжі;
- німецької, яка займалася пошуками закономірностей у структурі бюджету.

Серед представників німецької школи можна також виділити Е. Енгеля і А. Швабса. Енгелю належить заслуга у визначенні одиниці спостереження в бюджетній статистиці: оскільки сім'ї розрізняються за розмірами та складом, то їх бюджети неспівставні, а тому необхідна нова споживча одиниця.

Великим досягненням бюджетної статистики було виведення двох законів:

- закону Енгеля – із зростанням доходу в сім'ї її витрати на харчування зростають абсолютно, але зменшуються відносно;
- закон Швабса – із зростанням доходу сім'ї її витрати на житло ростуть абсолютно, але зменшуються відносно.

На початку XIX ст. практично в усіх країнах були відсутні спеціальні статистичні установи. Тільки у Швеції в 1748 р. було створено Табельну комісію – державну установу, яка організовувала статистику щодо населення в країні. На рубежі XVIII – XIX ст. було створено статистичне бюро у Франції. Перша відомча статистична установа з'явилась у 1841 р. у Бельгії за наполяганням А. Кетле – Центральна статистична комісія.

Основна частина робіт з практичної статистики виконувалася адміністративними органами або окремими дослідницькими установами.

Створення спеціальних статистичних установ свідчить про високу оцінку значення статистики та визнання її справою державної важливості. Найбільший прогрес відбувався у сфері статистики населення. У міру розвитку перепису як джерела інформації з'явилася тенденція її диференціації і спеціалізації.

До кінця XIX ст. був накопичений багатий статистичний матеріал, за яким почали видаватися збірки-щорічники, що включають динамічні ряди за основними показниками.

Розвитку міжнародних зв'язків статистики сприяло створення в 1839 р. Американської статистичної асоціації, мета якої – збирання, зберігання і оброблення статистичної інформації в різних країнах.

Наприкінці XIX ст. зародилися теорії стійкості, кореляції та регресії. Теорія стійкості, автором якої є німець В. Лексис (1837 – 1914 рр.), виникла як спроба пояснити повторюваність даних у динаміці, виявлену А. Кетле. В опублікованій у 1879 р. роботі "Про теорію стабільності статистичних рядів" Лексис виклав висунуту ним теорію стійкості, яка породила багато послідовників і супротивників, а також вплинула на розвиток статистики в XX ст.

В. Лексис розумів стійкість ряду динаміки як випадковість відхилень рівня динамічного ряду від загальної середньої. У його теорії (на відміну від Кетле, якому нормальний розподіл здавався повсюдним) нормальний розподіл трактується як результат класифікації даних. Щоб виключити варіацію даних, викликану зовнішніми причинами, та досягти нормальної стійкості, науковець запропонував використовувати метод найменших квадратів.

В. Лексис також ввів розмежування типів динаміки:
еволюторний – головний прояв основних тенденцій;
ундуляторний – хвилеподібний розвиток у часі;
періодичний – правильне повторення хвиль;
осциляторний – безладне коливання рівнів.

Дослідник розробив формули, які дозволяли виміряти дисперсію трьох видів: *загальну, внутрішньогрупову та міжгрупову*. Також вчений зробив значний внесок у розроблення методів демографії, удосконалення складання таблиць смертності. Створений ним *графічний метод демографічного аналізу* отримав назву *демографічної сітки Лексиса*. Послідовниками В. Лексиса були В. І. Борткевич, А. А. Чупров, А. А. Марков.

У XIX ст. остаточно визначилися основні риси статистичного методу: масове спостереження, узагальнення даних, аналіз. Надалі ідеї Ф. Гальтона, К. Пірсона та В. Лексиса отримали теоретичне узагальнення в концепції стохастичної теорії статистики.

Уперше методи статистики у вирішенні проблем нової еволюційної теорії застосував Ф. Гальтон, двоюрідний брат Ч. Дарвіна. Він пов'язав кореляцію із суттю еволюційного учення на прикладі спадкоємства зростання, а також вивчив зв'язок антропологічних змінних з інтелектуальними здібностями людини.

У 1896 р. Гальтон дав визначення кореляції, побудувавши теоретичну модель зміни для змінних, ввів поняття лінії регресії і кореляційного індексу R. Теоретичне обґрунтування висновків Ф. Гальтона та їх практична перевірка належать К. Пірсону.

Отже, праці вчених XIX ст. підготували швидкий розвиток математико-статистичних методів, їх обґрунтування і застосування.

Сучасна статистика – це особлива наука, що має свій *об'єкт, предмет і специфічні методи дослідження*.

Об'єктом вивчення статистики є масові явища та процеси будь-якої природи, у тому числі в економіці. Це означає, що статистичні показники завжди є результатом узагальнення деякої сукупності фактів [3; 9; 41].

Масовому явищу притаманні такі *ознаки*:

- складається з безлічі індивідуальних елементів, кожен елемент – це реально існуючі матеріальні об'єкти, що володіють деякими загальними властивостями, які іменуються в статистиці ознаками;
- характеризується варіацією, тобто відмінністю в числових значеннях окремих одиниць сукупності;
- часткова або повна незалежність елементів один від одного.

Статистика вивчає масові суспільно-економічні явища в конкретних умовах місця і часу.

Предметом статистики виступають розміри та кількісні співвідношення масових суспільних явищ у нерозривному зв'язку з їх якісним аспектом з метою виявлення закономірностей їх розвитку [3; 9; 41].

Предмету статистики притаманні три основні риси:

- 1) статистика вивчає суспільні явища, тому вона наука суспільна;
- 2) статистика вивчає кількісний бік соціально-економічних явищ і процесів у зв'язку з їх якісною сутністю. Ця риса відрізняє статистику від

математики, оскільки статистика оперує не абстрактними, а конкретними числами;

3) статистика вивчає масові явища та процеси, оскільки тільки в масі спостережень можна виявити закономірності розвитку.

Свій предмет статистика вивчає на основі особливої методології, що є сукупністю загальних правил (принципів) і спеціальних прийомів та методів статистичного дослідження.

Загальні правила статистичного дослідження виходять із положень економічної теорії та політекономії. Вони складають **теоретичну основу статистики**.

Економічна теорія і політекономія пояснюють сутність досліджуваних явищ або процесів, закони їхнього розвитку в конкретних умовах. Водночас статистика збагачує соціально-економічні науки отриманими в дослідженнях числовими даними, що використовуються для підтвердження або спростування висунутих теоретичних гіпотез.

Методологічною основою статистики є діалектичний метод пізнання та специфічні методи статистики, відповідно до яких статистика вивчає масові явища у їх взаємозв'язку, в русі та зміні, виявляючи їх кількісні та якісні характеристики. Для цього використовують такі спеціальні статистичні методи та прийоми, як, наприклад, зведення і групування, індексний метод, кореляційно-регресійний метод та ін. [5; 9; 86].

Статистика пов'язана з математикою. У них загальні методи оброблення й оцінювання даних, але різні предмети пізнання. Математична статистика вивчає закономірності масових явищ в абстрактній формі. Статистика як суспільна наука характеризує розміри та співвідношення суспільних явищ у конкретних умовах їх існування і розвитку. Передумовою використання статистичних методів у конкретному дослідженні має бути визначення сутності досліджуваного явища, його істотних властивостей. Теоретичний аналіз дає всебічне уявлення про природу та логіку предмета пізнання. Це – об'єктивна основа методологічних рішень. Особливості статистичної методології пов'язані: з точними вимірюваннями та кількісним описом масових суспільних явищ; з використанням узагальнювальних показників для характеристики об'єктивних статистичних закономірностей.

Статистичний аналіз масових явищ і процесів є необхідною ланкою в системі управління економікою і державою в цілому. За допомогою статистики здійснюється "зворотний зв'язок", тобто потік інформації йде

від об'єкта до суб'єкта управління – керівництву підприємств, об'єднань, територіальних, галузевих і центральних органів влади. Без всебічної та своєчасної інформації ефективні управлінські рішення неможливі.

1.3. Етапи статистичного дослідження та специфічні прийоми статистичного аналізу

У загальному сенсі під дослідженням розуміють розвиток знань або систематичне розслідування з метою встановлення фактів. **Статистичне дослідження** – це збирання інформації і вивчення процесу причинно-наслідкового зв'язку в спостережуваних явищах з метою отримання інформації для ухвалення обґрунтованих управлінських рішень. Кінцевою ціллю статистичного дослідження є визначення тенденцій розвитку соціально-економічних явищ і процесів на всіх рівнях економіки [68].

Інформаційне забезпечення статистичного дослідження включає такі етапи.

Перший етап. *Інформаційна компонента статистичного дослідження.* На цьому етапі передбачається відокремлення носіїв інформації в рамках обраної мети дослідження, що дозволяє вибрати джерела та почати збирання початкової інформації.

У результаті з'являються потоки необробленої інформації, яка в процесі збирання накопичується, створюючи масив первинної і вторинної інформації: *первинна інформація* – отримана дослідником необроблена інформація безпосередньо від об'єкта дослідження; *вторинна інформація* – отримана дослідником аналітична інформація з робіт інших авторів або опитувань. Отже, відбувається розроблення програмно-методологічних та організаційних аспектів статистичного спостереження за об'єктом дослідження.

Другий етап. *Формування інформаційної підтримки процесу оброблення статистичної інформації.* Він містить: проведення підбору кваліфікованих кадрів для збирання і подальшого оброблення статистичної інформації; розроблення методології збирання і оброблення статистичної інформації відповідно до статистичного дослідження, що проводиться; адаптацію наявних інформаційних технологій під завдання, що вирішуються у статистичному дослідженні; проведення поглибленого та всебічного аналізу статистичних даних за допомогою пакетів прикладних програм і формування вихідного масиву оброблених даних у вигляді баз даних;

систематизацію та формування потоків обробленої інформації для подальшого її використання.

Третій етап. *Оцінювання та інтерпретація результатів статистичного дослідження.* Потоки обробленої статистичної інформації порівнюються, перевіряються і оцінюються відповідно до поставленої мети та завдань статистичного дослідження. Кінцеві результати наводяться у вигляді аналітичних висновків, довідок чи звітів, у яких аналізуються основні тенденції розвитку об'єкта дослідження. Для наочності відображення отриманих результатів використовують засоби інформаційної візуалізації результатів статистичного аналізу. Поєднання цих видів інтерпретації інформації дозволяє приймати зважені управлінські рішення.

Отже, інформаційне забезпечення статистичного дослідження доцільно проводити поетапно. Це дозволить чітко сформулювати мету, об'єкт і предмет дослідження, обґрунтовано підійти до вибору методів, за допомогою яких буде проводитись аналіз, а отримані результати достовірно відобразатимуть основні тенденції розвитку соціально-економічних процесів.

Спираючись на теоретичну базу, статистика застосовує *специфічні методи*, які знаходять своє вираження у **трьох стадіях (етапах) статистичного дослідження:**

- статистичне спостереження, яке полягає у збиранні первинної інформації про окремі факти досліджуваного явища;
- групування та зведення зібраного матеріалу, що дозволяє провести його систематизацію та класифікацію. Це здійснюється на основі переходу від характеристики окремих елементів дослідження до узагальнюючих показників у формі абсолютних, відносних і середніх величин;
- оброблення статистичних показників, отриманих у результаті групування і зведення, аналіз отриманих результатів з метою формування висновків про стан і закономірності розвитку досліджуваного явища. На цьому етапі формуються науково обґрунтовані висновки, що спрямовані на розроблення заходів і формування та ухвалення ефективних управлінських рішень.

Стадії статистичного дослідження об'єднуються метою дослідження. На кожній з них застосовуються ті методи, які можуть дати глибоку та всебічну характеристику досліджуваних явищ. Так, на *першій стадії* використовується метод масового статистичного спостереження, який забезпечує всебічність, повноту та репрезентативність початкової інформації,

тобто дає інформаційну базу для статистичних узагальнень і характеристики об'єктивних закономірностей. Статистичні дані мають беззаперечну доказову силу саме тому, що вони спираються не на окремі факти, а на їх сукупність [9; 11; 82].

На *другій стадії* здійснюється узагальнення даних масового спостереження: елементи сукупності класифікують за певними ознаками за допомогою методів зведення, класифікації та статистичного групування. Наприклад, народжених можна класифікувати за статтю та місцем народження, видобуток вугілля – за шахтами або за датою. Упорядковану таким чином статистичну сукупність називають **статистичними рядами**. Залежно від способу класифікації розрізняють ряди розподілу та ряди динаміки. *Ряд розподілу* – це результат класифікації, групування елементів сукупності в статистиці (станом на певний момент або за певний проміжок часу). За допомогою групування виділяють характерні властивості і різноманітні типи явищ. *Ряд динаміки* відбиває значення статистичних показників у часі (за періодами або моментами часу), описує динаміку розвитку масового процесу.

На *третьій стадії* здійснюється аналіз й інтерпретація результатів статистичного дослідження: проводиться всебічний кількісно-якісний аналіз масових явищ та процесів на підставі методів вивчення варіації ознак, використання показників динаміки, індексного методу, встановлення тенденцій і прогнозування розвитку, виявлення взаємозв'язків і взаємозалежності між ознаками.

Вивчаючи різноманітні суспільні явища та процеси, статистичні методи пристосовуються до їх особливостей. Але в будь-якому дослідженні виявляються властиві статистичному методу особливості – масовість даних, кількісне вимірювання, узагальнення.

Аналітичні можливості статистичних методів поглиблюються завдяки використанню компактної і раціональної форми подання результатів узагальнення інформації, а також аналізу виявлених закономірностей. Такими формами є статистичні таблиці та графіки [1].

1.4. Поняття та категорії в статистиці

Статистика вивчає свій предмет за допомогою певних **категорій**, тобто понять, що відображають найбільш загальні й істотні властивості, ознаки, зв'язки та відношення предметів і явищ об'єктивного світу.

У статистиці таких основних понять – п'ять:

- статистична сукупність;
- статистична одиниця (одиниця статистичної сукупності);
- статистична ознака;
- статистична закономірність;
- статистичний показник.

Статистична сукупність – це множина елементів (одиниць сукупності), які мають спільні риси або ознаки. Кожний елемент сукупності характеризується певними значеннями ознаки, яка варіює. Це може бути сукупність мешканців міста Харкова, сукупність студентів ХНЕУ ім. С. Кузнеця тощо. Статистична сукупність є об'єктом статистичного вивчення.

Запропоноване визначення статистичної сукупності дозволяє виділити основні її **властивості**:

- **нерозкладність** – часткове виникнення або часткове зникнення елементів статистичної сукупності не руйнує її якісної основи, всі її якісні характеристики зберігаються. Студенти ХНЕУ ім. С. Кузнеця, як статистична сукупність, не змінять своєї якісної характеристики незалежно від того, що частина студентів (випускники) щорічно залишають заклад вищої освіти (ЗВО), і частина (першокурсники) вливається до їх складу. Ця властивість заснована на масовості статистичної сукупності;

- статистична сукупність *формується як множина об'єктів, явищ, об'єднаних загальною ознакою*. Усі елементи статистичної сукупності мають цю загальну властивість. Однак загальне не означає однакове. Значення загальної ознаки в різних одиницях сукупності, як правило, відрізняються одне від одного. Однорідність статистичної сукупності встановлюється в кожному конкретному статистичному дослідженні відповідно до його цілей і завдань;

- **варіація** – кількісна зміна значення статистичної ознаки з переходом від одного її елемента до іншого. Якщо значення ознаки в усіх елементів однакові, немає сенсу вивчати всю статистичну сукупність, достатньо розглянути лише один її елемент, щоб отримати знання про все явище. Варіація виникає під впливом певного комплексу умов і причин. Статистика не займається виявленням і з'ясуванням цих причин. Це прерогатива спеціальної економічної дисципліни, статистика кількісно оцінює вплив кожної причини на варіацію конкретної ознаки, що дозволяє враховувати вказаний вплив на ухвалення управлінських рішень різного рівня.

Усі статистичні сукупності можна розділити на **групи**:

- *створені самим життям*, вони створюють єдність незалежно від того, чи піддаються вони статистичному дослідженню (статистичні сукупності працівників підприємства, статистичні сукупності промислових підприємств Харкова тощо). Це реально існуючі статистичні сукупності, вони мають певний конкретний розмір;
- *створені спеціально для цілей статистичного дослідження* (сукупність покупців певного товару в маркетингових дослідженнях);
- *стохастичні сукупності* (гіпотетичні безлічі) – уявні, нереальні, передбачувані сукупності (сукупність небесних тіл у Галактиці; сукупність нескінченно великого числа кидання монети, що падає "орлом" або "решкою").

Статистична одиниця, або **одиниця статистичної сукупності**, – нерозкладний первинний незалежний елемент статистичної сукупності, який є носієм певної статистичної ознаки. Одиниця статистичної сукупності є межею дроблення сукупності, за якого зберігаються всі властивості досліджуваного явища або процесу, її окремим випадком. Вибір одиниці сукупності визначається метою та рівнем проваджуваного статистичного дослідження. Наприклад, можна вивчати продуктивність праці на рівні галузі, підприємства, цеху, дільниці, бригади. У кожному випадку одиниці сукупності будуть різними: підприємство галузі, працівник певного підприємства; робітник цеху, бригади. Одиниці сукупності обов'язково повинні бути якісно однорідними.

Загальну кількість одиниць сукупності називають **обсягом статистичної сукупності**.

Статистична ознака – характерна властивість, певна якісна характеристика одиниць статистичної сукупності. Наприклад, статистичними ознаками підприємств можуть бути: форма власності, чисельність працюючих, величина статутного капіталу, вартість активів та ін. Статистичні ознаки можна класифікувати за безліччю підстав. Класифікацію зручно подати у вигляді табл. 1.2.

За характером вираження розрізняють атрибутивні та кількісні ознаки:

- *атрибутивні (якісні, описові)* – виражаються у вигляді понять, найменувань, тобто словесно. Наприклад, стать, національність, освіта та ін. За ними можна отримати підсумкові відомості про кількість статистичних одиниць, що мають певні значення ознаки;

Класифікація статистичних ознак [68]

Підстави для класифікації				
Характер вираження	Характер варіації	Відношення до часу	Характер взаємозв'язку	Ступінь впливу
Атрибутивні (якісні) Кількісні	Альтернативні Дискретні Безперервні	Моментні Інтервальні	Факторні Результативні	Суттєві Несуттєві

- *кількісні* – виражаються числовою мірою (вік, стаж роботи, обсяг продажу, розмір доходу тощо). За ними можна отримати підсумкові дані про кількість одиниць, що володіють конкретним значенням ознаки, і сумарне або середнє значення ознаки за сукупністю.

За **характером варіації** ознаки розподіляють на:

- *альтернативні*, які можуть приймати тільки одне з двох можливих значень. Це ознаки володіння чи неволодіння чим-небудь. Наприклад, стать, сімейний стан; у маркетингових чи політологічних дослідженнях – відповідь на запитання у формі "так" чи "ні";

- *дискретні* – кількісні ознаки приймають тільки окремі значення, без проміжних між ними; як правило цілочисельні. Наприклад, розряд робітника, кількість дітей у сім'ї, кількість автомобілів тощо;

- *безперервні* – кількісні ознаки, які приймають будь-які значення в певних межах. Як правило, вони округлюються відповідно до прийнятої точності (наприклад: бухгалтерський прибуток за балансом у гривнях, податок з податкових реєстрів – у тис. грн).

За **відношенням до часу** розрізняють:

- *моментні* ознаки, які характеризують одиниці сукупності на критичний момент часу. Наприклад: вартість основних виробничих фондів (ОВФ) визначається на 01.01. і 31.12. відповідного року як вартість ОВФ на початок і кінець звітного року;

- *інтервальні* ознаки, які характеризують явище за певний часовий період (рік, квартал, місяць та ін.). Наприклад, денна виручка, річний обсяг продажів тощо [22].

За **характером взаємозв'язку** ознаки розподіляють на:

- *факторні*, що викликають зміни інших ознак або створюють можливості для змін значень інших ознак. Факторні ознаки підрозділяють, відповідно, на *ознаки причини* й *ознаки умови*;

- *результативні* (ознаки наслідків), що залежать від варіації інших ознак. Наприклад, вартісний обсяг випуску продукції є результативною ознакою, величина якої залежить від факторних ознак – чисельності працівників і продуктивності праці.

За **ступенем впливу** на досліджувану статистичну одиницю ознаки розрізняють на:

- *суттєві* (основні), тобто ознаки, які є головними для досліджуваного явища. Наприклад, для підприємства такими будуть кількість виробленої продукції, кількість працівників;

- *несуттєві*, тобто такі, що не пов'язані безпосередньо із сутністю досліджуваного явища. Наприклад, назва підприємства, його територіальна належність.

Важливою категорією статистики є статистична закономірність. Як філософська категорія будь-яка закономірність є формою прояву причинного зв'язку, що виражається в послідовності, регулярності, повторюваності подій із високим ступенем вірогідності, якщо причини, що породжують події, не змінюються або змінюються незначно [22; 29].

Статистична закономірність – одна з форм прояву закономірного зв'язку в масових явищах і процесах, що перебувають під впливом багатьох факторів, які постійно змінюються. Статистично закономірність виявляється на рівні статистичної сукупності та встановлюється за допомогою статистичних методів. Закономірність виникає як результат впливу великої кількості постійно та випадково діючих причин. *Постійно діючі причини* надають змінам в явищах регулярність і повторюваність, *випадкові* – викликають відхилення в цій регулярності. На рівні статистичних одиниць закономірність проявляється не завжди. Наприклад, відомо, що середня тривалість життя у жінок більше, ніж у чоловіків; але це не означає, що кожна жінка живе довше, ніж кожен чоловік (серед чоловіків зустрічається більше довгожителів) [42; 82].

Оскільки статистична закономірність виявляється в результаті масових статистичних даних, це обумовлює її взаємозв'язок із законом великих чисел – статистичні закономірності є наслідком дії цього закону.

Закон великих чисел у найпростішому формулюванні твердить, що в масових явищах і процесах випадкові другорядні ознаки в спостережуваних одиницях взаємопогашаються, в результаті чого виразно проявляються найбільш істотні ознаки, закономірності розвитку таких явищ [22; 29]. Отже, закон виражає діалектику випадкового та необхідного. Наприклад, на 100 дівчаток народжується 104 – 106 хлопчиків; проте в різних сім'ях і навіть в невеликих населених пунктах це співвідношення може бути абсолютно іншим. Відповідно до природи масової закономірності тенденції, розкриті за допомогою закону великих чисел, мають силу тільки як масові тенденції, але не як закони, що фіксують стійкий, загальний характер причинно-наслідкового зв'язку явищ. У статистичній закономірності ці зв'язки менш стійкі, не мають загального характеру, а стосуються певного простору та часу, справедливі лише для певних умов існування досліджуваного явища.

Термін "статистична закономірність" уперше почали застосовувати в природничих науках (на противагу поняттю "динамічна закономірність"), щоб виражати таку форму зв'язку, за якої конкретним значенням будь-яких факторів завжди строго відповідають певні значення залежних від цих факторів ознак. У випадку динамічної закономірності кількісні співвідношення між величинами залишаються справедливими для кожної окремої статистичної одиниці. Наприклад, площа кола змінюється зі зміною її радіуса, і ця залежність виражається формулою $S = 2\pi r^2$, яка справедлива для будь-якого кола.

Для опису досліджуваних масових явищ і процесів, як інструмент їх пізнання, використовують статистичні показники. **Статистичним показником** називають узагальнену числову характеристику будь-якого масового явища (процесу) з його якісною визначеністю в конкретних умовах місця і часу. Цим він відрізняється від індивідуальних значень ознаки (варіант). Наприклад, середня заробітна плата працівників конкретного підприємства за певний період часу – статистичний показник, а заробітна плата конкретного працівника – це індивідуальне значення ознаки (варіанти). На відміну від індивідуального значення ознаки статистичний показник може бути отриманий тільки розрахунковим шляхом. Це можуть бути простий підрахунок одиниць сукупності, підсумовування їх значень ознаки або більш складні розрахунки. Наприклад, кількість працівників на підприємстві на початок року, обсяг виробленої продукції за квартал, собівартість, рентабельність виробництва тощо.

Відповідно до визначення статистичний показник має якісний і кількісний аспекти. Якісний бік статистичного показника визначається ознакою, що підлягає вивченню і відображається в назві показника, кількісний закладений у числовому значенні показника.

На практиці для відображення різноманітних аспектів соціально-економічних явищ і процесів використовують статистичні показники, які можна класифікувати певним чином (табл. 1.3).

Таблиця 1.3

Класифікація статистичних показників [68]

Підстави для класифікації		
Виконувана функція	Охоплювані одиниці сукупності	Форми вираження
Планові Облікові Прогностичні	Індивідуальні Зведені	Абсолютні Відносні Середні

Планові показники – відображають директивну функцію, орієнтовані на виконання поставлених завдань. *Облікові* – показують реальний стан досліджуваного явища. *Прогностичні* – його можливий стан у майбутньому.

Якщо статистичний показник стосується конкретного явища (наприклад, окремого підприємства), то його називають *індивідуальним*. Якщо ж показник характеризує сукупності явищ (наприклад, однорідних підприємств регіону), то його називають *узагальнювальним* або *зведеним*. У статистиці використовують також систему статистичних показників – сукупність взаємопов'язаних і розташованих у логічній послідовності узагальнених даних.

Абсолютні – вихідна первинна форма вираження статистичних показників. *Відносні* – похідні, вторинні показники відносно абсолютних, що виражають певні співвідношення між кількісними характеристиками статистичних сукупностей. *Середні* – найбільш поширена форма статистичних показників, що характеризують типовий рівень явища. Середні показники розраховують на одиницю статистичної сукупності.

1.5. Організація та завдання статистики в сучасних умовах

Початок процесу організації міжнародної статистики було покладено групою європейських статистиків, які в 1853 р. запропонували створити Міжнародний статистичний конгрес з метою вироблення на науковій основі єдиних методів, однакових правил і програм спільних робіт. Для організації скликання Конгресу багато зробили видатні статистики того часу А. Кетле і Е. Енгель. У розвиток цього міжнародного форуму було вирішено заснувати **Міжнародний статистичний інститут** – МСІ (*International Statistical Institute – ISI*), котрий було створено в 1885 р. Відповідно до Статуту основним завданням МСІ є розвиток і вдосконалення статистичних методів у різних країнах світу. Фактично МСІ проводить курс на досягнення міжнародної порівнянності статистичних показників шляхом вироблення однакових методів їх обчислення і розроблення загальноприйнятих класифікацій таких показників у різних галузях статистики з метою обміну думками між науковими та практичними працівниками статистики та їхніми організаціями [68].

МСІ є автономною міжнародною організацією, проте має консультативний статус при ЕКОСОП ООН і ЮНЕСКО. Він не входить у систему спеціалізованих установ ООН, але регулярно бере участь у роботі Статистичної комісії ООН, не будучи її членом. Практична діяльність МСІ координується його виконавчим органом – Постійним бюро, розташованим у м. Воорбург (Нідерланди).

У сучасному механізмі міжнародної статистики простежується свого роду розподіл праці: одні статистичні організації розробляють статистичну методологію, програми спостереження; інші безпосередньо здійснюють збирання, оброблення та публікацію статистичних даних. До першого типу організацій належать ті, що проводили раніше міжнародні статистичні конгреси (МСК) і нині діючий Міжнародний статистичний інститут. До другого типу – статистичний апарат ООН і його спеціалізовані установи та статистичні служби інших численних міжнародних організацій.

Основна практична діяльність в сфері міжнародної статистики здійснюється спеціалізованими службами ООН. Найбільшу за обсягом роботу проводить Статистична комісія ООН і Статистичне бюро Секретаріату ООН, які виконують функції секретаріату Комісії. Статистична комісія ООН (створена в 1946 р.) є однією з так званих функціональних комісій Економічної

та соціальної ради ООН (ЕКОСОР), що скеровує і контролює діяльність усіх підрозділів ООН у галузі статистики.

Сьогодні Статистична комісія здійснює свою діяльність за такими напрямами:

- аналіз досвіду окремих країн у розробленні статистичної методології;
- поліпшення порівнянності даних і розроблення міжнародних стандартів;
- застосування в міжнародній статистиці ЕОМ і розвиток інтегрованих систем збирання і оброблення міжнародної інформації;
- розроблення заходів з надання допомоги країнам, що розвиваються, у розвитку національної статистики.

Кінцевою метою роботи Статистичної комісії є досягнення єдиної системи збирання, оброблення та розповсюдження міжнародної статистичної інформації органами й установами ООН.

Функціями Статистичного бюро Секретаріату ООН є:

- підготовка всіх заходів для сесій Статистичної комісії;
- підготовка й остаточне оброблення методологічних доповідей та висновків від країн і різноманітних організацій зі статистичних питань; збирання, аналіз, публікування і оновлення статистичних відомостей, отримуваних від держав-членів і спеціалізованих установ ООН;
- поліпшення порівнянності даних шляхом організації додаткових досліджень і розрахунків; координація статистичної діяльності ООН, спеціалізованих установ, національних статистичних і міжнародних економічних організацій;
- сприяння розвитку та поліпшення статистики у цілому та насамперед у країнах, що розвиваються; підготовка необхідних матеріалів та обслуговування сесій Статистичної комісії.

Статистичні органи ряду міжнародних організацій випускають *статистичні збірники*. Значну роботу було проведено з побудови системи основних взаємопов'язаних показників народно-господарського процесу в цілому – системи національних рахунків і щодо зіставлення її елементів з балансом народного господарства окремих країн. У рекомендаціях Статистичної комісії ООН за стандартами національних рахунків помітний вплив передового національного досвіду в галузі балансових розробок. Про це свідчать, зокрема, спроби об'єднання національних рахунків

і фінансових звітів, пропозиції про включення до системи національних рахунків показників національного багатства та трудових ресурсів, деяка зміна методології обчислення національного доходу та ін.

Найважливішими статистичними збірниками ООН та її спеціалізованих установ з 1946 р. є: *"Статистичний щорічник"* ("Statistical yearbook"), *"Демографічний щорічник"* ("Demographic yearbook"), *"Щорічник зі статистики міжнародної торгівлі"* ("Yearbook of international trade statistics"), *"Щорічник ООН"* ("Yearbook of the United Nations") та ін. Продовольча та сільськогосподарська організація ООН (ФАО) видає *"Щорічник зі статистики продовольства та сільського господарства"* ("Yearbook of food and agricultural statistics"), а також щорічники зі статистики рибальства та лісового господарства; Організація Об'єднаних Націй із питань освіти, науки та культури (ЮНЕСКО) видає *"Міжнародний щорічник з освіти"* ("International yearbook of education") і загальний Статистичний щорічник ("Statistical yearbook") [68].

Зусиллями багатьох країн створено *Глобальну статистичну систему*, до якої входять:

- Статистична комісія ООН;
- галузеві статистичні підрозділи ООН;
- система статистичних видань ООН та інших міжнародних організацій;
- спеціальні установи ООН: ФАО – Комісія ООН з продовольства; ЮНЕСКО – Комісія ООН зі співробітництва в галузі науки, культури та освіти; ВООЗ – Всесвітня організація охорони здоров'я; СБ – Світовий Банк, МВФ – Міжнародний валютний фонд; СОТ – Світова організація торгівлі; МОТ – Міжнародна організація праці;
- регіональні статистичні організації, наприклад, Статистичне управління європейських співтовариств (Євростат), Статистичний комітет Країн Союзу Незалежних Держав (СНД).

Між статистичними органами немає суворої підпорядкованості, проте Статистична комісія ООН має офіційний статус "першої серед рівних". Вона координує генеральний перелік міжнародних стандартів і класифікацій і несе відповідальність за їх передання ряду держав.

Основна мета створення Глобальної статистичної системи полягає в ефективному використанні ресурсів для здійснення статистичної діяльності на національному та міжнародному рівні. У табл. 1.4 наведено адреси **веб-сайтів** міжнародних статистичних організацій.

Адреси веб-сайтів міжнародних статистичних організацій

Найменування організації	Адреси веб-сайтів
Статистичний офіс Європейського Союзу (Eurostat))	https://ec.europa.eu/eurostat
Статистичний комітет СНД	http://www.cisstat.com
Міжнародна організація праці (ILO)	https://www.ilo.org
Статистичний портал Організації економічного співробітництва та розвитку (OECD)	http://www.oecd.org/statistics
База даних Продовольчої організації при ООН (FAOSTAT)	http://www.fao.org/statistics
Міжнародний валютний фонд (IMF)	http://www.imf.org
Інститут статистики UNESCO	http://data.uis.unesco.org
Економічна комісія для Латинської Америки і Карибського регіону при ООН (UN ECLAC)	https://www.cepal.org
Статистичний сайт ООН	http://unstats.un.org/unsd
Організація Об'єднаних Націй	http://www.un.org
Економічна комісія для Азіатсько-Тихоокеанського регіону Америки при ООН (UN ESCAP)	http://www.unescap.org/stat
Міжнародна промислова статистика (UNIDO)	http://www.unido.org
Статистичні дані Всесвітньої організації охорони здоров'я (WHO)	http://www3.who.int/whosis/menu.cfm
Світовий Банк (World Bank)	http://www.worldbank.org
Світова організація торгівлі (WTO)	http://www.wto.org

Система статистичних органів України відповідає державному устрою і адміністративно-територіальному розподілу країни.

Організація статистичної роботи в різних галузях здійснюється Державною службою статистики України (Держстат), очолюваною Головою, який призначається на посаду та звільняється з посади Кабінетом Міністрів України за поданням Прем'єр-міністра України, внесеним на підставі пропозицій Міністра економічного розвитку і торгівлі.

Діяльність Служби регламентується Положенням про Державну службу статистики України, затвердженим постановою Кабінету Міністрів України. Для розгляду наукових рекомендацій та проведення фахових консультацій з основних питань діяльності в Держстаті можуть утворюватися інші

постійні або тимчасові консультативні, дорадчі та інші допоміжні органи. Рішення про утворення чи ліквідацію колегії, інших постійних або тимчасових консультативних, дорадчих та інших допоміжних органів, їх кількісний та персональний склад, положення про них затверджує Голова Держстату [53].

Структурні підрозділи Державної служби статистики України:

Департамент національних рахунків та макроекономічної статистики;

Департамент статистики виробництва;

Департамент структурної статистики та статистики фінансів підприємств;

Департамент статистики послуг;

Департамент статистики сільського господарства та навколишнього середовища;

Департамент обстежень домогосподарств;

Департамент статистики торгівлі;

Департамент статистики праці;

Департамент статистики населення та регіональної статистики;

Департамент статистики цін;

Департамент інформаційних технологій;

Департамент планування та координації статистичної діяльності;

Департамент статистичної інфраструктури;

Департамент поширення інформації та комунікацій;

Департамент фінансово-економічного забезпечення;

Департамент управління персоналом та організаційного забезпечення;

Управління міжнародного співробітництва та європейської інтеграції;

Управління бухгалтерського обліку та звітності;

Управління справами;

Відділ правового забезпечення;

Сектор режимно-секретної роботи;

Сектор внутрішнього аудиту;

Головний спеціаліст із питань запобігання та виявлення корупції.

Держстат України проводить свою роботу через відповідні територіальні органи – Головні управління статистики в кожній області. Низовими організаціями державної статистики, які перебувають у безпосередньому підпорядкуванні статистичним управлінням, є районні та міські відділи статистики.

Ще одним із найважливіших інститутів у формуванні офіційної статистики країни є *Національний банк України*, який забезпечує збирання, складання та поширення даних фінансового та зовнішнього секторів економіки. Ці дані стосуються: грошово-кредитної та фінансової статистики (огляди фінансових корпорацій, депозити, кредити, процентні ставки депозитних корпорацій, цінні папери, фінансові рахунки сектора фінансових корпорацій); статистики платіжного балансу, міжнародної інвестиційної позиції та зовнішнього боргу; міжнародних резервів і курсів обміну валют; статистики індикаторів фінансової стійкості [55].

Національна академія статистики, обліку та аудиту та Державне підприємство "Інформаційно-аналітичне агентство" (Держаналітінформ) є функціональними органами державної статистики. Метою діяльності "Держаналітінформ" є створення умов для забезпечення потреб органів державної влади, органів місцевого самоврядування, інших юридичних, а також фізичних осіб в Україні та за її межами у статистичній інформації та доступу до неї, надання інформаційно-аналітичних і експертних послуг з економічних проблем на платній основі.

Основним завданням усіх органів статистики є збирання, перевірка достовірності, розроблення, узагальнення, аналіз та подання в установлені терміни законодавчим, виконавчим, управлінським та господарським органам науково обґрунтованих статистичних даних. Така інформація характеризує економічний і соціальний розвиток країни, виконання загальнодержавних і регіональних програм, зростання ефективності суспільного виробництва, економічні реформи, використання природних, трудових і матеріальних ресурсів, динаміку життєвого рівня населення тощо.

Засади організації статистики відповідно до Конституції України визначаються виключно законами України. Закон України "Про державну статистику" [59] регулює правові відносини в галузі державної статистики, визначає права та функції органів державної статистики, організаційні засади здійснення державної статистичної діяльності з метою отримання всебічної та об'єктивної статистичної інформації щодо економічної, соціальної, демографічної та екологічної ситуації в Україні та її регіонах і забезпечення нею держави та суспільства. Остання редакція цього Закону відбулась 19.04.2014 р. у зв'язку з ухваленням законів України "Про інформацію" та "Про доступ до публічної інформації" [58].

Завдяки створеній правовій базі з'явилися можливості наблизити статистику до міжнародних стандартів і світової практики. Найбільш істотними

кроками в цьому напрямі стали: впровадження системи національних рахунків (СНР); показників рівня життя населення, зовнішньоекономічної діяльності, функціонування ринку праці, нових структур підприємницької діяльності; розрахунок індексу споживчих цін (інфляція).

Перший етап реформування, який було в основному завершено у 1997 р., відповідно до Державної програми переходу України на міжнародну систему обліку та статистики характеризувався концентрацією ресурсів у найважливіших галузях статистики або у таких, що не відповідали міжнародній практиці. Протягом 1993 – 1997 рр. було створено основні елементи системи національних рахунків і платіжного балансу, статистику зовнішньої торгівлі, грошово-кредитну та банківську статистику, нову статистику ринку праці. Відповідно до міжнародних стандартів запроваджено розрахунок індексів інфляції, значно вдосконалено галузеву статистику, започатковано моніторинг фінансового стану підприємств та організацій (у тому числі таких нових формувань, як довірчі товариства, страхові організації, позабюджетні фонди тощо). Організовано статистичне обстеження фондового ринку, створено Єдиний державний реєстр підприємств та організацій України (ЄДРПОУ), розроблено основні положення Державної системи класифікації і кодування техніко-економічної та соціальної інформації, розпочато роботу зі створення банку статистичних даних і комп'ютерної мережі органів державної статистики.

Сьогодні національна статистика України здійснює черговий етап своїх реформ. 27 лютого 2019 р. Кабінет Міністрів України ухвалив розроблену Мінекономрозвитку разом із Державною службою статистики Програму розвитку державної статистики до 2023 р. [53; 54]. Програма визначає стратегічні напрями та завдання розвитку державної статистики, а також очікувані результати. Вона передбачає модернізацію державної статистики через імплементацію міжнародних стандартів і регламентів ЄС та повний перехід до процесно-орієнтованої системи виробництва статистичної інформації (GSBPM), а також використання інформаційних технологій (спеціального програмного забезпечення для подання та оброблення статистичної звітності, конструкторів електронних форм звітності).

Іншим важливим напрямом реформування державної статистики є забезпечення відкритості та доступності статистичної інформації. Зокрема, це означає створення сучасного, більш зручного для користувача (user-friendly), веб-порталу Держстату та його мобільної версії, забезпечення широкого доступу користувачів до мікроданих (щодо зайнятості

населення, стану домогосподарств тощо), активізацію комунікацій щодо результатів діяльності Держстату та даних його досліджень, а також у цілому підвищення рівня статистичної грамотності суспільства через відповідні навчальні програми в закладах середньої та вищої освіти, інформування в ЗМІ та ін.

Програма передбачає зменшення звітного навантаження на респондентів, що надають статистичні дані для подальшого аналізу. Досягти цього можна через збільшення кількості несуцільних статистичних спостережень (вибіркових) і запровадження методів моделювання, використання сучасних методів і технологій збирання даних, розширення використання адміністративних даних (у тому числі на регіональному рівні), запровадження безкоштовного звітування респондентами до органів державної статистики в електронному вигляді.

Серед важливих нововведень Програми розвитку статистики – запуск двох нових вибіркових обстежень домогосподарств: "Статистика доходів і умов життя" (європейський аналог – EU-SILC) та обстеження бюджетів домогосподарств (HBS), а також модернізація показників підрахунку національних рахунків, у тому числі показника ВВП за паритетом купівельної спроможності (ПКС). Це забезпечить гармонізацію і уможливить порівняння фізичних обсягів валового внутрішнього продукту за категоріями кінцевого використання країн ЄС і кандидатів на членство в ЄС.

Важливим компонентом реалізації Програми є розвиток людського капіталу та підвищення рівня професійної компетенції співробітників органів державної статистики. Зокрема, передбачено гармонізацію освітніх програм підготовки фахівців у галузі статистики з європейськими стандартами, зокрема з Магістерською програмою з офіційної статистики ЄС (EMOS), удосконалення системи підвищення кваліфікації персоналу та вивчення мов країн Ради Європи [53].

Завдання статистики на сучасному етапі розвитку національної економіки нерозривно пов'язані з трансформаційними процесами в державі. В якості основних завдань статистики в Україні виділяють:

усебічне дослідження глибоких перетворень соціальних та економічних процесів, що виникають у суспільстві, на основі науково обґрунтованих показників;

узагальнення і прогнозування тенденцій розвитку держави;

виявлення резервів ефективності суспільного виробництва;

своєчасне забезпечення управлінських органів статистичною інформацією;

розроблення заходів щодо зближення вітчизняної методології статистичних досліджень із методологією і стандартами міжнародної статистики;

формування і функціонування системи моніторингу (спеціально організованих спостережень);

розроблення і впровадження загальнодержавної автоматизованої системи збирання, нагромадження, оброблення й аналізу статистичної інформації;

удосконалення статистичної методології та системи статистичних показників, збирання і оброблення статистичної інформації, економіко-статистичного аналізу соціально-економічних явищ і процесів.

Важливі поняття

Методологічна основа статистики – діалектичний метод пізнання та специфічні методи статистики.

Об'єкт вивчення статистики – масові явища та процеси будь-якої природи.

Предмет статистики – розміри та кількісні співвідношення масових суспільних явищ у нерозривному зв'язку з їх якісним аспектом із метою виявлення закономірностей їх розвитку.

Статистика – суспільна наука, яка розробляє методологію статистичного дослідження, що використовується іншими науками.

Статистична закономірність – одна з форм прояву закономірного зв'язку в масових явищах і процесах, що перебувають під впливом багатьох факторів, які постійно змінюються.

Статистична одиниця, або одиниця статистичної сукупності, – нерозкладний первинний незалежний елемент статистичної сукупності, який є носієм певної статистичної ознаки.

Статистична ознака – характерна властивість, певна якісна характеристика одиниць статистичної сукупності.

Статистична сукупність – множина елементів (одиниць сукупності), які мають спільні риси або ознаки. Кожний елемент сукупності характеризується певними значеннями ознаки, яка варіює.

Статистичний показник – узагальнена числова характеристика будь-якого масового явища (процесу) з його якісною визначеністю в конкретних умовах місця і часу.

Теоретична основа статистики – положення економічної теорії та політекономії.

Типові завдання

Питання для обговорення на семінарі

1. Охарактеризуйте основні етапи розвитку статистики та поясніть, на основі чого відбувся перехід від обліку до статистики як науки.
2. Сформууйте перелік показників, якими можна охарактеризувати: населення країни, споживчий ринок, галузь промисловості, банківську установу.
3. Розкрийте сутність етапів статистичного дослідження.
4. Охарактеризуйте тематичну спрямованість інтернет-ресурсів міжнародної статистики.

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Предмет, методи та завдання статистики" подані у роботі [67]. Лабораторна робота націлена на формування теоретичних знань і розуміння методологічних засад статистики шляхом розкриття питань семінарського завдання.

Завдання для самодіагностики

1. Яке походження має термін "статистика", ким і коли він був введений у науковий обіг?
2. Охарактеризуйте етапи розвитку статистики.
3. Що розуміють під статистикою у сучасних умовах?
4. Сформулюйте об'єкт і предмет вивчення статистики.
5. Назвіть основні категорії статистики.
6. Дайте визначення статистичної сукупності й одиниці сукупності. Наведіть приклади.
7. Опишіть стадії (етапи) статистичного дослідження.

8. За якими підставами можуть бути класифіковані статистичні ознаки?

9. Наведіть приклади моментних та інтервальних ознак.

10. Які існують класи статистичних показників?

11. Опишіть систему державних органів статистики України та назвіть завдання статистики на сучасному етапі розвитку.

Питання для критичного переосмислення (есе)

1. Укажіть, які сукупності можна виділити на підприємстві або в організації для статистичного вивчення. Назвіть одиницю сукупності. Відповідь обґрунтуйте.

2. Укажіть, якими кількісними й атрибутивними ознаками можна охарактеризувати сукупність студентів ЗВО, сукупність страхових організацій, сукупність робітників високої кваліфікації, сукупність країн – імпортерів палива.

3. Укажіть основні ознаки, що визначають варіацію: успішності студентів, прибутку підприємств харчової промисловості, заробітної плати робітників різних галузей економіки.

4. До якого виду (кількісні або якісні, дискретні або неперервні) належать такі ознаки:

а) кількість робітників фірми;

б) родинні зв'язки членів сім'ї;

в) стать і вік людини;

г) соціальний статус вкладників банку;

д) кількість дітей у сім'ї;

е) роздрібний товарообіг крамниць?

Відповідь обґрунтуйте.

2. Статистичне спостереження

Основні питання:

2.1. Сутність і завдання статистичного спостереження.

2.2. Форми, види та способи статистичного спостереження.

2.3. План статистичного спостереження.

2.4. Помилки статистичного спостереження та забезпечення вірогідності даних.

2.1. Сутність і завдання статистичного спостереження

Для дослідження соціально-економічних явищ і процесів необхідно зібрати первинні статистичні дані (інформацію), під якими розуміють сукупність характеристик масових явищ і процесів, отриманих у результаті статистичного спостереження. Ці дані є вихідним матеріалом для формування узагальнювальних показників і висновків про тенденції їх розвитку.

Зібрана інформація про явище, що вивчається не завжди вважається статистичною. *Статистична інформація* повинна відповідати певним **вимогам** [13; 31; 68]:

повнота:

охоплювати всі одиниці статистичної сукупності або таку їх частину, котра дає змогу робити висновки про сукупність у цілому;

охоплювати всі істотні аспекти явища, його властивості, внутрішні та зовнішні зв'язки;

збиратися за максимально тривалий термін, що сприятиме ослабленню впливу випадкових факторів і виявленню закономірностей розвитку явища;

достовірність – зібрані в процесі спостереження дані повинні відповідати фактичному стану природи явища;

зіставність – з метою забезпечення подальшого зіставлення даних їх слід збирати в установленій термін, за єдиною програмою, з використанням однакових методів;

своєчасність подання (особливо для ухвалення управлінських рішень);

доступність, що особливо є актуальним в умовах ринкової економіки, коли низький рівень відповідальності недержавних структур щодо подання звітів призводить до порушення вимог повноти та своєчасності реєстрації даних.

Основними **властивостями** статистичної інформації є її **масовість** і **стабільність**. Перша властивість пов'язана з особливостями предмета статистики, друга – з незмінністю зібраної інформації, її здатністю застарівати та з необхідністю отримання нової інформації для формування раціональних управлінських рішень. Практичний менеджмент потребує постійного оновлення статистичних даних. Інформація, що є достовірною, повною, але несвоєчасною, стає практично не потрібною для ухвалення рішень на будь-якому рівні управління.

Отримання якісної статистичної інформації значно залежить від того, на якому рівні вона збирається. В Україні діють дві **системи збирання інформації**:

- 1) *централізована* (загальнодержавна);
- 2) *децентралізована* (відомча, окремих економічних структур).

Централізована система збирання має ширші можливості для якісного спостереження, наукову методику, кваліфіковані кадри, технічне забезпечення тощо. Проте децентралізована система є більш оперативною завдяки меншому проміжку часу між збиранням даних і використанням готової статистичної інформації. Для децентралізованої системи проблема наукової обґрунтованості методів статистичного спостереження та їх практичного застосування є більш гострою.

Склад статистичної інформації визначається потребами розвитку суспільства. У сучасних економічних умовах господарювання її споживачами є як державні органи, так і різні недержавні структури. Основним джерелом статистичної інформації є видання органів державної статистики.

Надання статистичної інформації є основним завданням органів державної статистики, а сама статистична інформація – продукцією їх діяльності, що має, як і будь-яка інша, свою вартість. Найбільшою вартістю наділена та інформація, отримання якої виходить за межі програми роботи органів державної статистики.

Основним джерелом отримання статистичної інформації і першим етапом статистичного дослідження є статистичне спостереження.

Статистичне спостереження – це планомірне, науково організоване збирання даних або відомостей про масові явища та процеси, яке полягає в реєстрації відібраних ознак для кожної одиниці сукупності [68; 74; 88].

До статистичних відносять ті спостереження, що вивчають статистичні закономірності, які проявляються в масових процесах, у великій кількості одиниць сукупності шляхом реєстрації (запису) відповідних ознак явищ (процесів). Отже, спостереження вважається статистичним, якщо воно забезпечує реєстрацію встановлених фактів у облікових документах для подальшого узагальнення і відповідає таким умовам, як [45; 68; 75]:

- планомірність;
- масовість;
- систематичність.

Планомірність статистичного спостереження припускає, що воно проводиться за заздалегідь розробленим планом. У такий план включають

питання методології, організації, техніки збирання інформації, контролю її якості та достовірності, оформлення підсумкових результатів.

Масовий характер статистичного спостереження означає, що воно охоплює таку кількість випадків прояву досліджуваного явища, котра є достатньою для отримання достовірних статистичних даних, які характеризують сукупність у цілому.

Систематичність спостереження передбачає регулярність його проведення. Це дає змогу вивчати тенденції та закономірності соціально-економічних явищ і процесів.

Залежно від **рівня реєстрації** статистичні спостереження можуть бути первинними або вторинними.

Первинне спостереження – це реєстрація вихідних даних, що надходять від об'єкта, який їх продукує (поточний облік кількості зареєстрованих шлюбів і розлучень у відповідній установі; опитування населення щодо процесу приватизації майна).

Вторинне спостереження – це збирання раніше зареєстрованих та оброблених даних (матеріали банківських звітів, результатів аудиторської перевірки, підсумків біржових торгів).

Процес проведення статистичного спостереження складається з таких **етапів**:

I етап – підготовка спостереження;

II етап – проведення масового збирання даних;

III етап – підготовка даних спостереження до оброблення;

IV етап – розроблення пропозицій щодо вдосконалення проведення статистичного спостереження.

На *першому етапі* вирішуються програмно-методологічні й організаційні питання плану статистичного спостереження, питання організації проведення спостереження.

Другий етап пов'язаний із безпосереднім проведенням спостереження і включає такі роботи, як: розсилка бланків, анкет, форм статистичної звітності, переписних листів; їх заповнення і подання до органів, які проводять спостереження.

На *третьому етапі* зібрана інформація перевіряється на повноту, піддається арифметичному та логічному контролю з метою виявлення і виключення помилок.

На *четвертому етапі* аналізуються причини виникнення помилок у заповненні статистичних формулярів, розробляються пропозиції щодо вдосконалення проведення статистичного спостереження.

2.2. Форми, види та способи статистичного спостереження

Статистичне спостереження як метод дослідження відрізняється від інших методів збирання інформації характером і масовістю даних і способами їх отримання.

Подамо інформацію щодо сутності процесу спостереження в табличному вигляді (табл. 2.1) [44; 64; 68].

Таблиця 2.1

Форми, види та способи статистичного спостереження

Форми статистичного спостереження	Види статистичного спостереження		Способи статистичного спостереження
	за часом реєстрації даних	за ступенем охопленням одиниць	
1. Статистична звітність. 2. Спеціально організоване спостереження: а) переписи; б) обліки; в) спеціальні обстеження; г) опитування. 3. Реєстр	1. Поточне (безперервне). 2. Перервне: а) періодичне; б) одноразове	1. Суцільне. 2. Несуцільне: а) вибіркове; б) обстеження основного масиву; в) монографічне; г) анкетне; д) моніторинг	1. Безпосередній облік. 2. Документальний облік. 3. Опитування: а) експедиційний спосіб; б) самореєстрація; в) кореспондентський спосіб; г) анкетний спосіб

Розрізняють три організаційні **форми** здійснення статистичного спостереження: статистична звітність, спеціально організоване спостереження та реєстр [4; 17; 68; 74; 75].

Статистична звітність – основна форма статистичного спостереження, за допомогою якої статистичні органи у визначені терміни отримують необхідні дані у вигляді встановлених у законному порядку звітних документів.

Звітність характеризується такими *властивостями*, як обов'язковість, систематичність, достовірність.

Обов'язковість – звітність неодмінно подають усі зареєстровані суб'єкти діяльності, дотримуючись уніфікованої форми та затвердженого переліку показників і вказуючи свої реквізити: назву, адресу, прізвище та підпис відповідальної особи, дату складання звітності.

Систематичність передбачає регулярне, своєчасне складання та подання звітності в затверджені терміни.

Достовірність – дані, наведені у звітності, повинні відповідати дійсності та виключати будь-які викривлення (приховування і приписки). За достовірність наведених даних суб'єкти діяльності несуть адміністративну та судову відповідальність. Завдяки цьому вдається попереджати неподання даних або подання перекручених фактів.

Звітність складається на підставі первинних даних оперативного та бухгалтерського обліку.

Класифікаційні ознаки звітності:

за рівнем затвердження і призначення:

зовнішня – затверджують і збирають органи державної статистики, міністерства та відомства;

внутрішня – розробляють самі суб'єкти діяльності для власних оперативних, управлінських і аналітичних цілей. Наприклад: аналіз ринкової ситуації, визначення власних ресурсів, прогнозування діяльності. Ґрунтується на даних внутрішньої звітності багатоструктурних суб'єктів діяльності, акціонерних товариств, концернів, асоціацій;

за частотою подання:

періодична – місячна, квартальна: охоплює показники поточної діяльності суб'єктів;

річна – підбиває основні підсумки фінансово-виробничої діяльності суб'єктів господарювання за рік;

за терміновістю надання:

поштова – передається поштою;

електронна – передається електронною поштою.

Спеціально організоване статистичне спостереження організується і проводиться з метою отримання відомостей, які відсутні в звітності, або для перевірки її даних. Наприклад: Всеукраїнський перепис населення потребує спеціально підготовлених людей, спеціально розробленої програми, є трудомістким і затратним.

Спеціально організоване спостереження може проводитись у вигляді переписів, обліків, спеціальних обстежень, опитувань.

Перепис – це суцільне або вибіркоче спостереження масових явищ із метою визначення їхнього розміру та складу на певну дату. Перепис здійснюється періодично (як правило, з рівним інтервалом) або одноразово. Так, перепис населення в більшості країн світу відбуваються раз на десять років. Особливістю перепису є те, що він проводиться одночасно на всій території за єдиною для всіх одиниць програмою. Так, у Програмі останнього перепису населення України (5 грудня 2001 року) були включені питання, що стосуються окремої людини (вік, стать, національність, рідна мова, шлюбний стан, освіта, громадянство, джерела засобів існування), а також домогосподарства в цілому (склад домогосподарства, зайнятість його членів, житлові умови тощо).

Облік – це суцільне спостереження масових явищ, що ґрунтується на даних обстеження, опитування і документальних записах. Наприклад, облік земельного фонду за видами угідь, якості ґрунту, категоріями господарств.

Спеціальні обстеження – це несцільні спостереження окремих масових явищ відповідно до певної тематики, яка виходить за межі звітності. Такі обстеження можуть бути періодичними або одноразовими. Наприклад: обстеження з питань неформальної зайнятості населення, бюджетні обстеження домогосподарств, маркетингові обстеження.

Опитування – це найчастіше несцільні спостереження думок, оцінок, мотивів певної частини респондентів, які реєструються з їх слів. Винятком виступає *референдум* – суцільне опитування всього населення з різноманітних соціально-економічних або політичних питань.

Реєстр – це форма безперервного статистичного спостереження за довготривалими процесами. Реєстр є системою, що постійно стежить за станом одиниці спостереження і оцінює силу впливу різних факторів на досліджувані показники. У реєстрі кожна одиниця спостереження характеризується сукупністю показників: тих, що залишаються незмінними протягом часу спостереження і реєструються один раз; інші показники, періодичність зміни яких невідома, оновлюється у разі зміни; треті – є динамічними рядами показників із заздалегідь відомим періодом оновлення.

До найбільш відомих реєстрів належать: *реєстри населення та реєстри підприємств та організацій*.

Реєстр населення – це поименований і регулярно оновлюваний перелік мешканців країни. Спостереження проводяться за такими ознаками: стать, дата та місце народження (постійні дані), шлюбний стан (змінна ознака). Інформація в реєстр заноситься на кожну народжену людину або яка прибула із-за кордону. У випадку смерті або виїзду на постійне місце проживання в іншу країну дані з реєстру вилучають.

Реєстр підприємств та організацій – це перелік суб'єктів всіх видів економічної діяльності із зазначенням їх реквізитів та основних показників. В Україні прийнято Єдиний державний реєстр підприємств і організацій (ЄДРПОУ). Цей Реєстр дозволяє налагодити єдиний інформаційний простір, до якого входять суб'єкти ринку.

Інформаційний фонд ЄДРПОУ містить:

реєстровий код суб'єкта;

відомості про галузеву, територіальну належність суб'єкта, його підпорядкованість;

види власності;

організаційні форми господарювання;

довідкові відомості (прізвище керівників, адресу, номери телефонів, факсів, відомості про засновників тощо);

економічні показники.

Види статистичного спостереження відрізняються за:

характером реєстрації даних у часі:

поточне – коли зміни фіксуються в міру їх надходження (реєстрація народження, шлюбу). Таке спостереження проводиться з метою вивчення динаміки будь-кого явища;

періодичне – спостереження проводяться за схожою програмою через певні проміжки часу (перепис населення, реєстрація цін на певні види товарів, облік успішності);

одноразове – відомості про явище, що вивчається, отримують один раз на момент спостереження в міру виникнення потреби (маркетингове дослідження місцевого ринку товарів);

повнотою охоплення одиниць сукупності:

суцільне – охоплює всі одиниці сукупності (звітність, перепис населення країни);

несуцільне – припускає, що обстеженню підлягає лише частина одиниць досліджуваної сукупності.

Несуцільне спостереження, в свою чергу, підрозділяють на:

вибіркове – випадковий відбір тих одиниць сукупності, які необхідно дослідити (опитування суб'єктів малого та середнього бізнесу);

метод основного масиву – обстеження найістотніших, найбільш великих одиниць сукупності, які є визначальними для характеристики об'єкта спостереження (фінансовий рейтинг найбільш впливових банків);

монографічне – детальному обстеженню піддаються окремі одиниці сукупності, зазвичай представники будь-яких нових типів явищ із метою їх досконального вивчення (обстеження діяльності МВФ);

анкетне – збирання відомостей за допомогою анкет (опитування учасників рекламної акції);

моніторинг – процес систематичного або безперервного збирання інформації про параметри складного об'єкта (діяльності) із метою визначення тенденцій зміни останнього (спостереження за метеоумовами, за біржовими цінами).

За **способами отримання інформації** розрізняють такі спостереження:

безпосереднє спостереження – інформацію збирають спеціально підготовлені особи, у завдання яких входить отримання відомостей шляхом особистого обліку одиниць сукупності (підрахунку, зважування, вимірювання значення ознаки). Це найбільш точний і надійний спосіб отримання даних, але і найбільш трудомісткий та затратний;

документальне спостереження (облік) використовується підприємствами, організаціями, установами для складання звітності на основі документів первинного обліку; є досить достовірним і надійним джерелом інформації;

опитування – спосіб спостереження, за якого відомості отримують зі слів респондента (опитуваного). Опитування використовується для формування інформації про процеси та явища, що піддаються безпосередньому прямому спостереженню. У статистичній практиці застосовують такі основні способи опитування:

усне (експедиційне) опитування – спеціально підготовлені реєстратори на основі опитування обстежуваної особи заповнюють переписні листи (фіксують факти). Реєстратор одночасно контролює достовірність отриманих відомостей. Усне опитування забезпечує досить точні результати, але є витратним способом отримання даних;

спосіб самореєстрації полягає в тому, що формуляри заповнюють самі опитувані особи (респонденти); реєстратори роздають опитувальні бланки, інструктують респондентів, збирають заповнені формуляри, контролюючи правильність заповнених відомостей;

кореспондентський спосіб полягає в тому, що формуляри заповнюють і надсилають опитувані особи на адресу організації, яка проводить дослідження, без участі реєстраторів на основі інструкцій щодо їх заповнення. Цей спосіб опитування вимагає найменших витрат, але не забезпечує високої якості отриманих даних, тому що не завжди є можливість перевірити точність отриманих відомостей;

анкетний спосіб передбачає збирання інформації у вигляді анкет. Анкети поширюються різними способами, заповнюються на добровільних засадах, як правило, анонімно. Кількість анкет, отриманих організацією, що проводить обстеження, завжди значно менша кількості розісланих анкет. Таке опитування дає приблизні, орієнтовні дані, які використовують зазвичай для вивчення громадської думки з різних питань.

2.3. План статистичного спостереження

Будь-яке статистичне спостереження проводиться за заздалегідь розробленим планом. **План статистичного спостереження** складається із двох груп питань: програмно-методологічних та організаційних [4; 13; 31; 68].

Розроблення програмно-методологічних питань плану спостереження полягає в науково-практичному обґрунтуванні та визначенні сутності явища, умов його формування та прояву. Крім того, підбирається система ознак – характеристик явища, враховується можливість їх кількісного оброблення та перевірки на точність.

Комплекс програмно-методологічних питань можна подати за послідовністю їх виникнення та розв'язання (рис. 2.1).

Програмно-методологічні питання плану статистичного спостереження **містять** [45; 46; 68; 75]:

- визначення мети та завдань спостереження;
- вибір об'єкта, одиниці спостереження та одиниці сукупності;
- розроблення програми спостереження;
- вибір форми, виду та способу проведення спостереження.



Рис. 2.1. Програмно-методологічні питання плану статистичного спостереження

Кожне статистичне спостереження проводиться з метою отримання достовірних даних про досліджувані процеси та явища. **Мета статистичного спостереження** має бути конкретною та чітко сформульованою, виходити із загальних завдань, поставлених перед статистичним дослідженням. Відповідно до принципів системного підходу завдання спостереження повинні відповідати поставленій меті, виходити з неї. Мета та завдання зумовлюють шляхи розроблення програми та вибір форм проведення спостереження, а конкретність їх формулювання забезпечує повноту отримання даних.

Залежно від мети та поставлених завдань визначають **об'єкт спостереження** – статистичну сукупність соціально-економічних явищ чи процесів, що підлягають обстеженню. Установлення об'єкта спостереження означає визначення точних меж і складу сукупності. Наприклад, під час перепису населення необхідно встановити, яке населення слід

реєструвати: наявне (населення, що фактично перебуває в момент перепису в обстежуваній місцевості) або постійне (населення, що проживає на цій території постійно, але на момент перепису може бути відсутнім).

Інколи для обмеження об'єкта спостереження використовують поняття цензу. **Ценз** – це ряд ознак, кількісне значення яких у проведенні статистичного спостереження виступає підставою для обліку (або невіднесення) одиниці до досліджуваної сукупності. Наприклад, якщо об'єктом спостереження є робота малих підприємств, то згідно з чинним законодавством до них відносять підприємства з кількістю не більше 50 працівників.

Якщо об'єкт спостереження буде визначений не досить чітко, то спостереження може охопити явища, які не належать до розглядуваної сукупності, і навпаки, деяка частина одиниць випаде з обстеження.

Будь-який об'єкт спостереження як статистична сукупність складається з одиниць сукупності. **Одиниця сукупності** – це елемент статистичної сукупності, який є носієм ознак, що підлягають реєстрації. Наприклад, у проведенні демографічних обстежень одиницею сукупності може бути як людина, так і сім'я, а в бюджетних обстеженнях – домогосподарство або сім'я. Кількість одиниць сукупності називають **обсягом сукупності**. Так, наприклад, у проведенні перепису населення *статистичною сукупністю* виступає все населення країни, а *одницею сукупності* – окрема людина; в торгівлі *статистичну сукупність* складають торговельні підприємства, а *одницю сукупності* – окреме торговельне підприємство; якщо треба дослідити якість води, то *сукупністю* виступають усі водні джерела, а *одницею сукупності* – окреме водне джерело.

Одиниці сукупності володіють безліччю різних ознак. *Статистична ознака* – це конкретна властивість, якість, відмінна риса одиниці сукупності. Наприклад, у працівника підприємства – це вік, стать, освіта, сімейний стан тощо. Дослідника цікавлять певні ознаки, тому в програмі спостереження наводиться перелік чи формулювання ознак, за якими необхідно зареєструвати дані.

Уточнення та формулювання ознак одиниці сукупності здійснюється за такими *правилами*:

ознаки відбираються з урахуванням цілей дослідження, можливостей їх оброблення й аналізу отриманих даних;

відібраних ознак не повинно бути багато;

ознаки необхідно комбінувати, щоб вони взаємно доповнювали одна одну;

відібрані ознаки повинні враховувати можливості дослідника.

Однак не кожна одиниця сукупності може надати про себе інформацію. Тому в ході обстеження виділяють *одиницю спостереження (звітну одиницю)*. **Одиниця спостереження** – це первинна одиниця, від якої отримують інформацію. Так, під час перепису населення одиницею спостереження є домогосподарство, а також кожен його член. Коли мова йде про реєстрацію проданих квартир на біржах нерухомості, то одиницею спостереження є біржа. Отже, *одиниця сукупності й одиниця спостереження можуть співпадати або не співпадати*.

Коли спостереження передбачає подання звітності, то *звітна одиниця* (одиниця спостереження) може співпадати з одиницею сукупності. У бюджетних обстеженнях господарств такими одиницями є домашні господарства; в ході вивчення громадської думки населення – кожен респондент, тобто людина, яка висловлює свою думку. Проте може не співпадати. Так, у ході вивчення платоспроможного попиту одиницею сукупності є покупець; коли дослідник звертається до нього з запитанням, то покупець виступає одиницею, що інформує (звітує), але за даними про продаж товару можна звернутися до менеджера магазину з продажу або до касира.

Відповідно до поставленої мети, визначених завдань, обраного об'єкта розробляють програму спостереження.

Програма спостереження – це перелік ознак, що підлягають реєстрації (у безпосередньому спостереженні) або це перелік запитань, за допомогою яких збирають відомості (в опитуваннях). Складання програми спостереження є складним і відповідальним процесом, що забезпечує якість зібраної інформації. Склад і зміст програми спостереження визначаються завданнями дослідження та особливостями суспільно-економічного явища, що досліджується. Будь-яке явище має багато ознак. Збирати інформацію щодо всіх – недоцільно та неможливо. Тому необхідно відібрати найбільш важливі ознаки, що відповідають поставленим завданням і відповідній меті спостереження.

До програми спостереження ставляться такі **вимоги** [17; 46; 68]:

включення тільки істотних ознак, що безпосередньо характеризують досліджуване явище, його тип, основні риси та властивості (другорядні питання, що не пов'язані з рішенням поставлених завдань, не повинні включатися до програми);

не включення питань, що можуть мати завідомо неточні відповіді;

включення питань контрольного характеру задля перевірки інформації, що збирається (логічно пов'язані питання про вік, сімейний стан, освіту, наявність дітей);

орієнтація всіх питань програми на певну форму відповіді: *цифрову, альтернативну, багатоваріантну*. За цифровою – відповідь дається в кількісній формі (про вік, стаж, заробіток); за *альтернативної* – у формі "так" чи "ні" (стать, сімейний стан); за *багатоваріантної* – вибір одного або декількох варіантів із запропонованого меню (на питання про стан у шлюбі можливі такі варіанти відповідей: перебуває у шлюбі; ніколи не перебував; вдівець; розлучений);

числове кодування відповіді для полегшення оброблення;

визначення складу та послідовності запитань (логіка розташування питань сприяє отриманню більш достовірних даних: рік народження та кількість повних років, що виповнилися опитуваному).

Одночасно з програмою спостереження розробляється *інструментарій* у вигляді *статистичних формулярів* та *інструкцій* щодо їх заповнення.

Статистичний інструментарій забезпечує не тільки вхідну, але і вихідну частину інформаційної бази спостереження. Тобто, визначаючи ознаки, складаючи блоки запитань, одночасно готують макети вихідних таблиць, де немає цифрової інформації. За макетами таблиць можна визначити, наскільки кожне запитання (ознака) узгоджується з іншими, передбачити, як "працює" те чи інше запитання (ознака), вилучити мало-інформативні з них, обґрунтувати методику подальшого статистичного оброблення інших запитань.

Статистичний формуляр – це документ єдиного зразка, що містить програму та результати спостереження. Він може мати різні назви: бланк обстеження, переписний лист, анкета, звіт.

Обов'язковими *елементами статистичного формуляра є титульна та адресна частини*. У *титульній частині* вказують: назву статистичного спостереження та органу, що його проводить; номер формуляра, а також, хто і коли його ухвалив. В *адресній* – адресу звітної одиниці, її підпорядкованість.

Інструкція передбачає порядок проведення спостереження чи заповнення формуляра. Залежно від складності програми спостереження це може бути документ у вигляді окремої брошури, або підказки у відповідях, або роз'яснення на зворотній стороні бланка.

Програма спостереження передбачає також визначення виду та способу реєстрації даних. Зазвичай вид і спосіб спостереження (див. п. 2.2)

залежать від його мети, сутності об'єкта спостереження, обсягу та ступеня точності очікуваних результатів.

Другою частиною плану статистичного спостереження є комплекс **організаційних питань**, до яких відносять [68; 71; 74; 88]:

визначення органу (виконавця) спостереження – суб'єкт спостереження (спостереження може проводитися власними силами або організаціями, що спеціалізуються на проведенні спостереження);

визначення *часу спостереження*: дати початку, дати закінчення спостереження, критичної дати. *Термін (період)* спостереження встановлюється виходячи із обсягу роботи та чисельності персоналу, задіяного у збиранні даних. *Критичною датою (моментом)* вважається конкретний день року (час дня), станом на який проводиться реєстрація ознак за кожною одиницею статистичної сукупності;

визначення *місця (території)* проведення спостереження. Наприклад, якщо визначається вартість споживчого кошика населення у м. Харків, то місцем проведення спостереження буде територія міста. Вибір місця проведення спостереження визначається його метою.

Загальна схема організаційних питань плану статистичного спостереження подана на рис. 2.2.



Рис. 2.2. Організаційні питання плану статистичного спостереження

Залежно від масштабності об'єкта спостереження, а також зацікавленості щодо його результатів **суб'єктами спостереження** можуть бути:

центральні органи державної статистики: Державна служба статистики України та її регіональні відділення, що здійснюють державні обстеження на макrorівні (переписи населення, земельного фонду; обстеження міграційних потоків населення, неформальної зайнятості, бюджетів домогосподарств, діяльності суб'єктів бізнесу);

статистичні відділи міністерств і відомств, що здійснюють державні обстеження локального характеру за певною тематикою (обстеження Державної фіскальної служби, Державної прикордонної служби тощо);

спеціальні інститути, агентства, міжнародні організації, що проводять обстеження, які засновані на вивченні: громадської думки або мотивації, поведінки та оцінок окремих суб'єктів суспільно-економічного життя (Міжнародний інститут соціології, Міжнародна організація праці – МОП, Міжнародний центр дослідження проблем тенденцій діяльності економічного характеру, Генеральне агентство у справах економіки та фінансів при комітеті ЄС тощо);

аналітичні відділи окремих економічних структур (підприємств, організацій, фірм, банків, бірж, страхових товариств тощо), що проводять на мікрорівні обстеження маркетингового або контрольного характеру (маркетингові обстеження туристичних фірм, які мають на меті визначити обсяг споживчого ринку; обстеження фактичних і потенційних клієнтів комерційних банків з метою аналізу потреби в різних видах банківських послуг).

Обґрунтування **місця обстеження** передбачає визначення пункту, в якому знаходиться одиниця спостереження і реєструються дані. Наприклад, під час телефонного опитування телеглядачів місцем обстеження є місце проживання респондентів. У разі обстеження осіб, які прибули із-за кордону, про мету їх приїзду – місцем обстеження буде районне відділення віз і реєстрацій (ВВІР).

Для того щоб спостереження дало ймовірні та своєчасні дані, необхідно вирішити питання часу та періоду спостереження.

Час спостереження (об'єктивний час) – це час, до якого належать дані спостереження. Коли об'єктом спостереження є *процес*, то вибирають *інтервал часу*, протягом якого накопичуються дані. Якщо об'єктом спостереження є певний *стан*, то вибирають **критичний момент** – момент часу, станом на який реєструються дані. Наприклад: кількість угод, укладених на універсальній біржі, може реєструватися як за деякий період (протягом місяця), так і за станом на певні моменти – дні біржових торгів,

що відбуваються, як правило, один раз на тиждень. У такому випадку угоди на момент часу фіксуються первинним спостереженням, а за період – вторинним.

Критичний момент використовують під час проведення перепису населення (коли потрібно точно зареєструвати його стан в один "момент"). Так, критичним моментом останнього Всеукраїнського перепису населення був час 24.00 з 4 на 5 лютого 2001 року. Зрозуміло, що переписати все населення країни в одну мить неможливо, тому крім часу спостереження встановлюється **період спостереження (суб'єктивний час)**, протягом якого реєструються дані. Якщо період перепису населення становить, скажімо, 10 днів (з 05.02 до 14.02 включно), то, наприклад, 10 лютого всі дані про членів певного домогосподарства реєструються станом на початок перепису (05.02). Коли в домогосподарстві на період реєстрації народилася дитина, то в переписному аркуші її не враховують, тому що на критичний момент перепису дитини не було.

Час спостереження вибирають в найкращий чи нейтральний для об'єкта спостереження період. Наприклад: перепис плодових насаджень здійснюють у період їх цвітіння і наявності плодів; перепис населення призначають на період найменшої його міграційної активності (для України це – зимові місяці).

Для проведення будь-якого статистичного спостереження необхідно відповідне **матеріально-технічне забезпечення**: засоби для друкування, обчислювальна та розмножувальна техніка, транспортні засоби, статистичний інструментарій та рекламні носії. Останнє важливо в разі підготовки макроспостережень, що вимагають роз'яснень і повідомлень для широкого кола громадян.

Одночасно методологічним (як засіб запобігання помилок) і організаційним (як засіб виявлення та виправлення помилок) питанням плану спостереження є **контроль даних спостереження**.

Складовою організаційних завдань плану статистичного спостереження є також висвітлення його результатів у засобах масової інформації.

2.4. Помилки статистичного спостереження та забезпечення вірогідності даних

Під час проведення статистичного спостереження можуть виникати помилки, тобто розбіжності між реальною ситуацією й отриманими результатами спостереження. Виправлення помилок проводять за допомогою контролю.

Контроль означає перевірку даних обстежень щодо їх повноти та достовірності. Повнота даних контролюється, як правило, візуально: перевіряють їх наявність за всіма одиницями та позиціями. Достовірність даних перевіряють засобами *логічного та арифметичного контролю* [4; 17; 68].

Логічний контроль – це перевірка сумісності даних, яка полягає в порівнянні взаємозалежних ознак. Наприклад, контроль відповідей респондентів, що отримані під час їх опитування: порівняння віку респондента з його сімейним станом, виду діяльності з джерелом засобів існування, матеріального становища господарства зі структурою бюджетних витрат. Логічний контроль установлює лише наявність помилки, а не її розмір. Але іноді вдається визначити наближені межі помилки, порівнюючи дані спостереження з аналогічними даними інших спостережень або дані в динаміці (помилку в даних про котирування цін на біржі визначають, знаючи курс цін і межі його коливань; помилку в розмірі процентних ставок за кредитами знаходять із відомостей про облікову та кредитну ставки НБУ).

Установити розмір помилки та виправити її можна засобами *арифметичного контролю*: прямим або побічним перерахунком зареєстрованих даних (розмір сукупного доходу домогосподарства можна перевірити відносно всіх його витрат за основними статтями; розмір акціонерного капіталу товариства можна проконтролювати, знаючи кількість акціонерів і розмір їх середнього внеску). Якщо помилки виходять за межі логічного та арифметичного контролю, то встановити достовірність даних неможливо, тоді доцільно враховувати природу помилок.

Класифікація помилок статистичного спостереження здійснюється за такими ознаками:

за причинами виникнення розрізняють помилки:

реєстрації – виникають у разі будь-якого спостереження внаслідок викривлення фактів або неправильного їх запису;

репрезентативності – виникають під час вибіркового спостереження через несучільність реєстрації даних і порушення принципів випадковості відбору;

за природою виникнення помилки реєстрації можуть бути:

випадковими – виникають унаслідок збігу випадкових обставин (через неухважність реєстратора або незосередженість респондента). Вони перекручують дані спостереження, проте завдяки масовості випадків дія таких помилок врівноважується, і вони істотно не впливають на результат;

систематичними – виникають у результаті постійних спотворень в одному напрямі (тенденція округляти розміри доходів і витрат до цілих чисел: доходів у бік зменшення, а витрат – у бік збільшення). Такі помилки суттєво зміщують результати спостереження в один бік (збільшення або зменшення).

Систематичні помилки, в свою чергу, підрозділяють на:

ненавмисні – що виникають унаслідок необґрунтованості програми спостереження, некомпетентності реєстраторів. Наприклад: спостерігаючи вплив наслідків аварії на ЧАЕС на стан здоров'я населення, отримані результати значно завищили рівень захворюваності, оскільки було обстежено все населення, яке перебувало в медичних установах; класичний приклад – жіноче кокетування (зменшення віку), старече кокетування (збільшення віку) або округлення віку до чисел, кратних 5 або 10, – так звана вікова акумуляція;

зловмисні – що виникають через свідоме перекручування фактів із метою очорнити або прикрасити дійсність. Наприклад, масові факти приховування доходів від оподаткування; приписки в звітах; приховування виручки від реалізації в комерційній торговельній мережі, пунктах громадського харчування тощо.

Значна вірогідність статистичних даних зумовлюється дієвою системою заходів, спрямованих на зменшення або уникнення помилок. Серед них виділяють: якісний первинний облік, розроблення науково-методичних рекомендацій з питань перевірки вірогідності даних, добір кваліфікованих кадрів статистиків-аналітиків, комп'ютерне оброблення інформації і автоматизацію статистичних робіт.

Важливі поняття

Об'єкт спостереження – статистична сукупність соціально-економічних явищ чи процесів, що підлягають обстеженню.

Одиниця сукупності – елемент статистичної сукупності, який є носієм ознак, що підлягають реєстрації.

Одиниця спостереження – первинна одиниця, від якої отримують інформацію.

План статистичного спостереження – сукупність програмно-методологічних та організаційних питань.

Помилка спостереження – розбіжність між реальною ситуацією і отриманими результатами спостереження.

Програма спостереження – перелік ознак, що підлягають реєстрації (у безпосередньому спостереженні) або перелік питань, за якими збираються відомості (в опитуваннях).

Реєстр – форма безперервного статистичного спостереження за довготривалими процесами.

Спеціально організоване спостереження – спостереження, яке організовується і проводиться з метою отримання відомостей, які відсутні в звітності, або для перевірки її даних.

Статистична звітність – основна форма статистичного спостереження, за допомогою якої статистичні органи у визначені терміни отримують необхідні дані у вигляді встановлених у законному порядку звітних документів.

Статистичне спостереження – планомірне, науково організоване збирання даних або відомостей про масові явища та процеси, яке полягає у реєстрації відібраних ознак щодо кожної одиниці сукупності.

Суб'єктивний час – час, протягом якого реєструються дані.

Час спостереження (об'єктивний час) – час, до якого належать дані спостереження.

Типові завдання

Завдання 2.1. Необхідно підготувати проєкт програми статистичного спостереження для визначення тенденції до зміни кількості крадіжок у регіоні за 2010 – 2018 рр. і скласти проєкт статистичного формуляра.

Розв'язання.

Відповідно до мети спостереження необхідно скласти таблиці-формуляри, до яких буде занесено зібрану інформацію (табл. 2.2).

Таблиця 2.2

Програма статистичного спостереження динаміки щодо кількості крадіжок, зареєстрованих у регіоні за період 2010 – 2018 рр.

Показники кількості крадіжок	2010 р.	...	2018 р.	...	Зміни	
	подія, одиниць	відхилення, %	подія, одиниць	відхилення, %	подія, одиниць	відхилення, %
1	2	3	4	5	6	7
Авто						
Квартирні						

1	2	3	4	5	6	7
Офісні						
У громадському транспорті						
На залізничному транспорті						
На міських ринках						
Крадіжки з розбійним нападом						

За результатами спостереження можна провести динамічний аналіз (табл. 2.3).

Таблиця 2.3

**Проект статистичного формуляра динаміки
щодо кількості крадіжок, зареєстрованих у регіоні
за період 2010 – 2018 рр.**

Роки	Кількість крадіжок, одиниць	Абсолютний приріст, одиниць		Відносний приріст, %	
		за роками	порівняно з 2010 р.	за роками	порівняно з 2010 р.
2010					
2011					
...					
2018					

Завдання 2.2. Подано анкету для отримання посади майстра цеху на підприємстві:

Дата подання: 25.05.2018 р.

ПІБ: Іванов Петро Олексійович

Дата народження: 30.02.1976 р.

Сімейний стан: одружений

Діти: 2

у т. ч. неповнолітні: не має

Освіта: вища технічна

Стаж роботи: 32 років.

Місце попередньої роботи: майстер ділянки гальванічного цеху.

Необхідно провести логічний та арифметичний контроль наведених даних.

Розв'язання.

Логічний контроль проводиться за змістом і відповідністю кожної з отриманих відомостей.

Не пройшла логічний контроль відомість про дату народження: 30.02.1976 р. – у лютому не може бути більше 29 днів, тому наявною є помилка в даті або місяці народження.

Арифметичний контроль проводиться шляхом розрахунків за зв'язаними показниками.

Не проходить арифметичний контроль відомість про стаж роботи:

Іванов Петро Олексійович на дату подання відомостей мав вік 42 повних років. Указаний стаж складає 32 роки, тому виходить, що офіційно Іванов П. О. почав працювати з $42 - 32 = 10$ років, що суперечить чинному законодавству.

Завдання 2.3. Визначте об'єктивний та суб'єктивний час спостереження:

а) дані обліку кредиторської заборгованості банків на початок кварталу мають бути подані протягом десяти днів із початку наступного кварталу;

б) строк подання річних звітів з експорту продукції експортоорієнтованих підприємств – не пізніше 15-го числа з початку наступного року.

Розв'язання.

Виходячи із визначення об'єктивного та суб'єктивного часу спостереження (під якими розуміють час, до якого, відповідно, належать дані спостереження та час, протягом якого здійснюється реєстрація даних) маємо:

а) об'єктивний час – квартал, оскільки інформація щодо кредиторської заборгованості має бути подана на початок кварталу, тобто за попередній квартал; суб'єктивний час – 10 днів, оскільки протягом цього часу інформація має бути оброблена та подана у відповідні органи;

б) об'єктивний час – рік, оскільки інформація щодо обсягів експорту продукції має бути подана за рік (річний звіт); суб'єктивний час – 15 днів, оскільки інформація повинна бути подана протягом 15-ти днів.

Завдання 2.4. Визначте організаційну форму та спосіб спостереження:

а) перелік усіх зареєстрованих економічних структур із зазначенням їх реквізитів, виду та сфери діяльності;

б) обстеження відгуків клієнтів, що обслуговуються на станціях автосервісу.

Розв'язання.

а) організаційна форма спостереження – реєстр, оскільки він є переліком одиниць певного об'єкта спостереження. У цьому випадку – сукупності зареєстрованих економічних структур із зазначенням необхідних ознак (реквізитів), які постійно оновлюються або поповнюються – таких, як назва, адреса, телефон тощо. Спосіб спостереження – документальний облік, оскільки реєстрація або оновлення ознак економічних структур відбувається за даними, наведеними у певних документах;

б) організаційна форма – спеціально організоване спостереження, оскільки така форма спостереження охоплює сфери життя, не відображені у звітності. Так, для обстеження відгуків клієнтів на станціях автосервісу не існує форм звітності, тому це спеціально організоване спостереження. Спосіб спостереження – опитування, яке є несудцільним спостереженням думок, оцінок щодо якості обслуговування на станціях автосервісу, що реєструються зі слів клієнтів цих станцій.

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Огляд можливостей Excel" подані у роботі [67]. Лабораторна робота націлена на опанування та закріплення основних прийомів роботи з електронними таблицями: введення та редагування даних, форматування, проєктування таблиць.

Завдання для самодіагностики

1. Що таке статистичне спостереження, в чому полягають його сутність і основні завдання?

2. Назвіть основні програмно-методологічні й організаційні питання статистичного спостереження.

3. Що таке об'єкт, одиниця спостереження та одиниця сукупності?

4. Що таке критичний момент і час спостереження?

5. Яких вимог дотримуються під час формування програми статистичного спостереження?

6. Сформулюйте об'єкт, одиницю і мету статистичного спостереження і розробіть програму: а) обстеження діяльності експорто-орієнтованих підприємств; б) аналізу витрат підприємства на рекламу; в) аналізу використання робочої сили та робочого часу на підприємстві.
7. Назвіть основні форми статистичного спостереження.
8. Розкажіть про види статистичного спостереження: а) за ступенем охоплення одиниць статистичної сукупності; б) за часом реєстрації даних.
9. Що таке помилки статистичного спостереження та які їх види?
10. Як здійснюється контроль результатів спостереження? Назвіть способи усунення помилок статистичного спостереження.

Питання для критичного переосмислення (есе)

1. Складіть перелік запитань, що входять до програми спостережень:
- а) опитування студентів магістратури щодо практичної спрямованості навчальних дисциплін, які викладаються;
 - б) вибіркового обстеження бюджету молодих сімей регіону;
 - в) обліку кредитних операцій банків.
2. Яким чином створення автоматизованих банків даних впливає на розроблення програми спостереження?
3. Розробіть програму проведення статистичного спостереження та здійснення контролю отриманих результатів щодо обстеження обсягів виробництва продукції та факторів виробництва промислових підприємств регіону.
4. Наведіть переваги та недоліки форм статистичного спостереження. Відповідь обґрунтуйте.
5. Охарактеризуйте будь-які три різноманітні статистичні обстеження за доцільними видами та способами їх здійснення.

3. Подання статистичних даних: таблиці, графіки, карти

Основні питання:

- 3.1. Роль і значення графічного методу.
- 3.2. Основні елементи графіка. Правила побудови статистичних графіків.
- 3.3. Види статистичних графіків і таблиць. Способи їх побудови.

3.1. Роль і значення графічного методу

Галузь візуалізації інформації з'явилася внаслідок досліджень взаємодії людини та комп'ютера, комп'ютерних наук, графіків, дизайну, психології та економіки. Вона все частіше застосовується як найважливіший компонент у наукових дослідженнях, цифрових бібліотеках, для інформаційно-аналітичного аналізу даних тощо. Візуалізація інформації спрямована на створення нових і більш наочних підходів до передання даних в інтуїтивно зрозумілі способи.

Серед фундаментальних підходів до аналізу даних особливе місце належить **візуалізації статистичних даних** – візуальному аналізу даних, що спирається на пізнавальні навички аналітиків. **Графічний формат** статистичних даних спрощує розуміння великої кількості інформації та зв'язків між різними серіями даних. Саме тому аналітик повинен застосовувати різні методи та інструменти, щоб інтерпретувати будь-який соціально-економічний процес за допомогою візуалізації статистичних даних.

Графіки є важливим засобом вираження та аналізу статистичних даних, оскільки наочне уявлення полегшує сприйняття інформації. Графіки дозволяють миттєво охопити й осмислити сукупність показників – виявити найбільш типові їх співвідношення та зв'язки, визначити тенденції розвитку, охарактеризувати структуру, ступінь виконання плану, оцінити географічне розміщення об'єктів. Цим пояснюється застосування графіків для пропаганди статистичної інформації, що характеризує результати розвитку різних сфер національної економіки та соціальних відносин.

Статистичний графік – це креслення, на якому статистичні сукупності, що характеризуються певними показниками, описуються за допомогою умовних геометричних образів або знаків [18; 30; 40; 51].

Подання табличних даних у вигляді графіка справляє сильніше враження, ніж цифрове, дозволяє краще осмислити результати статистичного спостереження, правильно їх тлумачити, значно полегшує розуміння статистичного матеріалу, робить його наочним і доступним. Проте графіки мають не лише ілюстративне значення. Вони дають нове знання про предмет дослідження, будучи методом узагальнення та візуалізації початкової інформації.

У зв'язку з упровадженням в аналітичну роботу нових економіко-математичних методів, сучасної обчислювальної техніки та пакетів прикладних програм комп'ютерної графіки неможливо уявити наукові дослідження без використання графічного методу з його широким арсеналом засобів візуалізації.

Особливо актуальним і різноплановим є використання графічного методу в статистичних дослідженнях для унаочнення складних взаємозв'язків між досліджуваними показниками, тенденцій та закономірностей масових соціально-економічних явищ і процесів у динаміці та просторі. За допомогою графіків візуалізують структуру явищ, вивчають результати порівняння, контролю за виконанням плану, досліджують ступінь розповсюдженості на території тих чи інших процесів та явищ. Більш того, графічний метод широко використовується у міжнародних порівняннях і зіставленнях, для пропаганди передового досвіду, наочності прогресивних технологій і наукових досягнень тощо.

Отже, значення графічного методу в аналізі й узагальненні даних досить велике. Графічне зображення дозволяє здійснити контроль достовірності статистичних показників, оскільки, подані на графіку, вони яскравіше висвітлюють допущені неточності, пов'язані або з наявністю помилок спостереження, або із сутністю самого явища, що вивчається. За допомогою графічного зображення можливі вивчення закономірностей розвитку явища, встановлення існуючих взаємозв'язків.

Просте зіставлення даних не завжди дає можливість вловити структуру причинних залежностей, тоді як їх графічне зображення сприяє виявленню причинних зв'язків, особливо в разі встановлення первинних гіпотез, що підлягають подальшому дослідженню. Графіки широко використовують для вивчення структури явищ, їх зміни в часі та розміщення в просторі. У них виразніше виявляються порівняльні характеристики та увиразнюються основні тенденції розвитку та взаємозв'язку, властиві процесу або явищу, що вивчається.

3.2. Основні елементи графіка.

Правила побудови статистичних графіків

У ході побудови графічного зображення слід дотримуватися низки *вимог*. Перш за все графік має бути достатньо наочним, оскільки весь сенс графічного зображення як методу аналізу в тому і полягає, щоб наочно

зобразити статистичні показники. Крім того, графік має бути виразним, дохідливим і зрозумілим.

Для виконання наведених вимог кожен графік повинен включати ряд основних елементів:

- 1) графічний образ;
- 2) поле графіка;
- 3) просторові орієнтири;
- 4) масштабні орієнтири;
- 5) експлікацію графіка [28; 68; 74; 85].

Розглянемо докладніше кожен із вказаних елементів.

Графічний образ (основа графіка) – це геометричні знаки, тобто сукупність крапок, ліній, фігур, за допомогою яких зображуються статистичні показники [68]. Важливо правильно вибрати графічний образ, який повинен відповідати меті графіка та сприяти найбільшій виразності відображуваних статистичних даних. Графічними є лише ті образи, в яких властивості геометричних знаків – фігура, розмір ліній, розташування частин – мають істотне значення для виразу змісту зображуваних статистичних величин, причому кожній зміні зображуваного змісту відповідає зміна графічного образу.

Поле графіка – це частина площини, де розташовані графічні образи. Поле графіка має певні розміри, які залежать від його призначення [68; 85].

Просторові орієнтири графіка задаються у вигляді системи координатних сіток. Система координат необхідна для розміщення геометричних знаків у полі графіка. Найбільш поширеною є *система прямокутних координат*.

У полярній системі координат один із променів (зазвичай правий горизонтальний) береться за вісь координат, щодо якої визначається кут променя. За другу координату вважається її відстань від центру сітки, звана **радіусом**. У радіальних графіках промені позначають моменти часу, а кола – величини явища, що вивчається [28].

На статистичних картах просторові орієнтири задаються контурною сіткою (контури річок, берегова лінія морів та океанів, межі держав) і визначають ті території, до яких віднесені статистичні величини.

Масштабні орієнтири статистичного графіка визначають за масштабом і системою масштабних шкал. **Масштаб статистичного графіка** – це міра перекладу числової величини в графічну [51; 68].

Масштабною шкалою називають лінію, окремі точки якої можуть бути прочитані як певні числа [28; 74].

Шкала має велике значення в графіку. Вона включає три елементи: лінію (носій шкали); певну кількість помічених рисками крапок, які розташовані на носіїв шкали в певному порядку; цифрове позначення чисел, відповідних окремим поміченим крапкам. Як правило, цифровим позначенням забезпечуються не всі помічені крапки, а лише деякі з них, розташовані в певному порядку. Згідно з правилами, числове значення необхідно розміщувати строго навпроти відповідних крапок, а не між ними.

Носій шкали може бути як прямою, так і кривою лінією. Тому розрізняють **шкали прямолінійні** (наприклад, міліметрова лінійка) і **криволінійні** – дугові та кола (циферблат годинника).

Графічні та числові *інтервали бувають рівними або нерівними*. Якщо протягом всієї шкали рівним графічним інтервалам відповідають рівні числові, така шкала є *рівномірною*. Коли рівним числовим інтервалам відповідають нерівні графічні інтервали, та навпаки, – шкала буде *нерівномірною*.

Масштабом рівномірної шкали називають **довжину відрізання** (графічний інтервал), прийнятого за одиницю і виміряного в деяких одиницях вимірювання. Чим менше масштаб, тим густіше розташовуються на шкалі крапки з однаковим значенням [10; 85].

Побудувати шкалу – означає на заданому носіїв шкали розмістити крапки та позначити їх відповідними числами згідно з умовами завдання. Як правило, масштаб визначається зразковою прикидкою можливої довжини шкали і її меж. Наприклад, на полі в 20 клітин треба побудувати шкалу від 0 до 850. Оскільки 850 без залишку не ділиться на 20, то округлимо число 850 до найближчого зручного числа.

З нерівномірних шкал найбільше розповсюдження має логарифмічна шкала. Методика її побудови дещо інша, оскільки на цій шкалі відрізки пропорційні не зображуваним величинам, а їх логарифмам.

Останній елемент графіка – **експлікація**. Кожен графік повинен мати словесний опис його змісту. Опис включає назву графіка, яка в короткій формі передає його зміст; підписи вздовж масштабних шкал і пояснення до окремих частин графіка.

Сьогодні розроблені та широко використовуються пакети прикладних програм комп'ютерної графіки, веб-додатки та бібліотеки, які полегшують завдання аналітика в практичному застосуванні графіків.

Найбільш поширеними є ППП "Harvard graphics", "Statgraf", "STATISTICA", "Excel", "SPSS", "R", Tableau, Project BI, Plotly, Chart.js, Raw, Dygraphs, FusionCharts та ін. [94; 101 – 102].

Для графічного зображення статистичних даних використовують різноманітні графіки, класифікація яких наведена на рис. 3.1.

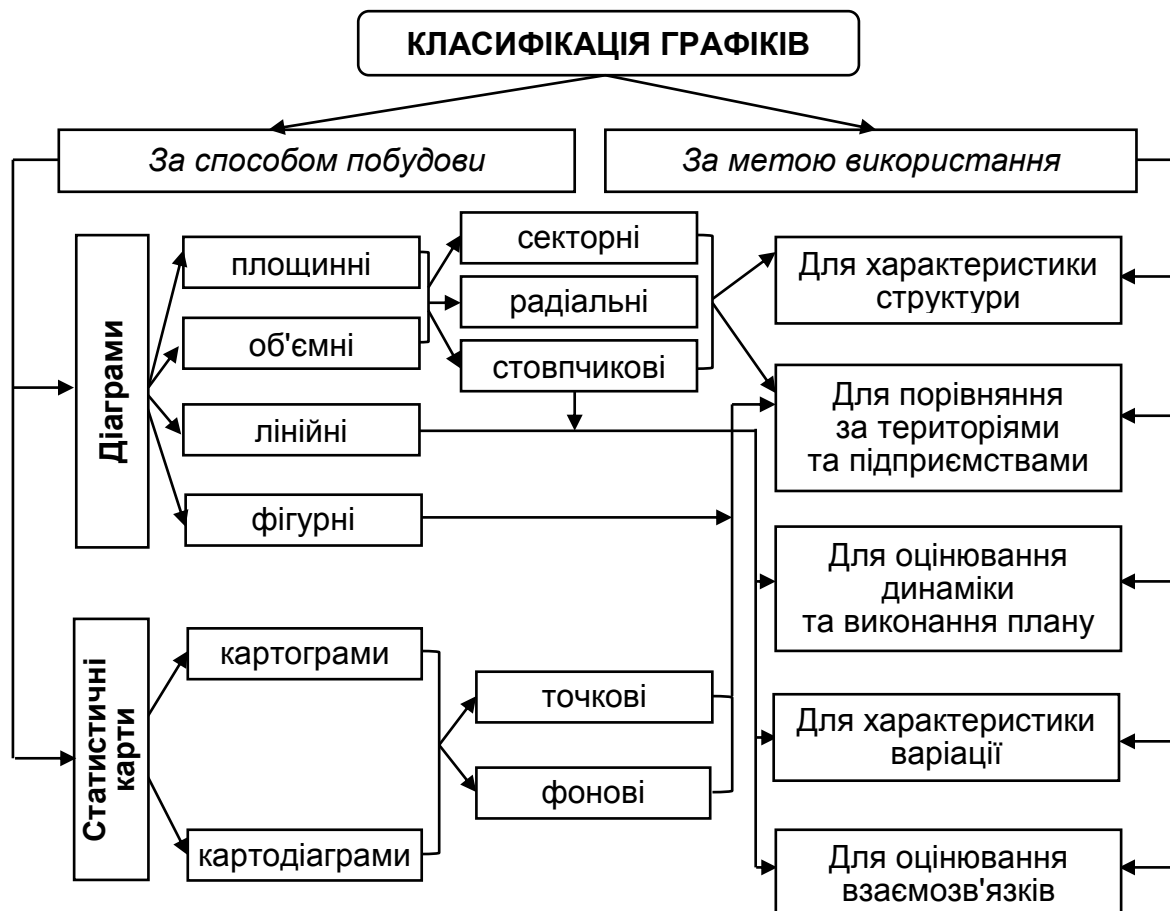


Рис. 3.1. Класифікація графіків [47; 62; 68]

Доцільно зазначити, що зображена на рис. 3.1 класифікація не є вичерпною, оскільки бібліотеки сучасних програмних продуктів візуалізації даних дозволяють використовувати широкий асортимент інфографіки відповідно до мети аналізу та наявних статистичних даних.

Незважаючи на розмаїття видів графічних зображень, щодо їх побудові виконуються загальні правила:

- відповідно до мети використання обирається відповідний тип графіка;
- визначається поле графіка, той простір, в якому розміщуються геометричні знаки;

задаються масштабні орієнтири за допомогою масштабних шкал (рівномірних або нерівномірних);

обирається система координат, необхідна для розміщення геометричних знаків у поле графіка [10; 68; 74].

Дотримання зазначених правил дозволить побудувати якісну візуалізацію.

3.3. Види статистичних графіків і таблиць. Способи їх побудови

Розглянемо найбільш поширені види графіків.

Один із способів графічного подання статистичних даних є діаграма.

Діаграма – це спосіб візуалізувати інформацію, наведену у вигляді таблиці чисел [15; 37]. Графічний формат діаграми спрощує розуміння великої кількості інформації та зв'язків між різними серіями даних. Діаграма надає огляд ситуації, що дозволяє згрупувати дані та виокремити важливі тенденції.

Лінійні діаграми застосовують для характеристики динаміки, тобто оцінювання зміни явищ у часі; для характеристики варіаційних рядів; для оцінювання виконання плану; для оцінювання взаємозв'язку між явищами [14; 40; 92]. Вони будуються на прямокутній системі координат, де на осі абсцис відкладають проміжки часу, а на осі ординат – рівні ряду динаміки або темпи зміни показників. Отримані точки з'єднують ламаним лініями. На одному графіку можна розмістити кілька діаграм, що дозволяє порівнювати динаміку різних показників. У варіаційних рядах такі діаграми називають **полігонами розподілу**.

Для тих же цілей можна використовувати **стовпчикові діаграми**. Стівпчики мають однакову основу, а їх висота відповідає числовій характеристиці значень ознаки. За висотою стівпчиків визначають співвідношення між рівнями досліджуваних показників. У варіаційних рядах такі графіки називаються **гістограмами**.

Стівпчикові діаграми також можуть використовуватися для просторових зіставлень: порівняння за територіями, країнами, підприємствами. Крім того, вони широко використовуються для вивчення структури явищ. Правила побудови стівпчикових діаграм допускають одночасне розташування на одній горизонтальній осі зображень декількох показників. У цьому

випадку стовпчики мають у своєму розпорядженні групи, для кожної з яких може бути прийнята різна розмірність ознак, що варіюють.

Різновидами стовпчикових діаграм є так звані **стрічкові** або **смугові діаграми**. Їх відмінність полягає в тому, що масштабна шкала розташована на горизонталі зверху або знизу; вона визначає величину смуг за довжиною. Сфера застосування стовпчикових і смугових діаграм однакова, оскільки ідентичні правила їх побудови. Одновимірність зображуваних статистичних показників і їх одномасштабність для різних стовпчиків і смуг вимагають виконання єдиного положення: дотримання відповідності (стовпчиків – за висотою, смуг – за довжиною) і пропорційності зображуваним величинам. Для виконання цієї вимоги необхідно: щоб шкала, за якої встановлюється розмір стовпчика (смуги), починалася з нуля; ця шкала має бути безперервною, тобто охоплювати всі числа певного статистичного ряду; розриви шкали та, відповідно, стовпчиків (смуг) не допускаються. Невиконання вказаних правил призводить до спотвореного графічного зображення аналізованого статистичного матеріалу [13; 18; 68].

Різновидом стовпчикових (стрічкових) діаграм є *спрямовані діаграми*. Вони відрізняються від звичайних двостороннім розташуванням стовпчиків або смуг і мають початок відліку за масштабом у середині. Зазвичай такі діаграми застосовують для зображення величин протилежного якісного значення. Порівняння між собою різноспрямованих стовпчиків (смуг) менш ефективно, ніж розташованих поряд, односпрямованих. Попри це аналіз спрямованих діаграм дозволяє робити достатньо змістовні висновки, оскільки особливе розташування додає графіку яскраве зображення.

До групи двосторонніх належать *діаграми числових відхилень*. У них смуги скеровані в обидва боки від вертикальної нульової лінії: вправо – для приросту; вліво – для зменшення. За допомогою таких діаграм зручно зображати відхилення від плану або деякого рівня, прийнятого за базу порівняння. Важливою перевагою цих діаграм є можливість бачити розмах коливань аналізованої статистичної ознаки, що має велике значення для економічного аналізу.

Для простого порівняння незалежних один від одного показників можна використовувати діаграми, принцип побудови яких полягає в тому, що порівнювані величини зображуються у вигляді правильних геометричних фігур, побудованих так, щоб їх площі співвідносились між собою як кількості. Іншими словами, ці *діаграми виражають величину явища розміром*

своєї зображеної площі. Для отримання діаграм цього типу використовують різноманітні геометричні фігури: квадрат, коло, рідше – прямокутник. Відомо, що площа квадрата дорівнює квадрату його сторони, а площа кола визначається пропорційно квадрату його радіусу.

Тому для побудови діаграм необхідно спочатку з порівнюваних величин отримати квадратний корінь. Потім на базі отриманих результатів визначити сторону квадрата або радіус кола відповідно до прийнятого масштабу.

Для правильної побудови діаграм квадрата або кола необхідно розташувати на однаковій відстані від одного відстані, а в кожній фігурі вказати числове значення, яке вона зображає, не приводячи масштабу вимірювання. До цього виду діаграм відносять **графічне зображення**, отримане шляхом побудови один в іншому квадратів, кіл або прямокутників з різною штриховкою або зафарбовуванням. Такі діаграми також дозволяють порівнювати між собою низку досліджуваних величин.

Найбільш виразним і легко сприйнятним є спосіб побудови **діаграм порівняння** у вигляді фігур-знаків. У цьому випадку статистичні сукупності зображуються не геометричними фігурами, а символами або знаками, які певним чином відтворюють зовнішній образ статистичних даних [62; 65; 90]. Перевага такого способу графічного зображення полягає у високому ступені наочності, в отриманні подібного відображення, що віддзеркалює змістовність порівнюваних сукупностей.

Найважливіша ознака будь-якої діаграми – масштаб. Тому, щоб правильно побудувати фігурну діаграму, необхідно визначити **одиницю рахунку**. За таку приймається окрема фігура (символ), якій умовно присвоюється конкретне числове значення, а досліджувана статистична величина зображується окремою кількістю однакових за розміром фігур, послідовно розташованих на рисунку [36]. Проте в більшості випадків не вдається зобразити статистичний показник цілою кількістю фігур. Останню з них доводиться ділити на частини, оскільки за масштабом один знак є дуже крупною одиницею вимірювання. Зазвичай ця частина визначається "на око". Складність точного її визначення є недоліком фігурних діаграм. Проте якщо велика точність подання статистичних даних не переслідується, то отримані результати будуть цілком задовільними.

Для характеристики структури соціально-економічних явищ використовують також **секторні діаграми**. Аналіз структури проводиться на основі зіставлення різних частин цілого за допомогою площ, утворених секторами кола [37].

Радіальні діаграми використовують для зображення явищ, які періодично змінюються в часі (наприклад, залежно від сезонних коливань).

Для їх побудови використовують полярну систему координат. Круг ділиться на 12 рівних частин, кожна з яких означає місяць. На радіусі, починаючи з центру, в масштабі відкладають відрізки, які відображають місячні рівні ознаки. Їх кінці з'єднують відрізком між собою, в результаті чого з'являється дванадцятигранник, який характеризує сезонність явищ.

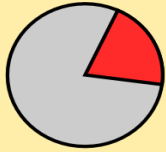

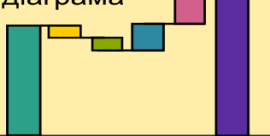



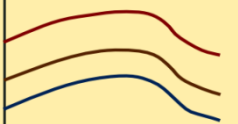
У табл. 3.1 наведена класифікація видів діаграм відповідно до наявних даних і мети візуалізації [37; 40; 44; 92].

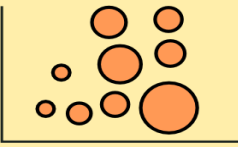


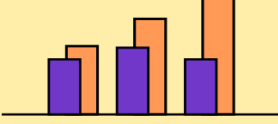

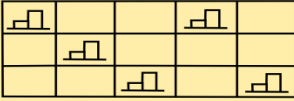
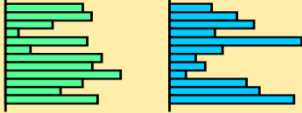
Таблиця 3.1

Види діаграм відповідно до мети та типу наявних даних

Мета візуалізації статистичних даних		Характеристика наявних даних	Рекомендований вид діаграми
1	2	3	4
Розподіл	Одна змінна	Небагато точок даних / категорій / інтервалів	<div style="border: 1px solid black; padding: 5px;"> <p>Стовпчикова гістограма</p>  <p>Column Histogram</p> </div>
		Багато точок даних	<div style="border: 1px solid black; padding: 5px;"> <p>Лінійна гістограма</p>  <p>Line Histogram</p> </div>
Розподіл. Зв'язок чи залежність	Дві змінні	Не важливо	<div style="border: 1px solid black; padding: 5px;"> <p>Діаграма розкиду</p>  <p>Scatter Chart</p> </div>
Розподіл	Три змінні	Не важливо	<div style="border: 1px solid black; padding: 5px;"> <p>Тривимірна ареальна діаграма</p>  <p>3D Area Chart</p> </div>

Продовження табл. 3.1

1	2	3	4
Композиція (структура досліджуваного явища чи процесу)	Статична	Проста частина в цілому	<p>Кругова діаграма</p>  <p>Pie Chart</p>
		Складові компонентів	<p>Стосова 100% стовпчаста діаграма з підкомпонентами</p>  <p>Stacked 100% Column Chart with Subcomponents</p>
Композиція (структура досліджуваного явища чи процесу)	Статична	Накопичення додавань і віднімань у цілому	<p>Водоспадна діаграма</p>  <p>Waterfall Chart</p>
Інтенсивність змін явища (процесу) протягом часу	Небагато періодів	Важливі тільки відносні відмінності	<p>Стосова 100% стовпчаста діаграма</p>  <p>Stacked 100% Column Chart</p>
		Важливі абсолютні та відносні відмінності	<p>Стосова стовпчаста діаграма</p>  <p>Stacked Column Chart</p>
	Багато періодів	Важливі тільки відносні відмінності	<p>Стосова 100% ареальна діаграма</p>  <p>Stacked 100% Area Chart</p>
		Важливі і абсолютні, і відносні відмінності	<p>Стосова ареальна діаграма</p>  <p>Stacked Area Chart</p>

1	2	3	4
Зв'язок або залежність	Три змінних	Не важливо	<p>Бульбашкова діаграма</p>  <p>Bubble Chart</p>
Порівняння	У часі	Багато періодів – циклічні дані	<p>Кругова діаграма «Павутиння»</p>  <p>Circular Area Chart</p>
Порівняння	У часі	Багато періодів та ациклічні дані	<p>Лінійна діаграма</p>  <p>Line Chart</p>
		Мало періодів, багато категорій	
	У часі та між об'єктами	Мало періодів, одна чи декілька категорій	<p>Стовпчаста діаграма</p>  <p>Column Chart</p>
		Одна змінна на об'єкт, мало категорій, мало об'єктів	
	Між об'єктами	Дві змінні на об'єкт	<p>Стовпчаста діаграма зі змінною шириною стовпчиків</p>  <p>Variable Width Column Chart</p>
		Одна змінна на об'єкт, багато категорій	<p>Таблиця або таблиця з вбудованими діаграмами</p>  <p>Table or Table with Embedded Charts</p>
		Одна змінна на об'єкт, мало категорій, багато об'єктів	<p>Горизонтальна стовпчаста діаграма</p>  <p>Bar Chart</p>

Для оцінювання географічного розміщення явищ і порівняльного аналізу територій застосовують статистичні карти. Вони досить часто використовуються в публікаціях ООН.

Статистичні карти є видом графічних зображень статистичних даних на схематичній географічній карті; характеризують рівень або ступінь розповсюдження того або іншого явища на певній території [15; 47; 68; 94]. Засобами зображення територіального розміщення є штрихування, фонове розфарбовування або геометричні фігури. Розрізняють *картограми* та *картодіаграми*.

Картограма показує територіальний розподіл досліджуваного явища за окремими районами; використовується для виявлення закономірностей цього розподілу. **Картограма** – це схематична географічна карта, на якій штрихуванням різної густини, крапками або забарвленням певного ступеня насиченості показується порівняльна інтенсивність якого-небудь показника в межах кожної одиниці нанесеного на карту територіального ділення (наприклад, щільність населення за областями або регіонами, розподіл районів за врожайністю зернових культур тощо). Картограми діляться на *фонові* та *точкові*.

Картограма фонові – вид картограми, на якій штрихуванням різної густини або забарвленням певного ступеня насиченості показують інтенсивність якого-небудь показника в межах територіальної одиниці.

Картограма точкові – вид картограми, де рівень обраного явища зображується за допомогою крапок [51; 68]. Крапка зображує одну одиницю сукупності або деяку їх кількість, показуючи на географічній карті щільність або частоту прояву певної ознаки.

Фонові картограми, як правило, використовують для зображення середніх або відносних показників, *точкові* – для об'ємних (кількісних) показників (чисельність населення, поголів'я худоби тощо).

На рис. 3.2 наведено приклад фонові картограми.

Другу велику групу статистичних карт складають **картодіаграми**, що є поєднанням діаграм із географічною картою. Як образотворчі знаки в картодіаграмах використовують діаграмні фігури (стовпчики, квадрати, кола, фігури, смуги), які розміщують на контурі географічної карти. Картодіаграми дають можливість географічно відобразити складніші статистико-географічні побудови, ніж на картограмах. Серед картодіаграм слід виділити картодіаграми простого порівняння, графіки просторових переміщень, ізоліній [14; 40; 65].

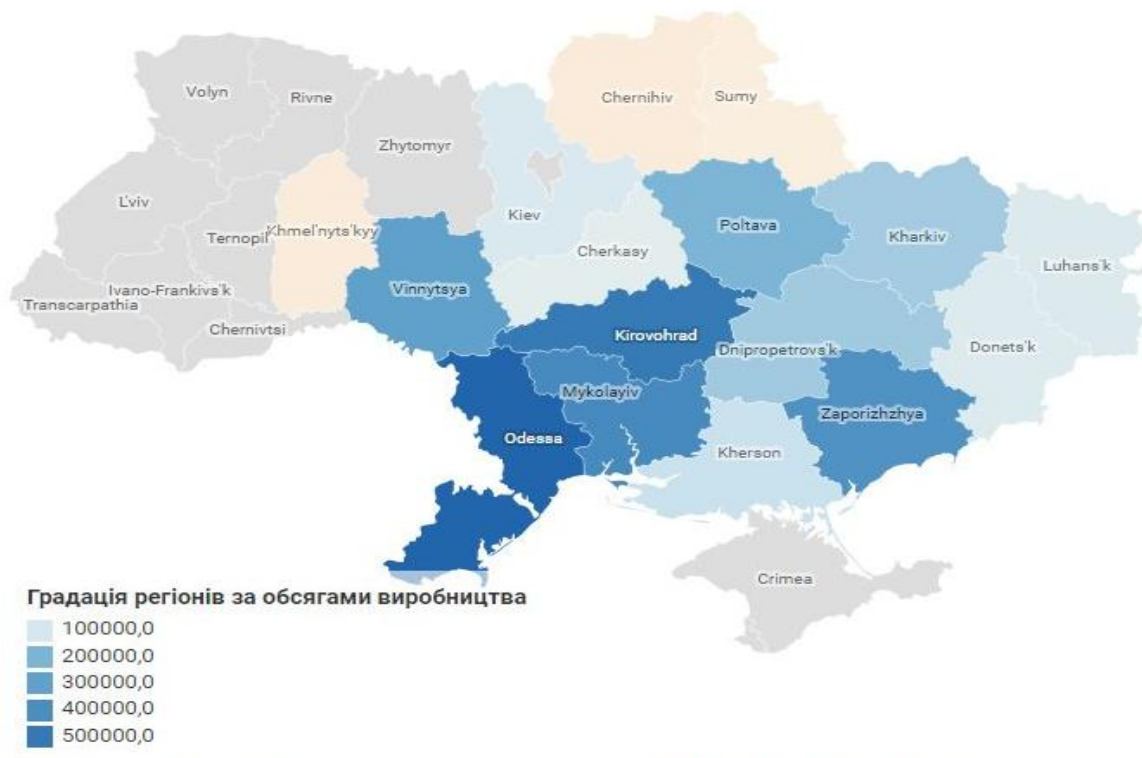


Рис. 3.2. Картограма регіонів України за обсягами виробництва соняшникової олії у 2018 р.
(розроблено автором)

На картодіаграмі *простого порівняння* (на відміну від звичайної діаграми) діаграмні фігури, що зображають величини досліджуваного показника, розташовані не в ряд, як на звичайній діаграмі, а розносяться на всій карті відповідно до того району, області або країни, яку вони представляють. Елементи простої картодіаграми можна виявити на політичній карті, де міста відрізняються різними геометричними фігурами залежно від кількості жителів.

Графіки просторових переміщень використовують для зображення на карті різноманітних просторових переміщень: природних (діяльність вітрів, морські течії за деякий проміжок часу) та соціально-економічних (міграція населення, рух капіталу, обсяг перевезених вантажів). Використовуючи різні графічні засоби (стрілки різного кольору, форми та ширини), на таких графіках можна подати напрям і швидкість переміщень, кількість, якість та інші характеристики досліджуваних статистичних явищ і процесів, що переміщуються в просторі.

Заголовок таблиці відображає зміст таблиці, місце та час, з яким співвіднесені її дані, одиниці вимірювання, якщо вони є спільними для наведених даних.

Підметом таблиці є перелік одиниць сукупності або групи, тобто об'єкт вивчення; **присудком** – цифрові дані, що характеризують підмет, тобто результати зведення [62; 68; 85].

Зазвичай *підмет* розташовується зліва у вигляді назв рядків, а *присудок* – зверху у вигляді назв графів.

За **змістом підмета** всі статистичні таблиці можна розподілити на такі групи:

прості таблиці, у підметах котрих відсутні групування. Вони містять узагальнювальні показники, що належать до переліку: одиниць сукупності (перелікові таблиці); хронологічних дат (хронологічні таблиці); територій (територіальні таблиці);

групові таблиці, в підметі котрих досліджуваний об'єкт роз'єднаний на групи за певною ознакою. Кожна група може бути охарактеризована низкою показників;

комбінаційні таблиці, у підметах котрих надане групування одиниць сукупності за двома та більше ознаками, взятими в комбінації [68].

Присудок таблиці може бути простим або складним. *Простий присудок* передбачає паралельне розташування показників (графи 1, 2 у макеті таблиці), а *складний* – комбіноване (графи 3, 4, 5).

Одним із різновидів статистичних таблиць є *таблиці спряженості* (таблиця співзалежності, факторна таблиця), що широко використовується як універсальний засіб вивчення статистичних зв'язків між змінними. Такий вид статистичних таблиць дає змогу побачити основну картину взаємозв'язку між двома змінними, допомагає знайти взаємодію між ними. Це дозволило їм також знайти своє призначення в бізнес-аналітиці, наукових та інженерних дослідженнях.

Таблицею взаємної спряженості називають таблицю, яка містить зведену числову характеристику сукупності, що вивчається за двома та більше атрибутивними ознаками або комбінацією кількісних та атрибутивних ознак. Таблиця спряженості має вигляд матриці, що показує спільний розподіл частоти двох змінних [51; 65].

Макет таблиці спряженості розмірністю $i \times j$, де $i = 1, 2 \dots k$ – число варіантів значень однієї ознаки (A); $j = 1, 2 \dots n$ – число варіантів значень другої ознаки (B), наведено в табл. 3.2.

Загальна схема таблиці спряженості

	B_1	B_2	...	B_i	Разом
A_1	f_{11}	f_{12}	...	f_{1j}	f_{10}
A_2	f_{21}	f_{22}	...	f_{2j}	f_{20}
...
A_i	f_{i1}	f_{i2}	...		
Разом	f_{01}	f_{02}	...	f_{0j}	f_{00}

У побудові статистичних таблиць необхідно дотримуватись певних **правил** щодо їх оформлення:

1) таблиця повинна бути компактною, легко доступною для огляду. Її не слід завантажувати зайвими подробицями, що затрудняють аналіз;

2) заголовок таблиці повинен ясно та стисло висловлювати її зміст. Заголовки рядків підмета та графів присудка також треба формулювати точно та лаконічно;

3) у таблиці бажано давати нумерацію графів. Це полегшує користування таблицею, що показує спосіб розрахунку чисел у графах. Графи, які містять підмет, позначаються великими літерами алфавіту; графи, які містять присудок, нумеруються арабськими числами. Не допускається у заголовках підмета та присудка скорочення слів;

4) якщо одиниці вимірювання різні, то їх указують у назвах рядків і графів;

5) наведені в підметі та присудку ознаки треба розташовувати в логічному порядку з урахуванням необхідності їх спільного розгляду. Інформація розміщується від часткового до загального, тобто спочатку показують складові, а наприкінці підводять підсумки;

6) якщо в таблиці наведено не всі дані, а тільки найбільш значущі з них, то спочатку показують підсумок, а потім виділяють найважливіші частини за допомогою оборотів "у тому числі", "з них";

7) слід розрізняти "Разом" і "Усього". "Разом" є підсумком для певної частини сукупності, а "Усього" – підсумок для всієї сукупності;

8) в оформленні таблиці застосовують такі позначення:

прочерк (–) – якщо явище відсутнє;

символ "x" – якщо явище не має осмисленого змісту;

три крапки (...) – якщо відсутні відомості (або робиться запис "немає відомостей");

якщо відомості є, але їх числове значення менше прийнятої в таблиці точності, воно виражається дробовим числом 0,0;

9) округлення чисел, наведених у таблиці, має проводитися з однаковою мірою точності;

10) якщо одна величина перевершує іншу багатократно, то отримані відносні показники краще висловлювати не у відсотках, а в кількості разів [68].

Дотримання наведених правил побудови й оформлення статистичних таблиць робить їх основним засобом подання, оброблення та узагальнення статистичної інформації.

Важливі поняття

Графічний образ (основа графіка) – геометричні знаки, тобто сукупність крапок, ліній, фігур, за допомогою яких зображуються статистичні показники.

Довжина відрізання (графічний інтервал) – масштаб рівномірної шкали, прийнятий за одиницю і зміряний в деяких вимірниках.

Експлікація – словесний опис змісту графіка.

Заголовок таблиці – поняття, яке відображає зміст таблиці, місце та час, з яким співвіднесені її дані, одиниці вимірювання, якщо вони є спільними для наведених даних.

Ізолнії – лінії рівного значення певної величини в її розповсюдженні на поверхні, зокрема на географічній карті або графіку.

Картограма – діаграма, яка показує територіальний розподіл досліджуваного явища за окремими районами; використовується для виявлення закономірностей цього розподілу.

Картодіаграма – діаграма, яка є поєднанням діаграми з географічною картою.

Лінійні діаграми – застосовують для характеристики динаміки, тобто оцінювання зміни явищ у часі; для характеристики варіаційних рядів; для оцінювання виконання плану; для оцінювання взаємозв'язку між явищами.

Масштаб статистичного графіка – міра перекладу числової величини в графічну.

Масштабна шкала – лінія, окремі точки якої можуть бути трактовані як певні числа.

Підмет таблиці – перелік одиниць сукупності або групи, тобто об'єкт вивчення; *присудок* – цифрові дані, що характеризують підмет, тобто результати зведення.

Поле графіка – це частина площини, де розташовані графічні образи.

Радіальні діаграми – діаграми, які використовуються для зображення явищ та які періодично змінюються в часі.

Секторні діаграми – діаграми, які можуть використовуватися для характеристики структури соціально-економічних явищ.

Статистична таблиця – форма раціонального та наочного викладення статистичних даних про досліджувані явища.

Статистичні карти – вид графічних зображень статистичних даних на схематичній географічній карті, які характеризують рівень або ступінь розповсюдження того або іншого явища на певній території.

Стовпчикові діаграми – діаграми, які можуть використовуватися для просторових зіставлень: порівняння за територіями, країнами, фірмами та для вивчення структури явищ.

Таблиця взаємної спряженості – таблиця, яка містить зведену числову характеристику сукупності, що вивчається за двома та більше атрибутивними ознаками або комбінацією кількісних та атрибутивних ознак.

Типові завдання

Завдання 3.1. Є дані про міграційних рух населення України за період 2013 – 2018 рр. (табл. 3.3). Використовуючи ППП Microsoft Excel, графічно надайте динаміку зміни кількості прибулого та вибулого населення. Обґрунтуйте вибір типу діаграми.

Таблиця 3.3

Міграційних рух населення України

Роки	Кількість прибулих, осіб	Кількість вибулих, осіб
1	2	3
2013	761 842	621 842
2014	622 506	519 914
2015	613 278	519 045

1	2	3
2016	556 808	506 188
2017	442 287	430 290
2018	595 333	561 656

Розв'язання.

Microsoft Excel – це багатофункціональна програма, яка використовується в будь-якій галузі (сфері) діяльності. Різні інструменти MS Excel для розрахунку та аналізу дозволяють обробляти велику кількість даних і візуально оформлювати результати. Зокрема, MS Excel дозволяє створювати велику кількість різних діаграм. Діаграми створюються на основі даних, що містяться в робочих аркушах книги MS Excel. Більш того, всі типи діаграм у MS Excel є динамічними, тобто зі зміною вихідних даних, на яких побудована діаграма, вони автоматично оновлюються і на вже побудованій діаграмі.

Для побудови стандартної діаграми в MS Excel необхідно виконати такі дії:

1) виділити вихідний інформаційний простір, тобто задати діапазон даних (таблиця з вихідними даними на робочому аркуші MS Excel);

2) запустити "Майстер діаграм" або відкрити вкладку "Вставка" панелі інструментів; вибрати меню "Діаграми" й обрати необхідний тип діаграми (гістограма, графік, кругова, лінійчаста, секторна, точкова, біржова, поверхнева, кільцева, бульбашкова, пелюсткова);

3) задати необхідні параметри діаграми (підписи осей, сітка, ряди даних, легенда);

4) створити необхідний дизайн (обрати фон, заливку). Модифікація, форматування та редагування діаграми здійснюється за допомогою контекстного меню елементів, команди "Діаграма" або вкладки панелі інструментів з назвою "Конструктор".

Для візуалізації міграційних потоків оберемо лінійну діаграму (графік) (рис. 3.4), що дозволить представити графічно загальну тенденцію розвитку процесу з плином часу (за роками) та порівняти значення за досліджуваними категоріями.

Більш того, на одній лінійній діаграмі можна побудувати декілька кривих, які дозволять порівняти динаміку різних показників або того самого показника за різними регіонами, галузями та ін. Цей тип доцільно використовувати, якщо є динамічний ряд.

Міграційний рух населення

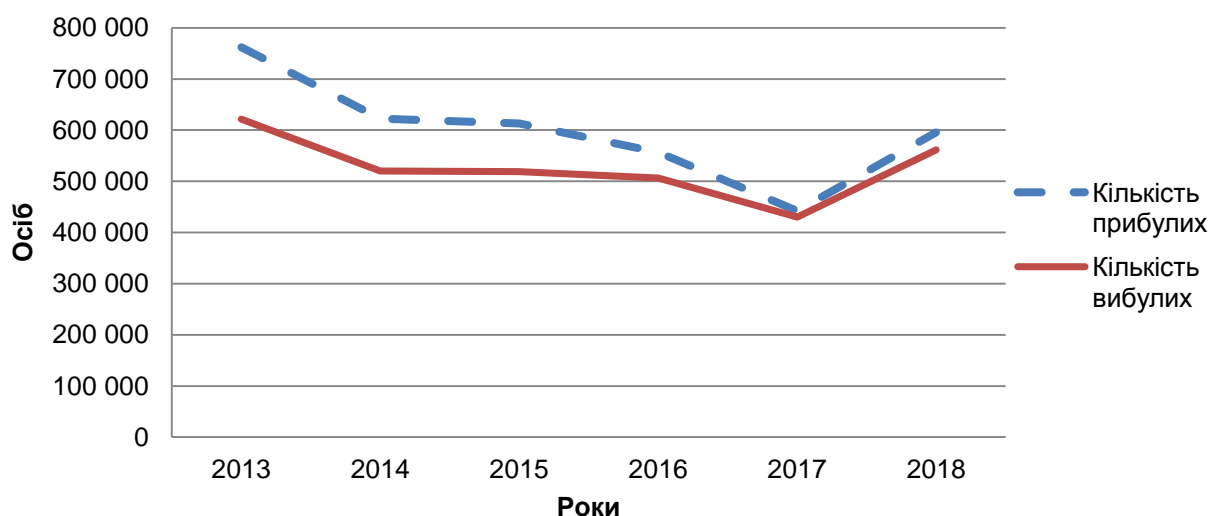


Рис. 3.4. Динаміка міграційного руху населення за період 2013 – 2018 рр.

Завдання 3.2. На основі даних, наведених у табл. 3.4, побудуйте комбіновану діаграму.

Таблиця 3.4

Середньомісячна вартість витрат на одне домогосподарство

Показники	Роки						
	2012	2013	2014	2015	2016	2017	2018
Середньомісячна вартість витрат на одне домогосподарство, грн, у тому числі:	3 073,3	3 458	3 592,1	3 820,3	4 048,9	4952	5 720,4
харчові та безалкогольні напої	1 585,8	1 773,9	1 799,6	1 913,9	2 101,4	2 629,5	2 848,8
алкогольні напої, тютюнові вироби	104,45	117,6	125,7	133,7	137,7	163,4	165,9
непродовольчі товари та послуги	1 072,6	1 224,1	1 336,3	1 398,2	1 469,8	1 807,5	2 316,8
інші витрати	310,4	342,3	330,5	374,4	340,1	351,6	388,9

Графічно подайте середньомісячну вартість витрат на одне домогосподарство. Для зображення загального розміру витрат однієї родини

за період 2012 – 2018 рр. побудуйте комбіновану діаграму, що скрадатиметься з стовпчастої гістограми та графіка з маркерами (для візуалізації видів витрат одного домогосподарства).

Розв'язання.

Для побудови комбінованої діаграми засобами MS Excel, використаємо описаний у завданні 1 алгоритм створення діаграми.

На першому етапі виділяємо весь масив вихідних даних та обираємо тип діаграми "Гістограма стовпчаста". Потім виділяємо кожен ряд даних, що містить інформацію про окремий вид витрат одного домогосподарства та через контекстне меню змінюємо тип діаграми для ряду даних на "Графік із маркерами" (рис. 3.5).

Загальні середньомісячні витрати на одне домогосподарство

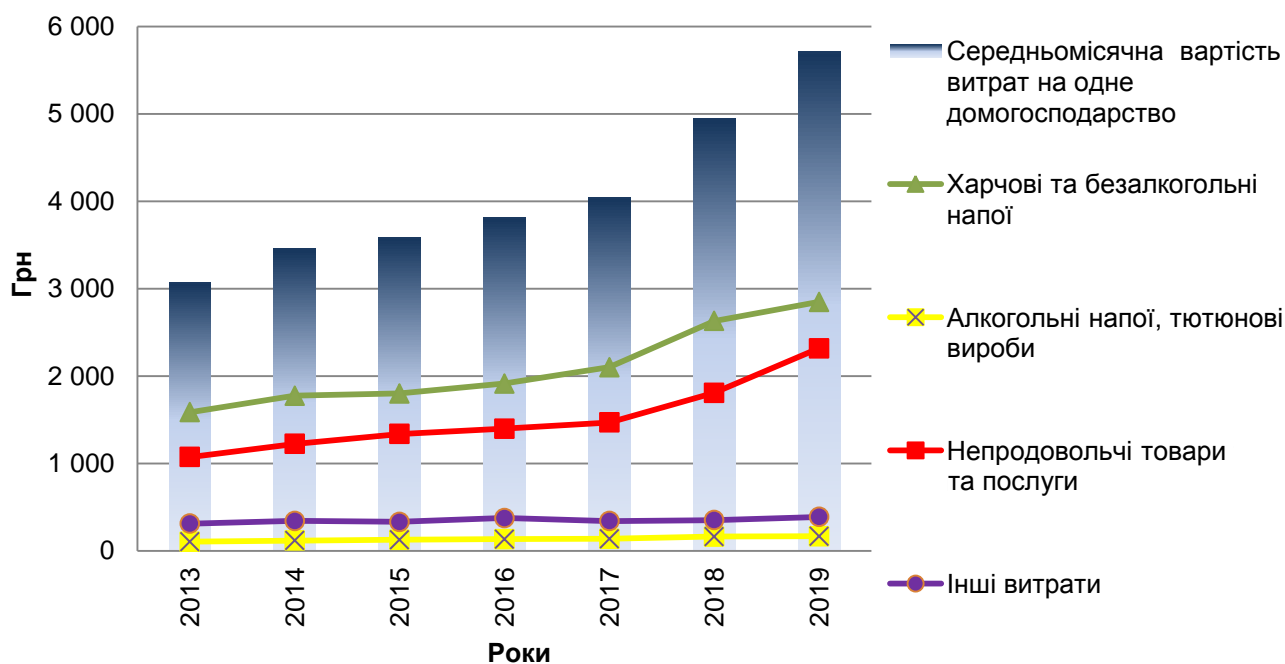


Рис. 3.5. Комбінована діаграма розподілу середньомісячних витрат на одне домогосподарство за період 2012 – 2018 рр.

Для редагування та форматування дизайну створеної комбінованої діаграми використовуємо вкладки "Конструктор", "Макет", "Формат" MS Excel.

Завдання 3.3. Є дані про обсяги доходів певного агропромислового підприємства, що спеціалізується на вирощуванні та реалізації сільськогосподарської продукції (табл. 3.5).

Обсяги доходу від реалізації сільськогосподарської продукції у 2018 р.

Види сільськогосподарської продукції	Обсяг отриманого від реалізації доходу, млн дол. США
Соняшникова олія наливом	1 067,5
Соняшникова олія бутильована	140,7
Фермерські продукти	381,3
Зерно, в тому числі:	877,14
ячмінь	129,26
пшениця	360,09
кукурудза	387,79

Необхідно побудувати вторинну кругову діаграму, що відобразить структуру доходів агропромислового підприємства.

Розв'язання.

Кругова діаграма показує як абсолютну величину, так і внесок (питому вагу) кожного елемента ряду даних у загальну суму. Кругова діаграма використовується для графічного зображення структури та пропорційності досліджуваного явища чи процесу. Однак на круговій діаграмі можна розмістити лише один ряд даних. Вторинна кругова діаграма застосовується для деталізації одного з секторів кругової діаграми. Такий спосіб візуалізації допомагає спростити та деталізувати перегляд невеликих областей (секторів) кругової діаграми.

У ході побудови вторинної кругової діаграми важливо дотримуватися таких правил:

1) головна та вторинна діаграми – частини одного ряду даних. Їх не можна редагувати незалежно один від одного;

2) сектори на вторинному колі теж показують частки, що і на звичайній діаграмі. Однак сума процентів не дорівнює 100, а становить загальну величину значення на секторі головної кругової діаграми (від якої відокремлена вторинна);

3) за замовчуванням на вторинному колі відображається остання третина даних. Якщо, наприклад, у вихідній таблиці 7 рядків (для діаграми – 7 секторів), то останні три значення будуть візуалізовані на вторинній (другорядній) діаграмі (рис. 3.6).

Структура доходів агропромислового підприємства у 2018 р.

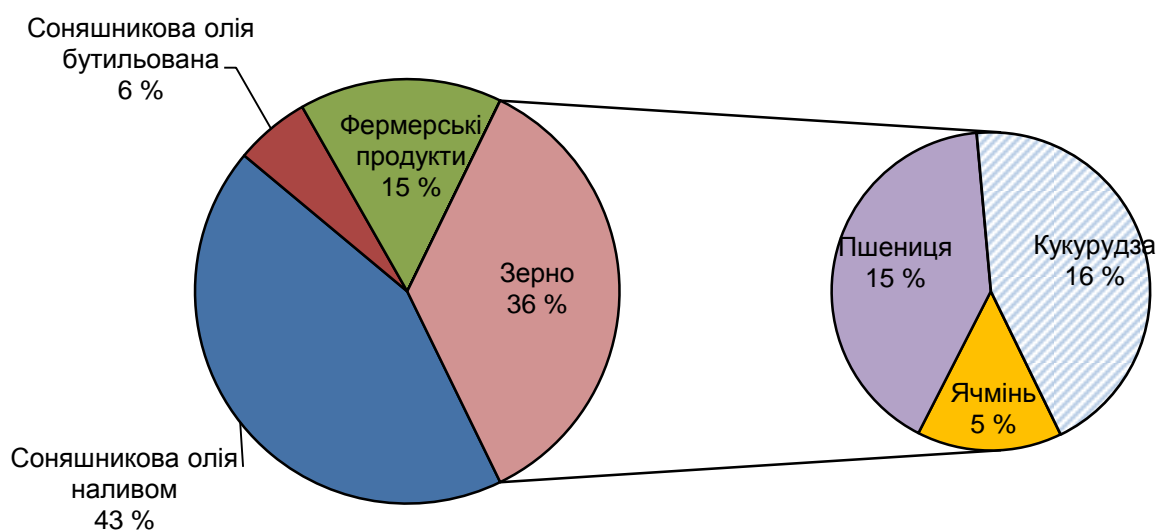


Рис. 3.6. Вторинна кругова діаграма структури доходів підприємства

Проаналізуйте отримані результати.

Завдання 3.4. На основі даних, наведених в табл. 3.6, побудуйте пелюсткову діаграму.

Таблиця 3.6

Обсяги продажів садового центру, тис. грн

Місяці	Вид товару			
	цибулини	зерно	квіти	дерева та кущі
Січень	–	140	–	–
Лютий	–	294	–	52
Березень	–	462	71	110
Квітень	–	313	84	195
Травень	–	178	286	189
Червень	–	52	381	90
Липень	–	–	412	22
Серпень	–	–	345	–
Вересень	280	–	157	121
Жовтень	420	–	87	302
Листопад	193	–	–	297
Грудень	–	–	–	178

Розв'язання.

Зобразимо сезонність продажів садового центру за допомогою пелюсткової діаграми (рис. 3.7). Пелюсткова діаграма виступає аналогом

графіка, зображеного в полярній системі координат, і відображає розподіл значень відносно початку координат. Осі спрямовані назовні від центру діаграми. Значення кожної точки зазначається на відповідній осі.

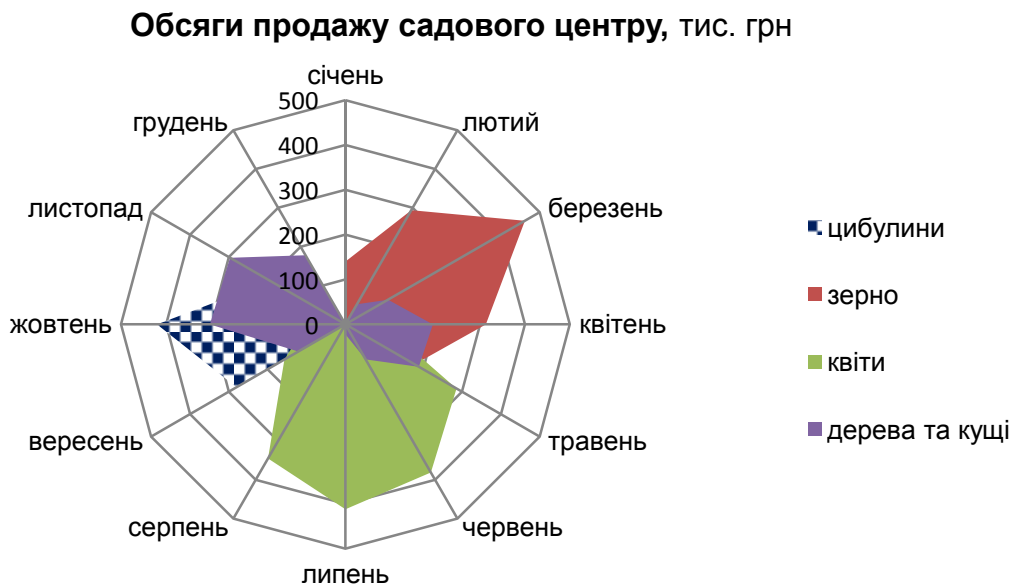


Рис. 3.7. Пелюсткова діаграма розподілу продажів садових культур

Пелюсткова діаграма застосовується, якщо категорії неможливо порівняти безпосередньо. Так, подана на рис. 3.7 пелюсткова діаграма в якості категорій містить 12 місяців року, а значеннями виступають сезонні продажі садових культур. Отже, за допомогою MS Excel було побудовано заповнену пелюсткову діаграму (див. рис. 3.7), яка відображає сезонний характер продажів садових культур.

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Візуалізація соціально-економічної інформації: побудова та аналіз різних видів діаграм" подані у роботі [67]. Лабораторна робота націлена на ознайомлення з типами графіків та отримання практичних навичок візуалізації та аналізу соціально-економічної інформації за допомогою графічного редактора пакету MS Excel.

Запитання для самодіагностики

1. Що називають статистичною таблицею? Які завдання ставляться перед статистичними таблицями?
2. Що таке макет статистичної таблиці? Назвіть його складові.

3. Що таке таблиця спряженості? Для чого її використовують?
4. За якими правилами будують статистичні таблиці?
5. Назвіть основні вимоги до побудови статистичних графіків.
6. Які основні елементи графіка?
7. Що таке діаграма? Які види діаграм вам відомі?
8. Як будують радіальну діаграму? Що зображують за допомогою радіальної діаграми?
9. Як будують квадратні та кругові діаграми?
10. Що таке картограми та картодіаграми? У чому їх відмінність?

Питання для критичного переосмислення (есе)

1. Розкрийте сутність графічного метода та його значення в статистичному аналізі даних.
2. З якою метою використовують лінійні, стовпчикові та секторні діаграми? Наведіть приклади їх практичного використання в економіці.
3. Які сучасні пакети прикладних програм, онлайн-редактори чи додатки для візуалізації статистичних даних вам відомі? Охарактеризуйте їх.
4. Надайте розгорнуту класифікацію видів діаграм відповідно до мети візуалізації статистичних даних і наявного вихідного інформаційного простору.
5. Зробіть порівняльний аналіз практичного застосування в економіці видів статистичних карт – картограм і картодіаграм.

4. Зведення і групування статистичних даних

Основні питання:

- 4.1. Сутність статистичного зведення та його види.
- 4.2. Статистичні групування та їх види.

4.1. Сутність статистичного зведення та його види

Статистичні дані, зібрані в процесі спостереження, не дозволяють отримати узагальнювальні характеристики досліджуваної сукупності, виявити закономірності її розвитку, тому що в процесі спостереження фіксуються характеристики тільки окремих одиниць сукупності. Для отримання узагальнювальних характеристик зібрану інформацію необхідно систематизувати, перетворити її в упорядковану систему статистичних

показників. Систематизація отриманої інформації та узагальнення спостережуваних факторів є *змістом другої стадії* статистичного дослідження, що називають *зведенням і групуванням* [1; 5; 13].

Статистичні дані є комплексом послідовних операцій з узагальнення конкретних поодиноких фактів, що утворюють сукупність, для виявлення типових рис і закономірностей, властивих досліджуваному явищу. Під **зведенням** розуміють сукупність прийомів наукового узагальнення і оброблення даних статистичного спостереження з метою отримання статистичних показників і подальшого їх аналізу. На основі цих показників надається характеристика чисельності сукупності, розміру притаманних їй ознак, структури сукупності та її якісний склад, встановлюються специфічні особливості та закономірності досліджуваного явища, взаємозв'язку між ознаками та ін. [28; 31; 43; 49].

Метою зведення є отримання підсумкових даних шляхом підрахунку одиничних відомостей.

Складовими статистичної зведення є [70; 78; 84; 85; 87]:

- 1) розроблення програми систематизації і групування даних;
 - 2) обґрунтування системи показників для характеристики груп і сукупності в цілому;
 - 3) проєктування макетів таблиць, у яких подають результати зведення;
 - 4) визначення технологічних схем оброблення інформації, програмного забезпечення;
 - 5) підготовка даних до оброблення на комп'ютері, формування автоматизованих банків даних;
 - 6) безпосереднє зведення, узагальнення, розрахунок показників.
- Види зведень наведено в табл. 4.1.

Таблиця 4.1

Класифікація статистичного зведення [93; 98 – 100; 103]

Класифікаційні ознаки	Вид зведень
За глибиною оброблення матеріалу	Прості; складні
За формою оброблення інформації	Централізовані; децентралізовані
За технікою виконання	Автоматизована; ручна

За **глибиною опрацювання матеріалу** розрізняють **прості та складні зведення** [14; 51; 70; 78; 84; 85; 87; 97].

Простим зведенням називають операцію з підрахунку загальних підсумків за сукупністю одиниць спостереження, тобто визначення розміру досліджуваного явища.

Складним зведенням є комплекс операцій, що включають групування одиниць спостереження, підрахунок підсумків за кожною групою і сукупності в цілому, а також подання результатів групування в табличній формі [14; 51; 70; 78; 84; 85; 87; 97].

За **формою оброблення матеріалу** зведення розподіляють на централізовані та децентралізовані.

За **централізованого зведення** весь первинний матеріал надходить в одну організацію, де і піддається обробленню за прийнятою програмою, за єдиною методикою (наприклад, у Державному комітеті зі статистики України або територіальних управліннях статистики).

За **децентралізованого зведення** розроблення статистичного матеріалу здійснюється за ієрархічною системою управління, піддаючи відповідному обробленню на кожному рівні. Наприклад, підприємства здають звіти до районних відділів статистики, які складають зведення щодо свого району та відправляють узагальнену інформацію до регіональних управлінь або комітетів. Ті свої зведення надсилають до Державного комітету зі статистики України, де і визначаються показники в цілому щодо народного господарства країни.

За **ступенем автоматизації оброблення даних** відомості буває **автоматизована**, яка проводиться з використанням ПК, і **ручна** [14; 51; 70; 78; 84; 85; 87; 97].

Організація статистичного зведення здійснюється за спеціальною програмою, яка повинна складатися одночасно з розроблення *плану та програми проведення спостереження*.

Програма статистичного зведення містить [1; 5; 13; 23; 28; 31; 43; 49; 51]:

- вибір ознак групування;
- визначення порядку формування груп;
- розроблення системи статистичних показників для характеристики виділених груп і сукупності в цілому;
- розроблення макетів таблиць для подання результатів зведення.

План статистичного зведення містить вказівки щодо термінів і послідовності виконання окремих етапів зведення, її виконавців, порядку подання його результатів [1; 5; 13; 23; 28; 31; 43; 49; 51].

На першому етапі статистичного дослідження отримують статистичну інформацію, яку потім систематизують, зводять, обробляють, аналізують та узагальнюють.

На другому етапі статистичного дослідження здійснюється розподіл сукупності на групи, однорідні в тому чи іншому розумінні. Для цього використовують найважливіші положення такого розподілу: в одну групу об'єднують елементи сукупностей, у певній мірі подібні між собою; міра подібності між елементами одної групи значно вища, ніж між елементами, що належать до різних груп. У кожному конкретному статистичному дослідженні вирішують такі завдання: що взяти за основу групування; скільки груп необхідно сформуувати; як розмежувати групи.

Основою розмежування елементів сукупності на групи може бути будь-яка ознака (атрибутивна чи кількісна), що має якісно відмінну характеристику. Таку ознаку називають **групувальною**. Залежно від складності явища та мети дослідження групувальних ознак може бути декілька [1; 5; 13; 23; 28; 31; 43; 49; 51].

Якщо поділ елементів сукупності на групи здійснюється за атрибутивними ознаками, то такий вид групування називають **класифікацією**, або **номенклатурою**. Вони розробляються міжнародними та національними статистичними органами та рекомендуються як статистичний стандарт.

Класифікація у статистиці – це систематизований розподіл явищ та об'єктів на певні групи, класи, розряди на підставі їх подібності або відмінності. Різновидом класифікацій є товарні номенклатури як стандартизований перелік об'єктів і груп. Види статистичних класифікацій подані на рис. 4.1.

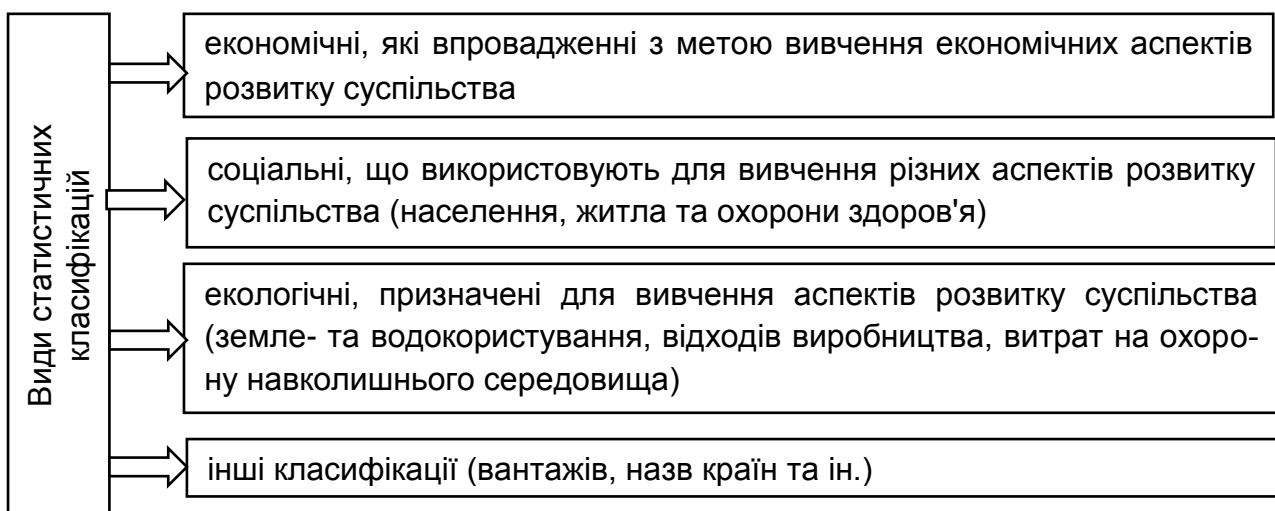


Рис. 4.1. Види статистичних класифікацій [43; 49; 51]

Прикладами сучасних класифікацій національного рівня є такі, що повністю узгоджені з міжнародними стандартами [93; 97 – 99; 103]:

"Класифікація видів економічної діяльності" (КВЕД) – в якості ознаки класифікації прийнята одна з трьох ознак: призначення виробленої продукції; єдність технології виробництва; однорідність використаної сировини;

"Класифікація форм власності" (КФВ) – в якості об'єктів класифікації встановлюються форми власності за чинним законодавством України (державна, колективна, приватна власність та ін.);

"Українська класифікація товарів зовнішньоекономічної діяльності" (УКТ ЗЕД) – відповідає потребам статистичних служб, митних органів зовнішньоекономічної діяльності;

"Класифікація організаційно-правових форм господарювання" (ПФГ) – здійснена класифікація суб'єктів підприємницької діяльності (державне, колективне, приватне підприємство та ін.), організацій, що провадять підприємницьку діяльність (заклад, установа тощо), відокремлених підрозділів (філія, представництво).

Спільним у національних і міжнародних класифікаціях є те, що варіація їх ознак фіксується у певному системному вигляді з використанням кодів (шифрів) позицій, що класифікуються.

Для класифікації товарів і послуг, для характеристики ознак видів економічної діяльності сьогодні застосовують такі класифікаційні системи (номенклатури):

Міжнародна стандартна галузева класифікація всіх видів економічної діяльності (МСГК);

Стандартна міжнародна торгова класифікація ООН (СМТК);

Гармонізована система опису і кодування товарів (ГС);

Класифікатор основних продуктів (КОП).

4.2. Статистичні групування та їх види

Науковою основою зведення є **статистичне групування** як процес утворення однорідних груп на основі розчленування (розподілу) статистичної сукупності на частини або об'єднання досліджуваних статистичних одиниць у частковій сукупності за істотними для них ознаками [14; 78; 84; 85; 87; 93; 97 – 100; 103].

Метод групування є основою для застосування інших методів статистичного аналізу з метою виділення основних аспектів і характерних особливостей явищ, що досліджуються. За своєю роллю у процесі дослідження метод групування виконує функції, аналогічні функціям експерименту в природничих науках: за допомогою групування за окремими ознаками та комбінації самих ознак статистика допомагає виявити закономірності та взаємозв'язки явищ в умовах, що у певній мірі нею визначаються. З використанням такого методу можна простежити взаємовідносини різних факторів і визначити силу їх впливу на результативні показники.

Статистичні групування здійснюють у кілька послідовних **етапів** [84; 85; 87; 93; 97 – 100; 103]:

I етап – теоретичний аналіз явища або процесу, що досліджуються;

II етап – вибір групувальної ознаки (ознак);

III етап – визначення кількості груп і величини інтервалу; побудова інтервального ряду розподілу одиниць сукупності за групувальною ознакою (ознаками), що досліджуються;

IV етап – визначення та обґрунтування системи статистичних показників для виділення і характеристики типових груп; складання макетів таблиць;

V етап – обчислення абсолютних, відносних і середніх показників;

VI етап – табличне та графічне оформлення результатів групування;

VII етап – аналіз отриманих результатів; формулювання висновків і пропозицій.

Залежно від ступеня складності досліджуваного явища та поставлених завдань статистичні групування можуть виконуватися за одним або кількома **ознаками групування**. Групування називають **простим** (одновимірним), якщо однорідні групи формуються за однією ознакою одночасно. Якщо однорідні групи утворюються за двома та більше ознаками, то угруповання є **складними** (рис. 4.2). Типи простих (одновимірних) групувань докладно розглянуті в роботах [1; 5; 13; 14; 23; 28; 31; 78; 84; 85; 97; 100].

Структурні групування, які використовуються для вивчення внутрішньої будови статистичної сукупності і характеристики структурних зрушень. Вони дають інформацію про поточний стан масових явищ і застосовуються з метою оперативного управління.

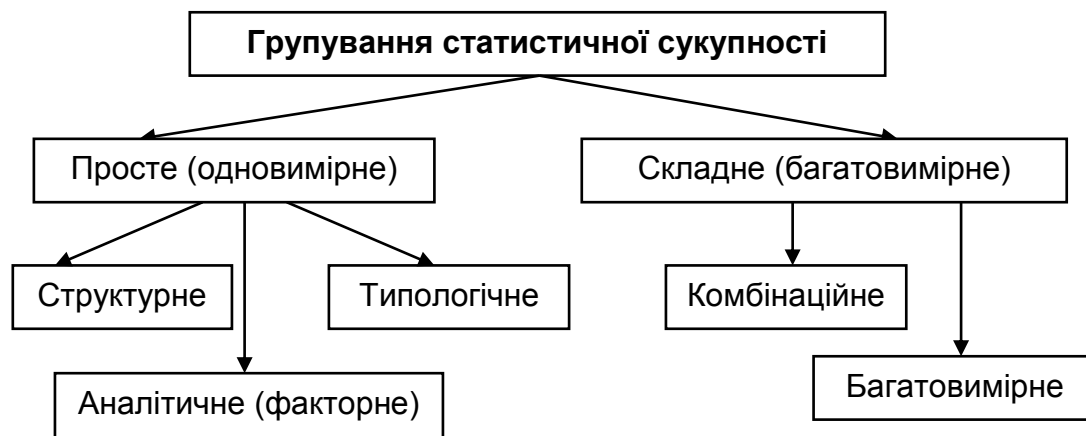


Рис. 4.2. Види групувань

Структурне групування виконується в кілька *етапів*:

- вибір ознаки групування;
- визначення необхідного кількості груп;
- визначення параметрів груп;
- розподіл одиниць спостереження за виділеними групами;
- розрахунок структурних характеристик;
- формулювання висновків [1; 5; 13; 14; 23; 28; 31; 78; 84; 85; 97; 100].

Метою **типологічного групування** є вивчення поширеності різних типів економічних явищ у статистичній сукупності. Типологічні групування застосовують, як правило, до неоднорідної сукупності та здійснюють за допомогою складних різноінтервальних групувань. Результатом типологічних групувань є поділ сукупності на класи, соціально-економічні типи, однорідні групи одиниць [1; 5; 13; 14; 23; 28; 31; 78; 84; 85; 97; 100].

За своєю сутністю типологічне групування – це групування-класифікатор. Такі групування часто ґрунтуються на стійкому переліку груп, які не змінюються або змінюються незначно в часі. Прикладом може слугувати групування підприємств за формою власності (державна, муніципальна, приватна, змішана) або секторами економіки.

У ході виконання типологічних групувань важливо правильно вибрати підставу для групування. Для цього необхідно попередньо виявити можливий тип явища на основі аналізу сутності та закономірностей його розвитку. Число груп і їх параметри встановлюють неформально на основі виділених якісних закономірностей, часто з залученням кількісних ознак. Наприклад, для групування населення за віковим складом виділяють чотири вікові групи: 1) дошкільний вік – до 7 років; 2) шкільний – з 7 до 17 років; 3) робочий – з 17 до 55 (60) років; 4) пенсійний – з 55 (60) років [93; 97 – 99; 103].

За технікою виконання типологічне групування схоже зі структурним, за винятком перших етапів – ознака групування, кількість груп, їх параметри визначаються на основі якісного аналізу. У таких групуваннях дуже часто застосовують спеціалізовані інтервали. Типологічні групування оформлюють у табличній формі, об'єктом аналізу в них є показники структури.

Аналітичні групування призначені для виявлення зв'язку між ознаками, що досліджуються. Вони дозволяють виявити наявність і напрямок зв'язку, а також виміряти їх тісноту та силу. Усі досліджувані ознаки в цьому випадку розподіляють на дві групи: факторні; результативні. Взаємозв'язок між ними виявляється в тому, що зі зміною середнього значення факторної ознаки систематично змінюється середнє значення результативної ознаки [1; 5; 13; 14; 23; 28; 31; 78; 84; 85; 97; 100].

Аналітичні групування відрізняються від структурних і типологічних за технікою виконання, яка полягає в тому, що:

проводиться групування одиниць сукупності за факторною ознакою, вона виконується як структурна;

у кожній виділеній групі відбираються відповідні значення результативної ознаки, і на їх основі розраховується деякий узагальнювальний показник – середнє значення;

аналізуються зміни узагальнювального показника – середнього значення результативної ознаки за групами; формується висновок про наявність або відсутність взаємозв'язку і його напрям. Якщо зі зміною значень факторної ознаки, покладених в основу групування, змінюється величина результативних, то визнається наявність зв'язку між ознаками. Якщо зі збільшенням значень факторної ознаки збільшується значення результативної, то зв'язок належить до прямого; в іншому випадку – до зворотного.

За кількістю групувальних ознак, покладених в основу **складного (багатовимірного)** групування, розрізняють прості та комбінаційні групування. **Простим** називають групування, яке проводиться за однією ознакою. **Комбінаційні групування** виконують за кількома ознаками послідовно. Послідовність установлюється виходячи з логіки взаємозв'язку показників. Як правило, групування починають з атрибутивної ознаки. У комбінаційному групуванні сукупність логічно послідовно розбивається на однорідні частини за окремими ознаками: на групи – за однією ознакою, потім всередині кожної групи за другою ознакою – на підгрупи і т. д. [1; 5; 13; 14; 23; 28; 31; 78; 84; 85; 97; 100].

Такі групування призначені для більш глибокого аналізу досліджуваного явища. Вони дозволяють виявити та порівняти відмінності та зв'язки між аналізованими ознаками, які неможливо за кожною з них встановити на основі ізольованих груп. Проте слід мати на увазі, що у вивченні впливу великої кількості ознак застосування комбінаційного групування неможливе: це призводить до дроблення інформації, а отже, до затушовування проявів закономірності. Навіть за наявності великих обсягів інформації доводиться обмежуватися двома-чотирма ознаками.

Комбінаційне групування за двома ознаками (X , Y) оформляється у вигляді шахової таблиці, в якій значення однієї ознаки X відкладаються на рядках, а значення другої ознаки Y – на стовпцях. На перетині j -го стовпця та i -го рядка матриці знаходяться частоти спільного прояву значення ознаки Y в j -му стовпці та значення ознаки X в i -му рядку [1; 5; 13; 14; 23; 28; 31; 78; 84; 85; 97; 100].

Багатовимірні групування. До них належать групування, виконані за кількома ознаками групування одночасно. Метою багатовимірних групувань є класифікація даних на основі безлічі ознак, тобто виділення груп статистичних одиниць, однорідних за кількома ознаками одночасно. У процесі такого групування вирішуються, наприклад, завдання типізації – виділяються самостійні економічні чи соціальні типи явищ. Так, прийомами багатовимірної класифікації можна всю сукупність промислових підприємств розбити на "дрібні", "середні" та "великі", використовуючи такі ознаки: чисельність промислово-виробничого персоналу, обсяг продукції, споживання матеріальних ресурсів та інше. Можна виділити типи підприємств за фінансовим станом на основі таких показників, як: розмір прибутку, рівень рентабельності виробництва, рівень капіталізації, рівень ліквідності цінних паперів тощо. У психології багатовимірні групування використовують для виділення типів людей за ступенем їх професійної придатності, в медицині – для діагностування хвороб на основі безлічі симптомів [78; 84; 85; 97; 100].

Для виконання багатовимірних групувань можна використовувати **два основних підходи** [1; 5; 14; 49; 51; 70; 78; 84; 85; 87; 93; 98 – 100; 103]:

- перший полягає в тому, що розраховується узагальнювальний показник за сукупністю ознак групування і проводиться просте групування за ним;
- другий підхід полягає у використанні метода кластерного аналізу.

Представником першого підходу є **метод багатовимірної середньої**. Алгоритм групування за цим методом передбачає виконання послідовних етапів [1; 5; 14; 49; 51; 70; 78; 84; 85; 87; 93; 98 – 100; 103].

I етап – складається матриця абсолютних значень ознак за всіма статистичними одиницями: x_{ij} , $i = n, 1$ – статистичні одиниці, $j = k, 1$ – ознаки.

II етап – абсолютні значення ознак замінюються рівнями, нормованими за середнім значенням:

$$P_{ij} = \frac{x_{ij}}{\bar{x}_j}, \quad (4.1)$$

де P_{ij} – нормоване значення j -ї ознаки у i -ї статистичної одиниці;

x_{ij} – статистична одиниця;

$$\bar{x}_j \text{ – середнє значення } j\text{-ї ознаки, } \bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}.$$

III етап – для кожної статистичної одиниці розраховується багатовимірна середня:

$$\bar{P}_i = \frac{\sum_{j=1}^k P_{ij}}{k}, \quad (4.2)$$

де k – число підстав групування.

IV етап – відповідно до значень багатовимірної середньої сукупність розділяється на однорідні групи, тобто виконується просте групування за багатовимірною середньою.

Поряд із первинним групуванням, від якого розглянуто за етапами, у статистиці застосовують вторинне, яке проводять на основі раніше здійсненого. Таке групування використовують для покращення характеристики досліджуваного явища, якщо первинне групування не дає чіткої характеристики розподілу одиниць сукупності.

За допомогою групувань у статистиці вирішують **завдання** [51; 70]:

- вивчення складу статистичних сукупностей;
- виділення окремих типів явищ усередині сукупності;

- виявлення причинно-наслідкових зв'язків між різними ознаками всередині сукупності;
- класифікація одиниць сукупності за безліччю ознак.

Основними категоріями методу групувань є ознака групування (підстава для групування) та інтервал. Принципове значення для побудови групувань має вибір групувальної ознаки, на основі якої виділяють різні типи, групи та підгрупи. За групувальні приймають найістотніші ознаки. Групувальною може бути атрибутивна (якісна) або кількісна ознака.

Ознакою групування (підставою для групування) називають ознаку, за якою відбувається виділення однорідних груп. Вибір ознаки групування здійснюється відповідно до цілей статистичного дослідження. Як ознака групування зазвичай виступає істотна ознака. Обов'язковою умовою виконання будь-якого групування є впорядкування статистичної сукупності за значеннями ознаки групування.

Вибір ознаки групування завжди засновують на аналізі якісної природи досліджуваного явища. Групувальною зазвичай обирається одна з істотних, легко розпізнаваних ознак, що мають як атрибутивний (якісний), так і кількісний характер.

Для *атрибутивної* ознаки число груп відповідає числу різновидів ознаки. Так, для розподілу підприємств за формами власності виділяють такі форми: державна, комунальна, приватна, колективна, міжнародних організацій та юридичних осіб інших держав.

У групуванні за *кількісною (варіаційною)* ознакою виникає питання стосовно кількості груп та інтервалів групування.

Інтервал – це сукупність значень ознаки, що варіюють у групі. Він визначає кількісні межі груп, а його ширина є проміжком між максимальним і мінімальним значеннями ознаки в групі.

Для виконання групування використовують такі **типи інтервалів** [28; 32]:

- **рівні** – в усіх виділених групах ширина інтервалу є однаковою;
- **нерівні** – в кожній групі ширина інтервалу різна; її зміна може змінюватися закономірно (наприклад, рівномірно зростати) або довільно, тобто бути вільною;
- **закриті** – якщо відомі верхня і нижня межі інтервалів (максимальне та мінімальне значення ознаки в групах);
- **відкриті** – якщо відома тільки одна межа інтервалу, верхня або нижня.

У процесі групування необхідно виходити з принципів: відмінності між одиницями спостереження, віднесеними до однієї групи, повинні бути менше, ніж між одиницями, віднесеними до різних груп. Визначення необхідного числа груп відбувається за тим правилом, що число груп повинно бути достатнім для об'єктивного подання досліджуваної сукупності. За великої кількості груп відмінності між ними стають малопомітними, а в самих групах з причини їх малої наповнюваності перестає діяти закон великих чисел і можливі прояви випадковості. За незалежної кількості до однієї групи можуть потрапити статистичні одиниці з ознаками, які істотно розрізняються за значеннями.

На кількість виділених груп впливають такі **фактори** [51; 70]:

- *рівень коливання ознаки групування* – чим значніше варіація ознаки, тим більшу кількість груп необхідно виділяти за інших рівних умов;
- *розмір досліджуваної статистичної сукупності* – чим більше розмір досліджуваної сукупності, тим більшу кількість груп необхідно виділяти.

Виділені групи повинні бути достатньо заповненими. Наявність порожніх груп чи мала кількість статистичних одиниць у них свідчать про неправильне визначення їх кількості. Орієнтовно кількість груп можна визначити, використовуючи емпіричну залежність – формулу Стерджеса [13; 14; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 93; 97 – 100; 103]:

$$m = 1 + 3,322 \times \lg N, \quad (4.3)$$

де m – кількість груп;

N – чисельність одиниць статистичної сукупності.

Залежність Стерджеса дає добрі результати, якщо сукупність складається з великої кількості одиниць, розподіл близький до нормального та застосовані рівні інтервали. У практичних розрахунках можна використовувати співвідношення, отримані на підставі формули Стерджеса (табл. 4.2).

Таблиця 4.2

Співвідношення, отримані на підставі формули Стерджеса [51; 70]

N	15 – 24	25 – 44	45 – 89	90 – 179	180 – 359	360 та більше
m	5	6	7	8	9	10

Існує ще один спосіб визначення кількості виділених груп. Він пов'язаний із застосуванням середньоквадратичного відхилення рівними і нерівними σ : якщо ширина інтервалу дорівнює $0,5\sigma$, то виділяється 12 груп, якщо $2/3\sigma$ – то 9 груп, якщо σ – то 6 груп.

Визначення параметрів групи. У кожній виділеній групі розраховуються такі параметри [1; 5; 13; 23; 43; 49; 51; 70; 78; 84; 85; 87; 93; 98 – 100; 103]:

- верхня межа інтервалу x^B і нижня межа інтервалу x^H ;
- ширина інтервалу h ;
- середина інтервалу x_i .

Нижня межа інтервалу x^H є найменшим значенням ознаки в групі.

Верхня межа інтервалу x^B – найбільше значення ознаки в групі.

Інтервали групування бувають **рівними** та **нерівними** (прогресивно зростаючі, прогресивно спадні, довільні, спеціалізовані). Якщо варіація ознаки проявляється у порівняно вузьких межах, і розподіл статистичних одиниць має достатньо рівномірний характер, то будують групування з рівними інтервалами [70; 78; 84; 85; 87; 93; 98 – 100; 103]:

$$h = \frac{X_{\max} - X_{\min}}{m}. \quad (4.4)$$

На основі розрахованої ширини інтервалу h послідовно визначаються межі інтервалів. Визначення меж починається з першої групи. Нижня межа інтервалу приймається дорівненою мінімальному значенні ознаки в сукупності, $x_1^H = x_{\min}$. Верхня межа першого інтервалу розраховується $x_1^B = x_1^H + h$. Надалі розрахунок інтервалів проводиться за формулами [70; 85; 87; 93; 98 – 100; 103]:

$$x_i^H = x_{i-1}^B; x_i^B = x_i^H + h. \quad (4.5)$$

Середина інтервалу (центральна варіанта) визначається як напівсума верхньої та нижньої границь інтервалу. Параметр середини інтервалу використовується для розрахунку узагальнювальних характеристик досліджуваної сукупності.

Розподіл одиниць сукупності за групами. Основним завданням цього етапу є підрахунок числа одиниць, що потрапили в кожну з виділених груп n_i . У розподілі одиниць спостереження за виділеними групами (особливо якщо ознака групування є безперервна) має місце невизначеність: до якої групи відносити одиниці зі значеннями ознаки, що збігаються з межами інтервалів? Для усунення невизначеності використовують принцип однаковості. Такі одиниці включають в групу, в якій нижня межа збігається зі значенням ознаки. Наприклад, є групи підприємств за обсягом виробництва, млн грн: 400 – 450; 450 – 500; 500 – 550; 550 – 600; 600 – 650. До якої групи слід віднести підприємства з обсягом виробництва у 500 млн грн? Відповідно до принципу однаковості – до другої групи.

Розрахунок структурних характеристик. Розрахунок полягає у визначенні для кожної групи питомої ваги (частки) її одиниць у загальному обсязі статистичної сукупності. Як і будь-яка відносна величина, цей показник може бути визначений у вигляді коефіцієнтів [93; 99; 100; 103]:

$$d_i = \frac{n_i}{N} \text{ або відсотків } d_i = \frac{n_i}{N} \times 100 \%. \quad (4.6)$$

Розрахувавши такі частки для всіх груп, отримуємо структуру досліджуваної статистичної сукупності, яка дорівнює повному набору часток: $\sum d_i = 1$.

Формулювання висновків про склад сукупності. Для структурних групувань у висновках зберігаються два положення:

- які значення ознаки зустрічаються в сукупності найбільш часто, які – найбільш рідко;
- який характер зміни структури залежно від зміни значення ознаки. Зі збільшенням x частка може збільшуватися або зменшуватися. Це досить типово для економічних показників.

Висновки обов'язкові, інакше нівелюється сенс групування. Дані структурних групувань зазвичай подають у формі відповідної таблиці.

Приклад 4.1. Виконання структурного групування: потрібно зробити групування з рівними інтервалами за даними про стаж роботи співробітників компанії "ABC". У головному офісі працює 50 співробітників, зі стажем роботи від 0 до 26 років.

За формулою Стерджеса розрахуємо оптимальну кількість груп [70; 78; 84; 85; 87; 93; 98 – 100; 103]:

$$m = 1 + 3,322 \times \lg 50 \approx 7.$$

Далі розраховується постійна ширина інтервалів [70; 78; 84; 85; 87; 93; 98 – 100; 103]:

$$h = \frac{x_{\max} - x_{\min}}{m} = \frac{26 - 0}{7} = 4.$$

Структурне групування виконано за параметрами, що були розраховані за табл. 4.3.

Таблиця 4.3

Групування співробітників компанії "АВС" за стажем роботи
[70; 78; 84; 85; 87; 93; 98 – 100; 103]

№ групи	Параметри груп, роки				Кількість співробітників у групі	Питома вага співробітників у групах, %
	Межі інтервалів		Ширина інтервалу	Середина інтервалу		
	нижня	верхня				
1	0	4	4	2	6	12
2	4	8	4	6	8	16
3	8	12	4	10	11	22
4	12	16	4	14	13	26
5	16	20	4	18	6	12
6	20	24	4	22	4	8
7	24	28	4	26	2	4
Усього	0	28	28	14	50	100

За проведеним групуванням можна зробити такі висновки:

найбільша кількість співробітників компанії у головному офісі має стаж від 12 до 16 років, вони становлять 26 % від загальної чисельності співробітників компанії; найменша – зі стажем від 24 до 28 років, їх частка складає 4 %;

із збільшенням стажу роботи число співробітників компанії спочатку зростає, досягає максимуму у 4-й групі, а потім знижується.

Приклад типологічного групування наведено в табл. 4.4.

Таблиця 4.4

**Групування населення територіальної громади
за віковими категоріями**

Вікові категорії населення	Границі інтервалів		Чисельність у групі, осіб	Показник структури, %
	верхня межа	нижня межа		
Дошкільний вік	до	7	192	14,3
Шкільний вік	7	17	218	16,3
Робочий вік	17	55 (60)	574	42,8
Пенсійний вік	55 (60)	і більше (∞)	357	26,6
Разом			1 340	100,0

Приклад 4.2. Виконання *аналітичного* групування: необхідно встановити залежність між стажем роботи та розміром заробітної плати робітників ділянки. Щодо кожного робітника відомі стаж роботи та місячна заробітна плата. Тут факторною ознакою є стаж роботи, результативною – розмір заробітної плати. За факторною ознакою раніше було проведене структурне групування. Співробітники компанії були розбиті на 7 однорідних груп. Додатково в кожній групі була розрахована сумарна заробітна плата співробітників за місяць і середня її величина за формулою середньої арифметичної простої [1; 5; 13; 14; 70; 78; 84; 85; 87; 93; 97 – 100; 103]:

$$\bar{y}_i = \frac{\sum_{i=1}^n y_i}{n}, \quad (4.7)$$

де y_i – заробітна плата i -го співробітника.

Результати розрахунків приведені в табл. 4.5. Об'єктом аналізу в табл. 4.5 є середнє значення результативної ознаки – середньомісячна заробітна плата співробітників у групах і середина інтервалу – середній стаж роботи. Якщо середнє значення результативної ознаки, встановлене за групами, має певну різницю, то зв'язок між ознаками можна вважати встановленим. Якщо середній результат з переходом від однієї групи до іншої практично не змінюється, то зв'язок між ознаками відсутній.

**Дослідження залежності заробітної плати співробітників
компанії від стажу роботи**

[1; 5; 13; 23; 28; 31; 43; 49; 51; 70; 84; 85; 87; 93; 98 – 100; 103]

№ групи	Групи співробітників компанії за стажем, років			Число співробітників компанії у групі, осіб	Загальна зарплата співробітників компанії у групі, грн	Середня зарплата групи, грн
	Межі інтервалів		Середина інтервалу			
	нижня	верхня				
1	0	4	2	6	18 000	3 000
2	4	8	6	8	28 000	3 500
3	8	12	10	11	41 800	3 800
4	12	16	14	13	59 800	4 600
5	16	20	18	6	34 800	5 800
6	20	24	22	4	27 200	6 800
7	24	28	26	2	14 800	7 400
Усього	0	28	14	50	224 400	4 488

У розглянутому прикладі зміна стажу роботи приводить до зміни заробітної плати. Отже, за допомогою аналітичного групування можна встановити наявність зв'язку між ознаками, але не описати його. Для опису необхідно використовувати апарат *кореляційно-регресійного аналізу*.

Приклад 4.3. Виконання групування на основі *багатовимірної середньої*: необхідно виділити однорідні групи статистичних одиниць за трьома ознаками. Обсяг статистичної сукупності складає 10 об'єктів, кожен з яких характеризується умовними значеннями ознак. Вихідні дані та розрахунок багатовимірної середньої подані в табл. 4.6.

Отже, для кожного з 10 об'єктів розрахована умовна характеристика – багатовимірна середня, що замінює три первинних ознаки. За багатовимірною середньою як ознакою групування необхідно виконати одновимірне структурне групування.

Для цього визначається:

кількість однорідних груп $m = 1 + 3,322 \lg 10 \approx 4$;

ширина інтервалів $h = \frac{x_{\max} - x_{\min}}{m} = \frac{1,7 - 0,3}{4} = 0,33$.

Таблиця 4.6

Розрахунок багатовимірної середньої [93; 98; 103]

Номер об'єкта	Абсолютні значення ознак			Нормовані значення ознак			Розрахунок багатовимірної середньої	
	X ₁₁	X ₁₂	X ₁₃	P ₁₁	P ₁₂	P ₁₃	$\sum_{j=1}^k P_{ij}$	\bar{P}_i
1	2	18	62	0,57	1,67	1,0	3,24	1,08
2	1	5	40	0,29	0,46	0,64	1,39	0,46
3	2	7	40	0,57	0,65	0,64	1,86	0,62
4	6	15	77	1,71	1,39	1,24	4,34	1,4
5	1	9	43	0,29	0,83	0,69	1,81	0,60
6	6	20	95	1,71	1,85	1,53	5,1	1,7
7	5	9	62	1,42	0,83	1,0	3,25	1,08
8	1	1	46	0,29	0,10	0,74	1,13	10,38
9	8	15	84	2,29	1,39	1,35	5,03	1,68
10	3	9	72	0,86	0,83	1,16	2,85	2,95
Усього	35	108	621	10,00	10,00	10,00	10,00	10,00
\bar{X}_j	3,5	10,8	62,1	1,0	1,0	1,0	1,0	1,0

Структурне групування подане в табл. 4.7.

Таблиця 4.7

Групування об'єктів за багатовимірною середньою [93; 98]

№ групи	Параметри групи			Номер об'єкта	Кількість об'єктів у групі, n _i
	нижня межа	верхня межа	ширина інтервалу		
1	0,38	0,71	0,33	2, 3, 5, 8	4
2	0,71	1,04	0,33	10	1
3	1,04	1,37	0,33	1,7	2
4	1,37	1,7	0,33	4, 6, 9	3
Усього	0,38	1,7	1,32	–	10

У виконанні багатовимірних групувань кожна одиниця сукупності, що має набір з k ознак, розглядається як крапка в k -вимірному просторі – просторі ознак; кожній ознаці надається сенс координати. Задача класифікації в цьому випадку зводиться до виділення згущення об'єктів у цьому просторі. Для цього використовують різні алгоритми, але завжди однорідні групи виокремлюють на підставі наближеності об'єктів за сукупністю ознак. Мірою наближеності об'єктів, тобто мірою подібності одиниць сукупності, можуть бути різні критерії.

Виділяють **три типи** заходів подібності [1; 5; 13; 14; 70; 78; 84; 85; 87]:

- коефіцієнти подібності;
- коефіцієнти зв'язку (кореляції);
- показники відстані.

Коефіцієнти подібності використовують для вимірювання ступеня наближеності між парою об'єктів, кожна з ознак яких приймає значення 0 або 1.

Коефіцієнти кореляції використовують як вимірювачі сили зв'язку між статистичними одиницями або між ознаками. Для вимірювання тісноти зв'язку кількісних ознак застосовують коефіцієнти лінійної кореляції.

У кластерному аналізі мірою подібності є *міра (показник) відстані* між двома об'єктами i та j . Для кількісних ознак використовується евклідова відстань.

Іноді залежно від цілей економічних досліджень необхідно перегрупувати дані, щоб забезпечити порівнянність структур двох сукупностей з однією ознакою. Результат перегрупування називають **вторинним групуванням**. Перегрупування виконується або об'єднанням, або розбиванням інтервалів первинного групування.

Якщо межі інтервалів первинних і вторинних груп збігаються, то частоти (частки) об'єднаних інтервалів просто підсумовуються. Коли виконують розбиття інтервалу первинного групування, частоти розподіляють між новоствореними групами пропорційно до співвідношення частин довжини вихідного інтервалу.

Передбачається, що всередині інтервалу розподіл рівномірний. Техніку перегрупування даних слід розглянути на прикладі розподілу працівників за розміром середньомісячної заробітної плати у двох галузях промисловості (табл. 4.8).

Розподіл працюючих за рівнем середньорічної заробітної плати

[1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 93; 98 – 100]

Галузь А		Галузь Б	
Зарплата, грн	Частка працівників, %	Зарплата, грн	Частка працівників, %
До 160	15	До 160	12
160 – 180	20	160 – 190	30
180 – 200	26	190 – 220	21
200 – 220	23	220 – 250	18
220 – 240	9	250 – 280	13
240 і більше	7	280 і більше	6
Разом	100	Разом	100

Результати первинного групування безпосередньо порівняти не можна, оскільки інтервали групування різні: в галузі А ширина інтервалу 20, в галузі Б – 30 грн. Перегрупуємо дані, утворивши п'ять груп з інтервалом $h = 40$ грн. Очевидно, інтервали розділення в галузі А потрібно об'єднати, а в галузі Б – розбити.

Результати вторинного групування наведені у табл. 4.9.

Таблиця 4.9

Вторинне групування працюючих за рівнем середньомісячної заробітної плати

[1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 100]

Зарплата, грн	Частка робітників, %	
	Галузь А	Галузь Б
До 160	15	12
160 – 200	$20 + 26 = 46$	$30 + \frac{1}{3} \cdot 21 = 37$
200 – 240	$23 + 9 = 32$	$\frac{2}{3} \cdot 21 + \frac{2}{3} \cdot 18 = 26$
240 – 280	7	$\frac{1}{3} \cdot 18 + 13 = 19$
280 і більше	–	6
Усього	100	100

Порівнявши частки вторинного групування, бачимо, що в галузі В сукупність працівників за рівнем заробітної плати більш диференційована. Перегрупуванням даних можна перейти від структурного групування до типологічного.

Важливі поняття

Децентралізоване зведення – розроблення статистичного матеріалу здійснюється за ієрархічною системою управління, піддаючись відповідному обробленню на кожному рівні.

Інтервали типологічного групування – інтервали, які формуються не за математичними принципами, а за соціально-економічною змістовністю. Межа інтервалу розглядається як умовна межа переходу кількості в нову якість.

Класифікація – більш стійке розмежування одиниць сукупності (спостереження), ніж за групуванням. Класифікації використовуються протягом тривалого часу, хоча з часом вони можуть піддаватися більш-менш суттєвим змінам. Класифікації затверджуються у формі як національного та міжнародного стандартів.

Коефіцієнти кореляції – коефіцієнти, які використовують як вимірники сили зв'язку між статистичними одиницями або між ознаками. Для вимірювання тісноти зв'язку кількісних ознак застосовують коефіцієнти лінійної кореляції.

Коефіцієнти подібності – коефіцієнти, які використовують для вимірювання ступеня наближеності між парою об'єктів, кожен з ознак яких приймає значення 0 або 1.

Міра відстані між двома об'єктами i і j – міра подібності, яка використовується у кластерному аналізі. Для кількісних ознак використовується евклідова відстань.

Просте зведення – операція з підрахунку загальних підсумків за сукупністю одиниць спостереження, тобто визначення розміру досліджуваного явища.

Складне зведення – комплекс операцій, що включають групування одиниць спостереження, підрахунок підсумків за кожною групою і сукупністю в цілому, а також подання результатів групування в табличній формі.

Спеціалізовані інтервали – інтервали, що застосовуються для виділення із сукупності одних типів з однією ознакою для явищ, що перебувають у різних умовах.

Статистичне групування – процес утворення однорідних груп на основі розчленування (розподілу) статистичної сукупності на частини або об'єднання досліджуваних статистичних одиниць у частині сукупності за істотними для них ознаками.

Централізоване зведення – процес, коли весь первинний матеріал надходить в одну організацію, де і піддається обробленню за прийнятою програмою, за єдиною методикою.

Типові завдання

Завдання 4.1. Є інформація про кількість книг, отриманих студентами за абонементом за минулий навчальний рік. Необхідно побудувати ранжований і дискретний варіаційні ряди розподілу, позначивши елементи ряду (табл. 4.10).

Таблиця 4.10

Вихідні дані

2	4	4	7	6	5	2	2	3	4
4	3	6	5	4	7	6	6	5	3
2	4	2	3	5	7	4	3	3	2
4	5	6	6	10	4	3	3	2	3

Розв'язання.

Розглянута сукупність є безліччю варіантів кількості отриманих студентами книг. Слід підрахувати кількість таких варіантів і впорядкувати їх у вигляді варіаційного ранжованого та варіаційного дискретного рядів розподілу (табл. 4.11, 4.12).

Ранжований ряд розподілу

Кількість отриманих книг	Кількість студентів, які отримали книги
2	7
3	9
4	9
5	5
6	6
7	3
10	1
Усього	40

Таблиця 4.12

Дискретний ряд

Кількість отриманих книг	Питома вага студентів, які отримали книги, у загальній сукупності
2	$7/40 = 0,175$
3	$9/40 = 0,225$
4	$9/40 = 0,225$
5	$5/40 = 0,125$
6	$6/40 = 0,150$
7	$3/40 = 0,075$
10	$1/40 = 0,025$
Усього	1

Проаналізуйте отримані результати.

Завдання 4.2. Є дані про вартість основних фондів у 50 підприємств, (тис. грн). Необхідно побудувати ряд розподілу, виділивши 5 груп підприємств (з рівними інтервалами). Вхідні дані наведено в табл. 4.13.

Вихідні дані

18,8	16,0	12,6	20,0	30,0	16,4	14,6	18,4	11,6	17,4
10,4	26,4	16,2	15,0	23,6	29,2	17,0	15,6	21,0	12,0
10,2	13,6	16,6	15,4	15,8	18,0	20,2	16,0	24,0	28,0
16,4	19,6	27,0	24,8	11,0	15,8	18,4	21,6	24,2	24,8
25,8	25,2	13,4	19,4	16,6	21,6	30,0	14,0	26,0	19,0

Розв'язання.

Для розрахунку обираємо найбільше та найменше значення вартості основних фондів підприємств. Це 30,0 і 10,2 тис. грн.

Знайдемо розмір інтервалу: $h = (30,0 - 10,2) : 5 = 3,96$ тис. грн.

Тоді в першу групу будуть входити підприємства, розмір основних фондів яких становить від 10,2 до $10,2 + 3,96 = 14,16$ тис. грн. Таких підприємств буде 9. До другої групи ввійдуть підприємства, розмір основних фондів яких складе від 14,16 до $14,16 + 3,96 = 18,12$ тис. грн. Таких підприємств буде 16. Аналогічно знайдемо число підприємств, що входять у третю, четверту та п'яту групи.

Отриманий ряд розподілу помістимо в табл. 4.14.

Таблиця 4.14

Групування підприємств за вартістю основних фондів

Групи підприємств за розміром основних фондів, тис. грн	Кількість підприємств
10,2 – 14,16	9
14,16 – 18,12	16
18,12 – 22,08	11
22,08 – 26,04	8
26,04 – 30,0	6

Проаналізуйте отримані результати.

Завдання 4.3. Маємо інформацію щодо підприємств легкої промисловості (табл. 4.15).

Таблиця 4.15

Вихідні дані

№ п/п	Середньо-облікова кількість робітників, осіб	Основні фонди, тис. грн	Обсяг виробленої продукції, млн грн	№ п/п	Середньо-облікова кількість робітників, осіб	Основні фонди, тис. грн	Обсяг виробленої продукції, млн грн
1	100	369	560	16	102	256	410
2	140	473	760	17	96	220	370
3	94	251	440	18	98	240	330
4	83	280	520	19	84	106	210
5	157	590	800	20	76	180	160
6	195	1 200	960	21	96	250	300
7	54	160	310	22	85	230	240
8	120	480	570	23	110	370	240
9	180	970	820	24	112	350	230
10	125	400	440	25	67	125	150
11	43	120	100	26	63	140	130
12	256	900	990	27	250	1 150	920
13	182	670	870	28	212	790	650
14	124	500	410	29	184	290	340
15	110	379	310	30	137	275	320

Проведіть групування підприємств за кількістю робітників, утворивши 6 груп з рівними інтервалами. Підрахуйте за кожною групою: кількість підприємств; кількість робітників; обсяг виробленої продукції за рік; середню фактичну продуктивність праці одного робітника; обсяг основних фондів; середній розмір основних фондів одного підприємства; середню величину виробленої продукції одним підприємством. Результати розрахунків оформіть у вигляді таблиці. Зробіть висновки.

Розв'язання.

Для розрахунків виберемо найбільше та найменше значення середньооблікового числа робітників на підприємстві. Це 43 і 256.

Знайдемо розмір інтервалу: $h = (256 - 43) : 6 = 35,5$.

Тоді в першу групу будуть входити підприємства, середньооблікова кількість робітників на яких становить від 43 до $43 + 35,5 = 78,5$ осіб. Таких підприємств буде 5. До другої групи ввійдуть підприємства, середньооблікова кількість робітників на яких складе від 78,5 до $78,5 + 35,5 = 114$ осіб. Таких підприємств буде 12. Аналогічно знайдемо число підприємств, що входять в третю, четверту, п'яту та шосту групи.

Отриманий ряд розподілу помістимо в табл. 4.16 і обчислимо необхідні показники за кожною групою.

Таблиця 4.16

Групування підприємств за кількістю робітників

Групи підприємств за кількістю робітників	Кількість підприємств, одиниць	Кількість робітників, осіб	Обсяг виробленої продукції, млн грн	Середня фактична продуктивність праці одного робітника, грн (ст. 4 : ст. 3)	Обсяг основних фондів, тис. грн	Середній розмір основних фондів одного підприємства (ст. 6 : ст. 2)	Середню величину виробленої продукції одним підприємством, млн грн (ст. 4 : ст. 2)
43 – 78,5	5	303	850	2,805	725	145	170
78,5 – 114	12	1 170	4 160	3,556	3 301	375	346,667
114 – 149,5	5	646	2 500	3,870	2 128	426	500
149,5 – 185	4	703	2 830	4,026	2 520	630	707,500
185 – 220,5	2	407	1 610	3,956	1 990	995	805
220,5 – 256	2	506	1 910	3,775	2 050	1 025	955

Як видно з табл. 4.16, друга група підприємств є найчисленнішою. У неї входять 12 підприємств. Найбільш чисельною є п'ята та шоста групи (по два підприємства). Це найбільші підприємства (за кількістю робітників).

Оскільки друга група найчисленніша, обсяг виробленої продукції за рік підприємствами цієї групи й обсяг основних фондів значно вище за інших. Разом з тим середня фактична продуктивність праці одного робітника на підприємствах цієї групи не є найбільшою. Тут лідерами є підприємства четвертої групи. На цю групу припадає і досить великий обсяг основних фондів.

У висновку зазначимо, що середній розмір основних фондів і середня величина виробленої продукції одного підприємства прямо пропорційні розмірам підприємства (за кількістю робітників).

Завдання 4.4. За даними табл. 4.17 складіть: а) комбінаційне групування підприємств за цими ознаками, утворивши по три групи з рівними інтервалами; наведіть отримані результати графічно; надайте економічну інтерпретацію отриманим даним; б) аналітичне групування, яке показало б залежність продуктивності праці від фондоозброєності.

Таблиця 4.17

Вихідні дані

№ п/п	Обсяг проданого товару, тис. грн	Обсяг вкладених коштів, тис. грн	№ п/п	Обсяг проданого товару, тис. грн	Обсяг вкладених коштів, тис. грн
1	9,1	46,6	13	6,4	11,4
2	8,8	26,0	14	5,8	10,8
3	4,5	14,2	15	6,0	19,7
4	5,8	15,2	16	5,8	21,4
5	12,4	42,2	17	8,9	34,5
6	16	54,0	18	12,6	42,1
7	10,2	30,6	19	4,7	18,6
8	6,4	44,2	20	5,7	37,2
9	10,2	38,2	21	12,2	49,5
10	6,7	23,4	22	8,9	37,6
11	10,4	22,5	23	10,6	23,8
12	5,4	20,2	24	11,6	21,5

Розв'язання.

Комбінаційне групування здійснимо за двома ознаками: обсягом проданого товару й обсягом вкладених коштів. Застосувавши рівні інтервали, визначимо їх ширину та сформуємо інтервали.

Для обсягу проданого товару:

$$h = \frac{x_{\max} - x_{\min}}{m} = \frac{16 - 4,0}{3} = 4 \text{ тис.грн};$$

інтервали: 4 – 8; 8 – 12; 12 – 16.

Для обсягу вкладених коштів:

$$h = \frac{54 - 10,8}{3} = 14,4 \text{ тис. грн};$$

інтервали: 10,8 – 25,2; 25,2 – 39,6; 39,6 – 54.

Комбінаційне групування підприємств наведено в табл. 4.18.

Таблиця 4.18

Комбінаційне групування підприємств

Обсяг проданого товару, тис. грн	Обсяг вкладених коштів, тис. грн			Разом
	10,8 – 25,2	25,2 – 39,6	39,6 – 54	
4 – 8	9	2	–	11
8 – 12	3	4	2	9
12 – 16	–	–	4	4
Усього	12	6	6	24

Дані табл. 4.18 показують на наявність прямого зв'язку між обсягом проданого товару й обсягом вкладених коштів, що також підтверджується виглядом діаграми розсіювання наявних значень змінних (рис. 4.3).

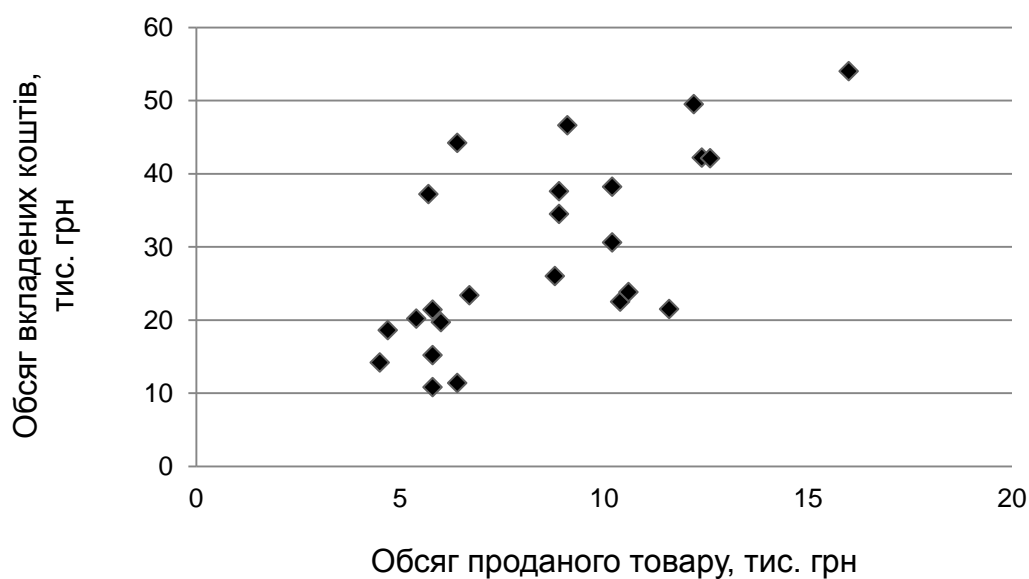


Рис. 4.3. Залежність обсягу проданого товару від обсягу вкладених коштів

Для підтвердження наявності зв'язку між показниками проведемо аналітичне групування (табл. 4.19).

Таблиця 4.19

Аналітичне групування підприємств

Обсяг проданого товару, тис. грн	Кількість підприємств	Обсяг вкладених коштів, тис. грн	
		Разом	На 1 підприємство
4 – 8	11	236,3	21,48
8 – 12	9	282,3	31,37
12 – 16	4	187,8	46,95
Усього	24	706,4	х
У середньому	х	235,5	29,43

Зіставлення двох показників за допомогою аналітичного групування підтверджує гіпотезу про наявність зв'язку між обсягом проданого товару й обсягом вкладених коштів. Цей зв'язок є прямо пропорційним, тобто з ростом обсягом вкладених коштів зростає також і обсяг проданого товару.

Завдання 4.5. Маємо дані з оцінювання якості роботи банківських працівників щодо обслуговування клієнтів банку "СТД" за останній місяць. Дані подані за дванадцятибальною системою, із групуванням за чотирибальною системою (табл. 4.20).

Таблиця 4.20

Вихідні дані

Оцінка	Кількість працівників банку
10 – 12 балів	25
7 – 9 балів	56
4 – 6 балів	42
1 – 3 бали	14

Необхідно зробити вторинне групування утворивши такі групи за системою (у балах): відмінно – 11 – 12; дуже добре – 10; добре – 7 – 9; задовільно – 5 – 6; незадовільно – 4; дуже незадовільно – 3; вкрай незадовільно – 1 – 2.

Результати та порядок розрахунків наведемо у вигляді табл. 4.21.

Таблиця 4.21

Розрахункові дані

Групи	Порядок розрахунків	Кількість працівників банку
Відмінно	$(25/3) \times 2 = 17$	17
Дуже добре	$(25/3) \times 1 = 8$	8
Добре	56	56
Задовільно	$(42/3) \times 2 = 28$	28
Незадовільно	$(42/3) \times 1 = 14$	14
Дуже незадовільно	$(14/3) \times 1 = 5$	5
Вкрай незадовільно	$(14/2) \times 2 = 9$	9

Кількість працівників банку у групах під час розрахунків округляється, тому що не може бути неціла кількість працівників банку. Отже, завдяки перегрупуванню ми отримали групування якості роботи банківських працівників щодо обслуговування клієнтів банку "СТД" згідно із запропонованою шкалою.

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Зведення і групування статистичних даних" подані у роботі [67]. Лабораторна робота націлена на опанування навички групування даних у пакеті MS Excel. Завданням роботи є групування статистичних даних за допомогою надбудови MS Excel "Аналіз даних".

Запитання для самодіагностики

1. У чому полягає завдання зведення статистичних даних?
2. Назвіть прийоми групування статистичних даних.
3. Назвіть основні статистичні класифікації.
4. Наведіть завдання групування та їх значення в статистичному дослідженні.
5. Окресліть групувальні ознаки та принципи їх вибору.
6. Надайте визначення кількості груп і розміру інтервалів групування.
7. Охарактеризуйте види групувань.
8. Опишіть функції групувань.
9. Опишіть принцип методу вторинного групування.
10. У чому полягають принципи утворення інтервалів у групуванні?

Питання для критичного переосмислення (есе)

1. Сутність зведення статистичних даних у процесі створення інформаційно-аналітичної бази дослідження.
2. Відмінність між класифікацією та групуванням: практичний аспект.
3. Функції в економіко-статистичному аналізі, виконувані за допомогою групування.
4. Поясніть відмінності між багатовимірним та комбінаційним групуваннями.
5. Сутність практичної складової побудови типологічного, структурного й аналітичного групування.

Розділ 2

Статистичні індикатори та ряди розподілу

5. Узагальнювальні статистичні показники

Основні питання:

- 5.1. Сутність, види та значення абсолютних показників.
- 5.2. Сутність, види та форми вираження відносних показників.
- 5.3. Сутність, види та методи розрахунку середніх показників.

5.1. Сутність, види та значення абсолютних показників

Інформація про розміри, пропорції, зміни в часі, інші закономірності соціально-економічних явищ створюється, передається і зберігається у вигляді статистичних показників. З філософської точки зору **статистичний показник** – це міра, тобто єдність якісного та кількісного відображення певної властивості соціально-економічного явища чи процесу.

Якісна змістовність показника визначається сутністю явища, відображаючись у назві: народжуваність, врожайність, прибутковість тощо. Кількісний аспект подається числом і його вимірником. Оскільки статистика вивчає суспільні явища в конкретних умовах простору та часу, то значення будь-якого показника визначається щодо цих атрибутів. У цьому сенсі говорять, наприклад, що капітал певної фірми на початок 2018 р. склав 4400 млн грн.

Сполучною ланкою між якісною змістовністю і числовим вираженням є правило побудови – **модель показника**, що розкриває його статистичну структуру, встановлюючи, що, де, коли і як має бути виміряне. У моделі обґрунтовуються одиниці, взяті для вимірювання, технологія збирання даних, обчислювальні операції [13; 49; 68]. Модель показника має надзвичайно важливе значення для забезпечення достовірності статистичної інформації.

Показники розрізняють **за способом обчислення, ознакою часу й аналітичними функціями** (рис. 5.1).



Рис. 5.1. Класифікація показників

За **способом обчислення** розглядають *первинні* та *похідні* показники.

Первинні визначаються зведенням даних статистичного спостереження і подаються у формі абсолютних величин (кількість і сума вкладів в ощадний банк). *Похідні* показники обчислюються на базі первинних або похідних показників. Вони мають форму середніх чи відносних величин (середня заробітна плата, індекс середньої заробітної плати).

За **ознакою часу** показники розподіляють на *інтервальні* та *моментні*. *Інтервальні* характеризують явище за певний час (день, декаду, місяць, рік). Прикладом може бути обсяг виробленої продукції, введення в дію житла, споживання свіжої води та ін. До *моментних* відносять показники, що дають кількісну характеристику явищ на певний момент часу: площа виноградних і цитрусових насаджень, протяжність нафтопроводів на кінець року тощо.

Інтервальні та моментні показники можуть бути як *первинними*, так і *похідними*. Наприклад, площа зрошуваних земель – первинний моментний показник, а частка таких земель у загальній площі – похідний моментний

показник; спожита електроенергія в галузі – первинний інтервальний показник, а з розрахунку на одиницю робочого часу – похідний.

Інтервальні показники залежать від тривалості періоду, за який вони обчислюються. Особливістю первинних інтервальних показників є *адитивність*, тобто можливість підсумовування. Похідні показники в основному *неадитивні*.

Серед статистичних показників існують пари *взаємозворотних показників*, які паралельно характеризують те саме явище та розподіляються за ознакою оборотності. Прямий показник x зростає з посиленням явища, зворотний $1/x$, навпаки, зменшується. Прикладом можуть слугувати такі показники:

купівельна спроможність грошової одиниці – прямий показник, ціна одиниці товару – зворотний;

продуктивність праці за одиницю часу – прямий показник, трудомісткість одиниці продукції – зворотний і т. д.

За способом виконання своїх функції (аналітичні функції) показники можуть відображати обсяг явища, його середній рівень, інтенсивність прояву, структуру, зміни в часі або в просторі. У зв'язку із цим, виділяють такі групи показників [18; 42; 68]:

абсолютні та відносні;

середні величини;

показники варіації.

Абсолютні показники характеризують чисельність сукупності або обсяг досліджуваного явища в конкретних межах простору та часу, тобто відображають рівень розвитку явища, його розмір.

Абсолютний показник можна отримати одним із **двох способів**:

шляхом підрахунку одиниць сукупності, що володіють конкретним значенням ознаки (наприклад: кількість транспортних підприємств у м. Харкові на конкретну дату, чисельність промислово-виробничого персоналу підприємства тощо);

шляхом підсумовування значення ознаки за всією статистичною сукупністю.

Абсолютні показники завжди є іменованими числами. Залежно від соціально-економічної сутності досліджуваних явищ вони виражаються у натуральних, вартісних і трудових одиницях вимірювання.

Натуральні вимірники використовують у тих випадках, коли одиниці вимірювання відповідають споживчим властивостям досліджуваних

явищ (наприклад: виробництво автомобілів вимірюється в штуках, виробництво сталі – у тоннах, врожайність – у центнерах).

Натуральні одиниці вимірювання можуть бути *складними*. Такі одиниці застосовують, коли для характеристики досліджуваного явища однієї одиниці вимірювання недостатньо, тоді використовується твір двох одиниць. Наприклад, виробництво електроенергії вимірюється у кіловат-годинах, вантажообіг – у тонно-кілометрах.

До групи натуральних включають також **умовно-натуральні** одиниці вимірювання. Вони використовуються, коли який-небудь продукт має кілька різновидів і загальний обсяг можна отримати тільки виходячи із загальної для всіх різновидів споживчої властивості.

Щоб отримати узагальнені підсумки, один із різновидів продукту приймається як одиниця вимірника, а інші приводяться до нього за допомогою відповідних коефіцієнтів перерахунку.

Вартісні вимірники дозволяють дати грошову оцінку досліджуваним явищам і процесам. Ці вимірники використовують для узагальнення даних, починаючи з рівня підприємства та до рівня народного господарства, та для оцінювання неоднорідних статистичних сукупностей. У вартісних одиницях вимірюється обсяг випущеної продукції підприємства, доходи населення тощо. Схема отримання загального обсягу статистичної ознаки у вартісному вираженні виглядає таким чином:

$$Q = \sum_{i=1}^m p_1 \times q_1, \quad (5.1)$$

де p_1 – ціна (вартісна оцінка одиниці ознаки);

q_1 – обсяг ознаки в натуральному вираженні;

m – кількість ознак.

Показники, виражені в вартісних одиницях, можна підсумувати, отримати за ними підсумкові дані, але в їх використанні необхідно враховувати зміну цін у часі. Для усунення зазначеного недоліку вартісних вимірників слід застосовувати "незмінні" або "зіставні" ціни одного періоду [13; 68; 52].

Трудові одиниці вимірювання застосовують для оцінювання загальних витрат праці та трудомісткості окремих операцій технологічного процесу. До них належать людино-години, людино-дні (оцінка витрат робочого часу), нормо-хвилини (оцінка трудомісткості).

На рис. 5.2 наведено класифікацію абсолютних показників.

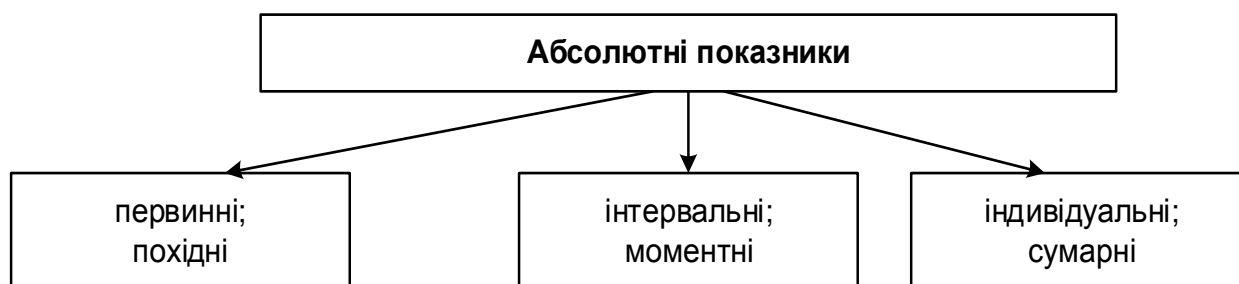


Рис. 5.2. Класифікація абсолютних показників

Уся сукупність абсолютних величин включає як індивідуальні (характеризують значення окремих одиниць сукупності), так і сумарні (характеризують підсумкове значення декількох одиниць сукупності або підсумкове значення істотної ознаки з тієї чи іншої частини сукупності) показники.

Абсолютні показники слід також підрозділити на моментні та інтервальні. *Моментні абсолютні показники* характеризують факт існування явища або процесу, його розмір (обсяг) на певну дату/час. *Інтервальні абсолютні показники* характеризують загальний обсяг явища за той чи інший період часу (наприклад, випуск продукції за квартал або за рік), допускаючи подальше підсумовування [42; 49; 68].

Як такі абсолютні показники не дають повного уявлення про явище, що вивчається, не показують його структуру, розвиток у часі, співвідношення між його частинами: на їх основі складно проводити порівняння з іншими подібними явищами. Ці аналітичні функції виконують відносні показники.

5.2. Сутність, види та форми вираження відносних показників

Хоча абсолютні показники відіграють важливу роль у практичній і пізнавальній діяльності людини, аналіз фактів обов'язково приводить до необхідності різного роду зіставлень. І тоді абсолютні показники, що характеризують ті або інші досліджувані явища, розглядаються не тільки відокремлено, але й у порівнянні з іншими показниками, прийнятими за масштаб оцінювання або базу порівняння.

Відносним статистичним показником є узагальнювальна характеристика, виражена у вигляді числової міри співвідношення двох абсолютних величин, що зіставляються. Такий показник отримують шляхом ділення одного абсолютного показника на інший. Логічна схема розрахунку відносного показника виглядає таким чином:

$$\text{Відносна величина} = \frac{\text{Величина порівняння}}{\text{База порівняння}}. \quad (5.2)$$

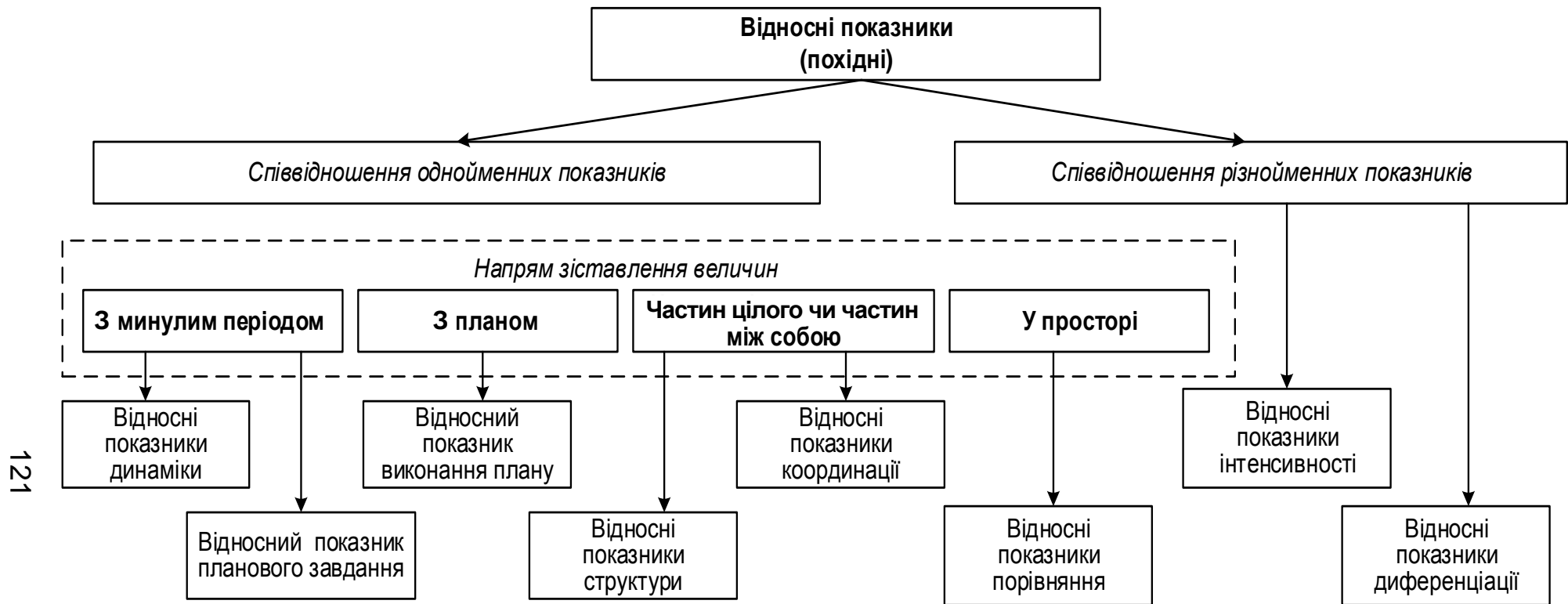
Залежно від величин, які стоять у чисельнику або в знаменнику, відносні величини можуть бути виражені в таких формах: коефіцієнтах (частинах), відсотках (%), проміле (‰), продецимілле (‱), коли за базу порівняння беруть, відповідно, 1, 100, 1 000, 10 000 одиниць. Проміле широко застосовують у демографічній статистиці для характеристики народжуваності, смертності населення та відображення інших демографічних процесів. ПрODEцимілле використовують для оцінювання забезпечення населення лікарняними ліжками, місцями у закладах вищої освіти. Варто зауважити, що безрозмірним за формою відносним показникам може бути приписана конкретна одиниця вимірювання. Наприклад, показники природного руху населення – коефіцієнт народжуваності, коефіцієнт смертності; обчислюється в проміле, але показує кількість народжених або померлих за рік, розраховуючи на 1 000 осіб.

Розмаїтість співвідношень у реальному житті вимагає застосування різних за змістовністю та статистичною природою відносних показників. Залежно від своїх функцій, у процесі аналізу відносні величини класифікують на види (рис. 5.3).

Розглянемо змістовність кожного з виділених видів відносних показників.

1. Відносний показник динаміки. Динамікою в статистиці називають зміну соціально-економічного явища або процесу в часі. Відносний показник динаміки (ВПД) характеризує напрям та інтенсивність зміни показника в часі; визначається співвідношенням його значення або за два періоди, або за моменти часу.

Базою порівняння можуть бути: попередній рівень, що змінюється, – розрахунок ланцюговим способом; постійний, вилучений у часі рівень – розрахунок базисним способом. ВПД називають **темпами росту** [34; 68].



121

Рис. 5.3. Класифікація відносних величин

Приклад 5.1. Обсяг товарообігу в Україні становить (млн грн): 2015 р. – 544 153; 2016 р. – 720 731; 2017 р. – 949 864. Розрахуємо темпи росту даного показника:

1) *ланцюговим способом:*

$$\text{ВПД} = \frac{\text{показник поточного року}}{\text{показник попереднього року}}. \quad (5.3)$$

Темп росту товарообігу країни склав: в 2016 р. $\frac{720\,731}{544\,153} = 1,33$ (або 133 %); у 2017 р. $\frac{949\,864}{720\,731} = 1,32$ (або 132 %). Темп росту товарообігу країни в 2018 р. збільшився в порівнянні з 2015 р. на 33 %, а темп росту товарообігу країни в 2017 р. порівняно з 2016 р. збільшився на 32 %. Спостерігається зниження темпу росту аналізованого показника;

2) *базисним способом:*

$$\text{ВПД} = \frac{\text{показник поточного року}}{\text{показник базисного року}}. \quad (5.4)$$

За базу порівняння приймаємо 2015 р. Тоді темп росту товарообігу країни в 2016 р. співпаде з розрахованим за ланцюговим способом і складе 32 %; темп росту товарообігу країни в 2017 р. складе $\frac{949\,864}{544\,153} = 1,75$. Тобто спостерігається збільшення темпу росту товарообігу країни в 2017 р. порівняно з 2015 р. на 75 %.

2. Відносні показники структури (ВПС) характеризують частки (питомі ваги) частин сукупності в загальному її обсязі. Вони показують структуру сукупності, її будову. Розрахунок відносних показників структури полягає в обчисленні питомих ваг окремих частин у всій сукупності:

$$\text{ВПС} = \frac{\text{показник частини сукупності}}{\text{показник всієї сукупності}}. \quad (5.5)$$

ВПС звичайно виражають у формі коефіцієнтів або відсотках. Сума коефіцієнтів повинна становити 1, а сума відсотків – 100, тому що питомі ваги зведені до загальної підстави. Відносні показники структури використовують для вивчення складних явищ, що розпадаються на частини, (для

вивчення складу населення за різними ознаками – вік, освіта, національність тощо). Приклад такої сукупності наведено в табл. 5.1.

Таблиця 5.1

Структура роздрібного товарообігу міста

Показники	Сума, млрд грн	У % до підсумку
Роздрібний товарообіг, усього	83,1	100
у тому числі:	42,4	51
організації роздрібної торгівлі		
неторгові організації	20,8	25
фізичні особи на ринках	19,9	24

3. Відносні показники координації (ВПК) характеризують відношення частин даних статистичної сукупності до однієї з них, взятої за базу порівняння. ВПК показують, у скільки разів одна частина сукупності більше іншої або скільки одиниць однієї частини сукупності доводиться на 1,10,100 і т. д. одиниць іншої частини.

За базу порівняння вибирається частина, що має найбільшу питому вагу або є пріоритетною в сукупності. Так, прийнявши за базу в попередньому прикладі товарообіг організацій роздрібної торгівлі, можна розрахувати ВПК для неторгових організацій: $VPK = 25 / 51 = 0,49$. Це означає, що на кожну гривню товарообігу роздрібної торгівлі доводиться 49 копійок товарообігу неторгових організацій.

Відносні показники координації відіграють важливу роль в економічному аналізі, тому що з їх допомогою наявні в сукупності співвідношення сформовані більш чітко та наочно.

4. Відносний показник плану (прогнозу) (ВПП) і відносний показник виконання плану (ВПВП) використовують усі суб'єкти фінансово-господарської діяльності, що здійснюють поточне та стратегічне планування. Ці показники розраховують таким чином:

$$VPP = \frac{\text{плановий показник поточного періоду}}{\text{фактичний показник попереднього періоду}}; \quad (5.6)$$

$$VPVP = \frac{\text{фактичний показник поточного періоду}}{\text{плановий показник поточного періоду}}. \quad (5.7)$$

Відносний показник плану (відносний показник планового завдання) характеризує *напруженість* планового завдання, а *відносний показник виконання плану* – *ступінь* його виконання.

Приклад 5.2. Розрахунок ВПП і ВПВП: фактичний оборот фірми в 2017 р. склав 22 млн грн; аналіз ринку показав, що за 2018 р. реально довести оборот до 2,6 млн грн; фактичний же оборот – 2,5 млн грн.

$$\text{ВПП} = \frac{2,6}{2,2} = 1,18;$$

$$\text{ВПВП} = \frac{2,5}{2,6} = 0,96.$$

Розрахунки, що показують планове завдання на 2018 р., у 1,3 раза перевищує фактичний рівень 2017 р., але план 2018 р. виконаний тільки на 96 %.

Відносний показник динаміки (ВПД), розрахований ланцюговим способом, відносні показники плану (ВПП) і виконання плану (ВПВП) зв'язані між собою залежністю:

$$\begin{aligned} \text{ВПД} &= \text{ВПП} \times \text{ВПВП} = \\ &= \frac{\text{Плановий показник поточного року}}{\text{Фактичний показник попереднього року}} \times \frac{\text{Фактичний показник поточного року}}{\text{Плановий показник поточного року}} = \\ &= \frac{\text{Фактичний показник поточного року}}{\text{Фактичний показник попереднього року}}. \end{aligned} \quad (5.8)$$

5. Відносні показники порівняння (ВППр) характеризують порівняльні розміри однойменних абсолютних показників, що належать до різних об'єктів або територій, але за однаковий період часу. Їх обчислюють як частки від розподілу однойменних абсолютних показників, що характеризують різні об'єкти, що віднесені до того ж періоду або моменту часу.

$$\text{ВППр} = \frac{\text{Показник, що характеризує об'єкт А}}{\text{Показник, що характеризує об'єкт Б}}. \quad (5.9)$$

За допомогою показників порівняння можна зіставляти продуктивність праці в різних країнах і визначати, де й у скільки разів вона вище;

порівнювати ціни на різні товари, економічні показники різних підприємств тощо.

Існує така класифікація ВППСр: відносні показники просторового порівняння; відносні показники порівняння зі стандартом.

Відносні показники просторового порівняння (ВППСр) – це відношення розмірів або рівнів однойменних показників за різними територіями або об'єктами. Найчастіше їх використовують для порівняння показників економічного розвитку або життєвого рівня країн і регіонів. Базою порівняння в цьому випадку може виступати будь-який об'єкт, але методика розрахунку показників повинна бути єдиною [18; 52; 68].

Відносні показники порівняння зі стандартом вимагають порівняння фактичних значень показників з певним еталоном – стандартом, нормативом, оптимальним рівнем, галузевим нормативом тощо. Наприклад, визначення відхилення коефіцієнта поточної ліквідності підприємства від нормативних значень [18; 52; 68].

6. Відносні показники інтенсивності та рівня економічного розвитку (ВПІ) характеризують ступінь поширення або рівень розвитку досліджуваних явищ і процесів у певному середовищі; утворюються як результат порівняння різнойменних, але певним чином зв'язаних між собою величин. Зазначені показники розраховують таким чином:

$$\text{ВПІ} = \frac{\text{Показник, що характеризує явище}}{\text{Показник, що характеризує середовище розподілу явища}}. \quad (5.10)$$

ВПІ обчислюють, розраховуючи на 100, 1 000, 10 000 і т. д. одиниць досліджуваної сукупності, та використовують у тих випадках, коли неможливо за значенням абсолютного показника визначити масштаб поширення явища. Так, для вивчення демографічних процесів розраховують показники народжуваності, смертності, природного приросту (убування) населення як відношення числа народжених (померлих) або величини природного приросту за рік до середньорічної чисельності населення певної території на 1 000 або 10 000 осіб; забезпечення лікарською допомогою розраховується на 10 000 осіб; захворюваність або злочинність – на 100 000 населення.

Приклад 5.3. За станом на 01.01.2019 р. у Харкові народилося 1 596 немовлят, у Сумах – 574. Зіставлення абсолютних показників не дозволяє оцінити рівень народжуваності та визначити, де цей рівень

вище. Це можна зробити через ВПІ – коефіцієнти народжуваності в Харкові та Сумах. Населення Харкова в січні 2019 р. склало – 2 674,6 тис. осіб, у Сумах – 1 080,7.

$$\text{ВПІ}_{\text{Харків}} = \frac{1596}{2674,6} = 0,59 \text{ осіб / тис. осіб};$$

$$\text{ВПІ}_{\text{Суми}} = \frac{574}{1080,7} = 0,53 \text{ осіб / тис. осіб}.$$

Висновок: народжуваність у Харкові вища, ніж у Сумах.

7. Відносні показники диференціації (ВПдиф) розраховують шляхом порівняння двох структурних рядів, один із яких характеризує співвідношення сукупності за чисельністю одиниць, а другий – за величиною якої-небудь ознаки. Наприклад, порівняння питомої ваги господарств за чисельністю і питомою вагою валової продукції, основних фондів, працівників тощо [49; 68].

Відносні показники мають важливе значення в практичній діяльності, але їх не можна розглядати у відриві від абсолютних показників, через які вони розраховуються. У протилежному випадку можна прийти до неправильних висновків. Отже, тільки спільне використання абсолютних і відносних показників дозволяє провести якісний аналіз різних явищ соціально-економічного життя.

5.3. Сутність, види та методи розрахунку середніх показників

Одним із головних завдань статистичного дослідження є виявлення закономірностей масових явищ. Закономірності можна виявити лише узагальнюючи однорідні явища та даючи узагальнювальну характеристику одиницям явища. Тому з метою виявлення значення ознаки, яке є характерним для всієї сукупності одиниць, що вивчаються, розраховують середню величину.

Вагомий внесок в обґрунтування і розвиток теорії середніх величин здійснив А. Кетле (1796 – 1874), член Бельгійської академії наук, член-кореспондент Петербурзької академії наук. В обчисленні середньої величини за множиною одиниць вплив випадкових причин погашається. Середня, абстрагуючись від індивідуальних особливостей окремих одиниць сукупності, виражає загальні властивості, притаманні всім одиницям.

А. Кетле підкреслював, що статистичні середні є *не просто методом математичного вимірювання, а категорією об'єктивної дійсності*. **Принципова сутність статистичного пізнання** полягає в погашенні випадкового, викликаного дією індивідуальних причин, і у виявленні закономірностей, обумовлених загальними причинами [52; 68].

Можливістю переходу від одиничного до загального, від випадкового до закономірного пояснюється *важливість методу середніх величин* і його широке застосування в статистичних дослідженнях.

Умовами наукового використання середніх величин є таке [68]:

- сукупність повинна складатися з якісно однорідних одиниць;
- метод середніх треба поєднувати з методом групувань;
- сукупність має бути достатньо великою;
- можливість використання системних середніх.

Отже, **середня величина** є узагальнювальною характеристикою ознаки досліджуваної сукупності, котра відображає її типовий рівень у розрахунку на одиницю сукупності в конкретних умовах місця і часу.

Середні величини застосовуються: для оцінювання досягнутого рівня показника, що вивчається; в аналізі та плануванні виробничо-господарської діяльності підприємств (об'єднань), фірм, банків та інших господарських одиниць; для виявлення взаємозв'язків явищ; у прогнозуванні та розрахунку нормативів.

Коректність використання середніх величин в аналітичній практиці обумовлена насамперед наявністю якісно однорідної сукупності, за якою обчислено середню.

Перед тим як обчислювати середні величини, необхідно провести угруповання одиниць досліджуваної сукупності, виділивши якісно однорідні групи.

Середня, що розрахована за сукупністю в цілому, є *загальною середньою*; *середні*, що обчислені для кожної групи, – *груповими середніми*. Загальна середня відображує загальні риси досліджуваного явища; групова середня дає характеристику розміру явища, що складається в конкретних умовах даної групи.

Крім того, *середня величина завжди є іменованою*, тобто має ту саму розмірність (одиницю вимірювання), що і ознака в окремих одиницях сукупності.

У соціально-економічному аналізі використовують два класи середніх величин: *степеневі середні*; *структурні середні*.

До **степеневих середніх** належать кілька видів середніх, що побудовані за одним загальним принципом:

$$\bar{x} = \sqrt[k]{\frac{\sum_{i=1}^n x_i^k}{n}}, \quad (5.11)$$

де x_i – варіанта;

n – обсяг статистичної сукупності;

k – показник ступеня, від якого залежить вид середньої.

Поєднання індивідуальних значень ознаки з визначальними властивостями сукупності обумовлює вид ступеневої середньої.

На практиці використовують декілька значень показника ступеня k : за $k = -1$ використовують середню гармонічну; за $k = 0$ – середню геометричну; за $k = 1$ отримують середню арифметичну; за $k = 2$ – середню квадратичну.

Степеневі середні залежно від форми подання вихідних даних можуть бути *простими* та *зваженими*.

Якщо вихідні дані подані простим переліком значень ознаки у статистичних одиниць, то використовується формула **ступеневої середньої простої**:

$$\bar{x} = \sqrt[k]{\frac{\sum_{i=1}^n x_i^k}{n}}.$$

Якщо дані попередньо згруповані (подані рядом розподілу), то використовується формула **ступеневої середньої зваженої**:

$$\bar{x} = \sqrt[k]{\frac{\sum_{i=1}^m x_i^k f_i}{\sum_{i=1}^m f_i}},$$

де f_i – частота повторення індивідуальних значень ознаки;

m – кількість однорідних груп.

Необхідно детально розглянути види середніх величин.

1. Середня арифметична є найбільш поширеним видом степеневих середніх. Вона використовується у випадках, коли обсяг усереднюваної ознаки є адитивною величиною (утворюється як сума окремих значень ознаки). Якщо індивідуальні значення ознаки у статистичних одиниць замінити середньою арифметичною, то сумарний обсяг ознаки за сукупністю в цілому буде зберігатися незмінним.

Середня арифметична проста, використовується в роботі з *незгрупованими даними* та розраховується за формулою:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}. \quad (5.12)$$

Якщо у вихідних даних окремі значення усереднюваної ознаки будуть повторюватися, то розрахунок середньої здійснюватиметься за згрупованими даними або варіаційними рядами. У цьому випадку кількість одиниць за варіантами ознаки не є однаковою; її подають у вигляді ваги. У цьому разі для розрахунку необхідно застосовувати *середню арифметичну зважену* (середню згрупованих величин).

$$\bar{x} = \frac{\sum_{i=1}^m x_i f_i}{\sum_{i=1}^m f_i} \quad \text{або} \quad x = \sum_{i=1}^m x_i d_i, \quad (5.13)$$

де $d = \frac{f_i}{\sum_{i=1}^m f_i}$ – частість, тобто питома вага статистичних одиниць, що во-

лодіють певним значенням ознаки в загальному обсязі сукупності.

Наведені формули розрахунку середньої арифметичної використовують як для дискретних, так і для інтервальних рядів. Причому інтервальний ряд треба перетворити в дискретний шляхом знаходження значення середини кожного інтервалу (центру). Якщо мають місце інтервали з нечітко виділеними межами (нижній або верхній інтервали відкриті), то для визначення середньої величини інтервали необхідно перевести в закриті: для першого інтервалу беруть розмір кроку другого інтервалу, а для останнього – розмір кроку попереднього інтервалу.

Середня арифметична має ряд корисних **властивостей**, до найважливіших з яких відносять таке:

1) середня арифметична постійної величини дорівнює цій величині:

$$A = A \text{ за } A = \text{const};$$

2) алгебраїчна сума відхилень варіант від їх середньої арифметичної дорівнює нулю:

$$\begin{aligned} \sum(x_i - \bar{x}) &= \sum x - n\bar{x} = 0 && \text{для простої середньої;} \\ \sum(x - \bar{x})f &= \sum xf - \bar{x}\sum f = 0 && \text{для зваженої середньої;} \end{aligned}$$

3) сума квадратів відхилень від середньої величини завжди буде менше, ніж сума квадратів відхилень від будь-якої іншої величини:

$$\sum(x - \bar{x})^2 f < (x - A)^2 f;$$

4) якщо всі варіанти зменшити (збільшити) на постійне число A , то середня арифметична з них зменшиться (збільшиться) на це саме число:

$$\frac{\sum(x_i \pm A) \times f_i}{\sum f_i} = \bar{x} \pm A;$$

5) якщо всі варіанти однаково збільшити (зменшити) в ту саму кількість разів, то середня арифметична збільшиться (зменшиться) в стільки ж разів:

$$\frac{\sum(x_i \times A) \times f_i}{\sum f_i} = \bar{x} \times A;$$

6) якщо всі ваги середньої однаково збільшити (зменшити) в кілька разів, то середня арифметична не зміниться:

$$\frac{\sum x_i \left(\frac{f_i}{A}\right)}{\sum \frac{f_i}{A}} = \bar{x};$$

7) помноження середніх на суму частот дорівнює сумі помножень варіантів на частоти:

$$\bar{x} \sum f = \sum xf;$$

8) загальна середня дорівнює середній частинних середніх, зважених на чисельність відповідних груп сукупності:

$$\bar{X}_{\text{заг}} = \frac{\bar{x}_1 f_1 + \bar{x}_2 f_2 + \dots + \bar{x}_n f_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum \bar{x}_i f_i}{\sum f_i}.$$

$$\text{Якщо } d_f = \frac{f}{\sum f}, \text{ то } \bar{X}_{\text{заг}} = \frac{\sum \bar{x}_i d_f}{\sum d_f}.$$

2. Середня гармонічна має більш складну конструкцію, ніж середня арифметична. Вона використовується в тих випадках, коли статистична інформація не містить *частот* за окремими значеннями ознаки, а подана помноженням значення ознаки на частоту [18; 68].

$$\bar{x} = \sqrt[n]{\sum_{i=1}^n \frac{1}{x_i}}. \quad (5.14)$$

Залежно від форми подання вихідних даних середня гармонічна може бути розрахована як *проста* та як *зважена*.

Якщо вихідні дані не згруповані, то застосовується *середня гармонічна проста*:

$$\bar{x} = \frac{n}{\sum_{i=1}^m \frac{1}{x_i}}. \quad (5.15)$$

У роботі зі згрупованими даними використовується **середня гармонічна зважена**:

$$\bar{x} = \frac{\sum_{i=1}^m M_i}{\sum_{i=1}^m \frac{M_i}{x_i}}, \quad (5.16)$$

де M_i – статистична вага; $M_i = x_i \times f_i$.

Середня геометрична застосовується в тих випадках, коли загальний обсяг усереднюваної ознаки є мультиплікативною величиною (коли визначальна властивість формується як добуток індивідуальних значень ознаки) [13; 18].

$$\bar{x}_{\text{геом}} = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} = \sqrt[n]{\prod_{i=1}^n x_i}. \quad (5.17)$$

Форма *середньої геометричної зваженої* у практичних розрахунках не застосовується. У соціально-економічних дослідженнях середня геометрична застосовується в аналізі рядів динаміки для визначення середнього коефіцієнта зростання, коли задана послідовність відносних величин динаміки.

Середня геометрична може використовуватися для визначення рівновіддаленої величини від максимального та мінімального значення ознаки.

Середня квадратична використовується в тих випадках, коли у разі заміни індивідуальних значень ознаки на середню величину необхідно зберегти незмінною суму квадратів вихідних величин. Головна сфера її використання – вимірювання ступеня коливання індивідуальних значень ознаки щодо середньої арифметичної: середньоквадратичне відхилення. Крім цього, середня квадратична застосовується, якщо необхідно обчислити середню величину ознаки, що виражена в квадратних або кубічних одиницях вимірювання (в обчисленні середньої величини квадратних ділянок, середніх діаметрів труб, стовбурів та ін.).

Середня квадратична може бути:

$$\text{проста: } \bar{x}_{\text{КВ}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}; \quad (5.18)$$

$$\text{зважена: } \bar{x}_{\text{КВ}} = \sqrt{\frac{\sum_{i=1}^m x_i^2 \cdot f_i}{\sum_{i=1}^m f_i}}. \quad (5.19)$$

Усі **степеневі середні** розрізняються між собою значеннями показника ступеня: чим вище показник ступеня, тим більше кількісне значення середнього показника:

$$\bar{x}_{\text{гарм}} \leq \bar{x}_{\text{геом}} \leq \bar{x}_{\text{арифм}} \leq \bar{x}_{\text{кв}} \quad (5.20)$$

Таку залежність степеневих середніх називають **властивістю мажорантності середніх**.

З метою *узагальнення даних рядів динаміки* розраховують середній рівень ряду динаміки, використовуючи **середню хронологічну** для моментного ряду з рівними проміжками часу поміж датами:

$$\bar{x} = \frac{\frac{1}{2}x_1 + x_2 + \dots + \frac{1}{2}x_n}{n-1} \quad (5.21)$$

У випадку, коли усереднена ознака є відношенням між логічно зв'язаними величинами, постає питання про вибір виду середньої.

Основний методологічний принцип вибору виду середньої – **забезпечення логіко-змістовної сутності показника**, що вивчається.

Формально цей принцип можна записати, як показано в табл. 5.2.

Таблиця 5.2

Основні методологічні принципи вибору виду середньої

Показники	Прямі	Оборотні
Первинні	Проста арифметична	Проста гармонічна
Похідні	Зважена	Зважена гармонічна

Середня величина може бути розрахована й у тому випадку, коли окремі значення варіант не реєструються, а відомі лише підсумки. Так, не підраховуються дані про витрати палива на кожну кіловат-годину

електроенергії, врожайність на кожному окремому гектарі посівної площі тієї чи іншої сільськогосподарської культури тощо.

Важливі поняття

Абсолютні показники – показники, що характеризують чисельність сукупності або обсяг досліджуваного явища в конкретних межах простору та часу, тобто відображають рівень розвитку явища, його розмір.

Відносний показник динаміки (ВПД) – показник, який характеризує напрям та інтенсивність зміни показника в часі; визначається співвідношенням його значення або за два періоди, або за моменти часу.

Відносний показник плану (відносний показник планового завдання) – показник, який характеризує *напруженість* планового завдання, а відносний *показник виконання плану* – ступінь його виконання.

Відносний статистичний показник – узагальнювальна характеристика, виражена у вигляді числової міри співвідношення двох абсолютних величин, що зіставляються.

Відносні показники диференціації (ВПдиф) – показники, які розраховують шляхом порівняння двох структурних рядів, один із яких характеризує співвідношення сукупності за чисельністю одиниць, а другий – за величиною якої-небудь ознаки.

Відносні показники координації (ВПК) – показники, які характеризують відношення частин даних статистичної сукупності до однієї з них, узяті за базу порівняння; показують, у скільки разів одна частина сукупності більше іншої або скільки одиниць однієї частини сукупності доводиться на 1, 10, 100 тощо одиниць іншої частини.

Відносні показники порівняння (ВППр) – показники, які характеризують порівняльні розміри однойменних абсолютних показників, що належать до різних об'єктів або територій, але за однаковий період часу.

Відносні показники структури (ВПС) – показники, які характеризують частки (питомі ваги) складових сукупності в загальному її обсязі.

Середня величина – узагальнювальна характеристика ознаки, що вивчається, в досліджуваній сукупності, котра відображає її типовий рівень у розрахунку на одиницю сукупності в конкретних умовах місця і часу.

Статистичний показник – міра, тобто єдність якісного та кількісного відображення певної властивості соціально-економічного явища чи процесу.

Типові завдання

Завдання 5.1. Необхідно провести розрахунок абсолютних величин за такими даними: в Україні в січні 2017 р. було вироблено такі види рибних консервів: 100 тис. банок ємністю 240 см³; 250 тис. банок ємністю 160 см³; 75 тис. банок ємністю 480 см³. Переведіть вироблену продукцію в умовно-натуральні одиниці, якщо ємність умовної консервної банки складає 353,4 см³.

Розв'язання.

1. Визначимо коефіцієнти переведення:

$$\frac{240}{353,4} = 0,679, \quad \frac{160}{353,4} = 0,453, \quad \frac{480}{353,4} = 1,358.$$

2. Розрахуємо загальний обсяг виробництва рибних консервів:

$$100 \times 0,679 + 250 \times 0,453 + 75 \times 1,358 = 283 \text{ тис. ум. банок.}$$

Отже, у січні 2017 р. в Україні було вироблено 283 тис. умовних банок рибних консервів.

Завдання 5.2. За даними табл. 5.3 необхідно визначити відносні величини структури та координації.

Таблиця 5.3

Вихідні дані

Показники	Млн грн
Зовнішньоторговельний обіг, усього	765,5
експорт	356,2
імпорт	409,3

Обґрунтуйте отримані результати.

Розв'язання.

Відносна величина структури – це відношення частини до цілого, або питома вага частини одиниць у загальному обсязі сукупності (табл. 5.4).

Розрахункові дані

Показники	Млн грн
Зовнішньоторговельний обіг, усього	100,0
експорт	46,5
імпорт	53,5

Відносна величина координації характеризує відношення різних структурних одиниць тієї самої сукупності.

Так, на основі приведених вище даних ми можемо вчислити, що на кожний мільйон гривень експорту припадало 1,15 млн грн імпорту (409,3/356,2).

Завдання 5.3. Маємо дані про діаметр трьох труб: 20, 40, 60 см. Потрібно визначити середній розмір діаметра труб.

Розв'язання.

Оскільки вихідні дані подано у вигляді квадратних функцій, середній розмір діаметра труби визначимо за формулою середньої квадратичної.

$$\bar{x} = \sqrt{\frac{x^2}{n}} = \sqrt{\frac{20^2 + 40^2 + 60^2}{3}} = \sqrt{\frac{400 + 1\,600 + 3\,600}{3}} = \sqrt{\frac{5\,600}{3}} = \sqrt{1\,866,7} = 43,21 \text{ см.}$$

Отже, середній діаметр трьох труб дорівнює 43,21 см.

Завдання 5.4. Маємо дані зі "Звіту про чисельність персоналу" промислового підприємства міста (табл. 5.5).

Дані звітності

Квартали	Середня чисельність робітників, осіб
I	1 000
II	1 050
III	1 295
IV	1 820

Потрібно визначити середній коефіцієнт зростання чисельності робітників за квартал.

Розв'язання.

Для визначення середнього коефіцієнта зростання чисельності робітників обчислимо поквартальні коефіцієнти зростання як відношення кожного наступного рівня до попереднього. Отримаємо ланцюгові коефіцієнти зростання:

$$\text{II квартал порівняно з I кварталом: } K_1 = \frac{1050}{1000} = 1,05.$$

$$\text{III квартал порівняно з II кварталом: } K_2 = \frac{1295}{1050} = 1,23.$$

$$\text{IV квартал порівняно з III кварталом: } K_3 = \frac{1820}{1295} = 1,41.$$

Загальний обсяг явища в цьому випадку визначимо як добуток коефіцієнтів зростання. Таким чином, чисельність робітників за весь період (рік) збільшилася в 1,82 рази ($1,05 \times 1,23 \times 1,41$).

Середній квартальний темп зростання чисельності робітників обчислимо за формулою середньої геометричної, яку застосовують у тих випадках, коли загальний обсяг явища є не сумою, а добутком значень ознак:

$$\bar{K} = \sqrt[3]{K_1 \times K_2 \times K_3} = \sqrt[3]{1,05 \times 1,23 \times 1,41} = \sqrt[3]{1,821} = 1,221.$$

Отже, в середньому чисельність робочих за кожний квартал зростала в 1,221 рази, або на 22,1 %.

Такий результат можна дістати за іншою формулою обчислення середніх коефіцієнтів:

$$\bar{K} = \sqrt[n-1]{\frac{x_n}{x_1}},$$

де n – кількість періодів;

x_0 і x_n – початковий і кінцевий рівні ряду динаміки.

$$\bar{K} = \sqrt[n-1]{\frac{x_n}{x_1}} = \sqrt[3]{\frac{1820}{1000}} = 1,221.$$

Отримуємо той самий результат.

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Набуття навичок розрахунку відносних і середніх показників у MS Excel" подані у роботі [67]. Лабораторна робота націлена на ознайомлення з видами відносних і середніх величин та отримання практичних навичок їх розрахунку за допомогою пакету MS Excel.

Запитання для самодіагностики

1. У чому полягає сутність статистичних показників, їх функції та види?
2. Опишіть умови застосування абсолютних і відносних величин.
3. У чому полягає практичне застосування відносних і середніх величин?
4. Охарактеризуйте проблему гармонізації та зіставності показників у статистиці.
5. Розкрийте принцип використання математичних властивостей середніх величин для спрощення їх обчислення.
6. Зазначте особливості подання абсолютних статистичних величин.
7. Назвіть види відносних величин за їх аналітичною функцією.
8. Окресліть умови наукового застосування середніх величин.
9. Надайте порівняння одиниць вимірювання абсолютних і відносних величин.
10. Які аналітичні функції виконують відносні величини? Чи можна порівняти різнойменні показники?

Питання для критичного переосмислення (есе)

1. Поясніть сутність статистичного показника та його роль у статистичному аналізі.
2. Принципи узагальнювальних статистичних показників, які без зміни їх визначення забезпечують співвіднесеність із реальними соціально-економічними явищами та процесами.
3. Значення розрахунків відносних величин координації, порівняння, інтенсивності для проведення статистичного дослідження соціально-економічних процесів.
4. Змістовність взаємозв'язку між відносними величинами планового завдання, виконання плану, динаміки (наведіть приклади).
5. Види середніх, найчастіше використовуваних у статистичному аналізі. Принципи та критерії вибору виду середньої.

6. Аналіз рядів розподілу

Основні питання:

- 6.1. Закономірність розподілу та її частотні характеристики.
- 6.2. Характеристики центру розподілу.
- 6.3. Показники варіації.

6.1. Закономірність розподілу та її частотні характеристики

Результати статистичних зведень і угруповань можна подати у вигляді **статистичних рядів** – упорядкованих сукупностей значень аналізованого показника, тобто статистичної ознаки. За своєю змістовністю статистичні ряди розподіляють на ряди динаміки та ряди розподілу.

Рядом динаміки називають систематизовану сукупність числових даних, що характеризують зміни досліджуваних явищ у часі [68].

Ряд розподілу є систематизованою послідовністю статистичних одиниць, згрупованих за конкретною ознакою. Він характеризує склад досліджуваного явища, дозволяє судити про однорідність сукупності, закономірності розподілу статистичних одиниць. Зазвичай ряд розподілу є результатом **структурного групування** [68].

Ряд розподілу вважається побудованим, якщо відомо, яким чином змінюються в сукупності значення ознаки і як часто зустрічаються окремі значення ознаки.

Для різних статистичних ознак будують ряди розподілу різного типу:

- **атрибутивні** – будуються за описовими ознаками в порядку зростання або зменшення спостережуваних значень ознаки. Прикладом атрибутивних рядів можуть слугувати розподіли: населення – за національністю, за професіями, за статтю; підприємств – за формами власності;
- **варіаційні** – будуються за кількісними ознаками (наприклад: розподіл робітників за рівнем кваліфікації, за заробітною платою, студентів – за успішністю).

Варіаційні ряди поділяються на дискретні й інтервальні.

У **дискретних рядах** ознака приймає тільки цілі значення (наприклад, розмір сім'ї, тарифний розряд).

Інтервальні ряди засновані на безперервних ознаках, що приймають будь-які, в тому числі дробові значення. Залежно від того, яке структурне групування закладене в основу інтервального ряду, розрізняють

рівноінтервальні та нерівноінтервальні ряди. У рівноінтервальних рядах ширина інтервалу є величиною постійною, в нерівноінтервальних вона різна для різних груп [68].

Основними **елементами рядів розподілу** є:

1) значення ознаки (варіанти):

- X_i – дискретне в дискретних рядах;
- $X_i^B - X_i^H$ – інтервал для інтервальних рядів, де $i = 1 \div n_i$ – частота;

2) частота f_i – число одиниць сукупності, що володіють певним значенням ознаки. Частота показує, скільки разів це значення ознаки зустрічається у сукупності. Сума всіх частот завжди дорівнює обсягу статистичної сукупності, тобто $\sum f_i = N$.

Дослідження рядів розподілу здійснюється у два **етапи** [68; 81]:

- **емпіричне дослідження**, метою якого є отримання узагальнювальних характеристик досліджуваної сукупності;
- **теоретичне дослідження** з метою виявлення закономірності розглядуваного розподілу і його теоретичного опису.

Емпіричне дослідження починається з визначення частотних характеристик ряду розподілу.

Частотні характеристики рядів розподілу слід розглянути більш ретельно.

Вихідною частотною характеристикою будь-якого ряду розподілу є частота n_i . На її основі можна розрахувати такі характеристики:

- **частість** – питома вага (частка) одиниць сукупності, що мають певне значення ознаки. Тобто це частота, виражена у вигляді відносної величини (частки одиниці або відсотка):

$$d_i = \frac{f_i}{N}, \sum_{i=1}^m d_i = 1, \text{ де } i = \overline{1, m}. \quad (6.1)$$

Ця характеристика має важливе значення для дослідження рядів розподілу, тому що дозволяє зв'язати показники рядів розподілу з відповідними показниками та апаратом теорії ймовірностей. У теорії ймовірностей q_i є вірогідність того, що це значення ознаки зустрінеться в сукупності. Частість використовується для зіставлення рядів розподілу, що містять рівну кількість статистичних одиниць;

• **накопичена частота** – кількість одиниць сукупності, у яких значення ознаки не перевищує даного x^* , тобто це частота з наростаючим підсумком:

$$\sum_{i=1}^{m^*} f_i = S_{x^*}; \quad S_{x^*} = N_{x^*}, \quad (6.2)$$

де x^* – це значення ознаки в i -ї групі, для якої розраховується накопичена частота.

За накопиченими частотами можна побудувати *кумулятивний ряд розподілу* – ряд значень числа одиниць сукупності зі значеннями ознаки, меншими та дорівненими верхній межі відповідного інтервалу;

• **накопичена частість** – питома вага (частка) одиниць, у яких значення ознаки не перевершує даний x^* , тобто це частість із наростаючим підсумком:

$$Q_{x^*} = \sum_{i=1}^{m^*} f_i = 1; \quad (6.3)$$

• **щільність розподілу** – універсальна частотна характеристика, що дозволяє перейти від емпіричного до теоретичного розподілу. Для рядів з нерівними інтервалами тільки ця характеристика дає правильне уявлення про характер розподілу. Вона розраховується у двох варіантах:

як абсолютна щільність розподілу φ_i , що показує число одиниць сукупності, що припадають на одиницю ширини інтервалу значення ознаки:

$$\varphi_i = \frac{f_i}{h_i}; \quad (6.4)$$

як відносна щільність розподілу φ'_i , що показує питому вагу одиниць сукупності, що припадають на одиницю ширини інтервалу:

$$\varphi'_i = \frac{d_i}{h_i}. \quad (6.5)$$

Щільність розподілу забезпечує порівнянність різних рядів розподілу. Різні ряди розподілу характеризуються різним набором частотних характеристик: мінімальним – для атрибутивних рядів (частота f_i і частість d_i), для дискретних використовують чотири характеристики (частота n_i , частість q_i , накопичена частота S_i , накопичена частість Q_i), для інтервальних – всі п'ять (частота f_i , частість q_i , накопичена частота S_i , накопичена частість Q_i ; абсолютна φ_i і відносна φ_i' щільності розподілу).

Графіки є наочною формою відображення рядів розподілу. Для зображення рядів застосовуються *лінійні графіки* та *діаграми площинні*, побудовані в прямокутній системі координат.

Для графічного подання *атрибутивних рядів* розподілу використовуються різні діаграми: стовпчикові, лінійні, колові, фігурні, секторні тощо.

Для *дискретних варіаційних рядів* графіком є полігон розподілу.

Полігоном розподілу називають ламану лінію, що з'єднує точки з координатами $(x_i; f_i)$ або $(x_i; d_i)$, де x_i – дискретне значення ознаки, f_i – частота, d_i – частість [81].

Графік будується в прийнятому масштабі. Вигляд полігону розподілу наведено на рис. 6.1.

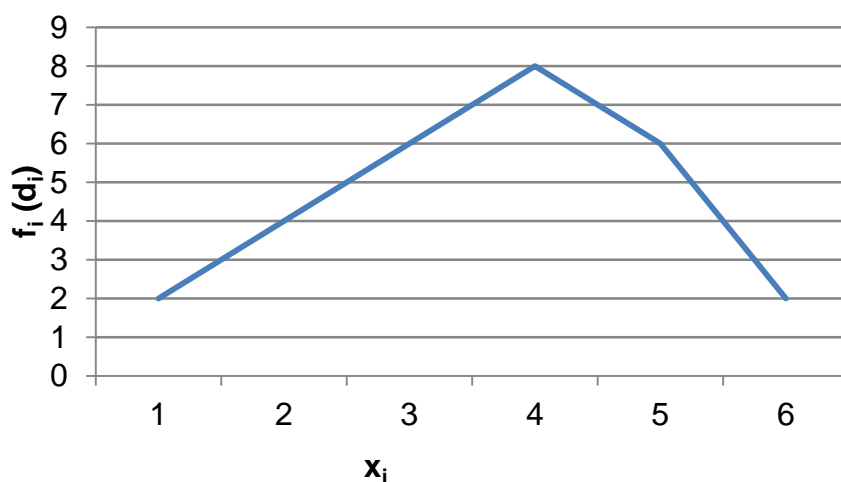


Рис. 6.1. Полігон розподілу

Для відображення **інтервальних варіаційних рядів** застосовують гістограми – ступінчасті фігури, що складаються з прямокутників. Основи цих прямокутників дорівнюють ширині інтервалу h_i , а висота – частоті f_i

(частоті d_i) рівноінтервального ряду чи щільності розподілу нерівноінтервального φ_i, φ_i' . Побудова діаграми аналогічна побудові стовпчикової діаграми. Загальний вигляд гістограми наведено на рис. 6.2.

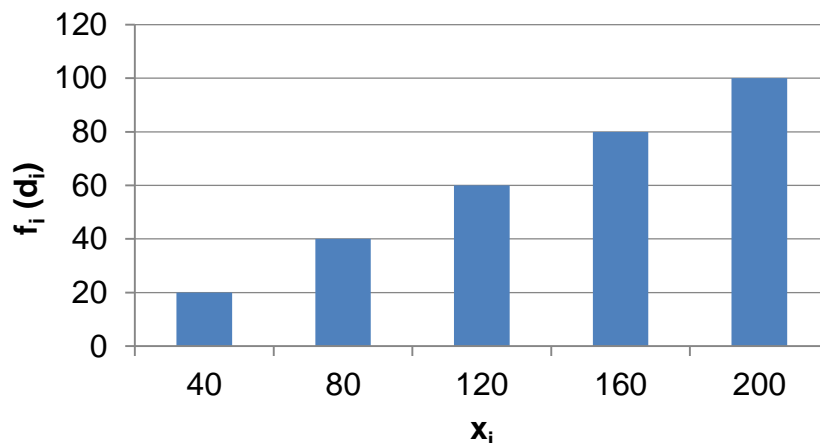


Рис. 6.2. Гістограма розподілу

У формі гістограми можна побудувати полігон розподілу, якщо за відповідні точки взяти середину виділених інтервалів.

Для графічного подання варіаційних рядів можна використовувати **кумуляту** – ламану лінію, складену з накопичених частот (частостей). Накопичені частоти наносять у вигляді ординат, поєднуючи вершини окремих ординат відрізками прямої. Створюється ламана лінія, що має неспадний вид. Координатами точок на графіку для дискретного ряду є $\{x_i; N_i\}$; для інтервального ряду – $\{X_i^B; N_i\}$. Початкова точка графіка має координати $(x_1^H; 0)$; найвища точка – $(x_m^B; N)$ Загальний вигляд кумуляти наведено на рис. 6.3. Використання кумуляти особливо зручно для проведення порівнянь варіаційних рядів [19; 81].

Для побудови графіків рядів розподілу велике значення має співвідношення масштабів на осі абсцис і осі ординат. У цьому випадку необхідно керуватися "правилом золотого перетину", відповідно до якого висота графіка повинна бути приблизно у два рази меншою від його основи.

Аналіз закономірностей розподілу передбачає оцінювання ступеня однорідності сукупності, асиметрії і ексцесу розподілу.

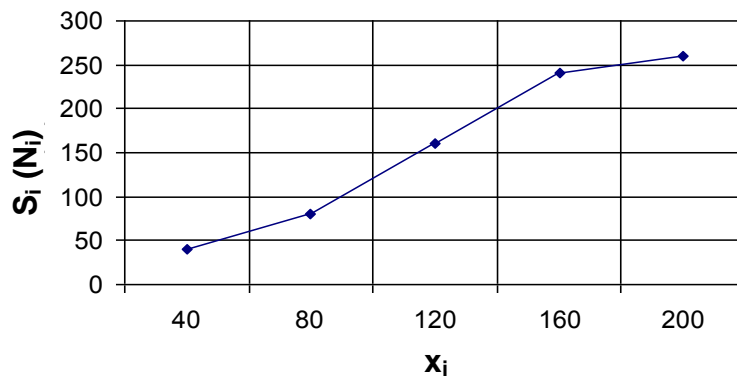


Рис. 6.3. Графік кумуляти

Однорідність сукупності – це передумова використання інших статистичних методів (середніх величин, регресійного аналізу тощо). *Однорідними* вважаються такі сукупності, елементи яких мають загальні властивості та належать до одного типу, класу. Водночас однорідність не означає повну тотожність властивостей елементів, а лише наявність у них спільного в суттєвому, головному. Критерієм однорідності сукупності вважається квадратичний коефіцієнт варіації, який завдяки властивостям у симетричному розподілі становить $V_{\sigma} = 0,33$.

Встановлення асиметрії та ексцесу дозволяє виявити симетричність розподілу випадкової величини.

З'ясування загального характеру розподілу передбачає не тільки оцінювання ступеня його однорідності, а й дослідження форми розподілу, тобто оцінювання симетричності й ексцесу.

6.2. Характеристики центру розподілу

У проведенні емпіричного дослідження ряду розподілу розраховують і аналізують такі **групи показників**:

- показники положення центру розподілу;
- показники ступеня однорідності ряду розподілу;
- показники форми розподілу.

До **показників положення центру розподілу** відносять статистичну середню у вигляді середньої арифметичної і структурні середні – мода та медіана.

Середня арифметична для дискретного ряду розподілу розраховується за формулою [68; 81]:

$$\bar{x} = \sqrt[k]{\frac{\sum_{i=1}^m x_i^k f_i}{\sum_{i=1}^m f_i}}, \quad (6.6)$$

де x_i – варіанти значень ознаки;

f_i – частота повторення індивідуальних значень ознаки;

m – кількість однорідних груп.

У інтервальному варіаційному ряду середня арифметична визначається за формулою:

$$\bar{x} = \frac{\sum_{i=1}^m x_i f_i}{\sum_{i=1}^m f_i}, \quad (6.7)$$

де x_i – середина відповідного інтервалу.

На відміну від середньої арифметичної, що розраховується на основі всіх варіант, мода та медіана характеризують значення ознаки у статистичній одиниці, що займає певне положення у варіаційному ряду.

Медіана (Me) – значення ознаки у статистичній одиниці, що стоїть у середині рангового ряду, розділяючи сукупність на дві рівні за чисельністю частини [68].

Мода (Mo) – значення ознаки, що найчастіше зустрічається в сукупності. Мода широко використовується в статистичній практиці для вивчення купівельного попиту, реєстрації цін та ін. [68].

Для дискретних варіаційних рядів Mo та Me вибирають відповідно до визначень: моду – як значення ознаки з найбільшою частотою f_i ; положення медіани за непарного обсягу сукупності визначається її номером

$n_{Me} = \frac{N+1}{2}$, де N – обсяг статистичної сукупності. За парного обсягу

та ряду медіана дорівнює середній з двох варіантів, що знаходяться в середині ряду.

Медіану використовують як найбільш надійний показник типового значення неоднорідної сукупності, оскільки вона нечутлива до крайніх значень ознаки, які можуть значно відрізнитися від основного масиву значень показника. Крім цього медіана знаходить практичне застосування внаслідок особливої математичної властивості: $\sum |x_i - Me| \rightarrow \min$.

У **інтервальному ряду** значення Mo та Me обчислюються більш складним шляхом. Мода визначається таким чином:

- за максимального значення частоти визначається інтервал, в якому знаходиться значення моди. Його називають модальним.
- усередині модального інтервалу значення моди обчислюється за формулою:

$$Mo = x_{Mo}^i + h_{Mo} \times \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} + f_{Mo+1})}, \quad (6.8)$$

де x_{Mo}^i – нижня межа модального інтервалу;

h_{Mo} – ширина модального інтервалу;

f_{Mo} , f_{Mo-1} , f_{Mo+1} – відповідно, частоти модального, передмодального (що передує модальному) та постмодального інтервалів.

Для розрахунку медіани в інтервальних рядах використовуються такі підходи:

- за накопиченими частотами знаходять медіанний інтервал. Медіанним називають інтервал, що містить центральну одиницю сукупності;
- усередині медіанного інтервалу значення Me визначається за формулою:

$$Me = x_{Me}^i + h_{Me} \times \frac{\frac{N}{2} - S_{Me-1}}{f_{Me}}, \quad (6.9)$$

де x_{Me}^i – нижня межа медіанного інтервалу;

h_{Me} – ширина медіанного інтервалу;

N – обсяг статистичної сукупності;

S_{Me-1} – накопичена частота передмедіанного інтервалу;

f_{Me} – частота медіанного інтервалу.

У нерівноінтервальних рядах для обчислення Mo використовують іншу частотну характеристику – **абсолютна щільність розподілу**:

$$Mo = x_{Mo}^H + h_{Mo} \times \frac{\varphi_{Mo} - \varphi_{Mo-1}}{(\varphi_{Mo} - \varphi_{Mo-1}) + (\varphi_{Mo} - \varphi_{Mo+1})}, \quad (6.10)$$

де φ_{Me} – абсолютна щільність розподілу модального інтервалу;

φ_{Mo-1} – абсолютна щільність розподілу перед модального інтервалу;

φ_{Mo+1} – абсолютна щільність розподілу після модального інтервалу.

Моду можна визначити графічно: за полігоном розподілу в дискретних рядах, за гістограмою розподілу – в інтервальних; медіану – за кумулятою.

Для знаходження моди в інтервальному ряду праву вершину модального прямокутника треба з'єднати з правим верхнім кутом попереднього прямокутника, а ліву вершину – з лівим верхнім кутом подальшого прямокутника. Абсциса точки перетину цих прямих і буде модою розподілу.

Для визначення медіани висоту найбільшої ординати кумуляти, що відповідає загальній чисельності сукупності, ділять навпіл. Через отриману точку проводять пряму, паралельну осі абсцис, до перетину її з кумулятою. Абсциса точки перетину є медіаною.

Отже, для характеристики положення центру ряду розподілу можна використовувати три показники: середнє значення ознаки, моду, медіану. Для вибору виду та форми конкретного показника центру розподілу необхідно виходити з таких рекомендацій:

- для стійких соціально-економічних процесів у якості показника центру використовують середню арифметичну. Такі процеси характеризуються симетричними розподілами, в яких $x_{cp} = Mo = Me$;

- для нестійких процесів положення центру розподілу характеризується за допомогою Mo або Me . Для асиметричних процесів переважною

характеристикою центру розподілу є медіана, оскільки займає положення між середньою арифметичною і модою.

6.3. Показники варіації

Після встановлення середньої величини ($x_{\text{ср}}$, M_o , M_e) виникає питання, у якій мірі індивідуальні значення ознаки відрізняються між собою і від середньої їх величини.

Варіацією ознаки називають різницю в числових значеннях одиниць сукупності та їх коливання поблизу середньої величини, що характеризує цю сукупність. Чим менша варіація, тим більш однорідною є сукупність і тим більш надійною є середня величина [68; 81].

Для вимірювання варіації в статистиці використовують абсолютні та відносні показники.

До абсолютних показників варіації відносяться:

- розмах варіації R ;
- середнє лінійне відхилення \bar{I} ;
- середній квадрат відхилень (дисперсія) σ^2 ;
- середньоквадратичне відхилення σ .

Розмах варіації R є найбільш простим показником варіації, розраховуваним за формулою:

$$R = x_{\text{max}} - x_{\text{min}}. \quad (6.11)$$

Цей показник становить різницю між максимальним і мінімальним значеннями ознак і характеризує розкид (розмах) елементів сукупності. Розмах вловлює тільки крайні значення ознаки в сукупності, не враховуючи повторюваність його проміжних значень, а також не відображає відхилень всіх варіантів значень ознаки.

Розмах часто використовується в практичній діяльності (наприклад, відмінність між max і min пенсій, заробітної плати в різних галузях).

Середнє лінійне відхилення (\bar{I}) ураховує відмінності всіх одиниць досліджуваної сукупності. Середнє лінійне відхилення є середньою арифметичною абсолютних значень відхилень окремих варіантів від їх середньої

арифметичної [68]. Цей показник розраховується за формулами простої і зваженої середньої арифметичної:

$$\text{для незгрупованих даних} - \bar{I} = \frac{\sum_{i=1}^N |x_i - \bar{X}|}{N};$$

$$\text{для згрупованих даних} - \bar{I} = \frac{\sum_{i=1}^m |x_i - \bar{X}| \times f_i}{\sum_{i=1}^m f_i},$$

де x_i – середина відповідного інтервалу;

\bar{X} – середнє значення за усією сукупністю.

У практичних розрахунках середнє лінійне відхилення використовується для оцінювання ритмічності виробництва та рівномірності поставок.

Оскільки модулям притаманні погані математичні властивості, то на практиці часто застосовують інші показники середнього відхилення від середньої – дисперсію і середньоквадратичне відхилення.

Дисперсія ознаки σ^2 – середній квадрат відхилень варіантів від їх середньої величини; є загальноприйнятою мірою варіації. Залежно від вихідних даних дисперсія обчислюється за формулами простої і зваженої середньої арифметичної:

$$\text{для незгрупованих даних} - \sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N};$$

$$\text{для згрупованих даних} - \sigma^2 = \frac{\sum_{i=1}^m (x_i - \bar{X})^2 \times f_i}{\sum_{i=1}^m f_i}.$$

У разі використання зваженої середньої для розрахунку дисперсії в інтервальних рядах розподілу як варіанти значень ознаки обирають серединні значення x_i (середини інтервалів), які не є середнім значенням у групі. У результаті отримують наближене значення дисперсії.

Існують простіші підходи в обчисленні дисперсії. Найбільш часто використовується скорочений спосіб розрахунку дисперсії (метод моментів),

відповідно до якого дисперсія σ^2 є різницею між середнім із квадратів значень ознаки x^2 і квадратом їх середньої (\bar{x}^2):

$$\sigma^2 = \bar{x}^2 - (\bar{x})^2, \quad (6.12)$$

де $\bar{x}^2 = \frac{\sum x_i^2}{N}$ – для незгрупованих даних;

$\bar{x}^2 = \frac{\sum x_i^2 \cdot f_i}{\sum f_i}$ – для згрупованих даних.

Цей спосіб дозволяє провести обчислення дисперсії за вихідними даними без попереднього розрахунку відхилень.

Розмірність дисперсії відповідає квадрату розмірності досліджуваної ознаки, тому цей показник не має економічної інтерпретації. Для збереження економічного сенсу розраховують ще один показник варіації – середнє квадратичне відхилення.

Середнє квадратичне відхилення σ – середня квадратична з відхилень окремих значень ознаки від їх середньої арифметичної [68; 81]:

$$\text{а) для незгрупованих даних} - \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (x - \bar{x})^2}{n}}; \quad (6.13)$$

$$\text{б) для згрупованих даних} - \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^m (x - \bar{x})^2 f_i}{\sum f_i}}.$$

Середнє квадратичне відхилення є іменованою величиною, має розмірність усередненої ознаки, економічно добре інтерпретується. Воно також використовується для оцінювання надійності середньої: чим менше середнє квадратичне відхилення σ , тим надійніше середнє значення ознаки x , тим краще середня характеризує досліджувану сукупність.

Для розподілів, близьких до нормального розподілу, між середньо-квадратичним відхиленням і середнім лінійним відхиленням існує наступна залежність: $\sigma \approx 1,25 \times \bar{I}$.

Середнє квадратичне відхилення грає важливу роль в аналізі рядів розподілу. В умовах нормального розподілу існує наступна залежність між величиною середнього квадратичного відхилення і кількістю спостережень: у межах $x \pm 1\sigma$ лежать 0,683 (або 68,3 %) кількості спостережень; у межах $x \pm 2\sigma$ – 0,954 або 95,4 % спостережень; в межах $x \pm 3\sigma$ знаходяться 0,997 (або 99,7 %) спостережень. На практиці не зустрічаються відхилення, які перевищують 0,9973, це називають *правилом 3 σ* (трьох сигм).

Якщо в якості показника центру розподілу використовується медіана, то для характеристики варіації можна застосовувати так зване кватильне відхилення [68]:

$$Q = \frac{Q_3 - Q_1}{2 Me}, \quad (6.14)$$

де Q_1, Q_3 – відповідно, перший і третій кватилі розподілу.

Цей показник можна застосовувати замість розмаху варіації, щоб усунути недоліки, пов'язані з використанням крайніх значень ознаки.

Між середнім квадратичним відхиленням, середнім лінійним відхиленням, кватильним відхиленням, розмахом варіації в нормально розподіленій сукупності існує таке співвідношення:

$$6\sigma \approx 7,5\bar{I} \approx 9Q \approx R \quad \text{або} \quad \sigma \approx 1,25\bar{I} \approx 1,5Q \approx 1/6R. \quad (6.15)$$

Відносні показники варіації призначені для оцінювання і порівняння варіації сукупностей, які мають різні одиниці вимірювання або різні значення середніх величин. Ці показники розраховують як відношення абсолютних показників варіації до середньої арифметичної (медіани).

Найпоширенішим відносним показником варіації є коефіцієнт варіації V_σ – відношення середнього квадратичного відхилення до середньої арифметичної, виражене у відсотках:

$$V_\sigma = \frac{\sigma}{\bar{x}} \times 100 \%. \quad (6.16)$$

Коефіцієнт варіації використовується для характеристики однорідності досліджуваної сукупності. Статистична сукупність вважається кількісно однорідною, якщо коефіцієнт варіації не перевищує 33 %.

Крім коефіцієнта варіації як відносних показників оцінки варіабельності рядів розподілу виступають:

$$\text{коефіцієнт осциляції: } V_R = \frac{R}{\bar{X}} \times 100 \%$$

$$\text{відносне лінійне відхилення: } V_d = \frac{\bar{d}}{\bar{X}} \times 100 \%$$

відносне кватильне відхилення:

$$V_Q = \frac{Q}{\bar{X}} \times 100 \% \text{ або } V_Q = \frac{Q}{Me} \times 100 \%, \text{ або } V_Q = \frac{Q_3 - Q_1}{2Me}$$

Дисперсійний аналіз (від лат. dispersio – розсіювання) – статистичний метод, що дозволяє аналізувати вплив різних факторів на досліджувану змінну. Метод, розроблений біологом Р. Фішером у 1925 році і спочатку застосовувався для оцінювання експериментів у рослинництві. У подальшому з'ясувалася загальнонаукова значущість дисперсійного аналізу для експериментів у психології, педагогіці, медицині та ін. [68].

Метою дисперсійного аналізу є перевірка значущості відмінності між середніми за допомогою порівняння дисперсій. Дисперсію вимірюваної ознаки розкладають на незалежні складові, кожна з яких характеризує вплив того чи іншого фактора або їх взаємодії. Подальше порівняння таких складових дозволяє оцінити значущість кожного досліджуваного фактора, а також їх комбінації.

У разі суттєвості нульової гіпотези (про рівність середніх за декількома групами спостережень, вибраних з генеральної сукупності) оцінка щодо дисперсії, пов'язаної з внутрішньогруповою мінливістю, повинна бути близькою до оцінки міжгрупової дисперсії.

Дисперсія та її властивості є базовими поняттями дисперсійного аналізу, тому зупинимося на їх означеннях більш детально.

Дисперсія (англ. variance), або центральний момент другого порядку, є мірою відхилення значень випадкової величини від середнього значення (середньої величини). Більші значення дисперсії свідчать про більші відхилення значень випадкової величини від середнього. Якщо

дисперсія дорівнює 0, то всі реалізації випадкової величини знаходяться в одній точці.

Дисперсія як базовий показник варіації має ряд математичних властивостей, які допомагають зрозуміти її природу та спростити її розрахунок.

До властивостей дисперсії відносять такі [68; 81]:

1) дисперсія постійної величини дорівнює нулю:

$$\sigma_{\text{Aconst}}^2 = 0. \quad (6.17)$$

Ця властивість заснована на тому, що дисперсія є показником розсіювання варіант близько середньої арифметичної, а середня арифметична постійної величини дорівнює нулю;

2) дисперсія не змінюється, якщо всі варіанти збільшити або зменшити на те саме число A:

$$\sigma_{(x_i - A)}^2 = \sigma^2. \quad (6.18)$$

Це означає, що дисперсію можна розрахувати не за даними значень ознаки, а щодо відхилення від якого-небудь постійного числа;

3) якщо всі варіанти поділити (помножити) на число A, то дисперсія зменшиться (збільшиться) в A^2 раз, середнє квадратичне відхилення зміниться в A разів:

$$\sigma_{x/A}^2 = \sigma^2 / A. \quad (6.19)$$

Це означає, що всі значення ознаки можна поділити на постійне число (наприклад, на величину інтервалу), обчислити середньоквадратичне відхилення, а потім помножити його на постійне число;

4) якщо обчислити середній квадрат відхилень якої-небудь величини A в тій чи іншій мірі відмінності від середньої арифметичної (\bar{x}), то він завжди буде більше середнього квадрата відхилень, обчисленого від середньої арифметичної:

$$\sigma_A^2 > \sigma^2. \quad (6.20)$$

При чому він збільшений на певну величину – квадрат різниці між середнім значенням (середньою величиною) і цією умовною величиною, тобто на:

$$(\bar{x} - A)^2, \quad \sigma_A^2 = \sigma^2 + (\bar{x} - A)^2$$

або

$$\sigma_A^2 = \sigma^2 + (\bar{x} - A)^2 = \frac{\sum_{i=1}^m (\bar{x} - A)^2 f_i}{\sum_{i=1}^m f_i} + (\bar{x} - A)^2. \quad (6.21)$$

Це означає, що дисперсія від середнього значення (середньої величини) завжди менше дисперсій, обчислених від будь-яких довільних величин, тобто вона має властивість мінімальності;

5) якщо ряд спостережень складається з двох груп спостережень, то дисперсія всього ряду дорівнює сумі середньої арифметичної групових дисперсій і середньої арифметичної квадратів відхилень групових середніх від середньої всього ряду, причому в обчисленні середніх арифметичних вагами є обсяги груп.

Розрізняють такі види дисперсій: *загальна*, міжгрупова і внутрішньогрупова.

Загальна дисперсія (σ^2) вимірює варіацію ознаки у всій статистичній сукупності під впливом усіх факторів, що викликають цю варіацію [68]. Вона розраховується за формулою:

$$\sigma^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 f_i}{\sum_{i=1}^m f_i}. \quad (6.22)$$

Міжгрупова дисперсія (δ^2) характеризує зміну ознаки зумовлену чинниками, покладеними в основу групування. Отже, міжгрупова дисперсія є дисперсією локальних середніх. Її розрахунок проводиться за формулою [68]:

$$\delta^2 = \frac{\sum_{i=1}^m (\bar{x}_i - \bar{x})^2}{m}, \quad (6.23)$$

де \bar{x}_i – локальна середня в кожній групі;

m – кількість груп у сукупності.

Внутрішньогрупова дисперсія характеризує випадкову варіацію, тобто коливання ознаки, що виникають під впливом неврахованих факторів і незалежної від варіації ознаки – фактора, покладеного в основу групування [68]. Внутрішньогрупова дисперсія розраховується для кожної однорідної групи:

$$\sigma_i^2 = \frac{\sum_j^{m_i} (x_i - \bar{x}_i)}{f_i}. \quad (6.24)$$

На підставі внутрішньогрупової дисперсії розраховується середня із внутрішньогрупових дисперсій (залишкова) $\overline{\sigma_i^2}$:

$$\overline{\sigma_i^2} = \frac{\sum_{i=1}^{m_i} \sigma_i^2}{m}. \quad (6.25)$$

Ці види дисперсій пов'язані між собою таким відношенням:

$$\sigma^2 = \delta^2 + \overline{\sigma_i^2}. \quad (6.26)$$

Зазначене співвідношення називають *правилом додавання дисперсій*. Очевидно, що, чим більше величина міжгрупової дисперсії, тим більш якісно проведене групування, тим сильніше факторна ознака впливає на загальну варіацію. Крім цього, користуючись зазначеним правилом, можна за двома відомими дисперсіями розрахувати невідому третю дисперсію.

Оцінювання щільного зв'язку ґрунтується на *правилі складання дисперсій*. У моделі аналітичного групування мірою щільного зв'язку є відношення міжгрупової дисперсії до загальної, яке називають **кореляційним відношенням** [68; 81]:

$$\eta^2 = \frac{\delta^2}{\sigma^2}, \quad (6.27)$$

де η^2 – кореляційне відношення;

δ^2 – міжгрупова дисперсія;

σ^2 – загальна дисперсія.

Також виділяють *дисперсію альтернативної ознаки*. Серед ознак, що варіюють, зустрічаються ознаки, варіація яких виявляється в тому, що в одних одиниць сукупності вони зустрічаються, а в інших – ні. Ознаки, якими володіють одні об'єкти, але не володіють інші називаються *альтернативними ознаками*. Відсутність альтернативної ознаки приймають за 0, а її наявність за – 1. Частка об'єктів, що володіють ознакою в чисельності всієї сукупності, позначається P , а частка одиниць, без ознаки – Q , якщо $p + q = 1$.

Дисперсія альтернативної ознаки розраховується за формулою [68; 81]:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f} = \frac{(1-p)^2 p + (0-q)^2 q}{p+q} = \frac{pq(q+p)}{p+q} = pq, \quad (6.28)$$

де $x = \frac{\sum x f}{f} = \frac{1 \times p + 0 \times q}{p+q}$.

Дисперсійний аналіз є одним з основних методів статистичного оцінювання результатів експерименту. Широке застосування він отримує в аналізі економічної інформації. Перевагою цього методу є те, що він дає досить надійні висновки щодо вибірок досить невеликої чисельності.

Досліджуючи варіацію результативної ознаки під впливом одного або кількох факторів, за допомогою дисперсійного аналізу можна отримати (крім загальних оцінок істотності залежностей) оцінку відмінностей у величині середніх, що формуються за різних рівнів факторів, та істотності взаємодії факторів. Дисперсійний аналіз застосовується для вивчення залежностей як кількісних, так і якісних ознак, а також їх поєднання.

Сутність цього методу полягає в статистичному вивченні вірогідності впливу одного або кількох факторів, а також їх сукупного впливу на результативну ознаку [68]. Відповідно до цього за допомогою дисперсійного аналізу вирішуються три основні завдання:

- 1) загальне оцінювання істотності відмінностей між груповими середніми;
- 2) оцінювання вірогідності взаємодії факторів;
- 3) оцінювання істотності відмінностей між парами середніх.

Найчастіше такі завдання вирішують у процесі вивчення впливу кількох факторів на результативну ознаку.

Проведення дисперсійного аналізу передбачає виконання таких умов: результати спостережень вихідного параметра системи Y є незалежними випадковими величинами, що мають нормальний закон розподілу; випадкові помилки спостережень підкоряються нормальному закону розподілу;

вихідні досліджувані фактори x_i впливають лише на зміну середніх значень, а дисперсія спостережень залишається постійною;

експерименти мають проходити з однаковою точністю.

Перевірка зазначених умов перед проведенням дисперсійного аналізу обов'язкова.

Принципова схема дисперсійного аналізу включає: встановлення основних джерел варіювання результативної ознаки та визначення обсягів варіації (сум квадратів відхилень) за джерелами її утворення; визначення числа ступенів свободи, що відповідають компонентам загальної варіації; обчислення *дисперсія як відношення відповідних обсягів* варіації до їх числа ступенів свободи; аналіз співвідношень між дисперсіями; оцінювання вірогідності різниці між середніми; формулювання висновків. Зазначена схема зберігається як для простих моделей дисперсійного аналізу, коли дані групуються за однією ознакою, так і щодо складних, коли дані групуються за двома та більшим числом ознак. Однак із збільшенням числа групувальних ознак ускладнюється процес розкладання загальної варіації за джерелами її утворення. Відповідно до принципової схеми дисперсійний аналіз можна подати у вигляді п'яти послідовно виконуваних етапів (рис. 6.4) [68].

В основу розв'язування задач дисперсійного аналізу закладений *закон розкладання (додавання) варіації*. Відповідно до нього загальна варіація (коливання) результативної ознаки розподіляється на дві: варіацію, зумовлену дією досліджуваного фактора (факторів); варіацію, викликану дією випадкових причин. Щоб оцінити вірогідність відмінностей між груповими середніми, необхідно визначити міжгрупову та внутрішньогрупову варіації. Якщо міжгрупова (факторна) варіація значно перевищує внутрішньогрупову (залишкову) варіацію, то фактор впливає на результативну ознаку, істотно змінюючи значення групових середніх величин.

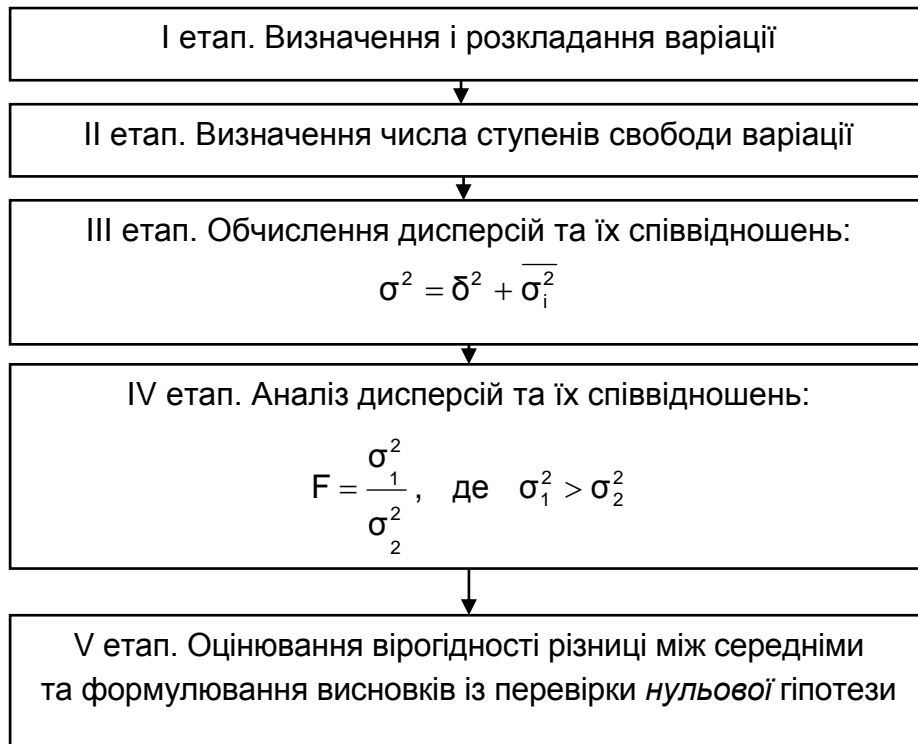


Рис. 6.4. Етапи дисперсійного аналізу

Для оцінювання істотності відмінностей між середніми та формулювання висновків із перевірки нульової гіпотези ($H_0 : \bar{x}_1 = \bar{x}_2 = \dots = \bar{x}_n$) у дисперсійному аналізі використовується своєрідний норматив – ***F-критерій***, закон розподілу якого встановив Р. Фішер [68; 74; 81]. Цей критерій є відношенням двох дисперсій: факторної, породжуваної дією досліджуваного фактора, та залишкової, зумовленої дією випадкових причин:

$$F = \frac{\sigma_1^2}{\sigma_2^2}, \quad \text{де } \sigma_1^2 > \sigma_2^2. \quad (6.29)$$

Дисперсії σ_1^2 і σ_2^2 є оцінками дисперсії генеральної сукупності. Якщо вибірки з дисперсіями σ_1^2 і σ_2^2 зроблені з тієї самої генеральної сукупності, де варіація величин мала випадковий характер, то розбіжність у величинах σ_1^2 і σ_2^2 також випадкова.

Для обчислення фактичного значення F-критерію в чисельнику береться більша з дисперсій, тому $F > 1$. Очевидно, що чим більше критерій F , тим значнішим є розбіжності між дисперсіями. Якщо $F = 1$, то необхідність про оцінювання істотності відмінностей дисперсій відкидається.

Якщо $F_{\text{табл}}$ менше $F_{\text{факт}}$, то відмінності вибірових дисперсій визначається не лише випадковими факторами; вони істотні. Нульову гіпотезу ($H_0 : \bar{X}_1 = \bar{X}_2 = \dots = \bar{X}_n$) у цьому випадку відхиляють; є підстави стверджувати, що дія фактора на результативну ознаку не є істотною. Критерій F дозволяє встановити наявність або відсутність істотних зв'язків між груповими середніми в цілому.

У ході оброблення методом дисперсійного аналізу соціально-економічних даних необхідно мати на увазі, що через багату чисельність факторів та їх взаємозв'язок важко навіть за ретельного вирівнювання умов установити ступінь об'єктивного впливу кожного окремого фактора на результативну ознаку [68]. Тому рівень залишкової варіації зумовлюється не тільки випадковими причинами, але і суттєвими факторами, які не були враховані в побудові моделі дисперсійного аналізу. Унаслідок цього залишкова дисперсія, як база порівняння, інколи стає не адекватною своєму призначенню, якщо вона вища на величину та не може виступати як критерій сутності впливу факторів. У зв'язку із цим у побудові моделей дисперсійного аналізу актуалізується проблема відбору найважливіших факторів та вирівнювання умов для проявлення дії кожного з них. Крім того, застосування дисперсійного аналізу передбачає нормальний або близький до нормального розподіл досліджуваних статистичних сукупностей. Якщо ця умова не дотримується, то оцінки, отримані в дисперсійному аналізі, будуть перебільшеними.

Важливі поняття

Атрибутивні ряди розподілу – ряди розподілу, побудовані за описовими ознаками в порядку зростання або зменшення спостережуваних значень ознаки.

Варіаційні ряди розподілу – ряди розподілу, які будуються за кількісними ознаками.

Варіація ознаки – різниця в числових значеннях одиниць сукупності та їх коливання поблизу середньої величини, що характеризує цю сукупність.

Внутрішньогрупова дисперсія – дисперсія, яка характеризує випадкову варіацію, тобто коливання ознаки, що виникають під впливом неврахованих факторів і незалежної від варіації ознаки – фактора, покладеного в основу групування.

Дисперсійний аналіз – статистичний метод, що дозволяє аналізувати вплив різних факторів на досліджувану змінну.

Дисперсія, або центральний момент другого порядку – міра відхилення значень випадкової величини від середнього значення (середньої величини). Значні значення дисперсії свідчать про значні відхилення значень випадкової величини від середнього значення. Якщо дисперсія дорівнює 0, то всі реалізації випадкової величини знаходяться в одній точці.

Дисперсія альтернативної ознаки – дисперсія, яка характеризує варіацію ознаки, якою володіють одні об'єкти та не володіють інші.

Дисперсія ознаки σ^2 – середній квадрат відхилень варіантів від їх середньої величини; є загальноприйнятою мірою варіації.

Загальна дисперсія – дисперсія, яка вимірює варіацію ознаки в усій статистичній сукупності під впливом усіх факторів, що викликають цю варіацію.

Медіана (Me) – значення ознаки у статистичної одиниці, що стоїть у середині рангового ряду та розділяє сукупність на дві рівні за чисельністю частини.

Міжгрупова дисперсія – дисперсія, яка характеризує зміну ознаки, обумовлену факторами, покладеними в основу групування. Таким чином, міжгрупова дисперсія є дисперсією локальних середніх.

Мода (Mo) – значення ознаки в сукупності, яке зустрічається найчастіше.

Накопичена частість – питома вага (частка) одиниць, у яких значення ознаки не перевершує даний x^* . Тобто це частість із наростаючим підсумком.

Накопичена частота – кількість одиниць сукупності, у яких значення ознаки не перевищує даного x^* . Тобто це частота з наростаючим підсумком.

Однорідні сукупності – сукупності, елементи яких мають загальні властивості та належать до одного типу, класу.

Однорідність сукупності – передумова використання інших статистичних методів (середніх величин, регресійного аналізу тощо).

Ряд динаміки – систематизована сукупність числових даних, що характеризують зміни досліджуваних явищ у часі.

Ряд розподілу – систематизована послідовність статистичних одиниць, згрупованих за конкретною ознакою.

Середнє квадратичне відхилення – середня квадратична з відхилень окремих значень ознаки від їх середньої арифметичної.

Середнє лінійне відхилення – характеристика варіації ознаки, що враховує відмінності всіх одиниць досліджуваної сукупності.

Статистичний ряд – упорядкована сукупність значень аналізованого показника, тобто статистичної ознаки.

Частість – питома вага (частка) одиниць сукупності, що мають певне значення ознаки. Тобто це частота, виражена у вигляді відносної величини (частки одиниці або відсотка).

Щільність розподілу – універсальна частотна характеристика, що дозволяє перейти від емпіричного до теоретичного розподілу.

Типові завдання

Завдання 6.1. Маємо дані про розподіл верстатів на підприємствах машинобудівного комплексу за терміном експлуатації (табл. 6.1).

Таблиця 6.1

Вихідні дані

Групи за терміном експлуатації	Кількість верстатів	Накопичені частоти
1 – 3	43	43
3 – 6	56	99
6 – 9	48	147
9 – 12	32	179
12 – 15	18	197
Усього	197	

Необхідно визначити показники центральної тенденції і показники варіації.

Розв'язання.

Для розрахунку показників доцільно скласти допоміжну таблицю (табл. 6.2).

$$\bar{x} = \frac{1255,5}{197} = 6,37 \text{ років} - \text{середній термін роботи.}$$

Для визначення моди та медіани в інтервальному ряду спочатку знаходять інтервал, який містить ці показники, а за тим розраховують конкретні значення цих показників.

Таблиця 6.2

Допоміжна таблиця розрахунків

Групи за терміном експлуатації	Середина інтервалу, x_i	Кількість верстатів, f_i	Накопичені частоти, S	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
до 3	1,5	43	43	64,5	-4,87	23,72	1 019,96
3 – 6	4,5	56	99	252	-1,87	3,50	196
6 – 9	7,5	48	147	360	1,13	1,28	61,44
9 – 12	10,5	32	179	336	4,13	17,06	545,92
12 – 15	13,5	18	197	243	7,13	50,83	914,94
Усього		197		1 255,5			2 738,26

Модальним інтервалом у цьому випадку буде інтервал 3 – 6 років, тобто найбільша кількість верстатів перебуває в цьому інтервалі. Мода дорівнюватиме:

$$M_o = x_{M_o}^H + h_{M_o} \times \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})} = 3 + 3 \times \frac{56 - 43}{(56 - 43) + (56 - 48)} = 4,86 \text{ років.}$$

Отже, у досліджуваній сукупності найбільша кількість верстатів має термін експлуатації 4,86 роки.

Медіанним інтервалом також буде інтервал 3 – 6, тому що середину варіаційного ранжованого ряду (98,5) знаходять за накопиченими частотами там, де їх сума дорівнює 99. Мода складатиме:

$$M_e = x_{M_e}^H + h_{M_e} \times \frac{\frac{N}{2} - S_{M_e-1}}{f_{M_e}} = 3 + 3 \times \frac{\frac{197}{2} - 43}{56} = 5,97 \text{ років.}$$

Отже, термін експлуатації дорівнює 5,97 років і є тою варіантою, яка ділить варіаційний ряд розподілу на дві рівні частини (98 верстатів мають термін експлуатації нижче 5,97 років і 99 верстатів – понад 5,97 років).

Середнє квадратичне відхилення складатиме:

$$\sigma = \sqrt{\frac{2\,738,26}{197}} = 3,73 \text{ роки.}$$

Середній термін експлуатації становить 6,37 років, а середнє відхилення (що показує, як у середньому відхиляються індивідуальні значення ознаки від середньої арифметичної) дорівнює 3,73 роки.

Коефіцієнт варіації:

$$V\sigma = \frac{3,73}{6,37} = 0,585, \text{ або } 58,5 \%.$$

Відповідно до отриманих результатів сукупність не є однорідною, тому що отримане значення коефіцієнта варіації більше порогового значення 33 %.

Завдання 6.2. Маємо дані про розподіл працівників підприємства за стажем роботи (табл. 6.3).

Таблиця 6.3

Розподіл робітників за стажем роботи

Стаж роботи, років	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24	24 – 28
Кількість робітників за стажем роботи, осіб	6	8	11	13	6	4	2

Розв'язання.

Розрахунок показників варіації розглянемо на прикладі ряду розподілу робітників підприємства за стажем роботи. Для цього складемо допоміжну табл. 6.4.

Розрахунок середнього стажу робітників:

$$\bar{x} = \frac{\sum x_i \times f_i}{\sum f_i} = \frac{600}{50} = 12 \text{ років.}$$

Розмах показує загальний діапазон зміни стажу, він становить 28 років.

Таблиця 6.4

Розрахунок показників варіації для розподілу робітників за стажем роботи

№ групи	Стаж роботи, років			f_i	$f_i \times x_i$	Розрахунок середнього лінійного відхилення		Розрахунок дисперсії	
	x_i^H	x_i^B	x_i			$ x_i - \bar{x} $	$ x_i - \bar{x} \times f_i$	x_i^2	$x_i^2 \times f_i$
1	0	4	2	6	12	10	60	4	24
2	4	8	6	8	48	6	48	36	288
3	8	12	10	11	110	2	22	100	1 100
4	12	16	14	13	182	2	26	196	2 548
5	16	20	18	6	108	6	36	324	1 944
6	20	24	22	4	88	10	40	484	1 936
7	24	28	26	2	52	14	28	676	1 352
Разом	–	–	–	50	600	–	260	–	9 192

Середнє лінійне відхилення становить:

$$\bar{i} = \frac{\sum |x_i - \bar{x}| \times f_i}{\sum f_i} = \frac{260}{50} = 5,2 \text{ роки.}$$

Дисперсія для розглядуваного ряду становить:

$$\sigma^2 = \bar{x}^2 - (\bar{x})^2 = \frac{\sum x^2 \times f_i}{\sum f_i} - \left(\frac{\sum x_i \times f_i}{\sum f_i} \right)^2 = \frac{9192}{50} - 12^2 = 183,84 - 144 = 39,84.$$

Показник із такою розмірністю неможливо інтерпретувати, тому розрахуємо середнє квадратичне відхилення.

Середнє квадратичне відхилення становить: $\sigma = \sqrt{39,84} = 6,3$ роки

Перевіримо співвідношення між середнім лінійним відхиленням і середнє квадратичне відхилення: $\sigma \approx 1,25 \times \bar{i} = 1,25 \times 5,2 \approx 6,5$. Можна зробити висновок, що розподіл робітників за стажем наближений до нормального.

Коефіцієнт варіації становить: $V\sigma = \frac{6,3}{12} \times 100 \% = 53 \%$, що свідчить

про високий рівень коливання ознаки в сукупності.

Завдання 6.3. У табл. 6.5 наведені депозитні відсоткові ставки за десятима відділеннями банків.

Визначте різні види дисперсій депозитної відсоткової ставки, покажіть їх зв'язок, надайте інтерпретації та зробіть висновки про тісноту залежності рівня депозитної ставки від відділення банку.

Таблиця 6.5

Вихідні дані

Відділення комерційних банків	Депозитні відсоткові ставки у відділеннях, %					
	Центральні	22	25	24	27	–
Філії	20	19	21	22	18	24

Розв'язання.

Знайдемо середнє значення депозитної ставки щодо кожного відділення за формулою середньої арифметичної простої:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_m}{m} = \frac{\sum_{i=1}^m x_i}{m};$$

$$\bar{x}_1 = \frac{22 + 25 + 24 + 27}{4} = 24,5 \%;$$

$$\bar{x}_2 = \frac{20 + 19 + 21 + 22 + 18 + 24}{6} = 20,7 \%;$$

$$\bar{x}_0 = \frac{24,5 \times 4 + 20,7 \times 6}{10} = 22,22 \%.$$

Для центрального відділення комерційного банку внутрішньо-групова дисперсія дорівнює:

$$\sigma_1^2 = \frac{\sum_{i=1}^m x_i (\tilde{x}_i - \bar{x}_1)}{f_i} = \frac{(22 - 24,5)^2 + (25 - 24,5)^2 + (24 - 24,5)^2 + (27 - 24,5)^2}{4} = 3,25$$

та показує варіацію депозитних відсоткових ставок у центральному відділенні комерційного банку.

Аналогічно розраховуємо і для філії:

$$\begin{aligned}\bar{\sigma}_2^2 &= \frac{(20 - 20,7)^2 + (19 - 20,7)^2 + (21 - 24,5)^2 +}{6} = \\ &= \frac{+ (22 - 20,7)^2 + (18 - 20,7)^2 + (24 - 20,7)^2}{6} = 5,92.\end{aligned}$$

Середня із групових дисперсій дорівнює:

$$\bar{\sigma}_i^2 = \frac{\sum_{i=1}^m \sigma_i^2}{m} = \frac{3,25 \times 4 + 5,92 \times 6}{10} = 4,85,$$

що відображає варіацію депозитних відсоткових ставок у середньому для всієї сукупності. Варіація обумовлена різними факторами, окрім відмінностей між відділеннями комерційного банку.

Міжгрупова дисперсія

$$\bar{\delta}^2 = \frac{\sum_{i=1}^m (x_i - \bar{x}_i)^2}{m} = \frac{(24,5 - 22,2)^2 \times 4 + (20,7 - 22,2)^2 \times 6}{10} = 3,47$$

характеризує варіацію групових середніх, обумовлену відмінностями в відділеннях комерційного банку (депозитна відсоткова ставка відрізняється на 3,47).

Згідно з правилами складання дисперсій, загальна дисперсія дорівнює: $\sigma^2 = \bar{\delta}^2 + \bar{\sigma}_i^2 = 4,85 + 3,47 = 8,32$. Вона характеризує вплив усіх факторів на варіацію депозитних відсоткових ставок.

Завдання 6.4. На фірмі проведено аналіз якості виробленої в партії продукції. У результаті перевірки виявлено, що 2 % усієї продукції в партії є бракованою. Необхідно визначити дисперсію альтернативної ознаки.

Розв'язання.

Отже, $p = 2 \% = 0,02$.

Оскільки $p + q = 1$, то $q = 1 - p$.

Відповідно, 98 % деталей в партії якісні, тобто $q = 98 \% = 0,98$.

Дисперсія частки браку $\sigma^2 = pq = 0,02 \div 0,98 = 0,0196$.

Це означає, що варіація браку від середньої кількості бракованих виробів складає 0,14 ($\sigma = \sqrt{0,0196}$).

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Аналіз рядів розподілу" подані в роботі [67]. Лабораторна робота націлена на отримання вмінь щодо аналізу рядів розподілу за допомогою MS Excel. Завдання лабораторної роботи – провести аналіз статистичних рядів розподілу за допомогою надбудови "Аналіз даних".

Запитання для самодіагностики

1. Що таке варіація ознак? Наведіть приклади.
2. Які показники використовують для вимірювання варіації? Назвіть їх і наведіть формули.
3. Як знайти моду та медіану графічним способом?
4. Які недоліки притаманні розмаху варіації і середньому лінійному відхиленню?
5. Розкажіть про дисперсію і середнє квадратичне відхилення та їх місце у системі показників варіації.
6. Назвіть основні математичні властивості дисперсії.
7. Назвіть формули спрощених розрахунків дисперсії і поясніть їх сутність.
8. Як вимірюють варіацію альтернативної ознаки?
9. Які види дисперсій ви знаєте? Розкрийте їх сутність.
10. Розкажіть про правило додавання (розкладання) варіації. Де воно застосовується?

Питання для критичного переосмислення (есе)

1. Обґрунтуйте, чому середню розглядають як типовий рівень ознаки в сукупності. Як співвідноситься середня з іншими характеристиками центру розподілу?
2. Розкрийте послідовність розрахунків загальної, міжгрупової і внутрішньогрупової дисперсій для альтернативних ознак.
3. Наведіть математичні властивості дисперсії, які допомагають спростити вираження її величини.

4. Розкрийте сутність поняття "моменти статистичного розподілу" та поясніть його використання.

5. Поясніть необхідність розрахунку емпіричного кореляційного відношення та доцільність його використання в оцінюванні тісноти та сили зв'язку між факторною та результативною ознаками.

7. Вибірковий метод

Основні питання:

7.1. Загальне поняття про вибіркове спостереження. Види та способи відбору одиниць із генеральної сукупності.

7.2. Помилки вибіркового спостереження для різних схем відбору.

7.3. Особливості малої вибірки та визначення необхідної чисельності вибірки.

7.1. Загальне поняття про вибіркове спостереження.

Види та способи відбору одиниць із генеральної сукупності

У практиці статистичних досліджень, як найбільш науково обґрунтованому виді несучільного спостереження, широке визнання і застосування має **вибіркове спостереження**. Діалектична єдність одиничного, особливого та загального дає можливість за одиничним та особливим міркувати про загальне, а за умов правильного виявлення взаємозв'язку – за часткою судити про ціле.

Вибіркове спостереження дає найбільш оптимальний результат тільки в тому випадку, коли воно буде організовано та приведено у відповідність до **наукових принципів теорії вибіркового методу** [13; 25; 26; 46; 68; 74; 75]:

забезпечення випадковості відбору одиниць (шляхом відбору кожної із одиниць сукупності, що вивчається, забезпечується однакова їх можливість потрапити у вибірку);

забезпечення достатньої кількості відібраних одиниць (чим більше одиниць буде обстежено, тим точнішим буде уявлення про досліджувану сукупність і тим меншою буде похибка вибірки).

Важлива роль у формуванні вибіркового методу належить роботам Я. Бернуллі. Вагомий внесок у розроблення теоретичних основ розгляданого методу внесли математики П. Чебишев, О. Ляпунов, А. Марков, Р. Фішер, В. Госсет. Теорія вибіркового методу отримала розвиток у працях статистиків О. Чупрова, А. Ковалевського.

Класифікацію форм вибіркового спостереження дали такі статистики, як А. Боярський, Б. Ястремський.

Вибірковим називають вид несучільного спостереження, за характеристикою відібраної частини одиниць якого судять про всю сукупність у цілому. Тобто у ході вибіркового спостереження обстежуються не всі одиниці досліджуваного явища (процесу), а лише деяка їх частина, що відібрана певним способом. Однак спостереження організовано таким чином, що ця частина відібраних одиниць у зменшеному масштабі представляє (репрезентує) всю сукупність.

Вибіркове спостереження застосовують для вивчення різноманітних закономірностей суспільних явищ.

Вибіркове спостереження має *істотні переваги* порівняно з іншими методами отримання статистичних даних:

досить висока точність результатів обстеження завдяки залученню до процесу більш кваліфікованих кадрів, що приводить до зменшення помилок реєстрації;

економія часу та коштів у результаті скорочення обсягу роботи;

велика оперативність в отриманні даних про результати обстеження;

можливість дослідження дуже великих статистичних сукупностей;

зведення до мінімуму псування (знищення) досліджуваних об'єктів;

можливість дослідження повністю недоступних сукупностей.

За вибіркового дослідження вивчається порівняно невелика частина статистичної сукупності (5 – 10 %, рідше 20 – 25 % обсягу її одиниць).

Вибіркове спостереження часто застосовують у поєднанні із суцільним спостереженням – для поглибленого дослідження, уточнення та контролю результатів суцільного спостереження.

Проведення вибіркового спостереження містить такі **етапи** [7; 17; 37; 68; 88]:

обґрунтування мети статистичного спостереження;

складання програми спостереження і розроблення даних;

вирішення організаційних аспектів спостереження ;

визначення відсотка та способу відбору;

здійснення відбору;
реєстрація ознак досліджуваних одиниць;
узагальнення даних спостереження та визначення їхніх вибірових характеристик;
визначення помилки вибірки;
перевірка надійності результатів вибірового спостереження;
поширення вибірових характеристик на всю сукупність.

Для постановки завдання вибірового дослідження необхідно ввести такі *поняття* [17; 24; 68; 71; 75]:

генеральна сукупність (N) – досліджувана сукупність, з якої проводиться відбір одиниць, що підлягають вивченню (вона може мати кінцеве значення або не мати його);

вибірова сукупність (вибірка) (n) – частина одиниць генеральної сукупності, що відібрана для вивчення.

Якість результатів вибірового дослідження залежить від того, наскільки склад вибірки представляє генеральну сукупність, тобто наскільки вибірка *репрезентативна*.

Під **репрезентативністю вибірки** розуміють відповідність її властивостей і структури властивостям і структурі генеральної сукупності. Репрезентативність вибірки може бути забезпечена тільки у разі об'єктивності відбору даних, що гарантована принципами випадковості відбору одиниць.

Принцип випадковості передбачає, що на включення або виключення статистичної одиниці з вибірки не може вплинути жоден інший фактор, крім випадку. Застосування принципу випадковості у формуванні вибірки дозволяє надалі для оброблення даних використовувати апарат теорії ймовірності.

Найчастіше за допомогою вибірового дослідження визначають такі **характеристики генеральної сукупності**:

середнє значення ознаки в сукупності, тобто генеральна середня;
частка альтернативної ознаки в сукупності. Альтернативною вважається ознака, що приймає два значення. Якщо одне з них змінюється як задане, то частка альтернативної ознаки буде характеризувати питому вагу статистичних одиниць, котрі володіють заданим значенням альтернативної ознаки (наприклад, частка забракованих виробів у партії виготовленої продукції);

дисперсія ознаки в сукупності, тобто генеральна дисперсія – як показник варіації.

Метою вибіркового спостереження є виділення з генеральної сукупності n одиниць і розрахунок на їх основі невідомих параметрів генеральної сукупності.

У загальному вигляді **завдання вибіркового дослідження** формулюється таким чином [68]. Нехай існує деяка генеральна сукупність відомого обсягу (N одиниць), що має певні статистичні характеристики:

$D = \frac{P}{N}$ – **генеральна частка** (питома вага статистичних одиниць

генеральної сукупності, що володіють певними значенням ознаки), де P – число одиниць генеральної сукупності, яким притаманна певна ознака;

\bar{X} – **генеральна середня** (середнє арифметичне значення ознаки у генеральній сукупності);

σ^2 – **генеральна дисперсія** (дисперсія ознаки, що досліджується в генеральній сукупності);

σ – **генеральне середнє відхилення** (середнє квадратичне відхилення ознаки, що досліджується в генеральній сукупності).

Для їх визначення сформована вибірка сукупність обсягом n статистичних одиниць ($n < N$), що володіє аналогічними характеристиками:

w – **вбіркова частка** (питома вага статистичних одиниць, що володіють даним значенням ознаки у вибірковій сукупності);

\tilde{X} – **вбіркова середня** (середнє арифметичне значення ознаки у вибірковій сукупності);

σ_v^2 – **вбіркова дисперсія** (дисперсія ознаки, що досліджується в вибірковій сукупності);

σ_v – **вбіркове середнє квадратичне відхилення** (середнє квадратичне відхилення ознаки, що досліджується у вибірковій сукупності).

Необхідно на основі відомих характеристик вибірки отримати статистичні оцінки характеристик генеральної сукупності.

Відбирання одиниць із генеральної сукупності можна здійснювати по-різному – залежно від певних умов. Систему організації відбирання одиниць із генеральної сукупності називають **способом відбору**. Залежно від того, чи беруть участь в подальшій виборці вже відібрані одиниці, розглядають **методи відбору**: повторний і безповторний [17; 44; 45; 75].

У разі **повторного відбору** кожна одиниця бере участь у вибірці стільки разів, скільки відбирається одиниць. Тобто після реєстрації ознак

одиниця повертається у генеральну сукупність, і в подальшому є ймовірність того, що вона може знову потрапити до вибіркової сукупності.

У **безповторному відборі** кожна відібрана одиниця бере участь у відборі лише один раз і після реєстрації ознак у генеральну сукупність не повертається.

Повторний і безповторний методи відбору (залежно від характеру одиниці відбору) застосовуються в узгодженні з іншими видами відбору.

У практичній діяльності використовують три **види відбору**:

індивідуальний – відбирання одиниць сукупності;

груповий – відбирання груп одиниць;

комбінаційний – комбінація першого та другого видів.

Різні види відбору можуть здійснюватися різними способами. **Спосіб здійснення вибірки** визначений *правилом утворення вибіркової сукупності*. У практиці вибірових обстежень застосовуються такі вибірки, як: випадкова, механічна, типічна, серійна, комбінована [4; 23; 28; 37; 38; 68; 71].

Залежно від того, як змінюється одиниця відбору за умов послідовного проведення кількох відборів, відрізняють *одноступеневий* й *багатоступеневий* відбір.

Кожний із вказаних способів відбору має свої особливості здійснення вибірки та методи розрахунку середньої помилки вибірки.

Технічно **власне випадковий відбір** проводять методом жеребкування або за таблицею випадкових чисел. Можна очікувати, що серед відібраних одиниць є представники різних станів, якими характеризується ознака в загальній сукупності. У такому випадку середнє значення досліджуваної ознаки буде відбито досить точно.

Власне випадковий відбір у "чистому вигляді" застосовується рідко, але він є вихідним для всіх інших видів відбору.

Як зазначалось, випадковий відбір може бути повторним або безповторним.

Механічний відбір полягає у відбиранні одиниць із генеральної сукупності в будь-якому механічному порядку. Найчастіше механічний відбір застосовується у випадку об'єктивної послідовності розташування одиниць відбору. Відбір здійснюється безповторним засобом через рівні інтервали. З кожного інтервалу до вибірки потрапляє лише одна одиниця.

Для проведення механічного відбору необхідно встановити крок відліку (відстань між відібраними одиницями) та початок відліку (номер одиниці, яка повинна бути обстежена першою). Крок відліку встановлюється

виходячи з передбачуваного відсотка відбору. Наприклад, для десятивідсоткового відбору відбирається кожна десята одиниця, для двадцятивідсоткового – кожна двадцята.

Особливістю механічного відбору є те, що за його застосування можлива поява систематичних помилок, які пов'язані з випадковим збігом обраного інтервалу та циклічних закономірностей у розташуванні одиниць генеральної сукупності. Щоб уникнути систематичних помилок, слід відбирати статистичну одиницю, що знаходиться всередині кожного інтервалу. Цей спосіб дуже зручний у тих випадках, коли не можна заздалегідь скласти перелік одиниць генеральної сукупності (вибірка береться із сукупності, що постійно формується в часі).

Вивчаючи складні сукупності, які можна розбити на декілька якісно однорідних груп за суттєвими ознаками щодо мети дослідження доцільно використовувати *стратифіковану (розшаровану) вибірку*.

Усередині кожної групи такої вибірки проводиться власне випадковий або механічний відбір. Отримані групи за чисельністю одиниць, як правило, не дорівнені між собою, тому відбір одиниць здійснюється пропорційно обсягу групи. Отже, число спостережень за кожною групою визначається за формулою [68]:

$$n_{iB} = N \frac{n_j}{N}, \quad (7.1)$$

де n_{iB} – число спостережень з i -ї групи генеральної сукупності;

N – обсяг генеральної сукупності;

n_j – обсяг i -ї групи генеральної сукупності.

Якщо пропорції між групами у вибірці збігаються з пропорціями між групами в генеральній сукупності, то відбір називають **типовим**. За типового відбору сукупність попередньо розбивають на однорідні групи, типи, класи. Сутність типового відбору полягає у типовому районуванні статистичної сукупності на однорідні групи, з яких здійснюється відбір за випадковим або механічним принципом.

Серійний відбір застосовується в тих випадках, коли одиниці статистичної сукупності об'єднані в невеликі *групи* або *серії (гнізда)*. В якості таких серій можуть розглядатися, наприклад, пакунки з певною кількістю готової продукції.

Для відбору серій застосовують власне випадковий або механічний відбір. Спостереженню підлягають усі одиниці відібраної серії. Серійний відбір має велике практичне значення, тому що обстежується незначне число серій, і це скорочує витрати на проведення спостереження.

Розглянуті способи формування вибірки можуть застосовуватися в "чистому вигляді" або комбінуватися в різних поєднаннях і послідовності. Використання декількох способів формування вибірки в одному вибірковому дослідженні називають **комбінованим відбором (добором)**. Такий відбір проводиться в декілька етапів, і на кожному з них застосовується свій спосіб відбору. Наприклад, під час обстеження сімейних доходів вибіркоче обстеження проводиться в такій послідовності:

установлюються населені пункти, що потрапляють під обстеження. Для цього застосовують розшарований відбір, що дає змогу відібрати великі міста, середні міста, та інші населені пункти;

у кожному населеному пункті встановлюються місця, де проживають сім'ї: вулиці, будинки. Для цього використовується механічний відбір (за списком вулиць і нумерацією будинків);

у кожному місці проживання родин відбираються конкретні родини, для чого застосовується власне випадковий неповторний або механічний відбір. Для відбору використовують перелік номерів квартир або списки сімей.

Методи формування вибірки впливають не тільки на точність статистичних оцінок через помилки вибірки, а і на обсяг вибіркової сукупності.

7.2. Помилки вибіркового спостереження для різних схем відбору

Статистичною оцінкою, або **статистичною характеристикою** (параметром) генеральної сукупності називають наближене значення шуканої характеристики (параметру), що отримане за даними вибірки.

У статистиці використовують два **типи вибіркових оцінок** – точкові та інтервальні [7; 13; 26; 37; 44].

Точкова оцінка характеризує значення параметра й обчислена на основі даних вибірки. **Інтервальна оцінка** – довірчий інтервал значень параметра із заданою ймовірністю.

Якість статистичних оцінок визначається такими їхніми **властивостями**:

1) спроможність. Оцінка вважається спроможною, якщо за необмеженого збільшення обсягу вибірки $n \rightarrow \infty$ (N) її помилка прагне до 0 [68]:

$$\lim_{n \rightarrow \infty} (\tilde{a} - a) = 0, \text{ тому що } n \uparrow \lim_{n \rightarrow \infty} \tilde{a} = a, \quad (7.2)$$

де a – значення характеристики генеральної сукупності;

\tilde{a} – значення характеристики вибірки;

$\tilde{a} - a$ – помилка вибірки;

2) незміщеність. Оцінка вважається незміщеною, якщо із заданого обсягу вибірки n математичне очікування помилки дорівнює 0.

Для незміщеної оцінки її математичне очікування точно буде дорівнювати математичному очікуванню характеристики вибірки [68]:

$$M[\tilde{a} - a] = 0 \text{ або } M[\tilde{a}] = M[a]. \quad (7.3)$$

Незміщена оцінка не завжди дає добре наближення оцінюваного параметра, тому що можливі значення отриманої оцінки можуть бути сильно розсіяні навколо свого середнього значення. У такому випадку оцінка повинна відповідати ще одній вимозі – *ефективності*;

3) ефективність. Оцінка вважається ефективною, якщо її помилка, яку називають *помилкою вибірки*, є величиною мінімальною.

У математичній статистиці помилка вибірки визначається як [68]:

$$\mu(\tilde{a}) = \sqrt{M^2[\tilde{a} - a] + S^2}, \quad (7.4)$$

де $M^2[\tilde{a} - a]$ – квадрат математичного очікування помилки вибірки;

S^2 – вибіркова дисперсія.

Оцінка є ефективною, якщо виконується умова: $\mu(\tilde{a}) \rightarrow \min$.

Для **точкової оцінки** справедливі такі твердження:

точковою оцінкою *генеральної частки* є *вибіркова частка*: $p \approx W$;

точковою оцінкою *генеральної середньої* є *вибіркова середня*: $\bar{x} \approx \check{x}$.

Отже, заздалегідь відомо, що оцінки для зазначених параметрів є *спроможними та незміщеними*.

Для *інших параметрів* генеральної сукупності це твердження *не є справедливим*; отже: $\sigma^2 \neq \sigma_B^2$ і $\sigma \neq \sigma_B$.

У математичній статистиці доводиться, що точковою оцінкою генеральної дисперсії є вибіркова дисперсія, яка відкорегована на співвідношення $\frac{n}{n-1}$. Тобто $\sigma^2 = \sigma_B^2 \frac{n}{n-1}$, зі збільшенням n $\frac{n}{n-1} \rightarrow 1$. Тому у вибірках з обсягом більше за 30 одиниць спостереження вказаним відношенням можна знехтувати.

Аналогічно, точковою оцінкою генерального середньоквадратичного відхилення є вибіркове середньоквадратичне відхилення, відкориговане на $\frac{n}{n-1}$, тобто $\sigma = \sigma_B \frac{n}{n-1}$.

Основним **недоліком точкових оцінок** є те, що вони не враховують помилки вибірки, тобто не є ефективними. Тому більш доцільними є інтервальні оцінки параметрів генеральної сукупності, що враховують такі помилки.

Інтервальні оцінки відповідають усім трьом вимогам якості статистичної оцінки.

У математичній статистиці доводиться, що:

інтервальною оцінкою генеральної частки є її вибіркова частка з урахуванням помилки вибіркової частки, тобто: $p \approx w \pm \mu_w$, де μ_w – помилка вибіркової частки;

інтервальною оцінкою генеральної середньої є середня вибіркова з урахуванням помилки вибіркової середньої, тобто: $\bar{x} \approx \check{x} \pm \mu_x$, де μ_x – помилка вибіркової середньої.

Застосування інтервальних оцінок означає, що характеристики генеральної сукупності вкладаються в певний діапазон значень, який визначається з урахуванням відповідної помилки вибірки.

У разі правильного формування вибірки величину її помилки можна розрахувати заздалегідь. У загальному випадку під *помилкою вибірки* розуміють розбіжність, що об'єктивно виникає між характеристиками вибірки та генеральної сукупності.

Помилки вибірки підрозділяють на помилки реєстрації та помилки репрезентативності [7; 17; 24; 80; 88].

Помилки реєстрації виникають унаслідок неправильних або неточних відомостей. Їх джерелом є неуважність реєстратора, неправильне

заповнення формулярів, описки або нерозуміння сутності досліджуваної проблеми.

Помилки репрезентативності виникають через невідповідність структури вибірки структурі генеральної сукупності. Джерелом їх появи є різна варіація ознаки статистичних одиниць, у результаті якої розподіл одиниць у вибірковій сукупності відрізняється від розподілу одиниць у генеральній сукупності. Тобто *помилка репрезентативності* – це розбіжність між певною характеристикою генеральної сукупності (часткою, середньою, дисперсією тощо) та її вибірковою оцінкою.

За причинами виникнення помилки репрезентативності розподіляють на систематичні та випадкові.

Систематичні помилки репрезентативності виникають через неправильне формування вибірки, від чого порушується основний принцип її наукової організації (принцип випадковості). Такі помилки є односпрямованими та призводять до зміщення результатів обстеження.

Випадкові помилки репрезентативності виникають унаслідок випадковості добору та пов'язаних із цим розбіжностей між структурами вибіркової та генеральної сукупності.

Математичну основу для обчислення та врегулювання розміру випадкових помилок вибірки дає **теорія вибіркового методу**.

Наприклад, порівнюються дані про рівень успішності студентів факультету (генеральна сукупність) за двома десятивідсотковими випадковими вибірками (табл. 7.1).

Таблиця 7.1

Розподіл студентів за рівнем успішності

Оцінка за національною шкалою	Кількість студентів, осіб		
	Генеральна сукупність	Перша вибірка	Друга вибірка
2 (незадовільно)	100	9	13
3 (задовільно)	500	50	49
4 (добре)	380	30	32
5 (відмінно)	120	21	16
Усього	1 100	110	110

Обсяг генеральної сукупності N склав 1 000 студентів, обсяг кожної вибірки складає 10 % від N , тобто $n = 0,1 \times 1\,000 = 100$ осіб. Вихідні дані подані дискретними рядами розподілу студентів за рівнем успішності, тому середній рівень успішності розрахуємо за формулою середньої арифметичної зваженої. Він становить:

$$\text{за генеральною сукупністю: } \bar{x} = \frac{100 \times 2 + 500 \times 3 + 380 \times 4 + 120 \times 5}{1100} = 3,47;$$

$$\text{за першою вибіркою: } \tilde{x}_1 = \frac{2 \times 9 + 3 \times 50 + 4 \times 30 + 5 \times 21}{110} = 3,57;$$

$$\text{за другою вибіркою: } \tilde{x}_2 = \frac{2 \times 13 + 3 \times 49 + 4 \times 32 + 5 \times 16}{110} = 3,46.$$

Частка студентів, які отримали оцінки "добре" і "відмінно", дорівнює:

$$\text{за генеральною сукупністю: } w = \frac{380 + 120}{1100} = 0,45;$$

$$\text{за першою вибіркою: } w_1 = \frac{30 + 21}{110} = 0,46;$$

$$\text{за другою вибіркою: } w_2 = \frac{32 + 16}{110} = 0,44.$$

Різниця між показниками вибіркової та генеральної сукупності є випадковою помилкою репрезентативності. Помилки репрезентативності для середньої становлять:

$$\tilde{x}_1 - \bar{x} = 3,57 - 3,47 = +0,10;$$

$$\tilde{x}_2 - \bar{x} = 3,46 - 3,47 = -0,01.$$

Відповідно, помилки репрезентативності для частки дорівнюють:

$$w_1 - w = 0,46 - 0,45 = +0,01;$$

$$w_2 - w = 0,44 - 0,45 = -0,01.$$

З наведених розрахунків видно, що вибіркова середня та вибіркова частка є випадковими величинами, які можуть приймати різні значення залежно від одиниць, що потрапили у вибірку. Помилки вибірки також

можна вважати випадковими величинами. Вони можуть приймати різні значення, тому визначають середню (стандартну) з можливих помилок.

Величина помилки вибірки залежить від таких **факторів** [45; 74; 75; 80]:

ступінь коливання ознаки у генеральній сукупності (чим більш однорідною є досліджувана сукупність, тим менше величина середньої помилки за тою самою чисельністю вибірки);

обсяг (чисельність) вибірки (збільшуючи або зменшуючи обсяг вибірки n , можна регулювати величину середньої помилки: чим більше одиниць буде включено до вибірки, тим менше буде величина помилки, тому що тим точніше у вибірці буде представлена генеральна сукупність);

спосіб відбору одиниць у вибіркову сукупність (у практичній діяльності використовують різні способи формування вибіркової сукупності, але принципове значення має їх розподіл на способи випадкового повторного та неповторного відбору).

Під **середньою (стандартною) помилкою вибірки** μ розуміють таке розходження між середньою вибіркової та середньою генеральної сукупностей $(\bar{x} - \bar{x})$, яке не перевищує $\pm \sigma$.

У **власне випадковому повторному відборі** статистична одиниця, що потрапила у вибірку, після реєстрації досліджуваної ознаки повертається в генеральну сукупність і може знову потрапити у вибірку. Отже, для всіх одиниць генеральної сукупності забезпечується рівна ймовірність відбору.

У математичній статистиці середню помилку вибірки визначають за формулою [68]:

$$\mu = \sqrt{\frac{\sigma^2}{n}}, \quad (7.5)$$

де σ^2 – генеральна дисперсія.

Генеральна дисперсія, як інші параметри генеральної сукупності, є невідомою величиною, але відоме співвідношення між генеральною і вибірковою дисперсією $\sigma^2 = \sigma_B^2 \frac{n}{n-1}$. Тоді за досить великих обсягів вибірки ($n > 30$) $\frac{n}{n-1} \rightarrow 1$ можна вважати, що $\sigma^2 \approx \sigma_B^2$. У випадках малої

вибірки з $n < 30$ необхідно враховувати співвідношення $\frac{n}{n-1}$ і розраховувати середню помилку малої вибірки за формулою [68]:

$$\mu_{MB} = \sqrt{\frac{\sigma^2}{n-1}}. \quad (7.6)$$

Отже, для середньої кількісної ознаки **середня помилка** вибірки $\mu_{\bar{x}}$ буде дорівнювати [68]:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma_{\bar{x}}^2}{n}}, \quad (7.7)$$

де $\sigma_{\bar{x}}^2 = \frac{\sum(x_i - \bar{x})^2}{n}$ – вибіркова дисперсія кількісної ознаки.

Середня помилка вибірки для частки визначається за формулою [68]:

$$\mu_w = \sqrt{\frac{\sigma_w^2}{n}}, \quad (7.8)$$

де $\sigma_w^2 = w(1-w)$ – вибіркова дисперсія частки альтернативної ознаки.

Застосування простої випадкової повторної вибірки на практиці дуже обмежено. Це пов'язано з тим, що практично недоцільне, а іноді неможливе повторне спостереження тих самих одиниць (уже обстежена одиниця повторному обліку не піддається). Тому доцільно застосовувати безповторний відбір.

У **власне випадковому безповторному відборі** загальна кількість статистичних одиниць генеральної сукупності в процесі формування вибірки змінюється (зменшуючись щоразу на одиницю, що потрапила у вибірку), оскільки відібрані одиниці в генеральну сукупність не повертаються. Таким чином, ймовірність потрапляння окремих одиниць у вибірку за безповторного випадкового відбору також змінюється (для решти одиниць вона зростає). У цілому ймовірність потрапляння будь-якої статистичної одиниці у вибірку в разі безповторного відбору може бути

визначена як: $1 - \frac{n}{N}$. На цю величину повинна бути скоригована і середня помилка вибірки із застосуванням безповторного відбору.

Отже, розрахункові формули середньої помилки вибірки у безповторном відборі приймають вигляд [68]:

$$\text{для середньої кількісної ознаки: } \mu_{\bar{x}} = \sqrt{\frac{\sigma_{\bar{x}}^2}{n}} \times \left(1 - \frac{n}{N}\right); \quad (7.9)$$

$$\text{для частки альтернативної ознаки: } \mu_w = \sqrt{\frac{w \times (1-w)}{n}} \times \left(1 - \frac{n}{N}\right). \quad (7.10)$$

На практиці застосуванням вибіркового методу визначають межі, за які не вийде величина конкретної помилки вибіркового дослідження. Величина меж конкретної помилки визначається ступенем ймовірності.

Помилку вибірки, обчислену із заданим ступенем ймовірності, називають граничною помилкою вибірки.

Гранична помилка вибірки (Δ) – це максимум помилки за заданої ймовірності її появи. Це означає, що із заданою ймовірністю гарантується, що будь-яка помилка вибірки не перевищить граничну помилку. Така ймовірність є **довірчою**.

Гранична помилка вибірки Δ розраховується за формулою [68]:

$$\Delta = t\mu, \quad (7.11)$$

де t – коефіцієнт довіри, значення якого визначаються довірчою ймовірністю p_t .

Значення коефіцієнта довіри t задаються в таблицях нормального розподілу ймовірностей. Найчастіше використовують поєднання, показані в табл. 7.2 [68].

Таблиця 7.2

Значення коефіцієнта довіри t

t	p_t	t	p_t
1	0,683	2,5	0,988
1,5	0,866	3,0	0,997
2,0	0,954	3,5	0,999

Так, якщо $t = 1$, то з ймовірністю 0,683 можна стверджувати, що розбіжність між вибірковими характеристиками та параметрами генеральної сукупності не перевищить однієї середньої помилки.

Граничні помилки вибірки Δ для різних параметрів за різних методів відбору статистичних одиниць розраховуються за формулами, наведеними в табл. 7.3 [68].

Таблиця 7.3

Граничні помилки вибірки

Методи відбору	Граничні помилки вибірки	
	для середньої	для частки
Повторний	$\Delta_{\bar{x}} = t \times \sqrt{\frac{\sigma_{\bar{x}}^2}{n}}$	$\Delta_p = t \times \sqrt{\frac{w \times (1-w)}{n}}$
Безповторний	$\Delta_{\bar{x}} = t \times \sqrt{\frac{\sigma_{\bar{x}}^2}{n} \times \left(1 - \frac{n}{N}\right)}$	$\Delta_p = t \times \sqrt{\frac{w \times (1-w)}{n} \times \left(1 - \frac{n}{N}\right)}$

Знаючи величину граничної помилки вибірки, можна розрахувати **інтервали** для характеристик генеральної сукупності [7; 44; 48; 68; 88]: довірчий інтервал для генеральної середньої дорівнює $\bar{x} \pm \Delta_{\bar{x}}$; для генеральної частки – $w \pm \Delta_p$.

Розглянемо знаходження середньої та граничної помилки вибірки, визначення довірчих інтервалів для середньої та частки на прикладі.

Приклад 7.1. За звітний період на території регіону мешкало 2 000 сімей. У ході оцінювання попиту на товар А було проведено п'ятивідсоткове безповторне обстеження сімей. Було з'ясовано, що в 90 із 100 обстежених сімей зазначений товар споживається. У середньому кожна із обстежених сімей споживала 5 одиниць товару ($\bar{x} = 5$) зі стандартним відхиленням 0,5 одиниці ($\sigma = 0,5$ од.). З ймовірністю $p = 0,954$ слід установити частку сімей, які споживають товар А та межі його споживання.

Для отримання статистичних оцінок параметрів генеральної сукупності виконаємо такі розрахунки.

1. Визначаємо характеристики вибіркової сукупності:

а) *вибіркову частку* w (питома вага сімей, що споживають товар А, у виборці): $w = \frac{90}{100} = 0,9$;

б) *вибіркову середню* \tilde{x} (середній обсяг споживання товару А однією сім'єю у виборці): $\tilde{x} = 5$ одиниць.

2. Визначаємо *граничні помилки* вибірки:

а) для частки:

$$\Delta_p = t \times \sqrt{\frac{w \times (1-w)}{n} \times \left(1 - \frac{n}{N}\right)} = 2 \times \sqrt{\frac{0,9 \times (1-0,9)}{100} \times \left(1 - \frac{100}{2000}\right)} = 0,0585,$$

де $N = \frac{n \times 100}{5} = 2000$ родин;

б) для середньої:

$$\Delta_{\tilde{x}} = t \times \sqrt{\frac{\sigma_{\tilde{x}}^2}{n} \times \left(1 - \frac{n}{N}\right)} = 2 \times \sqrt{\frac{0,5^2}{100} \times \left(1 - \frac{100}{2000}\right)} \approx 0,1.$$

3. Розрахуємо *довірчі інтервали характеристик генеральної сукупності*:

а) для частки:

$$\begin{aligned} w - \Delta_p &\leq p \leq w + \Delta_p, \\ 0,9 - 0,059 &\leq p \leq 0,9 + 0,059, \\ 0,841 &< p < 0,959; \end{aligned}$$

б) для середньої:

$$\begin{aligned} \tilde{x} - \Delta_{\tilde{x}} &\leq \bar{x} \leq \tilde{x} + \Delta_{\tilde{x}}, \\ 5 - 0,1 &\leq \bar{x} \leq 5 + 0,1, \\ 4,9 &< \bar{x} < 5,1. \end{aligned}$$

Отже, з імовірністю 0,954 можна стверджувати, що в генеральній сукупності частка родин, які споживають товар А, буде не менше 84,1 %,

але не більше 95,5 %; середнє споживання товару в родинях перебуває у межах від 4,9 до 5,1 одиниць.

На підставі проведених розрахунків можна визначити межі споживання (попиту) товару А в регіоні:

$$4,9 \times 0,841 \times 2\,000 < Q < 5,1 \times 0,959 \times 2\,000;$$
$$8\,240 < Q < 9\,780.$$

Тобто з ймовірністю 0,954 можна стверджувати, що попит на товар А не буде нижчим 8 240 одиниць, але не перевищить 9 780.

Розглянемо розрахунок помилок вибірки для різних схем відбору.

Для визначення *середньої помилки механічної вибірки* використовують формули середньої помилки для власно випадкового відбору.

Типова вибірка забезпечує більш точні результати порівняно з іншими способами відбору одиниць, оскільки дозволяє виключити вплив міжгрупової дисперсії на середню помилку вибірки. Тому середня помилка типової вибірки буде залежати тільки від середньої з внутрішньогрупових дисперсій. Типова вибірка може бути повторною або неповторною.

Середня помилка типової вибірки в повторному відборі визначається за формулами [24; 48]:

для середньої кількісної ознаки:

$$\mu_{\bar{x}} = \sqrt{\frac{\bar{\sigma}_B^2}{n}}, \quad (7.12)$$

де $\bar{\sigma}_B^2 = \frac{\sum \sigma_i^2 \times n_i}{\sum n_i}$ – середня із внутрішньогрупових вибірових дисперсій;

для частки альтернативної ознаки:

$$\mu_w = \sqrt{\frac{\overline{w_i \times (1 - w_i)}}{n}}, \quad (7.13)$$

де $\overline{w_i \times (1 - w_i)}$ – середня з внутрішньогрупових дисперсій частки альтернативної ознаки у виборці:

$$\overline{w_i \times (1 - w_i)} = \frac{\sum w_i(1 - w_i)n_i}{\sum n_i}. \quad (7.14)$$

Середня помилка типової вибірки у безповторному відборі:
для середньої кількісної ознаки:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma_B^2}{n} \times \left(1 - \frac{n}{N}\right)}, \quad (7.15)$$

де n – обсяг вибірки з типової групи;

N – обсяг типової групи;

для частки альтернативної ознаки:

$$\mu_w = \sqrt{\frac{w_i \times (1 - w_i)}{n} \times \left(1 - \frac{n}{N}\right)}. \quad (7.16)$$

У серійному відборі випадкова помилка дещо більша, ніж за інших способів відбору. Оскільки всередині серій обстежуються всі без винятку статистичні одиниці, то величина середньої помилки серійного відбору залежить тільки від міжгрупової (міжсерійної) дисперсії.

Середня помилка серійної вибірки для повторного відбору визначається таким чином [24; 48]:

1) для середньої кількісної ознаки:

$$\mu_{\bar{x}} = \sqrt{\frac{\delta_{\bar{x}}^2}{r}}, \quad (7.17)$$

де $\delta_{\bar{x}}^2 = \frac{\sum (\tilde{x}_i - \tilde{\bar{X}})^2}{r}$,

\tilde{x}_i – середнє значення i -ї серії;

$\tilde{\bar{X}}$ – середня величина для всієї вибірки;

r – число відібраних серій;

2) для частки альтернативної ознаки:

$$\mu_w = \sqrt{\frac{\delta_w^2}{r}}, \quad (7.18)$$

де $\delta_w^2 = \frac{\sum (w_i - \bar{w})^2}{r}$ – міжгрупова дисперсія частки серійної вибірки;

w_i – частка ознаки в i -ій частці;

\bar{w} – загальна частка ознаки в усій виборці.

Середня помилка серійної вибірки в безповторному відборі буде визначена [24; 48]:

1) для середньої кількісної ознаки:

$$\mu_x = \sqrt{\frac{\delta_x^2}{r} \times \left(1 - \frac{r}{R}\right)}, \quad (7.19)$$

де R – загальне число серій в генеральній сукупності;

2) для частки альтернативної ознаки:

$$\mu_w = \sqrt{\frac{\delta_w^2}{r} \times \left(1 - \frac{r}{R}\right)}. \quad (7.20)$$

Щоб розрахувати граничну помилку вибірки для середньої ознаки та для частки з використанням розглянутих способів відбору, треба середню помилку вибірки помножити на коефіцієнт довіри t , значення якого залежить від рівня обраної ймовірності.

7.3. Особливості малої вибірки та визначення необхідної чисельності вибірки

Одним із факторів, що впливає на величину помилки вибірки, є її *чисельність*. Чим більша чисельність вибірки, тим менше помилка. З іншого боку, з обсягом вибірки пов'язані витрати на проведення дослідження: чим більше обсяг, тим більші витрати. Отже, вибірка має бути оптимальною за чисельністю, щоб забезпечити достовірність результатів дослідження і не викликати додаткових витрат.

Чисельність вибірки може бути визначена виходячи із допустимої помилки у вибірковому спостереженні та методу відбору статистичних одиниць. У загальному випадку гранична помилка вибірки пов'язана з її чисельністю таким співвідношенням [68]:

$$\Delta = t \times \mu = t \times \sqrt{\frac{\sigma_B^2}{n}}, \quad \text{звідки} \quad n = \frac{t^2 \times \sigma_B^2}{\Delta^2}. \quad (7.21)$$

Тобто зі збільшенням передбачуваної помилки значно зменшується необхідний обсяг вибірки та навпаки. За умови різних характеристик і різних методів формування вибірок, необхідна чисельність вибірки буде визначатися за формулами, поданими в табл. 7.4 [24; 48; 68].

Таблиця 7.4

Чисельність вибірки за різних методів відбору

Методи відбору	Формули визначення обсягу вибірки	
	Для середньої	Для частки
Повторний	$n = \frac{t^2 \times \sigma_{\bar{x}}^2}{\Delta_{\bar{x}}^2}$	$n = \frac{t^2 \times w \times (1 - w)}{\Delta_p^2}$
Безповторний	$n = \frac{t^2 \times \sigma_{\bar{x}}^2 \times N}{N \times \Delta_{\bar{x}}^2 + t^2 \times \sigma_{\bar{x}}^2}$	$n = \frac{t^2 \times w \times (1 - w) \times N}{N \times \Delta_p^2 + t^2 \times w \times (1 - w)}$

Приклад 7.2. Розрахуємо оптимальний обсяг вибірки для середньої ознаки шляхом безповторного відбору на основі таких даних: $N = 2\ 000$ одиниць, $\sigma_{\bar{x}} = \pm 1$ мм, $\Delta_{\bar{x}} = 0,5$ мм, $t = 2$ за $P = 0,954$.

$$n_{\bar{x}} = \frac{N \times t^2 \times \sigma_{\bar{x}}^2}{\Delta_{\bar{x}}^2 \times N + t^2 \times \sigma_{\bar{x}}^2} = \frac{2\ 000 \times 2^2 \times 1^2}{0,5^2 \times 2\ 000 + 2^2 \times 1^2} = 148 \text{ од.}$$

Отже, вибірка у 148 одиниць забезпечить задану точність результатів за безповторного відбору.

У статистичному аналізі часто виникає потреба в порівнянні помилок вибірки різних ознак або тієї самої ознаки в різних сукупностях. Такі порівняння здійснюють за допомогою **відносної помилки вибірки** (коефіцієнта варіації) u_{μ} . **Коефіцієнт варіації** – це відносна стандартна помилка, яка визначається шляхом ділення стандартної (середньої) помилки

в абсолютному вираженні на середнє значення оцінки показника; виражається як коефіцієнт або у відсотках [68]:

$$u_{\mu} = \frac{\mu_{\bar{x}}}{\bar{x}} \times 100, \quad (7.22)$$

де $\mu_{\bar{x}}$ – стандартна (середня) помилка вибірки.

Обчислення відносної стандартної помилки рекомендується для додатних кількісних показників і не рекомендується для відносних величин (особливо часток, величин динаміки, тому що в цих випадках дуже легко хибно інтерпретувати її змістовність).

Відносна помилка u_{μ} показує, на скільки відсотків може відхилитися вибіркова оцінка від параметра генеральної сукупності.

Вибіркове спостереження, чисельність одиниць якого не перевищує 30, називають **малою вибіркою**.

Для визначення можливих меж помилки використовується *критерій Стьюдента* [68]:

$$t = \frac{\bar{x} - \bar{x}}{\mu_{MB}}, \quad (7.23)$$

де $\mu_{MB} = \frac{\sigma}{\sqrt{n-1}}$ – стандартна помилка малої вибірки.

Гранична помилка малої вибірки визначається як [68]:

$$\Delta_{MB} = t \times \mu_{MB}, \quad (7.24)$$

де t – критерій Стьюдента.

На основі встановленої закономірності розподілу помилок малих вибірок складено спеціальні таблиці, в яких наведено значення t -критерію Стьюдента та відповідних рівнів ймовірності щодо різної кількості одиниць вибірки. Відповідно до таблиці значень ймовірності для t -критерію Стьюдента визначається ймовірність того, що фактичне значення t для випадкових величин не буде більше табличного за абсолютною величиною.

Кінцевою метою будь-якого вибіркового спостереження є розповсюдження його характеристик на генеральну сукупність.

Методи поширення даних вибіркового спостереження [13; 48; 68; 74]:

спосіб прямого перерахунку: середній розмір ознаки, що визначений шляхом вибіркового спостереження, множать на обсяг генеральної сукупності;

спосіб коефіцієнтів: зіставляючи дані вибіркового спостереження із суцільним, обчислюють коригувальний коефіцієнт, який використовують для внесення поправок у матеріали суцільного спостереження.

Коригувальний коефіцієнт розраховують як [68]:

$$Y_1 = Y_0 \times \frac{y_1}{y_0}, \quad (7.25)$$

де Y_1 – чисельність сукупності з поправкою на недооблік;

Y_0 – чисельність сукупності без цієї поправки;

y_1 – чисельність сукупності в контрольних точках за первісними даними;

y_0 – чисельність сукупності в тих же точках за даними контрольних заходів.

Прямий перерахунок даних вибірки на всю генеральну сукупність застосовується, коли метою дослідження є визначення обсягу ознаки генеральної сукупності за наявності лише кількості її одиниць.

У випадку, коли вибірконе спостереження проводять з метою уточнення результатів суцільного спостереження, застосовують спосіб коефіцієнтів.

Важливі поняття

Вибіркова сукупність (вибірка) (n) – частина одиниць генеральної сукупності, що відібрана для вивчення.

Власне випадковий відбір – проводять методом жеребкування або за таблицею випадкових чисел.

Генеральна сукупність (N) – досліджувана сукупність, з якої проводиться відбір одиниць, що підлягають вивченню.

Гранична помилка вибірки – максимум помилки за заданою ймовірністю її появи.

Інтервальна оцінка – довірчий інтервал значень параметра за заданою ймовірністю.

Комбінована вибірка – використання декількох способів формування вибірки в одному вибірковому дослідженні.

Мала вибірка – вибірка, чисельність одиниць якої не перевищує 30.

Методи відбору (повторний, безповторний) – визначення, які одиниці (вже відібрані чи ні) беруть участь у подальшій виборці.

Методи поширення даних вибіркового спостереження – метод прямого перерахунку та метод коефіцієнтів.

Механічна вибірка – відбір одиниць із генеральної сукупності відбувається у довільному механічному порядку.

Наукові принципи теорії вибіркового методу – забезпечення випадковості відбору одиниць; забезпечення достатньої кількості відібраних одиниць.

Помилка репрезентативності – розбіжність між певною характеристикою генеральної сукупності (часткою, середньою, дисперсією тощо) та її вибірковою оцінкою.

Середня (стандартна) помилка вибірки – розходження між середньою вибіркової та середньою генеральної сукупностей, яке не перевищує $\pm\sigma$.

Серійна вибірка – застосовується в тих випадках, коли одиниці статистичної сукупності об'єднані в невеликі групи (серії).

Спосіб відбору – система організації відбирання одиниць із генеральної сукупності.

Статистична оцінка (параметра) генеральної сукупності – наближене значення шуканої характеристики (параметра), яке отримане за даними вибірки.

Типова вибірка – генеральна сукупність попередньо розбивається на однорідні групи, з яких проводять випадковий або механічний відбір.

Точкова оцінка – оцінка, яка характеризує значення параметра й обчислена на основі даних вибірки.

Якість статистичних оцінок – визначається такими їхніми властивостями, як спроможність, незміщеність, ефективність.

Типові завдання

Завдання 7.1. Для визначення середнього періоду користування кредитом в банку була здійснена п'ятивідсоткова механічна вибірка, до якої потрапило 100 рахунків. У результаті обстеження встановлено,

що середній термін користування кредитом – 30 днів за середньоквадратичного відхилення у 9 днів. З ймовірністю 0,997 необхідно визначити межі, в яких буде перебувати термін користування кредитом у генеральній сукупності.

Розв'язання.

Середній термін користування кредитом у банку знаходиться в межах:

$$\tilde{x} - \Delta\tilde{x} \leq \bar{x} \leq \tilde{x} + \Delta\tilde{x}.$$

Оскільки вибірка механічна, то граничну помилку вибірки знаходять таким чином:

$$\Delta\tilde{x} = t \times \sqrt{\frac{\sigma_{\tilde{x}}^2}{n} \times \left(1 - \frac{n}{N}\right)} = 3 \times \sqrt{\frac{9^2}{100} \times \left(1 - \frac{100}{2000}\right)} = 2,625 \approx 3 \text{ дні.}$$

Отже, з імовірністю 99,7 % можна стверджувати, що термін користування кредитом у генеральній сукупності перебуває в межах $30 - 3 \leq \bar{x} \leq 30 + 3$, тобто від 27 до 33 днів.

Завдання 7.2. Маємо дані про середньооблікову чисельність робітників на підприємстві – 500 осіб. За матеріалами спеціального вибіркового обстеження стажу роботи працівників підприємства виявлено, що 50 з них мають стаж роботи понад 20 років. Вибір проводився способом випадкової неповторної вибірки.

Потрібно з ймовірністю $p = 0,954$ визначити: а) частку (питому вагу) робітників зі стажем роботи понад 20 років у вибірковій сукупності; б) середню і граничну помилки вибіркової частки; в) межі, в яких перебуває частка робітників зі стажем роботи понад 20 років у генеральній сукупності. За результатами розрахунків зробіть висновки.

Розв'язання.

1. Визначимо вибірку частку (питому вагу) робітників зі стажем роботи понад 20 років:

$$w = \frac{n}{N} = \frac{50}{500} = 0,10, \text{ або } 10,0 \%$$

2. Обчислимо середню помилку вибіркової частки:

$$\mu_p = \sqrt{\frac{w \times (1-w)}{n} \times \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0,1 \times (1-0,1)}{50} \times \left(1 - \frac{50}{500}\right)} = \sqrt{\frac{0,1 \times 0,9 \times 0,9}{50}} = 0,04$$

або 4,0 %.

3. Для проведення інтервального оцінювання і побудови інтервалу для генеральної частки визначимо граничну помилку вибіркової частки. За таблицю знайдемо, що заданій ймовірності $p = 0,954$ відповідає нормоване відхилення $t = 2$.

4. Гранична помилка вибіркової частки становитиме:

$$\Delta w = t \times \mu_p = 2,0 \times 0,04 = 0,08, \text{ або } 8,0 \%$$

5. Побудуємо інтервал, в якому із заданою ймовірністю перебуває генеральна частка робітників підприємства зі стажем роботи понад 20 років:

$$\begin{aligned} w - \Delta w < p < w + \Delta w; \\ 0,10 - 0,08 < p < 0,10 + 0,08; \\ 0,02 < p < 0,18; \\ \text{або } 2\% < p < 18\%. \end{aligned}$$

Отже, з ймовірністю $p = 0,954$ можна стверджувати, що частка робітників підприємства зі стажем роботи понад 20 років у генеральній сукупності перебуває в інтервалі від 2,0 до 18,0 %.

Завдання 7.3. У цеху підприємства працює 10 бригад робітників. З метою вивчення їх продуктивності праці була здійснена двадцятивідсоткова серійна вибірка, до якої потрапило 2 бригади. У результаті обстеження встановлено, що середня продуктивність праці робітників у бригадах становила 4,6 і 3 тис. ум. од. З ймовірністю 0,954 визначте межі, в яких перебуватиме середня продуктивність робітників цеху.

Розв'язання.

Середня продуктивність робітників цеху перебуває в межах:

$$\tilde{x} - \Delta \tilde{x} \leq \bar{x} \leq \tilde{x} + \Delta \tilde{x}.$$

Вибіркова середня серійної вибірки складе:

$$\tilde{x} = \frac{4,6 + 3}{2} = 3,8 \text{ тис. ум. од.}$$

Дисперсію серійної вибірки розрахуємо таким чином:

$$\delta_{\tilde{x}}^2 = \frac{\sum(\tilde{x}_i - \tilde{x})^2}{r},$$

де \tilde{x}_i – вибіркова середня серії;

\tilde{x} – вибіркова середня серійної вибірки;

r – кількість серій.

Отже, значення дисперсії складе: $\delta_{\tilde{x}}^2 = \frac{(4,6 - 3,8)^2 + (3 - 3,8)^2}{2} = 0,64.$

Граничну помилку вибірки для середньої розрахуємо таким чином:

$$\Delta_{\tilde{x}} = t \sqrt{\frac{\delta_{\tilde{x}}^2}{r} \times \left(1 - \frac{r}{R}\right)} = 2 \sqrt{\frac{0,64}{2} \times \left(1 - \frac{2}{10}\right)} = 1 \text{ тис. ум. од.}$$

Отже, з імовірністю 0,954 можна стверджувати, що середня продуктивність праці робітників цеху перебуває в межах $3,8 - 1 \leq \bar{x} \leq 3,8 + 1$, тобто від 2,8 до 4,8 тис. ум. од.

Завдання 7.4. Відділ маркетингових досліджень вирішив провести дослідження попиту на мийний засіб А серед сукупності домогосподарок міста. Визначте обсяг вибіркової сукупності з ймовірністю 0,954, щоб помилка вибірки не перевищувала 2 %, якщо за даними попередніх обстежень відомо, що дисперсія дорівнює 0,08.

Розв'язання.

Оскільки в завданні не відома чисельність генеральної сукупності домогосподарок міста, то буде використовуватись формула розрахунку необхідної чисельності вибірки для повторного відбору:

$$n = \frac{t^2 \times \sigma_{\tilde{x}}^2}{\Delta_{\tilde{x}}^2} = \frac{2^2 \times 0,08}{0,02^2} = 800 \text{ осіб.}$$

Отже, для проведення вказаного спостереження, щоб задовільнити певним вимогам, необхідно обстежити 800 домогосподарок міста.

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Набуття навичок проведення вибіркового спостереження за допомогою MS Excel" подані у роботі [67]. Лабораторна робота націлена на отримання навичок із проведення вибіркового обстеження та оброблення його результатів за допомогою табличного процесору MS Excel.

Запитання для самодіагностики

1. У чому полягає сутність вибіркового спостереження, необхідність і доцільність його застосування?
2. Які наукові умови застосування вибіркового спостереження?
3. Які ви знаєте методи відбору одиниць у вибірку сукупність?
4. У чому полягає сутність механічного та типового відбору?
5. У чому полягає сутність серійного та комбінованого відбору?
6. Охарактеризуйте середню помилку вибірки, наведіть алгоритм її розрахунку.
7. Охарактеризуйте граничну помилку вибірки, наведіть алгоритм її розрахунку.
8. Що таке точкова й інтервальна оцінка генеральної середньої, довірчий інтервал?
9. Як розраховується необхідна чисельність вибірки та від яких факторів вона залежить?
10. Що таке мала вибірка та в чому полягають особливості визначення помилок у таких вибірках?

Питання для критичного переосмислення (есе)

1. Порівняйте між собою повторну та безповторну вибірки, визначте їх переваги та обмеження.
2. Проведено обстеження:
 - а) кожної сотні сім'ї працівників галузі промисловості з метою вивчення життєвого рівня населення промислового регіону;
 - б) п'яти багатодітних сімей працівників промислового підприємства з метою вивчення їх житлових умов.

Як ви вважаєте, яке з цих обстежень є вибірковим і чому? Відповідь обґрунтуйте.

3. Як ви вважаєте, чи погоджуються вибіркові дані з твердженням, що 30 % підприємницьких структур не сплачують податки в повному обсязі, якщо відомо, що із 450 обстежених бізнес-структур у 125 виявлено порушення в заповненні податкової декларації стосовно суми доходу, який підлягає оподаткуванню? Визначте частку підприємницьких структур, які приховують частину доходів від оподаткування, та довірчі межі цієї частки в генеральній сукупності. Висновок зробіть з ймовірністю у 99,7 %.

4. Проєктується опитування підприємців з приводу оцінювання економіко-правових умов їхньої діяльності. Як ви вважаєте, яким повинен бути обсяг вибірки, щоб з імовірністю 0,954 відносна похибка вибірки не перевищувала 8 %? За результатами попереднього обстеження відомо, що 75 % підприємців були незадоволені умовами діяльності.

5. У соціологічному дослідженні взяли участь 350 респондентів. За результатами опитування виявлено, що: 15 % уникає ризику; 12 % задіяні в операціях із цінними паперами. На вашу думку, скільки становить середня помилка вибірки для частки опитуваних, що уникає ризику, та якою вона є для частки респондентів, які працюють на ринку цінних паперів? Відповідь обґрунтуйте.

8. Аналіз концентрації, диференціації та подібності рядів розподілу

Основні питання:

8.1. Порядкові характеристики розподілу.

8.2. Статистичні показники концентрації та централізації.

8.3. Статистичне оцінювання структурних змін у часі та просторі.

8.1. Порядкові характеристики розподілу

Квантилі призначені для більш глибокого вивчення структури ряду розподілу. **Квантиль** – це значення ознаки, що займає певне місце у впорядкованій за цією ознакою сукупності.

Розрізняють такі види квантилів [14; 43; 49; 51; 70; 78; 84; 85; 87; 100]:

- **квартилі** ($Q_{1/4}$, $Q_{2/4} = Me$, $Q_{3/4}$) – значення ознаки, що ділять впорядковану сукупність на чотири рівні частини;
- **децилі** (D_1, D_2, \dots, D_9) – значення ознаки, що ділять сукупність на десять рівних частин;
- **перцентелі** – значення ознаки, що ділять сукупність на сто рівних частин.

Якщо дані згруповані, то значення квантиля визначається за накопиченими частотами: номер групи, яка містить i -й квантиль. Визначається як номер першої групи від початку ряду, в якому сума накопичених частот дорівнює або перевищує $i \times S$, де i – індекс квантиля.

Якщо ряд інтервальний, то значення квантиля визначається за формулою [14; 43; 49; 51; 70; 78; 84; 85; 87; 100]:

$$Q_i = x_{Q_i}^n + h_{Q_i} \times \frac{i \times S - S_{Q_{i-1}}}{f_{Q_i}}, \quad (8.1)$$

де $x_{Q_i}^n$ – нижня межі інтервалу, в якому перебуває i -й квантиль;

$S_{Q_{i-1}}$ – сума накопичених частот інтервалів, які передують інтервалу, в якому знаходиться i -й квантиль;

f_{Q_i} – частота інтервалу, в який потрапив i -й квантиль.

Розрахуємо квартали для ряду розподілу співробітників фірми за стажем роботи [14; 43; 49; 51; 70; 78; 84; 85; 87; 100]: нижній квантиль $Q_{1/4}$ відповідає 13-й одиниці, верхній квантиль $Q_{3/4}$ – 38-й. Це, відповідно, друга та четверта групи [78; 84; 85; 87; 100].

$$Q_{1/4} = 4 + 4 \times \frac{\frac{50}{4} - 6}{8} = 4 + 4 \times 0,8 \approx 7,2 \text{ роки};$$
$$Q_{3/4} = 12 + 4 \times \frac{\frac{3 \times 50}{4} - 25}{13} = 12 + 4 \approx 16 \text{ років.}$$

Отже, у чверті співробітників фірми стаж менше 7 років і у чверті – більше 16 років.

Значення ознаки, що ділять ряд на п'ять рівних частин, називають **квінтиле**, на десять частин – **децилями**, на сто частин – **перцентелями**.

Формула для розрахунку першого та дев'ятого децилів має вигляд [84; 85; 87]:

$$D_1 = x_{D_1} + \frac{i}{f_{D_1}} \times \left(\frac{\sum_{i=1}^k f_i}{10} - S_{D_1-1} \right); \quad (8.2)$$

$$D_9 = x_{D_9} + \frac{i}{f_{D_9}} \times \left(\frac{9 \times \sum_{i=1}^k f_i}{10} - S_{D_9-1} \right), \quad (8.3)$$

де x_{D_1} – нижня межа інтервалу, в якому перебуває перший дециль;

x_{D_9} – нижня межа інтервалу, в якому знаходиться дев'ятий дециль;

$\sum_{i=1}^k f_i$ – кількість спостережень;

S_{D_1-1} – накопичена частота в інтервалі, що передує першому децильному;

S_{D_9-1} – накопичена частота в інтервалі, що передує дев'ятому децильному;

f_{D_1} – частота в першому децильному інтервалі;

f_{D_9} – частота в дев'ятому децильному інтервалі;

i – величина інтервалу.

Якщо центр розподілу поданий медіаною, то за відносну міру варіації беруть *квартильний коефіцієнт варіації* [70; 78; 84; 85; 87; 93; 103]:

$$V_Q = \frac{Q_3 - Q_1}{2 \times Me}. \quad (8.4)$$

Відносне квартильне відхилення [70; 78; 84; 85; 87; 93; 103]:

$$V_Q = \frac{Q}{\bar{x}} \times 100\%, \text{ або } V_Q = \frac{Q}{Me} \times 100\%, \text{ або } V_Q = \frac{Q_3 - Q_1}{2Me}.$$

Для оцінювання ступеня варіації застосовують також співвідношення децилів. Так, *коефіцієнт децильної диференціації* показує кратність співвідношення дев'ятого (D_9) і першого (D_1) децилів [70; 78; 84; 85; 87; 93; 103]:

$$V_D = \frac{D_9}{D_1}. \quad (8.5)$$

Аналіз закономірностей розподілу передбачає оцінювання ступеня однорідності сукупності, асиметрії і ексцесу розподілу. З'ясування загального характеру розподілу передбачає не тільки оцінювання ступеня його однорідності, а і дослідження форми розподілу, тобто оцінювання симетричності й ексцесу.

З математичної статистики відомо, що зі збільшенням обсягу статистичної сукупності ($\infty \rightarrow N$) і одночасним зменшенням інтервалу групування ($x_i \rightarrow 0$) полігон або гістограма розподілу інтенсивно наближається до деякої плавної кривої, що є для зазначених графіків межею. Цю криву називають *емпіричною кривою розподілу*; вона подається графічним зображенням у вигляді безперервної лінії зміни частот, функціонально пов'язаної зі зміною варіант.

У статистиці розрізняють такі ***види кривих розподілу*** [1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70; 78]: одновершинні криві; багатoverшинні криві.

Однорідні сукупності описуються ***одновершинними*** розподілами. ***Багатoverшинність*** розподілу свідчить про неоднорідність досліджуваної сукупності або про неякісне виконання угруповання.

Одновершинні криві розподілу поділяють на ***симетричні, помірно асиметричні*** та ***вкрай асиметричні***.

Розподіл є симетричним, якщо частоти будь-яких двох варіантів, рівновіддалених в обидва боки від центру розподілу, рівні між собою. У таких розподілах $\bar{x} = M_0 = M_e$.

Для характеристики асиметрії використовують ***коефіцієнти асиметрії***.

Найбільш часто використовують ***коефіцієнт асиметрії Пірсона*** [1; 5; 13; 23; 28; 31]:

$$A_s = \frac{\bar{x} - M_0}{\sigma}. \quad (8.6)$$

У одновершинних розподілах величина цього показника змінюється від -1 до $+1$. У симетричних розподілах $A_s = 0$.

У разі $A_s > 0$ спостерігається правостороння асиметрія (рис. 8.1). У розподілі з правобічною асиметрією $M_o \leq M_e \leq \bar{x}$.

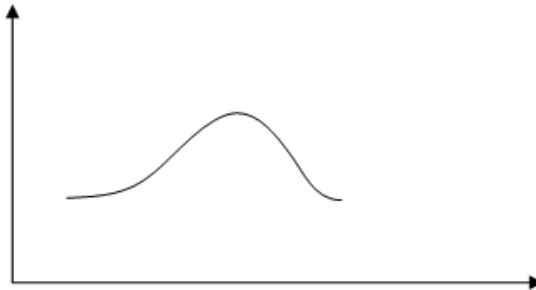


Рис. 8.1. Правостороння асиметрія [1; 5; 13; 14; 23; 70; 78; 84; 85; 87]

За $A_s < 0$ – асиметрія негативна лівостороння, $M_o > M_e > \bar{x}$ (рис. 8.2).

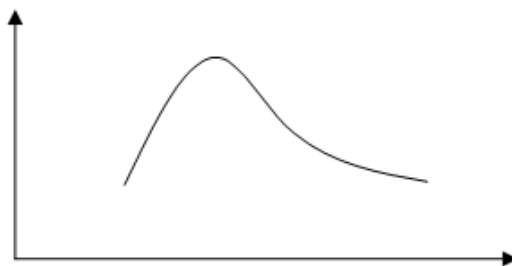


Рис. 8.2. Лівостороння асиметрія [1; 5; 13; 14; 23; 70; 78; 84; 85; 87]

Чим ближче по модулю A_s до 1, тим асиметрія істотніше:

- якщо $|A_s| < 0,25$, то асиметрія вважається незначною;
- якщо $0,25 < |A_s| < 0,5$, то асиметрія вважається помірною;
- якщо $|A_s| > 0,5$ – асиметрія значна.

Коефіцієнт асиметрії Пірсона характеризує асиметрію тільки в центральній частині розподілу, тому більш поширеним і більш точним є **коефіцієнт асиметрії**, розрахований на основі центрального моменту третього порядку [1; 5; 13; 14; 23; 70; 78; 84; 85; 87]:

$$A_s = \mu_3 : \sigma^3, \quad (8.7)$$

де μ_3 – центральний момент третього порядку;

σ^3 – середнє квадратичне відхилення в третьому ступені.

Алгебраїчний центральний момент розподілу – це середня арифметична k -го ступеня відхилення індивідуальних значень ознаки від середньої:

а) центральний момент для незгрупованого ряду [78; 84; 85; 87]:

$$\mu_k = \frac{\sum_{i=1}^m (x_i - \bar{x})^k}{n}; \quad (8.8)$$

б) центральний момент для згрупованого ряду [1; 5; 13; 14; 23; 70; 78; 84; 85; 87]:

$$\mu_k = \frac{\sum_{i=1}^m (x_i - \bar{x})^k f_i}{\sum_{i=1}^m f_i}. \quad (8.9)$$

Очевидно, що момент другого порядку є дисперсією, яка характеризує варіацію. Моменти третього та четвертого порядків характеризують, відповідно, асиметрію та ексцес.

Відповідно, формули для визначення центрального моменту третього порядку мають такий вигляд:

а) центральний момент для незгрупованого ряду [1; 5; 13; 14; 23; 70; 78; 84; 85; 87]:

$$\mu_3 = \frac{\sum_{i=1}^m (x_i - \bar{x})^3}{n}; \quad (8.10)$$

б) центральний момент для згрупованого ряду [1; 5; 13; 23; 70; 84; 87]:

$$\mu_3 = \frac{\sum_{i=1}^m (x_i - \bar{x})^3 f_i}{\sum_{i=1}^m f_i}. \quad (8.11)$$

У симетричному розподілі $\mu_3 = 0$. Чим більше скошеність ряду, тим більше значення μ_3 .

Теоретично коефіцієнт асиметрії не має меж, однак на практиці його значення не буває занадто великим і в помірному розподілу не перевищує одиниці.

Для оцінювання суттєвості розрахованого другим способом коефіцієнта асиметрії визначається його середня квадратична помилка [1; 5; 13; 23; 28]:

$$\sigma_{As} = \sqrt{\frac{6 \times (N - 1)}{(N + 1) \times (N + 3)}} \quad (8.12)$$

Якщо $\frac{|A_s|}{\sigma_{as}} > 3$, то асиметрія є істотною.

Для одновершинних розподілів розраховується ще один показник оцінки його форми – **ексцес**. Ексцес є показником гостровершинності розподілу. Він розраховується для симетричних розподілів на основі центрального моменту четвертого порядку [14; 70; 78; 84; 85; 87; 99; 100]:

$$E_x = \frac{\mu^4}{\sigma^4} - 3 \quad (8.13)$$

У симетричному, близькому до нормального розподілу $E_x = 0$. Очевидно, у гостровершинному розподілі $E_x > 0$, у плосковершинному $E_x < 0$.

Для оцінювання суттєвості коефіцієнта ексцесу розраховується середня квадратична помилка за формулою [1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70]:

$$\sigma_{Ex} = \sqrt{\frac{24 \times n \times (n - 2) \times (n - 3)}{(n - 1)^2 \times (n + 3) \times (n + 5)}} \quad (8.14)$$

Результат оцінювання суттєвості показників асиметрії і ексцесу дозволяє зробити висновок про те, чи можна віднести цей емпіричний розподіл до типу кривих нормального розподілу.

Розрахуємо показники асиметрії і ексцесу для ряду розподілу співробітників компанії за стажем роботи. Маємо такі характеристики для поданого ряду: $\bar{X} = 12$ років, $M_0 = 12,9$ років, $\sigma = 6,3$ роки.

Коефіцієнт асиметрії Пірсона дорівнює [85; 87; 93; 99; 100]:

$$A_s = \frac{\bar{x} - Mo}{\sigma} = \frac{12 - 12,9}{6,3} \approx -0,14 < 0,$$

що говорить про наявність незначної лівосторонньої асиметрії в центральній частині розподілу.

Коефіцієнт асиметрії, розрахований через центральний момент третього порядку [1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 99; 100]:

$$A_s = \frac{\mu_3}{\sigma^3} = \frac{\sum(x_i - \bar{x}) \cdot f_i}{\sum f_i} = \frac{61,44}{6,3^3} = \frac{61,44}{250} = 0,24 > 0.$$

Це означає, що в цілому в усьому ряді спостерігається правостороння асиметрія.

Розрахунок центрального моменту третього порядку μ^3 наведено в допоміжній табл. 8.1.

Таблиця 8.1

Розрахунок центральних моментів третього і четвертого порядку
[1; 5; 13; 23; 28; 31; 93; 98 – 100; 103]

№ п/п	x_i	f_i	$x_i - \bar{x}$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^3 * f_i$	$(x_i - \bar{x})^4$	$(x_i - \bar{x})^4 * f_i$
1	2	6	-10	-1 000	-6 000	10 000	60 000
2	6	8	-6	-216	-1 728	1 296	10 368
3	10	11	-2	-8	-88	16	176
4	14	13	2	8	104	16	208
5	18	6	6	216	1 296	1 296	7 776
6	22	4	10	1 000	4 000	10 000	40 000
7	26	2	14	2 744	5 488	38 416	76 832
Разом	14	50	-	-	3 072		195 360

Показник ексцесу [1; 5; 78; 84; 85; 87; 93; 98 – 100; 103]:

$$\sum x = \frac{\mu_4}{\sigma^4} - 3 = \frac{195\,360}{50} \div 6,3^4 - 3 = \frac{3\,907,2}{1575,3} - 3 = 2,5 - 3 = -0,5$$

свідчить про те, що розподіл плосковершинний.

8.2. Статистичні показники концентрації та централізації

У статистичному аналізі поряд із характеристиками нерівномірності розподілу певної ознаки між окремими складовими сукупності важливим є оцінювання концентрації значень ознаки в окремих її частинах (розподіл майна або доходів між окремими групами населення, кількості зайнятих за окремими видами економічної діяльності, площ сільськогосподарських угідь між окремими агропромисловими комплексами).

Тобто, *оцінювання відмінностей* двох розподілів у просторі та часі базується на *порівнянні часток* цих розподілів [14; 43; 49; 51; 70; 78; 84]: частки розподілу елементів сукупності, частки розподілу значень ознаки.

Приклад 8.1. Наведені в табл. 8.2 дані свідчать про нерівномірний розподіл територіальних громад за обсягом отриманих "теплих кредитів" і за обсягами спожитої електроенергії.

До першої групи належить 20 % територіальних громад, а частка спожитої електроенергії становить 4 %. Проте остання група охоплює лише 3 % територіальних громад, які споживають 46 % електроенергії. На відхиленнях частин двох розподілів – за кількістю елементів сукупності d_j та обсягом значень ознак D_j – ґрунтується *оцінка концентрації*.

Таблиця 8.2

Розрахункові дані

[14; 70; 78; 84; 85; 87; 93; 97 – 100; 103]

Обсяг отриманих "теплих кредитів", тис. грн	У % до всього		Модуль відхилення часток $\frac{1}{100} d_j - D_j $
	Частка територіальних громад d_j	Частка спожитої електроенергії D_j	
До 5	20	4	0,16
5 – 10	38	5	0,33
10 – 20	22	8	0,14
20 – 50	13	12	0,01
50 – 100	4	25	0,21
100 і більше	3	46	0,43
Усього	100	100	1,28

Якщо розподіл значень ознаки в сукупності *рівномірний*, то частки однакові: $d_j = D_j$. Відхилення часток свідчать про певну концентрацію.

Верхня межа суми відхилень $\sum |d_j - D_j| = 2$, а тому **коефіцієнт концентрації** обчислюється як напівсума модулів відхилень [1; 5; 13; 23]:

$$K = \frac{1}{2} \sum_{j=1}^m |d_j - D_j|, \quad (8.15)$$

де d_j – частка розподілу елементів сукупності;

D_j – частка розподілу значень ознаки.

Значення коефіцієнта концентрації коливаються в межах від нуля (рівномірний розподіл) до одиниці (повна концентрація). Чим більше ступінь концентрації, тим більше значення коефіцієнта K . У нашому прикладі $K = 1,28 : 2 = 0,64$, що свідчить про високий ступінь концентрації споживання електроенергії промисловими підприємствами.

Коефіцієнти концентрації широко використовують у регіональному аналізі для оцінювання рівномірності територіального розподілу виробничих потужностей, фінансових ресурсів та ін.

Локалізацію значень ознаки в окремих складових сукупності визначають за допомогою **коефіцієнта локалізації**, що характеризує співвідношення часток [23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 99; 100]:

$$L_j = \frac{D_j}{d_j} \times 100. \quad (8.16)$$

За даними табл. 8.3 коефіцієнти локалізації свідчать про нерівномірність отримання замовлень на послуги компаній певною мірою про сезонність цих послуг.

Таблиця 8.3

Розрахункові дані

[14; 51; 70; 78; 84; 85; 87; 93; 98 – 100]

Компанії	Частка відвідувань сайту компанії, d_j	Частка отриманих замовлень на послуги, D_j	Коефіцієнт локалізації L_j , %
А	30	34	113
В	50	42	84
С	20	24	120
Усього	100	100	*

За аналогією з коефіцієнтом концентрації обчислюється **коефіцієнт подібності (схожості) структур** двох сукупностей [93; 98; 103]:

$$P = 1 - \frac{1}{2} \sum_1^m |d_j - d_k|, \quad (8.17)$$

де d_j і d_k – частка розподілу елементів сукупності j -ї та k -ї сукупності.

Якщо структури сукупностей однакові, то $P = 1$; якщо абсолютно протилежні, то $P = 0$. Чим більше схожі структури, тим більше значення P .

За наведеними в табл. 8.4 даними щодо структури куплених турів за часом за двома їх видами розрахуємо коефіцієнт подібності структур. Структури куплених турів подібні за умови, що отримане значення P буде наближатися до одиниці [1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 93; 99; 100]:

$$P = 1 - \frac{|36 - 25| + |24 - 42| + |40 - 33|}{2} = 1 - 0,18 = 0,82.$$

Таблиця 8.4

Структура куплених турів за часом, %

Види турів	Частка куплених турів за часом		
	Міжсезонний	Свята та вихідні	Сезон
Пляжний	36	24	40
Екскурсійний	25	42	33

Порівняння структур на основі відхилень частин доцільне для рядів із нерівними інтервалами, особливо в атрибутивних рядах. Аналіз варіації в рядах розподілу необхідно доповнювати розрахунком показника диференціації.

Коефіцієнт диференціації ($K_{\text{диф}}$) визначається як співвідношення двох середніх, отриманих з 10 % найбільших і найменших значень ознаки, що досліджується.

Приклад 8.2. Маємо дані про розміри прибутку 20 комерційних банків. Найменше значення прибутку мають 2 банки (10 % від загальної сукупності),

що відповідає: 3,7 і 4,3 млн грн; найбільше значення мають ще два банки, чий розмір показника складе: 7,9 і 8,1 млн грн. Середнє значення за найменшим значенням прибутку буде дорівнювати 4,0 млн грн; середнє значення в групі найбільш прибуткових банків – 8,0 млн грн. У цьому випадку [1; 28; 31; 43; 49; 51; 70; 78; 93; 98 – 100; 103]:

$$(K_{\text{диф}}) = \frac{\bar{X}_{\text{найбільші}}}{\bar{X}_{\text{найменші}}} = \frac{8}{4} = 2. \quad (8.18)$$

Тобто розмір прибутку у 10 % банків із найвищими доходами вдвічі перевищує розмір прибутку 10 % банків із найменшими доходами.

Централізація означає зосередженість (скупченість) обсягу ознак окремих одиниць (наприклад, капіталу в окремих комерційних банках, продукції якогось виду на окремих підприємствах та ін.).

Узагальнювальний показник централізації I_z розраховується за формулою [14; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 93; 98 – 100; 103]:

$$I_z = \sum_{i=1}^n \left(\frac{m_i}{\sum_{i=1}^n m_i} \right)^2, \quad (8.19)$$

де m_i – значення ознаки i -ї одиниці сукупності;

$\sum_{i=1}^n m_i$ – обсяг ознаки всієї сукупності;

n – обсяг сукупності (кількість одиниць, що входять у сукупність).

Максимального значення I_z досягає тільки за умови, що сукупність складається тільки із однієї одиниці, якій належить весь обсяг ознаки. Мінімальне значення цього показника наближається до нуля, але ніколи його не досягає.

Оцінювання нерівномірності розподілу між окремими складовими сукупності ґрунтується на порівнянні часток двох розподілів – за кількістю елементів сукупності d_i і обсягом значень ознаки D_i . Якщо розподіл значень ознаки рівномірний, то $d_i = D_i$; відхилення часток свідчать про певну нерівномірність, яка вимірюється коефіцієнтами локалізації та концентрації [1; 5; 13; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 93; 98 – 100; 103].

Різновидом кумуляти є **крива концентрації**, або **графік Лоренца**. Для побудови кривої концентрації на обидві осі прямокутної системи координат наноситься масштабна шкала у відсотках від 0 до 100. На осі абсцис указують накопичені частоти, що характеризують розподіл одиниць сукупності, а на осі ординат – накопичені значення частки (у відсотках) за обсягом ознаки. На графіку рівномірному розподілу ознаки відповідає діагональ квадрата. За нерівномірного розподілу графік є увігнутою кривою – залежно від рівня концентрації ознаки (рис. 8.3).

Графіком функції (в прямокутній системі координат) є крива Лоренца, яка опукла вниз і проходить під діагоналлю одиничного квадрата, що розташований у першій координатній чверті.

Кожна точка на кривій Лоренца відповідає твердженню на зразок: 20 % найбіднішого населення отримують 7 % від його сукупного доходу. У випадку абсолютно рівного розподілу кожна група населення має дохід, який пропорційний її чисельності. Такий випадок описується кривою рівності (line of perfect equality), що є насправді прямою, яка з'єднує початок координат і точку (1;1), тобто – діагональ одиничного квадрата.



Рис. 8.3. **Крива Лоренца (крива концентрації)** [43; 49; 70; 84; 85]

У випадку повної нерівності розподілу (коли лише один член суспільства отримує дохід) крива (line of perfect equality) спочатку "прилипає" до осі абсцис, а потім із точки (1;0) стрибкоподібно переходить у точку (1;1). Будь-яка інша крива Лоренца розташується між кривою абсолютної рівності та кривою повної нерівності [1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 97; 99; 100].

Розподіли Лоренца застосовують не лише для моделювання розподілів доходів суспільства, але і майна домогосподарств, часток ринку для окремих підприємств галузі, природних ресурсів окремих країн. Зустріти розподіл Лоренца можна і за межами економічної науки.

8.3. Статистичне оцінювання структурних змін у часі та просторі

Структура будь-якої статистичної сукупності динамічна. Змінюються склад і технічний рівень виробничих фондів, вікова та професійна структура працівників, склад і якість залучених до виробництва природних ресурсів, асортимент і якість виробленої продукції, структура споживчого бюджету та інше. Зміна часток окремих складових сукупності свідчить про структурні зрушення.

Абсолютні та відносні показники зміни окремих частин цілого непропорційні між собою: меншим абсолютним змін можуть відповідати великі відносні зміни, а великим абсолютним змін – менші відносні. Саме тому аналіз змін у структурі будь-якої сукупності слід розраховувати і абсолютні, і відносні показники змін структур для отримання більш точного уявлення про структурні зміни порівнюваних структур [1; 5; 13; 23; 28; 31; 43; 49; 51; 70; 78; 93; 97 – 99; 103].

Переходячи до узагальнювальних показників, звернемо увагу на такий момент. Якщо загальний обсяг досліджуваної сукупності зростає, то відносні показники зміни за окремими елементами сукупності можуть бути більше або менше одиниці, тобто вони можуть зростати та скорочуватися. Причому якщо відносний показник зміни окремого елемента більше відносної зміни в усій сукупності, то це означає, що питома вага цього елемента в сукупності зростає.

Відповідно, якщо відносний показник зміни якого-небудь елемента або частини сукупності менше аналогічного показника всієї сукупності в цілому, то питома вага цієї частини в загальному обсязі скорочується. Таким чином, зміна структури цілого є наслідком нерівномірної інтенсивності зміни окремих частин, тобто відмінностей у відносних змінах питомих ваг.

Для аналізу змін структури досить часто потрібна узагальнена характеристика цих змін. Для цього можуть бути використані такі показники, як: сума абсолютних змін питомих ваг; індекс структурних змін; середньолінійне відхилення, середньоквадратичне відхилення, лінійний коефіцієнт структурних зрушень.

Сума абсолютних змін питомих ваг [14; 70; 78; 84; 85; 87; 93; 97 – 100; 103]:

$$A = \sum_{i=1}^n |d_{i1} - d_{i0}|, \quad (8.20)$$

де d_{i1} – частка показника у структурі у звітному році;
 d_{i0} – частка показника у структурі в базисному році;
 n – кількість елементів (груп) у сукупності.

Сума абсолютних змін питомих ваг виражається у відсоткових пунктах. Ця величина характеризує сумарний обсяг відхилень однієї структури від іншої.

Індекс структурних змін. Для зручності оцінювання використовують інший показник [1; 5; 13; 23; 28; 31; 78; 84; 85; 87; 93; 97 – 100; 103]:

$$I_{\text{стр.змін}} = \frac{1}{2} \sum_{i=1}^n |d_{i1} - d_{i0}|. \quad (8.21)$$

Індекс структурних змін розраховується через питому вагу, виражається у відсотках. Він може приймати значення від 0 до 100 %, наближення до нуля означає відсутність змін, наближення до максимуму – свідчить значні зміни структури.

Приклад 8.3. За даними табл. 8.5, структура спожитого в регіоні палива (в перерахунку на умовне) змінилася: зменшилися частки газу та мазуту, зросла частка вугілля та інших видів палива.

Таблиця 8.5

Розрахункові дані щодо спожитого в регіоні палива

Види палива	2013 р., d_0	2018 р., d_1	Відхилення часток, $d_1 - d_0$	Модулі відхилень, $ d_1 - d_0 $	Квадрати відхилень, $(d_1 - d_0)^2$
Вугілля	29	42	13	13	169
Газ	23	16	- 7	7	49
Мазут	45	36	- 9	9	81
Інші види	3	6	+ 3	3	9
Усього	100	100	0	32	308

Проведемо оцінювання інтенсивності структурних зрушень за допомогою **середньолінійного або середньоквадратичного відхилень** часток [1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 93; 97 – 100; 103]:

$$\bar{l}_d = \frac{\sum_{j=1}^m |d_{j1} - d_{j0}|}{n}; \quad (8.22)$$

$$\sigma_d = \sqrt{\frac{\sum_{j=1}^m (d_{j1} - d_{j0})^2}{n}}, \quad (8.23)$$

де d_{j0} і d_{j1} – частки, відповідно, базисного та поточного періодів;
 n – число складових сукупності.

Лінійний коефіцієнт структурних зрушень становить $\bar{l}_d = \frac{32}{4} = 8$,

тобто частки окремих видів палива змінилися в середньому на 8 відсоткових пунктів. Завдяки своїм математичним властивостям квадратичний коефіцієнт структурних зрушень є дещо більшим [1; 3; 6; 7; 12; 16; 17; 24; 27; 28; 32; 37; 40; 41; 43; 89; 90; 101; 103; 104]:

$$\sigma_d = \sqrt{\frac{308}{4}} = 8,8 \text{ відсоткових пунктів.}$$

З використанням розглянутих показників аналіз зміни структур відбувається без урахування розміру бази, від якої ця зміна відбулася. Більш точну оцінку може дати використання не абсолютних, а відносних змін. Зокрема, можна розрахувати відносний середньолінійний коефіцієнт структурних зрушень, як середню величину з відносних лінійних відхилень (тобто темпів приросту), взятих за модулем [1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 93; 97 – 100; 103]:

$$\bar{l}_{\text{відн}} = \frac{\sum_{j=1}^m \left| \frac{d_{j1} - d_{j0}}{d_{j0}} \right|}{n}. \quad (8.24)$$

Найбільш удосконалені аналітичні властивості (порівняно з лінійним й середньоквадратичним показником) має індекс Салаї та інтегральний коефіцієнт структурних розбіжностей Гатєва.

Індекс Гатєва розраховується як [1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 93; 97 – 100; 103]:

$$K_{\text{Гатєва}} = \sqrt{\frac{\sum_{i=1}^n (d_{i1} - d_{i0})^2}{\sum_{i=1}^n d_{i1}^2 + \sum_{i=1}^n d_{i0}^2}}, \quad (8.25)$$

де d_{i1} – частка показника у структурі у звітному році;

d_{i0} – частка показника у структурі у базисному році.

Індекс приймає більш високе значення, якщо, наприклад, досліджувані партії товарів мають приблизно однаковий розмір. Водночас чим більше кількість партій товарів, а їх розмір менше, тим значення індексу буде вище.

Індекс Салаї має такий вигляд [1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 93; 97 – 100; 103]:

$$I_{\text{Салаї}} = \sqrt{\frac{\sum_{i=1}^n \left(\frac{d_{i1} - d_{i0}}{d_{i0} + d_{i1}} \right)^2}{n}}. \quad (8.26)$$

Індекс Салаї дуже чутливий до некоректного подання малих партій, яке помітно відрізняється від всіх інших. Чим ближче отримане значення до одиниці, тим істотніше відбулися структурні зміни. Коефіцієнт Салаї приймає близькі до одиниці значення, коли в сумі велика кількість одиниць.

Індекс Рябцева. Значення цього показника не залежать від числа градацій структур. Оцінювання проводять на основі максимально можливої величини розбіжностей між компонентами структури. Отже, відбувається співвідношення фактичних розбіжностей окремих компонентів структур із максимально можливими значеннями [1; 5; 13; 14; 23; 28; 31; 43; 49; 51; 70; 78; 84; 85; 87; 93; 97 – 100; 103]:

$$K_{\text{Рябцева}} = \sqrt{\frac{\sum (d_1 - d_0)^2}{\sum (d_1^2 + d_0^2)}}. \quad (8.27)$$

Отже, розглянуті показники складають узагальнену характеристику структурних змін, проте не дають уявлення про величину цих змін.

Важливі поняття

Алгебраїчний центральний момент розподілу – середня арифметична k -го ступеня відхилення індивідуальних значень ознаки від середньої.

Дециль (D_1, D_2, \dots, D_9) – значення ознаки, що ділять сукупність на 10 рівних частин.

Критерії згоди – спеціальні статистичні показники, що використовують для отримання об'єктивної оцінки розбіжності між емпіричною та теоретичною кривими розподілу (засновані на використанні різних відстаней між емпіричним і теоретичним розподілом).

Квантиль – значення ознаки, що займає певне місце в упорядкованій за цією ознакою сукупності.

Коефіцієнт децильної диференціації – показує кратність співвідношення дев'ятого (D_9) і першого (D_1) децилів.

Коефіцієнт диференціації ($K_{\text{диф}}$) – визначається як співвідношення двох середніх, отриманих із 10 % найбільших і найменших значень ознаки, що досліджується.

Оцінювання відмінностей двох розподілів у просторі та часі – порівняння частки розподілу елементів сукупності та частки розподілу значень ознаки.

Перцентель – значення ознаки, що ділять сукупність на 100 рівних частин.

Рівномірність розподілу – випадок, коли частки розподілу елементів сукупності та частки розподілу значень ознаки однакові.

Теоретичні розподіли – залежності між щільністю розподілу та значеннями ознаки, які відбивають закономірності розподілу.

Централізація – зосередженість (скупченість) обсягу ознаки в окремих одиницях (наприклад, капіталу в окремих комерційних банках, продукції якогось виду на окремих підприємствах та ін.).

Типові завдання

Завдання 8.1. Маємо дані про розподіл підлітків за часом, який вони витрачають на комп'ютерні ігри (табл. 8.6). За даними табл. 8.6 визначте перший і третій квантилі та перший і дев'ятий децилі.

Вихідні дані

Групи за часом, що витрачається на комп'ютерні ігри, годин	Кількість підлітків, осіб	Накопичені частоти
1 – 3	43	43
3 – 6	56	99
6 – 9	48	147
9 – 12	32	179
12 – 15	18	197
Усього	197	–

Розв'язання.

Розрахуємо перший кuartиль:

$$Q_1 = 3 + 3 \times \frac{0,25 \times 197 - 43}{56} = 5,97.$$

Розрахуємо третій кuartиль:

$$Q_3 = 9 + 3 \times \frac{0,75 \times 197 - 147}{32} = 9,07.$$

Розрахуємо перший дециль:

$$D_1 = 0 + 3 \times \frac{0,1 \times 197 - 0}{43} = 1,37.$$

Розрахуємо дев'ятий дециль:

$$D_9 = 12 + 3 \times \frac{0,9 \times 197 - 179}{18} = 11,72.$$

Отже, перша чверть (49 підлітків) витрачає на комп'ютерні ігри до 3,33 години. Остання чверть (49 підлітків) витрачає більше 9,07 годин; 10 % підлітків витрачає на комп'ютерні ігри до 1,37 години, 10 % – більше 11,72 годин.

Завдання 8.2. Рівень рентабельності підприємств легкої та харчової промисловості характеризується даними, наведеними в табл. 8.7.

Таблиця 8.7

Вхідні дані

Рівень рентабельності, %	% до підсумку	
	Легка промисловість	Харчова промисловість
До 5	3	8
5 – 10	8	15
10 – 15	16	21
15 – 20	22	26
20 – 25	24	17
25 – 30	18	9
30 і вище	9	4
Усього	100	100

Для кожної галузі визначте квартилі рівня рентабельності, поясніть їх змістовність. Порівняйте варіацію, зробіть висновки.

Розв'язання.

Розрахунок квартилів і децилів ґрунтується на кумулятивних частках: $S_1 = d_1$, $S_2 = d_1 + d_2$, $S_3 = d_1 + d_2 + d_3$ та інші (табл. 8.8).

Таблиця 8.8

Кумулятивні частки рівня рентабельності, %

Рівень рентабельності, %	% до підсумку			
	Легка промисловість	Кумулятивна частка (Sd)	Харчова промисловість	Кумулятивна частка (Sd)
До 5	3	3	8	8
5 – 10	8	11	15	23
10 – 15	16	27	21	44
15 – 20	22	49	26	70
20 – 25	24	73	17	87
25 – 30	18	91	9	96
30 і вище	9	100	4	100
Усього	100	–	100	–

Перший та третій квартилі визначаються за формулами:
перший квартиль:

$$Q_1 = x_0 + h \times \frac{0,25 \times \sum_i^m f_j - S_{Q1-1}}{f_{Q1}};$$

третій квартиль:

$$Q_3 = x_0 + h \times \frac{0,75 \times \sum_i^m f_j - S_{Q3-1}}{f_{Q3}}.$$

а) для легкої промисловості:

перший квартильний інтервал: 10 – 15, оскільки $s_3 = 27 > 100/4$:

$$Q_1 = 10 + 5 \times \frac{0,25 \times 100 - 11}{16} = 14,38\%.$$

другий квартильний інтервал: 20 – 25, оскільки $s_5 = 73 > 100/2$:

$$Q_2 = 20 + 5 \times \frac{0,5 \times 100 - 49}{16} = 20,2\%;$$

третій квартильний інтервал: 25 – 30, оскільки $S_6 = 91 > 3/4 \times 100$:

$$Q_3 = 25 + 5 \times \frac{0,75 \times 100 - 73}{18} = 25,56\%.$$

б) для харчової промисловості:

перший квартильний інтервал: 10 – 15, оскільки $s_3 = 44 > 100/4$:

$$Q_1 = 10 + 5 \times \frac{0,25 \times 100 - 23}{21} = 10,48\%;$$

другий квартильний інтервал: 15 – 20, оскільки $s_4 = 70 > 100/2$:

$$Q_2 = 15 + 5 \times \frac{0,5 \times 100 - 44}{26} = 16,5\%;$$

третій квартильний інтервал: 20 – 25, оскільки $S_5 = 87 > 3/4 \times 100$:

$$Q_3 = 20 + 5 \times \frac{0,75 \times 100 - 70}{17} = 21,47 \%$$

Для легкої промисловості чверть усіх підприємств мають рівень рентабельності до 14,38 %; половина – до 20,2 %; 25 % усіх підприємств із найвищим рівнем рентабельності, найнижчий мають 25,56 %. Для харчової промисловості чверть усіх підприємств має рівень рентабельності до 10,48 %, половина – до 16,5 %; 75 % – до 21,47 %.

Центр розподілу поданий медіаною, тому використаємо квартильний коефіцієнт варіації:

$$V_Q = \frac{Q_3 - Q_1}{2 \times Me}$$

Для легкої промисловості:

$$V_Q = \frac{25,56 - 14,38}{2 \times 20,2} = 0,29, \%$$

Для харчової промисловості:

$$V_Q = \frac{21,47 - 10,48}{2 \times 16,5} = 0,48, \%$$

Більшу варіацію мають підприємства харчової промисловості.

Завдання 8.3. Проведіть аналіз концентрації та локалізації розподілів за даними (табл. 8.9).

Таблиця 8.9

Вихідні дані

Види економічної діяльності	Кількість підприємств за видами діяльності	Інвестиції в основний капітал, млн грн
1	2	3
Сільське господарство, мисливство, лісове господарство	14 736	5 199,3
Добувна промисловість	12 983	8 580,8
Переробна промисловість	98 230	21 811,5

1	2	3
Виробництво та розподілення Зелектроенергії, газу та води	2 654	5 158,7
Будівництво	35 478	2 991,1
Торгівля; ремонт автомобілів, по- бутових виробів та предметів осо- бистого вжитку	115 231	7 737,1
Діяльність транспорту та зв'язку	35 621	14 790,5
Освіта	15 112	665,3
Охорона здоров'я та надання со- ціальної допомоги	23 427	907,2
Інші види діяльності	27 317	19 114,1
Усього	380 789	86 955,6

Розв'язання.

Для розрахунку коефіцієнтів локалізації та концентрації необхідно побудувати розрахункову таблицю (табл. 8.10).

Таблиця 8.10

Розрахункова таблиця

Види економічної діяльності	Кількість підприємств за видами діяльності	Інвестиції в основний капітал, млн грн	d_j	D_j	K_n
Сільське господарство, мис- ливство, лісове господарство	14 736	5 199,3	3,87	5,98	1,55
Добувна промисловість	12 983	8 580,8	3,41	9,87	2,89
Переробна промисловість	98 230	21 811,5	25,80	25,08	0,97
Виробництво та розподілення електроенергії, газу та води	2 654	5 158,7	0,70	5,93	8,51
Будівництво	35 478	2 991,1	9,32	3,44	0,37
Торгівля; ремонт автомобілів, побутових виробів та предме- тів особистого вжитку	115 231	7 737,1	30,26	8,90	0,29
Діяльність транспорту та зв'язку	35 621	14 790,5	9,35	17,01	1,82
Освіта	15 112	665,3	3,97	0,77	0,19
Охорона здоров'я та надання соціальної допомоги	23 427	907,2	6,15	1,04	0,17
Інші види діяльності	27 317	19 114,1	7,17	21,98	3,06
Усього	380 789	86 955,6	100	100	-

Коефіцієнт локалізації розраховується за формулою $K_{л} = \frac{D_j}{d_j}$. Ре-

зультати розрахунків наведено в останній колонці розрахункової таблиці. Максимальна локалізація інвестицій в основний капітал спостерігається в галузі виробництва та розподілення електроенергії, газу та води.

Розрахуємо коефіцієнт концентрації:

$$K_{к} = \frac{1}{2} \sum |D_j - d_j| = \frac{1}{2} \sum |5,98 - 3,87| + |9,87 - 3,41| + |25,08 - 25,8| + |5,93 - 0,7| + |3,44 - 9,32| + |8,9 - 30,26| + |17,01 - 9,53| + |0,77 - 3,97| + |1,04 - 6,15| + |21,8 - 7,17| = 36,27.$$

Отримане значення коефіцієнта концентрації свідчить про високу концентрацію інвестицій за видами економічної діяльності.

Завдання 8.4. Проведіть аналіз подібності структури успішності студентів першої групи до інших (табл. 8.11).

Таблиця 8.11

Вихідні дані

Успішність	Група 1, %	Група 2, %	Група 3, %	Група 4, %
Відмінно	12	16	8	4
Добре	44	54	36	30
Задовільно	40	20	42	48
Незадовільно	4	10	14	18

Розв'язання.

Коефіцієнт подібності структур розраховується за формулою:

$$K_{\text{под}} = 100 - \frac{1}{2} \sum |d_{jk} - d_{js}|.$$

Розрахуємо коефіцієнт подібності структури другої групи до першої:

$$K_{\text{под}21} = 100 - \frac{1}{2} \sum (|16 - 12| + |54 - 44| + |20 - 40| + |10 - 4|) = 80 \text{ \%}.$$

Розрахуємо коефіцієнт подібності структури третьої групи до першої:

$$K_{\text{под}_{21}} = 100 - \frac{1}{2} \sum (|8 - 12| + |36 - 44| + |42 - 40| + |14 - 4|) = 88 \%$$

Розрахуємо коефіцієнт подібності структури четвертої групи до першої:

$$K_{\text{под}_{21}} = 100 - \frac{1}{2} \sum (|4 - 12| + |30 - 44| + |48 - 40| + |18 - 4|) = 78 \%$$

Результати аналізу подібності структур показують, що найбільш подібною до першої групи є успішність студентів третьої групи.

Завдання 8.5. За даними про віковий склад розмитнених легкових і вантажних автомобілів (табл. 8.12) визначте квадратичний коефіцієнт структурних зрушень, проведіть порівняльний аналіз.

Таблиця 8.12

Вихідні дані

Вікова група, років	Структура розмитнених автомобілів, %			
	легкові		вантажні	
	2012 р.	2019 р.	2012 р.	2019 р.
До 10	60	45	56	51
10 – 20	26	34	24	30
20 і старші	14	21	20	19
Разом	100	100	100	100

Розв'язання.

Квадратичний коефіцієнт структурних зрушень: $\sigma_d = \sqrt{\frac{\sum_{j=1}^m (d_{j1} - d_{j0})^2}{n}}$:

1) для металорізального устаткування:

$$\sigma_d = \sqrt{\frac{(45 - 60)^2 + (34 - 24)^2 + (21 - 14)^2}{3}} = \sqrt{\frac{338}{3}} = 10,61;$$

2) для ковальсько-пресового устаткування:

$$\sigma_d = \sqrt{\frac{(51 - 56)^2 + (30 - 24)^2 + (19 - 20)^2}{3}} = \sqrt{\frac{62}{3}} = 4,55.$$

Зрушення у структурі легкових автомобілів більше ніж удвічі перевищують зміни структури вантажних протягом зазначеного періоду. Тобто в структурі розмитнених легкових автомобілів вдвічі переважають більш "молодші" авто.

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Аналіз концентрації, диференціації та подібності розподілів" подані у роботі [67]. Лабораторна робота має за мету опанування студентами прийомами аналізу концентрації, диференціації та подібності статистичних розподілів засобами пакету MS Excel.

Запитання для самодіагностики

1. Які існують порядкові характеристики розподілу?
2. Які показники характеризують ступінь однорідності сукупності, асиметрії й ексцесу розподілу?
3. Якими способами можна оцінити асиметрію розподілу?
4. Як розраховується та що характеризує коефіцієнт асиметрії?
5. Якими способами можна оцінити ексцес розподілу?
6. Який показник є мірою концентрацією і як він обчислюється?
7. Як визначити подібність розподілів?
8. Які показники дозволяють оцінити інтенсивність структурних зрушень?
9. Яким чином визначають теоретичні частоти?
10. Які критерії узгодження ви знаєте?

Питання для критичного переосмислення (есе)

1. Обґрунтуйте переваги та недоліки використання коефіцієнта Джинні в дослідженні диференціації доходів населення.
2. Яка економічна змістовність вкладається в розрахунок показників подібності розподілів?
3. Доведіть або спростуйте твердження, що індекс структурних зрушень дозволяє описати якісні структурні зміни.
4. Яким чином посилення ліво- або правосторонньої асиметрії розподілу впливатиме на розвиток явища або процесу, що досліджується?
5. Поясніть, що відбувається з явищем або процесом у гостровершинному ексцесі розподілу.

Розділ 3

Методи аналізу взаємозв'язків явищ і процесів

9. Статистичні методи вимірювання взаємозв'язків

Основні питання:

- 9.1. Поняття та види зв'язків у статистиці.
- 9.2. Модель аналітичного групування.
- 9.3. Рівняння регресії і визначення його параметрів.
- 9.4. Показники тісноти та істотності кореляційного зв'язку.
- 9.5. Побудова множинних кореляційно-регресійних моделей.
- 9.6. Методи вивчення зв'язку соціальних явищ.
- 9.7. Непараметричні показники зв'язку. Рангові коефіцієнти зв'язку.

9.1. Поняття та види зв'язків у статистиці

Серед багатьох форм зв'язків найважливішим є причинний, тобто такий, що визначає всі інші форми. Сутність причинності полягає у породженні одного явища іншим. Водночас причина як така ще не визнає слідства, вона залежить також від умов, в яких протікає дія причини.

Причинно-наслідкові відносини – це зв'язок явищ і процесів, коли зміна одного з них – причини веде до зміни іншого – слідства. Ознака, що характеризує слідство, є **результативною**, а ознаки, що характеризують причини, – **факторними**.

За статистичною природою зв'язки розділяють на функціональні та стохастичні. У **функціональному зв'язку** кожному значенню факторної ознаки відповідає одне чітко визначене значення результативної ознаки. Функціональні зв'язки звичайно виражаються формулами. Найчастіше такі зв'язки спостерігаються в математиці, фізиці, хімії. Мають місце функціональні зв'язки і в економічних процесах, але доволі рідко, оскільки відображають взаємозв'язки тільки окремих аспектів складних явищ суспільного життя. Наприклад, продуктивність праці – це співвідношення обсягу продукції та чисельності працівників.

Функціональний зв'язок можна подати у вигляді рівняння:

$$y_i = f(x_i), \quad (9.1)$$

де y_i – результативна ознака ($i = 1, \dots, n$);

$f(x_i)$ – відома функція зв'язку результативної та факторної ознак;

x_i – факторна ознака.

Функціональна залежність з однаковою силою проявляється у всіх одиницях сукупності незалежно від зміни інших ознак даного явища. Але в масових явищах суспільного життя через різноманітність факторів виникає широка варіація результативної ознаки. Це свідчить про те, що зв'язок між ознаками-факторами та результативною ознакою не повний, а виявляється лише у загальному, середньому. Такі зв'язки називають стохастичними.

У **стохастичному зв'язку** кожному значенню ознаки x відповідає певна множина значень ознаки y , які варіюють і утворюють ряд розподілу (умовний). Важливою особливістю стохастичних зв'язків є те, що вони спостерігаються не в одиничних випадках, а в масі. Вони потребують для свого дослідження масових спостережень, тобто статистичних даних.

Модель стохастичного зв'язку може бути зображена в загальному вигляді рівнянням:

$$\tilde{y}_i = f(x_i) + \varepsilon_i, \quad (9.2)$$

де \tilde{y}_i – розрахункове значення результативної ознаки;

$f(x_i)$ – частина результативної ознаки, що сформувалася під впливом відомих і врахованих факторних ознак (однієї або кількох), що перебувають у стохастичному зв'язку з результативною ознакою;

ε_i – частина результативної ознаки, що виникла внаслідок дії неврахованих факторів.

Підвидом стохастичного зв'язку є **кореляційний зв'язок**, коли зі зміною факторної ознаки x змінюються групові середні результативної ознаки y , тобто замість умовних розподілів порівнюються середні значення цих розподілів.

Прикладом стохастичного, зокрема кореляційного, зв'язку є розподіл ділянок виробництва за втратами від браку продукції (y) та відсотком порушень технологічної дисципліни (x) (табл. 9.1).

Таблиця 9.1

Розподіл виробничих ділянок за ознаками

Відсоток порушення технологічної дисципліни, x	Кількість ділянок з втратами від браку продукції, тис. грн						Середні втрати від браку, тис. грн, \bar{y}_i
	0,9 – 1,14	1,14 – 1,38	1,38 – 1,62	1,62 – 1,86	1,86 – 2,1	Разом, f_i	
1,2 – 1,575	2	3	1	–	–	6	1,22
1,575 – 1,95	–	–	6	–	–	6	1,5
1,95 – 2,325	–	–	2	3	–	5	1,644
2,325 – 2,7	–	–	–	–	5	5	1,98
У цілому	2	3	9	3	5	22	1,565

Наведені в табл. 9.1 дані складають кореляційну таблицю, в якій подано групування за двома взаємозв'язаними ознаками: факторною та результативною. Концентрація частот навколо діагоналі матриці свідчить про наявність кореляційного зв'язку між ознаками. Як видно з даних табл. 9.1, розподіл ділянок виробництва спостерігається із верхнього лівого кута матриці до нижнього правого. Характер концентрації частот на діагоналі свідчить про наявність прямого кореляційного зв'язку між розглядуваними ознаками.

Кожній групі за факторною ознакою відповідає свій розподіл y, який відрізняється від інших груп і від безумовного підсумкового розподілу. Отже, спостерігається стохастичний зв'язок між ознаками.

Умовні розподіли можна замінити середніми значеннями результативної ознаки, які обчислюються як середня арифметична зважена:

$$\bar{y}_i = \frac{\sum_{i=1}^m y_i f_i}{\sum_{i=1}^m f_i} \tag{9.3}$$

Поступова зміна середніх \bar{y}_i від однієї групи до іншої свідчить про наявність кореляційного зв'язку між ознаками.

Кореляційний зв'язок – поняття більш вузьке, ніж стохастичний зв'язок. Останній може відобразитись не тільки в зміні середньої величини, а й у варіації однієї ознаки залежно від іншої, тобто будь-якої характеристики варіації. Отже, кореляційний зв'язок є частковим випадком стохастичного зв'язку.

Кореляційний зв'язок між ознаками може виникнути внаслідок:

причинної залежності результативної ознаки (її варіації) від варіації факторної ознаки;

спряженості, що виникає за наявності загальної причини. Наприклад, відомий статистик О. О. Чупров [61] проаналізував, що якщо в ролі ознаки x взяти число пожежних команд у місті, а в якості ознаки y – суму збитків від пожеж за рік, то між цими ознаками існує пряма кореляція: в середньому чим більше пожежників у місті, тим більше і збитків від пожеж. Але ж цей факт не природний. Таку кореляцію не можна інтерпретувати як зв'язок причини та слідства. У цьому випадку обидві ознаки – слідства загальної причини – розміру міста. Зрозуміло, що у великих містах більше пожежних частин, але більше пожеж, а отже, і збитків від них на рік, ніж у малих містах;

взаємозв'язку ознак, кожна із яких і причина, і наслідок. Наприклад, кореляція між рівнями продуктивності праці та рівнем оплати 1 години праці: з одного боку, чим вище продуктивність праці, тим вище оплата, а з іншого – оплата праці відіграє стимуляційну роль для підвищення продуктивності праці. Отже, в такій системі ознак кожна виступає в якості незалежної змінної x і залежної змінної y .

Залежно від напрямку дії функціональні та стохастичні зв'язки можуть бути прямими та зворотними. У *прямому зв'язку* напрямок зміни результативної ознаки співпадає з напрямком зміни ознаки – фактора, тобто з підвищенням факторної ознаки підвищується і результативна, та навпаки. У зворотному випадку – між величинами, що розглядаються існують *зворотні зв'язки*. Наприклад, чим вище стаж роботи, тим вище заробітна плата – прямий зв'язок. А чим вище собівартість продукції, тим нижче прибуток – зворотній зв'язок.

За аналітичним вираженням зв'язки бувають прямолінійні та криволінійні. Якщо статистичний зв'язок між явищами може бути виражено

рівнянням прямої, то його називають *лінійним зв'язком*. Якщо ж він виражається рівнянням будь-якої кривої лінії (параболи, гіперболи, степеневої, експоненціальної, показової і т. д.), то такий зв'язок буде **нелінійним** або **криволінійним**.

Слід зазначити, що тільки функціональні зв'язки точно виражаються аналітичним рівнянням, а кореляційні – лише приблизно, за умови абстрагування від впливу всіх інших факторів. Тому на графічному зображенні кореляційного зв'язку спостерігається розкидання крапок (y і x) навколо теоретичної лінії.

Залежно від кількості досліджуваних ознак розрізняють парну (просту) та множинну кореляцію. У *парній кореляції* вивчають зв'язок між двома ознаками (результативною і факторною), у *множинній кореляції* – зв'язок між трьома та більшим числом ознак (результативною і двома та більшим числом факторних ознак). У випадку багатофакторного зв'язку всі фактори діють комплексно, тобто одночасно та у взаємозв'язку.

9.2. Модель аналітичного групування

Будуючи аналітичне групування, необхідно утворити таку кількість груп, за якою у варіації групових середніх у максимальному ступені проявиться вплив групової ознаки. Загальне правило твердить, що чим більше утворюється груп, тим більше підвищується міжгрупова варіація. Проте не доцільна велика кількість груп, особливо за невеликої кількості одиниць сукупності: у цьому випадку групи будуть малочислені, середні із них будуть мати випадковий характер, міжгрупова варіація відобразить не тільки вплив розглядуваного, але й інших факторів. Це означає, що необхідно вибирати оптимальне число груп для кожного конкретного випадку, коли групові середні будуть без випадкового характеру та найповнішим чином себе виявлять груповою ознакою.

Характеристикою кореляційного зв'язку є *лінія регресії*, яка розглядається у двох моделях: аналітичного групування та регресійного аналізу.

У *моделі аналітичного групування* – це емпірична лінія регресії, що утворюється з групових середніх значень результативної ознаки \bar{y}_i для кожного значення (інтервалу) x_i . Інакше кажучи, якщо на кореляційному полі точки з'єднати відрізками прямої лінії, то отримаємо ламану лінію з деякою тенденцією, яку називають *емпіричною лінією регресії*.

Метод аналітичного групування полягає в тому, що всі елементи сукупності групують за факторною ознакою x і в кожній групі обчислюють середні значення результативної ознаки y , тобто лінія регресії оцінюється лише в окремих точках, які відповідають певному значенню x .

Аналітичне групування дає змогу встановити кількісні співвідношення між розглядуваними ознаками. Ефекти впливу x на y визначаються як відношення приростів середніх групових значень $\Delta y : \Delta x$, де $\Delta y = \bar{y}_i - \bar{y}_{i-1}$, $\Delta x = x'_i - x'_{i-1}$. За даними табл. 9.1 прирости Δx у всіх групах однакові – 0,375 %, а середні втрати від браку продукції збільшуються за групами таким чином: $\Delta y_2 = 1,5 - 1,22 = 0,28$ тис. грн; $\Delta y_3 = 0,144$ тис. грн; $\Delta y_4 = 0,336$ тис. грн. Отже, зі збільшенням відсотка порушень технологічної дисципліни на 1 %, втрати від браку продукції у середньому зростають, відповідно, на $\Delta y_2 : \Delta x_2 = 0,28 : 0,375 = 0,74$ тис. грн і на 0,384 і 0,896 тис. грн.

Наступним етапом аналітичного групування є *вимірювання тісноти зв'язку*, яке ґрунтується на правилі складання дисперсій. У моделі аналітичного групування мірою щільності зв'язку є відношення міжгрупової дисперсії до загальної, тобто емпіричний коефіцієнт детермінації:

$$\eta^2 = \frac{\sigma_M^2}{\sigma^2}, \quad (9.4)$$

де σ^2 – загальна дисперсія, яка характеризує варіацію числових значень результативної ознаки, пов'язану з варіацією всіх факторів, що на неї впливають;

σ_M^2 – міжгрупова дисперсія, яка характеризує варіацію числових значень результативної ознаки, яка пов'язана з варіацією групової ознаки.

Коефіцієнт детермінації змінюється від 0 до 1. Якщо $\eta^2 = 0$, міжгрупова дисперсія дорівнює нулю. Це можливо лише за умови, що всі групові середні однакові, а кореляційний зв'язок між ознаками відсутній.

За $\eta^2 = 1$ міжгрупова дисперсія дорівнює загальній, середня з групових – нулю. У цьому випадку кожному значенню факторної ознаки відповідає єдине значення результативної ознаки, тобто зв'язок між ознаками функціональний.

Слід зауважити, що значення $n^2 > 0$ не завжди є доказом наявності кореляційного зв'язку між ознаками. Відмінний від нуля коефіцієнт детермінації може з'явитись і за випадкового розподілу сукупності на групи.

За даними табл. 9.2 загальна дисперсія втрат від браку продукції становить:

$$\sigma^2 = \bar{y}^2 - (\bar{y})^2 = \frac{\sum (y_i')^2 f_i}{\sum f_i} - \left(\frac{\sum y_i' f_i}{\sum f_i} \right)^2 =$$

$$= (1,02^2 \times 2 + 1,26^2 \times 3 + 1,5^2 \times 9 + 1,74^2 \times 3 + 1,98^2 \times 5) / 22 - 1,565^2 = 0,086.$$

У табл. 9.2 наведено аналітичне групування виробничих ділянок, яке описує залежність втрат від браку продукції на цих ділянках від відсотка порушень технологічної дисципліни. Там же подано розрахунок міжгрупової дисперсії.

Таблиця 9.2

Аналітичне групування

Відсоток порушень технологічної дисципліни, %	Кількість виробничих ділянок, f_i	Середні втрати від браку продукції, тис. грн, \bar{y}_i	$\bar{y}_i - \bar{y}$	$(\bar{y}_i - \bar{y})^2 f_i$
1,2 – 1,575	6	1,220	-0,345	0,714
1,575 – 1,95	6	1,500	-0,065	0,025
1,95 – 2,325	5	1,644	0,079	0,031
2,325 – 2,7	5	1,98	0,415	0,861
У цілому	22	1,565	–	1,631

Міжгрупова дисперсія дорівнює:

$$\sigma_M^2 = \frac{\sum (\bar{y}_i - \bar{y})^2 f_i}{\sum f_i} = \frac{1,631}{22} = 0,074.$$

Коефіцієнт детермінації становить:

$$\eta^2 = \frac{0,074}{0,086} = 0,86,$$

тобто варіація втрат від браку продукції на 86 % пояснюється варіацією відсотка порушень технологічної дисципліни і на 14 % – варіацією інших факторів. Отже, зв'язок між ознаками дуже сильний.

Сильний зв'язок між ознаками може виникнути випадково, тому необхідно перевірити його істотність, тобто довести не випадковість зв'язку.

Перевірка істотності зв'язку здійснюється за допомогою критеріїв математичної статистики. Вона ґрунтується на порівнянні фактичного значення η^2 з критичним $\eta_{1-2}^2(k_1, k_2)$ для певного рівня істотності α та числа ступенів свободи $k_1 = m - 1$ і $k_2 = n - m$, де m – число груп; n – обсяг сукупності.

Критичне значення η^2 є тим максимально можливим значенням коефіцієнта детермінації, яке може виникнути випадково за відсутності кореляційного зв'язку.

Якщо $\eta^2 > \eta_{1-\alpha}^2(k_1, k_2)$, то зв'язок між результативною і факторною ознаками вважається істотним. Якщо фактичне значення η^2 менше критичного, то наявність кореляційного зв'язку між ознаками не доведена, і зв'язок вважається неістотним.

Критичне значення вибирають таким чином, щоб імовірність отримання значення η^2 більшого від критичного (за умови відсутності зв'язку між ознаками) була достатньо мала. Таку ймовірність називають *рівнем істотності* α . Найчастіше в економіко-статистичних дослідженнях застосовують такі рівні істотності, як $\alpha = 0,05$ і $\alpha = 0,01$. Критичні значення кореляційного відношення для $\alpha = 0,05$ наведені в табл. А.1 в додатку А.

У нашому прикладі $k_1 = 4 - 1 = 3$, $k_2 = 22 - 4 = 18$. Критичне значення за цих ступенів свободи для рівня значущості $\alpha = 0,05$ становить $\eta_{0,95}^2(3,18) = 0,345$.

Оскільки $\eta^2 = 0,86 > 0,345$, то зв'язок визнається істотним з імовірністю 95 %.

Для перевірки істотності зв'язку використовують також функціонально зв'язану з η^2 характеристику *F-критерій* (критерій Фішера), який обчислюється за формулою:

$$F = \frac{\eta^2}{1 - \eta^2} \times \frac{k_2}{k_1}. \quad (9.5)$$

Критичні значення *F-критерію* наведені в табл. Б.1 додатка Б і табл. В.1 в додатку В.

У нашому прикладі фактичне значення *F* становить:

$$F = \frac{0,86}{1 - 0,86} \times \frac{18}{3} = 36,85,$$

що значно більше від критичного $F_{0,95}(3,18) = 3,16$. Це свідчить про істотність кореляційного зв'язку між аналізованими ознаками.

Комбінаційні аналітичні групування використовують для аналізу зв'язку результативної ознаки з двома та більше факторними ознаками. Вони дають можливість аналізувати залежність результативної ознаки від кожного з факторів за умови фіксованих значень інших. Методи вимірювання такого типу зв'язку та перевірки його істотності називають *багатофакторними дисперсійними комплексами*.

Недоліком методу аналітичних групувань є те, що він не дозволяє визначити форму (аналітичне вираження) впливу факторних ознак на результативні.

У загальному вигляді задача статистики в сфері вивчення взаємозв'язків полягає не тільки в кількісному оцінюванні їх наявності, напрямку та сили зв'язку, а й у визначенні форми (аналітичного вираження) впливу факторних ознак на результативні. Для її вирішення застосовують методи кореляційно-регресійного аналізу.

Кореляційно-регресійний аналіз як загальне поняття включає вимірювання тісноти, напрямку зв'язку та встановлення аналітичного вираження (форми) зв'язку (регресійний аналіз).

Метою кореляційного аналізу є кількісне визначення тісноти зв'язку між двома ознаками (за парного зв'язку) та між результативною і множиною факторних ознак (за багатофакторного зв'язку). Тіснота зв'язку кількісно виражається величиною коефіцієнтів кореляції. Такі коефіцієнти

дають можливість визначити "вигідність" факторних ознак у побудові рівнянь множинної регресії. Величина коефіцієнта кореляції слугує також оцінкою відповідності рівняння регресії виявленим причинно-наслідковим зв'язкам.

Метою регресійного аналізу є визначення аналітичного вираження зв'язку, встановлення ступеня впливу незалежних змінних на залежну та визначення розрахункових значень залежної змінної (функції регресії). За допомогою рис. 9.1 пояснимо різницю між кореляцією та регресією.

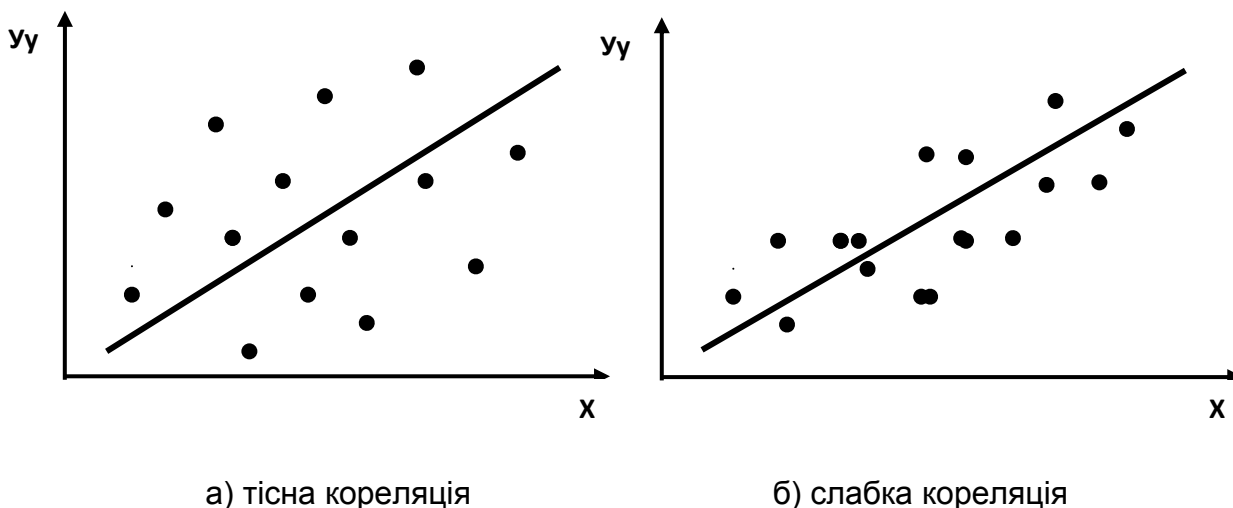


Рис. 9.1. Регресія за різної інтенсивності кореляції

Як видно із рис. 9.1, кут нахилу лінії регресії відносно осі абсцис однаковий для частини а та б. Але на частині а точки кореляційного поля концентруються навколо лінії регресії, а на б точки кореляційного поля розкидані. Це означає, що тіснота зв'язку, тобто міра кореляції між x та y , у випадку а буде високою, а у випадку б – низькою. Отже, рівняння регресії у випадку а буде статистично значущим, а у випадку б – статистично незначущим. Отже, випадки а та б розрізняють за величиною коефіцієнтів кореляції ($r_{yxa} \neq r_{yxb}$), водночас вони мають однакові коефіцієнти регресії ($b_{yxa} = b_{yxb}$).

У статистиці розрізняють такі **варіанти залежностей**:

парна кореляція (регресія) – зв'язок між двома ознаками (результативною та факторною або двома факторними);

часткова кореляція (регресія) – залежність між результативною та однією із факторних ознак за умови фіксованого значення інших факторних ознак;

множинна кореляція (регресія) – залежність результативної та двох або більше факторних ознак, що включені у дослідження.

Щоб результати кореляційно-регресійного аналізу знайшли практичне застосування і дали науково обґрунтовані результати, повинні виконуватись певні вимоги відносно об'єкта дослідження і якості вихідної статистичної інформації. Основні із цих **вимог** такі:

достатнє число спостережень, оскільки зв'язки між ознаками виявляються тільки внаслідок дії закону великих чисел. Кількість одиниць спостереження повинна в 6 – 10 разів перевищувати кількість включених у модель факторів;

якісна однорідність досліджуваної сукупності, що передбачає близькість формування результативних і факторних ознак. Необхідність виконання цієї умови впливає зі змістовності параметрів рівняння зв'язку, які є середніми величинами. В якісно однорідній сукупності вони будуть типовими характеристиками, в якісно різномірній – спотвореними, що перевертають характер зв'язку. Кількісна однорідність сукупності полягає у відсутності одиниць спостереження, які за своїми числовими характеристиками суттєво відрізняються від основної маси даних. Такі одиниці спостереження слід виключати із сукупності та вивчати окремо;

випадковість і незалежність окремих одиниць сукупності одна від одної, яка означає, що значення ознак одних одиниць сукупності не повинні залежати від значень інших одиниць цієї самої сукупності;

стійкість і незалежність дії окремих факторів;

сталість дисперсії результативної ознаки зі зміною факторних ознак;

нормальний розподіл ознак.

Якщо обсяг досліджуваної сукупності достатньо великий ($n > 50$), то нормальність розподілу може бути підтверджена на основі розрахунку й аналізу критеріїв Пірсона, Ястремського, Боярського, Колмогорова та ін. Якщо $n < 50$, то закон розподілу вихідних даних визначається на основі побудови та візуального аналізу поля кореляції. Водночас якщо в розташуванні крапок має місце лінійна тенденція, то можна передбачити, що сукупність вихідних даних $(y, x_1, x_2, \dots, x_n)$ підлягає нормальному розподілу.

Використовуючи методи кореляційно-регресійного аналізу, слід пам'ятати, що рівняння кореляційного зв'язку вимірює залежність між варіаціями результативної та факторної ознак. Міри тісноти зв'язку показують частку варіації результативної ознаки, яка пов'язана з варіацією факторної ознаки (ознак). Інтерпретувати кореляційні показники необхідно в термінах варіації відхилень від середньої величини. Якщо завдання дослідження

полягає у вимірюванні зв'язку не між варіацією двох ознак у сукупності, а між змінами ознак об'єкта в часі, то метод кореляційно-регресійного аналізу потребує змін. Це означає, що не можна трактувати кореляцію ознак як причинний зв'язок їх рівней. Інакше кажучи, метод кореляційно-регресійного аналізу не може пояснити ролі факторних ознак у створенні результативної ознаки. Це є обмеженням цього методу.

У побудові кореляційної моделі важливе значення має відбір факторів, у ході здійснення якого слід дотримуватися таких *рекомендацій*:

1) у рівняння регресії необхідно включати лише ті фактори, які є істотними для вивчення проблеми. Тому математичному моделюванню повинен передувати теоретичний аналіз. Невідповідність якісного та кількісного аналізів призводить до помилкової кореляції першого виду. Якщо порушується передбачення щодо необхідних властивостей розподілу вихідних даних, то має місце помилкова кореляція другого виду. З виявленням помилкової кореляції необхідно розширити коло питань, що розглядаються;

2) фактори, що використовуються в регресійному аналізі, повинні бути вимірні кількісно, тобто необхідно використовувати кількісні, а не якісні ознаки. Кореляційно-регресійний аналіз – це метод дослідження кількісного аспекту зв'язків, тому кожна ознака повинна мати кількісну (метричну) розмірність. Якщо в певній задачі не можна не враховувати атрибутивні фактори, то сукупність спочатку групують за атрибутивними ознаками. Таке групування здійснюється на основі експертних оцінок або методу таксономії. Різницю між виділеними групами перевіряють методом дисперсійного аналізу. Якщо в межах кожної із груп необхідно дослідити ще й кількісні закономірності, то здійснюється оброблення даних методом регресійного аналізу щодо кожної групи окремо або застосовується коваріаційний аналіз.

Штучне приведення атрибутивних ознак до кількісного вигляду шляхом ранжування, надання різних кодів та умовних чисел не приводить до задовільних результатів;

3) кожний фактор у рівнянні множинної регресії повинен бути поданий лише однією ознакою (натуральною або вартісною, абсолютною або відносною, але не обома одночасно). Недотримання цієї умови призводить до моделей, що не мають економічної інтерпретації.

Відбір ознаки, яка найкраще відображає кількісну величину реального фактора здійснюється на основі розглядання парних зв'язків результативної та всіх можливих факторних ознак. Цей аналіз здійснюється за кореляційною матрицею парних зв'язків.

Не можна у рівнянні регресії одночасно використовувати сумарні (агреговані) ознаки та їх складові.

Якщо зв'язок достатньо тісний та стійкий, то можна використовувати аналітичні фактори, підвищуючи конкретність моделі. Якщо ж зв'язок недостатньо тісний та стійкий, перевага віддається сумарному (синтетичному) показнику. З переходом від аналітичних показників до синтетичних тіснота та стійкість зв'язків підвищується за рахунок втрати конкретності (це один із прояв закону великих чисел).

Якщо в модель необхідно включити синтетичний показник і його частину (аналітичний показник), то величину синтетичного показника зменшують на величину аналітичного, що включений у модель. Але такий прийом може порушити мультиколінеарність факторів, що є допустимою;

4) у рівняння регресії рекомендується включати тільки такі фактори, які в ланцюгу "причина – наслідок" перебувають на одному рівні. Це одна з найважливіших вимог багатofакторного моделювання. Якщо вона не виконується, то коефіцієнти регресії не мають економічної змістовності;

5) якщо в моделі в якості змінних використати відносні величини, необхідно, щоб знаменником виступав той самий показник або зіставні показники (однаковий базис). З порушенням цієї вимоги параметри моделі трудніше піддаються економічній інтерпретації;

6) фактори, що включаються в модель, повинні бути декорельовані, тобто перебувати в слабкому зв'язку між собою. Вважають, що два фактори колінеарні, якщо парний коефіцієнт кореляції за абсолютною величиною більше 0,8. Ця умова застосовується, коли сукупність має середній обсяг, а кореляція окремих факторів та результативної ознаки є тісною. У загальному випадку межа статистичної колінеарності залежить не тільки від тісноти взаємозв'язку факторів, але і від зв'язку результативної ознаки з факторами. У практичних цілях можна вважати, що фактори не мають статистичного колінеарного зв'язку, якщо парні коефіцієнти кореляції відповідають таким вимогам: $r_{oi} > r_{ij}$; $r_{oj} > r_{ij}$. Тут 0 – це номер результативної ознаки; i, j – номери факторних ознак.

Якщо ці нерівності (або одна з них) не виконуються або близькі до рівності, то для багатofакторного аналізу рекомендується використовувати лише один із двох розглядуваних факторів (з вищим парним коефіцієнтом кореляції з результативною ознакою). Віддавати перевагу краще

використанню синтетичного показника, який відображає комплекс обох (або декількох) мультиколінеарно пов'язаних факторів;

7) слід дотримуватись логічних вимог, що враховуються у відборі показників, для регресійного аналізу (наприклад, таких, що відповідають вимогам до аналітичних групувань за факторною, а не результативною ознакою.

Необхідно додержуватись необерненості змінних, а в якості незалежної змінної не використовувати ознаку, яка є логічним наслідком ознаки, використаної в ролі залежної змінної;

8) використовувати мінімальне, але достатнє число факторів, що визначають середнє значення результативної ознаки. Підвищення числа факторів з метою отримання більш змістовної та конкретної моделі призводить до зниження її стійкості та ускладнення економічної інтерпретації. Тому включати в модель другорядні фактори не рекомендується. Ця вимога будується на принципі простоти. У регресійному аналізі цей принцип реалізується через алгоритм багатокрокового виключення змінних; він ураховується також і для вибору форми зв'язку;

9) фактори, включені в модель, повинні бути доступні та вірогідні. Не треба вводити фактори, якщо дані про них не можна отримати із звітності, а здобути лише шляхом спеціально організованого спостереження, яке є досить трудомістким і затратним.

Задачі, які розв'язуються за допомогою кореляційно-регресійного аналізу:

1) виділення найважливіших факторів, що впливають на результативну ознаку (тобто на варіацію її значень у сукупності). Ця задача розв'язується на базі мір тісноти зв'язку ознак-факторів із результативною ознакою;

2) оцінювання господарської діяльності за ефективністю використання факторів виробництва. Ця задача вирішується шляхом розрахунку для кожної одиниці сукупності тих величин результативної ознаки, які були б отримані від середньої у сукупності ефективності використання факторів у порівнянні їх з фактичними результатами виробництва;

3) прогнозування можливих значень результативної ознаки за заданими значеннями факторних ознак. Ця задача вирішується шляхом підстановки очікуваних значень факторних ознак у рівняння зв'язку та розрахунку очікуваних значень результативної ознаки.

Для розв'язання зворотної задачі – розрахунку необхідних значень факторних ознак для забезпечення необхідного значення результативної

ознаки – застосовується методика розв'язання оптимізаційної задачі зі знаходження найкращого із можливих варіантів;

4) підготовка даних, необхідних в якості вихідних для розв'язання оптимізаційних задач. Показники, які включають у розв'язання оптимізаційної задачі, можуть бути отримані на основі кореляційно-регресійної моделі.

Використання розглянутих завдань допоможуть сформулювати алгоритм проведення кореляційно-регресійного аналізу в ході дослідження різних явищ і процесів.

9.3. Рівняння регресії і визначення його параметрів

У моделі регресійного аналізу характеристикою кореляційного зв'язку є **теоретична лінія регресії**, що описується функцією $y = f(x)$, яку називають **рівнянням регресії**. На відміну від емпіричної, теоретична лінія регресії неперервна.

Різні явища по-різному реагують на зміну факторів. Для того щоб відобразити характерні особливості зв'язку конкретних явищ, статистика використовує різні за функціональним видом регресійні рівняння. Якщо зі зміною фактора x результат y змінюється більш-менш рівномірно, такий зв'язок описують лінійною функцією $Y = a + bx$. За нерівномірного співвідношення варіацій взаємопов'язаних ознак (наприклад, коли прирости значень y зі зміною x прискорені чи сповільнені або напрям зв'язку змінюються) використовують нелінійні регресії, зокрема:

$$\text{степеневу: } Y = ax^b; \quad (9.6)$$

$$\text{гіперболу: } Y = a + \frac{b}{x}; \quad (9.7)$$

$$\text{параболу: } Y = a + bx + cx^2. \quad (9.8)$$

Вибір та обґрунтування функціонального виду регресії базуються на теоретичному аналізі сутності зв'язку. Наприклад, зв'язок між урожайністю та кількістю опадів подають у вигляді параболи, а зв'язок між собівартістю y й обсягом продукції x – рівнянням гіперболи, де a – пропорційні витрати на одиницю продукції, b – постійні витрати на весь випуск.

Якщо кривизна регресії невелика, то в межах фактичної варіації ознак зв'язок між ними досить точно описується лінійною функцією. Цим

у значній мірі пояснюється широке використання лінійних рівнянь регресії: $Y = a + bx$.

Параметр b (**коефіцієнт регресії**) – величина іменована, має розмірність результативної ознаки та розглядається як ефект впливу x на y . Він показує середнє значення зміни результативної ознаки y зі зміною факторної ознаки x на одну одиницю її вимірювання, тобто варіацію y , що припадає на одиницю варіації x . Знак b показує напрямок цієї зміни.

Приклад 9.1. Маємо рівняння регресії $Y = 4,5 + 0,8x$, що характеризує залежність продуктивності праці Y від стажу роботи x . Оскільки параметр $b = 0,8 > 0$, то з підвищенням стажу продуктивність також підвищується. Підвищення на 1 рік стажу роботи приводить до підвищення продуктивності праці в середньому на 0,6 одиниць.

Параметр a – вільний член рівняння регресії, це значення y за $x = 0$. Якщо межі варіації x не містять нуля, то цей параметр має лише розрахункове значення. Характеристикою відносної зміни Y за рахунок x є коефіцієнт еластичності:

$$K_{\text{ел}} = b \frac{\bar{x}}{\bar{y}}, \quad (9.9)$$

який показує, на скільки відсотків y в середньому змінюється результативна ознака зі зміною факторної на 1 %.

Математично доведено, що значення параметрів a та b за умови лінійної залежності визначають із системи нормальних рівнянь:

$$\begin{aligned} \Sigma y &= na + b\Sigma x, \\ \Sigma xy &= a\Sigma x + b\Sigma x^2. \end{aligned} \quad (9.10)$$

Розв'язавши цю систему, знаходять такі значення параметрів:

$$\begin{aligned} b &= \frac{n \Sigma xy - \Sigma x \Sigma y}{n \Sigma x^2 - \Sigma x \Sigma x}, \\ a &= \bar{y} - b\bar{x}, \end{aligned} \quad (9.11)$$

де y – рівень фактичного ряду (результативна ознака);

n – число членів ряду;

x – факторна ознака.

Рівняння регресії відбиває закон зв'язку між x і y не для окремих елементів сукупності, а для сукупності в цілому; закон, який абстрагує вплив інших факторів, виходить із принципу "за інших однакових умов".

Вплив інших факторів, окрім x , зумовлює відхилення емпіричних значень y від теоретичних \tilde{y} в той чи інший бік. Відхилення $(y - \tilde{y})$ називають *залишками* та позначають символом " ε ".

Загальна дисперсія розраховується за формулою:

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y - \bar{y})^2, \quad (9.12)$$

а залишкова:

$$\sigma_\varepsilon^2 = \frac{1}{n} \sum_{i=1}^n (y - \tilde{y})^2. \quad (9.13)$$

У невеликих за обсягом сукупностях коефіцієнт регресії схильний до випадкових коливань. Тому слід перевірити його істотність. У лінійному зв'язку істотність коефіцієнта регресії перевіряють за допомогою *t-критерію Стьюдента*:

$$t = \frac{|b|}{\mu_b}, \quad (9.14)$$

де b – коефіцієнт регресії;

μ_b – стандартна похибка.

Стандартна похибка коефіцієнта регресії залежить від варіації факторної ознаки σ_x^2 , залишкової дисперсії σ_ε^2 і числа ступенів свободи $k = n - 2$:

$$\mu_b = \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_x^2(n-2)}}, \quad (9.15)$$

де $\sigma_x^2 = \overline{x^2} - (\bar{x})^2$, $\overline{x^2} = \frac{\sum x^2 \times f}{\sum f}$, $(\bar{x})^2 = \frac{\sum x \times f^2}{\sum f}$.

Якщо $t > t_{кр(1-\alpha)}(k)$, то вплив ознаки x на ознаку y визнається істотним.

Для коефіцієнта регресії визначають також довірчі межі $b \pm t\mu_b$.

Приклад 9.2. Щодо десяти комерційних банків є дані про суму активів і депозитів (табл. 9.3). Необхідно провести кореляційно-регресійний аналіз зв'язку між цими ознаками. Для характеристики зазначеного зв'язку слід визначити: форму та математичну модель зв'язку; параметри рівняння регресії; тісноту зв'язку.

Таблиця 9.3

Розрахункова таблиця для визначення параметрів рівняння регресії

Номер банку	Назва банку	Сума депозитів, млрд грош. од., x	Сума активів, млрд грош. од., y	Розрахункові дані				
				x^2	y^2	$x \times y$	\bar{y}	$(y_i - \bar{y})^2$
1	А	32	81	1024	6561	2592	80,614	0,1489
2	Б	25	75	625	5625	1875	76,043	1,0878
3	В	28	79	784	6241	2212	78,002	0,9960
4	Г	18	73	324	5329	1314	71,472	2,3347
5	Д	30	80	900	6400	2400	79,308	0,4788
6	Е	24	75	576	5625	1800	75,390	0,1521
7	Ж	29	83	841	6889	2407	78,655	18,8790
8	З	17	69	289	4761	1173	70,819	3,3087
9	К	24	73	576	5329	1752	75,390	5,7121
10	Л	18	70	324	4900	1260	71,472	2,1667
Разом		246	758	6263	57660	18785	758,00	35,2648
У середньому		24,6	75,8	626,3	5766	1878,5	75,800	3,5264

Щоб визначити форму зв'язку між сумою активів (y) та сумою депозитів (x), побудуємо графік – кореляційне поле (рис. 9.2).

Як видно з рис. 9.2, залежність між результативною і факторною ознаками прямолінійна і її можна виразити рівнянням прямої:

$$\hat{y} = a_0 + a_1x.$$

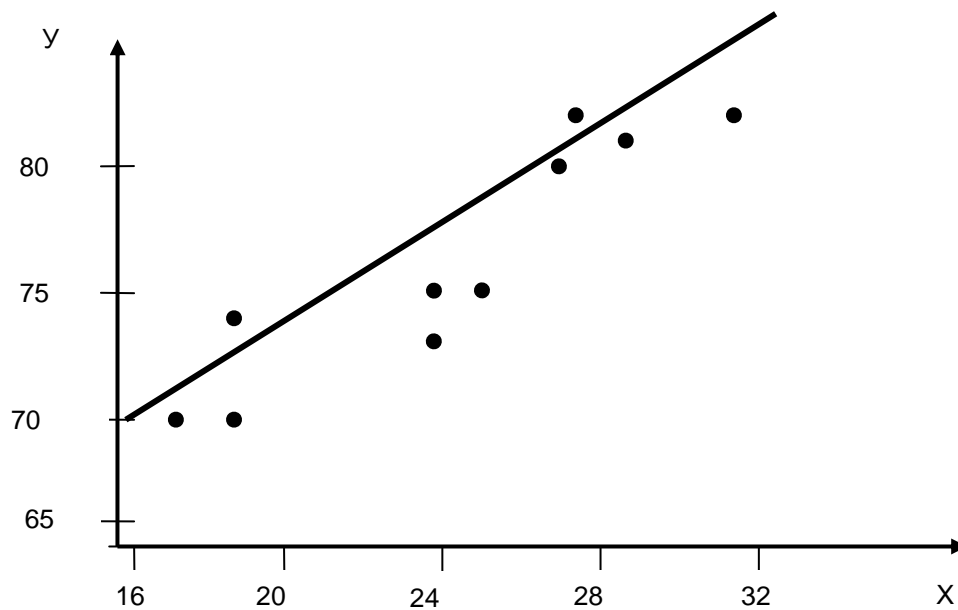


Рис. 9.2. Кореляційне поле залежності суми активів (y) від суми депозитів (x)

Параметри (a_0 , a_1) цього рівняння прямої визначимо на основі методу найменших квадратів, розв'язавши систему нормальних рівнянь:

$$\sum y = na_0 + a_1 \sum x,$$

$$\sum yx = a_0 \sum x + a_1 \sum x^2,$$

$$a_0 = \frac{\sum y \sum x^2 - \sum yx \sum x}{n \sum x^2 - \sum x \sum x},$$

$$a_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - \sum x \sum x}.$$

За даними табл. 9.3 система нормальних рівнянь прийме вид:

$$758 = 10a_0 + a_1 246,$$

$$18785 = a_0 246 + a_1 \times 6263,$$

звідси:

$$a_0 = \frac{758 \times 6\,263 - 18\,785 \times 246}{10 \times 6\,263 - 246 \times 246} = 59,718,$$
$$a_1 = \frac{10 \times 18\,785 - 246 \times 758}{10 \times 6\,263 - 246 \times 246} = 0,653.$$

Отже, рівняння регресії, яке виражає зв'язок між обсягом активів та обсягом депозитів, має вигляд:

$$\bar{y} = 59,718 + 0,653x.$$

Якщо параметри рівняння регресії визначені правильно, то повинна дотримуватись рівність сум теоретичних та емпіричних значень обсягу активів (див. табл. 9.3).

Коефіцієнт регресії $a_1 = 0,653$ показує, що з підвищенням обсягу депозитів на 1 млрд грош. од., обсяг активів зростає на 0,653 млрд грош. од. у середньому для сукупності банків.

Для оцінювання впливу факторної ознаки на результативну можна використовувати коефіцієнт еластичності:

$$K_{\text{ел}} = a_1 \times \frac{\bar{x}}{\bar{y}};$$
$$K_{\text{ел}} = 0,653 \times \frac{24,6}{75,8} = 0,21.$$

Коефіцієнт еластичності показує, що зі збільшенням обсягу депозитів на 1 %, обсяг активів підвищується на 0,21 %.

Приклад 9.3. Перевіримо адекватність однофакторної регресійної моделі та істотність показників тісноти кореляційного зв'язку.

У невеликих за обсягом сукупностях (у цьому прикладі має місце мала вибірка, $n < 30$) коефіцієнт регресії (a_1), схильний до випадкових коливань. Перевіримо його істотність за допомогою t-критерію Стьюдента:

$$t = \frac{|a_1|}{\mu_{a_1}},$$

де μ_{a_1} – середня помилка параметра a_1 ;

$$\mu_{a_1} = \sqrt{\frac{\sigma_{\text{зал}}^2}{\sigma_x^2(n-2)}},$$

1) де $\sigma_{\text{зал}}^2$ – залишкова дисперсія, яка розраховується за формулою (9.13) і дорівнює:

$$\sigma_{\text{зал}}^2 = \frac{32,2648}{10} = 3,52648,$$

2) де σ_x^2 – дисперсія факторної ознаки, яка дорівнює згідно з формулою (9.15):

$$\sigma_x^2 = \frac{6263}{10} - \left(\frac{246}{10}\right)^2 = 11,14.$$

Середня помилка параметра a_1 за формулою (9.15) складе:

$$\mu_{a_1} = \sqrt{\frac{3,52648}{11,14(10-2)}} = 0,1989 \text{ млрд грош. од.}$$

Фактичне значення t -критерію Стьюдента дорівнює:

$$t = \frac{0,653}{0,1989} = 3,28.$$

За таблицею критичних точок t -розподілу Стьюдента, за $\alpha = 0,05$ і числу ступенів свободи $k = n - m = 10 - 2 = 8$, критичне значення $t_{0,05} = 2,31$.

Оскільки $t_{\text{факт}} > t_{\text{крит}}$ ($3,28 > 2,31$), то вибірковий коефіцієнт регресії $a_1 = 0,653$ є ймовірним і значущим.

Визначимо інтервал, в якому із заданих рівнем значущості перебуває коефіцієнт регресії в генеральній сукупності:

$$a_1 - t \times \mu_{a_1} \leq a_1 \leq a_1 + t \times \mu_{a_1};$$

$$0,653 - 2,31 \times 0,1989 \leq a_1 \leq 0,653 + 2,31 \times 0,1989;$$

$$0,194 \leq a_1 \leq 0,112.$$

Отже, з рівнем значущості $\alpha = 0,05$ (з імовірністю помилитися в п'яти випадках із ста) можна стверджувати, що величина коефіцієнта регресії, що характеризує зв'язок між обсягом активів та обсягом депозитів у генеральній сукупності, перебуває від 0,194 до 1,112 млрд грош. од.

9.4. Показники тісноти та істотності кореляційного зв'язку

Перевірка адекватності регресійної моделі може бути доповнена кореляційним аналізом. Для цього необхідно визначити тісноту кореляційного зв'язку між змінними x та y . Тіснота кореляційного зв'язку, як і будь-якого іншого, може бути виміряна **емпіричним кореляційним відношенням** η_e , коли міжгрупова дисперсія σ_M^2 характеризує відхилення групових середніх результативної ознаки від загальної дисперсії: $\eta_e = \sqrt{\sigma_M^2/\sigma^2}$.

Слід відрізнити від емпіричного кореляційного відношення теоретичне η . Воно показує відносну величину, яку отримують у результаті порівняння середнього квадратичного відхилення значень результативної ознаки σ_M із середнім квадратичним відхиленням емпіричних (фактичних) значень результативної ознаки σ :

$$\eta = \sqrt{\sigma_M^2/\sigma^2},$$

$$\text{де } \sigma_M = \sqrt{\frac{\sum(\tilde{y} - \bar{y})^2}{n}} = \sqrt{\sigma_{M\tilde{y}}^2};$$

$$\sigma = \sqrt{\frac{\sum(y - \bar{y})^2}{n}} = \sqrt{\sigma_y^2}.$$

$$\text{Тоді: } \eta = \sqrt{\frac{\sum(\tilde{y} - \bar{y})^2}{\sum(y - \bar{y})^2}}. \quad (9.16)$$

Зміна значення η пояснюється впливом факторної ознаки.

В основу розрахунку кореляційного відношення покладене правило складання дисперсій, тобто $\sigma^2 = \sigma_M^2 + \sigma_\varepsilon^2$, де σ_ε^2 відображає варіацію усіх факторів, окрім x , тобто є залишковою дисперсією:

$$\sigma_\varepsilon^2 = \sigma_\varepsilon^2 = \frac{\sum(y - \tilde{y})^2}{n}.$$

Тоді формула теоретичного кореляційного відношення прийме вид:

$$\eta = \sqrt{\frac{\sigma_M^2}{\sigma^2}} = \sqrt{\frac{\sigma^2 - \sigma_{\text{зап}}^2}{\sigma^2}} = \sqrt{1 - \frac{\sigma_{\varepsilon}^2}{\sigma^2}}, \quad (9.17)$$

або:

$$\eta = \sqrt{1 - \frac{\sum (y - \tilde{y})^2}{\sum (y - \bar{y})^2}}. \quad (9.18)$$

Підкореневе вираження кореляційного відношення є **коефіцієнтом детермінації** (міри причинності). Цей коефіцієнт показує частку варіації результативної ознаки під впливом варіації ознаки – фактора.

Теоретичне кореляційне відношення застосовується для вимірювання тісноти зв'язку у лінійній та криволінійній залежностях між результативною та факторною ознакою. У криволінійних зв'язках теоретичне кореляційне відношення, що розраховується за формулами (9.17) та (9.18), називають *індексом кореляції R*.

Як видно з формул (9.17) і (9.18) кореляційне відношення може перебувати в межах від 0 до 1, тобто $0 \leq \eta \leq 1$. Чим ближче η до 1, тим зв'язок між ознаками тісніше.

Продемонструємо розрахунок теоретичного кореляційного відношення як міри тісноти зв'язку на прикладі даних табл. 9.4, для якого за рівнянням прямої регресії $\tilde{y} = 4 + 0,6x$ розраховані значення денної продуктивності праці одного робочого залежно від стажу роботи.

Теоретичне кореляційне відношення розрахуємо двома способами (див. табл. 9.4):

$$\text{за формулою (9.16)} \quad \eta = \sqrt{\frac{29,7}{32,1}} = \sqrt{0,925} = 0,962;$$

$$\text{за формулою (9.18)} \quad \eta = \sqrt{1 - \frac{2,4}{29,7}} = \sqrt{0,92} = 0,96.$$

Розрахункові значення, необхідні для розрахунку η

$y - \bar{y}$	$(y - \bar{y})^2$	$\tilde{y} - \bar{y}$	$(\tilde{y} - \bar{y})^2$	$y - \tilde{y}$	$(y - \tilde{y})^2$
1	2	3	4	5	6
-3,3	10,89	-2,7	7,29	-0,6	0,36
-2,3	5,29	-2,1	4,41	-0,2	0,04
-1,3	1,69	-1,5	2,25	0,2	0,04
-0,3	0,09	-0,9	0,81	0,6	0,36
-0,3	0,09	-0,3	0,09	0,0	0,00
0,7	0,49	0,3	0,09	0,4	0,16
0,7	0,49	0,9	0,81	-0,2	0,04
1,7	2,89	1,5	2,25	0,2	0,04
2,7	7,29	2,1	4,41	0,6	0,36
1,7	2,89	2,7	7,29	-1,0	1,00
Усього	32,10	-	29,70	-	2,40

Отримане значення теоретичного кореляційного відношення свідчить про дуже тісний зв'язок між ознаками (відповідно до шкали Чеддока, табл. Г.1 в додатку Г).

Коефіцієнт детермінації дорівнює 0,925. Звідси маємо 92,5 % загальної варіації продуктивності праці обумовлено варіацією фактора – стажу роботи; тільки 7,5 % загальної варіації не можна пояснити варіацією стажу роботи.

Крім того, за лінійної форми зв'язку застосовується інший показник тісноти зв'язку – **лінійний коефіцієнт кореляції**, що був запропонований англійським математиком К. Пірсоном:

$$r = \frac{\overline{xy} - \bar{x} \times \bar{y}}{\sigma_x \sigma_y} = \frac{\sum(x - \bar{x}) \times (y - \bar{y})}{n \times \sigma_x \sigma_y}, \quad (9.19)$$

де n – число спостережень.

Для практичних розрахунків з малим числом спостережень, а саме для ($n \leq 20 - 30$), лінійний коефіцієнт кореляції зручніше розраховувати за формулою:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n} \right) \times \left(\sum y^2 - \frac{(\sum y)^2}{n} \right)}}. \quad (9.20)$$

Отже, коефіцієнт перебуває у межах від -1 до 1 .

Від'ємне значення показує зворотний зв'язок, додатне – прямий. Якщо $r = 0$, лінійний зв'язок відсутній. Чим ближче коефіцієнт кореляції до 1 , тим тісніше зв'язок між ознаками. Якщо $r = \pm 1$, зв'язок функціональний.

Приклад 9.4. Визначимо тісноту кореляційного зв'язку між обсягом активів і депозитів за допомогою лінійного коефіцієнта кореляції і коефіцієнта детермінації, дані для розрахунку яких наведені в табл. 9.3.

Лінійний коефіцієнт кореляції, розрахований за формулою (9.20), буде дорівнювати:

$$r = \frac{18\,785 - \frac{246 \times 758}{10}}{\sqrt{\left(6\,263 - \frac{246^2}{10}\right) \times \left(57\,660 - \frac{758^2}{10}\right)}} = 0,6662.$$

Коефіцієнт кореляції показує, що між обсягом активів та обсягом депозитів спостерігається помітний зв'язок (за шкалою Чеддока).

Коефіцієнт детермінації $r^2 = 0,6662^2 = 0,4438$ показує, що 44,38 % загальної варіації обсягу активів банків зумовлено варіацією в обсягах депозитів, а 55,62 % ($100\% - 44,38\%$) – іншими факторами, які в цьому прикладі не були враховані.

Квадрат лінійного коефіцієнта кореляції r^2 називають **лінійним коефіцієнтом детермінації**, значення якого перебуває в межах від 0 до 1. Його інтерпретація аналогічна коефіцієнту детермінації η^2 .

Значення η і r однакові лише за наявності прямолінійного зв'язку. Якщо вони різні, то це свідчить, що зв'язок між ознаками криволінійний. Установлено: якщо різниця квадратів η^2 і r^2 не перевищує 0,1, то гіпотезу про прямолінійну форму зв'язку можна вважати підтвердженою.

Показники тісноти зв'язку, розраховані за даними невеликої за кількістю сукупності, зазнають впливу випадкових величин. Це викликає необхідність перевірки їх істотності, що дає право розповсюджувати висновки за результатами вибірки на генеральну сукупність.

Приклад 9.5. Вірогідність вибіркового коефіцієнта кореляції перевіримо за допомогою критичних значень вибіркового коефіцієнта кореляції за різних чисел ступеня свободи і рівнях значущості (табл. Д.1 в додатку Д). Для прикладу (див. табл. 9.3), табличне значення коефіцієнта кореляції

за $\alpha = 0,05$ і $k = 8$ становить $r_{0,05} = 0,632$. Оскільки $r_{\text{факт}} > r_{0,05}$ ($0,6662 > 0,632$), то можна вважати вибірковий коефіцієнт кореляції ймовірним.

Адекватність регресійної моделі за $n < 30$ оцінюється за допомогою F-критерію Фішера:

$$F = \frac{\sigma_{y_x}^2}{\sigma_{\varepsilon}^2} \times \frac{n-m}{m-1}, \quad (9.21)$$

де m – число параметрів моделі;

$\sigma_{\bar{y}}^2$ – факторна дисперсія, що характеризує варіацію результативної ознаки під впливом ознаки-фактора, що включена в модель.

Факторна дисперсія розраховується за формулою:

$$\sigma_{\bar{y}}^2 = \frac{\sum(\tilde{y} - \bar{y})^2}{n} \quad (9.22)$$

або:

$$\sigma_{\bar{y}}^2 = \sigma_y^2 - \sigma_{\varepsilon}^2 = \frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2 - \sigma_{\varepsilon}^2. \quad (9.23)$$

$$\sigma_{\bar{y}}^2 = \frac{57\,660}{10} - \left(\frac{758}{10}\right)^2 - 3,526\,48 = 16,83;$$

$$F = \frac{16,83}{3,526\,48} \times \frac{10-2}{2-1} = 38,179.$$

Табличне значення F з рівнем значущості $\alpha = 0,05$ і числом ступенів свободи $k_1 = n - m = 10 - 2 = 8$ та $k_2 = m - 1 = 2 - 1 = 1$ дорівнює 5,32. Оскільки $F_{\text{розрах}} > F_{\text{табл.}}$, то рівняння регресії можна вважати адекватним.

Для перевірки надійності регресійної моделі розрахуємо помилку апроксимації:

$$\bar{\varepsilon} = \frac{1}{n} \sum \frac{|y - \tilde{y}|}{y} \times 100\%, \quad (9.24)$$

$$\bar{\varepsilon} = \frac{1}{10} \times 0,19855 \times 100\% = 1,98\%.$$

Оскільки $\bar{\epsilon}$ не перевищує 12 – 15 %, то регресійною моделлю можна користуватися на практиці.

9.5. Побудова множинних кореляційно-регресійних моделей

Парна кореляція використовується в тих випадках, коли серед факторів, що впливають на результативну ознаку, є домінуючий.

Проте в практиці економічного аналізу часто доводиться вивчати явища, які складаються під впливом не одного, а багатьох різних факторів, кожний з яких окремо може не справляти вирішального впливу. Сукупний же вплив факторів іноді виявляється достатньо сильним, щоб за їх змінами можна було робити висновки про величини показника досліджуваного явища. Методи вимірювання кореляційного зв'язку одночасно між двома, трьома та більше кореляційними ознаками створюють вчення про **множинну кореляцію** (питання множинної кореляції вперше досліджував англійський вчений Ф. А. Еджворт наприкінці ХІХ ст.).

У моделях множинної кореляції залежна змінна у розглядається як функція кількох (у загальному випадку n) незалежних змінних x .

Використання прикладних програм дає можливість розв'язувати кореляційно-регресійні моделі різних залежностей і вибирати із цієї множини таке рівняння, яке найточніше описує ступінь наближення фактичних даних до теоретичних і, відповідно, дає найменшу суму квадратів відхилень фактичних даних від розрахованих за рівнянням зв'язку.

Багатофакторний кореляційно-регресійний аналіз може бути застосований для:

- 1) розрахунку теоретичних (очікуваних) значень результативної ознаки;
- 2) оцінювання і порівняння фактичного та розрахункового значень результативної ознаки;
- 3) порівняльного аналізу різних сукупностей;
- 4) виявлення резервів виробництва та об'єктивного оцінювання результатів роботи об'єкта дослідження;
- 5) прогнозування суспільних явищ;
- 6) розроблення нормативів.

Парна кореляція через те, що разом з досліджуваним фактором на результативну ознаку впливають й інші фактори, не завжди дає правильне уявлення про зв'язок між результативною і факторною ознаками

(перебільшує або применшує міру залежності). Перевага багатфакторного кореляційно-регресійного аналізу порівняно з парною кореляцією полягає в тому, що він дає змогу оцінити ступінь впливу на результативну ознаку кожного з включених у модель факторів за умови фіксованого положення (на середньому рівні) решти факторів.

Відбір найістотніших факторів до кореляційної моделі є одним із найважливіших і принципових завдань багатфакторного кореляційно-регресійного аналізу. Природно, що всі фактори, які впливають на досліджувану результативну ознаку, до рівняння регресії включити не можна. З усього комплексу таких факторів необхідно відібрати найбільш важливі, істотні. Захоплення великою кількістю факторів за відносно невеликої чисельності сукупності може призвести до неякісних результатів. Крім того, із збільшенням у рівнянні регресії кількості параметрів значно утруднюється інтерпретація отриманих результатів.

Велику роль у відборі факторів відіграють завчасно побудовані та проаналізовані факторні групування. Дуже важливого значення тут набувають комбінаційні групування, які дозволяють визначити вплив на результативну ознаку фактора, що цікавить дослідника, з фіксованими значеннями інших факторів. Можна зробити висновок про те, що статистичні групування становлять основу кореляційного та дисперсійного аналізів; найбільшої ефективності останні досягають у поєднанні з методом групувань.

Як уже було зазначено, для забезпечення стійкості параметрів рівняння зв'язку кількість факторів, включених до моделі, має бути в 6 – 10 разів меншою від чисельності досліджуваної сукупності. Водночас сукупність, з якої відбирають фактори, повинна бути якісно однорідною.

Відбираючи фактори, потрібно виключати ті, що взаємно дублюють один одного та перебувають у функціональному зв'язку. Функціональний або близький до нього зв'язок між самими факторами вказує на *мультиколінеарність* (для двох – колінеарність). Наявність мультиколінеарності свідчить про те, що ці фактори відображають ту саму сторону впливу на результативну ознаку.

За високої корельованості факторів (тіснота зв'язку між двома факторами перевищує $r > 0,8$) вплив одного з них акумулює і вплив другого. Отримані кореляційні моделі стають нестійкими.

За умови мультиколінеарності включення до кореляційної моделі взаємопов'язаних факторів можливе тоді, коли тіснота зв'язку між ними

менша, ніж тіснота зв'язку результативної ознаки з кожним фактором. Потрібно, щоб кореляційна модель містила незалежні, не взаємодубльовані фактори. Небажаним є включення до однієї моделі часткових і загальних факторів. Повністю слід виключити фактори, функціонально пов'язані з результативною ознакою.

Вибір рівняння регресії для побудови багатофакторної моделі спирається на положення теорії досліджуваного явища або практичний досвід попередніх досліджень. Графічний метод у цьому випадку є малоефективним. Якщо попередніх даних немає, то будуються комбінаційні групування, використовуються експертні оцінки, вивчаються парні зв'язки між результативною ознакою і кожним фактором, перебираються функції різних типів, здійснюється послідовний перехід від лінійних рівнянь зв'язку до більш складних видів.

Виконання усіх цих прийомів пов'язане із значною кількістю зайвих підрахунків. Однак кореляційні зв'язки в більшості випадків відображаються функціями лінійного типу або степеневими, які шляхом логарифмування або заміни змінних можна звести до лінійного вигляду, тому рівняння множинної регресії можна будувати в лінійній формі.

За n змінних лінійне рівняння багатофакторного зв'язку має вигляд:

$$\tilde{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n, \quad (9.25)$$

де \tilde{y} – залежна змінна (результативна ознака);

x_1, x_2, \dots, x_n – незалежні змінні (фактори);

a_0 – початок відліку, який економічного сенсу не має або вказує на наявність не врахованих у моделі факторів;

a_1, a_2, \dots, a_n – коефіцієнти регресійного рівняння.

Рівнянням множинної регресії називають рівняння, за допомогою якого виражається кореляційний зв'язок між кількома ознаками. Параметри рівняння регресії, як і у випадку парної кореляції, знаходять методом найменших квадратів.

Коефіцієнти множинної регресії показують ступінь середньої зміни результативної ознаки зі зміною відповідної факторної ознаки на одиницю за умови, що всі інші фактори, які включені до рівняння регресії, залишаються фіксованими на одному (середньому) рівні.

Коефіцієнти множинної регресії, які характеризують зв'язок між результативною ознакою і фактором із фіксованим значенням інших факторів, називають **коефіцієнтами чистої регресії**, а коефіцієнти парної регресії – **коефіцієнтами повної регресії**.

Коефіцієнти чистої регресії, що мають різний фізичний смисл і одиниці вимірювання, не дають чіткого уявлення про те, які саме фактори найістотніше впливають на результативну ознаку. Крім того, величина коефіцієнтів регресії залежить від ступеня варіації ознаки.

Щоб привести коефіцієнти чистої регресії до порівнянного вигляду, їх виражають у стандартизованій формі у вигляді коефіцієнтів еластичності (E) та бета-коефіцієнтів (β).

Коефіцієнти еластичності показують, на скільки відсотків змінюється величина результативної ознаки зі зміною відповідної факторної ознаки на один відсоток за умови фіксованого значення інших факторів.

Коефіцієнти еластичності та коефіцієнти чистої регресії зв'язані між собою таким відношенням (9.26):

$$E_i = a_i \times \frac{\bar{x}_i}{\bar{y}}, \quad (9.26)$$

де a_i – коефіцієнт чистої регресії за i -го фактора;

\bar{x}_i та \bar{y} – середні значення відповідно i -го фактора та результативної ознаки.

Бета-коефіцієнти показують, на скільки середньоквадратичних відхилень σ_y зміниться результативна ознака зі зміною відповідного фактора на одне значення середньоквадратичного відхилення σ_x (за умови постійності інших факторів, включених до рівняння регресії).

Бета-коефіцієнти обчислюються за формулою:

$$\beta_i = a_i \frac{\sigma_{x_i}}{\sigma_y}, \quad (9.27)$$

де a_i – коефіцієнт чистої регресії за i -го фактора;

σ_{x_i} і σ_y – середні квадратичні відхилення, відповідно, за i -м фактором і результативною ознакою.

З наведеної формули випливає, що бета-коефіцієнти мають той самий знак (плюс, мінус), що і коефіцієнти чистої регресії.

Отже, бета-коефіцієнти характеризують фактори, в розвитку яких приховуються найбільші резерви поліпшення результативної ознаки.

У парному лінійному зв'язку коефіцієнт кореляції є бета-коефіцієнтом:

$$r = b \frac{\sigma_x}{\sigma_y} = \beta. \quad (9.28)$$

Після побудови багатофакторної регресійної моделі розраховують характеристики тісноти зв'язку між залежною та незалежною змінними: парні, часткові, множинні коефіцієнти кореляції; множинний коефіцієнт детермінації, – а потім перевіряють адекватність моделі.

Для вимірювання тісноти зв'язку між двома розглядуваними із сукупності змінних (без урахування їх взаємодії з іншими змінними) використовують **парні коефіцієнти кореляції**. Методика розрахунку таких коефіцієнтів та їх інтерпретація аналогічні методиці розрахунку лінійного коефіцієнта кореляції у випадку однофакторного зв'язку.

Парні коефіцієнти кореляції розраховують за формулами:

$$r_{y x_1} = \frac{\overline{x_1 y} - \bar{x}_1 \bar{y}}{\sigma_{x_1} \sigma_y}; \quad (9.29)$$

$$r_{y x_2} = \frac{\overline{x_2 y} - \bar{x}_2 \bar{y}}{\sigma_{x_2} \sigma_y}; \quad (9.30)$$

$$r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \bar{x}_1 \bar{x}_2}{\sigma_{x_1} \sigma_{x_2}}. \quad (9.31)$$

У реальних умовах усі змінні взаємозв'язані. Тіснота цього зв'язку визначається **частковими коефіцієнтами кореляції**, які характеризують ступінь і вплив одного із аргументів на функцію за умови, що інші незалежні змінні фіксуються на постійному рівні. Залежно від кількості змінних,

вплив яких елімінується, часткові коефіцієнти кореляції можуть бути різного порядку: з виключенням впливу однієї змінної отримуємо частковий коефіцієнт кореляції першого порядку; з виключенням впливу двох змінних – другого порядку і т. д. Парний коефіцієнт кореляції між функцією та аргументом звичайно не дорівнює відповідному частковому коефіцієнту.

Частковий коефіцієнт кореляції першого порядку між ознаками x_1 та y зі виключенням впливу ознаки x_2 розраховують за формулою:

$$r_{y_{x_1(x_2)}} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{(1-r_{yx_2}^2) \times (1-r_{x_1x_2}^2)}}; \quad (9.32)$$

залежність y від x_2 з виключенням впливу x_1 :

$$r_{y_{x_2(x_1)}} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{(1-r_{yx_1}^2) \times (1-r_{x_1x_2}^2)}}. \quad (9.33)$$

Можна розраховувати взаємозв'язок факторних ознак із виключенням впливу результативної ознаки:

$$r_{x_1x_2(y)} = \frac{r_{x_1x_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{(1-r_{yx_1}^2) \times (1-r_{yx_2}^2)}}, \quad (9.34)$$

де r – парні коефіцієнти кореляції між відповідними ознаками.

Вивчення парних і часткових коефіцієнтів кореляції дозволяє відібрати найбільш істотні фактори.

Показником тісноти зв'язку між результативною і двома або більше факторними ознаками є **сукупний коефіцієнт множинної кореляції** ($R_{yx_1, x_2, \dots, x_n}$). У випадку лінійного двофакторного зв'язку цей коефіцієнт розраховують за формулою:

$$R_{yx_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1x_2}}{1 - r_{x_1x_2}^2}}. \quad (9.35)$$

Сукупний коефіцієнт множинної кореляції вимірює одночасний вплив факторних ознак на результативну. Його значення перебувають у межах від -1 до $+1$. Чим менше емпіричні значення розглядуваного показника відхиляються від лінії множинної регресії, тим кореляційний зв'язок є більш інтенсивним, а отже, значення R ближче до одиниці.

Для вимірювання частки варіації результативної ознаки, що пояснюється впливом факторів, які включені в рівняння множинної регресії, використовують **сукупний коефіцієнт множинної детермінації** (R^2). Значення цього коефіцієнта перебувають у межах від 0 до 1 . Чим ближче R^2 до одиниці, тим варіація аналізованої ознаки у більшій мірі характеризується впливом відібраних факторів.

Але слід зауважити, що коефіцієнт R^2 не можна використовувати для перевірки адекватності рівняння регресії. Причина в тому, що багатofакторний регресійний аналіз оперує випадковими спостереженнями, але не обов'язково розподіленими за багатомірним нормальним законом (цьому закону повинні підпорядковуватись відхилення фактичних значень функції від розрахункових). Сукупний коефіцієнт множинної детермінації визначає тільки якість вирівнювання за рівнянням регресії.

Перевірку істотності рівняння регресії здійснюють на основі F -критерію Фішера (9.21). Вважають, що рівняння регресії можна використовувати на практиці не менше ніж чотирикратно.

Для оцінювання вагомості коефіцієнтів регресії за лінійної залежності у від x_1 і x_2 використовують t -критерій Стьюдента за умови $n - m - 1$ ступенів свободи:

$$t_{a_1} = \frac{a_1 \sigma_{x_1} \sqrt{1 - r_{x_1 x_2}^2} \sqrt{n - m - 1}}{\sigma_y \sqrt{1 - R_{yx_1 x_2}^2}} ; \quad (9.36)$$

$$t_{a_2} = \frac{a_2 \sigma_{x_2} \sqrt{1 - r_{x_1 x_2}^2} \sqrt{n - m - 1}}{\sigma_y \sqrt{1 - R_{yx_1 x_2}^2}} . \quad (9.37)$$

Значення a_1 та a_2 беруть за модулем. Параметри визначаються вагомими, якщо $t_{розр} > t_{табл}$ з рівнем значущості α і числом ступенів свободи $n - m - 1$.

Оцінку вагомості коефіцієнтів регресії за допомогою t-критерію використовують для завершення відбору істотних факторів у процесі багатокрокового регресійного аналізу. Його змістовність полягає в тому, що після оцінювання регресії із моделі виключають той фактор, коефіцієнт за яким незначний і має найменше значення критерію. Потім рівняння регресії будується без цієї факторної ознаки; знову проводиться оцінювання адекватності рівняння і вагомості коефіцієнтів регресії. Такий процес триває, доки всі коефіцієнти регресії не будуть вагомими, що свідчить про наявність у регресійній моделі тільки істотних факторів. У деяких випадках розрахункове значення $t_{\text{розрах}} \text{ перебуває близько } t_{\text{табл}}$, тому таку ознаку можна залишити для подальшої перевірки його значущості у сполученні з іншим набором факторів.

Послідовне виключення неістотних факторів складає основу багатокрокового регресійного аналізу.

Істотність сукупного коефіцієнта кореляції оцінюється також за допомогою t-критерію Стюдента за формулою:

$$t_{R_{y_{x_1, x_2}}} = \frac{R_{y_{x_1, x_2}}^2 \sqrt{n - m - 1}}{1 - R_{y_{x_1, x_2}}^2} \quad (9.38)$$

Якщо $t_{\text{розрах}} > t_{\text{табл}}$ із рівнем значущості 0,01 або 0,05 і числом ступенів свободи $(n - m - 1)$, то коефіцієнт множинної кореляції визнається істотним.

9.6. Методи вивчення зв'язку соціальних явищ

З використанням кореляційно-регресійних моделей (КРМ) для вирішення конкретних соціально-економічних завдань необхідно:

щоб побудова рівняння регресії була основана на розглянутих принципах і умовах;

на основі аналізу знаків коефіцієнтів регресії виявити ті змінні, коефіцієнти регресії яких мають знаки, що не відповідають теоретичним уявленням. Це може бути наслідком функціональних зв'язків між пояснювальними змінними, тобто наявністю колінеарності;

якщо розрахунки виконуються за допомогою прикладних програм за допомогою обчислювальної техніки, то програма видає значення t-критерію для кожного із коефіцієнтів регресії. Отже, якщо коефіцієнт регресії

перевищує середньоквадратичну помилку не менше ніж у 1,95 рази (тобто $t \geq 1,95$), то він вважається статистично вагомим.

Для аналізу адекватності рівняння регресії можливі такі варіанти:

побудована модель на основі її перевірки за F-критерієм Фішера у цілому є адекватною; усі коефіцієнти регресії є вагомими. Таку модель можна використовувати для ухвалення рішень і здійснення прогнозів;

модель за F-критерієм Фішера адекватна, але частина коефіцієнтів регресії невагома. У такому випадку модель може бути використана для ухвалення деяких рішень, але не для прогнозування;

модель за F-критерієм Фішера адекватна, але всі коефіцієнти регресії невагомі. У такому випадку модель повністю вважається неадекватною; на її основі не формуються рішення і не здійснюються прогнози;

регресію можна використовувати для аналізу діяльності одиниць сукупності зі значенням коефіцієнта детермінації близьким до одиниці та не нижчим 0,5;

регресія використовується для прогнозу: якщо значення змінних перебуває в середині інтервалу фактичних значень факторів – це точковий прогноз; якщо вони виходять за ці межі – дають інтервальну оцінку очікуваного значення результату.

Якщо необхідно вийти за реальні межі вимірювання факторної ознаки, то слід дотримуватись такого обмеження: не можна підставляти в рівняння регресії значення x , які значно відрізняються від тих, на основі яких це рівняння було отримано. Для вибору значень прогнозованої ознаки-фактора рекомендується не виходити за межі $1/3$ розмаху варіації як більше мінімального, так і більше максимального значення ознаки-фактора, що мали місце у вихідних даних.

Отже, можна пропонувати такі рекомендації щодо побудови кореляційно-регресійної моделі:

1) ознаки-фактори повинні перебувати в причинному зв'язку з результативною ознакою (наслідком);

2) ознаки-фактори не повинні бути складовими результативною ознакою або її функціями;

3) ознаки-фактори не повинні дублювати один одного, тобто бути колінеарними (з коефіцієнтом кореляції більше 0,8);

4) не треба включати у модель фактори різних рівней ієрархії, тобто фактор найближчого порядку та його субфактори. Наприклад, у модель собівартості продукції не слід включати і продуктивність праці, і вік робітників як субфактори самої продуктивності;

5) для результативної ознаки та факторних ознак бажано виконання єдності одиниці сукупності, до якої вони віднесені. Наприклад, якщо y – продукція підприємства, то і фактори x повинні відноситись до підприємства (рівень спеціалізації, рівень браку, чисельність кваліфікованих робітників тощо);

6) математична форма рівняння регресії повинна відповідати логіці зв'язку факторів із результатом у реальному середовищі. Виходячи із цього вибирають адитивну або мультиплікативну форму зв'язку;

7) перевага віддається тій моделі, в якій менше число факторів із тим же коефіцієнтом детермінації або навіть із несуттєво меншим;

8) якщо, крім кількісних факторів, у багатофакторному регресійному аналізі в рівняння включають і не кількісні, то їх наявність у одиниць сукупності позначають одиницею, а відсутність – нулем, тобто вводять фіктивні змінні $U = \begin{cases} 1 \\ 0 \end{cases}$. Число фіктивних змінних повинно бути на одиницю

менше числа градацій якісного (не кількісного) фактора. За допомогою цього прийому вимірюють вплив якісних факторів, ізолюючи їх від впливу кількісних факторів.

Головна відмінність методу кореляційно-регресійного аналізу від методу аналітичних групувань у тому, що кореляційно-регресійний аналіз дозволяє розділити вплив комплексу факторних ознак, аналізувати різні сторони складної системи взаємозв'язків. Якщо метод комбінаційного аналітичного групування не дає можливості аналізувати більш трьох факторів, то кореляційний метод за обсягу сукупності близько 100 одиниць дозволяє провести аналіз систем з 8 – 10 факторами та відокремити їх впливи.

Головними обмеженнями методу кореляційно-регресійного аналізу є те, що цей метод не може пояснити ролі факторних ознак у створенні результативної ознаки. Крім того, моделі на основі рівняння регресії володіють слабкими екстраполяційними властивостями, тому що не відображають тенденції розвитку соціально-економічних явищ і процесів і можуть застосовуватись для побудови лише короткострокових прогнозів, що мають імовірнісний характер.

Метод аналітичного групування і кореляційно-регресійний аналіз використовують основні параметри розподілу – середні величини та дисперсії, тому їх називають **параметричними**. У статистиці також широко використовують і **непараметричні методи** визначення взаємозв'язків,

які ґрунтуються на кількісних значеннях ознак і не потребують обчислення параметрів їх розподілів. Якщо в кореляційно-регресійному аналізі всі ознаки належать до метричної шкали, а в методі аналітичного групування це стосується результативної ознаки, то непараметричні методи використовують і тоді, коли є ознаки порядкової чи номінальної шкали. Але слід підкреслити, що ця перевага непараметричних методів досягається за рахунок меншої глибини аналізу взаємозв'язку; за їх допомогою лише визначають наявність зв'язку та вимірюють його тісноту.

Використання непараметричних методів оцінювання тісноти стохастичного зв'язку ґрунтується на порівнянні частот або часток умовних розподілів у таблицях взаємної спряженості.

Таблицею взаємної спряженості називають таблицю, яка містить зведену числову характеристику розглядуваної сукупності за двома та більше атрибутивними ознаками або комбінацією кількісних та атрибутивних ознак. Макет таблиці спряженості розмірністю $i \times j$, де $i = 1, 2, \dots, k$ – число варіантів значень однієї ознаки (A); $j = 1, 2, \dots, n$ – число варіантів значень другої ознаки (B) наведено в табл. 9.5.

Таблиця 9.5

Загальна схема таблиці спряженості

A \ B	B ₁	B ₂	...	B _j	Разом
A ₁	f ₁₁	f ₁₂	...	f _{1j}	f ₁₀
A ₂	f ₂₁	f ₂₂	...	f _{2j}	f ₂₀
...
A _i	f _{i1}	f _{i2}	...	f _{ij}	f _{i0}
Разом	f ₀₁	f ₀₂	...	f _{0j}	f ₀₀

Для визначення тісноти зв'язку двох якісних ознак, кожна із яких складається тільки із двох груп, застосовують **коефіцієнти асоціації та контингенції**:

$$K_a = \frac{f_{11} \times f_{22} - f_{12} \times f_{21}}{f_{11} \times f_{22} + f_{12} \times f_{21}}; \quad (9.39)$$

$$K_k = \frac{f_{11} \times f_{22} - f_{12} \times f_{21}}{\sqrt{(f_{11} + f_{12}) \times (f_{12} + f_{22}) \times (f_{11} + f_{21}) \times (f_{21} + f_{22})}}. \quad (9.40)$$

Значення коефіцієнта контингенції завжди менше значення коефіцієнта асоціації. Зв'язок між ознаками вважається підтвердженим, за умов $K_a \geq 0,5$ або $K_k \geq 0,3$.

Корисною мірою в аналізі 4-клітинкових таблиць взаємної спряженості є *відношення перехресних добутків*, або *відношення шансів*:

$$W = \frac{f_{11} \times f_{22}}{f_{12} \times f_{21}}. \quad (9.41)$$

Відношення шансів характеризує міру відносного ризику.

Приклад 9.6. За даними табл. 9.6 оцінимо щільність зв'язку між формою власності підприємства та рівнем задоволеності умовами життя робітників, що працюють на цих підприємствах, а також ефективність зміни підприємством форми власності.

Коефіцієнт контингенції свідчить про наявність стохастичного зв'язку:

$$K_k = \frac{13 \times 60 - 24 \times 3}{\sqrt{16 \times 63 \times 37 \times 84}} = 0,40.$$

Відношення шансів становить: $W = \frac{13 \times 60}{24 \times 3} = 10,8.$

Таблиця 9.6

Таблиця спряженості між двома ознаками

Форма власності підприємства	Кількість робітників		Разом
	Задоволені рівнем життя	Не задоволені рівнем життя	
Приватна	13	3	16
Державна	24	60	84
Разом	37	63	100

Тобто шанси бути задоволеними умовами життя серед робітників приватних підприємств у 10,8 раза вище, ніж серед робітників державних підприємств. Отже, для підприємства є сенс змінити форму власності, для робітника – змінити місце роботи.

Коли кожна із якісних ознак складається більше ніж із двох груп, то для визначення тісноти зв'язку застосовують **коефіцієнт взаємної спряженості Пірсона**, **коефіцієнт взаємної спряженості Чупрова** та модифікацію коефіцієнта спряженості Чупрова – **коефіцієнт Крамера**.

Коефіцієнт взаємної спряженості Пірсона:

$$K_p = \sqrt{\frac{\chi^2}{n + \chi^2}}, \quad (9.42)$$

де χ^2 – **коефіцієнт квадратичної спряженості хі-квадрат Пірсона**, який характеризує розбіжність між частотами (частками) умовного та безумовного розподілу; розраховується за формулою:

$$\chi^2 = n \sum_i \sum_j \frac{f_{ij}^2}{f_{i0} \times f_{0j}} - 1, \quad (9.43)$$

де n – кількість спостережень.

Фактичні значення χ^2 Пірсона порівнюються з критичними. Критичні значення χ^2 для $\alpha = 0,05$ і числа ступенів свободи $k = (m_x - 1) \times (m_y - 1)$ наведено в табл. Г.2, в додатку Г.

Якщо фактичне значення перевищує критичне, то з імовірністю 0,95 істотність зв'язку між ознаками доведена.

Коефіцієнт взаємної спряженості Чупрова:

$$K_c = \sqrt{\frac{\chi^2}{n \sqrt{(m_x - 1) \times (m_y - 1)}}}, \quad (9.44)$$

де m_x – число груп за ознакою x ;

m_y – число груп за ознакою y .

Оскільки за незалежності ознак $\chi^2 = 0$, то $K_c = 0$. У функціональному зв'язку $K_c = 1$ за умови, що $m_x = m_y$.

У разі, коли $m_x \neq m_y$, зручніше користуватися формулою Крамера:

$$K_k = \sqrt{\frac{\chi^2}{n(m_{\min} - 1)}}, \quad (9.45)$$

де m_{\min} – мінімальне число груп (m_x або m_y).

Якщо $m_x = m_y$, то значення коефіцієнтів, обчислені за формулами Чупрова та Крамера, збігаються.

Значення коефіцієнтів Чупрова, Крамера та Пірсона коливаються в межах від 0 до 1. Коефіцієнт Чупрова враховує кількість виділених груп за кожною ознакою і дає найбільш обережну оцінку зв'язку. Зі значенням $K_c \geq 0,3$ можна говорити про помірний або щільний зв'язок між ознаками.

Перевірка істотності зв'язку здійснюється на основі критерію χ^2 -квадрата Пірсона.

Приклад 9.7. За допомогою коефіцієнтів взаємної спряженості дослідимо щільність зв'язку між типом суб'єкта господарювання та видом вкладу залежно від валюти (табл. 9.7).

Таблиця 9.7

Вклади в кредитні організації регіону

Види вкладу	Усі вклади, млн грош. од.	у тому числі		
		підприємств та організацій	фізичних осіб	банків
Усього	702,7	69,3	580,6	52,8
у тому числі:				
у гривнях	222,1	18,0	200,7	3,4
в іноземній валюті	480,6	51,3	379,9	49,4

За допомогою коефіцієнтів Пірсона, Чупрова та Крамера визначимо тісноту зв'язку між ознаками:

$$K_p = \sqrt{\frac{18,84}{702,7 + 18,84}} = 0,16;$$

$$K_c = \sqrt{\frac{18,84}{702,7 \times (2 - 1) \times (3 - 1)}} = 0,12;$$

$$K_k = \sqrt{\frac{18,84}{702,7 \times (2 - 1)}} = 0,16.$$

Оскільки, розраховані коефіцієнти тісноти зв'язку мають низькі значення (навіть менше ніж 0,3), то можна зробити висновок, що між ознаками зв'язку немає, і вони не залежать одна від одної. Перевіримо істотність цього висновку за допомогою χ^2 -квадрата критерію Пірсона.

Фактичне значення коефіцієнта квадратичної спряженості – χ^2 -квадрат Пірсона складає:

$$\chi^2 = 702,7 \left[\frac{18^2}{222,1 \times 69,3} + \frac{200,7^2}{222,1 \times 580,6} + \frac{3,4^2}{222,1 \times 52,8} + \frac{51,3^2}{480,6 \times 69,3} + \frac{379,9^2}{480,6 \times 580,6} + \frac{49,4^2}{480,6 \times 52,8} - 1 \right] = 18,84,$$

що значно перевищує критичне для $\alpha = 0,05$ і числа ступенів свободи $k = (2 - 1) \times (3 - 1) = 2$, яке дорівнює 5,99. Отже, з імовірністю 95 % відсутність зв'язку між видом вкладу залежно від валюти та суб'єктом господарювання доведена.

Методи аналізу таблиць взаємної спряженості можна використати і для кількісних ознак.

Бісеріальний коефіцієнт кореляції запропоновано К. Пірсоном. Його призначено для дослідження кореляції в таблицях розміром $2 \times n$, які є дихотоміями за певною номінальною ознакою і класифікаціями за номінальною або порядковою ознакою, що класифікується за q класами та може бути впорядкованою або невпорядкованою. Вихідний розподіл має бути двовимірним нормальним. За класифікації за порядковою ознакою бісеріальний коефіцієнт розраховується за формулою:

$$r_b = \frac{(\bar{x}_1 - \bar{x})}{ns_x z_k}, \quad (9.46)$$

де \bar{x}_1 – середнє за першим рядком;

\bar{x} – загальне середнє за всією таблицею;

s_x – вибіркве середнє квадратичне відхилення;

n – загальна чисельність усіх вибірок;

z_k – ордината щільності нормального розподілу в точці k , де k – розв'язок рівняння:

$$1 - F(k) = n_1/n, \quad (9.47)$$

де n_1 – чисельність першого рядка.

Значення бісеріального коефіцієнта кореляції можуть змінюватися від -1 до $+1$. Його похибку можна визначити за формулою:

$$m_{rb} = \frac{1-r_b}{\sqrt{n}}. \quad (9.48)$$

Бісеріальний коефіцієнт кореляції має t -розподіл з кількістю ступенів свободи $(n - 2)$.

9.7. Непараметричні показники зв'язку. Рангові коефіцієнти зв'язку

В аналізі соціально-економічних явищ і процесів використовують умовні оцінки (наприклад, ранги). **Ранжування** – це процедура впорядкування об'єктів вивчення, яка виконується на основі переваги. **Ранг** – це порядковий номер значень ознаки, розташованих у порядку зростання або спадання їх величин.

Якщо значення ознаки мають однакову кількісну оцінку, то ранг усіх цих значень приймається дорівнює середній арифметичній від відповідних номерів місць, які визначають. Такі ранги називають *зв'язними*.

Для вимірювання зв'язку між ознаками порядкової шкали використовують **коефіцієнт кореляції рангів Спірмена**. Його розрахунок ґрунтується на різниці рангів $d = R_x - R_y$, де R_x, R_y – ранги елементів сукупності, відповідно, за першою і другою ознаками. Цей коефіцієнт обчислюють за формулою:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}, \quad (9.49)$$

де n – число елементів сукупності.

Коефіцієнт кореляції рангів, як і лінійний коефіцієнт кореляції, може набувати значень від -1 до $+1$. Якщо два рядки рангів повністю збігаються, то $\sum d^2 = 0$. Отже, має місце повний прямий зв'язок і $\rho = 1$. За повного зворотного зв'язку (ранги двох рядків розташовані у зворотному порядку) $\rho = -1$.

Для прикладу використаємо дані табл. 9.8, у якій наведені ранги десяти підприємств згідно з обсягом реалізованої продукції та розміром прибутку.

Таблиця 9.8

Розрахунок коефіцієнта кореляції рангів Спірмена

Номер підприємства	Ранжування				d	d ²
	Обсяг реалізованої продукції, млн грн	R _x	Розмір прибутку, тис. грн	R _y		
1	1,2	2	200	1	1	1
2	1,4	3	240	3	0	0
3	5,2	10	384	6	4	16
4	0,9	1	256	4	-3	9
5	1,9	4	215	2	2	4
6	2,7	6	395	7	-1	1
7	3,1	7	310	5	2	4
8	3,9	8	425	8	0	0
9	2,4	5	469	9	-4	16
10	4,0	9	471	10	-1	1

Підставивши необхідні дані у формулу (9.49), отримаємо:

$$\rho = 1 - \frac{6 \times 52}{10 \times (100 - 1)} = 0,68.$$

Це свідчить про прямий та помітний зв'язок між величиною прибутку й обсягом реалізованої продукції.

Істотність зв'язку можна перевірити, порівнявши критичне значення коефіцієнта рангової кореляції для рівня істотності $\alpha = 0,05$ і $n = 10$, яке становить $\rho_{0,95}(10) = 0,5515$ (табл. Е.1 додатка Е) з фактичним значенням. Оскільки фактичне ρ більше від критичного, то наявність та істотність зв'язку між ознаками доведена з імовірністю 95 %.

Слід зазначити, що для зворотних зв'язків із критичним значенням коефіцієнта рангової кореляції порівнюється абсолютна величина фактичного значення ρ .

Рангові коефіцієнти мають як переваги, так і недоліки порівняно з параметричними. Тут не треба дотримуватись певних математичних передумов відносно розподілу ознак (зокрема, передумови нормальності розподілу). Проте втрачається певна інформація про взаємозв'язок, оскільки використовуються не значення, а лише ранги ознак.

Метод вимірювання зв'язку та характеристик його щільності має ґрунтуватись на попередньому теоретичному аналізі сутності явищ, характеру взаємозв'язків, наявній інформації.

Для вивчення стохастичних (кореляційних) зв'язків використовується метод порівняння паралельних рядів двох показників, один з яких є факторним (X), а другий – результативним (Y). Основним завданням застосування цього методу є оцінювання тісноти (сили) взаємозв'язку та визначення його напрямку на основі розрахунку спеціальних коефіцієнтів.

Найпростішим показником є **коефіцієнт Фехнера (K_{ϕ})**, який розраховується за формулою:

$$K_{\phi} = \frac{C - H}{C + H}, \quad (9.50)$$

де C – число збігів знаків відхилень від середньої;

H – число напівзбігів знаків відхилень від середньої.

Якщо виконується нерівність $X \geq \bar{X}$ або $Y \geq \bar{Y}$, значенню присвоюється знак "+", у протилежному випадку – знак "-". Якщо в обох показниках знаки однакові, має місце їх збіг, а коли вони різні – незбіг. Коефіцієнт Фехнера перебуває в межах від -1 до $+1$. Якщо $|K_{\phi}| \rightarrow 0$, зв'язок між показниками слабкий, а за $|K_{\phi}| \rightarrow 1$ – зв'язок тісний. Цей коефіцієнт має додатне значення за наявності прямого зв'язку, а від'ємне – в оберненому.

Більш досконалим показником вважаються **коефіцієнти рангової кореляції r -Спірмена або Φ -Кендалла**, які використовуються для перевірки гіпотези про зв'язок двох змінних після попереднього ранжування. Для коректного обчислення обох коефіцієнтів (Спірмена та Кендалла) результати вимірювань повинні бути подані в шкалі рангів або інтервалів. Принципових відмінностей між цими критеріями не існує, але прийнято вважати, що коефіцієнт Кендалла є "змістовнішим", оскільки він більш повно та детально аналізує зв'язки між змінними, перебираючи всі можливі відповідності між парами значень. Коефіцієнт Спірмена точніше враховує саме кількісний ступінь зв'язку між змінними.

Коефіцієнт рангової кореляції Спірмена (Spearman) є непараметричним аналогом класичного коефіцієнта кореляції Пірсона, але для його розрахунку враховуються не пов'язані з розподілом показники порівнюваних змінних (середнє арифметичне та дисперсія), а ранги.

Обмеженням для застосування коефіцієнта r -Спірмена є:

за кожною змінною – не менше п'яти спостережень;

за великої кількості однакових рангів в одній або двох змінних коефіцієнт дає неточне значення.

Приклад 9.8. Нехай (x_i, y_i) , $i = 1, 2, \dots, n$ – вибірка спостережень двох змінних X і Y , обмірюваних у порядковій або кількісній шкалах. Припустимо, що серед елементів вибірки x_i і y_i ($i = 1, 2, \dots, n$) немає елементів, які співпадають. Упорядкуємо елементи x_i за зростанням, тобто напишемо варіаційний ряд $(x^{(1)}, x^{(2)}, \dots, x^{(n)})$, і кожному x_i поставимо у відповідність ранг x_i' – номер елемента x_i у варіаційному ряді. Очевидно, найменший елемент вибірки $x^{(1)}$ – ранг 1, найбільший елемент вибірки $x^{(n)}$ – ранг n . Аналогічно визначимо ранги y_i' елементів y_i , ($i = 1, 2, \dots, n$). Кожній парі (x_i, y_i) відповідають пари рангів (x_i', y_i') . Коефіцієнт рангової кореляції Спірмена обчислюється за формулою:

$$r_s = 1 - \frac{6 \sum (x_i' - y_i')^2}{n(n^2 - 1)}. \quad (9.51)$$

Отримане значення r_s називають *вибірковим коефіцієнтом рангової кореляції Спірмена*. Коефіцієнт r_s за модулем не перевищує одиницю. Більші значення вибіркового коефіцієнта r_s показують, що між випадковими величинами X і Y є залежність (у цьому випадку говорять, що коефіцієнт рангової кореляції Спірмена є значним).

Якщо $|r_s| < r(\alpha, n)$, де α – заданий рівень значущості, n – обсяг вибірки, то гіпотеза $H_0: r_s = 0$ приймається на рівні значущості 2α за альтернативною гіпотезою $H_1: r_s \neq 0$.

Коефіцієнт рангової кореляції τ -Кендалла (Kendall's tau-b) є самостійним оригінальним методом, що спирається на обчислення співвідношення пар значень двох вибірок, що мають однакові або відмінні тенденції (зростання або зниження значень). Цей коефіцієнт називають ще *коефіцієнтом конкордації*. Таким чином, основною ідеєю цього методу є те, що про напрям зв'язку можна судити, попарно порівнюючи між собою спостереження: якщо у пари спостережень зміна на X співпадає за напрямом із зміною на Y , то це свідчить про позитивний зв'язок; якщо не співпадає – то про негативний зв'язок.

Коефіцієнт рангової кореляції τ -Кендалла обчислюється за формулою:

$$\tau = 1 - \frac{4k}{n(n-1)}, \quad (9.52)$$

де k – число інверсій у ряді рангів другої змінної (y_i') за умови, що ранги першої змінної (x_i') упорядковані.

У випадку збіжних рангів для розрахунку рангових коефіцієнтів кореляції r_s і τ використовують скоректовані формули. Вибіркове значення коефіцієнта рангової кореляції Спірмена обчислюється за формулою:

$$r_s = \frac{\frac{1}{6}(n^3 - n) - \sum (x_i' - y_i')^2 - T_x - T_y}{\sqrt{\left[\frac{1}{6}(n^3 - n) - 2T_x\right] \left[\frac{1}{6}(n^3 - n) - 2T_y\right]}}, \quad (9.53)$$

де $T_x = \frac{1}{12} \sum_{t=1}^{m_x} [(n_t)^3 - n_t]$, $T_y = \frac{1}{12} \sum_{l=1}^{m_y} [(n_l)^3 - n_l]$, у яких:

m_x – число груп рангів, що збігаються, у послідовності рангів x_i' ;

n_t – число рангів, що збігаються, у групі з номером t ($t = 1, 2, \dots, m_x$);

m_y – число груп рангів, що збігаються, у послідовності рангів y_i' ;

n_l – число рангів у групі, що збігаються, з номером l , ($l = 1, 2, \dots, m_y$).

Скоректована формула для обчислення коефіцієнта рангової кореляції Кендалла має вигляд:

$$\tau' = \frac{\tau - \frac{2(U_x + U_y)}{n(n-1)}}{\sqrt{\left(1 - \frac{2U_x}{n(n-1)}\right) \left(1 - \frac{2U_y}{n(n-1)}\right)}}, \quad (9.54)$$

де τ – значення коефіцієнта рангової кореляції Кендалла, обчислене без виправлення;

$$U_x = \frac{1}{2} \sum_{t=1}^{m_x} n_t(n_t - 1); \quad (9.55)$$

$$U_x = \frac{1}{2} \sum_{l=1}^{m_x} n_l(n_l - 1). \quad (9.56)$$

Застосування коефіцієнта Кендалла є переважним, якщо в початкових даних є викиди.

Особливістю рангових коефіцієнтів кореляції є те, що максимальним за модулем ранговим коефіцієнтам (+1, -1) не обов'язково відповідають строгі прямо або зворотно пропорційні зв'язки між початковими змінними X і Y: достатній лише монотонний функціональний зв'язок між ними. Рангові коефіцієнти досягають свого максимального за модулем значення, якщо більшому значенню однієї змінної завжди відповідає більше значення іншої змінної (+1) або більшому значенню однієї змінної завжди відповідає менше значення іншої змінної, і навпаки (-1).

Статистична гіпотеза, що перевіряється, порядок ухвалення статистичного рішення і формулювання змістовного висновку ті самі, що і для випадку r-Спірмена або R-Пірсона.

Якщо статистично достовірний зв'язок не виявлений, але є підстави вважати, що зв'язок насправді є, то слід спочатку перейти від r-Спірмена до t-Кендалла (або навпаки), а потім перевірити можливі причини невірності зв'язку. Такими причинами можуть бути:

нелінійність зв'язку (для цього розглядають графік двовимірного розсіювання). Якщо зв'язок не монотонний, то розмежовують вибірку на частини, в яких зв'язок монотонний, або на контрастні групи та порівнюють їх за рівнем вираженості ознаки;

неоднорідність вибірки (розглядають графік двовимірного розсіювання). Слід спробувати розділити вибірку на частини, в яких зв'язок може мати різні напрями.

Якщо ж зв'язок статистично достовірний, то, перш ніж робити змістовний висновок, необхідно виключити можливість помилкової кореляції (за аналогією з метричними коефіцієнтами кореляції).

Важливі поняття

Бета-коефіцієнти – коефіцієнти, що показують, на скільки середньоквадратичних відхилень зміниться результативна ознака зі зміною відповідного фактора на одне значення середньоквадратичного відхилення (за постійності інших факторів, включених до рівняння регресії).

Емпіричне кореляційне відношення – коефіцієнт, за допомогою якого вимірюється тіснота кореляційного зв'язку.

Залишкова дисперсія – показник якості рівняння регресії.

Зворотний зв'язок – зв'язок, за якого напрямок зміни результативної ознаки не співпадає з напрямком зміни ознаки-фактора; тобто з підвищенням факторної ознаки знижується результативна, та навпаки.

Коефіцієнт взаємної спряженості Пірсона, коефіцієнт взаємної спряженості Чупрова, коефіцієнт Крамера – коефіцієнти тісноти зв'язку, які використовуються, коли кожна із якісних ознак складається більше ніж із двох груп.

Коефіцієнт детермінації – коефіцієнт, який показує частку варіації результативної ознаки під впливом варіації ознаки-фактора.

Коефіцієнт квадратичної спряженості χ^2 -квадрат Пірсона – коефіцієнт, який характеризує розбіжність між частотами (частками) умовного та безумовного розподілу; є критерієм перевірки істотності зв'язку.

Коефіцієнт кореляції рангів Спірмена – коефіцієнт, який використовують для вимірювання зв'язку між ознаками порядкової шкали.

Коефіцієнт рангової кореляції τ -Кендалла (Kendall's tau-b) – самостійний оригінальний метод, що спирається на обчислення співвідношення пар значень двох вибірок, які мають однакові або відмінні тенденції (зростання або зниження значень). Цей коефіцієнт називають ще коефіцієнтом конкордації.

Коефіцієнт регресії – іменована величина, яка має розмірність результативної ознаки та розглядається як ефект впливу x на y .

Коефіцієнт Фехнера – коефіцієнт, який характеризує елементарну ступінь тісноти зв'язку, який доцільно використовувати для встановлення факту наявності зв'язку, коли існує невеликий обсяг вихідної інформації

Коефіцієнти асоціації та контингенції – коефіцієнти, які використовують для визначення тісноти зв'язку двох якісних ознак, кожна із яких складається тільки із двох груп.

Коефіцієнти повної регресії – коефіцієнти парної регресії.

Коефіцієнти чистої регресії – коефіцієнти множинної регресії, які характеризують зв'язок між результативною ознакою і фактором за умови фіксованого значення інших факторів.

Кореляційний аналіз – кількісне визначення тісноти зв'язку між двома ознаками (у парному зв'язку) та між результативною і множиною факторних ознак (у багатофакторному зв'язку).

Кореляційний зв'язок – підвид стохастичного зв'язку, коли зі зміною факторної ознаки x змінюються групові середні результативної ознаки y . Тобто замість умовних розподілів порівнюються середні значення цих розподілів.

Лінійний зв'язок – зв'язок між явищами, який може бути виражений рівнянням прямої.

Лінійний коефіцієнт детермінації – квадрат лінійного коефіцієнта кореляції.

Лінійний коефіцієнт кореляції – показник тісноти зв'язку, що використовується у лінійній формі зв'язку (запропонований англійським математиком К. Пірсоном).

Множинна кореляція – вимірювання кореляційного зв'язку одночасно між двома, трьома та більше кореляційними ознаками.

Нелінійний, або криволінійний зв'язок – зв'язок між явищами, який виражається рівнянням будь-якої кривої лінії (параболи, гіперболи, степеневої, експоненціальної, показової тощо).

Непараметричні методи – методи визначення взаємозв'язків, які ґрунтуються на кількісних значеннях ознак і не потребують обчислення параметрів їх розподілів.

Параметричні методи – методи, в яких використовуються основні параметри розподілу – середні величини та дисперсії.

Парні коефіцієнти кореляції – коефіцієнти, що вимірюють тісноту зв'язку між двома із сукупності змінних (без урахування їх взаємодії з іншими змінними).

Прямий зв'язок – зв'язок, за якого напрямок зміни результативної ознаки співпадає з напрямком зміни ознаки-фактора; тобто з підвищенням факторної ознаки підвищується результативна, та навпаки.

Ранг – порядковий номер значень ознаки, розташованих у порядку зростання або спадання їх величин.

Ранжування – процедура впорядкування об'єктів вивчення, яка виконується на основі переваги.

Регресійний аналіз – визначення аналітичного вираження зв'язку, встановлення ступеня впливу незалежних змінних на залежну та визначення розрахункових значень залежної змінної (функції регресії).

Стохастичний зв'язок – зв'язок, за якого кожному значенню ознаки x відповідає певна множина значень ознаки y , які варіюють і утворюють ряд розподілу (умовний).

Сукупний коефіцієнт множинної детермінації – показник, що вимірює частку варіації результативної ознаки, яка пояснюється впливом факторів, що включені в рівняння множинної регресії.

Сукупний коефіцієнт множинної кореляції – це показник тісноти зв'язку між результативною та двома або більше факторними ознаками.

Таблиця взаємної спряженості – таблиця, яка містить зведену числову характеристику розглядуваної сукупності за двома та більше атрибутивними ознаками або комбінацією кількісних та атрибутивних ознак.

Теоретична лінія регресії – неперервна лінія, що описується функцією $y = f(x)$, яка називається рівнянням регресії.

Факторна ознака – ознака, що характеризує причину.

Функціональний зв'язок – зв'язок, за якого кожному значенню факторної ознаки відповідає одне чітко визначене значення результативної ознаки.

Часткові коефіцієнти кореляції – коефіцієнти, що характеризують ступінь і вплив одного із аргументів на функцію за умови, що інші незалежні змінні фіксуються на постійному рівні.

Типові завдання

Завдання 9.1. Виявіть наявність стохастичного зв'язку, скориставшись комбінаційним розподілом елементів статистичної сукупності.

Розв'язання.

Сукупність шахт регіону поділено на групи за двома ознаками: x – глибина розробка вугільних шарів; y – фондомісткість видобутку вугілля. Кожна група за глибиною розробки шару характеризується своїм особливим розподілом шахт за фондомісткістю видобутку вугілля. Це умовні розподіли. Порівняння умовних розподілів указує на тенденцію підвищення фондомісткості зі зростанням глибини розробки шарів. Звичайно, для кожної окремої шахти така залежність може не виявитись через вплив інших факторів. Певні межі варіації фондомісткості характерні для кожної групи. Так, на шахтах, де глибина розробки шарів 500 – 700 м, фондомісткість коливається в межах від 18 до 26 грн за тонну. Проте середній рівень фондомісткості в цій групі вищий порівняно з попередньою групою (300 – 500 м) і нижчий порівняно з наступною (700 і більше) (табл. 9.9).

Комбінаційний розподіл шахт за глибиною розробки шарів і фондомісткістю вугілля

Глибина розробки шару, м	Кількість шахт з рівнем фондомісткості, грн/т						Середній рівень фондомісткості, грн/т
	До 20	20 – 22	22 – 24	24 – 26	26 і більше	Разом	
До 300	9	7	1			17	20,0
300 – 500		8	27	5		40	22,9
500 – 700			6	15	4	25	24,8
700 і більше				8	10	18	26,1
За сукупністю у цілому	9	15	34	28	14	100	23,5

Результати розрахунків показника середньої фондомісткості за кожною групою такі:

$$\bar{y}_1 = \frac{19 \times 9 + 21 \times 7 + 23 \times 1}{17} = 20;$$

$$\bar{y}_2 = \frac{21 \times 8 + 23 \times 27 + 25 \times 5}{40} = 22,9;$$

$$\bar{y}_3 = \frac{23 \times 6 + 25 \times 15 + 27 \times 4}{25} = 24,8;$$

$$\bar{y}_4 = \frac{25 \times 8 + 27 \times 10}{18} = 26,1.$$

Середні рівні фондомісткості видобутку вугілля наведено в останній графі табл. 9.9. Зростання групових середніх від групи до групи свідчить про наявність кореляційного зв'язку між глибиною розробки шару та фондомісткістю вугілля. Отже, кореляційний зв'язок, як і стохастичний, – це властивість сукупності в цілому, а не окремих її елементів.

Отже, можна не лише стверджувати, що існує кореляційний зв'язок між факторною x і результативною у ознаками, а і визначати, як у середньому змінюється y зі зміною x на одиницю. *Ефекти впливу* x на y визначаються

відношенням приростів середніх групових величин $D_y:D_x$. Наприклад, у другій групі порівняно з першою глибина розробки вугільного шару більша на 200 м, а фондомісткість видобутку вугілля на $22,9 - 20,0 = 2,9$ грн/т. Звідси:

$$\frac{\Delta y}{\Delta x} = \frac{2,9}{200} = 0,0145.$$

Отже, зі зростанням глибини розробки шару на 100 м фондомісткість зростає в середньому на 1,45 грн/т. Аналогічно розраховані ефекти впливу глибини розробки шару на фондомісткість вугілля у третій групі становлять 0,95, у четвертій – 0,65 грн на тонну вугілля.

Завдання 9.2. Необхідно побудувати рівняння лінійного регресійного зв'язку між урожайністю зернових і кількістю внесених добрив (у центнерах без змін поживної речовини).

Розв'язання.

Розглянемо порядок обчислення параметрів лінійної регресії на прикладі зв'язку між урожайністю зернових і кількістю внесених добрив (у центнерах діючої поживної речовини – д. р.). Значення взаємозв'язаних ознак і необхідні для розрахунку параметрів величини наведено в табл. 9.10.

Таблиця 9.10

Розрахунок параметрів лінійної регресії, теоретичних значень і залишкових величин

Номер господарства	Кількість внесених добрив X, д. р	Урожайність зернових Y, ц/га	XY	X ²	\hat{Y}	Y – \hat{Y}	(Y – \hat{Y}) ²
1	1,1	23	25,3	1,21	24	-1	1
2	1,4	25	35,0	1,96	27	-2	4
3	1,2	26	31,2	1,44	25	1	1
4	2,0	33	66,0	4,00	33	0	0
5	1,5	27	40,5	2,25	28	-1	1
6	1,3	2,8	36,4	1,69	26	2	4
7	1,8	30	54,0	3,24	31	-1	1
8	1,7	32	54,4	2,89	30	2	4
Разом	12,0	224	342,8	18,68	224	×	16

Проміжні результати, отримані із розрахункової таблиці (див. табл. 9.10):

$$\begin{aligned}\sum X &= 12; \quad \sum Y = 224; \quad \sum XY = 342,8; \quad \sum X^2 = 18,68; \\ \bar{x} &= 12 : 8 = 1,5; \quad \bar{Y} = 224 : 8 = 28.\end{aligned}$$

Користуючись цими величинами, визначаємо:

$$\begin{aligned}b &= \frac{8 \times 342,8 - 12 \times 224}{8 \times 18,68 - 12 \times 12} = \frac{54,4}{5,44} = 10,0 \text{ (ц/га);} \\ x &= 28 - 10,0 \times 1,5 = 13,0.\end{aligned}$$

Отже, рівняння регресії має вигляд $Y = 13,0 + 10,0x$, тобто кожний центнер внесених добрив (у перерахунку на діючу поживну речовину) дає приріст урожайності в середньому 10 ц/га. Якщо добрива зовсім не вносити ($x = 0$), то врожайність зернових не перевищить 13,0 ц/га.

Рівняння регресії відбиває закон зв'язку між x і y не для окремих елементів сукупності, а для сукупності в цілому; закон, який абстрагує вплив інших факторів, виходить з принципу "за інших однакових умов". За цих умов очікувана врожайність зернових із внесенням добрив у обсязі 1,1 ц д. р. на 1 га становить $Y = 13 + 10 \times 1,1 = 24$ (ц/га). Для інших значень факторної ознаки x теоретичні рівні врожайності наведено в табл. 9.10. Вплив інших (окрім x -факторів) зумовлює відхилення емпіричних значень y від теоретичних у той чи інший бік. Відхилення $Y - \bar{Y}$ називають **залишками** та позначають символом e . Залишки, як правило, менші за відхилення від середньої, тобто $(y - Y) \leq (y - \bar{y})$.

$$\text{У нашому прикладі: } \sum_1^n (Y - \bar{Y})^2 = 84, \quad \sum_1^n (Y - \hat{Y})^2 = 16.$$

$$\begin{aligned}\text{Відповідно, загальна дисперсія врожайності: } \sigma_y^2 &= \frac{1}{n} \sum_1^n (y - \bar{y})^2 = \\ &= \frac{84}{8} = 10,5,\end{aligned}$$

$$\text{залишкова дисперсія: } \sigma_e^2 = \frac{1}{n} \sum_1^n (y - Y)^2 = \frac{16}{8} = 2.$$

У невеликих за обсягом сукупностях коефіцієнт регресії схильний до випадкових коливань. Тому слід перевірити його істотність. Коли зв'язок

лінійний, істотність коефіцієнта регресії перевіряють за допомогою t-критерію (Стюдента), статистична характеристика якого для гіпотези $H_0: b = 0$ визначається відношенням коефіцієнта регресії b до власної стандартної похибки μ_b , тобто $t = b / \mu_b$.

Стандартна похибка коефіцієнта регресії залежить від варіації факторної ознаки σ_x^2 , залишкової дисперсії σ_e^2 і числа ступенів свободи $k = n - m$, де m – кількість параметрів рівняння регресії:

$$\mu_b = \sqrt{\frac{\sigma_e^2}{\sigma_x^2(n - m)}}.$$

Для лінійної функції $m = 2$. За даними табл. 9.10 маємо:

$$\sigma_x^2 = \frac{18,68}{8} - 1,5^2 = 0,085, \quad \sigma_e^2 = 2.$$

Звідси $\mu_b = \sqrt{\frac{2}{0,085(8 - 2)}} \approx 2,0$ (ц/га), а $t = \frac{b}{\mu_b} = \frac{10}{2} = 5$, що перевищує

критичне значення двостороннього t-критерію $t_{0,95(6)} = 2,45$. Гіпотеза про випадковий характер коефіцієнта регресії відхиляється, а отже, з імовірністю 0,95 вплив кількості внесених добрив на врожайність зернових визнається істотним.

Для коефіцієнта регресії, як і для будь-якої іншої випадкової величини, визначаються довірчі межі $b \pm t\mu_b$. У нашому прикладі довірчі межі коефіцієнта регресії з імовірністю 0,95 ($t = 2,45$) становлять $10,0 \pm 2,45 \times 2,0$.

Важливою характеристикою регресійної моделі є відносний ефект впливу фактора x на результат y – **коефіцієнт еластичності**:

$$\gamma = b \frac{\bar{x}}{\bar{y}}.$$

Він показує, на скільки відсотків у середньому змінюється результат y зі зміною фактора x на 1%. У нашому прикладі $\gamma = 10,0 \frac{1,5}{28} = 0,8035$, тобто збільшення кількості внесених добрив на 1 % спричинює приріст урожайності зернових у середньому на 0,8 %.

Завдання 9.3. На підставі даних завдання 2 оцініть щільність зв'язку між кількістю внесених добрив і врожайністю зернових.

Розв'язання.

На практиці застосовують різні модифікації формули коефіцієнта кореляції. Для оцінювання щільності зв'язку між кількістю внесених добрив і врожайністю зернових скористаємося однією з модифікацій зазначеної формули:

$$r = \frac{\sum_{i=1}^n xy - n\bar{x}\bar{y}}{\sqrt{n\sigma_x^2\sigma_y^2}}.$$

За даними табл. 9.10 $\sum_{i=1}^n XY = 342,8$; $\bar{X} = 1,5$; $\sigma_x^2 = 0,085$; $\bar{y} = 28$;
 $\sigma_y^2 = 10,5$.

Згідно із цими значеннями коефіцієнт кореляції становить 0,900, що свідчить про вагомий вплив кількості внесених добрив на врожайність зернових:

$$r = \frac{342,8 - 8 \times 1,5 \times 28}{\sqrt{0,085 \times 10,5}} = 0,900.$$

Коефіцієнт кореляції, оцінюючи щільність зв'язку, вказує також на його напрям: коли зв'язок прямий, r – величина додатна, а коли він зворотний – від'ємна. Знаки коефіцієнтів кореляції і регресії однакові, величини їх взаємозв'язані функціонально:

$$r = b \frac{\sigma_x}{\sigma_y}; \quad b = r \frac{\sigma_y}{\sigma_x}.$$

Завдяки цьому один коефіцієнт можна обчислити, знаючи інший. Наприклад:

$$r = 10,0 \sqrt{\frac{0,085}{10,50}} = 0,900.$$

Відхилення індивідуального значення ознаки y від середньої ($Y - \bar{Y}$) можна розкласти на дві складові. У регресійному аналізі це відхилення від лінії регресії ($Y - \hat{Y}$) і відхилення лінії регресії від середньої ($\hat{Y} - \bar{Y}$).

Відхилення ($\hat{Y} - \bar{Y}$) є наслідком дії фактора X , відхилення $Y - \hat{Y}$ – наслідком дії інших факторів. Взаємозв'язок факторної та залишкової варіацій описується правилом декомпозиції варіації:

$$\sigma_y^2 = \delta_Y^2 + \sigma_e^2,$$

де $\sigma_y^2 = \frac{1}{n} \sum_1^n (Y - \bar{Y})^2$ – загальна дисперсія ознаки y ;

$\delta_Y^2 = \frac{1}{n} \sum_1^n (Y - \bar{Y})^2 = \frac{1}{n} (a \sum x + b \sum XY) - \bar{Y}^2$ – факторна дисперсія;

$\sigma_e^2 = \frac{1}{n} \sum_1^n (Y - \hat{Y})^2$ – залишкова дисперсія.

Очевидно, значення факторної дисперсії δ_Y^2 буде тим більшим, чим сильніший вплив фактора X на Y . Відношення факторної дисперсії до загальної розглядається як міра щільності кореляційного зв'язку; його називають **коефіцієнтом детермінації**:

$$R^2 = \frac{\delta_Y^2}{\sigma_y^2}.$$

Якщо за даними табл. 9.10 $\sigma_y^2 = 10,5$, $\sigma_e^2 = 2,0$, то $\delta_Y^2 = 10,5 - 2,0 = 8,5$.

Аналогічний результат дають такі обчислення:

$$\sigma_y^2 = \frac{1}{8} (13 \times 224 + 10 \times 342,8) - 28^2 = 8,5.$$

Коефіцієнт детермінації становить $R^2 = \frac{8,5}{10,5} = 0,81$, тобто 81 % варіації врожайності зернових залежить від варіації кількості внесених добрив, а 19 % припадає на інші фактори.

Корінь квадратний з коефіцієнта детермінації називають **індексом кореляції** R . Коли зв'язок лінійний, $R = |r|$, що підтверджують обчислення: $R = \sqrt{R^2} = \sqrt{0,81} = 0,90$. Тому за відомим лінійним коефіцієнтом кореляції r можна визначити внесок ознаки x у варіацію ознаки y .

Завдання 9.4. Необхідно визначити коефіцієнт рангової кореляції за даними експертних оцінок ефективності економіки та ступеня політичного ризику для семи країн з трансформаційною економікою (табл. 9.11).

Таблиця 9.11

Розрахунок коефіцієнта рангової кореляції

№ п/п	Експертні оцінки, балів		Ранги		$d_j = R_{x_j} - R_{y_j}$	d_j^2
	Ефективність економіки (max = 10)	Ступінь політичного ризику (max = 100)	R_{x_j}	R_{y_j}		
1	6,6	64,5	1	7	-6	36
2	5,8	57,8	2	6	-4	16
3	2,9	23,6	6	1	5	25
4	3,4	36,2	5	4	1	1
5	4,5	45,3	3	5	-2	4
6	2,7	28,4	7	2	5	25
7	4,2	32,7	4	3	1	1
Разом	×	×	×	×	×	108

Розв'язання.

Оскільки експертні оцінки подані балами, необхідно провести ранжування країн. За оцінками ефективності економіки країні з найбільшим балом надається ранг 1, з найменшим – ранг $n = 7$. За оцінками ступеня політичного ризику, навпаки, ранг 1 надається країні з найменшим ризиком, а ранг 7 – країні з найбільшим ризиком.

Сума квадратів відхилень рангів $\sum_1^n d_j^2 = 108$, а коефіцієнт рангової кореляції:

$$\rho = 1 - \frac{6 \times 108}{7(49 - 1)} = 1 - \frac{648}{336} = -0,928.$$

Значення коефіцієнта рангової кореляції свідчить про наявність зворотного та досить високого рівня зв'язку між ефективністю економіки та ступенем політичного ризику. Критичне значення коефіцієнта рангової кореляції (табл. 9.12) для рівня істотності $\alpha = 0,05$ і $n = 7$; $\rho_{0,95}(7) = 0,71$. Отже, з імовірністю 0,95 істотність зв'язку доведено.

Таблиця 9.12

**Критичні значення коефіцієнта рангової кореляції
Спірмена за $\alpha = 0,05$**

Обсяг вибірки n	5	6	7	8	9	10	11	12
$\rho_{0,95}(n)$	0,90	0,83	0,71	0,64	0,60	0,56	0,53	0,50

Якщо два та більше елементів сукупності мають однакові значення ознаки, їм надається середній ранг. Нехай, наприклад, друге за розміром значення ознаки мають три елементи сукупності (№ 2, 3, 4), тоді всім їм надається ранг $\frac{1}{3}(2 + 3 + 4) = 3$, а щільність зв'язку можна оцінити за формулою лінійного коефіцієнта кореляції.

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Опанування навиків проведення кореляційного, регресійного та дисперсійного аналізу у MS Excel" подані в роботі [67]. Лабораторна робота націлена на отримання практичних навичок у проведенні кореляційного, регресійного та дисперсійного аналізу за допомогою пакету MS Excel.

Запитання до самодіагностики

1. Дайте поняття функціонального та кореляційного зав'язків. Як проявляється кореляційний зв'язок?
2. Які задачі розв'язуються за допомогою кореляційно-регресійного аналізу?
3. Дайте визначення множинної кореляції.
4. Що характеризують коефіцієнти регресії в рівнянні множинної регресії?
5. Чим відрізняється коефіцієнт еластичності від коефіцієнта регресії?

6. У чому полягає сутність коефіцієнтів еластичності і β -коефіцієнта?
7. За допомогою яких коефіцієнтів проводиться кількісне оцінювання зв'язку соціальних явищ?
8. Що таке непараметричні показники зв'язку?
9. Які ви знаєте рангові коефіцієнти зв'язку?
10. Чим зумовлена необхідність перевірки істотності зв'язку? Як формується нульова гіпотеза? За яких умов вона приймається, а за яких – відхиляється?

Питання для критичного переосмислення (есе)

1. Кореляційно-регресійний аналіз інвестиційної привабливості підприємств.
2. Кореляційний аналіз фінансів підприємства.
3. Кореляційні дослідження в психології.
4. Кореляційний аналіз у соціології. Двовимірний розподіл двох ознак: наочне зображення, коефіцієнти кореляції.
5. Кореляційно-регресійний аналіз у дослідженні соціальних явищ.

10. Аналіз інтенсивності динаміки

Основні питання:

- 10.1. Поняття про ряди динаміки та їх види.
- 10.2. Аналітичні показники рядів динаміки.

10.1. Поняття про ряди динаміки та їх види

Суспільні явища безперервно змінюються як у просторі, так і в часі. Вивчення поступового розвитку та змін суспільних явищ – одне з основних завдань статистики. Вирішується воно на основі аналізу рядів динаміки (інколи їх називають *динамічними*, *хронологічними* або *часовими* рядами).

Ряд динаміки у статистиці – це послідовність значень статистичних показників, які характеризують зміну в часі величини суспільного явища чи процесу.

Кожний ряд динаміки складається з двох **елементів** [31; 39; 68]: **періоду часу (t)**, за який (або станом на який) приводиться числове значення;

рівня ряду (Y) – конкретного значення того чи іншого показника.

Рівні ряду мають певні *особливості*:

рівень майбутнього часу залежить від рівня, що був досягнутий у попередній час;

чим більшим є інтервал часу поміж подіями, тим більше відрізняється їх кількісний та якісний стан.

Наприклад, чисельність випускників ЗВО зі спеціальності "Економіка" освітньо-професійної програми "Бізнес-аналітика та менеджмент" першого (бакалаврського) рівня змінилась у 2017 році порівняно з 2010 роком. Але змінилась не тільки чисельність, а і кількість освітньо-професійних програм, які охоплює спеціальність "Економіка", якісний рівень підготовки студентів, вимоги до практичної підготовки спеціалістів.

Отже, побудова й аналіз рядів динаміки дають змогу виявити закономірності розвитку явищ суспільного життя та його особливості.

Для правильного аналізу рядів динаміки необхідно знати їх *види*, які виокремлюються групуванням елементів ряду за різними ознаками.

Класифікація рядів динаміки здійснюється за такими *ознаками*.

Статистична природа показника [21; 68; 71; 87]:

первинні (чисельність населення міста);

похідні (чисельність населення міста жіночої статі вікової групи 15 – 49 років).

Характер відображення динаміки:

моментні, що фіксують стан явища на певний момент часу (курс євро на кожен день, товарні запаси на початок доби, чисельність населення згідно з переписом),

інтервальні, що характеризують явища за певний період часу (товарообіг крамниці за квартал, доходи населення за місяць, виробництво продукції за роками).

Рівень інтервального ряду динаміки залежить від тривалості періоду часу, який він характеризує. Рівні моментних рядів динаміки не залежать від проміжку часу між датами.

Як первинні, так і похідні показники, обчислені на основі інтервальних рядів, на відміну від моментних, залежать від тривалості інтервалу часу (середньодобове чи середньорічне виробництво електроенергії на душу населення).

Якщо рівні інтервального ряду подані у вигляді абсолютної величини, то є можливість переходити від ряду динаміки з малими часовими

інтервалами до більш крупних проміжків часу, тобто їх рівні є сенс підсумовувати.

Шляхом послідовного додавання рівнів інтервального ряду отримують накопичені підсумки за певний період (нарахування заробітної плати у фірмі, обсяг відпрацьованих людино-годин). Підсумок рівнів моментного ряду динаміки як такий не можливий, адже здобуті величини не мають економічного значення (залишки товарів у торгівельній мережі змінюються щоденно, тому підсумок їх за робочий тиждень не має сенсу).

Засіб виразу рівнів ряду:

ряди абсолютних (розмір кредитних ресурсів банку, прибуток фірми);

середніх (середня ціна реалізації, середня заробітна плата);

відносних величин (зміна доходів домогосподарств, зміна структури капіталу).

Повнота часу:

повні, коли дати чи періоди слідуєть один за одним із рівними інтервалами;

неповні, коли у послідовності часу рівні інтервали не витримуються.

Кількість показників:

одновимірні, що характеризують зміну одного показника (видобуток золота);

багатовимірні, що характеризують зміну двох і більше показників.

Багатовимірні ряди динаміки у свою чергу розподіляють на:

паралельні, що відображають динаміку або того самого показника щодо різних об'єктів (зміна прибутку за підприємствами певної галузі), або різних показників щодо того самого об'єкта (видобуток вугілля, нафти та газу в регіоні);

ряди взаємозв'язаних показників.

Зв'язок між показниками багатовимірною динамічного ряду може бути *функціональним* (зміна чисельності населення, в тому числі чоловіків і жінок) або *кореляційним* (вплив зміни суми капіталу комерційного банку на величину його прибутку) [20; 68].

Аналіз динаміки суспільних явищ, як правило, здійснюється на підставі багатовимірних рядів динаміки. Вони дають змогу оцінити інтенсивність і описати характер розвитку всіх складових, провести порівняльний аналіз динаміки двох і більше явищ, оцінити вплив інтенсивності розвитку одних явищ на інші, побудувати науково обґрунтовані прогнози.

Динамічні ряди економічних показників, на відміну від рядів динаміки математичної статистики, в більшості своїй нестационарні, їм притаманна тенденція, яка відображує зміни економіки. Поряд із динамічністю економічні процеси мають таку властивість, як інерційність: зберігається механізм формування явищ і характер розвитку (темпи, напрям, коливання). Діалектична єдність мінливості та сталості, динамічності й інерційності формують характер динаміки, надаючи принципову можливість статистичного прогнозування соціально-економічного розвитку.

Основними **правилами побудови рядів динаміки** є такі [4; 5; 26; 36; 68]:

1) усі показники динамічного ряду повинні бути **вірогідними** та **науково обґрунтованими**;

2) рівні динамічного ряду повинні бути **зіставлені** поміж собою за: *територією* (за зміною кордонів країни, меж областей чи районів); *методикою розрахунку показників* (витрати часу на виконання певної роботи одним працівником і одним робітником за певний проміжок часу); *періодом, або моментом спостереження* (явища із сезонним характером рівнів – кількість реалізованих торговельною мережею сноубордів взимку збільшується – не коректно об'єднувати в один ряд динаміки за критичним моментом реєстрації, тому що його рівні віднесені до різних дат реєстрації);

об'єктом спостереження, тобто всі рівні динамічного ряду повинні стосуватись до того самого об'єкта спостереження;

одинацями спостереження (у дослідженні чисельності виробничих підприємств міста одиницею спостереження виступає підприємство. Тому необхідно чітко визначити, які підприємства належать до виробничих, а які ні);

ступенем охоплення одиниць сукупності (згідно з КВЕД будівництво віднесене до сфери послуг, а згідно з балансом народного господарства ця галузь належить до сфери матеріального виробництва);

структурою сукупності (показники народжуваності та смертності населення пов'язані з віковою структурою населення за різні роки);

одинацями вимірювання;

3) довжина динамічного ряду визначається **періодом вивчення попередньої історії** розвитку явища (чи процесу).

Так, для побудови прогнозного значення певної ознаки необхідно знати передісторію її розвитку. У цьому випадку не можна розглядати

весь динамічний ряд як єдиний період передісторії. Наприклад, для прогнозу величини ВВП у країні на 2019 р. попередня історія дослідження зміни цього показника визначає період 1992 – 2018 рр.; проте на той час грошовими одиницями в Україні були рублі, купони, купоно-карбованці, карбованці, гривні;

4) забезпечення зіставлення рівнів в ті самі проміжки часу за умов одночасного використання в аналізі декількох динамічних рядів.

Наприклад, для характеристики ефективності використання основних виробничих фондів проводиться порівняння за певний період часу прибутку підприємства (інтервальний ряд) і вартості основних виробничих фондів на кінець року (моментний ряд). Порівняння показників є можливим за умов перерахунку абсолютних рівнів моментного динамічного ряду в середні (тобто необхідно розрахувати середньорічну вартість основних виробничих фондів і використовувати зміни цього показника в подальших порівняннях).

Порівнянність даних – необхідна умова аналізу ряду динаміки.

Якщо основною причиною непорівнянності даних є [40; 68; 70]:

непорівнянність одиниць вимірювання чи зміна цін для вартісних показників, то *способом зведення даних до порівнянного вигляду є прямий перерахунок даних за допомогою коефіцієнтів переведення, індексів, курсів валют;*

зміна структури сукупності, то *способом зведення даних до порівнянного вигляду є використання стандартизованої структури;*

зміна кола об'єктів дослідження, території, методології обліку та розрахунку, то *способом зведення даних до порівнянного вигляду є застосування спеціальних прийомів зімкнення розірваного ряду динаміки – "статистичних ключів" (два ряди об'єднуються на основі співвідношення рівнів перехідного періоду або рівень перехідного періоду є базою для порівняння, щодо якої формується зімкнутий ряд відсоткових відношень).*

Завдання статистики полягає у тому, щоб шляхом аналізу рядів динаміки визначити інтенсивність розвитку соціально-економічних явищ і виявити й описати тенденції цього розвитку; оцінити структурні зрушення, сталість і коливання рядів; виявити фактори економічного зростання.

Для вивчення динаміки соціально-економічних явищ використовують певні статистичні характеристики.

10.2. Аналітичні показники рядів динаміки

Для оцінювання напряму та розміру змін рівнів ряду в часі статистика використовує взаємозв'язані характеристики. Серед них: абсолютний приріст, темп зростання, темп приросту й абсолютне значення 1 % приросту, показник прискорення (уповільнення). Порядок розрахунку базисних і ланцюгових характеристик динаміки схематизований на рис. 10.1.



Рис. 10.1. Аналітичні характеристики ряду динаміки

Розрахунок характеристик динаміки ґрунтується на зіставленні рівнів ряду. Рівень ряду динаміки, що порівнюється, називають *поточним* (Y_1), а рівень, з яким проводиться порівняння, – *базисним* (Y_0). Базою для зіставлення може бути або попередній рівень, або початковий. У першому випадку база порівняння змінна, в другому – постійна. База порівняння визначається залежно від мети дослідження. Характеристики динаміки, що обчислені зіставленням суміжних рівней, називають **ланцюговими**, а що обчислені зіставленням із постійною базою порівняння – **базисними**.

Порівняння рівнів ряду динаміки між собою складає методологічну основу оцінювання інтенсивності динаміки. Ланцюгові характеристики оцінюють інтенсивність змін в окремі проміжки часу в межах досліджуваного періоду, а базисні характеристики вказують на кінцевий результат змін за весь період у цілому.

Абсолютний приріст (ΔY) відображає абсолютну швидкість змінювання рівнів ряду за певний інтервал часу. Він обчислюється як різниця рівнів ряду, знаки + або – показують напрям динаміки.

Ланцюгові та базисні абсолютні прирости адитивно зв'язані: сума ланцюгових абсолютних приростів дорівнює загальному базисному приросту за весь період. Тобто:

$$\sum_1^n \Delta Y = \sum_1^n (Y_i - Y_{i-1}) = Y_n - Y_0. \quad (10.1)$$

Абсолютний приріст залежно від статистичної природи показника може бути відносною величиною. Наприклад, на певній фірмі частка коштів на забруднення водоймищ у загальному обсязі коштів на охорону довкілля у 2016 р. становила 82 %, у 2018 р. – 79 %, тобто зменшилась на три відсотки.

Інтенсивність зміни рівнів ряду оцінюється відносною величиною – **темпом зростання (T_p)**, який є кратним відношенням рівнів у формі коефіцієнта чи відсотка.

Темп зростання показує, у скільки разів рівень розглядуваного періоду більший або менший будь-якого рівня, прийнятого за базу порівняння.

Між ланцюговими та базисними темпами зростання існує мультиплікативний зв'язок: добуток ланцюгових темпів зростання дорівнює базисному темпу зростання:

$$\prod_1^n T_p = \prod_1^n \frac{Y_n}{Y_{n-1}} = \frac{Y_n}{Y_0}. \quad (10.2)$$

Співвідношення абсолютного приросту та базового рівня є *вимірником відносної швидкості зростання*. Алгебраїчні перетворення цього відношення дають відхилення темпу зростання від бази порівняння, яка становить 100 %.

Відносну швидкість зростання називають **темпом приросту** (ΔT_p), який (на відміну від темпу зростання) завжди виражають у відсотках.

$$\Delta T_p = \frac{\Delta Y}{Y_{i-1}} \times 100 = 100(T_p - 1). \quad (10.3)$$

Ланцюгові темпи приросту не мають таких властивостей, як адитивність чи мультиплікативність. З базисними темпами приросту вони співвідносяться через темпи зростання.

Абсолютне значення 1 % приросту (A) є часткою від ділення абсолютного приросту на відповідний показник темпу приросту. Алгебраїчно ця величина виражена сотою частиною попереднього рівня (прийнятого за базу порівняння) ряду динаміки:

$$A = \frac{\Delta Y}{\Delta T_p} = 0,01 \times Y_{i-1}. \quad (10.4)$$

Для базисних темпів приросту значення (A) однакові. Вага відсотка приросту залежить від базисного рівня.

Порівняльний аналіз інтенсивності динаміки базується на співвідношенні односпрямованих характеристик швидкості розвитку процесів. Якщо швидкість розвитку явища (процесу) в межах аналізованого періоду неоднакова, то зіставленням однойменних характеристик швидкості визначають прискорення або уповільнення динаміки.

Для оцінювання змін швидкості динаміки на різних етапах розвитку певного процесу (явища) застосовують показник абсолютного та відносного прискорення (уповільнення).

Абсолютні прирости вказують на швидкість зміни рівнів ряду в одиницю часу, а їх систематичне зростання констатує, що ряд розвивається з *прискоренням*. За умов систематичного зниження різниця абсолютних ланцюгових приростів характеризує *абсолютне уповільнення ряду*. **Абсолютним прискоренням (уповільненням)** у статистиці називають різницю між наступним та попереднім абсолютними приростами. Прискорення (уповільнення) показує, наскільки розглядувана швидкість більше (менше) попередньої, тобто абсолютне прискорення (уповільнення) є швидкістю зміни швидкості.

$$\Delta^* = \Delta Y_i - \Delta Y_{i-1}. \quad (10.5)$$

Якщо інтервали часу щодо показників ряду динаміки однакові, то можна зіставляти базисні характеристики швидкості; якщо неоднакові – слід користуватись середніми швидкостями. Абсолютне прискорення (зростання) характеризується додатною величиною ($\Delta^* > 0$), уповільнення – від'ємною ($\Delta^* < 0$) [12; 18; 27; 41; 68].

Зіставляючи темпи зростання (приросту), дістають *коефіцієнт прискорення (уповільнення) відносної швидкості розвитку*. Для наочності та зручності їх тлумачення дільником виступає більший за значенням темп зростання (приросту). Причому обидва мають бути одного напрямку.

Якщо систематично зростають (знижуються) ланцюгові темпи росту, то ряд розвивається з *відносним прискоренням (уповільненням)*:

$$\Delta\% = \frac{\Delta Y_i}{\Delta Y_{i-1}}. \quad (10.6)$$

Отримана величина виражається у відсоткових пунктах.

Пункти зростання (відсоткові пункти) можна підсумовувати: сума пунктів зростання дорівнює загальному темпу приросту за весь період.

Таким чином, **відносне прискорення (уповільнення)** є темпом приросту абсолютного приросту.

З метою порівняння відносної швидкості двох паралельних процесів чи явищ застосовують **коефіцієнт випередження ($K_{\text{вип}}$)**.

У порівняльному аналізі $K_{\text{вип}}$ визначається співвідношенням темпів зростання одного показника за різними об'єктами або різних показників за одним об'єктом.

Для декількох динамічних рядів, що відображають різні економічні явища за однакові відрізки часу, $K_{\text{вип}}$ буде дорівнювати:

$$K_{\text{вип}} = \frac{T_p \text{ більший за значенням для певного показника у відповідний період часу}}{T_p \text{ менший для показника, що порівнюється за цей же період часу}}. \quad (10.7)$$

Розрахунок цього показника дозволяє наочно встановити, для якого динамічного ряду інтенсивність змін рівнів є найбільшою.

Зіставлення темпів приросту взаємопов'язаних показників використовують для визначення **коефіцієнта еластичності ($K_{\text{ел.}}$)**:

$$K_{\text{ел.}} = \frac{\Delta T_{\text{рх}}}{\Delta T_{\text{ру}}}. \quad (10.8)$$

Коефіцієнт еластичності показує, на скільки відсотків змінюється результативна ознака Y за умов зміни факторної ознаки x на 1 %.

Наприклад, якщо ціна реалізації товару зросла на 13 %, а попит зменшено на 7 %, то цінова еластичність попиту на товар становитиме:

$$K_{\text{ел.}} = \frac{-7}{13} = -0,54.$$

Ступінь еластичності попиту за ціною (пряма еластичність попиту) показує, що зі зростанням ціни реалізації товару на 1 % попит на цей товар зменшується на 0,54 %. Цей показник в більшості своїй є від'ємним тому, що згідно із законом попиту ціна та попит змінюються в протилежному напрямі.

У дослідженні рядів динаміки, рівні яких змінюються і варіюють, постає потреба в обчисленні сталої, типової для розглядуваного періоду характеристики. Такою характеристикою виступає **середній рівень**.

Методи обчислення середніх рівнів рядів динаміки залежать від статистичної структури показника.

Середніми за рядами динаміки є [68]:

середній рівень ряду;

середній абсолютний приріст;

середній темп зростання;

середній темп приросту.

В економічній практиці використання середніх показників іноді є необхідною умовою для проведення аналітичного оцінювання. Наприклад, виробництво сільськогосподарської продукції залежить від погодних умов певного року, тому порівняння показників, що характеризують зміну виробництва цієї продукції за рік, стає недоцільним. Правильним буде порівняння середньорічних рівнів, середньорічних абсолютних приростів чи темпів зростання за певні проміжки часу.

До середньорічних показників доводиться вдаватись за неможливості зіставлення абсолютних даних. Наприклад, щоб визначити виробництво продукції на душу населення, необхідно абсолютний розмір виробництва поділити на чисельність населення, що для аналізованого проміжку часу не є постійною величиною. Так, на початок 2018 р. чисельність мешканців населеного пункту в області складала 257,8 тис. осіб, а на кінець – 260,0 тис. осіб. Щоб визначити розмір виробництва на душу

населення в 2018 р., треба віднести загальне виробництво до середньої чисельності мешканців населеного пункту, яка складе: 258,9 тис. осіб.

Відповідно до теорії середніх величин, їхнє обчислення необхідно проводити за однорідними у певному відношенні групами. Для явищ, які розвиваються в часі, це означає, що середня рівнів ряду повинна характеризувати в певній мірі час з однаковими умовами розвитку. Загальна середня за весь період дослідження може бути доповнена середніми за окремі проміжки часу цього періоду.

Порядок розрахунку середнього рівня для інтервального та моментного ряду динаміки відрізняється.

Середній рівень ряду (\bar{Y}):

1) для *інтервального ряду* динаміки:

а) коли рівні ряду однаково віддалені один від одного:

$$\bar{Y} = \frac{Y_1 + Y_2 + Y_3 + \dots + Y_n}{n} = \frac{\sum Y_n}{n}, \quad (10.9)$$

де Y_n – рівень ряду для n -го періоду;

n – число рівнів ряду динаміки.

Наприклад, за останні п'ять років прибуток підприємства складав (у млн грн): за перший рік – 12,0; за другий – 13,4; за третій – 14,0; за четвертий – 13,8; за п'ятий – 15,6.

Середньорічний обсяг прибутку підприємства складатиме:

$$\bar{Y} = \frac{\sum Y_n}{n} = \frac{12,0 + 13,4 + 14,0 + 13,8 + 15,6}{5} = 13,76 \text{ млн грн};$$

б) коли рівні ряду динаміки не однаково віддалені один від одного:

$$\bar{Y} = \frac{\sum Y_n \times t_n}{\sum t_n}, \quad (10.10)$$

де t_n – відрізок часу, протягом якого зберігається значення рівня Y_n .

Наприклад, середньооблікова чисельність робітників на фірмі за I квартал становила (в особах): 250; за II квартал – 245; за друге півріччя – 233.

Середньооблікова чисельність робітників на фірмі за рік складатиме:

$$\bar{Y} = \frac{\sum Y_n \times t_n}{\sum t_n} = \frac{250 \times 3 + 245 \times 3 + 233 \times 6}{12} = 240,25 \text{ осіб};$$

2) для *моментного ряду* динаміки:

а) з рівними інтервалами часу між датами:

$$\bar{Y} = \frac{\frac{1}{2}Y_1 + Y_2 + \dots + \frac{1}{2}Y_n}{n-1}, \quad (10.11)$$

де n – число рівнів динамічного ряду.

Цей вид середньої має назву – *середня хронологічна проста*.

Наприклад, залишки товарів у торговельній мережі складали (у тис. грн): на 01.01 – 140; на 01.04 – 132; на 01.07 – 150; на 01.10 – 147; на 01.01 наступного року – 142.

Середньоквартальний розмір залишків товарів становитиме:

$$\bar{Y} = \frac{\frac{1}{2}Y_1 + Y_2 + \dots + \frac{1}{2}Y_n}{n-1} = \frac{\frac{1}{2}140 + 132 + 150 + 147 + \frac{1}{2}142}{5-1} = 142,5 \text{ тис. грн};$$

б) з нерівними інтервалами часу між датами:

$$\bar{Y} = \frac{(Y_1 + Y_2) \times t_1 + (Y_2 + Y_3) \times t_2 + \dots + (Y_{n-1} + Y_n) \times t_{n-1}}{2(t_1 + t_2 + \dots + t_{n-1})} = \frac{\sum(Y_n + Y_{n+1}) \times t_n}{2 \sum t_n}, \quad (10.12)$$

де Y_n – рівень моментного ряду динаміки на дату n ;

t_n – період часу між датами (або період, протягом якого рівень Y_n є незмінним).

Цей вид середньої має назву – *середня хронологічна зважена*.

Наприклад, чисельність робітників фірми на деякі дати 2016 р. становила (в особах): на 01.01.2016 р. – 200; на 01.03.2016 р. – 195; на 01.04.2016 р. – 210; на 01.09.2016 р. – 215; на 01.01.2017 р. – 203.

Середньорічна чисельність робітників фірми за 2016 р. складатиме:

$$\bar{Y} = \frac{\sum(Y_n + Y_{n+1}) \times t_n}{2 \sum t_n} = \frac{(200+195) \times 2 + (195+210) \times 1 + (210+215) \times 5 + (215+203) \times 4}{2 \times (2+1+5+4)} = 208 \text{ осіб.}$$

Узагальнювальними характеристиками інтенсивності динаміки є середній абсолютний приріст, середній темп зростання, середній темп приросту.

Середній абсолютний приріст ($\bar{\Delta}$) характеризує середню швидкість зростання (зниження) рівня [12]. Він визначається для моментних та інтервальних динамічних рядів з рівними інтервалами часу між датами як середньоарифметична проста з ланцюгових приростів:

$$\bar{\Delta} = \frac{1}{n} \Delta Y_{\text{ланцюг}}, \quad (10.13)$$

де n – число абсолютних ланцюгових приростів.

Виходячи з того, що $\sum \Delta Y_{\text{ланцюг}} = \Delta Y_{\text{баз.}}$, отримуємо:

$$\bar{\Delta} = \frac{1}{n} (Y_n - Y_0), \quad (10.14)$$

де Y_n – останній рівень динамічного ряду;

Y_0 – рівень, що взятий за базу порівняння;

n – число рівнів ряду.

В обчисленні *середнього темпу росту* потрібно враховувати, що швидкість розвитку явищ відбувається за правилами складних відсотків, де накопичується приріст на приріст. Тому середній темп росту прийнято обчислювати за формулою *середньої геометричної*.

Середній темп (коефіцієнт) зростання (\bar{T}_p) показує, у скільки разів у середньому кожен рівень ряду більше (менше) від попереднього рівня [27]:

1) для динамічних рядів із рівними інтервалами часу між датами:

$$\bar{T}_p = \sqrt[n]{T_{p_1} \times T_{p_2} \times T_{p_3} \times \dots \times T_{p_n}}, \quad (10.15)$$

де $T_{p_1}, T_{p_2}, T_{p_3}, \dots, T_{p_n}$ – ланцюгові темпи зростання;

n – число ланцюгових темпів зростання;

2) коли відомі лише початковий та кінцевий рівні динамічного ряду (за умов нерівних інтервалів часу між датами):

$$\bar{T}_p = \sqrt[n-1]{\frac{Y_n}{Y_0}}, \quad (10.16)$$

де n – число рівнів динамічного ряду.

Середній темп приросту визначається як:

$$\Delta \bar{T}_p = \bar{T}_p - 1(100). \quad (10.17)$$

Розглянуті середні показники динаміки є простими та чітко інтерпретують результат.

Середній темп зростання можна обчислити на основі [68]:

ланцюгових темпів зростання;

кінцевого за весь період темпу зростання;

кінцевого та базисного рівнів ряду.

Наприклад, статутний капітал фірми на початок звітного року становив 10 млн грн, а на кінець – 11,025 млн грн. За перше півріччя капітал зріс на 2,52 %, за друге – на 7,8 %, а за рік – на 10,25 %.

Тоді середньорічний темп росту складе:

$$\bar{T}_p = \sqrt{1,0252 \times 1,078} = \sqrt{1,1025} = \sqrt{\frac{11,025}{10}} = 1,05, \text{ або } 105 \%,$$

тобто в середньому щорічний приріст статутного капіталу на фірмі склав 5 %.

Важливі поняття

Абсолютне прискорення (уповільнення) – різниця між наступним та попереднім абсолютними приростами.

Абсолютний приріст – показник, що відображає абсолютну швидкість змінювання рівнів ряду за певний інтервал часу.

Базисні характеристики динаміки – характеристики динаміки, що обчислені зіставленням з постійною базою порівняння.

Відносне прискорення (уповільнення) – темп приросту абсолютного приросту.

Елементи ряду динаміки – 1) *період часу* (t), за який (або станом на який) приводиться числове значення; 2) *рівень ряду* (Y) – конкретне значення того чи іншого показника.

Інтервальні ряди динаміки – ряди динаміки, які характеризують явища за певний період часу.

Коефіцієнт еластичності – зіставлення темпів приросту взаємопов'язаних показників.

Ланцюгові характеристики динаміки – характеристики динаміки, що обчислені зіставленням суміжних рівнів.

Моментні ряди динаміки – ряди динаміки, які фіксують стан явища на певний момент часу.

Порівнянність даних – необхідна методологічна основа аналізу ряду динаміки.

Ряд динаміки у статистиці – послідовність значень статистичних показників, які характеризують зміну в часі величини суспільного явища чи процесу.

Середній рівень ряду – рівень ряду, який узагальнює властивості, що притаманні ряду динаміки в певний період.

Темп зростання – показник, що характеризує інтенсивність зміни рівнів ряду; є кратним відношенням рівнів у формі коефіцієнта чи відсотка.

Темп приросту – показник, що характеризує відносну швидкість зростання.

Типові завдання

Завдання 10.1. Маємо дані про темпи росту обсягу виробленої підприємством продукції (табл. 10.1). Необхідно визначити середній темп росту та зробити висновки.

Таблиця 10.1

Темпи росту обсягу виробленої продукції за роками

Роки	2009	2010	2011	2012	2013	2014	2015	2016
Темпи росту	1,2006	1,1059	1,1839	1,2909	1,2792	1,2326	1,3245	1,3179

Розв'язання.

1. Знайдемо середній темп росту:

$$\bar{T}_p = \sqrt[n]{T_{p_1} \times T_{p_2} \times \dots \times T_{p_n}} = \sqrt[8]{1,2006 \times 1,1059 \times 1,1839 \times 1,2909 \times 1,2792 \times 1,2326 \times 1,3245 \times 1,3179} = 1,2399.$$

Отже, щорічно обсяг виробленої продукції на підприємстві збільшувався в середньому на 23,99 % за період 2009 – 2016 рр.

Завдання 10.2. Маємо дані про зміни витрат коштів виробничою фірмою на охорону довкілля (табл. 10.2). Необхідно розрахувати абсолютні та відносні показники динамічного ряду розподілу. Отримані результати подайте у вигляді таблиці.

Таблиця 10.2

Розрахункові показники ряду динаміки

Роки	Обсяг коштів на охорону довкілля, млн грн (Y)
2010	11,2
2013	7,7
2014	7,3
2015	6,9
2016	7,0
2017	7,1

Розв'язання.

За базу порівняння вибираємо 2013 р., тому що із цього року має місце щорічна зміна рівня ряду (Y). У табл. 10.3 подана вхідна інформація для проведення розрахунків.

Таблиця 10.3

Розрахункові показники ряду динаміки

Роки	Обсяг коштів на охорону довкілля, млн грн (Y)	Розрахункові показники						Абсолютне значення 1% приросту, (A), млн грн
		Абсолютний приріст (ΔY), млн грн		Темп зростання (T_p), %		Темп приросту (ΔT_p), %		
		ланцюговий	базисний	ланцюговий	базисний	ланцюговий	базисний	
2010	11,2	–	–	–	–	–	–	–
2013	7,7	–	–	–	–	–	–	–
2014	7,3	-0,4	-0,4	94,81	94,81	-5,19	-5,19	0,077
2015	6,9	-0,4	-0,8	94,52	89,61	-5,48	-10,39	0,073
2016	7,0	0,1	-0,7	101,45	90,91	1,45	-9,09	0,069
2017	7,1	0,1	-0,6	101,43	92,21	1,43	-7,79	0,070

За період 2013 – 2017 рр. обсяг коштів на охорону довкілля фірмою було скорочено на 0,6 млн грн, і щорічні порівняння з 2013 р. дають від'ємні показники. Тому ланцюговий абсолютний приріст коштів на охорону довкілля у 2016 і 2017 рр. складав однакову величину – 0,1 млн грн. За 2013 – 2017 рр. спостерігається зниження витрат коштів на охорону довкілля на 7,79 % (92,21 – 100). Лише у 2016 і 2017 рр. спостерігається незначний щорічний приріст у розмірі 1,45 і 1,43 %, відповідно.

Так, у 2014 р. (порівняно з 2013 р.) зниження відрахувань коштів на охорону довкілля на 1 % дорівнювало 0,077 млн грн; підвищення відрахувань на 1 % у 2016 р – 0,069 млн грн, відповідно. Отже, за розглянутий період обсяги витрат коштів на охорону довкілля за виробничою фірмою мали неоднозначну тенденцію до зниження.

Завдання 10.3. Використовуючи інформацію завдання 10.2 про обсяг коштів на охорону довкілля (див. табл. 10.3), визначимо середні показники ряду динаміки.

Розв'язання.

Розглянутий ряд динаміки за сутністю показника є інтервальним; тому середній рівень ряду за 2013 – 2017 рр. буде дорівнювати:

$$\bar{Y} = \frac{\sum Y_n}{n} = \frac{7,7 + 7,3 + 6,9 + 7,0 + 7,1}{5} = \frac{36}{5} = 7,2 \text{ млн грн,}$$

тобто середній рівень витрат на охорону довкілля на фірмі за період 2013 – 2017 рр. склав 7,2 млн грн.

Середній абсолютний приріст за визначений період складе:

$$\bar{\Delta} = \frac{1}{n} \Delta Y_{\text{ланцюг.}} = \frac{(-0,4) + (-0,4) + 0,1 + 0,1}{4} = \frac{-0,6}{4} = 0,15 \text{ млн грн,}$$

тобто за визначений період відбувалось у середньому абсолютне щорічне зниження витрат на охорону довкілля на 0,15 млн грн.

Середній темп (коефіцієнт) зростання розрахуємо за різні періоди часу:

1) 2010 – 2013 рр.:

$$\bar{T}_p = \sqrt[n-1]{\frac{Y_n}{Y_0}} = \sqrt[4-1]{\frac{7,7}{11,2}} = \sqrt[3]{0,6875} = 0,85, \text{ або } 85 \%,$$

тобто за визначений період щорічне зниження витрат на охорону довкілля на виробничій фірмі в середньому складало 15 %;

2) 2013 – 2017 рр.:

$$\bar{T}_p = \sqrt[n]{T_{p1} \times T_{p2} \times \dots \times T_{pn}} = \sqrt[4]{0,9481 \times 0,9452 \times 1,0145 \times 1,0143} = \\ = \sqrt[4]{0,9221} = 0,97993 = 0,98 \text{ (або 98 \%)}$$

або:

$$\bar{T}_p = \sqrt[n-1]{\frac{Y_n}{Y_0}} = \sqrt[5-1]{\frac{7,1}{7,7}} = \sqrt[4]{0,9221} = 0,98, \text{ або } 98 \%,$$

тобто за 2013 – 2017 рр. у середньому щорічне зниження витрат на охорону довкілля на фірмі складає 2 %;

3) 2010 – 2017 рр.:

$$\bar{T}_p = \sqrt[n-1]{\frac{Y_n}{Y_0}} = \sqrt[8-1]{\frac{7,1}{11,2}} = \sqrt[7]{0,63393} = 0,937 \text{ або } 93,7 \%,$$

тобто за визначений період щорічне зниження витрат на охорону довкілля в середньому складає 6,3 %.

Завдання 10.4. Маємо дані про кількість податкових надходжень за період 2013 – 2017 рр. (табл. 10.4). За допомогою абсолютних, відносних і середніх величин необхідно провести аналіз неподаткових надходжень. Результати розрахунків подати в табличному вигляді.

Таблиця 10.4

Розмір неподаткових надходжень

Роки	2013	2014	2015	2016	2017
Неподаткові надходження, млн грн	130	156	141	150	145

Розв'язання.

Результати розрахунків подано в табличному вигляді (табл. 10.5).

Результати обчислення показників динамічного ряду

Роки	Неподаткові надходження, млн грн	Абсолютний приріст, млн грн		Коефіцієнт зростання		Темп приросту, %		Абсолютне значення 1 % приросту, млн грн
		базисний	ланцюговий	базисний	ланцюговий	базисний	ланцюговий	
2013	130	0	–	1,000	–	0	–	–
2014	156	+26	+26	1,200	1,200	+20,0	+20,0	1,30 (↑)
2015	141	+11	–15	1,085	0,904	+8,5	–9,6	1,56 (↓)
2016	150	+20	+9	1,154	1,064	+15,4	+6,4	1,41 (↑)
2017	145	+15	–5	1,115	0,967	+11,5	–3,3	1,52 (↓)

Далі розраховуємо:

1) середньорічний рівень неподаткових надходжень як середню арифметичну просту:

$$\bar{y} = \frac{130+156+141+150+145}{5} = \frac{722}{5} = 144,4 \text{ млн грн};$$

2) середньорічний абсолютний приріст:

за базисними або за ланцюговими абсолютними приростами:

$$\bar{\Delta} = \frac{145 - 130}{5 - 1} = 3,75 \text{ млн грн};$$

$$\bar{\Delta} = \frac{+26 - 15 - 5}{4} = 3,75 \text{ млн грн};$$

3) середньорічний коефіцієнт зростання за формулою середньої геометричної простої:

а) за базисними або за ланцюговими абсолютними приростами:

$$K_3 = \sqrt[4]{1,115} = 1,028;$$

$$K_3 = \sqrt[4]{1,2 \times 0,904 \times 1,064 \times 0,967} = 1,028;$$

б) середньорічний темп приросту:

$$\overline{T_{\text{пр}}} = (1,028 - 1) \times 100 = 2,8 \%$$

або:

$$\overline{T_{\text{пр}}} = 102,8 - 100 = 2,8 \%$$

Отже, можна зробити висновки, що у 2014 р. порівняно з 2013 р. неподаткові надходження зросли на 26 млн грн, а у 2015 р. порівняно з 2014 р. – зменшилися на 15 млн грн. У цілому ж за період із 2013 – 2017 рр. неподаткові надходження зросли на 15 млн грн (або у 1,115 рази чи на +11,5 %). За цей самий період часу середньорічні надходження становили 144,4 млн грн, а їх середньорічний абсолютний приріст становив 3,75 млн грн. Тобто в середньому за рік обсяг неподаткових надходжень за період 2013 – 2017 рр. зростав на 2,8 %.

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Набуття навичок показників динаміки в MS Excel" подані в роботі [67]. Лабораторна робота націлена на отримання практичних навичок у проведенні аналізу рядів динаміки й надавати економічну інтерпретацію розрахованим показникам за допомогою пакету MS Excel.

Запитання для самодіагностики

1. Що таке ряди динаміки і яка їх роль у статистичному аналізі?
2. З яких елементів складається ряд динаміки?
3. Які ви знаєте види рядів динаміки? Наведіть приклади.
4. Яких умов слід дотримуватися в побудові рядів динаміки?
5. Який вид середніх величин використовується для розрахунку середнього рівня моментного й інтервального ряду динаміки?
6. Що таке базисні та ланцюгові показники динаміки?
7. Назвіть показники динаміки та роз'ясніть, як вони розраховуються.
8. Як розраховується середній темп зростання в рядах динаміки?
9. Які бувають ряди динаміки залежно від реєстрації фактів?
10. Назвіть правила побудови рядів динаміки.

Питання для критичного переосмислення (есе)

1. Якщо швидкість розвитку в межах розглядуваного періоду неоднакова, порівнянням однойменних характеристик швидкості вимірюється прискорення чи уповільнення динаміки.

2. Які взаємопов'язані характеристики доцільно використовувати для оцінювання властивостей динаміки? Відповідь обґрунтуйте.

3. Шляхом розрахунку різного роду середніх показників узагальніть притаманні динамічному ряду властивості.

4. За яких причин виникає непорівнюваність рівнів ряду динаміки? Наведіть приклади.

5. Поясніть взаємозв'язок абсолютного приросту та темпу приросту. Доведіть, що абсолютне значення одного відсотка приросту становить соту частину рівня, взятого за базу порівняння.

11. Аналіз тенденцій розвитку та коливань

Основні питання:

11.1. Прийоми виявлення основної тенденції розвитку в рядах динаміки.

11.2. Інтерполяція та екстраполяція.

11.3. Факторний аналіз рядів динаміки.

11.4. Аналіз сезонних коливань.

11.1. Прийоми виявлення основної тенденції розвитку в рядах динаміки

Ряди динаміки характеризують процеси розвитку соціально-економічних явищ. Цим процесам одночасно властиві динамічність та інерція. Динамічність виявляється у змінах, коливаннях рівнів ряду динаміки, а інерція – у сталості дії формуючих факторів, напрямі й інтенсивності розвитку. Ряд динаміки містить у собі залишки минулого, основи сьогодення і зародки майбутнього. У першу чергу необхідно виділити **фактори еволюційного характеру**, що здійснюють постійний вплив і визначають загальний напрямок розвитку явища, його довгострокову еволюцію.

Фактори постійної дії здійснюють на досліджуване явище визначальний вплив і формують у рядах динаміки основну *тенденцію розвитку* (тренд), як правило, довгострокову. Другою групою факторів є **фактори осцилятивного характеру**. Вплив цих факторів виявляється періодично – це викликає повторювані в часі коливання рівнів ряду динаміки – сезонні, циклічні (наприклад, цикли економічної кон'юнктури). Схематично циклічні коливання можна подати у вигляді синусоїди (значення ознаки спочатку зростає, досягає певного *max*, потім знижується, досягає свого *min*, знову зростає і так далі). Решта факторів, що впливають на ряд динаміки, – це фактори, що викликають нерегулярні коливання рівнів, які зумовлюють випадкові короткочасні зміни рівнів ряду динаміки. Ці фактори розподіляють на [13; 48; 68]:

ті, що спричиняють спорадичні зміни рівнів (війна, екологічні катастрофи, епідемії тощо);

випадкові, слабкого впливу. Це другорядні фактори, що призводять до випадкових різноспрямованих змін рівнів.

Будь-який соціально-економічний процес (Y) розглядають як функцію часу (t). Звичайно, час не є фактором конкретного соціально-економічного явища; він лише акумулює в собі всі умови та причини, що визначають це явище. Отже, будь-який ряд динаміки теоретично може бути поданий у вигляді таких складових [41; 48; 68]:

тренд – основна довгострокова тенденція розвитку динамічного ряду (до збільшення або зниження його рівнів) $f(t)$;

циклічні (періодичні) коливання $C(t)$;

сезонні коливання $S(t)$;

випадкові коливання $E(t)$.

Зв'язок між цими компонентами можна подати адитивно (сумарно), або мультиплікативно (добутком), або комбіновано:

$Y_t = f(t) + C_t + S_t + E_t$ – характеризується тим, що характер циклічних і сезонних коливань залишається постійним;

$Y_t = f(t) \times C_t \times S_t \times E_t$ – якщо характер циклічних і сезонних коливань залишається постійним лише стосовно тренду;

$$Y_t = f(t) + C_t \times S_t + E_t.$$

Розглянемо на прикладі різницю між адитивною та мультиплікативною сезонними компонентами. Графік обсягу продажів дитячих іграшок, імовірно, буде мати щорічні піки у листопаді-грудні та інший – суттєво менший за висотою – у літні місяці, що припадають на канікули. Така сезонна закономірність буде повторюватися щороку. За своєю природою сезонний компонент може бути *адитивним* або *мультиплікативним*. Так, щороку обсяг продажів деякої конкретної іграшки може збільшуватися в грудні на 3 млн грн. Тому можна врахувати ці сезонні зміни, додаючи до свого прогнозу на грудень 3 млн грн. Тут ми маємо адитивну сезонність. Проте у грудні обсяг продажів деякої іграшки може збільшуватися на 40 %, тобто множитися на множник 1,4. Це означає, що якщо середній обсяг продажів цієї іграшки невеликий, то абсолютне (у грошовому вираженні) збільшення цього обсягу в грудні також буде відносно невеликим (але у відсотковому відношенні воно буде постійним). Якщо ж іграшка продається добре, то й абсолютне (у гривнях) зростання обсягу продажів буде значним. Тут знову обсяг продажів зростає в число разів, що дорівнює певному множнику, а сезонний компонент – за своєю природою, мультиплікативний у цьому випадку дорівнює 1,4. Якщо перейти до графіка часових рядів, то розходження між цими двома видами сезонності буде проявлятися так: в адитивному випадку ряд буде мати постійні сезонні коливання, величина яких не залежатиме від загального рівня значень ряду; у мультиплікативному випадку величина сезонних коливань буде змінюватися залежно від загального рівня значень ряду.

Розглянутий приклад можна розширити, щоб проілюструвати поняття адитивності та мультиплікативності тренд-циклічних компонентів. У випадку з іграшками тренд моди може привести до стійкого зростання продажів (наприклад, це може бути загальний тренд у бік іграшок освітньої спрямованості). Як і сезонний компонент, цей тренд може бути за своєю природою адитивним (продажі щорічно збільшуються на 3 млн грн) або мультиплікативним (продажі щорічно збільшуються на 30 %, або в 1,3 раза). Крім того, обсяг продажів може містити циклічні компоненти. Повторимо ще раз, що циклічний компонент відрізняється від сезонного тим, що він звичайно має більшу часову довжину та проявляється через нерівні проміжки часу. Так, деяка іграшка може мати ажіотажний попит протягом літнього сезону (наприклад, агресивно рекламована лялька, що зображує персонаж популярного мультфільму). Як і в попередніх випадках, такий

циклічний компонент може змінювати обсяг продажів адитивно або мультиплікативно.

Особливістю вивчення розвитку соціально-економічних явищ у часі є те, що в одних рядах динаміки загальна тенденція розвитку (зростання або падіння показника) проявляється відразу через візуальний огляд інформації, шляхом побудови графіка вихідних даних, а в інших рядах динаміки тенденція безпосередньо не виявляється.

Вивчення тренду включає два основних етапи [68]:

ряд динаміки перевіряють на наявність тренду;

проводять **вирівнювання часового ряду** та безпосереднє виділення тренду з екстраполяцією отриманих результатів.

У соціально-економічних рядах динаміки можна спостерігати тенденцію трьох видів: середнього рівня; дисперсії; автокореляції.

Тенденція середнього рівня – аналітично виражена за допомогою математичної функції, навколо якої варіюють фактичні рівні досліджуваного явища. У цьому випадку значення тренду в окремі моменти часу будуть математичними очікуваннями ряду динаміки. Часто тенденцію середнього рівня називають *детермінованою* (невипадковою) *складового ряду динаміки*.

Тенденція дисперсії є тенденцією зміни відхилень між емпіричними рівнями та детермінованим компонентом ряду.

Тенденцією автокореляції є тенденція зміни зв'язку між окремими рівнями ряду динаміки. Графічно ця зміна не простежується.

Початковим етапом виділення і аналізу тренду є перевірка гіпотези про існування тренду. Існує біля десятка критеріїв перевірки наявності тренду. Розглянемо деякі з них [41; 48; 68].

1. **Перевірка істотності різниці середніх.** Ряд динаміки розбивають на дві рівні або майже рівні частини. Перевіряють гіпотезу про існування різниці середніх: $H_0: \bar{y}_1 = \bar{y}_2$. Оскільки число членів аналізованого ряду, як правило, мале, то для перевірки гіпотези користуються теорією малої вибірки. За основу перевірки беруть t_α -критерій Стьюдента. Якщо $t \geq t_\alpha$, гіпотеза про відсутність тренду відкидається, і навпаки: за $t < t_\alpha$ гіпотеза (H_0) приймається. Тут t – розрахункове значення, знайдене для аналізованих даних; t_α – табличне значення критерію за рівня ймовірності помилки, дорівненого α . У випадку рівності або несуттєвого розходження

дисперсій двох досліджуваних сукупностей ($\sigma_1^2 = \sigma_2^2$) визначення розрахункового значення t проводять за залежністю:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\delta \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad (11.1)$$

де \bar{y}_1, \bar{y}_2 – середні для першої і другої половини ряду динаміки;

n_1 і n_2 – число спостережень у цих рядах;

σ – середньоквадратичне відхилення різниці середніх, обумовлене залежністю:

$$\sigma = \sqrt{\frac{(n_1 - 1) \delta_1^2 + (n_2 - 1) \delta_2^2}{n_1 + n_2 - 2}}. \quad (11.2)$$

Дисперсії для першої і другої частин ряду розраховують за залежністю:

$$\sigma_i^2 = \frac{\sum (y_i - \bar{y}_i)^2}{n - 1}. \quad (11.3)$$

Перевірку гіпотези про рівність **дисперсій** здійснюють за допомогою F-критерію, основний сенс якого полягає у порівнянні розрахункового відношення з табличним. Розрахункове значення критерію визначають за формулою:

$$F = \frac{\sigma_1^2}{\sigma_2^2}. \quad (11.4)$$

Якщо розрахункове значення F менше табличного за заданого рівня значущості, то гіпотезу про рівність дисперсій приймають. Якщо F більше, ніж табличне значення, то гіпотезу про рівність дисперсій відхиляють і залежність для розрахунку t вважають не придатною для використання. Для виконання умови про рівність дисперсій визначають значення t_α і перевіряють гіпотезу (H_0). Для цього теоретичне значення t_α визначають із числом ступенів свободи дорівненим $n_1 + n_2 - 2$. Розглянутий метод дає позитивні результати для рядів із монотонною тенденцією. Коли ж ряд

динаміки змінює загальний напрям розвитку, то точка повороту тенденції виявляється близькою до середини ряду. Тому середні двох відрізків будуть близькі, а перевірка може не показати наявності тенденції.

2. Перевірку наявності тенденції проводять **за методом Фостера – Стюарта**. Цей метод (крім визначення наявності) тенденції дозволяє виявити тренд дисперсії рівнів ряду динаміки, який важливо знати для аналізу та прогнозування економічних явищ.

Нульова гіпотеза H_0 підтверджує існування тренду. Статистики критерію мають вигляд:

$$s = \sum_{i=2}^n S_i; \quad (11.5)$$

$$d = \sum_{i=2}^n d_i, \quad (11.6)$$

де $d_i = u_i - l_i$; $S_i = u_i + l_i$.

Якщо $x_i > x_{i-1}, \dots, x_1$, то $u_i = 1$, у протилежному випадку $u_i = 0$, якщо $x_i < x_{i-1}, \dots, x_1$, то $l_i = 1$; в іншому випадку $l_i = 0$.

Статистика S використовується для перевірки тренду в дисперсіях, статистика d – для виявлення тренду в середніх.

Очевидно, що $0 \leq S \leq n-1$; $-(n-1) \leq d \leq n-1$.

За відсутності тренду величини: $t = d/f$.

$$\bar{t} = \frac{S - f^2}{l}, \text{ де } l = \sqrt{2 \ln n - 3,4253}, f = \sqrt{2 \ln n - 0,8456}$$

мають розподіл Стюдента з $v = n$ ступенями свободи. Формули для f й l використовують за $n > 50$, їхні значення за $n < 50$ наведені в табл. 11.1.

Таблиця 11.1

Постійні f і l критерію Фостера – Стюарта

n	10	15	20	25	30	40	45	50
f	1,964	2,153	2,279	2,373	2,447	2,509	2,561	2,606
l	1,288	1,521	1,677	1,791	1,882	1,956	2,019	2,072

Якщо $|\bar{t}| > t_{\frac{1+\alpha}{2}}$, то з довірчою ймовірністю α нульова гіпотеза H_0 існу-

вання тренду приймається, у протилежному випадку гіпотеза H_0 відкидається; $(t_\gamma - \gamma)$ – квантиль розподілу Стюдента.

Після встановлення наявності тенденції в ряді динаміки проводять її опис.

У вивченні динаміки важливі не лише числові значення рівнів, а і їх послідовність. Як правило, часові інтервали між рівнями однакові (доба, декада, календарний місяць, квартал, рік).

Узявши будь-який інтервал за одиницю, послідовність рівнів запишемо так: $y_1, y_2, y_3, \dots, y_n$.

Залежно від статистичної природи показника-рівня розрізняють динамічні ряди первинні та похідні, ряди абсолютних, середніх і відносних величин. За ознакою часу динамічні ряди розподіляють на інтервальні та моментні. Рівень моментного ряду фіксує стан явища на певний момент часу t (наприклад кількість працюючих на початок року, студентів – на 1 вересня і т. д.). В інтервальному ряді рівень – це агрегований результат процесу; залежить від тривалості часового інтервалу: виробництво електроенергії за рік, вилов риби за сезон. Зауважимо, що похідні показники, обчислені на основі інтервальних рядів, на відміну від моментних, залежать від протяжності часу (середньодобове або середньорічне виробництво електроенергії на душу населення).

Соціально-економічні процеси є *динамічні*, якщо проявляються сталою зміною рівнів динамічного ряду. Поряд з динамічністю їм притаманна *інерційність*: зберігається механізм формування явищ і характер розвитку (темпи, напрям, коливання). У разі значної інерційності процесу та незмінності комплексу умов його розвитку правомірно очікувати в майбутньому ті властивості та характер розвитку, які були виявлені в минулому. Діалектична єдність мінливості та сталості, динамічності та інерційності формує характер динаміки, уможливлючи статистичне прогнозування соціально-економічних процесів [68].

У ході вивчення закономірностей соціально-економічного розвитку статистика розв'язує низку завдань: вимірює інтенсивність динаміки, виявляє і описує тенденції, оцінює структурні зрушення, сталість і коливання рядів; виявляє фактори, які спричиняють зміни.

Передумовою аналізу будь-якого динамічного ряду є порівнянність статистичних даних, які його формують. Непорівнянність даних може зумовлюватися різними причинами:

змiнами в методологiї обліку та розрахунку показника, зокрема використання різних одиниць для вимірювання;

змiнами в структурі сукупності, а також територіальними змiнами;

різними критичними моментами реєстрації даних чи тривалістю періодів, до яких належать рівні;

змiною цін для вартісних показників.

Порівнянність даних забезпечується на етапах їх збирання та оброблення. Використовують також спеціальні прийоми зведення даних до порівнянного вигляду – "статистичні ключі" зімкнення динамічних рядів.

Приклад 11.1. Припустимо, помісячні рівні витрат сировини на виробництво продукції в I півріччі непорівнянні, оскільки у квітні змінився порядок обліку витрат (табл. 11.2).

Таблиця 11.2

Зімкнення динамічних рядів

Місяці	Обсяг витрат, т		Зімкнений ряд	
	Старий порядок реєстрації	Новий порядок реєстрації	Відносних величин, %	Абсолютних величин, т
Січень	40	–	80	44,0
Лютий	45	–	90	49,5
Березень	48	–	96	52,8
Квітень	50	55	100	55,0
Травень	–	58	105	58,0
Червень	–	60	109	60,0

Подолати переривчастість ряду можна двома способами:

способом відносних рівнів, коли за базу порівняння для кожного ряду беруть квітневий рівень. Два ряди відносних рівнів об'єднуються в один;

способом, який ґрунтується на співвідношенні квітневих рівнів: $55 : 50 = 1,1$. Помноживши рівні першого ряду на цей коефіцієнт, дістанемо єдиний зімкнений (порівняний) ряд динаміки за весь період (остання графа табл. 11.2).

Важливим завданням статистики в аналізі рядів динаміки є визначення основної тенденції розвитку, притаманної тому чи іншому ряду динаміки. Під **основною тенденцією розвитку ряду динаміки** розуміють

зміну, яка визначає загальний напрям розвитку. Це є систематичною складовою довгострокової дії. У деяких випадках загальна тенденція чітко простежується в динаміці розглядуваного показника; в інших випадках вона може не проявлятися через відчутні випадкові коливання. Наприклад, в окремі моменти часу сильні коливання роздрібних цін можуть затулити наявність тенденції до зростання або зниження цього показника. Тому для виявлення основної тенденції розвитку в статистиці застосовують дві **групи методів** [68]:

згладжування, або механічне вирівнювання, окремих рівнів ряду динаміки з використанням фактичних значень сусідніх рівнів;

вирівнювання із застосуванням кривої, проведеної між конкретними рівнями таким чином, щоб вона відображала тенденцію, притаманну ряду, й одночасно звільняла його від незначних коливань.

Розглянемо методи кожної групи.

1. **Метод укрупнення інтервалів** заснований на укрупненні періодів часу, до яких належать рівні. Наприклад, ряд тижневих даних можна перетворити в ряд помісячної динаміки, ряд квартальних даних замінити річними рівнями. Рівні нового ряду можуть бути отримані шляхом підсумовування рівнів вихідного ряду або можуть представляти середні рівні.

Поширеним прийомом виявлення тенденції розвитку є згладжування ряду динаміки. Сутність різних прийомів згладжування зводиться до заміни фактичних рівнів ряду розрахунковими рівнями, які меншою мірою схильні до коливань. Це сприяє більш чіткому прояву тенденції розвитку.

2. До **методів згладжування** належать методи: усереднення, простої ковзної і зваженої ковзної середньої.

2.1. Метод *усереднення за лівою та правою половиною*. Ряд динаміки ділять на дві частини, знаходять для кожної з них середнє арифметичне значення і проводять через отримані точки лінію тренду на графіку.

2.2. Метод *простої ковзної (плинної) середньої*. У цьому методі вихідні рівні ряду замінюють середніми величинами, які отримують із даного рівня і декількох, симетрично розташованих навколо нього. Ціле число рівнів, з яких розраховують середнє значення, називають *інтервалом згладжування* (плинності). Однак розрахунок середньої проводять із поступовим виключенням з даного інтервалу згладжування першого рівня ряду

динаміки та включенням наступного: $\bar{y}_1 = \frac{y_1 + y_2 + y_3}{3}$; $\bar{y}_2 = \frac{y_2 + y_3 + y_4}{3}$ і т. ін.

Інтервал може бути непарним (3, 5, 7 і т. д. рівнів) або парним (2, 4, 6 і т. д. рівнів).

У *непарному згладжуванні* отримане середнє арифметичне значення відповідає середині розрахункового інтервалу, в *парному* середнє значення відповідає проміжку між датами.

Для *згладжування* ряду динаміки з *парним числом рівнів* виконують додаткову операцію, що називають *центруванням*, оскільки, в обчисленні ковзної середньої (наприклад, за чотирма рівнями) $\bar{y}_1 = \frac{y_1 + y_2 + y_3 + y_4}{4}$,

$\bar{y}_2 = \frac{y_2 + y_3 + y_4 + y_5}{4}$. Її значення відноситься до точки між моментами часу,

коли було зафіксовано фактичні рівні y_2 та y_3 . Схема обчислень і розташувань рівнів згладженого ряду ускладнюється:

$y_1, y_2, y_3, y_4, y_5, y_6 \dots$ – вихідні рівні;

$\bar{y}_1, \bar{y}_2, \bar{y}_3 \dots$ – згладжені рівні;

$\bar{y}_{1ц}, \bar{y}_{2ц} \dots$ – центровані згладжені рівні: $\bar{y}_{1ц} = \frac{\bar{y}_1 + \bar{y}_2}{2}$, $\bar{y}_{2ц} = \frac{\bar{y}_2 + \bar{y}_3}{2}$.

Розглянемо застосування ковзної середньої на прикладі динамічного ряду продажу смартфонів у торговій мережі за 2017 р. (табл. 11.3):

$$y_1 = \frac{23 + 25 + 21}{3} = 23; \quad y_2 = \frac{25 + 21 + 26}{3} = 24 \text{ і т. д.}$$

Таблиця 11.3

Динаміка продажу смартфонів у торговій мережі за 2017 р.

Місяці	Продано смартфонів, тис. шт.	Трьохрівневі ковзні середні	Чотирьохрівневі ковзні середні нецентровані	Чотирьохрівневі ковзні середні центровані
1	2	3	4	5
Січень	23	–	–	–
Лютий	25	23	23,8	–
Березень	21	24	25,0	24,4
Квітень	26	25	24,8	24,9
Травень	28	26	26	25,8
Червень	24	27	26,8	27
Липень	29	27	27,3	27,5

1	2	3	4	5
Серпень	28	29	27,8	28,4
Вересень	30	29	29	29,3
Жовтень	29	30	29,5	30,1
Листопад	31	31	30,8	–
Грудень	33	–	–	--

На рис. 11.1 видно згладжування сильно коливних рівнів, що здійснено за допомогою ковзної середньої.

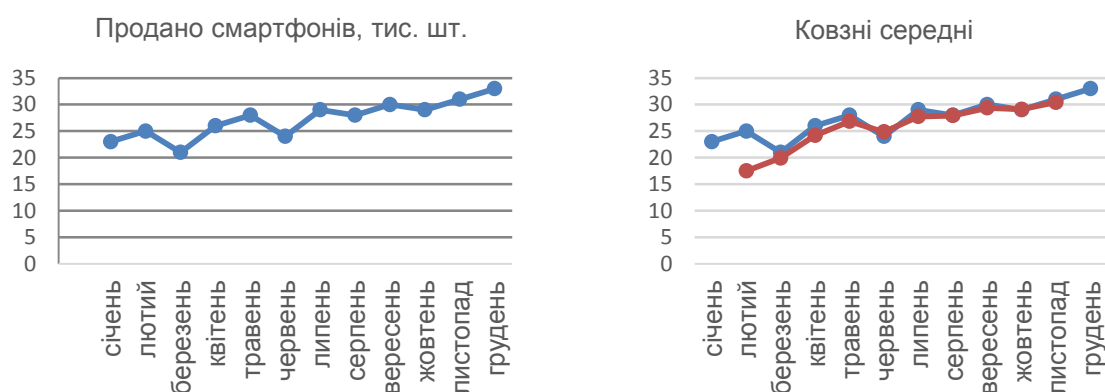


Рис. 11.1. Динаміка продажу смартфонів у торговій мережі за 2017 р.

2.3. Метод **зваженої ковзної середньої**. Основна відмінність від попереднього методу полягає в тому, що рівням, які входять в інтервал усереднення, надають різні ваги, оскільки апроксимація в межах інтервалу згладжування здійснюється з використанням рівнів, розрахованих за поліномом n -го порядку.

$$\bar{y}_i = a_0 + a_1 \times i + a_2 \times i^2 + \dots, \quad (11.7)$$

де i – порядковий номер рівня інтервалу згладжування.

Недолік методу простий ковзної середньої полягає в тому, що згладжений ряд динаміки скорочується через неможливість отримати згладжені рівні для початку та закінчення ряду. Цей недолік усувається застосуванням методу аналітичного вирівнювання для аналізу основної тенденції.

3. Метод **аналітичного вирівнювання** є найбільш ефективним способом визначення основної тенденції. Під цим розуміють визначення основної тенденції розвитку досліджуваного явища, що проявляється в часі. Аналітичне вирівнювання полягає в підборі для даного ряду динаміки теоретичної кривої, що виражає основні риси динаміки фактичного явища, тобто щонайкраще описує емпіричні дані.

У результаті вирівнювання ряду динаміки отримують найбільш загальний, сумарний, проявлений у часі результат дії всіх причинних факторів. Відхилення конкретних рівнів ряду від рівнів, що відповідають загальній тенденції, пояснюють дією факторів, що проявляються випадково або циклічно. Для висунування гіпотези про можливий тип розвитку потрібно використовувати графічний метод. Наочне зображення аналізованого ряду динаміки дозволяє побачити розміщення на полі графіка емпіричних рівнів. Це сприяє кращому осмисленню специфіки змін у ряді динаміки. Але дати узагальнену статистичну оцінку виявленого тренду графічний метод не може.

Загальний вид трендової моделі такий:

$$Y_t = f(t) + E_t, \quad (11.8)$$

де $f(t)$ – рівень, обумовлений тенденцією розвитку;

E_t – випадкове та циклічне відхилення від тенденції.

Добір адекватної функції здійснюється методом найменших квадратів – мінімальністю відхилень суми квадратів між теоретичними й емпіричними рівнями ряду:

$$\sum (y_{t_i} - y_i)^2 \rightarrow \min, \quad (11.9)$$

де y_{t_i} , y_i – відповідно теоретичні й емпіричні рівні ряду.

Найважливішою проблемою аналітичного вирівнювання є добір математичної функції, за якою розраховують теоретичні рівні тренду. Від правильності розв'язання цієї задачі залежать висновки й прогнози про закономірності тренду досліджуваного явища.

Вибір форми кривої можна проводити завдяки аналізу графічного зображення рівнів ряду динаміки. Для цього доцільно використовувати

графічне зображення згладжених рівнів, у яких погашені випадкові коливання. Для оцінювання близькості трендового рівняння емпіричному ряду динаміки застосовують критерій Фішера (F). Фактичний (розрахунковий) рівень F-критерій порівнюють із теоретичним (табличним) значенням:

$$F_{\text{факт}} = \frac{\eta_T^2}{1 - \eta_T^2} \times \frac{n - m}{m - 1} = \frac{V_1}{V_2}, \quad (11.10)$$

де m – число параметрів;

η_T^2 – теоретичний коефіцієнт детермінації.

$$\eta_T^2 = 1 - \frac{\sigma_{y-\hat{y}}^2}{\sigma_y^2}, \quad (11.11)$$

де $\sigma_{y-\hat{y}}^2$ – залишкова дисперсія;

σ_y^2 – загальна дисперсія.

Залишкова дисперсія розраховується за формулою:

$$\sigma_{y-\hat{y}}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}; \quad (11.12)$$

загальна дисперсія:

$$\sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}. \quad (11.13)$$

Для визнання моделі надійною необхідне дотримання умови:

$$F_{\text{факт}} > F_{\text{крит.}}$$

$F_{\text{крит}}$ знаходять за спеціальними таблицями розподілу Фішера при $v_1 = m - 1$, $v_2 = n - m$ і рівні значущості α .

Для динамічних рядів, що мають невелику довжину та значні коливання, не рекомендується використовувати метод аналітичного вирівнювання за допомогою часової функції, тому що апроксимація практично

не адаптується до умов, що змінює формування рівнів. Із появою нових даних потрібно будувати нові моделі. Для згладжування таких рядів динаміки використовують методи адаптивного моделювання і прогнозування. В основу зазначених методів покладена модель експоненційного згладжування. Часовий ряд згладжують за допомогою зваженої ковзної середньої, у якій ваги розподіляють за експонентним законом.

11.2. Інтерполяція та екстраполяція

Аналіз динаміки соціально-економічних явищ, виявлення і характеристика основної тенденції розвитку дають підставу для прогнозування – визначення майбутніх розмірів рівня економічного явища.

Процес прогнозування передбачає, що закономірність розвитку, яка діяла в минулому (всередині ряду динаміки), збережеться і в прогнозованому майбутньому, тобто прогноз заснований на **екстраполяції**. Екстраполяція, що проводиться в майбутнє, називається **перспективною**, і в минуле – **ретроспективною**. Зазвичай, говорячи про екстраполяцію рядів динаміки, мають на увазі переважно перспективну екстраполяцію. Первинні прогнози, як правило, зводяться до екстраполяції тенденції. Для цього можуть використовуватися різні методи – залежно від вихідної інформації. Можна виділити наступні *елементарні методи екстраполяції*: на основі середнього абсолютного приросту, середнього темпу зростання і екстраполяція на основі застосування методу найменших квадратів і подання розвитку явищ у часі у вигляді рівняння тренду, тобто математичної функції рівнів ряду (y) від фактора часу (t) [16; 48; 68; 71].

Прогнозування за середнім абсолютним приростом можливе в тому випадку, якщо доцільно вважати загальну тенденцію лінійною. Тобто метод заснований на припущенні про рівномірну зміну рівня (під рівномірністю розуміють стабільність абсолютних приростів).

У цьому випадку, щоб отримати прогноз на "i" кроків вперед (i – період упередження), достатньо використати формулу:

$$\hat{y}_{n+1} = y_n + i \times \bar{\Delta}, \quad (11.14)$$

де y_n – фактичне значення в останній n-ій точці ряду (кінцевий рівень ряду);

\hat{y}_{n+1} – прогнозна оцінка значення (n+1) рівня ряду;

$\bar{\Delta}$ – значення середнього абсолютного приросту, розраховане для ряду динаміки $y_1, y_2, y_3, \dots, y_n$.

Прогнозування за середнім темпом зростання можна здійснювати у випадку, коли є підстава вважати, що загальна тенденція ряду характеризується показовою (експоненційною) кривою. Для пошуку прогнозного значення на i кроків вперед необхідно використовувати таку формулу:

$$\hat{y}_{n+1} = y_n \times \overline{K}_p^i, \quad (11.15)$$

де \overline{K}_p^i – середній коефіцієнт зростання, розрахований для ряду $y_1, y_2, y_3, \dots, y_n$.

Прогнозування на основі аналітичного вирівнювання є найпоширенішим методом прогнозування. Для формування прогнозу використовують аналітичне рівняння тренду. Для цього, досить у моделі продовжити значення умовного показника часу. Формально операцію екстраполяції можна подати як визначення функції [68]:

$$Y_{t+v} = f(Y_t^0, v), \quad (11.16)$$

де Y_{t+v} – прогнозне значення на період упередження v ;

Y_t^0 – база екстраполяції, найчастіше – останній, визначений за рівнянням тренду, рівень ряду.

За допомогою методу екстраполяції отримують два види прогнозу: точкові й інтервальні. **Точковий прогноз** є конкретним числовим значенням рівня в прогнозований період (момент) часу. **Інтервальний прогноз** – діапазон числових значень, що приблизно містить прогнозоване значення рівня [68].

Інтервальні прогнози мають значні переваги перед точковими – вони враховують імовірність здійснення прогнозу. Очевидно, що точковий прогноз малоімовірний, оскільки тренду властиві коливання рівнів, а звідси виникають помилки параметрів. Джерелом цих помилок є обмежене число спостережень, кожен з яких містить випадковий компонент. Випадковий компонент спостерігається і за межами динамічного ряду, а отже, його треба враховувати. Для цього визначають довірчий інтервал, який би з певною ймовірністю встановив межі можливих значень

Y_{T+V} . Точковий прогноз перетворюється на інтервальний – ширина інтервалу залежить від варіації рівнів динамічного ряду навколо тренду та ймовірності висновку $(1 - \alpha)$:

$$Y_{t+V} = \pm t_{\alpha} \bar{\delta}_e, \quad (11.17)$$

де t_{α} – коефіцієнт довіри розподілу Стьюдента;

$$\bar{\delta}_e = \sqrt{\frac{\sum (y_i - y_{t_i})^2}{n - m}}, \quad (11.18)$$

де $\bar{\delta}_e$ – залишкове середньоквадратичне відхилення тренду, скориговане за числом ступенів свободи $(n - m)$;

n – число рівнів базисного ряду динаміки;

m – число параметрів адекватної моделі тренду.

У практиці статистичного аналізу тренду виділяють такі типи розвитку соціально-економічних явищ у часі:

1) **рівномірний розвиток**. Для цього типу динаміки характерні постійні абсолютні прирости:

$$\Delta_{y_t} \cong \text{const}. \quad (11.19)$$

Основну тенденцію в таких рядах описують за допомогою рівняння прямолінійної функції:

$$\hat{y}_t = a_0 + a_1 t, \quad (11.20)$$

де \hat{y}_t – теоретичне значення ряду динаміки;

a_0 і a_1 – параметри рівняння;

t – час (порядковий номер періоду, чи моменту часу).

Для лінійної залежності параметр a_0 звичайно інтерпретації немає, але іноді його розглядають як узагальнений початковий рівень ряду; a_1 – сила зв'язку, тобто параметр, що показує, наскільки зміниться результат за зміною часу на одиницю. Отже, a_1 можна подати як постійний теоретичний абсолютний приріст. Параметр a_1 є коефіцієнтом регресії, що визначає

напряму розвитку. Якщо $a_1 > 0$, то рівні ряду динаміки рівномірно зростають; якщо $a_1 < 0$, відбувається їхнє рівномірне зниження;

2) **рівноприскорений (рівноуповільнений) розвиток**. Для цього типу динаміки характерно постійне в часі збільшення (уповільнення) розвитку. Рівні таких рядів динаміки змінюються з постійними темпами приросту [68]:

$$T_{\rho_c} \cong \text{const}. \quad (11.21)$$

Основну тенденцію в таких рядах описують параболою другого порядку:

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2. \quad (11.22)$$

Параметри a_1 , a_0 ідентичні параметрам рівняння лінійної регресії. Параметр a_2 характеризує постійну зміну інтенсивності розвитку в одиницю часу. Якщо $a_2 > 0$, відбувається прискорення розвитку, а за $a_2 < 0$ зростання уповільнюється;

3) **розвиток зі змінним прискоренням (уповільненням)**. Таку динаміку описують за допомогою параболи 3-го порядку:

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3. \quad (11.23)$$

Параметр a_3 відображає зміну прискорення. Якщо $a_3 > 0$, прискорення зростає, а за $a_3 < 0$ прискорення сповільнюється.

4) **розвиток за експонентою**. Для цього типу динаміки характерні стабільні темпи зростання:

$$T_{\rho_c} \cong \text{const}. \quad (11.24)$$

Основна тенденція описується показовою функцією :

$$\hat{y}_t = a_0 a_1^t, \quad (11.25)$$

де a_1 – темп зростання (зниження) досліджуваного явища в одиницю часу, тобто інтенсивність розвитку;

5) *розвиток з уповільненням зростання наприкінці періоду.*

Для цього типу динаміки величини ланцюгових абсолютних приростів скорочується в кінцевих рівнях ряду динаміки.

$$\Delta_{y_t} \rightarrow 0. \quad (11.26)$$

Тенденцію розвитку в таких рядках динаміки описують за допомогою напівлогарифмічної функції:

$$\hat{y}_t = a_0 + a_1 \lg t. \quad (11.27)$$

Найбільш часто застосовують такі моделі регресій:

гіперболи: $\hat{y}_t = a_0 + \frac{a_1}{t}$; (11.28)

Таку модель називають зворотною. Як правило, її застосовують, коли необмежене збільшення пояснювальної змінної (у цьому випадку – часу t) асимптотично наближає залежну змінну Y до повної межі;

ступенева функція: $\hat{y}_t = a_0 + t_1^a$. (11.29)

Для аналітичного вирівнювання в рядах динаміки можна застосовувати й інші математичні функції. Як зазначалось, у дослідженні соціально-економічних явищ доводиться оперувати зі складним механізмом взаємодії факторів, що формують тренд. Тому на основі якісного аналізу не завжди можна отримати надійні висновки про тип розвитку у вигляді адекватної математичної функції. У найкращому разі може бути висунута робоча гіпотеза про можливі типи розвитку. Але вибір на цій основі конкретної математичної функції досить ускладнений. Особливо це стосується нелінійних функцій, теорія яких розроблена недостатньо.

Практика статистичного вивчення тренду з використанням засобів сучасної обчислювальної техніки уможлиблює підбір найбільш адекватної трендової моделі. Швидкодія сучасних ЕОМ з великою ємністю пам'яті надає всі необхідні для аналізу тренду показники, у тому числі застосовні для вибору адекватної математичної функції.

Один із показників адекватності математичної функції, використуваних у практиці статистичного вивчення тренду, є **стандартизована помилка апроксимації** [68]:

$$\sigma_{y_t} = \sqrt{\frac{\sum (y_{t_i} - y_t)^2}{n}}. \quad (11.30)$$

Застосування її у вивченні тренду базується на тому, що за найбільш адекватну беруть функцію, в якій стандартизована помилка апроксимації мінімальна.

Приклад 11.2. Розглянемо процес аналітичного вирівнювання ряду динаміки на базі валового збору зернових в Україні. З розглянутого вище графічного аналізу динаміки валового збору та виявлення тенденції за допомогою тричленної ковзної середньої видно, що явище має сильні коливання і тенденція проявляється не чітко. Тому проведемо аналітичне вирівнювання ряду за декількома функціями й оберемо з них найбільш доцільну (адекватну) реальним даним.

Здійснимо попередній аналіз ряду за допомогою ланцюгових приростів, темпів зростання і темпів приросту (табл. 11.4).

Таблиця 11.4

**Динаміка валового збору зернових, абсолютні приросту,
темпи зростання й темпи приросту**

Роки	Валовий збір зернових, тис. т (y)	$\Delta_{уц}$	$T_{рц}$	$T_{прц}, \%$
1	2	3	4	5
2001	38 674	–	–	–
2002	38 537	–137	1,00	–0,35
2003	45 623	7 086	1,18	18,39
2004	35 497	–10 126	0,78	–22,19
2005	33 930	–1 567	0,96	–4,41
2006	24 571	–9 359	0,72	–27,58
2007	35 472	10 901	1,44	44,37
2008	26 471	–9 001	0,75	–25,37

1	2	3	4	5
2009	24 581	-1 890	0,93	-7,14
2010	24 459	-122	1,00	-0,50
2011	39 706	15 247	1,62	62,34
2012	38 804	-902	0,98	-2,27
2013	20 234	-18 570	0,52	-47,86
2014	41 809	21 575	2,07	106,63
2015	38 016	-3 793	0,91	-9,07
2016	34 258	-3 758	0,90	-9,89
2017	29 295	-4 963	0,86	-14,49
2018	53 290	23 995	1,82	81,91

Для обчислення параметрів функції на основі вимог методу найменших квадратів складаємо систему нормальних рівнянь:

$$\begin{cases} na_0 + a_1 \sum t = \sum y \\ a_0 \sum t + a_1 \sum t^2 = \sum yt \end{cases} \quad (11.31)$$

Цю систему можна спростити, якщо провести кодування часу таким чином, що б $\sum t = 0$. Тоді система рівнянь матиме вигляд:

$$\begin{cases} na_0 = \sum y \\ a_1 \sum t^2 = \sum yt \end{cases} \quad (11.32)$$

звідки:

$$a_0 = \frac{\sum y}{n}, a_1 = \frac{\sum yt}{\sum t^2}. \quad (11.33)$$

Приклад кодування для парної кількості членів ряду динаміки наводиться в табл. 11.5. Якщо кількість рівнів ряду непарна, то центральний рівень ряду кодується як 0.

Розрахунок рівняння тренду

Роки	Валовий збір зернових, тис. т (y)	t_i	t_i^2	$t_i y_i$	y_{t_i}
2001	38 674	-17	289	-657 458	33 864
2002	38 537	-15	225	-578 055	33 953
2003	45 623	-13	169	-593 099	34 043
2004	35 497	-11	121	-390 467	34 132
2005	33 930	-9	81	-305 370	34 222
2006	24 571	-7	49	-171 997	34 311
2007	35 472	-5	25	-177 360	34 400
2008	26 471	-3	9	-79 413	34 490
2009	24 581	-1	1	-24 581	34 579
2010	24 459	1	1	24 459	34 668
2011	39 706	3	9	119 118	34 758
2012	38 804	5	25	194 020	34 847
2013	20 234	7	49	141 638	34 937
2014	41 809	9	81	376 281	35 026
2015	38 016	11	121	418 176	35 115
2016	34 258	13	169	445 354	35 205
2017	29 295	15	225	439 425	35 294
2018	53 290	17	289	905 930	35 383
Σ	623 227	0	1938	86 601	623227

Обчислюємо параметри а:

$$a_0 = 623\,227/18 = 34\,623,72;$$

$$a_1 = 86\,601/1938 = 44,686.$$

За обчисленими параметрами проводимо синтезування трендової моделі функції:

$$y_t = 34\,623,72 + 44,686t.$$

Далі визначаємо теоретичні рівні Y_t за кожен рік (замість t підставляємо відповідне значення коду року):

$$Y_{2001} = 34\,623,72 + 44,686 \times (-17) = 33\,864;$$

$$Y_{2002} = 34\,623,72 + 44,686 \times (-15) = 33\,953,$$

і так за прикладом проводяться розрахунки для кожного року.

$$Y_{2018} = 34\,623,72 + 44,686 \times 17 = 35\,383.$$

Правильність розрахунків перевіряємо за рівністю сум:

$$\sum y_i = \sum y_{t_i}.$$

Параметри даного рівняння регресії можна інтерпретувати в такий спосіб: $a_0 = 34\,623,72$ тис. т – це початковий рівень валового збору зернових за розглянутий період; параметр a_1 , трендової моделі показує, що валовий збір зернових в Україні зростає у середньому на 44,686 тис. т у рік.

Для параболі другого порядку $\bar{y}_t = a_0 + a_1 t + a_2 t^2$ за умови використання способу умовного позначення часу, коли $\sum t = 0$, параметри цієї функції знаходять за формулами:

$$a_0 = \frac{\sum t^4 \sum y - \sum t^2 \sum t^2 y}{n \sum t^4 - \sum t^2 \sum t^2}, \quad (11.34)$$

$$a_1 = \frac{\sum yt}{\sum t^2}; \quad (11.35)$$

$$a_2 = \frac{n \sum t^2 y - \sum t^2 \sum y}{n \sum t^4 - \sum t^2 \sum t^2}. \quad (11.36)$$

У результаті знаходження параметрів отримуємо:

$$y = 29\,437,64 + 44,685t + 48,167t^2.$$

Для параболи третього порядку $\bar{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$ за умови використання способу умовного позначення часу, коли $\sum t = 0$, параметри цієї функції знаходять за формулами:

$$a_0 = \frac{\sum t^4 \sum y - \sum t^2 \sum t^2 y}{n \sum t^4 - \sum t^2 \sum t^2}; \quad (11.37)$$

$$a_1 = \frac{\sum t^6 \sum ty - \sum t^4 \sum t^3 y}{\sum t^2 \sum t^6 - \sum t^4 \sum t^4}; \quad (11.38)$$

$$a_2 = \frac{n \sum t^2 y - \sum t^2 \sum y}{n \sum t^4 - \sum t^2 \sum t^2}; \quad (11.39)$$

$$a_3 = \frac{\sum t^2 \sum t^3 y - \sum t^4 \sum ty}{\sum t^2 \sum t^6 - \sum t^4 \sum t^4}. \quad (11.40)$$

Як результат отримання параметрів маємо:

$$y = 29\,437,64 - 60,729t + 48,167t^2 + 0,546t^3.$$

Для знаходження рівняння гіперболи $\hat{y}_t = a_0 + \frac{a_1}{t}$ кодування часу, за якого $\sum t = 0$, неефективне та можливе лише в тому випадку, коли ряд парний, тобто не містить значення часу рівне 0. Параметри рівняння знаходять із такої системи рівнянь:

$$\begin{cases} \sum y = a_0 n + a_1 \sum \frac{1}{t} \\ \sum y \frac{1}{t} = a_0 \sum \frac{1}{t} + a_1 \sum \frac{1}{t^2} \end{cases},$$

звідки:

$$a_0 = \frac{\sum y \sum \frac{1}{t^2} - \sum y \frac{1}{t} \sum \frac{1}{t}}{n \sum \frac{1}{t^2} - \sum \frac{1}{t} \sum \frac{1}{t}}; \quad (11.41)$$

$$a_1 = \frac{n \sum y \frac{1}{t} - \sum y \sum \frac{1}{t}}{n \sum \frac{1}{t^2} - \sum \frac{1}{t} \sum \frac{1}{t}}. \quad (11.42)$$

Отже, для знаходження оцінок параметрів a_0 , a_1 необхідно знайти чотири суми:

$$\sum y; \sum y \frac{1}{t}; \sum \frac{1}{t}; \sum \frac{1}{t^2}.$$

Дані для проведення розрахунку наведені в табл. 11.6.

Таблиця 11.6

Визначення даних для розрахунку параметрів рівняння гіперболи

Роки	Валовий збір зернових, тис. т (y)	t	1/t	1/t ²	y/t	Y _t
2001	38 674	-17	-0,05882	0,00346	-2 274,94	34 506
2002	38 537	-15	-0,06667	0,00444	-2 569,13	34 491
2003	45 623	-13	-0,07692	0,00592	-3 509,46	34 470
2004	35 497	-11	-0,09091	0,00826	-3 227,00	34 442
2005	33 930	-9	-0,11111	0,01235	-3 770,00	34 402
2006	24 571	-7	-0,14286	0,02041	-3 510,14	34 339
2007	35 472	-5	-0,20000	0,04000	-7 094,40	34 225
2008	26 471	-3	-0,33333	0,11111	-8 823,67	33 959
2009	24 581	-1	-1,00000	1,00000	-24 581,00	32 629
2010	24 459	1	1,00000	1,00000	24 459,00	36 618
2011	39 706	3	0,33333	0,11111	13 235,33	35 289
2012	38 804	5	0,20000	0,04000	7 760,80	35 023
2013	20 234	7	0,14286	0,02041	2 890,57	34 909
2014	41 809	9	0,11111	0,01235	4 645,44	34 845
2015	38 016	11	0,09091	0,00826	3 456,00	34 805
2016	34 258	13	0,07692	0,00592	2 635,23	34 777
2017	29 295	15	0,06667	0,00444	1 953,00	34 757
2018	53 290	17	0,05882	0,00346	3 134,71	34 741
Σ	623 227	0	0,00000	2,41190	4 810,3403	623 227

Знаходимо параметри a_0 , a_1 і синтезуємо трендову модель:

$$Y_t = 34\,623,7 + 1994,42 / t.$$

Після побудови рівняння регресії, розраховуємо помилку апроксимації (σ) для наших трендів (табл. 11.7).

Таблиця 11.7

**Визначення даних для розрахунку помилки апроксимації
рівнянь лінійного тренду та параболи**

Роки	Валовий збір зернових, тис. т (y)	Теоретичні рівні лінійного тренду (y_t)	Квадрати відхилень фактичних рівнів від теоретичних ($(y_t - y_i)^2$)	Теоретичні рівні параболи 2-го порядку (y_t)	Квадрати відхилень фактичних рівнів від теоретичних ($(y_t - y_i)^2$)
2001	38 674	33 864	23 136 100	3 8674	1540 1838
2002	38 537	33 953	21 013 056	3 8537	114 0925
2003	45 623	34 043	134 096 400	4 5623	74 406 018
2004	35 497	34 132	1 863 225	3 5497	522 125
2005	33 930	34 222	85 264	3 3930	985 906
2006	24 571	34 311	94 867 600	2 4571	47 804 356
2007	35 472	34 400	1 149 184	3 5472	25 538 768
2008	26 471	34 490	64 304 361	2 6471	10 667 375
2009	24 581	34 579	99 960 004	2 4581	23 620 794
2010	24 459	34 668	104 223 681	2 4459	25 720 055
2011	39 706	34 758	24 482 704	3 9706	94 105 339
2012	38 804	34 847	15 657 849	3 8804	63 023 467
2013	20 234	34 937	216 178 209	2 0234	141 055 291
2014	41 809	35 026	46 009 089	4 1809	65 085 917
2015	38 016	35 115	8 415 801	3 8016	5 100 801
2016	34 258	35 205	896 809	3 4258	15 2173 13
2017	29 295	35 294	3 5988 001	2 9295	135 739 128
2018	53 290	35 383	320 660 649	5 3290	84 128 641
Σ	623 227	623 227	121 2987 986	62 3227	829 264 056
δ	x	x	8 209	x	6 788

Визначення даних для розрахунку помилки апроксимації рівнянь параболи третього порядку та гіперболи наведено в табл. 11.8.

**Визначення даних для розрахунку помилки апроксимації
рівнянь параболи третього порядку й гіперболи**

Роки	Валовий збір зернових, тис. т (y)	Теоретичні рівні параболи 3-го порядку (y_t)	Квадрати відхилень фактичних рівнів від теоретичних ($(y_t - y_i)^2$)	Теоретичні рівні гіперболи (y_t)	Квадрати відхилень фактичних рівнів від теоретичних ($(y_t - y_i)^2$)
2001	38 674	41 707	9 199 905	34 506	17 368 860
2002	38 537	39 343	6495 87	34 491	16 372 050
2003	45 623	37 168	71 495 156	34 470	124 382 595
2004	35 497	35 207	84 099	34 442	11 121 57
2005	33 930	33 488	195 689	34 402	222 898
2006	24 571	32 036	55 720 694	34 339	95 410 023
2007	35 472	30 877	21 112 091	34 225	15 554 11
2008	26 471	30 039	12 727 713	33 959	56 068 892
2009	24 581	29 546	246 51 140	32 629	64 775 210
2010	24 459	29 426	24 667 371	36 618	147 844 677
2011	39 706	29 704	100 045 762	35 289	19 514 059
2012	38 804	30 406	70 518 546	35 023	14 298 943
2013	20 234	31 560	128 280 766	34 909	215 345 030
2014	41 809	33 191	74 272 414	34 845	48 492 781
2015	38 016	35 325	7 241 902	34 805	1 0310 310
2016	34 258	37 989	13 916 824	34 777	269 505
2017	29 295	41 208	141 916 832	34 757	29 829 985
2018	53 290	45 009	68 571 377	34 741	344 063 884
Σ	623 227	623 227	825 267 868	623 227	1 207 237 269
\bar{y}	x	x	6 771	x	8 190

Фактичні та вирівняні за прямою і параболою третього порядку дані подані графічно на рис. 11.2.

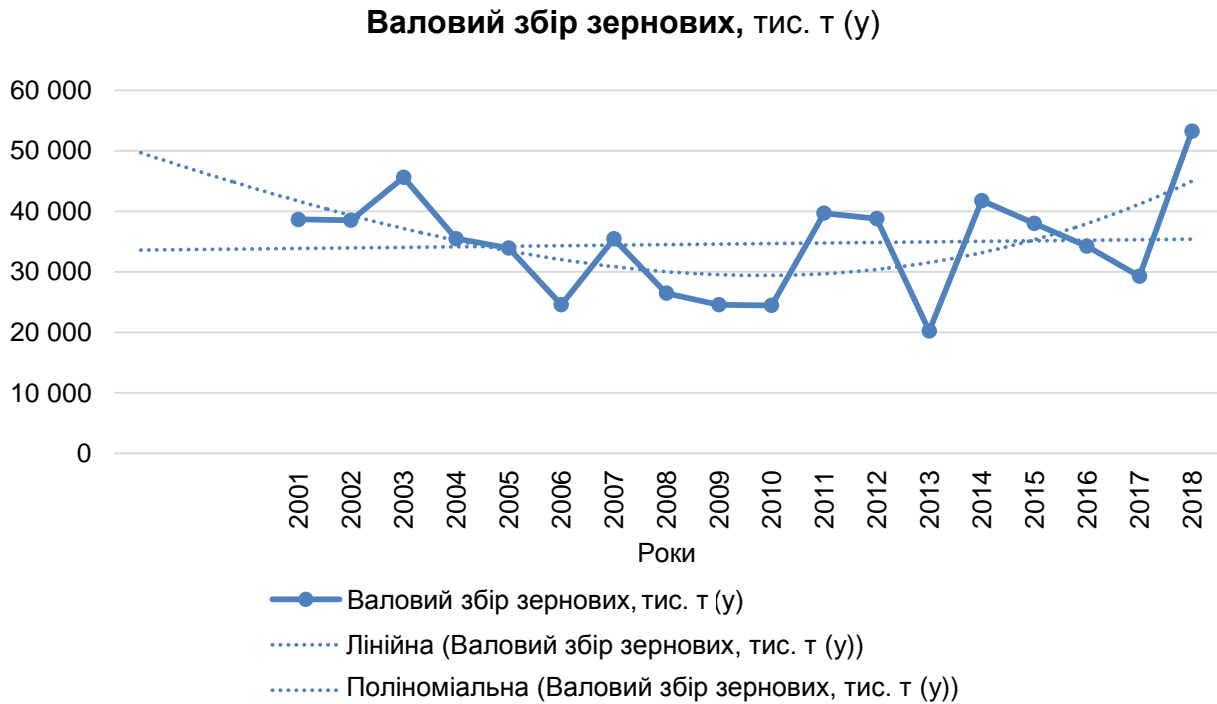


Рис. 11.2. Динаміка валового збору зернових в Україні 2001 – 2017 рр.

Підсумкові дані – графіками та значенням стандартної помилки апроксимації свідчать, що найадекватнішою моделлю є парабола третього порядку.

11.3. Факторний аналіз рядів динаміки

Важливе місце у вивченні динаміки соціально-економічних явищ належить **факторному аналізу**, метою якого є дослідження впливу окремих факторів на кількісні та якісні зміни явища в часі.

Для здійснення факторного аналізу рядів динаміки статистика використовує ряд методів і прийомів, а саме: приведення рядів динаміки до однієї основи, порівняння кількох паралельних рядів результативних і факторних показників, укрупнення періодів, розчленування досліджуваної сукупності на якісно однорідні групи та підгрупи. Тобто потрібні побудова простих і комбінаційних групувань, застосування дисперсійного та кореляційного методів аналізу. Якщо є необхідність у порівнянні відносної швидкості зміни (темлів зростання) різних явищ або показників, то найбільш поширеним є **метод приведення рядів динаміки до однієї основи**.

Для цього показники рядів динаміки виражають у % до першого рівня ряду (розраховуються базисні темпи росту), а потім обчислюють коефіцієнти випередження або відставання.

За допомогою даних коефіцієнтів порівнюють відносну швидкість динамічних рядів однакової змістовності щодо різних об'єктів (регіони, країни тощо) або різної – за одним об'єктом. Наприклад, за три роки фондоозброєність праці в одній галузі зросла на 50 %, в іншій – на 25 %. Коефіцієнт випередження темпу зростання фондоозброєності праці в першій галузі порівняно з другою становить $1,50 : 1,25 = 1,20$.

Можна порівняти динаміку фондоозброєності та продуктивності праці в кожній галузі. Якщо фондоозброєність зросла на 25 %, а продуктивність праці – на 37,5 %, то коефіцієнт випередження зростання продуктивності праці становить $1,375 : 1,250 = 1,10$.

Щодо темпів приросту, то їх співвідношення використовують лише для взаємозв'язаних показників x і y . Таке співвідношення називають **емпіричним коефіцієнтом еластичності** [68]; він показує, на скільки відсотків змінюється y зі зміною x на 1 %. Наприклад, ціна на товар А зросла на 2 %, а попит зменшився на 4 %. Цінова еластичність попиту на цей товар склала $-4/+2 = -2$, тобто зі зростанням цін на 1 % попит на товар зменшується на 2 %.

Фактичні рівні динамічних рядів під впливом різного роду факторів варіюють, відхиляючись від основної тенденції розвитку. В одних рядах коливання мають систематичний, закономірний характер, повторюються через певні інтервали часу; в інших – не мають такого характеру, тому їх називають **випадковими**. У конкретному ряду можуть поєднуватися систематичні та випадкові коливання.

Найпростішою оцінкою систематичних коливань є **коефіцієнти нерівномірності**, які обчислюються відношенням максимального та мінімального рівнів динамічного ряду до середнього. Чим більша нерівномірність процесу, тим більша різниця між цими двома коефіцієнтами [13; 16; 68; 71].

Приклад 11.3. Споживання питної води за добу становить $7\ 200\ \text{м}^3$, у середньому за годину $7\ 200 : 24 = 300\ \text{м}^3$. Найбільший рівень споживання води в період від 20.00 до 21.00 – $381\ \text{м}^3$, найменший – у період від 2.00 до 3.00 – $165\ \text{м}^3$.

Коефіцієнти нерівномірності такі:

$$K_{\max} = 381 : 300 = 1,27;$$

$$K_{\min} = 165 : 300 = 0,55.$$

Амплітуда коливань у розмірі 72 пункти (1,27 – 0,55) свідчить про істотну нерівномірність споживання води протягом доби [67].

11.4. Аналіз сезонних коливань

Основними завданнями статистичного вивчення коливань соціально-економічних процесів є такі:

вимірювання сили коливань;

вивчення типу коливань;

розкладання складних коливань на різномірні складові;

дослідження змін коливань у часі та динаміки коливань;

вивчення варіації коливань у просторовій або іншій сукупності об'єктів;

вивчення факторів коливань та їх статистико-математичне моделювання.

Окремим соціально-економічним процесам притаманні внутрішньорічні, сезонні піднесення і спади. Наприклад, виробництво й переробка сільськогосподарської продукції, нерівномірне завантаження транспорту, коливання попиту на товари тощо. **Сезонні коливання** виявляються і аналізуються на основі рядів щомісячних або щоквартальних даних.

Характер сезонних коливань описується "**сезонною хвилею**", яку утворюють індекси сезонності. У динамічних рядах, які не виявляють чіткої тенденції розвитку, **індекси сезонності** є відношенням фактичних місячних (квартальних) рівнів y_t до середньомісячного (середньо-квартального) за рік \bar{y} , %, визначається за формулою [68]:

$$I_c = 100 \frac{y_t}{\bar{y}}. \quad (11.43)$$

Залежно від способу вирівнювання вихідних даних розрізняють **методи розрахунку індексу сезонності**:

за простою середньою (спосіб постійної середньої);

за ковзною середньою;

аналітичним вирівнюванням (спосіб змінної середньої).

Індекси сезонності показують, у скільки разів фактичний рівень ряду в момент або інтервал часу і більший за середній рівень.

Спосіб постійної середньої застосовують для рядів з невираженою основною тенденцією розвитку (тренд відсутній) за формулою [68; 71]:

$$I_{\text{сез}} = \frac{Y_i}{\bar{y}}. \quad (11.44)$$

Спосіб змінної середньої застосовують для рядів із вираженою основною тенденцією розвитку й обчислюють за формулою:

$$I_{\text{сез}} = \frac{Y_i}{Y_t}. \quad (11.45)$$

В аналізі сезонності рівні часового ряду показують розвиток явища за місяцями (кварталами) одного або декількох років. На сезонні коливання можуть накладатися випадкові відхилення, для їхнього усунення проводять усереднення індивідуальних індексів однойменних внутрішньорічних періодів аналізованого ряду динаміки. Тому для кожного періоду річного циклу визначають узагальнені показники у вигляді середніх індексів сезонності:

$$\bar{I}_s = \frac{\sum i_{\text{сез}}}{n}. \quad (11.46)$$

Сукупність середніх індексів сезонності однойменних періодів створює модель сезонної хвилі. Якщо при побудові моделі сезонної хвилі випадкові коливання гасяться, то сума середніх індексів сезонності однойменних періодів = 1 200 % (якщо рівні бралися за місяць) і 400 %, якщо рівні були кварталними. Якщо умова не виконана проводять коректування моделі. Для цього розраховують коригувальний коефіцієнт:

$$K_{\text{кор}} = \frac{1\,200}{\sum I_{s(\text{середнє})}} \text{ – для місячних даних;}$$

$$K_{\text{кор}} = \frac{400}{\sum I_{s(\text{середнє})}} \text{ – для кварталних даних.}$$

На величину цього коефіцієнта коректують усі розраховані середні індекси сезонності.

Приклад 11.4. Розглянемо порядок обчислення сезонної хвилі на прикладі споживання електроенергії комунальним господарством регіону (табл.11.9).

Таблица 11.9

Щомісячна динаміка споживання електроенергії

Місяці року	Спожито електроенергії, y_t , млн кВт год	Індекс сезонності I_c , %	$I_c - 100$	$(I_c - 100)^2$
Січень	172	111,7	11,7	136,89
Лютий	161	104,5	4,5	20,25
Березень	158	102,6	2,6	6,76
Квітень	151	98,0	-2,0	4,00
Травень	147	95,5	-4,6	20,25
Червень	130	84,4	-15,6	243,36
Липень	124	80,5	-19,5	380,25
Серпень	146	94,9	-5,1	26,01
Вересень	149	96,8	-3,2	10,24
Жовтень	155	100,6	0,6	0,36
Листопад	168	109,1	9,1	82,81
Грудень	187	121,4	21,4	457,96
Разом	1848	100	0	1389,14

Середньомісячний обсяг споживання $\bar{y} = \frac{1848}{12} = 154$ (млн кВт год).

Індекси сезонності коливаються від 121,4 % у грудні ($\frac{187}{154} \times 100$)

до 80,5 % у липні ($\frac{124}{154} \times 100$). Амплітуда сезонних коливань становить

$R_t = 121,4 - 80,5 = 40,9$. Характер сезонної хвилі схематично ілюструє рис. 11.3.

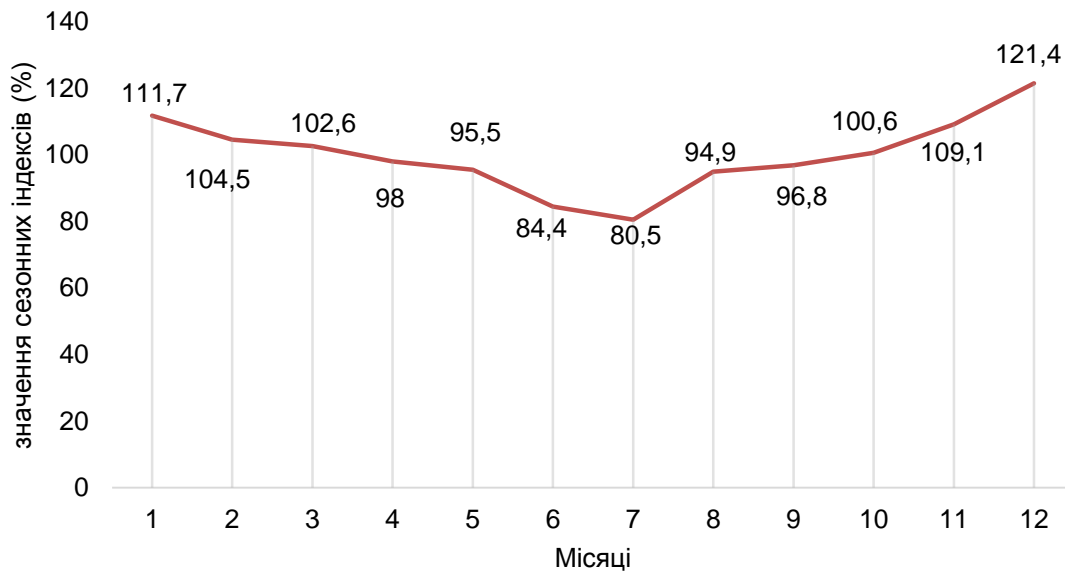


Рис. 11.3. Сезонна хвиля споживання електроенергії

Для порівняння інтенсивності сезонних коливань різних явищ чи того самого явища в різні роки використовують узагальнювальні характеристики варіації індексів сезонності:

$$\text{середнє лінійне відхилення } \bar{I}_c = \frac{1}{12} \sum_{1}^{12} |I_c - 100|;$$

$$\text{або середнє квадратичне відхилення } \sigma_t = \sqrt{\frac{1}{12} \sum_{1}^{12} (I_c - 100)^2}.$$

У динамічному ряду споживання електроенергії (див. табл. 11.8) середнє квадратичне відхилення:

$$\sigma_t = \sqrt{\frac{1389,14}{12}} = 10,8 \text{ п. п.}$$

Якщо спостерігається тенденція розвитку, попередньо проводиться згладжування чи вирівнювання динамічного ряду, визначаються теоретичні рівні для кожного місяця (кварталу) року, а індекс сезонності обчислюється як відношення фактичних рівнів ряду y_t до теоретичних Y_t , тобто

$$I_c = 100 \frac{y_t}{Y_t}.$$

Важливі поняття

Амплітуда (розмах коливань) – різниця між алгебраїчним, найбільшим за період відхиленням від тренду, та найменшим алгебраїчним відхиленням.

Аналітичне вирівнювання – підбір для даного ряду динаміки теоретичної кривої, що виражає основні риси динаміки фактичного явища, тобто щонайкраще описує емпіричні дані.

Вирівнювання часового ряду укрупненням інтервалів – перетворення рівнів ряду в ряди більш тривалих періодів.

Екстраполяція – поширення виявлених в аналізі рядів динаміки закономірностей розвитку досліджуваного явища на майбутнє.

Зважена ковзна середня – плинна середня, розрахована таким чином, що рівням, які входять в інтервал усереднення, надають різні ваги, оскільки апроксимація в межах інтервалу згладжування здійснюється з використанням рівнів, розрахованих за поліномом n -го порядку.

Згладжування – методи виявлення основної тенденції ряду динаміки.

Індекс сезонності – індикатор сезонних коливань, що показує, у скільки разів фактичний рівень ряду в момент або інтервал часу і більший за середній рівень.

Інтервальний прогноз – діапазон числових значень, що приблизно містить прогнозоване значення рівня.

Коефіцієнт нерівномірності – відношення максимального та мінімального рівнів динамічного ряду до середнього.

Коефіцієнт стійкості – величина, яка є доповненням коефіцієнта коливань до одиниці.

Колівання рівнів динамічних рядів – відхилення рівнів ряду від тренду.

Метод усереднення за лівою та правою половиною – метод, за якого ряд динаміки ділять на дві частини, знаходять для кожної з них середнє арифметичне значення і проводять через отримані точки лінію тренду на графіку.

Період коливання – тривалість циклу.

Прогнозування – припущення, що закономірність, яка діє усередині ряду динаміки та виступає як база прогнозування, збережеться надалі.

Проста ковзна (плинна) середня – середня арифметична величина, обрахунок якої ведеться з поступовим виключенням із даного інтервалу згладжування першого рівня ряду динаміки та включенням наступного.

Рівномірний розвиток – тип динаміки з постійними абсолютними приростами.

Рівноприскорений (рівновповільнений) розвиток – тип динаміки з постійними з постійними темпами приросту.

Розвиток за експонентою – тип динаміки зі стабільними темпами зростання.

Розвиток зі змінним прискоренням (уповільненням) – тип динаміки, що описують за допомогою параболи третього порядку.

Розвиток з уповільненням зростання наприкінці періоду – тип динаміки за якого величина ланцюгових абсолютних приростів скорочується у кінцевих рівнях ряду динаміки.

Сезонні коливання – повторювані стійкі внутрішньорічні коливання.

Тенденція (тренд) – напрям розвитку, основна довгострокова тенденція розвитку динамічного ряду.

Точковий прогноз – конкретне числове значення рівня в прогнозованій період.

Частота – кількість циклів в одиницю часу.

Типові завдання

Завдання 11.1. Маємо дані про прибуток підприємства (табл. 11.10). Необхідно визначити відсутній рівень динамічного ряду та зробити висновки.

Таблиця 11.10

Дані про прибуток підприємства

Роки	2012	2013	2014	2015	2016	2017	2018
Прибуток підприємства, млн грн	3,75	4,05	–	4,27	4,82	4,99	5,12

Розв'язання.

1. Знайдемо недостатній рівень за методом усереднення найближчих значень:

$$Y_i = \frac{Y_{i-1} + Y_{i+1}}{2};$$

$$Y_{2014} = \frac{4,05 + 4,27}{2} = 4,16 \text{ (млн грн).}$$

2. Знайдемо недостатній рівень за методом середнього абсолютного приросту:

$$Y_i = Y_{i-1} + \bar{\Delta};$$

$$\bar{\Delta} = \frac{Y_n - Y_1}{n-1} = \frac{5,12 - 3,75}{7-1} = 0,23;$$

$$Y_{2014} = 4,05 + 0,23 = 4,28 \text{ (млн грн).}$$

3. Знайдемо недостатній рівень за методом середнього темпу росту:

$$Y_i = Y_{i-1} \times \bar{T}_p;$$

$$\bar{T}_p = \sqrt[n-1]{\frac{Y_n}{Y_1}} = \sqrt[7-1]{\frac{5,12}{3,75}} = 1,053.$$

$$Y_{2014} = 4,05 \times 1,053 = 4,26.$$

Отже, розрахунок недостатнього рівня динамічного ряду трьома способами показав, що рівень прибутку підприємства в 2014 р. перебував у межах від 4,16 до 4,28 млн грн.

Завдання 11.2. Маємо дані про динаміку продажів туристичних путівок (табл. 11.11).

Таблиця 11.11

Динаміка продаж туристичних путівок

Квартали	Продано туристичних путівок		
	2016	2017	2018
I	170	161	159
II	175	215	242
III	190	191	207
IV	171	186	195

Для аналізу динаміки продажу туристичних путівок серед року необхідно визначити індекс сезонності та зробити висновки.

Розв'язання.

1. Визначимо темпи росту за роками (табл. 11.12).

Таблиця 11.12

Розрахункова таблиця

Рік	Кількість проданих туристичних путівок	Темпи росту, %	
		ланцюгові	базисні
2016	706	–	–
2017	753	106,66	106,66
2018	803	106,4	113,74

Бачимо, що ряд має чітку тенденцію до зростання. Для аналізу динаміки серед року в рядах із тенденцією до зростання, вивчення сезонності спирається на метод ковзної середньої:

$$i_{\text{Сез}} = \frac{y_i}{Y_i}$$

Для кожного періоду річного циклу визначають узагальнені показники у вигляді середніх індексів сезонності:

$$\bar{i}_s = \frac{\sum_{i=1}^n \sum i_{\text{ces}_i}}{n}$$

2. Визначимо теоретичні значення (y_t). Для цього розрахуємо параметри a_0 та a_1 . Побудуємо допоміжну таблицю (табл. 11.13).

Таблиця 11.13

Допоміжна таблиця

Періоди	Фактичні значення, y_i	t	t^2	$y_i \cdot t$
1	2	3	4	5
2016 р.				
I	170	-5,5	30,25	-935
II	175	-4,5	20,25	-787,5
III	190	-3,5	12,25	-665
IV	171	-2,5	6,25	-427,5

1	2	3	4	5
2017 р.				
I	161	-1,5	2,25	-241,5
II	215	-0,5	0,25	-107,5
III	191	0,5	0,25	95,5
IV	186	1,5	2,25	279
2018 р.				
I	159	2,5	6,25	397,5
II	242	3,5	12,25	847
III	207	4,5	20,25	931,5
IV	195	5,5	30,25	1 072,5
Разом	2 262	0	143	459

Таким чином:

$$a_0 = \frac{2\,262}{12} = 188,5;$$

$$a_1 = \frac{459}{143} = 3,21.$$

Рівняння прямої буде мати вигляд: $y_t = 188,5 + 3,21 \times t$.

3. Визначимо вирівняні значення ряду та розрахуємо індекси сезонності (табл. 11.14).

Таблиця 11.14

Розрахунок вирівняних значень ряду та індексу сезонності

Квартали	Фактичні значення			Вирівняні значення			Фактичні значення в % до вирівняних			Сума відсоткових відношень	Індекс сезонності
	2016	2017	2018	2016	2017	2018	2016	2017	2018		
I	170	161	159	170,8	183,7	196,5	99,5	86,2	80,9	266,6	88,9
II	175	215	242	174,1	186,9	199,7	100,5	115	121,2	336,7	112,2
III	190	191	207	177,3	190,1	202,9	107,2	100,4	102	309,6	103,2
IV	171	186	195	180,5	193,3	206,2	94,7	96,2	94,6	285,5	95,2

Отже, найбільша питома вага продажу туристичних путівок припадає на другий квартал.

Завдання 11.3. Необхідно побудувати лінійну трендову модель залежності видобутку нафти в регіоні від фактору часу.

Розв'язання.

Порядок обчислення параметрів лінійної функції розглянемо на прикладі динамічного ряду видобутку нафти в регіоні (табл. 11.15).

Таблиця 11.15

Динаміка видобутку нафти

Роки	y_t , млн т	Δ_t	Змінна часу, t	$y_t t$	$Y_t = 74,5 + 3,8t$
2012	63,5	–	–3	–190,5	63,1
2013	66,8	3,3	–2	–133,6	66,9
2014	71,0	4,2	–1	–71,0	70,7
2015	74,3	3,3	0	0	74,5
2016	76,9	2,6	1	76,9	78,3
2017	82,2	5,3	2	164,4	82,1
2018	86,8	4,6	3	260,4	85,9
Разом	521,5	×	0	106,6	521,5

Ланцюгові абсолютні прирости динамічного ряду практично стабільні, тому тенденцію можна описати лінійною функцією. Оскільки довжина ряду $n = 7$, то $\sum t^2 = 7(7^2 - 1) : 12 = 28$. Параметри трендового рівняння становлять:

$$a = \sum y_t : n = 521,5 : 7 = 74,5;$$

$$b = \sum y_t t : \sum t^2 = 106,6 : 28 = 3,8.$$

Лінійний тренд має вигляд $Y_t = 74,5 + 3,8t$, тобто середній рівень видобутку нафти становить 74,5 млн т, середньорічний приріст видобутку – 3,8 млн т.

В останній графі табл. 11.15 для кожного року наведено теоретичні рівні Y_t , тобто очікувані рівні видобутку нафти в t -му році, зумовлені дією основних факторів розвитку галузі: для 2012 р. $Y_1 = 74,5 + 3,8(-3) = 63,1$ млн т, для 2013 р. $Y_2 = 74,5 + 3,8(-2) = 66,9$ млн т і т. д.

Суми фактичних рівнів $\sum y_t$ і розрахованих за лінійним трендом теоретичних рівнів $\sum Y_t$ однакові: $\sum y_t = \sum Y_t = 521,5$ млн т.

Продовження виявленої тенденції за межі ряду динаміки називають **екстраполяцією тренду**. Це один із методів статистичного прогнозування, передумовою для використання якого є незмінність причинного комплексу, що формує тенденцію. Прогнозний, очікуваний рівень Y_{t+v} залежить від бази прогнозування та періоду упередження v . Так, припускаючи, що умови, в яких формувалась тенденція видобутку нафти, найближчим часом не зміняться, визначимо прогноз на 2020 р. Базою прогнозування є теоретичний рівень 2018 р., період упередження $v = 2$. Очікуваний в 2020 р. видобуток нафти досягне 93,5 млн т:

$$Y_{t+v} = 85,9 + 3,8 \times 2 = 93,5.$$

Метод екстраполяції дає точковий прогноз. На практиці, як правило, визначають довірчі межі прогнозного рівня $Y_{t+v} \pm ts_p$, де s_p – стандартна похибка прогнозу, t -квантиль розподілу Стьюдента.

Завдання 11.4. На підставі даних щодо розподілу путівок за місяцями в регіоні (табл. 11.16) необхідно визначити показники, які характеризують форму сезонних коливань та показники сили сезонних коливань.

Таблиця 11.16

Дані про розподіл путівок за місяцями в регіоні

Місяці	1	2	3	4	5	6	7	8	9	10	11	12
Кількість путівок	1 994	1 870	1 860	1 980	1 995	2 010	2 030	2 500	2 400	2 350	2 120	2 030

Розв'язання.

З метою визначення показників, які характеризують форму сезонних коливань, доцільно побудувати таблицю розрахунків (табл. 11.17).

Форму сезонних коливань характеризують абсолютні відхилення від середньомісячного рівня та індекси сезонних коливань.

Абсолютні відхилення знайдено шляхом вирахування від місячних рівнів середньомісячного рівня: $1\ 994 - 2\ 095 = -101$; $1\ 870 - 2\ 095 = -225$ і т. д.

Розрахункова таблиця

Місяці	Кількість путівок	Відхилення від середньомісячного рівня	Квадрати відхилень	Індекси сезонних коливань, %
1	1 994	-101	10 201	95,18
2	1 870	-225	50 625	89,26
3	1 860	-235	55 225	88,78
4	1 980	-115	13 225	94,51
5	1 995	-100	10 000	95,23
6	2 010	-85	7 225	95,94
7	2 030	-65	4 225	96,90
8	2 500	405	164 025	119,33
9	2 400	305	93 025	114,56
10	2 350	255	65 025	112,17
11	2 120	25	625	101,19
12	2 030	-65	4 225	96,90
	2 5139		10 201	

Індекси сезонних коливань розраховуємо таким чином:

$$\frac{1994}{2095} \times 100\% = 95,2\%; \quad \frac{1870}{2095} \times 100\% = 89,3\%.$$

Форму сезонних коливань можна надати у графічному вигляді (рис. 11.4):

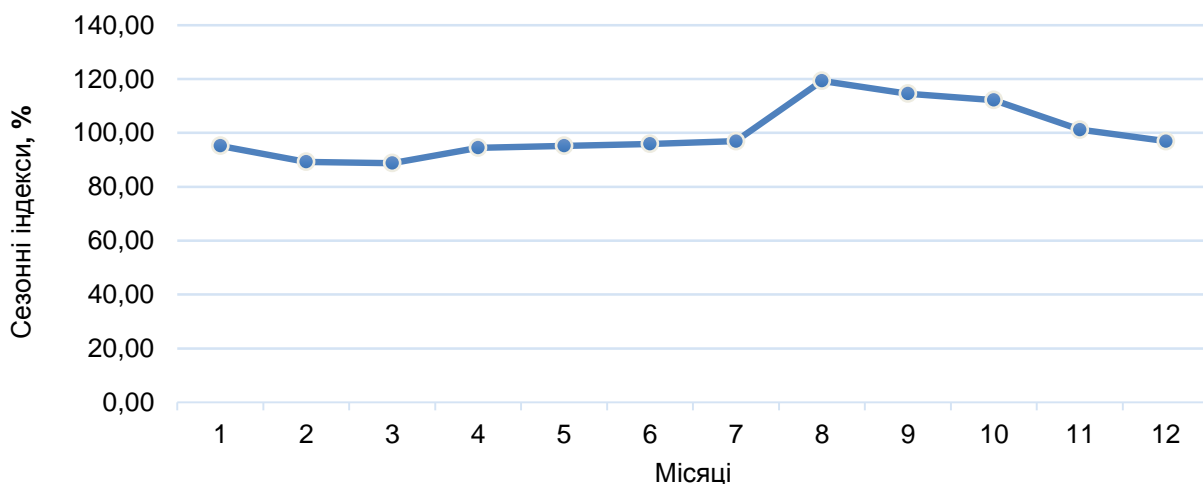


Рис. 11.4. Індекси сезонних коливань

Графік чітко показує, що збільшення попиту на путівки спостерігається з липня до жовтня, а в зимові місяці – зменшення.

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Аналіз тенденцій розвитку та коливань" подані у роботі [67]. Лабораторна робота націлена на придбання навиків використання екстраполяції та інтерполяції в рядах динаміки, побудови й аналізу моделей за допомогою MS Excel. Завдання лабораторної роботи: за допомогою MS Excel знайти і рівні динамічного ряду шляхом екстраполяції; побудувати прогноз та визначити, в яких межах перебуватиме прогнозне значення; побудувати різні моделі динаміки за даними часового ряду.

Запитання до самодіагностики:

1. Назвіть прийоми вирівнювання рядів динаміки.
2. У чому полягає сутність прийому приведення рядів динаміки до однієї основи та змикання рядів динаміки?
3. У чому полягає сутність і які умови для застосування прийому укрупнення періодів?
4. Що таке інтерполяція й екстраполяція рядів динаміки? У чому полягає їх значення і застосування?
5. Охарактеризуйте прийоми вирівнювання рядів динаміки способом ковзної середньої, за середнім абсолютним приростом і середнім коефіцієнтом зростання.
6. Як виконується прогноз на майбутнє за допомогою рівняння тренду?
7. Як виміряти сезонні коливання в рядах динаміки?
8. Як ви розумієте тенденцію розвитку? Наведіть приклади тенденції.
9. Яка різниця між згладжуванням і вирівнюванням динамічного ряду? Які методи використовують у тому й іншому випадку?
10. Опишіть методи інтерполяції динамічного ряду.

Питання для критичного переосмислення (есе)

1. Особливості дослідження динаміки соціально-економічних явищ.
2. Вимірювання сезонних коливань у рядах динаміки споживання населенням матеріальних благ.
3. Використання трендових моделей для прогнозування освітніх тенденцій в Україні та Європі.

4. Вимірювання сезонних коливань у сфері діяльності АПК.
5. Сезонність як один із визначальних факторів туристичного попиту.

12. Індексний метод

Основні питання:

- 12.1. Поняття про індекси та їх роль у статистико-економічному аналізі.
- 12.2. Агрегатний індекс як основна форма загального індексу. Середньозважені індекси.
- 12.3. Індекси зі змінними та постійними вагами.
- 12.4. Індекси середніх величин.
- 12.5. Територіальні індекси.
- 12.6. Індексний факторний метод аналізу.

12.1. Поняття про індекси та їх роль у статистико-економічному аналізі

У сучасній аналітичній практиці індекси розглядають як категорію, що відображає відносну зміну складних явищ, окремі елементи яких безпосередньо незіставні. Тому індекс, як особлива відносна величина, визначається методологією побудови та характером функцій.

У теорії дослідження індексів розглядають: синтетичну й аналітичну їх функції.

Згідно із **синтетичною** функцією, особливість індексів полягає у зведенні (агрегуванні) в ціле різнорідних одиниць статистичної сукупності.

Аналітичні властивості індексів дозволяють відстежити вплив окремих факторів на зміну розглядуваного явища.

Складність відображення зміни соціально-економічних явищ або процесів передбачає поєднання зазначених функцій.

Принципова відмінність індексів від відносних і середніх величин полягає в системному підході щодо дослідження явищ на основі об'єктивного взаємозв'язку між ними.

За системного підходу відносні величини неминуче перетворюються в показники, що характеризують динаміку зміни одиниці явища (індивідуальні індекси). За індивідуальними індексами обчислюються складні індекси, що відображають динаміку двох і більше взаємопов'язаних явищ.

Називаючи відносну величину індивідуальним індексом, статистик лише підкреслює, що ця величина призначена для розрахунку складного індексу.

Індекси не обов'язково визначають динаміку досліджуваних процесів. Нерідко вони використовуються для порівняння складних явищ у статистиці (територіальні індекси) або для оцінювання відхилення рівнів взаємопов'язаних явищ від деякого еталона чи оптимального варіанту.

Головне в теорії побудови індексів – перехід від аналізу кількісних відмінностей між елементами порівнюваних систем до аналізу кількісних відмінностей між системами у цілому.

Способи побудови індексів залежать від змістовності досліджуваного показника, методології розрахунку вихідних статистичних показників, наявності статистичних даних і мети дослідження. Індекси розраховують на вищій ступені статистичного узагальнення, результатом якого є зведення і оброблення даних [66; 84].

Отже, **індекс** – це відносна величина, що виходячи із системності розгляду явищ характеризує його зміну в часі та просторі або ступінь відхилення від певного стандарту (нормативу).

Назва індексу відбиває його змістовність, а числове значення – інтенсивність зміни або ступінь відхилення явища. Наприклад, індекс споживчих цін у звітному році становив 109,5 %.

Система показників, що використовується для побудови індексів має *умовні позначення*:

q – обсяг виробництва у натуральних одиницях вимірювання (фізичний обсяг реалізації);

p – ціна за одиницю;

Q або (**qp**) – вартість виробленої продукції (товарооборот);

z – собівартість одиниці продукції;

zq – загальні витрати на виробництво;

w – продуктивність праці;

f – середня заробітна плата;

T – чисельність працюючих (або відпрацьований час);

t – витрати часу на одиницю продукції.

Кожний індекс включає дані за два періоди: **звітний** (t_1) – поточний (порівнюваний) і **базисний** (t_0) – що використовується як база порівняння.

В **індексах динаміки** базою порівняння є *показник певного попереднього періоду* (моменту) часу, у **територіальних індексах** – *показник певного регіону* (об'єкта).

Індекс як відносний показник може приймати вид коефіцієнта (коли базовий рівень прийнято за одиницю) чи відсотка (коли базовий рівень прийнято за 100). Якщо індекс більше 1 (або 100 %), то рівень досліджуваного явища зростає, а коли менше за 1 (або 100 %) – то рівень явища буде знижуватися.

Будь-який індекс складається з таких *елементів*: індексована величина, тип (форма), вага індексів, строки розрахунку.

Залежно від:

індексованої величини – будують індекси фізичного обсягу продукції, індекси цін, індекси продуктивності праці та ін.;

форми індексів – розглядають агрегатні та середні індекси; залежно від форми середньої – середні арифметичні, середні гармонічні, середні геометричні та ін.;

вибору ваги індексу – відрізняють прості (незважені) та зважені індекси; серед зважених індексів – індекси з постійними (незмінними вагами) й індекси зі змінними вагами (які переглядаються у міру необхідності протягом певного часу);

строку розрахунку – розглядаються індекси базисні (з постійною, незмінною у часі базою порівняння) та ланцюгові індекси (зі змінною базою порівняння).

Методики побудови та розрахунку індексів для порівнянь у часі та просторі є однаковими.

Усі економічні індекси можна **класифікувати** за такими **ознаками**: база порівняння, ступінь охоплення елементів сукупності, форма побудови, характер об'єкта дослідження, вид статистичних ваг (сумірників), склад явища, строк розрахунку.

За базою порівняння розглядають індекси:

динамічні, що відображають зміну явища з урахуванням фактора часу (зміна вартості споживчого кошику населення країни у вересні порівняно із січнем поточного року);

територіальні, що відбивають результат порівняння в просторі (за різними об'єктами, регіонами, країнами). Наприклад, вартість споживчого кошику населення в Україні порівняно з відповідним показником у Болгарії;

міжгрупові, що характеризують відхилення явища від певного еталонного рівня (за базу порівняння приймається стандартне значення показника, мінімальне чи максимальне значення).

База порівняння має визначальне значення в обґрунтуванні показника індексу. Вибір бази порівняння визначається метою дослідження.

За ступенем охоплення елементів сукупності (рівнем агрегування інформації) індекси розподіляють на:

індивідуальні, що характеризують зміну в часі чи співвідношення в просторі одного індивідуального явища (зміна ціни на товар певного виду, зміна обсягу продажів товару певного виду);

зведені, що є співвідношенням рівнів складного явища, безпосередньо не вимірювані між собою (зміна кількості реалізованих товарів за різними товарними групами).

Зведені індекси, зі свого боку, підрозділяють на:

групові (субіндекси), якщо вивчається відносна зміна певної групи, що складає сукупність (зміна ціни на групу продовольчих або непродовольчих товарів);

загальні, якщо за наведених умов визначається зведений індекс за всією сукупністю (зміна ціни на групу продовольчих і непродовольчих товарів).

Обов'язковою умовою для визначення *індивідуального індексу* є максимальна однорідність того явища (об'єкта), для якого розраховується індекс.

Основним елементом індексного відношення є *величина*, що *індексується*, – значення ознаки статистичної сукупності, зміна якої є об'єктом вивчення [66; 68; 84].

Індивідуальні індекси позначають символом **(i)**. Так, наприклад:

індивідуальний індекс *фізичного обсягу* (i_q) реалізації товару (виробництва продукції):

$$i_q = \frac{q_1}{q_0}, \quad (12.1)$$

де q_1 і q_0 – обсяг продажу окремого виду товару в поточному та базисному періодах у натуральних вимірниках.

Цей індекс показує: у скільки разів зріс (зменшився) обсяг реалізації (випуску продукції) одиниці товару в звітному періоді порівняно з базисним; скільки відсотків складає зростання (зниження) обсягу реалізації певного товару (випуску продукції);

індивідуальний індекс *цін* (i_p):

$$i_p = \frac{p_1}{p_0}. \quad (12.2)$$

Цей індекс характеризує зміну ціни одиниці певного товару в поточному періоді (p_1) порівняно з базисним (p_0).

Зведені індекси позначають символом (I). Методика їх побудови є більш складною порівняно з індивідуальними.

Проблеми побудови зведених індексів стосуються:

зіставлення непорівнянних поміж собою чисельника та знаменника; вибору співмірника (показника, який дозволяє звести до порівняного виду сукупність різнорідних елементів) і ваги (показника, що визначає значущість співмірника в певній сукупності);

порядку індексації та фіксації співмірника та ваги.

Якщо порівнянність забезпечується за допомогою **агрегатів**, що порівнюються (добутку спряжених величин – співмірника та його ваги), то порядок індексації та фіксації ваг визначається відповідною системою побудови індексів: базисно-зваженою (Ласпейреса) чи поточно-зваженою (Пааше) [66; 68].

Групові індекси відображають закономірності в розвитку окремих частин розглядуваного явища. У таких індексах проявляється їх зв'язок методом групувань.

Залежно від *форми побудови* відрізняють індекси:

агрегатні, що є основною формою загальних індексів;

середньозважені, що залежно від форми середньої підрозділяють на:

арифметичні;

гармонічні.

Вибір форми розрахунку індексу залежить від:

мети дослідження;

економічної сутності аналізованого показника;

наявної інформації.

За характером об'єкта дослідження загальні індекси підрозділяють на індекси:

кількісних показників – у них міститься характеристика зміни обсягу певного явища, який виражають у відповідних одиницях вимірювання (індекс обсягу продажу доларів США на валютній біржі);

якісних показників – у них міститься характеристика зміни якісної ознаки, яка відображає особливості розвитку явища (індекс курсу євро).

В основу такого розподілу покладено вид індексованої величини.

За видами ваг розглядають індекси:

прості (незважені);

зважені, які, в свою чергу, підрозділяють на:

індекси з *постійними (незмінними) вагами*;

індекси з *змінними вагами* (що можуть переглядатися протягом певного часу).

Найбільш складним для побудови індексу є вибір ваги індексу та більш точний розрахунок ваги кожної групи, а іноді – кожної одиниці, яка входить у сукупність, що індексується. Система таких ваг повинна відбивати модель структури того соціально-економічного явища, динаміка якого знаходить числове вираження в індексі [66; 84]. Наприклад, ваги індексу цін повинні відбивати товарну структуру торгового обороту.

За складом явища розрізняють індекси:

змінного складу;

постійного (фіксованого) складу;

структурних зрушень.

За строком розрахунку розглядають індекси:

базисні (з постійною, незмінною за часом базою);

ланцюгові, де числові значення індексованої величини в кожний поточний строк зіставляються з їх значеннями в попередній строк (індекси зі змінною базою).

12.2. Агрегатний індекс як основна форма загального індексу.

Середньозважені індекси

Економічна змістовність індексу визначає методику його розрахунку.

Методика побудови агрегатного індексу передбачає розв'язання таких питань:

яка величина буде індексованою;

за яким складом різнорідних елементів явища необхідно розрахувати індекс;

що буде слугувати вагою для розрахунку індексу.

Залежно від характеру об'єкта дослідження визначають індекси об'ємних та індекси якісних показників.

До **групи об'ємних показників** відносять індекси фізичного обсягу продукції, національного доходу, роздрібного товарообороту, споживання та ін. За основу їх розрахунку взято величину показника обсягу.

До **групи якісних показників** належать відносяться індекси собівартості продукції, продуктивності праці, цін та ін. За основу їх розрахунку взято якісні показники.

З метою забезпечення зіставлення складових індексованих величин у розрахунок вводять показник вимірювання – **вагу**, що дозволяє забезпечити кількісну порівнянність за звітний та базисний періоди.

У виборі ваги індексу необхідно керуватися таким *правилом*: якщо будується індекс кількісного показника, то вага береться за базисний період; для побудови індексу якісного показника вага визначається за звітний період.

Найбільш типовим індексом кількісних показників є **індекс обсягу** (фізичного обсягу).

Агрегатний *індекс фізичного обсягу* має такий вид:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}, \quad (12.3)$$

де $\sum q_1 p_0$ – вартість продукції звітного періоду за базисними цінами;

$\sum q_0 p_0$ – вартість продукції за базисний період.

Цей індекс показує, у скільки разів зросла (зменшилась) вартість продукції за умов зростання (зниження) обсягу її виробництва або скільки відсотків складає зростання (зниження) вартості продукції внаслідок зміни фізичного обсягу її виробництва.

Якщо об'єктом дослідження є окреме підприємство, то індекс визначається за сукупністю вироблених товарів. Коли об'єктом дослідження виступає галузь промисловості, то індекс розраховується за сукупністю всіх товарів, що вироблені в цій галузі, або окремими їх групами залежно від мети аналізу. Якщо об'єктом дослідження є будь-який регіон, то індекс розраховується за товарами, що вироблені на підприємствах цього регіону.

У ринковій економіці особливе місце серед індексів якісних показників відводиться індексу цін.

Основною змістовністю **індексу цін** є оцінка динаміки цін на товари виробничого та невиробничого споживання. Крім цього, індекс цін відіграє роль загального вимірника інфляції у макроекономічних дослідженнях. Його використовують для: коригування законодавчо встановленого мінімального розміру оплати праці; формування податкових ставок; розроблення техніко-економічних обґрунтувань і проєктів; перерахунку основних показників СНР з фактичних цін у зіставлені.

Агрегатна формула загального **індексу цін** має вид:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}, \quad (12.4)$$

де $\sum p_1 q_1$ – вартість продукції звітного (поточного) періоду.

Цей індекс показує: у скільки разів зросла (зменшилась) вартість продукції (або товарообіг) через зміну цін; скільки відсотків складає зростання (зниження) вартості продукції (товарообігу) в результаті зміни цін.

Залежно від моделі розвитку ринкової економіки використовують різні системи індексів [66; 84].

Базові системи індексів:

Г. Пааше: $I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1}; I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1};$

Е. Ласпейреса: $I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}; I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}.$

Значення індексів цін за системами Пааше та Ласпейреса не співпадають, тому що вони мають різну економічну змістовність.

Індекс цін Пааше показує, наскільки товари у звітному періоді стали дорожче (дешевше), ніж у базисному.

Індекс цін Ласпейреса показує, у скільки разів товари базисного періоду подорожчали (подешевшали) через зміну цін на них у звітному.

Індекс Пааше характеризує вплив зміни цін на вартість товарів, реалізованих у звітному періоді. Індекс Ласпейреса показує вплив зміни цін на вартість товарів, реалізованих у базисному періоді.

Застосування індексів Пааше та Ласпейреса залежить від мети дослідження. Якщо аналіз проводиться для визначення економічного ефекту від зміни цін у звітному періоді в порівнянні з базисним, то застосовується індекс Пааше. Він відображає різницю між фактичною вартістю продажу товарів у звітному періоді ($\sum q_1 p_1$) і розрахунковою вартістю продажу цих же товарів за базисними цінами ($\sum q_0 p_1$). Якщо метою аналізу є визначення обсягу товарообігу продажів у майбутньому періоді такої ж кількості товарів, що й у базисному періоді, але за новими цінами, то застосовується індекс Ласпейреса.

Перемноживши ціни на відповідну кількість проданих товарів і підсумувавши добутки, отримуємо показник загального товарообороту. Відносну його зміну в динаміці характеризує *зведений індекс вартості (товарообороту)* у фактичних цінах:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}. \quad (12.5)$$

За відсутності певної інформаційної бази, що унеможлиблює проведення індексного аналізу в агрегатній формі, індекси можуть бути побудовані в формі *середніх з індивідуальних*.

Середньозважений індекс – середній з індивідуальних індексів, зважених на обсяги, що мають однакову розмірність і зафіксовані на незмінному рівні.

У разі такого визначення індексу постають питання щодо вибору форми середньої і визначення системи ваг.

Результат, отриманий за побудовою *середніх з індивідуальних* індексів, завжди має збігатися з результатом агрегатного індексу.

Середній індекс тотожний агрегатному.

Правила тотожності:

1) *середній арифметичний* індекс тоді буде тотожний агрегатному та буде давати такий же результат, коли за ваги індивідуального індексу у ньому взяти добутки знаменника агрегатного індексу.

Середній арифметичний індекс фізичного обсягу:

$$I_q = \frac{\sum i_{qp} q_0}{\sum p_0 q_0}; \quad (12.6)$$

2) *середній гармонійний* індекс тоді буде тотожний агрегатному та буде давати такий же результат, коли за ваги індивідуального індексу в ньому взяти добутки чисельника агрегатного індексу.

Середній гармонійний індекс цін:

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1}{i_p} \times q_1}. \quad (12.7)$$

Обчислення середніх індексів з іншими вагами не має економічної змістовності, хоч формально їх можна визначити.

Тобто якщо в чисельник агрегатного індексу замість індексованої величини підставити її значення з формули відповідного індивідуального індексу, то агрегатний індекс перетворюється на *середній арифметичний* (коли працюємо з кількісними показниками), а якщо таку зміну здійснити в знаменнику – то отримаємо *середній гармонічний* індекс (коли працюємо з якісними показниками).

Побудова індексів є процесом знаходження загальної міри, середньої величини зміни явища, а сам індексний метод є подальшим розвитком методу середніх величин.

У загальному вигляді класифікацію індексів можна звести до такого: агрегатним індексам із поточними вагами відповідають середні гармонічні індекси з поточними вагами;

агрегатним індексам із базисними вагами відповідають середні арифметичні індекси з базисними вагами.

Поняття "ваги" змінюється залежно від того, в якій формі будується індекс: агрегатній чи середній.

12.3. Індеси зі змінними та постійними вагами

Усі розглянуті індекси вимірюють зміну економічних явищ шляхом порівняння даних за два часові періоди. Якщо існує необхідність вивчення розвитку соціально-економічних явищ за певний період, використовують

систему індексів, яка послідовно характеризує зміни, що відбуваються протягом вибраного часового інтервалу. Система індексів містить низку складових, кількість яких дорівнює числу включених до аналізу періодів часу мінус одиниця.

Розглядають два *варіанти побудови* системи індексів [66; 68; 84]:

показники (рівні) кожного періоду, включеного до розглянутого інтервалу, порівнюються з рівнем одного періоду, вибраного за базу порівняння. Індокси, що входять у таку систему, називають **базисними**;

показники (рівні) порівнюються поміж собою послідовно, кожний рівень порівнюється з попереднім рівнем. Такі індокси називають **ланцюговими**.

Система ланцюгових і базисних індексів може складатися або з індивідуальних, або із загальних індексів.

У статистичній практиці вибір системи індексів визначається характером вирішуваних завдань. Базисні індокси дають більш чітке уявлення щодо загальної тенденції зміни розглядуваного явища, а ланцюгові – більш детальну картину послідовної зміни рівнів у часі.

Поміж *ланцюговими* та *базисними індоксами* існує певний **зв'язок**:

для індивідуальних індексів добуток ланцюгових показників (коефіцієнтів зростання зі змінною базою) завжди буде дорівнювати базисному індексу (коефіцієнту зростання з постійною базою);

відношення двох базисних індивідуальних індексів (коефіцієнтів зростання з постійною базою) дає ланцюговий показник (коефіцієнтів зростання зі змінною базою).

Приклад 12.1. Побудуємо ланцюгові та базисні індокси:

1) для *індивідуального індексу фізичного обсягу*:

а) **базисні**: $i_q = \frac{q_1}{q_0}; \frac{q_2}{q_0}; \frac{q_3}{q_0}; \frac{q_4}{q_0} \dots;$

б) **ланцюгові**: $i_q = \frac{q_1}{q_0}; \frac{q_2}{q_1}; \frac{q_3}{q_2}; \frac{q_4}{q_3} \dots$

У ході побудови рядів зведених (загальних) індексів виникає специфічне питання – вибір ваг цих індексів.

Індокси можуть бути розраховані не тільки з *постійними* та *змінними* базами порівняння, а й з *постійними* та *змінними вагами*,

2) для загального індексу фізичного обсягу:

а) **ланцюгові:**

$$\bullet \text{ зі змінною вагою } I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \frac{\sum q_2 p_1}{\sum q_1 p_1}; \frac{\sum q_3 p_2}{\sum q_2 p_2}; \frac{\sum q_4 p_3}{\sum q_3 p_3}; \quad (12.8)$$

$$\bullet \text{ з постійною вагою } I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \frac{\sum q_2 p_0}{\sum q_1 p_0}; \frac{\sum q_3 p_0}{\sum q_2 p_0}; \frac{\sum q_4 p_0}{\sum q_3 p_0}; \quad (12.9)$$

б) **базисні:**

$$\bullet \text{ зі змінною вагою } I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \frac{\sum q_2 p_1}{\sum q_0 p_1}; \frac{\sum q_3 p_2}{\sum q_0 p_2}; \frac{\sum q_4 p_3}{\sum q_0 p_3}; \quad (12.10)$$

$$\bullet \text{ з постійною вагою } I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \frac{\sum q_2 p_0}{\sum q_0 p_0}; \frac{\sum q_3 p_0}{\sum q_0 p_0}; \frac{\sum q_4 p_0}{\sum q_0 p_0}. \quad (12.11)$$

Для загальних індексів розглянута залежність відбуватиметься лише за умов, що ряд загальних індексів розраховано з тими ж вагами.

Система індексів із постійними вагами будується, як правило, для кількісних показників (індексів обсягу). Система індексів зі змінними вагами – для якісних показників (індексів цін, собівартості, витрат часу на одиницю продукції).

Індекси з постійними вагами є так званими *стандартизованими показниками*, що розраховуються до тієї самої "стандартної" структури сукупності явищ. Вони дозволяють елімінувати вплив зміни структури на динаміку індексованої величини.

12.4. Індекси середніх величин

На динаміку показників, рівні яких виражені середніми величинами, впливає зміна структури розглядуваного явища. Під зміною структури явища розуміють зміну частки окремих одиниць сукупності, з яких формуються середні, в загальній їх чисельності. Наприклад, на середню собівартість виробу в галузі впливає не тільки зміна собівартості цього виробу на галузевих підприємствах, а і зміна частки підприємств із різною собівартістю в загальному випуску цього виробу. Динаміка середньодушового

доходу населення залежить від зміни доходу кожної людини та зміни кількості людей із більш високими (або низькими) доходами в загальній чисельності населення.

Таким чином, на зміну середнього значення показника впливають два фактора одночасно: зміна значень індексованого показника та зміна структури явища. Знаходження кількісного впливу цих факторів будується на побудові індексів середнього рівня показника [66; 84].

Аналіз динаміки середнього рівня показників здійснюють на основі побудови системи взаємозалежних індексів. Індекс, що характеризує зміну середнього рівня показника за рахунок зміни всіх факторів у цілому, дорівнює добутку індексів-співмножників, кожний з яких характеризує зміну лише одного фактора та тим самим – вплив цієї зміни на динаміку середнього значення показника.

Індекс змінного складу розраховується як відношення середніх рівнів показника за поточний та базисний періоди.

Загальна схема визначення індекс змінного складу має вид:

$$I_{\text{зм.скл.}\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}, \quad (12.12)$$

де x_1, x_0 – рівні показника, що індексується у звітному та базисному періодах, відповідно;

f_1, f_0 – ваги (частоти) показника, що індексується у звітному та базисному періодах, відповідно.

Цей індекс показує вплив двох факторів одночасно на загальну динаміку середнього рівня.

Зміна середнього рівня показника за рахунок зміни індексованого показника (якісний фактор) визначається за допомогою **індексу постійного (фіксованого) складу**:

$$I_{\text{фікс.скл}\bar{x}} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1}. \quad (12.13)$$

Зміна середнього рівня показника за рахунок зміни частот (кількісний фактор) визначається за допомогою **індексу структурних зрушень**:

$$I_{\text{стр.зруш.}\bar{x}} = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}. \quad (12.14)$$

В якості ваг (частот) індексів середніх величин, поряд із абсолютними показниками f , можуть використовуватися відносні показники (частки) d . Тоді індекси динаміки середніх показників будуть виду:

$$I_{\bar{x}} = \frac{\sum x_1 d_1}{\sum x_0 d_0}; \quad (12.15)$$

$$I_{\bar{x}} = \frac{\sum x_1 d_1}{\sum x_0 d_1}; \quad (12.16)$$

$$I_{\bar{x}} = \frac{\sum d_1 x_0}{\sum d_0 x_0}, \quad (12.17)$$

де d_1, d_0 – частки одиниць із визначеним значенням ознаки в загальній сукупності у звітному та базисному періодах, відповідно, ($\sum d = 1$).

Між індексами середніх величин існує взаємозв'язок [66; 68]:

$$I_{\text{зм.скл.}} = I_{\text{фіксскл}} \times I_{\text{стрзр.}} \quad (12.18)$$

Отже, визначаючи той чи інший середній рівень показника, можна отримати характеристику його динаміки. Першим кроком в отриманні коректних характеристик динаміки певного середнього показника є розуміння його економічної змістовності.

Так, середній рівень:

заробітної плати працівників компанії визначається відношенням фонду заробітної плати за певний період дослідження в компанії до середньооблікової чисельності працівників у цій компанії за досліджуваний період;

цін реалізації певного виду товару в місті визначається відношенням загальної виручки від реалізації цього виду до кількості товару, реалізованого за весь період дослідження;

продуктивності праці працівників виробничої фірми визначається відношенням обсягу виробленої продукції всіма працівниками фірми до їх чисельності.

Різновидом індексів середніх величин є територіальні індекси, де середні рівні порівнюються за окремими територіями або об'єктами.

12.5. Територіальні індекси

Статистика широко застосовує метод порівняння показників у розрізі підприємств, міст, економічних районів, областей, країн. Узагальнювальні показники, які характеризують співвідношення рівнів складних економічних явищ у просторі (тобто в розрізі територій і об'єктів), називають **територіальними індексами**.

Побудова територіальних індексів має свої особливості порівняно з індексами, які характеризують динаміку явищ. Обчислення індивідуальних територіальних індексів достатньо легке, адже обчислюються звичайні відносні величини порівняння. Так, якщо порівнюється середня ціна на однотипну продукцію в двох регіонах, в першому із яких вона становить 250, а в другому – 270 грн, то зіставленням першого показника з другим створюється територіальний індекс, що буде дорівнювати 0,93. Тобто отриманий індивідуальний індекс, який показує, що середня ціна в першому регіоні на 7 % нижче, ніж у другому.

У побудові загальних територіальних індексів виникає проблема вибору *бази порівняння* та об'єкта, на рівні якого слід зафіксувати вагу індексу. У кожному конкретному випадку її вирішують виходячи з мети дослідження [66; 68; 84].

Порівняння показників можна здійснювати або за двома територіями (об'єктами), або за колом територій (об'єктів). У першому випадку базою може бути показник будь-якої з територій, а в другому – база порівняння повинна бути економічно обґрунтованою. Так, якщо порівнюється, наприклад, рентабельність за колом однотипних підприємств з приблизно однаковими техніко-економічними умовами виробництва, то за базу порівняння слід узяти підприємство, яке має найвищий рівень рентабельності.

Для *побудови територіальних індексів якісних показників* за вагу приймають:

середню величину кількісного показника за сукупністю одиниць порівнюваних територій;

кількісний показник, що належить до території, на якій інтенсивний показник є більш економічним;

кількісний показник, який прийнято за стандарт.

У побудові *територіальних індексів кількісних показників* вагою може слугувати середній рівень якісного показника:

для території, за якою здійснюється порівняння;

установлений для території, прийнятої за стандарт.

У динамічних порівняннях, як уже зазначалося, за ваги (наприклад, в агрегатному індексі цін) беруть кількість виробленої продукції у звітному періоді. Але для територіальних порівнянь поняття "звітний період" і "базисний період" мають умовне значення. Якщо порівнювати регіон А з регіоном Б, то базою буде рівень цін у регіоні Б, а за ваги треба брати кількість продукції в регіоні А.

У цьому випадку індекс цін матиме вигляд:

$$I_p = \frac{\sum p_A q_A}{\sum p_B q_A}. \quad (12.19)$$

Проте зовсім не обов'язково порівнювати регіон А з регіоном Б. На тій самій підставі можна порівнювати регіон Б з регіоном А. За такого зіставлення базою буде рівень цін регіону А, а вагою "звітного періоду" – кількість продукції регіону Б. Отже, індекс цін буде мати такий вигляд:

$$I_p = \frac{\sum p_B q_B}{\sum p_A q_B}. \quad (12.20)$$

Такі відмінності в територіальних індексах виникають унаслідок того, що порівнювані сукупності відрізняються за асортиментом продукції, структурою виробництва, структурою основних засобів, складом працівників тощо. У цих випадках для формування об'єктивних висновків і однозначної відповіді потрібно здійснити вирівнювання сукупностей за структурою.

У статистиці пропонуються різні способи розв'язання цієї проблеми, в тому числі *спосіб стандартних ваг*. Його сутність полягає в тому, що значення індексованої величини зважується не за вагами якого-небудь одного територіального підрозділу, а за вагами економічного району, регіону, країни, в яких знаходяться порівнювані території (об'єкти).

Так, якщо прийняти, що $q = q_A + q_B$, то, який би регіон не був взятий за базу порівняння, отримані результати не будуть суперечити один одному й індекс якісного показника (індекс цін) буде виду:

$$I_p = \frac{\sum p_A q}{\sum p_B q}. \quad (12.21)$$

У територіальних індексах кількісних показників за ваги можна обрати середні рівні відповідних якісних показників (цін, собівартості, фондодовіддачі, трудомісткості, урожайності тощо), обчислені в середньому за територіями, що порівнюються:

$$I_q = \frac{\sum q_A \bar{p}}{\sum q_B \bar{p}}, \quad (12.22)$$

$$\text{де } \bar{p} = \frac{p_A q_A + p_B q_B}{q_A + q_B}.$$

Територіальні індекси широко використовуються для здійснення міжнародних порівнянь.

12.6. Індексний факторний метод аналізу

За допомогою індексів визначається не тільки динаміка складного явища, але і вплив на неї окремих факторів. Індексний метод дозволяє оцінити вплив окремих факторів як у відносному, так і в абсолютному вираженні.

У статистичній практиці прийнято таке **правило факторного аналізу**: якщо результативний показник можна подати як добуток кількісного та якісного фактора, то, визначаючи вплив кількісного фактора на зміну результативного показника, якісний фактор фіксують на рівні базисного періоду; якщо ж визначається вплив якісного показника, то кількісний фактор фіксують на рівні звітного періоду [66; 84].

Найбільш розповсюдженими є такі **індексні системи**:

індекси фізичного обсягу (I_q) і цін (I_p) є факторними відносно до індексу вартості продукції:

$$I_{pq} = I_p \times I_q; \quad (12.23)$$

індекси собівартості (I_z) і фізичного обсягу продукції (I_q) є факторними відносно до індексу витрат на виробництво:

$$I_{zq} = I_z \times I_q; \quad (12.24)$$

індекси чисельності робітників (I_T) і заробітної плати (I_f) є факторними відносно до індексу фонду заробітної плати:

$$I_F = I_T \times I_f \quad (12.25)$$

або

$$\frac{\sum f_1 T_1}{\sum f_0 T_0} = \frac{\sum f_0 T_1}{\sum f_0 T_0} \times \frac{\sum f_1 T_1}{\sum f_0 T_1}; \quad (12.26)$$

індекси чисельності робітників (I_T) і продуктивності праці (I_w) є факторними відносно до індексу обсягу продукції:

$$I_Q = I_T \times I_w \quad (12.27)$$

або

$$\frac{\sum W_1 T_1}{\sum W_0 T_0} = \frac{\sum W_0 T_1}{\sum W_0 T_0} \times \frac{\sum W_1 T_1}{\sum W_0 T_1}; \quad (12.28)$$

індекси обсягу основних фондів (I_Φ) і фондівіддачі (I_v) є факторними відносно до індексу обсягу продукції:

$$I_Q = I_\Phi \times I_v \quad (12.29)$$

або

$$\frac{\sum V_1 \bar{\Phi}_1}{\sum V_0 \bar{\Phi}_0} = \frac{\sum V_0 \bar{\Phi}_1}{\sum V_0 \bar{\Phi}_0} \times \frac{\sum V_1 \bar{\Phi}_1}{\sum V_0 \bar{\Phi}_1}. \quad (12.30)$$

Наведені індексні системи називають **мультиплікативними моделями** [66; 68].

Розглянуті індексні системи є двофакторними. Але, якщо результативна ознака залежить від трьох і більше факторів, то будують багатofакторні індексні моделі.

Використовують два *методи розкладання загального індексу на факторні*:

метод ізольованого вивчення факторів;

метод ланцюгової підстановки (взаємозв'язаного вивчення факторів).

Оскільки в економічній практиці господарювання розглядувані явища є складними та, як наслідок, взаємозв'язаними, то доцільно застосувати другий метод, який потребує правильного розташування факторів у структурі моделі результативної ознаки (наприклад, $W = a \times b \times c \times d$).

На перше місце в моделі необхідно ставити якісний фактор. Збільшення ланцюга факторів на один фактор (наприклад, $a \times b$) кожного разу повинно приводити до показника з реальним економічним сенсом.

Для визначення впливу першого фактора всі інші фактори беруться на рівня звітного періоду (у чисельнику й у знаменнику).

Для визначення впливу другого фактора перший фактор зберігається на рівні базисного періоду, третій і всі наступні – на рівні звітного періоду і т. д.

Наприклад, $w = a \times b \times c \times d$. Звідси послідовно-ланцюгове розкладання факторів буде мати вигляд:

$$I_w = \frac{W_1}{W_0} = \frac{a_1 b_1 c_1 d_1}{a_0 b_0 c_0 d_0} = I_a \times I_b \times I_c \times I_d \quad (12.31)$$

Або

$$\frac{a_1 b_1 c_1 d_1}{a_0 b_0 c_0 d_0} = \frac{a_1 b_1 c_1 d_1}{a_0 b_1 c_1 d_1} \times \frac{a_0 b_1 c_1 d_1}{a_0 b_0 c_1 d_1} \times \frac{a_0 b_0 c_1 d_1}{a_0 b_0 c_0 d_1} \times \frac{a_0 b_0 c_0 d_1}{a_0 b_0 c_0 d_0} = \frac{a_1}{a_0} \times \frac{b_1}{b_0} \times \frac{c_1}{c_0} \times \frac{d_1}{d_0}. \quad (12.32)$$

Індексні системи застосовуються для визначення в абсолютному вираженні зміни складного явища за рахунок впливу окремих факторів. Розрахунки, що пов'язані із зазначеним визначенням в абсолютному вираженні зміни результативного показника за рахунок окремих факторів, називають *розкладанням абсолютного приросту (скорочення)* за факторами.

Так, розглянуту індексну систему чотирифакторного зв'язку (формули (12.31) – (12.32)) можна подати в абсолютних величинах:

$$\begin{aligned} & a_1 b_1 c_1 d_1 - a_0 b_0 c_0 d_0 = \\ & = (a_1 - a_0) b_1 c_1 d_1 + (b_1 - b_0) a_0 c_1 d_1 + (c_1 - c_0) a_0 b_0 d_1 + (d_1 - d_0) a_0 b_0 c_0. \end{aligned} \quad (12.33)$$

Таку модель називають **адитивною**.

Важливі поняття

Агрегатний індекс – основна форма побудови зведених індексів.

Вага індексу – величина, що слугує для вимірювання індексованих величин.

Індекс – відносна величина, що виходячи з системності розгляду явищ характеризує їх зміну в часі, просторі або ступінь відхилення від певного стандарту (нормативу).

Індекс змінного складу – динаміка середніх рівнів показника.

Індекс постійного (фіксованого) складу – індекс, який показує, як змінюється середній рівень показника за рахунок зміни якісного фактора (величини, що індексується).

Індекс структурних рушень – індекс, який показує, як змінюється середній рівень показника за рахунок зміни кількісного фактора (його структури).

Індексний метод – комплексна характеристика відносної зміни в часі, просторі або в порівнянні з яким-небудь еталоном таких явищ, які внаслідок наявності функціональної залежності між ними подані системою взаємопов'язаних показників за принципом подання інтегрального результату через його складові.

Індексована величина – ознака, зміну якої вивчають (ціна товарів, курс акцій, витрати робочого часу на виконання роботи, кількість проданих товарів).

Правило факторного аналізу – якщо результативний показник можна подати як добуток кількісного та якісного факторів, то, визначаючи вплив кількісного фактора на зміну результативного показника, якісний фактор фіксують на рівні базисного періоду; якщо ж визначається вплив якісного показника, то кількісний фактор фіксують на рівні звітного періоду.

Середньозважений індекс – середній з індивідуальних індексів, зважених на обсяги, що мають однакову розмірність і зафіксовані на незмінному рівні.

Територіальний індекс – узагальнювальний показник, який характеризує співвідношення рівнів складних економічних явищ у просторі (в розрізі територій чи об'єктів).

Типові завдання

Завдання 12.1. Аналітична фірма виконує роботу з оцінювання вартості послуг соціальних робітників (табл. 12.1). Необхідно визначити вплив

зміни кількості послуг, що здійснюють соціальні працівники, на загальні витрати часу на здійснення цих послуг всіма працівниками.

Таблиця 12.1

Показники діяльності фірми, що здійснює соціальні послуги

Види послуг	Загальні витрати часу на здійснення послуги у 2018 р., тис. людино-год	Зміна кількості послуг у 2019 р. порівняно з 2018 р., %
Умовні позначки	$q_0 t_0$	i_q
Прибирання	16	-8
Купівля продуктів	10	+2

Розв'язання.

Виходячи з економічної змістовності показника трудомісткості одиниці виконаних робіт (t) як витрат часу (T) на здійснення послуг соціальними робітниками (q) маємо запис $t = \frac{T}{q}$.

Тоді $T = t \times q$ за 2018 р. є базисним періодом порівняно з 2019 р. і загальні витрати часу за цей рік будуть мати позначку $q_0 t_0$.

Зміна кількості послуг надана у відсотках. База порівняння для відсотків – 100 одиниць. Тоді зміна кількості послуг із прибирання за визначений період складе: $\frac{100 - 8}{100} = \frac{92}{100} = 0,92$, з купівлі продуктів: $\frac{100 + 2}{100} = \frac{102}{100} = 1,02$.

Якщо постає питання щодо загальної зміни кількості послуг, треба розрахувати загальний індекс фізичного обсягу. Цей індекс в агрегатній формі за умовами задачі буде мати вид:

$$I_q = \frac{\sum q_1 t_0}{\sum q_0 t_0}$$

Але інформація про витрати часу за умов здійснення послуг у 2019 р.

$q_1 t_0$ – відсутня. Тоді, якщо $i_q = \frac{q_1}{q_0}$, то $q_1 = i_q \times q_0$.

Користуючись правилами тотожності індексів, перетворимо агрегатний індекс у середній:

$$I_q = \frac{\sum q_1 t_0}{\sum q_0 t_0}$$

Підставляємо $q_1 = i_q \times q_0$ у вираз $q_1 t_0$ і отримуємо новий індекс, який має назву *середній арифметичний індекс*: $I_q = \frac{\sum i_q q_0 t_0}{\sum q_0 t_0}$.

$$I_q = \frac{\sum i_q q_0 t_0}{\sum q_0 t_0} = \frac{0,92 \times 16 + 1,02 \times 10}{16 + 10} = \frac{24,92}{26} = 0,958, \text{ або } 95,8 \%$$

Отже, загальні витрати часу в фірмі, що здійснює соціальні послуги, за рахунок зміни їх кількості знизились на 4,2 %.

Завдання 12.2. Аналітичний відділ банку досліджує депозитні відсоткові ставки для юридичних і фізичних осіб (табл. 12.2). Завданням відділу є визначення динаміки середньої депозитної ставки у банківській установі.

Таблиця 12.2

Дані щодо депозитні відсоткові ставки

Вкладники	Сума залучених депозитів, тис. грн		Депозитна ставка, %	
	базисний період	звітний період	базисний період	звітний період
Умовні позначки	f_0	f_1	x_0	x_1
Юридичні особи	960	1040	35	29
Фізичні особи	210	360	22	18

Розв'язання.

Динаміку середніх рівнів визначають за допомогою індексів змінного складу, фіксованого складу та структурних зрушень.

Індекс змінного складу середньої депозитної ставки:

$$I_{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0},$$

де \bar{x}_1, \bar{x}_0 – середня депозитна ставка у звітному та базисному періодах;

x_1, x_0 – депозитна ставка у звітному та базисному періодах;

f_1, f_0 – сума залучених депозитів у звітному та базисному періодах.

$$I_{\bar{x}} = \frac{0,29 \times 1040 + 0,18 \times 360}{1040 + 360} : \frac{0,35 \times 960 + 0,22 \times 210}{960 + 210} = 0,2617 : 0,3266 = 0,80,$$

або 80 %.

Індекс фіксованого складу середньої депозитної ставки:

$$I_{\bar{x}} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1} = 0,2617 : \frac{0,35 \times 1040 + 0,22 \times 360}{1040 + 360} = 0,2617 : 0,3165 = 0,8268$$

або 82,68 %.

Індекс структурних зрушень середньої депозитної ставки:

$$I_{\bar{x}} = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} = 0,3165 : 0,3266 = 0,9690, \text{ або } 96,6 \%$$

Взаємозв'язок індексів: $I_{\bar{x}} = I_x \times I_f = 0,8268 \times 0,969 = 0,80$.

Отже, взаємозв'язок індексів забезпечує перевірку коректності розрахунків. Середня депозитна ставка у звітному періоді порівняно з базисним у банківській установі змінилася на 20 % (100 – 80). Це зниження забезпечено зниженням депозитної ставки за кожною групою вкладників на 17,32 % (100 – 82,68) і зниженням суми залучених депозитів за видами вкладників на 3,1 % (100 – 96,9).

Завдання 12.3. З метою вивчення купівельного попиту на товар А проводиться маркетингове дослідження (табл. 12.3). Необхідно визначити, в якому з регіонів і наскільки відрізняється рівень цін на товар А за умов різної за ним номенклатури.

Базові показники кон'юнктури ринку товару А

Номенклатура товару А	Регіон А		Регіон В	
	продано, кг	ціна 1 кг, грн	продано, кг	ціна 1 кг, грн
Умовні позначки	q_A	p_A	q_B	p_B
СХ	450	8,0	480	7,5
АН	240	5,0	210	4,5

Розв'язання.

Якщо в індексі цін фіксувати ваги того регіону, який порівнюється з іншим, то можна побудувати й обчислити два індекси:

$$I_p = \frac{\sum p_A q_A}{\sum p_B q_A} = \frac{8 \times 450 + 5 \times 240}{7,5 \times 450 + 4,5 \times 240} = \frac{4800}{4455} = 1,0774, \text{ або } 107,74\%,$$

тобто стосовно структури номенклатури товару А, проданого в регіоні А, рівень цін у регіоні А порівняно з регіоном В вищий на 7,74 %.

Однак, порівнюючи ціни регіону В із цінами регіону А, дістанемо:

$$I_p = \frac{\sum p_B q_B}{\sum p_A q_B} = \frac{7,5 \times 480 + 4,5 \times 210}{8 \times 480 + 5 \times 210} = \frac{4545}{4890} = 0,9294, \text{ або } 92,94\%,$$

тобто стосовно складу товару А, який було реалізовано в регіоні В, ціни в цьому регіоні нижчі, ніж у регіоні А на 7,06 %.

Отже, в кожному регіоні рівень цін на товар А є різним, якщо відношення рівнів цін вимірювати стосовно структури товару регіону, що порівнюється.

Завдання 12.4. Керівництво виробничого підприємства має за мету дослідити ефективність роботи робітників (табл. 12.4). Перед економічним відділом було поставлено завдання визначити, як змінився обсяг продукції, що випускає підприємство, в цілому та за рахунок зміни кількості працівників і продуктивності їх праці.

Дані щодо виробництва продукції за виробничими підрозділами

Виробничий підрозділ	Кількість виробленої продукції, од.		Чисельність робітників, осіб	
	базисний період	звітний період	базисний період	звітний період
Умовні позначки	q_0	q_1	T_0	T_1
1	4060	3800	200	150
2	2500	2000	100	100

Розв'язання.

Визначимо рівні продуктивності праці (W) за кожним із виробничих підрозділів підприємства за звітний та базисний періоди.

Виробничий підрозділ № 1:

$$W_1 = \frac{q_1}{T_1} = \frac{3800}{150} = 25,33 \text{ од.};$$

$$W_0 = \frac{q_0}{T_0} = \frac{4060}{200} = 20,30 \text{ од.}$$

Виробничий підрозділ № 2:

$$W_1 = \frac{2000}{100} = 20 \text{ од.};$$

$$W_0 = \frac{2500}{100} = 25 \text{ од.}$$

Тепер розрахуємо індекси.

Якщо $W = \frac{q}{T}$, то $q = W \times T$, тоді $I_q = I_w \times I_t$, або

$$\frac{\sum W_1 T_1}{\sum W_0 T_0} = \frac{\sum W_1 T_1}{\sum W_0 T_1} \times \frac{\sum W_0 T_1}{\sum W_0 T_0}.$$

За даними прикладу $I_q = I_w \times I_t = 1,045 \times 0,850 = 0,89$, або 89 %.

Отже, загалом на підприємстві випуск продукції зменшено на 11 %: за рахунок зменшення кількості працівників випуск зменшено на 15 %, а за рахунок збільшення продуктивності праці він зріс на 4,5 %.

Абсолютні прирости за рахунок окремих факторів обчислюють як різницю між чисельником і знаменником відповідних факторних індексів.

Так, загальний абсолютний приріст визначимо як:

$$\Delta q = W_1 T_1 - W_0 T_0 = q_1 - q_0.$$

Його можна розкласти за факторами:

$$\Delta q_{(w)} = W_1 T_1 - W_0 T_1 = (W_1 - W_0) \times T_1,$$

$$\Delta q_{(T)} = W_0 T_1 - W_0 T_0 = (T_1 - T_0) \times W_0.$$

Тоді маємо $\Delta q = \Delta q_{(w)} + \Delta q_{(T)}$.

Визначимо загальний абсолютний приріст та його приріст за рахунок впливу окремих факторів у виробничому підрозділі № 1: $\Delta q = 3\,800 - 4\,060 = -260$ одиниць, тобто у звітному періоді порівняно з базисним у виробничому підрозділі № 1 було вироблено продукції на 260 одиниць менше. Це зниження обумовлено:

1) зміною продуктивності праці:

$$\Delta q_{(w)} = (25,33 - 20,30) \times 150 = 755 \text{ одиниць, тобто за рахунок підвищення продуктивності праці випуск продукції зріс на 755 одиниць;}$$

2) зміною чисельності працівників:

2) зміною чисельності працівників:

$$\Delta q_{(T)} = (150 - 200) \times 20,3 = -1\,015 \text{ одиниць, тобто за рахунок зменшення чисельності робітників було не отримано 1\,015 одиниць продукції.}$$

Перевіримо взаємозв'язок: $755 + (-1\,015) = -260$ одиниць.

Перевіримо взаємозв'язок: $755 + (-1\,015) = -260$ одиниць.

Виробниче підприємство за визначений період дослідження знизило випуск продукції на 11 %, що в абсолютному виразі складає 260 одиниць. Зміна чисельності працівників сприяла зменшенню обсягу виробництва продукції на 15 %, тобто на 1 015 одиниць, тоді як підвищення продуктивності праці робітників на 4,5 % збільшило випуск на 755 одиниць.

Посилання на лабораторну роботу

Методичні рекомендації до виконання лабораторної роботи за темою "Набуття навиків аналізу статистичних даних за допомогою індексного

методу за допомогою MS Excel" подані в роботі [67]. Лабораторна робота націлена на придбання навиків аналізу статистичних даних за допомогою індексного методу в пакеті MS Excel.

Запитання до самодіагностики

1. У чому полягає сутність індексу?
2. Назвіть критерії, за якими здійснюється класифікація індексів.
3. Що є обов'язковою умовою для побудови індивідуальних індексів?
4. Визначте змістовну сутність агрегатних індексів.
5. У чому полягає відмінність індексних систем Пааше та Ласпейреса?
6. За яких умов застосовують середньозважені індекси?
7. За яких умов можна визначити взаємозв'язок між ланцюговими та базисними індексами?
8. Які відмінності існують у побудові індексів динаміки та територіальних індексів?
9. Чим відрізняються індекси змінного та фіксованого складу?
10. Як визначити абсолютний приріст результативного показника за рахунок окремих факторів співмножників?

Питання для критичного переосмислення (есе)

1. Відносні, середні величини та індекси: відмінність використання в економічних дослідженнях.
2. Базові системи індексів: умови поєднання в прикладних дослідженнях.
3. Особливості визначення ваг залежно від форми побудови індексів.
4. Практика застосування територіальних індексів.
5. Правила розкладання факторів у побудові багатофакторних моделей.

Рекомендована література

1. Акімов О. В. Статистика в малюнках та схемах : навч. посіб. / О. В. Акімов. – Київ : ЦУЛ, 2007. – 168 с.
2. Аксьонова І. В. Історія статистики : конспект лекцій для студентів спеціальності 7.050 110 усіх форм навчання / І. В. Аксьонова, О. В. Авраменко. – Харків : Вид. ХДЕУ, 2002. – 172 с.
3. Аксянова А. В. Теория и практика статистики: учеб. пособ. / А. В. Аксянова, Н. Н. Валеев, А. М. Гумеров. – Москва : Колосс, 2008. – 284 с.
4. Балдин К. В. Общая теория статистики : учеб. пособ. / К. В. Балдин. – Москва : Дашков и К^о, 2010. – 312 с.
5. Бек В. Л. Теорія статистики: курс лекцій : навч. посіб. / В. Л. Бек. – Київ : ЦУЛ, 2003. – 412 с.
6. Васильева Э. К. Статистика : учебник / Э. К. Васильева, В. С. Лялин. – Москва : ЮНИТИ-ДАНА, 2007. – 399 с.
7. Вуколов Э. А. Основы статистического анализа : практ. по стат. мет. и исследованию операций с использ. пак. STATISTICA и EXCEL. – Москва : Форум, 2010. – 464 с.
8. Глинский В. В. Статистический анализ : учеб. пособ. для студ. вузов экон. профиля / В. В. Глинский, В. Г. Ионин. – 3-е изд., перераб. и доп. – Москва : ИНФРА-М, 2002. – 241с.
9. Гмурман В. Е. Теория вероятностей и математическая статистика : учеб. пособ. для вузов / В. Е. Гмурман. – 12-е изд. – Москва : Высшее образование, 2006. – 479 с.
10. Годин А. М. Статистика : учебник / А. М. Годин. – Москва : Дашков и К^о, 2009. – 460 с.
11. Головач А. В. Статистичне забезпечення управління економікою: прикладна статистика : навч. посіб. / А. В. Головач, В. Б. Захожай, Н. А. Головач. – Київ : КНЕУ, 2005. – 334 с.
12. Гончарук А. Г. Основы статистики : навч. посіб. / А. Г. Гончарук. – Київ : ЦНЛ, 2004. – 148 с.
13. Горкавий В. К. Статистика : навч. посіб. / В. К. Горкавий. – Київ : Алерта, 2012. – 608 с.

14. Горна М. О. Статистика для маркетологів : дистанційний курс Moodle для студентів спеціалізацій "Маркетинг", "Підприємництво, торгівля та біржова діяльність" / М. О. Горна. – Київ : КНЕУ, 2018. – 140 с.
15. Григорович Б. А. Технології візуалізації даних / А. Г. Григорович, Б. А. Григорович // International Academy Journal Web of Scholar. – 2018. – 4 (22). – Vol. 1. – P. 23–28.
16. Григоруку П. М. Багатомірне економіко-статистичне моделювання : навч. посіб. / П. М. Григоруку. – Львів : Новий Світ-2000, 2006. – 148 с.
17. Громыко Г. Л. Теория статистики : учебник / Г. Л. Громыко. – Москва : ИНФРА-М, 2010. – 479 с.
18. Двігун А. О. Статистика : навч. посіб. для студентів вищих навчальних закладів / А. О. Двігун, П. А. Борисенко, К. І. Дерев'яноко. – Запоріжжя : Акцент Інвесттрейд, 2012. – 307 с.
19. Дегтяр А. О. Статистичні методи в державному управлінні : навч. посіб. / А. О. Дегтяр, М. В. Гончаренко. – 4-те вид., випр. і доп. – Харків : Вид-во ХарPI НАДУ "Магістр", 2018. – 176 с.
20. Дианов Д. В. Прикладная статистика: учебник / Д. В. Дианов. – Москва : Элит, 2006. – 768 с.
21. Дмитриева И. А. Общая теория статистики : учеб. пособ. / И. А. Дмитриева, С. Н. Лысенко. – Москва : Вузовский учебник, 2009. – 219 с.
22. Едророва В. Н. Общая теория статистики / В. Н. Едророва, М. В. Едророва. – Москва : Юристъ, 2001. – 511 с.
23. Економічна статистика : навч. посіб. / В. М. Соколов, Т. Г. Чала, О. С. Корепанов та ін. ; за ред. В. М. Соколова. – Харків : ХНУ ім. В. Н. Каразіна, 2017. – 388 с.
24. Елисеева И. И. Статистика : учебник / И. И. Елисеева. – Москва : Юрайт, 2010. – 565 с.
25. Єріна А. М. Організація вибіркового спостереження : навч. посіб. / А. М. Єріна. – Київ : КНЕУ, 2004. – 137 с.
26. Єріна А. М. Статистика : підручник / А. М. Єріна, З. О. Пальян. – Київ : КНЕУ, 2010. – 351 с.
27. Єріна А. М. Статистичне моделювання та прогнозування : підручник / А. М. Єріна, Д. Л. Єрін. – Київ : КНЕУ, 2014. – 170 с.

28. Єріна А. М. Теорія статистики : практикум / А. М. Єріна, З. О. Пальян. – Київ : Товариство "Знання", 2008. – 255 с.
29. Илышев А. М. Общая теория статистики : учебник / А. М. Илышев. – Москва : ЮНИТИ-ДАНА, 2008. – 535 с.
30. Ковалевский Г. В. Статистика : учебник / Г. В. Ковалевский. – Харьков : ХНАГХ, 2012. – 444 с.
31. Ковтун Н. В. Теорія статистики : підручник / Н. В. Ковтун. – Київ : Знання, 2012. – 399 с.
32. Колков С. В. Статистика : учеб. пособ. / С. В. Колков, К. Э. Плехотников. – Москва : Флинта, 2008. – 288 с.
33. Костин В. Н. Статистические методы и модели : учеб. пособ. / В. Н. Костин, Н. А. Тишина. – Оренбург : ГОУ ОГУ, 2004. – 138 с.
34. Кулинич О. І. Теорія статистики : підручник / О. І. Кулинич, Р. О. Кулинич. – 6-те вид., перероб. і доп. – Київ : Знання, 2013. – 239 с.
35. Лезина М. Л. Статистика : учеб. пособ. для ВУЗов / М. Л. Лезина. – Москва : Элит-2000, 2008. – 368 с.
36. Лугінін О. Є. Статистика : підручник / О. Є. Лугінін, С. В. Білоусова. – Київ : ЦНЛ, 2006. – 580 с.
37. Лялин В. С. Статистика: теория и практика в Excel : учеб. пособ. / В. С. Лялин. – Москва : Финансы и статистика, 2010. – 448 с.
38. Маличенко И. П. Общая теория статистики: практикум с решением : учеб. пособ. / И. П. Маличенко. – Ростов-на-Дону : Феникс, 2010. – 282 с.
39. Манцуров І. Г. Статистика економічного зростання та конкурентоспроможності країни / І. Г. Манцуров. – Київ : КНЕУ, 2006. – 388 с.
40. Марець О. Р. Представлення статистичної інформації за допомогою графічного методу / О. Р. Марець, О. М. Вільчинська // International Scientific Journal. – 2015. – № 9. – С. 118–125.
41. Мармоза А. Т. Практикум з теорії статистики : навч. посіб. / А. Т. Мармоза. – 3-тє вид., випр. – Київ : Ельга; Ніка-Центр, 2007. – 348 с.
42. Мармоза А. Т. Теорія статистики / А. Т. Мармоза. – Київ : Ельга; Ніка-Центр, 2003. – 392 с.
43. Матковський С. О. Теорія статистики : навч. посіб. / С. О. Матковський, О. Р. Марець. – Київ : Знання, 2010. – 535 с.
44. Моторин Р. М. Статистика для економістів : навч. посіб. / Р. М. Моторин, Е. В. Чекотовський. – Київ : Знання, 2009. – 430 с. ; компакт-диск.

45. Мхитарян В. С. Статистика : учебник / В. С. Мхитарян. – Москва : Academia, 2010. – 272 с.
46. Назаров М. Г. Общая теория статистики : учебник для ВУЗов / М. Г. Назаров. – Москва : Омега-Л, 2010. – 410 с.
47. Назаров М. Г. Практикум по общей теории статистики : учеб.-метод. пособ. / М. Г. Назаров. – Москва : Кнорус, 2010. – 178 с.
48. Опря А. Т. Математична статистика : навч. посіб. для студ. екон. спец. с.-г. вузів / А. Т. Опря. – Київ : Урожай, 1994. – 206 с.
49. Опря А. Т. Статистика (модульний варіант з програмованою формою контролю знань) : навч. посіб. / А. Т. Опря. – Київ : ЦУЛ, 2012. – 448 с.
50. Орлов А. И. Прикладная статистика : учебник для ВУЗов / А. И. Орлов. – Москва : Экзамен, 2006. – 672 с.
51. Осауленко О. Г. Статистичний словник / НТК статистичних досліджень : словник / О. Г. Осауленко, О. О. Васечко, М. В. Пугачова. – Київ : ДП "Інформ.-аналіт. агенство", 2012. – 250 с.
52. Офіційний сайт Головного управління статистики в Харківській області. – Режим доступу : <http://kh.ukrstat.gov.ua>.
53. Офіційний сайт Державної служби статистики України. – Режим доступу : www.ukrstat.gov.ua.
54. Офіційний сайт Міністерства економічного розвитку і торгівлі України. – Режим доступу : <http://me.gov.ua>.
55. Офіційний сайт Національного банку України. – Режим доступу : <https://bank.gov.ua>.
56. Петров Л. Ф. Методы динамического анализа экономики / Л. Ф. Петров. – Москва : Инфра-М, 2010. – 238 с.
57. Практикум по статистике : учеб. пособ. / под ред. А. П. Зинченко. – Москва : Колосс, 2001. – 392 с.
58. Про внесення змін до деяких законодавчих актів України у зв'язку з прийняттям Закону України "Про інформацію" та Закону України "Про доступ до публічної інформації" : Закон України № 1170-VII від 27.03.2014 р. [Електронний ресурс]. – Режим доступу : <https://zakon.rada.gov.ua/laws/show/1170-18>.
59. Про державну статистику : Закон України № 2614-XII від 17.09.1992 р. [Електронний ресурс]. – Режим доступу : <https://zakon.rada.gov.ua/laws/show/2614-12>.

60. Просветов Г. И. Статистика: задачи и решения : учеб. практ. пособ. / Г. И. Просветов. – Москва : Альфа-Пресс, 2008. – 496 с.
61. Протасов К. В. Статистический анализ экспериментальных данных / К. В. Протасов. – Москва : Мир, 2005. – 142 с.
62. Рудакова Р. П. Практикум по статистике / Р. П. Рудакова, Л. Л. Букин, В. И. Гаврилов. – Санкт-Петербург : ЛГУ им. А. С. Пушкина, 2006. – 251 с.
63. Салин В. Н. Статистика: учеб. пособ. / В. Н. Салин, Э. Ю. Чурилова, Е. П. Шпаковская. – Москва : КноРус, 2008. – 296 с.
64. Сєрова І. А. Організація статистичних спостережень : конспект лекцій / І. А. Сєрова, І. В. Аксьонова. – Харків : ХНЕУ, 2008. – 236 с.
65. Сигел Э. Ф. Практическая бизнес-статистика / Э. Ф. Сигел ; пер. с англ. – 4-е изд. – Москва : ИД "Вильямс", 2004. – 1051 с.
66. Статистика (модульний варіант з програмованою формою контролю знань) : навч. посіб. / А. Т. Опря, Л. О. Дорогань-Писаренко, О. В. Єгорова та ін. ; за ред. А. Т. Опрі. – 2-ге вид., перероб. і допов. – Київ : ЦНЛ, 2017. – С. 345–379.
67. Статистика : методичні рекомендації до лабораторних робіт для студентів усіх спеціальностей першого (бакалаврського) рівня [Електронний ресурс] / уклад. О. В. Раєвнєва, І. В. Аксьонова, І. А. Сєрова та ін. ; за ред. О. В. Раєвнєвої. – Харків : ХНЕУ ім. С. Кузнеця, 2019. – 104 с. – Режим доступу : <http://repository.hneu.edu.ua/handle/123456789/21537>.
68. Статистика : навч. посіб. / за ред. О. В. Раєвнєвої. – Харків : ВД "ІНЖЕК", 2011. – 504 с.
69. Статистика : навч.-метод. посіб. для самост. вивч. дисц. / А. М. Єріна, Р. М. Моторін, А. В. Головач та ін.; за заг. ред. А. М. Єріної, Р. М. Моторіна. – Київ : КНЕУ, 2002. – 448 с.
70. Статистика : підручник / за ред. Р. Я. Баран. – Чернівці : Наші книги. – 2008. – 240 с.
71. Статистика : підручник / С. С. Герасименко, А. В. Головач, А. М. Єріна та ін.; за наук. ред. д-ра екон. наук С. С. Герасименка. – 2-ге вид., перероб. і доп. – Київ : КНЕУ, 2000. – 467 с.
72. Статистика : учеб. пособ. / под ред. В. М. Симчеры. – Москва : Финансы и статистика, 2008. – 368 с.
73. Статистика : учебник / под ред. В. Г. Ионина. – Москва : Инфра-М, 2008. – 445 с.

74. Статистика : учебник / под ред. И. И. Елисейевой. – Москва : Проспект, 2010. – 448 с.
75. Статистика : учебник / под ред. Л. И. Ниворожкиной. – Москва : Дашков и К°, 2015. – 416 с.
76. Статистичні методи обробки та аналізу економічних даних : навч. посіб. для студ. вищих навч. закл. / за ред. Ю. В. Кулешкова. – Кіровоград : КДТУ, 2003. – 137 с.
77. Стешенко В. С. Птуха Михайло Васильович / В. С. Стешенко // Енциклопедія історії України : у 10 т. / редкол. : В. А. Смолій (голова) та ін. ; Інститут історії України НАН України. – Київ : Наукова думка, 2012. – Т. 9 : Прил. – С. 58.
78. Тарасенко І. О. Статистика : навч. посіб. / І. О. Тарасенко. – Київ : ЦНЛ, 2006. – 344 с.
79. Тарновская Л. И. Статистика : учеб. пособ. для студ. высш. учеб. заведений / Л. И. Тарновская. – Москва : Academia, 2008. – 320 с.
80. Теория статистики : учебник / под. ред. Р. А. Шмойловой. – 5-е изд. – Москва : Финансы и статистика, 2007. – 656 с.
81. Теория статистики с элементами эконометрики : В 2-х ч., ч. 2 : учебник для академического бакалавриата / отв. ред. В. В. Ковалев. – Москва : Изд. "Юрайт", 2017. – 348 с.
82. Теорія статистики : навч. посіб. / за ред. П. Г. Вашківа. – Київ : Либідь, 2001. – 320 с.
83. Теорія статистики : навч. посіб. / М. В. Макаренко, І. М. Гойхман, О. О. Гладчук та ін.; за ред. М. В. Макаренко. – Київ : Кондор, 2010. – 236 с.
84. Ткач Є. І. Загальна теорія статистики: підручник [для студ. вищ. навч. закл.] / Є. І. Ткач, В. П. Сторожук. – 3-тє вид. – Київ : ЦУЛ, 2017. – С. 302–333.
85. Тринько Р. І. Основи теоретичної і прикладної статистики : навч. посіб. / Р. І. Тринько, М. Є. Стадник. – Київ : Знання, 2011. – 400 с.
86. Уманець Т. В. Загальна теорія статистики : навч. посіб. / Т. В. Уманець. – Київ : Знання, 2006. – 239 с.
87. Штагрет А. М. Статистика : навч. посіб. / А. М. Штагрет. – Київ : ЦНЛ, 2005. – 232 с.
88. Щурик М. В. Статистика : навч. посіб. / М. В. Щурик. – 2-ге вид., оновл. і допов. – Львів : Магнолія, 2006, 2009. – 546 с.

89. Berenson M. Applied statistics. A first course Prentice Hall / M. Berenson. – New Jersey : Englewood Cliffs, 1988. – 324 p.
90. Business statistics: a decision – making approach / D. F. Groebner; Patrick W. Shannon, P. C. Fry, K. D. Smith. – New Jersey : Prentice Hall, 2008. – 1040 p.
91. Dowdy S. Statistics for research / S. Dowdy, S. Weardon, D. Chilko. – 3 ed. – Hoboken; New Jersey : Wiley-Interscience, 2004. – 627 p.
92. Few S. Show Me the Numbers: Designing Tables and Graphs to Enlighten / S. Few. – Oakland; California: Analytics Press. 2004. – 263 p.
93. Guillot C. Statistiques Descriptives [Electronic resource] / C. Guillot. – Courbevoie : Pôle Universitaire Léonard de Vinci, 2016. – 17 p. – Access mode : <http://www.mghassany.com/documents/stat-desc-fall-16-17.pdf>.
94. Introduction to R. Grammar of Graphics [Electronic resource]. – Access mode : <https://ramnathv.github.io/pycon2014-r/visualize/ggplot2.html>.
95. Keller G. Instructor's solutions manual for statistics for management and economics / G. Keller, B. Warrack. – 5 ed. – Duxbury : Pacific Grove, Ca., 2000. – 738 p.
96. Keller G. Statistics for Management and Economics / G. Keller, B. Warrack. – 5 ed. – Duxbury : Thomson Learning, 1999. – 920 p.
97. Leboucher L. Introduction à la statistique descriptive / L. Leboucher, M.-J. Voisin. – Toulouse : Cépaduès, 2013. – 206 p.
98. Mazerolle F. Statistique descriptive [Electronic resource] / F. Mazerolle. – Paris : Gualino éditeur, 2006. – 173 p. – Access mode : <https://epdf.pub/statistique-descriptive-serie-statistique-a-une-et-deux-variables-series-chronol.html>.
99. Monino J.-L. Statistique descriptive / J.-L. Monino. – 5 ed. – Malakoff Cedex : Dunod, 2017 – 343 p.
100. Py B. Statistique descriptive : Nouvelle méthode pour bien comprendre et réussir / B. Py. – 5 ed. – Paris : Economica, 2007. – 350 p.
101. Simon P. The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions / P. Simon // Wiley and SAS Business Series. – New Jersey : Wiley. – 2014. – 240 p.
102. Steele J. Beautiful Visualization: Looking at Data through the Eyes of Experts (Theory in Practice) / J. Steele, N. Iliinsky. – Boston : O'Reilly Media Inc., 2010. – 375 p.

103. Tille Y. Resume du Cours de Statistique Descriptive [Electronic resource] / Y. Tille. – Neuchâtel, 2010. – 172 p. – Access mode : https://www.unine.ch/files/live/sites/statistics/files/shared/documents/cours_statistique_descriptive.pdf.

104. Upton G. A dictionary of statistics : Invaluable reference work, covering all areas of statistics / G. Upton, I. Cook. – 2. ed. – New York : Oxford University Press, 2006. – 490 p.

Додатки

Додаток А

Таблиця А.1

**Критичні значення кореляційного відношення η^2 і коефіцієнта
детермінації R^2 за $\alpha = 0,05$**

k_2	k_1						
	1	2	3	4	5	6	7
5	0,569	0,699	0,764	0,806	0,835	0,854	0,872
6	0,500	0,632	0,704	0,751	0,785	0,811	0,831
7	0,444	0,575	0,651	0,702	0,739	0,768	0,791
8	0,399	0,527	0,604	0,657	0,697	0,729	0,754
9	0,362	0,488	0,563	0,618	0,659	0,692	0,719
10	0,332	0,451	0,527	0,582	0,624	0,659	0,687
12	0,283	0,394	0,466	0,521	0,564	0,600	0,630
14	0,247	0,348	0,417	0,471	0,514	0,550	0,580
16	0,219	0,312	0,387	0,429	0,477	0,507	0,538
18	0,197	0,283	0,345	0,394	0,435	0,470	0,501
20	0,179	0,259	0,318	0,364	0,404	0,432	0,468
22	0,164	0,238	0,294	0,339	0,377	0,410	0,439
24	0,151	0,221	0,273	0,316	0,353	0,385	0,414
26	0,140	0,206	0,256	0,297	0,332	0,363	0,391
28	0,130	0,193	0,240	0,279	0,314	0,344	0,371
30	0,122	0,182	0,227	0,264	0,297	0,326	0,353
40	0,093	0,139	0,176	0,207	0,234	0,259	0,282
50	0,075	0,113	0,143	0,170	0,194	0,216	0,235
60	0,063	0,095	0,121	0,144	0,165	0,184	0,202
80	0,047	0,072	0,093	0,110	0,127	0,142	0,156
100	0,038	0,058	0,075	0,090	0,103	0,116	0,128
120	0,032	0,049	0,063	0,080	0,087	0,098	0,109
200	0,019	0,030	0,038	0,046	0,053	0,060	0,067
400	0,010	0,015	0,019	0,023	0,027	0,031	0,034

Таблиця значень F для рівня значущості $\alpha = 0,05$

$k_2 \backslash k_1$	1	2	3	4	5	6	8	12	24	∞
1	161,5	199,5	215,7	224,6	230,2	233,9	238,8	243,9	249,1	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,91
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,79
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,72	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,47	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,53
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Таблиця значень F для рівня значущості $\alpha = 0,01$

$k_1 \backslash k_2$	1	2	3	4	5	6	8	12	24	∞
1	4052,1	4999,0	5403,5	5625,1	5764,1	5889,4	5981,3	6105,8	6234,2	6366,5
2	98,49	99,01	99,17	99,25	99,30	99,33	99,36	99,42	99,46	99,50
3	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,60	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	13,93	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,27	9,89	9,47	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91
11	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,02	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,59	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,43	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,08	2,65
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,01	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	2,92	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36
22	7,94	5,72	4,87	4,31	3,99	3,75	3,45	3,12	2,75	2,30
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,62	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,48	2,01
35	7,42	5,27	4,40	3,91	3,59	3,37	3,07	2,74	2,37	1,90
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,29	1,82
45	7,23	5,11	4,25	3,77	3,45	3,23	2,94	2,61	2,23	1,75
50	7,17	5,06	4,20	3,72	3,41	3,19	2,89	2,56	2,18	1,68
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,79	1,00

Таблиця Г.1

Шкала Чеддока

η	0,1 – 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9 – 0,99
Сила зв'язку	Слабка	Помірна	Помітна	Тісна	Дуже тісна

Таблиця Г.2

Квантилі χ^2 -розподілу за $\alpha = 0,05$

k	0,95	k	0,95
1	3,84	11	19,68
2	5,99	12	21,03
3	7,82	13	22,36
4	9,49	14	23,69
5	11,07	15	25,00
6	12,59	16	26,30
7	14,07	17	27,59
8	15,51	18	28,87
9	16,92	19	30,14
10	18,31	20	31,41

**Критичні значення вибіркового коефіцієнта кореляції за різних
числах ступенів свободи і рівнях значущості**

Число ступенів свободи $k = n - 2$	Рівень значущості		Число ступенів свободи $k = n - 2$	Рівень значущості	
	$\alpha = 0,05$	$\alpha = 0,01$		$\alpha = 0,05$	$\alpha = 0,01$
1	0,997	0,999	24	0,388	0,496,
2	0,950	0,990	25	0,381	0,487
3	0,878	0,959	26	0,374	0,478
4	0,811	0,917	27	0,367	0,470
5	0,754	0,874	28	0,361	0,463
6	0,707	0,834	29	0,355	0,456
7	0,666	0,798	30	0,349	0,449
8	0,632	0,765	35	0,325	0,418
9	0,602	0,735	40	0,304	0,393
10	0,576	0,708	45	0,288	0,372
11	0,553	0,684	50	0,273	0,354
12	0,532	0,661	60	0,250	0,325
13	0,514	0,641	70	0,232	0,302
14	0,497	0,623	80	0,217	0,283
15	0,482	0,606	90	0,205	0,267
16	0,468	0,590	100	0,195	0,254
17	0,456	0,575	125	0,174	0,228
18	0,444	0,561	150	0,159	0,208
19	0,433	0,549	200	0,138	0,181
20	0,423	0,537	300	0,113	0,148
21	0,413	0,526	400	0,098	0,128
22	0,404	0,515	500	0,088	0,115
23	0,396	0,505	1000	0,062	0,081

**Критичні значення коефіцієнту рангової кореляції Спірмена
для правосторонньої перевірки при α**

Обсяг вибірки	0,10	0,05	0,025	0,01	0,005	0,001
n = 4	0,8000	0,8000				
5	0,7000	0,8000	0,9000	0,9000		
6	0,6000	0,7714	0,8286	0,8857	0,9429	
7	0,5357	0,6786	0,7450	0,8571	0,8929	0,9643
8	0,5000	0,6190	0,7143	0,8095	0,8571	0,9286
9	0,4667	0,5833	0,6833	0,7667	0,8167	0,9000
10	0,4424	0,5515	0,6364	0,7333	0,7818	0,8667
11	0,4182	0,5273	0,6091	0,7000	0,7455	0,8364
12	0,3986	0,4965	0,5804	0,6713	0,7273	0,8182
13	0,3791	0,4780	0,5549	0,6429	0,6978	0,7912
14	0,3626	0,4593	0,5341	0,6220	0,6747	0,7670
15	0,3500	0,4429	0,5179	0,6000	0,6536	0,7464
16	0,3382	0,4265	0,5000	0,5824	0,6324	0,7265
17	0,3260	0,4118	0,4853	0,5637	0,6152	0,7083
18	0,3148	0,3994	0,4716	0,5480	0,5975	0,6904
19	0,3070	0,3895	0,4579	0,5333	0,5825	0,6737
20	0,2977	0,3789	0,4451	0,5203	0,5684	0,6586
21	0,2909	0,3688	0,4351	0,5078	0,5545	0,6455
22	0,2829	0,3597	0,4241	0,4963	0,5426	0,6318
23	0,2767	0,3518	0,4150	0,4852	0,5306	0,6186
24	0,2704	0,3435	0,4061	0,4748	0,5200	0,6070
25	0,2646	0,3362	0,3977	0,4654	0,5100	0,5962
26	0,2588	0,3299	0,3894	0,4564	0,5002	0,5856
27	0,2540	0,3236	0,3822	0,4481	0,4915	0,5757
28	0,2490	0,3175	0,3749	0,4401	0,4828	0,5660
29	0,2443	0,3113	0,3685	0,4320	0,4744	0,5567
30	0,2400	0,3059	0,3620	0,4251	0,4665	0,5479

Показчик

А

Абсолютне прискорення – 292
Абсолютний приріст – 285
Абсолютні показники – 117
Агрегатний індекс – 345, 359
Адитивна модель – 358
Алгебраїчний центральний момент розподілу – 200
Амплітуда (розмах коливань) – 331
Аналітичне вирівнювання – 310
Аналітичне групування – 91
Асиметрія – 198

Б

Багатовимірне групування – 91
База порівняння – 342
Базисні індекси – 350
Базисні характеристики динаміки – 284
Бета-коефіцієнти – 250, 251

В

Вага індексу – 359
Варіація ознаки – 159
Вибіркова сукупність (вибірка) – 170
Вирівнювання часового ряду укрупненням інтервалів – 302
Відносне прискорення – 287
Відносний показник динаміки – 120

Відносний показник плану (відносний показник планового завдання) – 124
Відносний статистичний показник – 120
Відносні показники диференціації – 126
Відносні показники координації – 123
Відносні показники порівняння – 124
Відносні показники структури – 122
Власне випадковий відбір – 172
Внутрішньогрупова дисперсія – 152
Вторинне групування – 93

Г

Генеральна сукупність – 170
Гістограма – 65
Глобальна статистична система – 28
Гранична помилка вибірки – 181
Графічний інтервал – 63
Графічний образ (основа графіка) – 62
Групові таблиці – 74

Д

Децентралізоване зведення – 86
Дециль – 197
Дисперсійний аналіз – 152
Дисперсія – 152

Дисперсія альтернативної
ознаки – 156
Дисперсія ознаки – 149
Діаграма – 65
Довжина відрізання (графічний
інтервал) – 63

Е

Експлікація графіку – 63
Екстраполяція – 312
Ексцес – 201
Емпіричне кореляційне
відношення – 267
Етапи інформаційного
забезпечення статистичного
дослідження – 17

З

Загальна дисперсія – 154
Заголовок таблиці – 74
Залишкова дисперсія – 159, 241
Зважена ковзна середня – 331
Зведені індекси – 343, 344
Зведення – 85
Зворотний зв'язок – 245, 268
Згладжування – 307

І

Ізолінії – 73
Індекс – 341
Індекс змінного складу – 352
Індекс Ласпейреса – 347, 348
Індекс Пааше – 347, 348
Індекс постійного складу – 352

Індексний метод аналізу – 349
Індексована величина – 342
Індекс сезонності – 327
Індекс структурних зрушень – 352
Індивідуальний індекс – 343
Інтервальна оцінка – 174
Інтервальний прогноз – 313
Інтервальні ряди динаміки – 292

К

Картограма – 71
Картограма точкова – 71
Картограма фоновна – 71
Картодіаграма – 71
Квантиль – 195
Квартильний коефіцієнт варіації – 196
Класифікація – 87
Класифікація статистичних ознак – 21
Коефіцієнт асиметрії – 198, 199
Коефіцієнт асоціації – 257
Коефіцієнт взаємної спряженості
Пірсона – 259
Коефіцієнт взаємної спряженості
Чупрова – 259
Коефіцієнт випередження – 287
Коефіцієнт детермінації – 226
Коефіцієнт децильної
диференціації – 198
Коефіцієнт диференціації – 212
Коефіцієнт еластичності – 288
Коефіцієнт квадратичної спряженості
хі-квадрат Пірсона – 259
Коефіцієнт контингенції – 258
Коефіцієнт концентрації – 204
Коефіцієнт кореляції рангів
Кендалла – 264

Коефіцієнт кореляції рангів
Спірмена – 262
Коефіцієнти повної регресії – 250
Коефіцієнти чистої регресії – 250
Коефіцієнт Крамера – 259
Коефіцієнт локалізації – 204
Коефіцієнт нерівномірності – 331
Коефіцієнт подібності
(схожості) – 102, 205
Коефіцієнт регресії – 236
Коефіцієнт структурних
зрушень – 208, 210
Коефіцієнт Фехнера – 264
Коливання рівнів динамічних
рядів – 299
Комбінаційна таблиця – 74
Комбінаційне групування – 91, 92
Комбінована вибірка – 174
Кореляційний аналіз – 242, 268
Кореляційний зв'язок – 222, 224

Л

Ланцюгові індекси – 342, 345
Ланцюгові характеристики
динаміки – 284
Лінійна діаграма – 65
Лінійний зв'язок – 269
Лінійний коефіцієнт
детермінації – 269
Лінійний коефіцієнт кореляції –
243, 244

М

Мала вибірка – 186
Масштабні орієнтири графіку – 62

Математична школа статистики – 7
Медіана – 145
Методи відбору: повторний
й безповторний – 171, 172
Методи поширення даних
вибіркового спостереження – 189
Методологічна основа
статистики – 16
Метод усереднення за лівою
та правою половиною – 307
Механічна вибірка – 172
Міжгрупова дисперсія – 154
Міжнародний статистичний
інститут – 26
Міра асиметрії – 198
Множинна кореляція – 230, 269
Мода – 145
Моментні ряди динаміки – 116, 293
Мультиплікативна модель – 357

Н

Наукові принципи теорії
вибіркового методу – 169, 170
Нелінійний або криволінійний
зв'язок – 269
Непараметричні методи – 256, 257

О

Об'єкт вивчення статистики – 15
Об'єкт спостереження – 46
Одиниця спостереження – 48
Одиниця сукупності – 47
Однорідність сукупності – 20, 144
Описова школа статистики – 7

П

Параметричні методи – 269
Парні коефіцієнти кореляції – 251
Період часу – 279
Перцентель – 196
Підмет таблиці – 74
План статистичного спостереження – 45
Поле графіка – 62
Полігон розподілу – 142
Помилка репрезентативності – 177
Помилка спостереження – 53
Порівнянність даних – 283
Предмет історії статистики – 11
Предмет статистики – 12
Присудок таблиці – 74
Прогнозування – 312, 331
Програма спостереження – 48
Проста ковзна (плинна) середня – 331
Просте зведення – 86
Прості таблиці – 74
Просторові орієнтири графіку – 62
Прямий зв'язок – 262, 269

Р

Радіальна діаграма – 68
Радіус – 62
Ранг – 262
Ранжування – 262
Регресійний аналіз – 229, 253, 269
Реєстр – 42
Результативна ознака – 221
Рівень ряду – 279, 280
Рівномірний розвиток – 314

Рівноприскорений (рівно-уповільнений) розвиток – 315
Розвиток зі змінним прискоренням (уповільненням) – 315
Розвиток за експонентою – 315
Розвиток з уповільненням зростання наприкінці періоду – 316
Ряд динаміки – 279
Ряд розподілу – 19, 139

С

Сезонні коливання – 300, 327
Секторна діаграма – 67
Середній рівень ряду – 133, 288
Середньозважені індекси – 348
Середня величина – 127
Середня (стандартна) помилка вибірки – 190
Серійна вибірка – 190
Складне зведення – 85
Смугова (стрічкова) діаграма – 66
Спеціально організоване спостереження – 41
Спосіб відбору – 179
Спосіб стандартних ваг – 355
Стадії (етапи) статистичного дослідження – 18, 19
Статистика – 8, 9, 10
Статистико-математичний напрям – 8
Статистична закономірність – 23
Статистична звітність – 40
Статистична карта – 71
Статистична одиниця або одиниця статистичної сукупності – 20, 21

Статистична ознака – 21
Статистична оцінка (параметру)
генеральної сукупності – 171
Статистична сукупність – 20
Статистична таблиця – 73
Статистичне групування – 88
Статистичне спостереження – 38
Статистичний графік – 60
Статистичний показник – 24, 134
Статистичний ряд – 161
Стовпчикова діаграма – 65
Стохастичний зв'язок – 222
Структурне групування – 89
Суб'єктивний час – 52
Сукупний коефіцієнт множинної
детермінації – 253
Сукупний коефіцієнт множинної
кореляції – 252, 253
Сумірник – 342

Т

Таблиця взаємної спряженості –
74, 257
Темп зростання – 284, 285
Темп приросту – 284, 286
Тенденція (тренд) – 300, 332
Теоретична лінія регресії – 235
Теоретична основа статистики – 35
Територіальний індекс – 354
Типова вибірка – 184

Типологічне групування – 90
Точкова оцінка – 174
Точковий прогноз – 313

Ф

Факторна ознака – 221, 270
Функції індексів – 340
Функції Статистичного бюро
Секретаріату ООН – 27
Функціональний зв'язок – 221

Ц

Централізація – 206
Централізоване зведення – 105

Ч

Час спостереження (об'єктивний
час) – 51
Часткові коефіцієнти кореляції – 252
Частота – 140

Ш

Шкала – 63

Я

Якість статистичних оцінок – 175

Зміст

Передмова	3
Розділ 1. Введення до статистики.....	5
1. Методологічні засади статистики	5
1.1. Поняття про статистику як суспільну науку, її виникнення та розвиток	5
1.2. Предмет і методологічні основи статистики	11
1.3. Етапи статистичного дослідження та специфічні прийоми статистичного аналізу	17
1.4. Поняття та категорії в статистиці	19
1.5. Організація та завдання статистики в сучасних умовах.....	26
Важливі поняття.....	34
Типові завдання	35
2. Статистичне спостереження	36
2.1. Сутність і завдання статистичного спостереження	37
2.2. Форми, види та способи статистичного спостереження	40
2.3. План статистичного спостереження	45
2.4. Помилки статистичного спостереження та забезпечення вірогідності даних	52
Важливі поняття.....	54
Типові завдання	55
3. Подання статистичних даних: таблиці, графіки, карти.....	59
3.1. Роль і значення графічного методу	60
3.2. Основні елементи графіка. Правила побудови статистичних графіків	61
3.3. Види статистичних графіків і таблиць. Способи їх побудови.....	65
Важливі поняття.....	76
Типові завдання	77
4. Зведення і групування статистичних даних.....	84
4.1. Сутність статистичного зведення та його види	84
4.2. Статистичні групування та їх види	88
Важливі поняття.....	104
Типові завдання	105
Розділ 2. Статистичні індикатори та ряди розподілу	115
5. Узагальнювальні статистичні показники	115

5.1. Сутність, види та значення абсолютних показників	115
5.2. Сутність, види та форми вираження відносних показників	119
5.3. Сутність, види та методи розрахунку середніх показників	126
Важливі поняття	134
Типові завдання	135
6. Аналіз рядів розподілу	139
6.1. Закономірність розподілу та її частотні характеристики	139
6.2. Характеристики центру розподілу	144
6.3. Показники варіації	148
Важливі поняття	159
Типові завдання	161
7. Вибірковий метод	168
7.1. Загальне поняття про вибіркове спостереження. Види та способи відбору одиниць із генеральної сукупності	168
7.2. Помилки вибіркового спостереження для різних схем відбору	174
7.3. Особливості малої вибірки та визначення необхідної чисельності вибірки	186
Важливі поняття	189
Типові завдання	190
8. Аналіз концентрації, диференціації та подібності рядів розподілу	195
8.1. Порядкові характеристики розподілу	195
8.2. Статистичні показники концентрації та централізації	203
8.3. Статистичне оцінювання структурних змін у часі та просторі	208
Важливі поняття	212
Типові завдання	212
Розділ 3. Методи аналізу взаємозв'язків явищ і процесів	221
9. Статистичні методи вимірювання взаємозв'язків	221
9.1. Поняття та види зв'язків у статистиці	221
9.2. Модель аналітичного групування	225
9.3. Рівняння регресії і визначення його параметрів	235
9.4. Показники тісноти та істотності кореляційного зв'язку	242
9.5. Побудова множинних кореляційно-регресійних моделей	247

9.6. Методи вивчення зв'язку соціальних явищ.....	254
9.7. Непараметричні показники зв'язку.	
Рангові коефіцієнти зв'язку.....	262
Важливі поняття.....	267
Типові завдання.....	270
10. Аналіз інтенсивності динаміки.....	279
10.1. Поняття про ряди динаміки та їх види.....	279
10.2. Аналітичні показники рядів динаміки.....	284
Важливі поняття.....	292
Типові завдання.....	293
11. Аналіз тенденцій розвитку та коливань.....	299
11.1. Прийоми виявлення основної тенденції розвитку в рядах динаміки.....	299
11.2. Інтерполяція та екстраполяція.....	312
11.3. Факторний аналіз рядів динаміки.....	325
11.4. Аналіз сезонних коливань.....	327
Важливі поняття.....	331
Типові завдання.....	332
12. Індексний метод.....	340
12.1. Поняття про індекси та їх роль у статистико-економічному аналізі.....	340
12.2. Агрегатний індекс як основна форма загального індексу. Середньозважені індекси.....	345
12.3. Індеси зі змінними та постійними вагами.....	349
12.4. Індеси середніх величин.....	351
12.5. Територіальні індекси.....	354
12.6. Індексний факторний метод аналізу.....	356
Важливі поняття.....	359
Типові завдання.....	359
Рекомендована література.....	367
Додатки.....	375
Показчик.....	381

НАВЧАЛЬНЕ ВИДАННЯ

Раєвнєва Олена Валентинівна
Аксьонова Ірина Вікторівна
Бровко Ольга Іванівна та ін.

СТАТИСТИКА

Навчальний посібник

*За загальною редакцією
д-ра екон. наук, професора О. В. Раєвнєвої*

Самостійне електронне текстове мережеве видання

Відповідальний за видання *О. В. Раєвнєва*

Відповідальний редактор *М. М. Оленич*

Редактор *Н. І. Ганцевич*

Коректор *Н. І. Ганцевич*

План 2019 р. Поз. № 18-ЕНП. Обсяг 389 с.

Видавець і виготовлювач – ХНЕУ ім. С. Кузнеця, 61166, м. Харків, просп. Науки, 9-А

*Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру
ДК № 4853 від 20.02.2015 р.*