

Кам'янець-Подільський національний університет імені Івана Огієнка

Економічний факультет

Кафедра управління персоналом та економіки праці

О. В. Рарок

СТАТИСТИКА

Конспект лекцій

Навчальний посібник

Кам'янець–Подільський

ФОП Сисин Я. І.

2017

Укладач: *Олександр Васильович Рарок* – кандидат економічних наук, старший викладач кафедри управління персоналом та економіки праці КПНУ імені Івана Огієнка.

Рекомендовано до друку вченою радою
Кам'янець-Подільського національного університету імені Івана Огієнка
(протокол № 16 від 26 грудня 2016 р.)

Рецензенти:

Місюк М. В., доктор економічних наук, професор кафедри економіки, підприємництва, торгівлі та біржової діяльності Подільського державного аграрно-технічного університету;

Рудик В. К., доктор економічних наук, завідувач кафедри фінансів, банківської справи та страхування Подільського державного аграрно-технічного університету;

Лисак В. Ю., кандидат економічних наук, доцент кафедри економіки підприємств Кам'янець-Подільського національного університету імені Івана Огієнка.

Статистика. Конспект лекцій : навчальний посібник / Укл. Рарок О. В.
С78 – Кам'янець-Подільський : ФОП Сисин І. Я., 2017. – 202 с.
ISBN 978-617-7052-71-4.

У навчальному посібнику у вигляді конспекту лекцій коротко розглянуто основний теоретичний матеріал, що стосується програми навчальної дисципліни «Статистика» для студентів економічного напрямку підготовки вищих навчальних закладів III-IV рівня акредитації. У матеріалах лекцій висвітлено методологічні основи статистики, основні поняття і методи статистичної науки; розглянуто питання, що виникають на стадії статистичного спостереження, зведення первинного статистичного матеріалу та його обробки і аналізу; розкрито суть статистичних показників та способи їх одержання; викладено методи статистичного аналізу (дисперсійний, кореляційно-регресійний, індексний, табличний і графічний), особливу увагу приділено статистичним методам аналізу взаємозв'язків між ознаками, які широко використовуються для прогнозування, виконаного за результатами наукових експериментів та математичної обробки виробничих і соціологічних показників.

До кожної лекції наведено питання для перевірки знань і самоконтролю, які можуть бути використані викладачами для складання матеріалів поточного і семестрового контролю знань студентів.

Посібник може бути корисним також для студентів інших спеціальностей, які вивчають цей предмет, а також для всіх читачів, що цікавляться проблемами статистики.

ЗМІСТ

Вступ	6
Лекція 1. Методологічні основи статистики	7
1.1. Становлення статистики як науки. Предмет статистики.....	7
1.2. Основні поняття і категорії статистичної науки.....	9
1.3. Статистичні показники. Закон великих чисел і статистичні закономірності	10
1.4. Метод статистики і статистична методологія.....	11
1.5. Галузі статистичної науки.....	12
1.6. Завдання статистики в сучасних умовах та її організація в Україні.....	14
1.7. Міжнародні статистичні організації.....	18
Питання для повторення і самоконтролю	20
Лекція 2. Статистичне спостереження	21
2.1. Суть, джерела та організаційні форми статистичного спостереження.....	22
2.2. Види статистичного спостереження та способи отримання даних.....	26
2.3. Загальний план статистичного спостереження.....	31
2.4. Програма і статистичний формуляр спостереження.....	32
2.5. Організаційний план статистичного спостереження, забезпечення точності даних.....	35
2.6. Помилки спостереження і контроль вірогідності даних.....	39
Питання для повторення і самоконтролю	39
Лекція 3. Зведення і групування статистичних даних	41
3.1. Суть і завдання статистичного зведення і статистичних групувань.....	41
3.2. Основні правила утворення груп.....	45
3.3. Типологічні, структурні та аналітичні групування.....	49
3.4. Прості та комбінаційні групування.....	53
3.5. Вторинні групування.....	56
3.6. Багатовимірні групування.....	60
3.7. Основні статистичні класифікації і класифікатори України.....	61
Питання для повторення і самоконтролю	63
Лекція 4. Узагальнюючі статистичні показники	64
4.1. Суть і види узагальнюючих статистичних показників.....	65
4.2. Абсолютні статистичні величини, їх види та одиниці виміру.....	67
4.3. Відносні величини. Види відносних величин, техніка їх обчислення та форми виразу.....	68
4.4. Взаємозв'язок між окремими видами відносних величин.....	70
4.5. Суть середніх величин та умови їх використання.....	72
4.6. Види середніх величин, їх обчислення. Багатовимірні середні.....	73
Питання для повторення і самоконтролю	79
Лекція 5. Аналіз рядів розподілу	81
5.1. Ряди розподілу, їх види і правила побудови. Закономірність	

розподілу.....	81
5.2. Характеристики центра розподілу.....	84
5.3. Характеристики варіації.....	87
5.4. Характеристика форми розподілу.....	89
5.5. Види та взаємозв'язок дисперсій.....	92
Питання для повторення і самоконтролю	94
Лекція 6. Аналіз концентрації, диференціації та подібності розподілу	95
6.1. Поняття коефіцієнтів концентрації і локалізації, їх обчислення та використання.....	95
6.2. Коефіцієнт подібності структур двох сукупностей.....	97
6.3. Інтенсивність структурних зрушень.....	98
Питання для повторення і самоконтролю	99
Лекції 7, 8. Статистичні методи вимірювання взаємозв'язків	99
7.1. Види взаємозв'язків. Основні поняття кореляційного та регресійного аналізу.....	100
7.2. Поняття про дисперсійний аналіз.....	102
7.3. Метод аналітичного групування і графічного зображення.....	106
7.4. Парна лінійна кореляція і парна регресія.....	108
7.5. Поняття про багатофакторну регресію.....	114
7.6. Нелінійна кореляція.....	115
7.7. Непараметричні методи вивчення зв'язків.....	120
7.8. Використання кореляційно-регресійного аналізу в соціально-економічних дослідженнях.....	124
Питання для повторення і самоконтролю	126
Лекція 9. Аналіз інтенсивності динаміки	127
9.1. Динамічний ряд як база аналізу і прогнозування соціально-економічного розвитку. Види рядів динаміки.....	127
9.2. Статистичні характеристики динамічних рядів, їх взаємозв'язок.....	131
9.3. Методи обчислення середнього рівня динамічного ряду.....	134
9.4. Економічна суть та техніка розрахунку середніх значень основних характеристик рядів динаміки.....	136
Питання для повторення і самоконтролю	137
Лекція 10. Аналіз тенденцій розвитку та коливань	138
10.1. Виявлення тенденцій розвитку явищ. Трендова компонента, її аналіз.....	139
10.2. Періодична компонента. Статистичний аналіз сезонних коливань.....	143
10.3. Факторний аналіз рядів динаміки.....	145
10.4. Інтерполяція та екстраполяція. Прогнозування на основі рядів динаміки.....	149
Питання для повторення і самоконтролю	152
Лекція 11. Індексний метод.	153
11.1. Суть індексів та їх роль у статистико-економічному аналізі.....	153
11.2. Класифікація індексів. Методологічні принципи їх побудови.....	154

11.3. Базисні та ланцюгові індекси, зв'язок між ними. Інтегральний індекс динаміки.....	155
11.4. Індекси агрегатної форми.....	158
11.5. Середньозважені індекси. Розкладання абсолютного приросту за факторами.....	160
11.6. Індекси змінного і фіксованого складу та структурних зрушень, їх взаємозв'язок.....	164
11.7. Індекси Ласпейреса, Пааше і Фішера. Індекс споживчих цін.....	165
Питання для повторення і самоконтролю	168
Лекція 12. Вибірковий метод.	169
12.1. Суть, значення і переваги вибіркового спостереження, способи формування вибірок.....	170
12.2. Обчислення похибок вибірки та визначення меж інтервалу для середньої величини і частки.....	172
12.3. Визначення необхідного обсягу вибірки.....	173
12.4. Способи поширення вибірових даних на генеральну сукупність.....	175
12.5. Застосування вибірових спостережень у соціально-економічних дослідженнях.....	176
Питання для повторення і самоконтролю	178
Лекція 13. Подання статистичних даних.	178
13.1. Поняття статистичної таблиці, її основні елементи.....	179
13.2. Види статистичних таблиць, правила їх побудови.....	181
13.3. Аналіз таблиць.....	183
13.4. Означення та основні елементи статистичного графіка.....	184
13.5. Види статистичних графіків.....	188
13.6. Проблема вибору необхідних видів графіків. Аналіз графіків.....	197
Питання для повторення і самоконтролю	199
Список використаної літератури.	200

Вступ

У період формування і становлення ринкових відносин між суб'єктами господарської діяльності нашої держави виникає нагальна потреба у вивченні закономірностей, що проявляються у масових соціально-економічних явищах, з метою прогнозування перспектив розвитку тих чи інших галузей економіки, соціальної інфраструктури, демографії тощо. Через це зростає інтерес до статистики як науки, яку у своїй практичній діяльності в першу чергу використовують фахівці-економісти, оскільки їхня праця пов'язана із збиранням, систематизацією, обробкою та аналізом статистичних матеріалів.

Реформи, що запроваджуються останнім часом у різні сфери людської діяльності в умовах децентралізації влади в нашій країні, спонукають до розширення і поглиблення знань з цієї науки та використання її методів на практиці. Зокрема, запровадження вибіркового обстеження, одноразових обліків чи опитування дозволить оптимізувати звітність за рахунок зменшення обсягу статистичного матеріалу і на цій основі проводити оперативний і поглиблений аналіз.

У запропонованому посібнику в короткій формі представлено системний виклад основних категорій, принципів і методів статистичної науки, теоретичних основ соціально-економічних методів аналізу і прогнозування із використанням дисперсійного, індексного, кореляційно-регресійного, табличного і графічного методів.

До складу навчального посібника входить теоретичний матеріал навчальної дисципліни «Статистика», представлений у вигляді конспекту 13 лекцій, зміст яких у стислій формі відповідає навчальній програмі цієї дисципліни для студентів економічного напрямку підготовки. Кожна лекція закінчується переліком контрольних питань для перевірки знань і проведення самоконтролю. Ці питання також можуть бути успішно використані викладачами для проведення як поточного, так і семестрового контролю знань студентів.

Лекція 1. Методологічні основи статистики

План

- 1.6. Становлення статистики як науки. Предмет статистики.
- 1.7. Основні поняття і категорії статистичної науки.
- 1.8. Статистичні показники. Закон великих чисел і статистичні закономірності.
- 1.4. Метод статистики і статистична методологія.
- 1.5. Галузі статистичної науки.
- 1.8. Завдання статистики в сучасних умовах та її організація в Україні.
- 1.9. Міжнародні статистичні організації.

1.1. Становлення статистики як науки. Предмет статистики

Термін «статистика» походить від латинського слова status – стан справ. У науку цей термін ввів німецький учений Готфрід Ахенвалль в 1746 році, запропонувавши змінити назву курсу «Державознавство», що викладалося в університетах Німеччини, на «Статистику», поклавши тим самим початок розвитку статистики як науки й навчальної дисципліни. Однак відомо, що статистичний облік вівся набагато раніше: проводився перепис населення в Древньому Китаю, здійснювалося порівняння військового потенціалу держав, проводився облік майна громадян у Древньому Римі тощо.

Початок статистичної практики відноситься приблизно до часу виникнення держави. Першою інформацією можна вважати глиняні таблички Шумерського царства (III-II тисячоліття до н.е.). Спочатку під статистикою розуміли опис економічного й політичного багатства держави або його частин. Однак поступово термін «статистика» став використовуватись ширше. За Наполеоном Бонапартом, «статистика – це бюджет речей». Тим самим статистичні методи були визнані корисними не тільки для адміністративного керування, але й застосування на рівні окремого підприємства.

В другій половині XIX – початку XX століть сформувалася наукова дисципліна – математична статистика, яка є частиною математики. У XX столітті статистику часто розглядають уже як самостійну науку: *статистика – це сукупність методів і принципів, згідно з якими проводиться збирання, аналіз, порівняння, подання й інтерпретація числових даних.*

Відразу після виникнення теорії ймовірностей (Паскаль, Ферма, XVII століття) ймовірнісні моделі стали використовуватись при обробці статистичних даних. Наприклад, вивчалася частота народження хлопчиків і дівчаток, було встановлено відмінність ймовірності народження хлопчика від 0,5 (її прийняли рівною 0,51). У 1794 році німецький математик Карл Гаусс розробив один із методів сучасної математичної статистики – метод найменших квадратів.

Перша третина XX століття пройшла під знаком параметричної статистики: вивчалися методи, що описувалися кривими сімейства Пірсона, серед них найбільш популярним виявився *нормальний розподіл.*

В даний час існує багато визначень терміну «статистика», але найчастіше він вживається в трьох значеннях: 1) статистика – це галузь практичної діяльності, яка має за мету збирання, обробку, аналіз та опублікування масових даних про ті чи інші явища; 2) статистика – це числова інформація, що характеризує різні сторони життя держави: культуру, населення, виробництво тощо; 3) статистика – це галузь знань, спеціальна наука, яка займається розробкою теоретичних положень і методів їх практичного використання.

Між статистичною наукою і практичною діяльністю існує тісний зв'язок і взаємозалежність: статистика використовує інформацію практичної діяльності підприємств, узагальнює її та розробляє методи проведення статистичних досліджень; з іншого боку, підприємства, організації та установи для практичної діяльності використовують теоретичні розробки і положення статистичної науки в практичних цілях.

Підсумовуючи сказане, предмет статистики на сучасному рівні можна визначити так: *статистика – це наука, яка вивчає кількісну сторону масових*

явищ соціально-економічного життя у нерозривному зв'язку з їх якісним змістом в конкретних умовах місця і часу.

1.2. Основні поняття і категорії статистичної науки

Первинним поняттям у статистиці є *статистична сукупність*, під якою розуміють певну множину однорідних елементів, поєднаних умовами існування і розвитку. Сукупність є *однорідною*, якщо одна або декілька її ознак (властивостей), що вивчаються, є спільними для всіх елементів. Кожен *елемент* називається інакше *одиницею* сукупності.

Кожна статистична сукупність характеризується якісними і кількісними ознаками (властивостями). *Якісні ознаки* (атрибутивні ознаки) виражаються у вигляді понять, визначень, які характеризують їхню суть, стан або якість. Наприклад, сорт продукції, професія, сімейний статус тощо. *Кількісні ознаки* виражають окремі значення якісних ознак у числовому виразі. У свою чергу кількісні ознаки поділяються на дискретні і неперервні. До *дискретних* відносяться ознаки, що виражаються окремими цілими числами, без проміжних значень. *Неперервні* – це ознаки, що можуть характеризуватися будь-якими дійсними числами. *Прямі ознаки* характеризують об'єкт дослідження безпосередньо (вік осіб, кількість присутніх студентів в аудиторії тощо). *Непрямі* – це ознаки, що не належать безпосередньо досліджуваному об'єкту (чи сукупності), а які стосуються іншої сукупності, яка входить до даної. *Багатоваріантні* ознаки – перш за все характеризуються рангами (шкалою рангів) у послідовності від більшого значення до меншого (наприклад, дуже низький, низький, середній, високий, дуже високий). *Альтернативні* ознаки приймають взаємовиключаючі значення: так-ні, позитивне-негативне. *Інтервальні* – це ознаки, які характеризують результат процесів. *Моментні* – характеризують об'єкт у певний момент часу. Окремі значення кількісних ознак називаються *варіантами*. *Первинні варіанти* характеризують одиницю сукупності в цілому: абсолютні значення, які виміряні чи обчислені. *Вторинні*

варіанти (похідні, розрахункові) – дані, які неможливо перевірити, оскільки вони взяті з певних джерел.

1.3. Статистичні показники. Закон великих чисел і статистичні закономірності

Статистичні показники – це числа в сукупності з набором ознак, що характеризують обставини, до яких вони відносяться. *Статистичний показник* – це кількісна характеристика соціально-економічних явищ і процесів в умовах якісної визначеності. *Статистичні дані* – це сукупність показників, отриманих внаслідок статистичного спостереження або обробки даних.

Статистична закономірність – це закономірність, в якій необхідність в кожному окремому явищі пов'язана з випадковістю, і лише в масовій сукупності явищ виявляє себе як закон. *Система статистичних показників* – це сукупність статистичних показників, які відображають взаємозв'язки, що об'єктивно існують між явищами.

Теоретичне обґрунтування статистичних закономірностей проводиться на підставі відомого в теорії ймовірностей *закону великих чисел*, що заключає в собі принцип, згідно з яким *закономірність масових явищ може проявлятися лише при достатньо великій кількості досліджуваних випадків*. Зокрема, відповідно до цього закону, що виражається кількома граничними теоремами, відносна частота випадкової події при зростанні кількості незалежних повторних випробувань наближається до ймовірності цієї події (теорема Бернуллі); середнє арифметичне великої кількості випадкових величин при певних умовах практично співпадає з математичним сподіванням цієї сукупності (теорема Чебишева); сумарна випадкова величина, що об'єднує в собі велику кількість випадкових величин з обмеженими дисперсіями, має закон розподілу, близький до нормального (теорема Ляпунова). Саме на підставі закону великих чисел за допомогою статистичних методів виникає можливість виявити статистичні закономірності, що проявляються у масових випадкових явищах, та оцінити точність і надійність одержаних результатів.

1.4. Метод статистики і статистична методологія

Метод статистики – це сукупність прийомів, за допомогою яких статистика досліджує свій предмет.

Метод статистики включає в себе три групи складових методів: *метод масових спостережень, метод угруповань і метод узагальнюючих показників*. Кожне статистичне дослідження проходить в три етапи:

1) На першому етапі за допомогою методу масових спостережень збирають первинний статистичний матеріал. Основний зміст цього етапу полягає в отриманні даних, що характеризують кожен одиницю спостереження.

2) На другому етапі статистичного дослідження проводиться первинна обробка, зведення та групування. Метод групувань дозволяє виділити однорідні сукупності, розділити їх на групи і підгрупи. Підсумок – це отримання підсумків в цілому по сукупності та по її групах і підгрупах. Результати групування і зведення викладаються у вигляді статистичних таблиць. Основний зміст цього етапу полягає в переході від характеристик кожної одиниці спостереження до зведених характеристик сукупності в цілому або її груп.

3) На третьому етапі отримані зведені дані аналізуються методом узагальнюючих показників (абсолютні, відносні і середні величини, показники варіації, індексні системи, методи математичної статистики, табличний метод, графічний метод тощо). Основний зміст цього етапу полягає у виявленні взаємозв'язків явищ, визначенні закономірностей їх розвитку та здійсненні прогнозних оцінок.

Сукупність методів і засобів, які розробляє статистика для своїх досліджень, формує *статистичну методологію*.

Загальною основою розробки і застосування статистичної методології є положення соціально-економічної теорії і принцип діалектичного методу пізнання явищ суспільного життя.

Згідно з діалектичним методом пізнання статистика вивчає всі явища і процеси у їх взаємозв'язку і взаємозалежності, в русі і зміні, виділяючи їх різні типи і форми, встановлює нове і прогресивне та визначає напрямки розвитку.

1.5. Галузі статистичної науки

Статистика, як самостійна наука, пройшла складний шлях свого становлення. У процесі розвитку в її складі виділилися: математична статистика, загальна теорія статистики, соціально-економічна статистика, галузеві статистики. Розглянемо коротко зміст кожного із видів статистики.

Математична статистика – це галузь математичних знань. Вона розробляє спеціальні прийоми збирання, систематизації і математичної обробки та аналізу статистичного матеріалу з метою вивчення характерних закономірностей для масових явищ та отримання науково-практичних висновків. У математичній статистиці більшість методів обробки статистичних даних ґрунтується на ймовірнісній природі цих даних. Галузь застосування таких статистичних методів обмежується вимогами, щоб досліджувані явища були підпорядковані достатньо визначеним ймовірнісним закономірностям. Математична статистика абстрагується від конкретного змісту цих масових явищ та використовує при дослідженні математичний апарат. Математична статистика є основою для використання математичних методів у будь-якій статистичній науці.

Загальна теорія статистики, якій присвячений даний посібник, містить принципи статистичної науки стосовно до різних сторін суспільного життя, тобто *загальні правила і методи статистичного дослідження*. Вона розробляє основні поняття і категорії статистичної науки, розглядає у загальному вигляді методи збирання, зведення, узагальнення та аналізу статистичних даних. Курс загальної теорії статистики побудований відповідно до стадій статистичного дослідження. Предметом пізнання загальної теорії статистики є найбільш загальні властивості кількісних відносин соціально-економічних явищ. До складу цієї навчальної дисципліни входять такі розділи: статистичне спостереження, статистичне групування, середні величини, вибіркоче спостереження, ряди динаміки, індекси, статистичні графіки. *Загальна теорія статистики* розробляє загальні показники і методи вивчення структури явищ і змінювання їх у часі, закономірностей і тенденцій їх розвитку і причинно-

наслідкових зв'язків між ними, а також принципи і методи статистичного моделювання і статистичного прогнозування. Показники і методи загальної теорії статистики використовуються всіма іншими галузями статистики.

Соціальна статистика – це галузь статистики, яка вивчає кількісну і якісну сторони масових *суспільних* явищ і процесів, що відбуваються у соціальному житті, і розробляє інтегровану систему показників здійснення соціальних процесів і явищ. Такі показники всебічно характеризують стан і розвиток соціальних умов життя, розкривають існуючі тенденції і закономірності розвитку суспільства. Використовуючи статистичні методи, соціальна статистика вивчає політичну, планову й ідеологічну сторони життя, такі його соціальні аспекти як формування особистості, сім'ї, добробут населення. До найважливіших показників соціальної статистики належать показники складу і устрою суспільства, структура і склад населення країни, рівень його життя, освіти і культури, стан здоров'я і медичного обслуговування, зайнятість трудових ресурсів, рівень реальних доходів, споживання матеріальних благ і послуг, житлово-комунальні й побутові умови, умови праці та відпочинку тощо.

Економічна статистика як галузь єдиної статистичної науки, використовуючи положення загальної теорії статистики, вивчає кількісну сторону масових *суспільних* явищ і процесів у сфері *матеріального виробництва* з метою виявлення пропорцій, тенденцій і закономірностей їх розвитку, тобто кількісно характеризує дію економічних законів, досліджує обсяг, структуру і динаміку явищ, показує взаємозалежності економічних процесів з урахуванням конкретних природних та історичних умов розвитку суспільства. Об'єктом вивчення економічної статистики є процеси економічного відтворення, його здійснення в умовах переходу до ринкових відносин і кінцеві результати для народного господарства в цілому. Її предметом, як галузі практичної діяльності держави, є кількісна сторона масових економічних явищ, які в сукупності характеризують народне господарство.

Галузеві статистики вивчають показники процесу виробництва в галузях матеріального виробництва (сільському господарстві, промисловості), в галузях, де продовжується процес виробництва у сфері обігу (торгівля, зв'язок, транспорт тощо); показники роботи галузей невиробничої сфери (житлово-комунального господарства, науки, фізичної культури і спорту тощо). До галузевих статистик належать також і деякі розділи статистики, пов'язані із функціональним аспектом диференціації економічних наук: статистика праці, статистика фінансів. Ці статистики розвивають і доповнюють методи і систему показників, розроблених загальною теорією статистики та економічною статистикою стосовно до особливостей конкретних галузей. В результаті інтеграції статистики з математичною статистикою і теорією ймовірностей намітився розділ статистики – статистичне моделювання і прогнозування.

1.6. Завдання статистики в сучасних умовах та її організація в Україні

Завдання статистичної науки тісно пов'язані з практичними потребами державного управління і керівництва розвитком економіки і соціальної сфери. Кожний новий етап розвитку суспільства висуває перед статистикою нові конкретні завдання. На сучасному етапі завдання статистики визначаються актуальними проблемами здійснення радикальних економічних реформ, переходу від командно-адміністративних форм управління до економічних, ринкової економіки тощо.

Статистика виконує важливу роль у механізмі управління економікою. Наявність об'єктивної і своєчасної інформації про соціально-економічні явища і процеси – необхідна умова ефективних управлінських рішень на державному і регіональних рівнях. Склад статистичної інформації в умовах ринкової економіки значною мірою визначається практичними потребами суспільства. Якість і достовірність статистичних даних – основа ефективних рішень, що сприяють успішному реформуванню економіки.

Перехід від командно-адміністративної економіки до ринкової потребує побудови принципово нової статистики – ринкової.

Найважливішими завданнями статистики є:

1. Збирання, розробка, узагальнення та аналіз статистичних даних, які характеризують стан і розвиток суспільного виробництва і соціальної сфери України.

2. Всебічне дослідження на основі науково обґрунтованої системи статистичних показників корінних перетворень економічних і соціальних процесів, що відбуваються у суспільстві. Реалізація державної політики в галузі статистики.

3. Своєчасне забезпечення вичерпною і достовірною інформацією законодавчих, виконавчих, управлінських та господарських органів, а також широкої громадськості.

4. Розробка статистичних даних, які показують хід виконання планів економічного розвитку, а також статистичних даних, необхідних для складання планів і розробки прогнозів розвитку суспільства.

5. Пропаганда передового досвіду, розробка даних, які характеризують ефективність нових технологій, форм організації праці і виробництва, ринкових відносин, міжнародних економічних зв'язків тощо.

6. Всебічне і глибоке вивчення пропорцій і взаємозв'язків у економіці, факторів і резервів підвищення ефективності суспільного виробництва, зростання продуктивності праці і життєвого рівня населення.

7. Удосконалення статистичної методології і системи статистичних показників, які характеризують стан і розвиток економіки і культури України, збирання та обробки статистичної інформації, економіко-статистичного аналізу соціально-економічних явищ і процесів, що відбуваються у суспільному житті України. Упровадження новітніх інформаційних технологій з обробки статистичної інформації.

8. Розробка питань створення загальнодержавної автоматизованої системи збирання, нагромадження, обробки та аналізу статистичної інформації на основі новітньої обчислювальної техніки. Збереження та захист статистичної інформації.

9. Забезпечення розробки, вдосконалення та впровадження системи державних класифікаторів техніко-економічної та соціальної інформації, які використовуються для проведення статистичних спостережень.

10. Розробка заходів щодо зближення вітчизняної методології статистичних досліджень з методологією і стандартами міжнародної статистики, яку здійснюють Статистична комісія ООН, Міжнародний статистичний інститут та інші міжнародні організації, впровадження у статистичну практику системи національних рахунків ООН.

Це найбільш загальні і важливі завдання національної статистики; на окремих етапах розвитку суспільства вони насичуються новим конкретним змістом.

Система статистичних органів України відповідає державному устрою та адміністративно-територіальному поділу країни.

Керівним організаційним і методологічним центром статистики в Україні, який здійснює централізоване керівництво справою обліку і статистики, є Державний комітет статистики України (Держкомстат України). Він проводить свою роботу через відповідні статистичні органи на місцях. У автономній республіці Крим таким органом є Державний комітет статистики Криму, а в областях – головні статистичні управління. Низовими органами державної статистики, які знаходяться в безпосередньому підпорядкуванні головних статистичних управлінь, є районні і міські управління статистики.

Структурні підрозділи Державного комітету статистики України мають такий склад:

Департамент макроекономічної статистики.

Департамент статистики виробництва.

Департамент міжгалузевої статистики підприємств.

Департамент статистики послуг.

Департамент статистики сільського господарства та навколишнього середовища.

Департамент обстежень домогосподарств. Департамент статистики торгівлі. Департамент статистики праці.

Департамент статистики населення та адміністративно-територіального устрою.

Департамент статистики цін. Департамент інформатизації. Департамент зведеної інформації.

Департамент планування та організації статистичних спостережень. Департамент фінансового забезпечення, бухгалтерського обліку та звітності.

Департамент кадрово-організаційного забезпечення. Адміністративно-господарський департамент.

Управління статистики зарубіжних країн та міжнародного співробітництва.

Сектор правового забезпечення та зв'язків з Верховною Радою України.

Сектор контрольно-ревізійної роботи. Сектор режимно-секторної роботи. Сектор контролю виконання.

Всередині департаментів є ряд відділів. Зазначені органи становлять єдину систему органів державної статистики України.

При Держкомстаті України функціонує Науково-дослідний інститут статистики, який веде узагальнення наукових досліджень з теорії та методології статистики.

Державний комітет статистики України здійснює державне управління всією, що знаходиться в його веденні, єдиною системою статистичних органів, справою статистики, обліку і звітності у всіх галузях економіки, створенням і функціонуванням статистичної інформаційної системи на основі єдиної наукової методології.

Права, обов'язки і функції органів державної статистики визначені Законом України "Про державну статистику" (1992 р.).

У Законі України "Про державну статистику" у ст. 1 зазначається, що державна статистика – це централізована система збирання, опрацювання, аналізу, поширення, збереження, захисту та використання статистичної

інформації, а державна статистична діяльність – сукупність заходів, пов'язаних з проведенням державних статистичних спостережень і наданням інформаційних послуг, спрямована на збирання, опрацювання, аналіз, поширення, збереження, захист і використання статистичної інформації, забезпечення її достовірності, а також удосконалення статистичної методології.

Головним елементом системи статистичних органів є державна статистика.

Поряд із загальнодержавною статистикою існує відомча статистика, яка обслуговує підприємства, відомства, міністерства. Вона виконує роботи, пов'язані із збиранням та аналізом статистичної інформації, потрібної для керівництва і планування їх діяльності. Для ведення статистики на підприємствах, об'єднаннях, концернах, асоціаціях, міністерствах утворені спеціальні статистичні служби.

1.7. Міжнародні статистичні організації

Процес інтеграції у світовій економічній простір потребує розширення міждержавних економічних зв'язків і контактів з міжнародними статистичними організаціями. Все більш актуальною проблемою стає потреба в міжнародних статистичних дослідженнях. Такі дослідження можуть здійснюватися тільки за умови використання порівнянних даних, одержаних за єдиною методологією, тобто вони повинні бути гармонізованими.

Перехід на міжнародні стандарти статистики та обліку зумовили зміни у вітчизняній системі показників. Це насамперед стосується запровадження національного рахівництва – системи національних рахунків (СНР), яка широко використовується у світовій практиці і відповідає особливостям і вимогам ринкових відносин. Саме СНР дає змогу розрахувати найважливіші макроекономічні показники відповідно до міжнародних стандартів.

Метрологічна робота Держкомстату України, а також підвідомчих організацій координується і проводиться за участі міжнародних статистичних організацій.

При Організації Об'єднаних Націй (ООН), починаючи з 1946 р., працює Статистична комісія ООН, форма роботи якої – регулярні сесії. При ній створена спеціальна робоча група, яка складається з цільових підрозділів, що відповідають основним напрямкам економічної статистики: 1) національне рахівництво; 2) статистика промисловості; 3) статистика міжнародної торгівлі; 4) статистика фінансів; 5) статистика цін; 6) статистика навколишнього середовища.

Діяльність Статистичної комісії ООН охоплює різні питання статистики та організації проведення найважливіших статистичних робіт: переписів населення і житлового фонду, обчислення валового внутрішнього продукту і національного доходу, розробки системи національних рахунків, міжнародних порівнянь. Комісія розробляє методологічні основи статистичного спостереження, стандартні рекомендації по збиранню та аналізу статистичних даних для ведення міжнародної статистики. Розроблені і схвалені робочими групами і групами експертів рекомендації включаються в проекти робочих документів для розгляду на сесіях Статистичної Комісії ООН та для наступного затвердження Комісією економічного і соціального розвитку ради ООН, а по найбільш важливим показникам – Генеральною Асамблеєю ООН.

Комісія проводить роботу по підвищенню міжнародної порівнянності статистичної інформації і координує статистичну діяльність спеціальних закладів ООН: ЮНЕСКО – Комісія ООН по співробітництву в галузі освіти, науки і культури; ФАО – Комісія ООН по продовольству; ВОЗ – Всесвітня організація охорони здоров'я; МОП – Міжнародна організація праці; ВТО – Всесвітня організація торгівлі; СБ – Світовий банк; МВФ – Міжнародний валютний фонд та ін. До регіональних статистичних організацій відносяться: Євростат – статистична організація країн Європейського Союзу, Статистичний комітет Співдружності Незалежних Держав (СНД), утворений в 1992 р. і який здійснює координацію діяльності статистичних служб країн-членів СНД.

Виконавчий орган Комісії – Статистичне бюро ООН, яке займається підготовкою статистичних публікацій, методологічних матеріалів, а також

підготовкою робочих документів для сесій Статистичної комісії ООН. Статистичне бюро ООН видає Статистичний щорічник ООН, Демографічний щорічник ООН, Щорічник національних рахунків, Щорічник промислової статистики та ін.

Між статистичними організаціями немає строгої підпорядкованості. Статистична комісія ООН має офіційний статус "першої серед перших". Вона координує генеральний перелік міжнародних стандартів і класифікацій і несе відповідальність за їх передачу ряду держав.

Міжнародний статистичний інститут зосереджує свою діяльність передусім на статистичній освіті та прикладному застосуванні статистики. Він регулярно публікує матеріали своїх сесій, випускає щорічники за певною програмою показників.

Спільними зусиллями багатьох країн, а також зазначених вище міжнародних статистичних організацій створена Глобальна статистична система, основна мета якої полягає в ефективному використанні наявних ресурсів для здійснення статистичної діяльності на національному і міжнародному рівнях.

Питання для повторення і самоконтролю

1. Від чого походить термін «статистика»? Що Ви знаєте про історію становлення статистики як науки і з чим пов'язують початок статистичної практики?

2. У яких значеннях найчастіше вживається термін «статистика»?

3. Як визначається статистика як наука на сучасному етапі?

4. Що таке статистична сукупність та одиниця цієї сукупності? Яка статистична сукупність називається однорідною та неоднорідною?

5. Що таке якісні та кількісні ознаки статистичної сукупності, дискретні та неперервні, прямі та непрямі, багатовимірні та альтернативні, інтервальні та моментні?

6. Що таке варіанта? Які особливості первинних і вторинних варіант?

7. Що таке статистичні показники і статистичні дані?

8. Що таке статистична закономірність і в чому її особливість?

9. В чому полягає суть закону великих чисел і як він використовується у статистиці?
10. Що таке метод статистики і які групи складових методів він включає в себе?
11. За якими етапами проходить кожне статистичне дослідження і що відбувається на кожному з них?
12. В чому полягає суть статистичної методології і що є загальною основою її розробки і застосування?
13. До якої галузі відноситься математична статистика і чим займається ця статистична наука?
14. Що вивчає загальна теорія статистики і чим вона відрізняється від інших галузей статистики?
15. Що вивчає соціальна статистика?
16. Що вивчає економічна статистика?
17. Які Ви знаєте галузеві статистики і в чому полягають їхні особливості?
18. Які найважливіші завдання повинна вирішувати статистика на сучасному етапі в нашій державі?
19. Що таке Держкомстат України та які структурні підрозділи входять до нього?
20. Що таке державна і відомча статистика і в чому полягають їхні функції?
21. Чим обумовлюється необхідність проведення міжнародних статистичних досліджень на сучасному етапі?
22. Які функції виконує статистична Комісія ООН і чим займається її виконавчий орган – Статистичне бюро ООН?

Лекція 2. Статистичне спостереження

План

- 2.1. Суть, джерела та організаційні форми статистичного спостереження.
- 2.2. Види статистичного спостереження та способи отримання даних.
- 2.3. Загальний план статистичного спостереження.
- 2.4. Програма і статистичний формуляр спостереження.

2.5. Організаційний план статистичного спостереження, забезпечення точності даних.

2.6. Помилки спостереження і контроль вірогідності даних.

2.1. Суть, джерела та організаційні форми статистичного спостереження

Статистичне спостереження – це спланована, науково організована реєстрація масових статистичних даних про економічні, соціальні та інші явища і процеси.

Статистичні дані – це масові системні кількісні характеристики соціально-економічних явищ та процесів.

Статистичні дані можна одержати різними шляхами і способами.

Залежно від організації статистичного спостереження розрізняють його три основні форми: 1) звітність; 2) спеціально організоване статистичне спостереження; 3) реєстри.

Звітність називають такий вид спостереження, при якому відомості надходять у статистичні органи від підприємств, установ та організацій у вигляді обов'язкових звітів про їхню діяльність. Подання звітності за затвердженою програмою у встановлені адреси і строки є обов'язковим для кожного підприємства, установи та організації. Статистична звітність – основна форма організації статистичного спостереження в Україні.

Джерелом відомостей для статистичної звітності є первинні облікові записи в документах бухгалтерського та оперативно-технічного обліку.

Звітність здійснюється за строго встановленими формами, затвердженими Державним комітетом статистики України. Звітність за незатвердженими статистичними органами формами, як і звітність, що здійснюється частіше встановленої періодичності, вважається незаконною. Органам статистики надано право відмінити незаконну звітність.

Перелік усіх форм із зазначенням їхніх реквізитів називається табелем звітності. Обов'язкова сукупність деяких зовнішніх елементів зафіксованих в кожній формі звітності називається реквізитом даної форми.

Основними реквізитами статистичної звітності є: 1) найменування форми; 2) номер і дата затвердження форми звітності; 3) адреси, в які подається звітність; 4) період, за який подаються відомості або на яку дату; 5) строки подання звітності; 6) назва підприємства або установи, яка подає звіт, і його адреса; 7) назва міністерства (відомства), якому підпорядковане підприємство; 8) підписи посадових осіб, відповідальних за складання звіту.

Звітність підрозділяють на *загальнодержавну* і *відомчу*. Загальнодержавна обов'язкова для всіх підприємств, установ і організацій, вона надходить і узагальнюється в органах загальнодержавної статистики для потреб державного управління. Відомча звітність збирається для своїх потреб міністерствами, відомствами (наприклад, звітність по сільському господарству, торгівлі тощо).

Розрізняють типову і спеціалізовану звітність. Типова має однакові показники для однієї або всіх галузей народного господарства. У спеціалізованій звітності показники відбивають специфіку тієї або іншої галузі і підгалузі (наприклад, звіти по птахівництву, рибництву, звірівництву тощо).

За періодичністю подання звітність підрозділяється на *річну* і *поточну*: *квартальну*, *місячну* і *тижневу*. Найбільш повною за складом показників, що подаються, є річна звітність.

За способами подання розрізняють звітність термінову, коли відомості передаються по телетайпу, телеграфу та іншими швидкими засобами, і поштову.

Поряд із звітністю важливим джерелом статистичних даних є *спеціально організоване статистичне спостереження*.

Спеціально організоване статистичне спостереження являє собою збирання відомостей за допомогою переписів, одночасних обліків і обстежень. Прикладом спеціально організованого статистичного спостереження є: переписи населення, устаткування, залишків матеріалів, багаторічних насаджень, обстеження домогосподарств, облік одержаної продукції в

особистих підсобних господарствах, переоцінка основних фондів у галузях економіки та ін.

У результаті спеціальних статистичних спостережень отримують дані про явища і процеси, які не охоплені статистичною звітністю, а в ряді випадків суттєво уточнюючих дані поточного обліку.

Статистичні переписи є другою за значенням організаційною формою статистичного спостереження в Україні. Перепис являє спеціально організоване статистичне спостереження, спрямоване на облік повної чисельності і складу певних об'єктів (явищ), а також на встановлення якісних характеристик їх сукупностей на певний момент часу.

Існує два види переписів. Перший вид переписів – це переписи, в яких статистичні формуляри заповнюються на основі даних первинного обліку підприємств, установ і організацій. Другий вид – переписи, при яких формуляри заповнюються на основі спеціально організованої реєстрації фактів (прикладом цього виду є перепис населення).

Відмінними особливостями переписів є: одночасність проведення їх на всій передбаченій території, єдність програми спостереження, стислість строків статистичного спостереження, реєстрація всіх одиниць спостереження станом на один і той самий момент часу – критичну дату перепису.

Серед статистичних переписів особливе місце належить переписам населення – спеціально організоване статистичне спостереження, метою якого є одержання даних про чисельність, склад і розміщення населення. Перепис населення 2001 р. дав змогу одержати повну і детальну статистичну інформацію про чисельність і склад населення країни за статтю, віком, сімейним складом, національністю, зайнятістю, суспільними групами, а також про розміщення населення по території країни. Програма перепису складалась з 19 запитань суцільного перепису.

1. Ваші родинні стосунки з особою, що записана першою в домогосподарстві.

2. Ваша стать.

3. Ви тимчасово: а) відсутній; б) проживаючий.
4. Ваша дата народження.
5. Ваше місце народження.
6. Ваше етнічне походження.
7. Ваші мовні ознаки.
8. Ваше громадянство.
9. Ваш сімейний стан.
10. Ваша освіта.
11. Ви закінчили професійно-технічний навчальний заклад.
12. Тип навчального закладу, в якому Ви навчаєтесь.
13. Ваші джерела існування.
14. Вид діяльності та повна назва установи, організації, підприємства, власної справи, де Ви були зайняті основною роботою.
15. Ваше заняття на основній роботі.
16. Місцезнаходження Вашої роботи.
17. Ваше положення у занятті (в якості кого Ви працювали).
18. У цьому населеному пункті Ви проживаєте безперервно з дня народження: а) так; б) ні.
19. Ви народжували дітей (для жінок у віці 15 років і старших):
а) так; б) ні.

До кожного питання наводились необхідні підказки.

Записи у переписній документації, що стосуються респондента, не підлягають поширенню без його згоди і використовуються лише для статистичних цілей у зведеному знеособленому вигляді. Переписному персоналу забороняється повідомляти будь-кому, крім респондента, зміст заповненої щодо нього переписної документації.

Статистичний реєстр (від лат. *registrum* – список, перелік) – це така організаційна форма спостереження, за якою факти стану окремих одиниць сукупності безперервно реєструються за передбаченими програмою ознаками.

Спостерігаючи за одиницями сукупності, передбачають, що процеси, які відбуваються там, мають початок, довготривале продовження і кінець. У реєстрі кожна одиниця спостереження характеризується сукупністю показників. Всі показники зберігаються до тих пір, поки одиниця спостереження знаходиться у реєстрі і не закінчила свого існування. Деякі показники залишаються незмінними весь час, поки одиниця знаходиться у реєстрі, інші можуть змінюватися час від часу.

Ведення реєстру забезпечує лише формування списку одиниць сукупності спостереження з його найважливішими описовими ознаками, таких як назва підприємства, його юридична адреса та місцезнаходження, основний вид діяльності.

Прикладами реєстрів, які ведуться в Україні, є: Єдиний державний реєстр підприємств і організацій України (ЄДРПОУ), реєстр виборців, реєстр акціонерного товариства, реєстр населення, банківський реєстр, реєстр юридичних і фізичних осіб - платників податків та ін.

Популярності сьогодні набуває *моніторинг* – спеціально організоване статистичне спостереження за станом об'єкта чи середовища. Основним завданням моніторингу є оцінка стану та динаміки, розвитку об'єкта чи середовища за визначеним переліком показників, які дають змогу контролювати, регулювати й управляти дослідженим об'єктом чи середовищем.

Відомими є приклади застосування моніторингу у вивченні і прогнозуванні гідрометеорологічних показників навколишнього природного середовища, його екологічного стану; рівня і динаміки цін, рейтингів, біржових індексів тощо.

2.2. Види статистичного спостереження та способи отримання даних

За часом проведення статистичне спостереження поділяють на безперервне і перервне.

Під *безперервним* (поточним) розуміють спостереження, при якому встановлення і реєстрація фактів проводиться мірою їх здійснення (наприклад, облік випуску продукції, виконання робіт, реєстрація народжень і смертей,

шлюбів і розлучень, доходів і витрат у сім'ях при вибірковому обстеженні домогосподарств тощо). Поточне спостереження проводиться на основі первинних документів, які містять інформацію, необхідну для достатньо повної характеристики досліджуваного явища.

Перервним називають таке спостереження, яке проводиться або регулярно через певні проміжки часу, або у разі потреби. Його поділяють на періодичне і одночасне.

Періодичним називають таке спостереження, яке здійснюють регулярно через певні проміжки часу (наприклад, щорічні обліки худоби, заключний облік посівних площ, який проводять один раз на рік після закінчення сівби ярих культур тощо).

Одночасне – таке спостереження, яке проводять у разі потреби для вирішення якогось завдання (наприклад, перепис багаторічних насаджень, переоцінка основних фондів у галузях економіки, перепис житлового фонду, перепис і бонітування тварин, паспортизація полів і т.п.).

Залежно від ступеня охоплення одиниць досліджуваної сукупності розрізняють два види статистичного спостереження: суцільне та несуцільне.

Суцільним називають таке спостереження, при якому обстеженню підлягають усі без винятку одиниці сукупності (наприклад, перепис населення, облік виходу продукції та ін.). Матеріали суцільного спостереження дають максимально повне уявлення про все розмаїття форм і можливість одержання точних характеристик про досліджувані соціально-економічні явища і процеси.

Однак через низку причин (велика трудомісткість, тривалість проведення, висока вартість тощо) суцільне спостереження часто буває економічно недоцільним або практично нездійсненним. Тому на практиці переважно застосовують несуцільне спостереження.

Несуцільним називають таке спостереження, при якому обстежується тільки частина сукупності (наприклад, визначення якості насіння, жирності молока, втрат урожаю, вивчення використання робочого часу і устаткування, цін на ринках тощо.).

Несуцільне спостереження базується на обліку деякої частини, як правило, достатньо масової частини одиниць спостереження, яка дає змогу на основі наукового відбору одиниць одержати стійкі узагальнюючі характеристики всієї сукупності.

Якість несцільного спостереження поступається результатам суцільного, однак досить очевидні і деякі переваги першого: за рахунок зменшення числа одиниць сукупності, що обстежуються, воно потребує менших витрат, сил і засобів, дає змогу застосувати більш детальну програму спостереження, швидше підводити підсумки обстежень і, отже, підвищує оперативність статистичного матеріалу.

Несуцільне спостереження своєю чергою підрозділяється на вибіркоче, монографічне, анкетне і основного масиву.

Вибірковим називають таке спостереження, при якому обстеженню підлягає *певна частина сукупності явищ*, яку отримали на основі ненавмисного випадкового відбору.

Монографічне спостереження – це докладний і всебічний опис окремих одиниць досліджуваної сукупності, які цікавлять дослідника. Монографічне обстеження може бути спрямоване на вивчення процесу розвитку окремого трудового колективу, узагальнення передового досвіду, опис нових технологій виробництва і форм організації праці та ін. Об'єктом монографічного опису можуть бути також сім'я, навчальний заклад, місто, регіон та інші об'єкти.

Анкетний спосіб спостереження ґрунтується на розсиланні анкет певному колу осіб або установ. Заповнення і повернення їх до органів, які проводять спостереження, є добровільним. Як правило, заповнених анкет повертається менше, ніж розсилається. Крім того, неможливо проконтролювати правильність відповідей на запитання анкети. Тому такий спосіб спостереження може застосовуватись у тих випадках, коли не вимагається висока точність відомостей, а потрібні приблизні характеристики. До нього удаються при проведенні соціологічних обстежень, у торгівлі для вивчення попиту на окремі товари і т.п.

Обстеження основного масиву являє собою спостереження за частиною найкрупніших одиниць, питома вага яких переважає в загальному обсязі досліджуваної сукупності. За принципом основного масиву в країні організоване спостереження за міською ринковою торгівлею. Число охоплених нею міст складає менше 5% усіх міст, однак в них мешкає більше половини чисельності всього міського населення країни. Таке спостереження дає змогу отримати достатньо надійний матеріал для загальної характеристики кон'юнктури міських ринків.

За способом отримання статистичних даних розрізняють три основних види спостереження: 1) безпосереднє; 2) документальне та 3) опитування.

Безпосереднім називають спостереження, яке здійснюється шляхом реєстрації досліджуваних одиниць та їхніх ознак на основі безпосереднього огляду, підрахунку, зважування, зняття показників приладів спеціальними особами, які проводять спостереження, інакше кажучи, реєстраторами, в завдання яких входить поряд з встановленням і оцінкою фактів фіксування їх у документах первинного обліку (наприклад, щорічний облік худоби, перепис багаторічних насаджень, інвентаризація матеріальних цінностей, метеорологічні спостереження – реєстрація температури повітря, ґрунту, снігового покриву, кількості опадів; облік одержаної продукції тощо).

Безпосереднє спостереження є досить точним і надійним джерелом статистичних даних, воно потребує великих затрат кваліфікованої праці.

Документальне спостереження ґрунтується на використанні різних документів первинного обліку підприємств, організацій установ (звітності, первинних бухгалтерських документів, річних звітів та ін.). Воно застосовується, наприклад, при переоцінці основних фондів підприємств, аналізі використання тракторів і автомобілів, урожайності і продуктивності тварин та ін. Цей вид спостереження також, як і безпосереднє спостереження, забезпечує найбільшу вірогідність статистичних даних.

Опитування – це спосіб спостереження, при якому відомості отримують зі слів опитуваних осіб. Опитування може бути усним та письмовим.

Розрізняють три способи опитування: експедиційний, самореєстрації і кореспондентський.

Експедиційний спосіб опитування полягає в тому, що спеціально підготовлені реєстратори на основі опитування обстежуваних осіб заповнюють переписні формуляри, одночасно контролюючи правильність отриманих відповідей (наприклад, перепис населення). Цей спосіб опитування потребує найбільших витрат, проте гарантує високу якість матеріалу і дає змогу включати до програми такі запитання, які ризиковано ставити при інших способах опитування. Найважливіші статистичні обстеження населення проводяться експедиційним способом.

При способі самореєстрації (самообчисленні) спеціально виділені особи безпосередньо вступають в контакт з тими, від кого потрібно одержати відомості, роздають бланки, інструктують про порядок їх заповнення і призначають час, до якого вони мають бути заповненими; потім в призначений час ці особи отримують заповнені бланки, перевіряють повноту і правильність їх складання. Прикладом такого спостереження є обстеження умов життя домогосподарств, при якому сім'ї самі ведуть записи про свої доходи і витрати, а статистики з проведення вибіркового обстеження населення регулярно (щоквартально) при відвідуванні сімей перевіряють повноту і правильність цих записів.

При способі самореєстрації забезпечується суттєва економія робочого часу, оскільки лічильники звільнені від необхідності заповнювати формуляри. Якість отриманих даних поступається усе ж даним, отриманим експедиційним способом.

Суть кореспондентського способу спостереження полягає в тому, що необхідні відомості надають особи, які добровільно виявили бажання відповісти на поставлені в анкетах запитання. Цей спосіб не потребує великих витрат, але він не забезпечує високої якості матеріалів, так як перевірити точність відомостей, що повідомляються, безпосередньо на місцях не завжди можливо.

Вимоги до якості даних статистичного спостереження:

1. Вірогідність даних – їх відповідність реальному стану.
2. Повнота даних – як за обсягом так і по суті.
3. Своєчасність даних – інформація не може бути застарілою.
4. Доступність даних.

2.3. Загальний план статистичного спостереження

План статистичного спостереження – це сукупність програмно-методологічних та організаційних питань. Він охоплює широке коло питань методики та організації збору статистичної інформації, контролю її якості та вірогідності.

Загальний план статистичного спостереження складається з двох частин: програмно-методологічної та організаційної.

Програмно-методологічна частина плану – це визначення мети, встановлення об'єкта, одиниць спостереження елементів сукупності, складання програми спостереження.

Мета спостереження визначається конкретними потребами в статистичних даних.

Об'єкт спостереження – це сукупність явищ, що вивчається. Слід чітко визначити його межі, істотні ознаки та характерні риси.

Одиниця статистичного спостереження є джерелом інформації.

Носіями ознак, що підлягають реєстрації, є *елементи сукупності*.

Організаційна частина плану спостереження визначає місце, час і органи спостереження, графік підготовки і інструктажу кадрів.

Місцем спостереження вважають пункт, де безпосередньо реєструються ознаки окремих одиниць сукупності в статистичних формулярах.

Час спостереження поділяють на об'єктивний і суб'єктивний. *Об'єктивним* називають той час, до якого відносяться дані спостереження.

Момент часу, станом на який проводиться реєстрація ознак елементів сукупності, набув назви *критичного*.

Період, протягом якого реєструються ознаки об'єкта спостереження, називають *суб'єктивним часом*.

2.4. Програма і статистичний формуляр спостереження

Програма статистичного спостереження являє собою перелік питань, на які треба одержати відповіді в процесі збирання статистичних зведень щодо кожної досліджуваної одиниці. Один і той самий об'єкт може бути обстежений з різних боків. Тому склад і зміст питань програми спостереження залежить від завдань дослідження і особливостей об'єкта. Вона повинна охоплювати широке і повне коло відомостей. Чим ширша програма, тим повніше висвітлюється досліджуване явище. Проте в неї не слід включати зайвих питань, які могли б ускладнити і розтягнути термін розробки даних. У той же час не слід складати програму надто вузько, адже в дослідження можуть не потрапити важливі питання.

При складанні програми велике значення має чітке формулювання питань, оскільки у більшості статистичних спостережень це складна і трудомістка робота, у виконанні якої беруть участь десятки і навіть сотні тисяч (при перепису населення) чоловік. Поставлені питання мають бути однаково зрозумілими для всіх.

Відповіді на питання програми спостереження записують у документ особливої форми – *статистичний формуляр*. Він являє собою первинний документ, в якому фіксують відповіді на питання програми по кожній з одиниць сукупності, це носій первинної інформації. Формуляри мають різні назви: форма первинного обліку або звітності, акт, бланк, таблиць, картка (фішка), анкета, опитувальний листок. Для всіх перелічених видів формулярів характерні деякі обов'язкові елементи: змістовна частина, яка включає перелік питань програми, зведена графа або декілька граф для запису відповідей і шифрів (кодів) відповідей, титульна і адресна частини. На титульній сторінці записується назва статистичного спостереження (наприклад, "Перепис худоби в приватному секторі"), назва організації, яка проводить спостереження, а також

зазначається, ким і коли затверджено формуляр або статистичне спостереження.

В адресну частину записують адресу обстежуваних (опитуваних) одиниць спостереження. Крім того, у формах статистичної звітності вказується, коли і куди треба надсилати одержану інформацію. Правильність даних обстеження стверджується підписами відповідальних осіб.

У практиці статистичного спостереження застосовують формуляри двох видів: картковий і списковий.

Картковим (або індивідуальним) називається статистичний бланк (фішка), який містить дані лише про одну одиницю спостереження. Загальна кількість карток повинна дорівнювати кількості одиниць досліджуваної сукупності.

Списковий формуляр – це статистичний бланк, в якому реєструються відомості по кількох одиницях спостереження. Наприклад, при перепису населення 1970 р. списковий формуляр було розраховано на реєстрацію у ньому зведень по шести особах, при перепису населення у 1989 р.– по двох.

Формуляри-картки зручні для ручної обробки занесених в них даних, але потребують значно більших трудових і матеріальних витрат, ніж формуляри-списки. Останні економічніші і зручніші для машинної обробки і контролю даних. Карткові формуляри використовуються при складанні звітів підприємств і закладів, які потребують широкої програми статистичного спостереження.

Спискові формуляри частіше використовуються при періодичних спостереженнях. Прикладом такого формуляра може бути переписний лист перепису населення 1989 р. Конструкція статистичних формулярів зумовлюється значною мірою впливом з боку технічних засобів обробки інформації. Це знаходить свій прояв у тому, що формуляри статистичного спостереження у деяких випадках виступають одночасно і безпосередніми носіями інформації при введенні їх в компютери. Саме за таких умов вирішується питання безперервної технології автоматизації статистичних робіт, коли первинна інформація заноситься в автоматизований-банк даних за

допомогою мережі Інтернет по каналах зв'язку, які з'єднують обчислювальні центри підприємств і державної статистики.

Вплив автоматизованої системи обробки статистичної інформації на конструкцію формулярів знаходить прояв у тому, що відповіді в них розташовані у зручному для шифрування (кодування) порядку.

Ефективність виконання розробленої програми спостережень значною мірою зумовлюється якістю інструктивного матеріалу. Для цього складають інструкцію (інколи її називають Статистичною інструкцією, або Доповідною інструкцією).

Інструкція – це документ, який пояснює питання програми статистичного спостереження, його мету, порядок заповнення статистичного формуляра і частково організаційні питання. Інструкція є одним з найважливіших документів спостереження. Вона може містити вказівки відносно тих питань, які виникають у процесі проведення спостереження: об'єкт і одиниця спостереження, час і строки проведення, критичний момент спостереження тощо.

У багатьох випадках необхідні додаткові пояснення того, як правильно розуміти дане питання і як вірно записати відповідь на нього. Як правило, інструкції пишуть для осіб, які здійснюють перепис або заповнюють форми статистичної звітності. Вони дуже важливі для забезпечення однакового розуміння питання у всіх спірних і сумнівних випадках. Наприклад, при перепису населення 1989 р. ставилось запитання "національність". Для запису відповіді особам, які здійснювали перепис, треба було дати певні вказівки, щоб запобігти різнобою у відповідях. Для цього були дані інструкції, де вказувалося: "Записується національність, котру вказує сам опитуваний. Національність дітей визначається батьками. Лише в тих сім'ях, де батько і мати належать до різних національностей і батьки утруднюються самі визначити національність дітей, слід надати перевагу національності матері". Таке уточнення порядку записів на запитання перепису дає придатний для обробки первинний матеріал.

Відсутність в інструкції тлумачень певного питання призведе до того, що кожна особа (лічильник) буде розуміти запитання по-своєму, внаслідок чого зібрані матеріали зовсім знеціняться.

Вказівки інструкції повинні бути конкретні й чіткі, а текст стислим і лаконічним. Інструкції до форм статистичної звітності здебільшого друкуються на самій формі документа.

Отже, головне призначення інструкції полягає у поясненні змісту питань програми, як треба давати на них відповіді і заповнювати формуляр. Найбільш типові ситуації повинні розглядатися на прикладі.

Слід зазначити, що методологія розробки програми статистичного спостереження в останні роки зазнала досить значних змін у зв'язку з функціонуванням автоматизованої системи державної статистики (АСДС), яка зумовила створення автоматизованих банків даних (АБД). Наявність останніх дала змогу запровадити в галузі матеріального виробництва реєстрову форму статистичного спостереження, яка полягає у створенні реєстра, або автоматизованої картотеки сукупності одиниць статистичного спостереження певного типу. Перехід до автоматизованої статистичної інформаційної системи (АСІС) створює широкі можливості удосконалення програми спостереження, а саме: відбувається об'єднання інформаційних баз державної статистики, галузевих і регіональних органів управління, підприємств, об'єднань та інших ланок управління.

2.5. Організаційний план статистичного спостереження, забезпечення точності даних

Організаційний план статистичного спостереження – це складова частина загального плану спостереження, в якій викладено порядок його організації і проведення. У ньому даються роз'яснення програмно-методологічних та організаційних питань. До перших належать формулювання мети і завдань спостереження, визначення його об'єкта і одиниці, розробка програми. До організаційних питань належать: фіксація місця, часу і строків спостереження, вказівка на те, хто його проводить, як воно проводиться і як здійснюється

постачання статистичними формулярами осіб, які виконують спостереження, способи доставки заповнених формулярів у відповідні статистичні органи. Сюди відносять також ряд специфічних підготовчих робіт, зокрема таких, як підбір та навчання кадрів, що залучаються до проведення спостереження, підготовка графічного матеріалу та ін.

В організаційному плані статистичного спостереження конкретизуються права і обов'язки окремих установ і організацій, які беруть участь у заходах спостереження.

При плануванні спостереження насамперед визначають органи спостереження – організаторів і виконавців робіт, а також права і обов'язки кожного співвиконавця.

Програму і план статистичного спостереження розробляють органи державної статистики на рівні Міністерства Державного комітету статистики України. Низові органи державної статистики переважно виконують роботу по збору статистичних даних, їх первинному контролю і зведенню по певній програмі.

У плані вказують строк проведення спостереження, тобто час початку і закінчення збирання зведень. Цей час не можна ототожнювати з часом спостереження, тобто часом, до якого належать зведення. Статистичні показники характеризують досліджуване явище або за певний період, або на певний момент часу. Наприклад, дані про кількість виробленої продукції можна взяти тільки за період (день, декаду, місяць, квартал, рік), а показники запасів матеріальних цінностей можуть бути представлені на певний момент часу (на початок місяця, на початок кварталу, на початок або кінець року тощо).

У плані має бути точно визначена територія, на якій здійснюється спостереження, а також особи і організації, відповідальні за проведення підготовчих робіт, збір, перевірку і обробку інформації по окремих ділянках території.

Серед організаційних питань значне місце у плані відводять проведенню підготовчих робіт. Насамперед треба скласти список звітних одиниць, тобто

тих, які будуть обстежені. Цей список (колективних підприємств, державних підприємств, орендних підприємств і т. ін.) необхідний для перевірки повноти зведень, що надходять, а також визначення обсягу робіт і розрахунку необхідної кількості робітників для проведення статистичного спостереження.

Важливим підготовчим заходом є розрахунок потреби в кадрах для проведення спостереження, їх добір та інструктаж. Необхідно заздалегідь віддрукувати і розіслати бланки документів та інструкції щодо їх заповнення. Інструктаж вважається однією з найважливіших підготовчих робіт статистичного спостереження. Успіх проведення останнього багато в чому залежить від рівня підготовленості кадрів.

При підготовці складних статистичних спостережень, як правило, в плані передбачається проведення пробних спостережень з метою перевірки на практиці проекту плану і програми основного спостереження. Матеріали таких пробних спостережень використовують для уточнення, доповнення і конкретизації програми і плану спостереження, а також інструкцій.

Серед підготовчих робіт чільне місце повинне належати (особливо це стосується перепису населення) пропаганді спостереження серед населення. Засоби масової інформації повинні вести роз'яснювальну роботу щодо завдань і мети спостереження, що значною мірою сприятиме підвищенню ефективності вирішення планово-організаційних питань спостереження. Роз'яснювальна робота про мету і завдання проведення спостереження здійснюється через пресу, радіо, телебачення та інші засоби масової інформації.

Таким чином, *організаційний план статистичного спостереження передбачає: визначення часу спостереження, часу і місця його проведення, порядок передачі матеріалів спостереження, комплекс підготовчих робіт, заходи, що забезпечують точність (вірогідність) статистичних даних.* Кожній з названих вище категорій і етапів статистичного спостереження можна дати такі визначення.

Час спостереження – це момент або період часу, якого стосується статистична інформація (дані). Наприклад, інформацію про виробництво

сільськогосподарської продукції збирають за певний період часу – день, декаду, місяць, квартал, рік.

Перепис населення передбачає належність даних до певного (критичного) моменту часу. Так, критичним моментом перепису населення 1989 р. було 24 години у ніч з 11 на 12 січня 1989р.

Час проведення спостереження, як правило, не збігається з часом спостереження. Наприклад, перепис населення 1989 р. здійснювався впродовж восьми днів – з 12 до 19 січня.

Місце проведення спостереження вибирають, виходячи зі способу спостереження і особливостей об'єкта спостереження. Так, інформацію про виробництво продукції сільськогосподарським підприємством реєструють у бухгалтерії підприємства, а при перепису населення - вдома, у поїздах, на вокзалах тощо.

Органами спостереження можуть бути різні установи системи державної статистики, відомства, підприємства і громадські організації. Серед них визначають організатора (установа системи державної статистики) і виконавця. Права і обов'язки кожного учасника статистичного спостереження повинні бути чітко визначеними.

Органи державної статистики повинні не тільки організувати проведення статистичного спостереження, а й ретельно перевіряти точність одержаних результатів при його здійсненні. Перевірка вірогідності даних - найважливіша умова успішної роботи у справі спостереження. Щоб забезпечити вірогідність даних, необхідно повсякденно, систематично контролювати, чи вірно зрозумілі і застосовуються статистичні програми та інструкції, чи забезпечується повнота одержаних зведень, правильність даних бухгалтерського і оперативного обліку.

У процесі здійснення спостереження можуть бути допущені помилки, їх називають помилками спостереження. Види і природа виникнення таких помилок будуть розглянуті далі.

2.6. Помилки спостереження і контроль вірогідності даних

Розбіжності між даними спостереження і дійсним значенням показників вважають помилками спостереження. Розрізняють *помилки реєстрації і репрезентативності*.

Помилками реєстрації називають ті, які виникли внаслідок неправильного встановлення фактів або неправильного їх запису. Вони допускаються випадково або систематично.

Випадкові помилки виникають внаслідок дії випадкових причин і спричиняють спотворення даних в той чи інший бік.

Систематичні помилки призводять до значних зміщень загальних підсумків статистичного спостереження, іноді вони виникають внаслідок не досить чіткого формулювання програми спостережень.

Помилки репрезентативності виникають лише в несудільному спостереженні тому, що відібрана і обстежена частина сукупності не повністю відтворює склад сукупності в цілому.

Питання для повторення і самоконтролю

1. В чому полягає суть статистичного спостереження і що таке статистичні дані?
2. Які існують основні форми статистичного спостереження?
3. Що таке статистична звітність і що є джерелом відомостей для неї?
4. Які основні реквізити має статистична звітність?
5. Що таке загальнодержавна і відомча, типова і спеціалізована звітність?
6. В чому суть спеціально організованого статистичного спостереження і в якій формі воно проводиться?
7. Яка особливість перепису населення як спеціально організованого статистичного спостереження?
8. Що таке статистичний реєстр і які саме реєстри використовуються в Україні?
9. В чому полягає завдання моніторингу при проведенні спеціально організованого статистичного спостереження?

10. Як проводиться безперервне, перервне, періодичне та одночасне статистичне спостереження?
11. Що таке суцільне і несуцільне статистичне спостереження та які вони мають переваги і недоліки?
12. Що таке вибіркове спостереження і яка його особливість?
13. В чому суть монографічного спостереження і в якій формі воно проводиться?
14. На чому ґрунтується анкетний спосіб спостереження і в яких випадках він застосовується?
15. Які основні види спостереження розрізняють за способом отримання статистичних даних і яка суть кожного з них?
16. Які вимоги пред'являються до якості даних статистичного спостереження?
17. Які функції виконує програмно-методологічна частина загального плану статистичного спостереження?
18. Що передбачає організаційна частина загального плану статистичного спостереження?
19. Що являє собою програма і формуляр статистичного спостереження? Що таке картковий і списковий формуляри та коли вони використовуються?
20. Яке значення має інструкція, що додається до розробленої програми статистичного спостереження?
21. Які питання відносяться до організаційного плану статистичного спостереження?
22. Які підготовчі роботи і заходи передбачаються в організаційному плані статистичного спостереження?
23. Які засоби передбачаються організаційним планом для забезпечення точності статистичного спостереження?
24. Які помилки можливі у процесі статистичного спостереження і в чому суть помилок реєстрації і репрезентативності?

Лекція 3. Зведення і групування статистичних даних

План

- 3.1. Суть і завдання статистичного зведення і статистичних групувань.
- 3.2. Основні правила утворення груп.
- 3.3. Типологічні, структурні та аналітичні групування.
- 3.4. Прості та комбінаційні групування.
- 3.5. Вторинні групування.
- 3.6. Багатовимірні групування.
- 3.7. Основні статистичні класифікації і класифікатори України.

3.1. Суть і завдання статистичного зведення і статистичних групувань

Результатом кожного статистичного спостереження є велика кількість інформації про кожну одиницю досліджуваної сукупності, яка не має систематизованого характеру. Для того, щоб провести наступні дослідження на основі одержаних статистичних даних необхідно їх упорядкувати, обробити та виконати зведену характеристику всієї сукупності за допомогою узагальнюючих статистичних показників. Таке завдання вирішується на другому етапі статистичного дослідження, який називається зведенням і групуванням статистичного матеріалу.

Отже, *статистичним зведенням* називається наукова обробка первинних даних статистичного спостереження з метою одержання узагальнюючих характеристик досліджуваного явища чи процесу за низкою суттєвих для них ознак.

Перед початком зведення первинного статистичного матеріалу його необхідно проаналізувати і виключити можливі втрати повноти охоплення обстежуваних одиниць та якості отриманих про них даних. Приступають до виконання зведення лише після того як весь статистичний матеріал проконтрольований і належним чином виправлений.

Розрізняють два типи зведень: просте і складне. *Просте зведення* виконується шляхом безпосереднього підрахунку отриманих числових даних, а

складне зведення передбачає групування даних, розробку системи показників, підрахунок групових і загальних підсумків та представлення результатів у вигляді статистичних таблиць чи графіків.

За формою організації зведення буває *централізоване і децентралізоване*. При централізованій формі організації зведення всі матеріали спостереження обробляють і синтезують в Держкомстаті України. При децентралізованій формі організації зведення статистичні матеріали обробляють та узагальнюють на місцях, а в центральні статистичні органи відправляють зведену інформацію по регіонах. Централізована форма організації зведення має переваги перед децентралізованою в тому, що при цьому використовується єдина методологія обробки даних. Однак децентралізована форма є дещо дешевшою та оперативнішою за централізовану. На практиці поєднують обидві форми зведення.

Для успішного здійснення статистичного зведення розробляється програма і складається план його проведення, у якому передбачається послідовність і строки виконання окремих частин зведення, оформлення його результатів у вигляді таблиць, статистичних збірників тощо.

У процесі обробки статистичних даних виникає потреба у виділенні серед них однорідних груп, типів, без яких неможливо їх описати за допомогою узагальнюючих кількісних характеристик: відносні і середні величини, індекси та інші.

Статистичне групування являє собою розподіл сукупності масових суспільних явищ на однорідні типові групи за суттєвими для них ознаками з метою всебічної характеристики їх стану, розвитку і взаємозв'язків. Наприклад, групування промислових підприємств за формами власності, групування населення за статтю, віком, групування комерційних банків за сумою активів балансу тощо.

За кількістю ознак, покладених в основу групування, розрізняють *прості і комбінаційні групування*. Групування, проведені за однією ознакою, називають

простими або одновимірними, а за двома і більшим числом ознак – комбінаційними або багатовимірними.

При побудові комбінаційного групування сукупність спочатку підрозділяється на групи за однією ознакою, а потім отримані групи поділяються своєю чергою на підгрупи за другою, третьою і т.д. ознаками. В цьому випадку можна прослідкувати вплив на результат сукупності факторів, покладених в основу групування, і оцінити вплив кожного окремо взятого фактора при вирівняних значеннях інших факторів.

Комбінаційне групування дає змогу також оцінити спільний вплив на результат групувальних ознак.

Значення і необхідність групувань випливають з самого предмету статистики. В кожному складному масовому суспільному явищі є якісно відмінні групи, є відмираюче старе і народжуване нове, яке треба виділити і вивчити. Наприклад, у сільському господарстві сформувались різні соціальні типи господарств: приватні і державні сільськогосподарські підприємства, кооперативи, підсобні господарства підприємств і населення тощо. Кожному з цих типів притаманні характерні риси і закономірності. Тому при групуванні не можна змішувати різні типи господарств, а виділяти їх якісно однорідні групи. Чим досконаліше статистика виділить такі групи, тим повніше вона їх охарактеризує, тим глибше розкриє суть процесів, що відбуваються, напрями і темпи їх розвитку тощо.

Найбільш відповідальним моментом групування є відбір ознак, які дозволять відокремити один від одного дійсно суттєво відмінні групи одиниць. Кожна одиниця спостереження володіє багатьма ознаками. Одні з цих ознак виражають суть, найбільш характерне у даному явищі, інші – другорядне, поверхнєве, нетипове.

Якщо в основу групувань покладені найбільш суттєві ознаки, то будуть виділені дійсно типові для даного явища групи. Якщо за основу групування узяти малозначущі ознаки, то ми отримаємо крайньо поверхнєве або взагалі перекручене уявлення про явище.

Метод статистичних групувань, який є одним з найбільш ефективних методів обробки масових даних, відкриває широкі можливості для вивчення взаємозв'язків між явищами, виявлення об'єктивних закономірностей досліджуваних явищ і процесів, встановлення на певному етапі переходу кількісних змін у якісні. Побудувавши групування досліджуваних об'єктів за будь-якою суттєвою ознакою і охарактеризувавши виділені групи різними показниками, можна прослідкувати залежність між ознаками, що покладені в основу групування, і вибраними показниками.

Метод статистичних групувань дає змогу так розробити первинний статистичний матеріал, щоб всі суттєві риси і особливості досліджуваних суспільних явищ отримали чітке вираження. Цим визначається роль групувань як наукової основи зведення.

В економіці важлива роль належить групуванням підприємств за виходом валової продукції на одиницю ресурсів виробництва, продуктивністю праці, собівартістю продукції, рентабельністю, урожайністю, продуктивністю тварин тощо.

Використання методу групувань створює умови для застосування багатьох інших статистичних методів наукового пізнання, перед усім відносних і середніх величин, індексного, кореляційного, дисперсійного методу та ін. Перераховані методи ефективні тільки на основі групувань і в поєднанні з ними.

За допомогою групувань вирішують різні завдання. Найважливішими з них є: 1) виділення і всебічна характеристика різних соціально-економічних явищ; 2) характеристика структури досліджуваних явищ; 3) вивчення взаємозв'язків між окремими ознаками сукупності.

Внаслідок проведення групування статистичних даних за певною ознакою складаються найпростіші таблиці, які називаються *рядами розподілу*.

Якщо групувальна ознака характеризується лише цілими числами (наприклад, вік працівника, стаж роботи, заробітна плата тощо), то вона називається кількісною, а її ряди розподілу – *варіаційними рядами*.

Якщо групувальні ознаки відображають якісну сторону сукупності (наприклад, національність, стать, освіту тощо), то вони називаються якісними, а відповідні ряди розподілу – *атрибутивними*.

Якщо групування одиниць сукупності виконується за територіальною ознакою, то одержані ряди розподілу називаються *територіальними*.

Таким чином, групування, будучи початковим етапом проведення статистичного аналізу, водночас є базою для виконання глибшого аналізу досліджуваного статистичного матеріалу.

3.2. Основні правила утворення груп

Принципове значення при побудові груп має вибір групувальної ознаки, визначення кількості груп і величини інтервалу. Вибір групувальної ознаки, тобто ознаки, на основі якої виділяють різні типи, групи, підгрупи, є одним з найважливіших моментів побудови групувань.

Вибір групувальної ознаки має бути оснований на аналізі якісної природи досліджуваного явища. Всебічний теоретично - економічний аналіз суті явища має бути спрямований на те, щоб у відповідності з метою і завданнями дослідження покласти в основу групування суттєві ознаки.

Групувальними ознаками можуть бути кількісні, атрибутивні (якісні), результативні та факторні ознаки.

Відібравши групувальну ознаку (ознаки) і побудувавши ранжирований ряд за цією ознакою, встановлюють кількість груп, на які буде поділено сукупність, що вивчається, і величину інтервалу.

Кількість груп залежить від загальної чисельності одиниць сукупності, характеру групувальної ознаки і виду групувань. Разом з тим при вирішенні цього питання слід дотримуватися двох важливих умов побудови групувань: 1) виділені групи мають відрізнятися якісною однорідністю; 2) кількість одиниць у кожній групі має бути досить великою. Ця вимога впливає із закону великих чисел.

Визначення числа груп та інтервалів у групуванні передусім залежить від того, якою є групувальна ознака – атрибутивною чи кількісною. Якщо

групування здійснюють за атрибутивною (якісною) ознакою (стать, сорт, порода, професія тощо), то виділяють стільки груп, скільки є градацій ознаки. Аналогічно виділяють групи і при групуванні за дискретною кількісною ознакою, яка змінюється в невеликих межах (кількість членів родини, оцінки студентів, кількість приплоду від однієї матки, кількість бригад у ТОВ тощо).

Якщо ж групують за кількісною ознакою (урожайність, собівартість тощо), що змінюється безперервно і набуває в певних межах будь - яких дрібних значень, то групи виділяють шляхом встановлення для кожної з них інтервалів, зазначених верхньою і нижньою межами величини ознаки для даної групи.

При встановленні числа груп і меж інтервалів важливо встановити за кількісними змінами якісні переходи, щоб виділити типи, не змішати суттєво відмінні одиниці спостереження в одній групі. Це завдання вирішується на основі теоретичного аналізу досліджуваного явища (процесу), порівняння груповальної ознаки з раніше оціненими величинами, для яких якісні переходи відомі.

Для побудови інтервального варіаційного ряду необхідно встановити число груп і величину інтервалу.

Питання щодо числа груп і величини інтервалу слід вирішувати з урахуванням множини обставин, перед усім виходячи з цілей дослідження, особливостей досліджуваного явища та ін. При цьому число груп і величину інтервалу слід встановити такими, які б дозволили більш рівномірно розподілити одиниці сукупності по групах і досягти при цьому їх представництва, якісної однорідності.

При визначенні числа груп потрібно брати до уваги розмах варіації ознаки $R = x_{max} - x_{min}$, тобто різницю між її максимальним і мінімальним значенням. Чим більший цей розмах, тим, як правило, більше утворюється груп. Необхідно також враховувати чисельність досліджуваної сукупності. Доцільно, щоб число груп не було занадто великим і малим і щоб в кожную групу попало достатньо велике число одиниць спостереження.

Число груп наближено можна визначити за формулою американського вченого Стерджесса:

$$n \approx 1 + 3,321 \cdot \lg N, \quad (3.1)$$

де N – чисельність (обсяг) сукупності.

Після встановлення числа груп визначають величину інтервалу. *Інтервалом групування* називають різницю між максимальним і мінімальним значенням ознаки в кожній групі. Однак цю величину можна визначити як різницю між верхніми і нижніми межами значень ознаки в суміжних групах. У практиці статистичних групувань правильне встановлення величини інтервалу має першорядне значення для утворення якісно однорідних груп.

Формула для визначення розміру інтервалу матиме вигляд:

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

де x_{max} та x_{min} – відповідно максимальна та мінімальна варіанта ознаки; n – число груп.

Якщо, наприклад, потрібно провести групування з рівними інтервалами 100 робітників підприємства за розміром заробітної плати, максимальне значення якої 3500 грн, а мінімальне – 2800 грн, то за формулою (3.1):

$$n \approx 1 + 3,321 \cdot \lg N = 1 + 3,321 \cdot \lg 100 = 7,642 \approx 8 \text{ груп};$$

$$\text{тоді за формулою (3.2) маємо: } h = \frac{3500 - 2800}{8} = 87,5 \text{ грн.}$$

Отже, у наведеному прикладі оптимальним розміром інтервалу може бути величина 87,5 грн, а число груп – 8.

Якщо розмір інтервалу визначається дробовим числом, то його заокруглюють в більшу сторону.

Формула Стерджесса не враховує особливостей і характеру варіації та розподілу досліджуваної ознаки. Тому механічне її застосування може призвести до неправильних висновків.

Визначення числа груп за формулою Стерджесса обґрунтоване в тих випадках, коли розподіл одиниць сукупності за даною ознакою наближається

до нормального, застосовуються рівні інтервали в групах і при незначній варіації ознаки.

За способом побудови інтервали можуть *бути рівними і нерівними, відкритими і закритими, спеціалізованими*. Вибір того або іншого виду інтервалу залежить від характеру розподілу одиниць досліджуваної сукупності.

Рівними називають інтервали, у яких різниці між верхньою і нижньою межами однакові. Групування з рівними інтервалами застосовуються тоді, коли варіація ознаки проявляється у порівняно вузьких межах і розподіл носить більш або менш рівномірний характер. Визначення величини інтервалу у випадку групування із застосуванням рівних інтервалів здійснюється за вище наведеною формулою. Припустимо, що площа хмелю в групі господарств коливається від 10 до 110 га. Необхідно побудувати групування за розміром площі, утворивши 5 груп з рівними інтервалами. Величина інтервалу становитиме:

Додаючи до мінімального значення ознаки (в даному випадку 10 га) знайдемо значення інтервалу, одержимо верхню межу першої групи: $10 + 20 = 30$ га. Додаючи далі величину інтервалу до верхньої межі першої групи, одержимо верхню межу другої групи: $30 + 20 = 50$ га і т.д. У результаті одержимо такі групи господарств за розміром площі хмелю (га): 10-30; 30-50; 50-70; 70-90; 90-110.

Нерівними називають інтервали, у яких різниця між верхньою і нижньою межами неоднакові. Нерівні інтервали обчислюються в тих випадках, коли досліджувана ознака змінюється в широких межах. Прикладом нерівних інтервалів може бути групування господарств за чисельністю великої рогатої худоби (голів): 1-99; 100-299; 300-599; 600-999; 1000-3000; понад 3000.

В соціально-економічній статистиці часто використовують групування з *нерівними інтервалами*. Застосування нерівних інтервалів, які прогресивно збільшуються або зменшуються, зумовлено самою природою більшості соціально-економічних явищ, коли в нижчих групах суттєве значення мають навіть малі відмінності в показниках, а у вищих групах такі відмінності є

неістотними. Так, наприклад, для нижчих груп при групуванні підприємств за чисельністю працюючих різниця в 50 чи 100 чоловік має велике значення, а для вищих груп, в яких зосереджені великі підприємства, така різниця несуттєва.

Відкритими називають інтервали, у яких наперед невідомі максимальне і мінімальне значення. Тому при групуванні перший і останній інтервал залишаються відкритими. Наприклад, групування господарств за урожайністю вівса (ц/ га): до 15; 15 - 20; 20 - 25; 25 - 30; понад 30.

Закритими називають інтервали, у яких максимальне і мінімальне значення відомі. Наприклад, групування працівників за стажем роботи (років): 05; 5-10; 10-15; 15-20; 20-25; 25-30; 30-35; 35-40.

У групуваннях, що мають за мету відобразити якісну своєрідність груп, застосовуються *спеціалізовані інтервали*. В цьому випадку в кожній групі є особливий свій зміст і межа інтервалу встановлюється там, де відбувається перехід від одної якості до другої. Кількість груп встановлюється відповідно з теорією питання. Наприклад, при характеристиці відгодівельних господарств за чисельністю поголів'я виділяють мілкі, середні і крупні господарства; за рівнем рентабельності – збиткові, низькорентабельні, середньо-рентабельні, високорентабельні.

3.3. Типологічні, структурні та аналітичні групування

Групування, що приводять до виділення соціально-економічних типів, класів, однокласних груп або сукупностей, називають типологічними. До них відносяться групування підприємств за формами власності, виробничим напрямком (зернові, молочні, відгодівельні та ін.); населення – за класовою належністю або соціальними групами; робітників – на зайнятих переважно фізичною і переважно розумовою працею, поділ підприємств на дрібні, середні та крупні, прибуткові та збиткові тощо.

Методологія типологічних групувань визначається тим, наскільки чітко виступають якісні відмінності в досліджуваних явищах. Наприклад, при групуванні галузей промисловості за економічним призначенням продукції

виділяють галузі, які виробляють продовольчі і непродовольчі товари, засоби виробництва і предмети споживання.

Типологічні групування широко застосовуються в економічних, соціологічних і демографічних дослідженнях. Вони не тільки служать меті виділення типів явищ, але й забезпечують можливість аналізу специфіки і особливостей розвитку окремих типів, зміни їх співвідношень в рамках загального економічного процесу. Проілюструємо це на прикладі (табл. 3.1).

Таблиця 3.1

Кількість діючих підприємств за організаційно-правовими формами господарювання в сільському господарстві України у 2007 році

Підприємства	Кількість підприємств	В % до підсумку
Господарські товариства	7428	12,7
Приватні підприємства	4229	7,2
Виробничі кооперативи	1262	2,2
Фермерські господарства	43475	74,5
Державні підприємства	360	0,6
Підприємства інших форм власності	1633	2,8
Разом	58387	100,0

З даних таблиці видно, що найбільшу питому вагу господарств з різними організаційно-правовими формами займають фермерські господарства і господарські товариства.

Структурні групування характеризують розподіл одиниць досліджуваної сукупності за будь-якою ознакою. За їх допомогою проводиться аналіз структури сукупності і структурних зрушень в розвитку соціально-економічних явищ і процесів. Порівняння структурних групувань у часі дає уявлення про структурні зрушення. До них відносяться групування населення за статтю, віком, сільськогосподарських підприємств – за площею

сільськогосподарських угідь, поголів'ям худоби тощо. Прикладом структурного групування можуть бути дані табл. 3.2.

Відомо, що явища суспільного життя знаходяться в певному взаємозв'язку і взаємозалежності. Тому важливим завданням статистичних групувань є вивчення взаємозв'язків і взаємозалежностей між ознаками. Таке завдання вирішується за допомогою аналітичних групувань.

Таблиця 3.2

Групування населення України за статтю в 2007 році (на 1.01.2008 р.)

Стать населення	Чисельність населення, млн. чол.	В % до підсумку
Чоловіки	21,3	46,1
Жінки	24,9	53,9
Разом:	46,2	100,0

Аналітичними називають групування, за допомогою яких вивчаються взаємозв'язки між окремими ознаками статистичної сукупності. Прикладом таких групувань можуть бути групування, в яких вивчається взаємозв'язок між собівартістю та її факторами, продуктивністю тварин та її факторами тощо. Аналітичні групування можуть бути побудовані за результативною (урожайністю, собівартістю, виходом продукції з одиниці земельної площі тощо) і факторною (якість ґрунту, кількість добрив, опадів тощо) ознакою. Відповідно одержимо результативні і факторні групування.

Факторною називається ознака, під впливом якої змінюється залежна від неї інша ознака.

Результативною називається ознака, яка змінюється під впливом факторної ознаки.

Наведемо приклад аналітичного групування за результативною ознакою (табл. 3.3).

Виконаємо аналіз групової таблиці 3.3. Для цього проведемо деякі зіставлення факторів і умов виробництва по окремих групах. Спочатку порівняємо показники крайніх груп – I і IV (нижчої і вищої).

З даних таблиці видно, що різниця в результативній ознаці (урожайності) між цими групами досягає 16,3 ц /га. Урожайність зернових культур в господарствах IV групи вища, ніж у господарствах I групи в 1,7 раз головним чином за рахунок більш високого рівня виробництва: у господарствах IV групи порівняно з господарствами I групи вносять мінеральних добрив на 1 га зернових більше на 1,34 ц діючої речовини, або в 1,95 раз, працівників на 100 га ріллі на 3,7 чол., або в 1,24 раз більше. Відрізняється ця група і більш високою якістю ґрунту – на 29 балів, або в 1,56 разів.

Таблиця 3.3

Залежність урожайності зернових культур від її факторів у господарствах лісостепової зони області

Групи господарств за урожайністю	Кількість господарств	Середня урожайність, ц/га	Якість ґрунту, балів	Внесено міндобрив на 1 га зернових культур, ц діючої речовини	Кількість середньорічних працівників на 100 га ріллі, чол.
I – до 25,0	16	22,1	52,0	1,41	15,6
II – 25,1-30,0	21	27,7	58,2	1,64	17,4
III – 30,1-35,0	38	32,6	69,4	2,12	17,5
IV – понад 35,0	15	38,4	81,0	2,75	19,3
У середньому	90	31,5	67,7	2,07	17,3

Спостерігаються суттєві відмінності в рівнях середньої урожайності по групах господарств. Зіставлення між собою показників окремих груп дозволяє зробити висновок, що підвищення рівня урожайності від групи до групи (від I до IV) зумовлюється за рахунок кращої забезпеченості добривами, робітниками, кращою якістю ґрунту. Так, наприклад, зіставлення урожайності і факторів та умов виробництва II та I групи показує, що урожайність в II групі господарств порівняно з I групою господарств вища в 1,25 раз. Ця група господарств має вищу забезпеченість працівниками (в 1,12 раз), більше вносять добрив (в 1,16 раз), має більш родючі ґрунти (в 1,12 раз).

Отже, в цілому можна зробити висновок про те, що більш високий рівень урожайності зернових культур забезпечується високим рівнем інтенсифікації виробництва, кращою якістю земель і раціональнішим та ефективнішим використанням самих факторів і умов виробництва.

Аналітичне групування може проводитись як за факторною, так і за результативною ознаками залежно від того, що є основним при статистичному дослідженні. Якщо вивчається вплив якоїсь однієї причини на різні явища, то групування проводять за факторною ознакою, а якщо вивчають вплив різних причин на яке-небудь одне явище, то групують сукупність за результативною ознакою.

Аналітичні групування дозволяють при більш глибокому аналізі знайти форму і виміряти силу зв'язку між варіаційними ознаками та на цій основі зробити важливі практичні висновки для планування і прогнозування.

3.4. Прості і комбінаційні групування

За кількістю ознак, покладених в основу групування, розрізняють *прості і комбінаційні групування*. Групування, проведені за однією ознакою, називають простими або одновимірними (такі групування вже розглянуті раніше), а за двома і більшим числом ознак – комбінаційними або багатовимірними.

При побудові комбінаційного групування сукупність спочатку підрозділяється на групи за однією ознакою, а потім отримані групи поділяються своєю чергою на підгрупи за другою, третьою і т.д. ознаками. В цьому випадку можна прослідкувати вплив на результат сукупності факторів, покладених в основу групування, і оцінити вплив кожного окремо взятого фактора при вирівняних значеннях інших факторів.

Комбінаційне групування дає змогу також оцінити спільний вплив на результат групувальних ознак.

Порядок побудови комбінаційного групування той самий, що і простого, за однією ознакою. Спочатку вирішується питання, про першу групувальну ознаку, потім про другу і т.д.

При комбінаційному групуванні важливо забезпечити достатнє число одиниць в кожній групі і підгрупі. При комбінації ознак число підгруп швидко зростає із збільшенням числа групувальних ознак: воно дорівнює добутку кількості груп, утворених за кожною ознакою із ознак окремо. Наприклад, якщо за першою ознакою виділити три групи, за другою – три і за третьою – дві, то в таблиці одержимо 18 рядків, не рахуючи підсумкових рядків по групах і загального підсумку. Звідси зрозуміло, що комбінаційне групування при невеликій чисельності доцільно будувати за 2-3 ознаками. При виділенні більшого числа групувальних ознак підгрупи можуть виявитися малочисельними, що утруднить виявлення існуючої між ознаками залежності.

Порівняно з простими комбінаційні групування володіють рядом додаткових аналітичних можливостей. Розглянемо це на прикладі комбінаційного групування 90 господарств лісостепової зони за якістю ґрунту і кількістю внесених мінеральних добрив (табл. 3.4). При цьому вивчимо вплив на ефективність виробництва зерна різних доз мінеральних добрив при вирівняній якості ґрунту.

Таблиця 3.4

Вплив якості ґрунту і різних доз добрив на економічну ефективність виробництва зерна

Групи господарств за якістю ґрунту, балів	Підгрупи господарств за внесенням добрив на 1 га зернових культур, ц діючої речовини	Кількість господарств	Урожайність, ц/га	Собівартість 1 ц зерна, грн	Затрати праці на 1 ц зерна, людино-годин
I – до 56	а) до 2,1	11	20,4	64,05	1,85
	б) понад 2,1	8	22,9	59,51	1,65
	У середньому:	19	21,5	60,90	1,71
II – 57-66	а) до 2,1	10	24,6	57,89	1,61
	б) понад 2,1	15	31,1	49,14	1,57
	У середньому:	25	28,0	52,85	1,58
III – 67-76	а) до 2,1	14	33,8	48,86	1,50
	б) понад 2,1	12	35,2	44,03	1,21
	У середньому:	26	34,7	46,62	1,38
IV – понад 76	а) до 2,1	7	36,0	43,96	1,39
	б) понад 2,1	13	43,7	39,48	1,16
	У середньому:	20	39,4	42,07	1,22
	Разом:	90	31,5	50,05	1,47
Підгрупи за добривами	а) до 2,1	42	28,5	53,20	1,57
	б) понад 2,1	48	33,1	47,74	1,40

З даних таблиці 3.4 чітко видно закономірність: з поліпшенням якості ґрунту і збільшенням доз добрив економічна ефективність виробництва зерна підвищується від групи до групи і від підгрупи до підгрупи (урожайність зростає, а собівартість і затрати праці на 1 ц знижуються).

Проведемо аналіз взаємозв'язку групувальних ознак з результативними показниками. Розглянемо взаємозв'язок урожайності з якістю ґрунту і дозами добрив на 1 га зернових культур. Для цього на основі таблиці 3.6 складемо більш зручну для аналізу і читання табл. 3.5.

Таблиця 3.5

Вплив якості ґрунту і доз добрив на урожайність зернових культур

Групи господарств за якістю ґрунту, балів	Підгрупи господарств за внесенням добрив на 1 га зернових культур, ц діючої речовини		У середньому
	а) до 2,1	б) понад 2,1	
I – до 56	20,4	22,9	21,5
II – 57-66	24,6	31,1	28,0
III – 67-76	33,8	35,2	34,7
IV - понад 76	36,0	43,7	39,4
У середньому:	28,5	33,1	31,5

Дані таблиці 3.5 свідчать про те, що на урожайність зернових культур суттєво впливає як якість ґрунту, так і добрива. Вплив якості ґрунту можна оцінити при порівнянні урожайності по групах в межах кожної підгрупи при однакових дозах добрив. Прибавки урожайності за рахунок якості ґрунту при порівнянні I і II груп склали: по I підгрупі 4,2 ц/ га (24,6 - 20,4), по II підгрупі 8,2 ц/ га (31,1 - 22,9).

Ці прибавки при порівнянні II і III груп складають: по I підгрупі 9,2 ц/га (33,8 - 24,6), по II підгрупі 4,1 ц/га (35,2 - 31,1), а між крайніми групами (I і IV) – відповідно по I підгрупі 15,6 ц/ га (36,0 - 20,4), по II підгрупі 20,8 ц/ га (43,7-22,9).

Вплив на урожайність доз добрив прослідковується при порівнянні її рівня по підгрупах у межах тієї чи іншої групи за якістю ґрунту. При переході від першої підгрупи до другої приріст урожайності становить:

- а) в I групі (до 56 балів) 2,5 ц/га (22,9 - 20,4);
- б) в II групі (57 - 66 балів) 6,5 ц/ га (31,1 - 24,6);
- в) в III групі (67 - 76 балів) 1,4 ц/ га (35,2 - 33,8);
- г) в IV групі (понад 76 балів) 7,7 ц/ га (43,7 - 36,0).

Порівняння одержаних приростів урожайності за рахунок кожного з факторів показує, що найвищі прибавки урожайності за рахунок добрив одержані в IV групі господарств, де найвища якість ґрунту.

За даними таблиці 3.7 оцінимо спільний вплив на врожайність двох групувальних ознак. Так, в підгрупі з найвищою якістю ґрунту і найвищими дозами добрив (46) урожайність порівняно з підгрупою з найнижчим рівнем факторів вища на 23,3 ц/га (43,7 - 20,4). Отже, за рахунок спільного впливу двох факторів урожайність підвищилась на 23,3 ц/ га.

Аналогічно можна провести аналіз взаємозв'язку інших результативних показників (собівартості і затрат праці на 1 ц зерна) з групувальними ознаками (якістю ґрунту і дозами добрив).

Таким чином, комбіноване групування за двома ознаками дозволяє вивчити вплив обох групувальних ознак на результативний показник.

3.5. Вторинні групування

Поряд з первинними групуваннями, виконаними на основі первинного статистичного матеріалу, у статистиці знаходять широке застосування вторинні групування.

Вторинним групуванням називають утворення нових груп на основі раніше проведеного групування.

Вторинне групування використовують для вирішення різних завдань, найважливішими з яких є: 1) утворення на основі групувань за кількісними ознаками якісно однорідних груп (типів); 2) приведення двох (або більше)

групувань з різними інтервалами до єдиного виду з метою порівнянності та аналізу; 3) утворення більш укрупнених груп, в яких ясніше проявляється характер розподілу.

Суть цього прийому полягає в одержанні порівнянних даних по різних групуваннях, для чого: чисельний склад групи (за процентом) фіксується на одному рівні у всіх групуваннях; по всіх групуваннях встановлюється також рівне число груп і однаковий зміст групових таблиць. Порівнянню і зіставленню підлягають не абсолютні показники по групах, а відносні величини, процентні відношення.

Розрізняють два способи вторинного групування: 1) шляхом перетворення інтервалів первинного групування (частіше простим укрупненням інтервалів) і 2) шляхом закріплення за кожною групою певної частини одиниць сукупності (часткове перегрупування). При використанні цих способів вторинного групування звичайно припускають, що розподіл ознаки всередині інтервалів буде рівномірним.

Застосування вторинного групування для приведення двох групувань з різними інтервалами до єдиного виду в цілях порівнянності проілюструємо на такому прикладі. Для цього використаємо дані первинного групування двох районів за чисельністю працівників тваринництва (табл. 3.6).

Таблиця 3.6

Групування господарств двох районів за чисельністю працівників тваринництва

Район I		Район II	
Групи господарств за чисельністю працівників, чол.	В % до підсумку	Групи господарств за чисельністю працівників, чол.	В % до підсумку
До 140	4	До 160	8
140-160	12	160-190	16
160-180	18	190-220	30
180-200	27	220-250	21
200-220	24	250-280	15
220-240	6	280-310	6
240-260	4	Понад 310	4
260-280	2	-	-
Понад 280	3	-	-
Разом:	100	Разом:	100

Безпосередньо дані групувань двох районів непорівнянні, так як господарства розподілені по групах з різними інтервалами: 20 чол. у районі I і 30 чол. у районі II. Число виділених груп також неоднакове.

Для приведення двох групувань в порівнянний вид виконаємо вторинне групування. З цією метою перегрупуємо матеріали в групи, єдині для обох районів: візьмемо інтервал 40 чол. (табл. 3.7).

Оскільки є можливість вторинне групування господарств району I здійснити способом простого укрупнення інтервалів (має місце збіг нижніх і верхніх інтервалів у двох групуваннях), використаємо цей спосіб для вирішення поставленого завдання.

Пояснимо послідовність розрахунків. У першу групу господарств з чисельністю працівників до 160 чол. увійдуть господарства I і II груп.

Таблиця 3.7

Вторинне групування господарств двох районів за чисельністю працівників тваринництва

Групи господарств за чисельністю працівників, чол.	В % до підсумку	
	район I	район II
до 160	16 (4+12)	8
160 - 200	45 (18+27)	26 (16+10)
200 - 240	30 (24+6)	34 (20+14)
240 - 280	6 (2+4)	22 (7+15)
понад 280	3	10 (6+4)
Разом:	100	100

Питома вага господарств цих груп у загальному підсумку становитиме 16% (4+12). У другу групу господарств з чисельністю працівників від 160 до 200 чол. увійдуть господарства III і IV груп. їхня питома вага в загальному

підсумку становитиме 45% (18+27). Аналогічно виконуються розрахунки при утворенні решти груп.

Перегрупуємо господарства району II. Оскільки укрупнення інтервалів для господарств району II не підходить і завдання не вирішує використаємо спосіб часткового перегрупування даних первинного групування.

У першу, заново створену групу господарств району II з чисельністю працівників тваринництва до 160 чол., повністю увійдуть господарства первинного групування з таким же інтервалом. Питома вага господарств цієї групи становить 8%.

У другу групу господарств вторинного групування з чисельністю працівників від 160 до 200 чол. повністю увійдуть господарства II групи (16%) і частина господарств III групи. Для визначення частини господарств, яку потрібно взяти з III групи, необхідно її розчленити на підгрупи з чисельністю працівників 190-200, 200-210, 210-220 чол. Показники питомої ваги господарств в цих підгрупах визначаються пропорційно діленню величини інтервалу. Величина інтервалу, яку ми розглядаємо, становить 30 чол. і ділиться на три рівні частини. Для одержання потрібного інтервалу 160 - 200 чол. до величини інтервалу II групи (160 - 190 чол.) слід додати одну третину величини інтервалу III групи (190 - 220 чол.) і таку ж частину господарств цієї групи.

Отже, в другу, заново створену групу господарств, увійдуть 16% господарств другої групи і одна третина III групи – 10% ($1/3 \times 30$), що становитиме 26% загальної чисельності господарств району II.

У III групу господарств вторинного групування (200 - 240 чол.) увійде частина господарств III групи (190 - 220 чол.), що залишилась, – 20% ($2/3 \times 30$) і дві третини господарств IV групи (220 - 250 чол.) - 14% ($2/3 \times 21$), тобто 34% всієї чисельності господарств району II.

Аналогічні розрахунки виконуються і при утворенні решти, заново створених груп господарств: 240 - 280 і понад 280 чол. Як би в табл. 3.4 поряд з даними про питому вагу господарств по групах були наведені дані про їх

чисельність, то розрахунки по заново створених групах виконувались би в тих самих співвідношеннях, що і за питомою вагою господарств.

Після вторинного групування первинний матеріал стає порівнянним, оскільки для двох районів узяті однакові групи за чисельністю працівників. З даних табл. 3.8 видно, що розподіл господарств за чисельністю працівників тваринництва в двох районах суттєво відрізняється: в районі I переважають господарства з чисельністю працівників тваринництва до 200 чол. (61% загальної чисельності господарств), в районі II – господарства з чисельністю працівників тваринництва понад 200 чол. (66% загальної чисельності господарств).

3.6. Багатовимірні групування

Багатовимірними називаються групування, коли статистична сукупність одночасно групується за багатьма ознаками. Набір таких ознак називається *однаковим простором*. Кожна ознака розглядається як координата цього простору, а кожний об'єкт – як точка в n – вимірному просторі, якщо в наборі є n ознак.

Завдання групування зводиться до виділення згущень цих точок простору. Такі згущення називаються *кластерами*, а багатовимірне групування – *кластерним аналізом*.

Алгоритми кластерного аналізу дають змогу поділити сукупність об'єктів на однорідні за певним формальним критерієм подібності групи (кластери). Основною властивістю цих груп є те, що об'єкти, які належать одному кластеру, подібніші між собою, ніж об'єкти з різних кластерів. Таку класифікацію можна виконувати одночасно за досить великою кількістю ознак. Наприклад, відомо чимало статистичних показників, які характеризують рівень соціально-економічного розвитку адміністративних районів країни: кількість населення, кількість безробітних, протяжність шосейних доріг, кількість квадратних метрів житла на одну людину тощо. Для організації опитування необхідно згрупувати райони у більші утворення (регіони), але варто зробити це так, щоб у кожному такому регіоні були райони, близькі за своїм соціально-

економічним розвитком. Це дасть змогу вибрати в такому регіоні один типовий район і результати опитування в ньому узагальнити щодо всього регіону. Таке групування може бути ефективно проведене методом кластерного аналізу, оскільки у даному разі враховується та узагальнюється велика кількість показників.

Групування, в основу яких покладено факторні ознаки, які є причиною зміни результативного показника, називаються факторними, а багатовимірне групування – *факторним аналізом*.

Суть факторного аналізу полягає в тому, що групу сильно скорельованих ознак можна пояснити та описати невеликою кількістю прихованих (латентних) факторів, які безпосередньо не спостерігаються, але розкривають значення ознак цієї групи. Наприклад, за такими ознаками, як «кількість прочитаних книг», «кількість книг у домашній бібліотеці», «кількість відвідувань театрів і музеїв», приховано фактор, який можна було б назвати «рівень культурного розвитку особистості». Факторний аналіз дає змогу виявити ці латентні фактори, описати залежність між ними та первинними ознаками, обчислити значення всіх побудованих таким чином факторів для кожного об'єкта. В результаті виникає можливість без значних втрат інформації перейти від аналізу великої кількості первинних ознак до аналізу порівняно невеликої кількості факторів.

3.7. Основні статистичні класифікації і класифікатори України

Особливим видом групувань у статистиці є класифікація.

Класифікацією у статистиці називається фундаментальне групування одиниць сукупності за атрибутивною ознакою на подібні і відмінні групи і підгрупи. Перелік цих груп і підгруп є своєрідним статистичним стандартом, затвердженим Держкомстатом України. Сюди, наприклад, відноситься класифікація галузей економіки, основних фондів, професій тощо.

Відмінною особливістю класифікацій є, по-перше, те, що в основу їх кладеться якісна ознака. По-друге, класифікації виступають в ролі своєрідного

статистичного стандарту, який встановлюється органами національної і міжнародної статистики на певний проміжок часу. Якщо в кожному конкретному дослідженні будується своє групування, то класифікація єдина для будь-якого дослідження, незалежно від того хто її проводить – органи державної статистики чи інші установи і відомства. По-третє, класифікації стійкі. Вони залишаються незмінними протягом тривалого часу. Проте, якщо з'являються нові групи одиниць, їх класи, розряди, то в класифікацію вносяться відповідні зміни і доповнення.

У класифікації кожному значенню групувальної ознаки присвоюють код, який слугує засобом ідентифікації елемента з відповідним значенням ознаки. Використання кодів дає змогу суттєво прискорити процедуру обробки даних за допомогою сучасних комп'ютерних технологій.

У класифікації точно визначені можливості групи, класи і розряди і є докладні показники, які допомагають віднести будь-яку одиницю об'єкта до тієї або іншої групи, класу, розряду в кожному конкретному випадку.

Класифікації мають фундаментальне значення для всього циклу статистичних робіт, особливо для складання системи національних рахунків.

У сучасній статистичній практиці розрізняють такі класифікації: *економічні, соціальні, екологічні та інші.*

Прикладами діючих класифікацій національного рівня є такі, що повністю узгоджені (гармонізовані) з міжнародними стандартами: Класифікація видів економічної діяльності (КВЕД), Класифікація форм власності (КФВ), Українська класифікація товарів зовнішньоекономічної діяльності (УКТ ЗЕД), Товарна номенклатура зовнішньоекономічної діяльності (ТН ЗЕД), Державний класифікатор продукції та послуг (ДКПП), Класифікація основних фондів (КОФ), Класифікатор об'єктів адміністративно-територіального устрою України (КОАТУУ), Класифікація інституційних секторів економіки (КІСЕ), Класифікація валют (КВ) та ін.

Термін "класифікація" застосовується неоднозначно. Наприклад, в міжнародній статистичній практиці класифікацією прийнято називати

вичерпний і структурований набір описаних категорій, які найчастіше представлені у вигляді ієрархії за допомогою цифрових та літерних кодів.

В національному законодавстві розрізняють поняття "класифікація" та "класифікатор". У цьому контексті класифікатор – це документ, в якому відповідно до прийнятих ознак класифікації та методів кодування об'єкти класифікації розподілені на угруповання і цим угрупованням та об'єктам класифікації надано коди. *Класифікація – це розподіл множини об'єктів на підмножини на підставі їх схожості чи несхожості.*

В той же час, у вітчизняній статистичній науці класифікатори та класифікації (або номенклатури) розрізняють залежно від того, що класифікується: об'єкт чи явище. Класифікатором є систематизований перелік об'єктів (продукції, адміністративно-територіальних одиниць тощо), кожному з яких надано певний код. Відповідно, класифікацією є систематизований перелік явищ, (наприклад, видів економічної діяльності), кожному з яких надано певний код.

Так, у державній статистичній діяльності класифікації застосовуються в системі ведення Єдиного державного реєстру підприємств та організацій України (ЄДРПОУ) та статистичних реєстрів, статистиці національних рахунків, інших галузях статистики.

Варто також зауважити, що система національних статистичних класифікацій не є незмінною структурою і потребує постійного супроводження та удосконалення. Зміни у чинному законодавстві, перегляд міжнародних модельних статистичних класифікацій, виникнення нових потреб у вивченні певних об'єктів та явищ потребують внесення відповідних змін до класифікацій.

Питання для повторення і самоконтролю

1. В чому полягає суть статистичного зведення? Що таке просте і складне, централізоване і децентралізоване зведення?

2. В чому полягає зміст статистичного групування. Що таке просте і комбінаційне групування?

3. Які основні завдання вирішують за допомогою статистичних групувань?
4. Що таке варіаційний та атрибутивний ряди розподілів?
5. Як відбувається вибір групувальної ознаки та якими ці ознаки можуть бути?
6. Які умови повинні виконуватись при визначенні числа груп та інтервалів групування?
7. Як визначається число груп і величина інтервалу?
8. Якими бувають типи інтервалів за способом їх побудови і що характерно для кожного з них?
9. Які групування називаються типологічними?
10. Які групування називаються структурними?
11. Які групування називаються аналітичними?
12. Що таке результативні і факторні ознаки та угруповання?
13. Що таке комбінаційне групування і які завдання воно вирішує?
14. В чому полягає суть вторинного групування?
15. Яке групування називається багатовимірним і в чому полягає суть кластерного аналізу?
16. В чому полягає суть факторного аналізу у багатовимірних групуваннях?
17. В чому полягає суть особливого виду групувань – класифікації?
18. Який державний орган розробляє національні класифікації і що таке класифікатор?
19. Які Ви знаєте приклади діючих класифікаторів України?

Лекція 4. Узагальнюючі статистичні показники

План

- 4.7. Суть і види узагальнюючих статистичних показників.
- 4.8. Абсолютні статистичні величини, їх види та одиниці виміру.
- 4.9. Відносні величини. Види відносних величин, техніка їх обчислення та форми виразу.
- 4.10. Взаємозв'язок між окремими видами відносних величин.

4.11. Суть середніх величин та умови їх використання.

4.12. Види середніх величин, їх обчислення. Багатовимірні середні.

4.1. Суть і види узагальнюючих статистичних показників

Після зведення і групування даних спостереження переходять до останнього – третього етапу статистичної методології. Він полягає в подальшій обробці статистичних таблиць шляхом обчислення *статистичних показників*.

Статистичний показник – це узагальнююча характеристика явища або процесу, яка характеризує всю сукупність одиниць обстеження і використовується для аналізу сукупності в цілому. За допомогою статистичних показників вирішується одна з головних задач статистики: визначається кількісна сторона явища чи процесу у поєднанні з якісною стороною. Кількісна сторона показника представляється числом з відповідною одиницею виміру для характеристики: розміру явищ (кількості робітників, обсягу товарообороту, капіталу фірми тощо); їх рівнів (рівня продуктивності праці робітників тощо); співвідношень (наприклад, між продавцями та іншими категоріями працівників магазину). Якісний зміст показника залежить від суті досліджуваного явища (процесу) і відображається у назві показника (прибутковість, народжуваність тощо).

Показники поділяються на види в залежності від способу їх обчислення, ознак часу, виконання своїх функцій.

За способом обчислення розрізняють первинні і похідні показники. Первинні визначаються шляхом зведення та групування даних і подаються у формі абсолютних величин (наприклад, кількість та сума вкладів громадян у банку). Похідні показники обчислюються на базі первинних і мають форму середніх або відносних величин (наприклад, середня заробітна плата, індекс цін).

Серед статистичних показників окрему групу становлять *взаємообернені показники* – пара характеристик, які існують паралельно і відповідають одному й тому ж явищу (процесу). Це прямий показник X , який змінюється у напрямі

зміни явища (наприклад, продуктивність праці за одну одиницю часу), та обернений $1/X$ – у протилежному напрямі (наприклад, трудомісткість одиниці продукції).

За ознакою часу показники поділяються на *інтервальні та моментні*. Інтервальні характеризують явище за певний період часу (місяць, квартал, рік): наприклад, середньомісячні сукупні витрати на душу населення. Моментні показники характеризують явище за станом на певний момент часу (дату): наприклад, залишок обігових коштів на початок місяця.

За способом виконання своїх функцій розглядають показники, що відбивають обсяг явища, його середній рівень, інтенсивність прояву, структуру, зміну в часі або порівнянні у просторі.

За суттю досліджуваних явищ розрізняють об'ємні показники, що характеризують розміри явищ, процесів, та якісні показники, що характеризують кількісні співвідношення, характерні властивості досліджуваних явищ.

За ступенем агрегування явищ можна розглядати індивідуальні показники, що виражають розміри ознаки окремих одиниць сукупності та загальні (узагальнювальні) показники, що виражають розміри ознаки окремих груп або всієї сукупності.

В статистиці використовують декілька різновидів статистичних показників: *абсолютні та відносні величини; середні величини; показники варіації*.

Функції статистичних показників: *пізнавальна, управлінська, контрольна, стимулювальна*.

Статистичні показники є базою для аналізу та прогнозування соціально-економічного розвитку держави, її окремих галузей і регіонів, стану й розвитку досліджуваних явищ, напряму та інтенсивності процесів, що відбуваються в суспільстві. Вивчаючи явища кількісно, економіст аналізує їх, пізнає якісно, проникає в сутність. У цьому проявляється *пізнавальна функція статистичних показників*.

Управлінська функція статистичних показників полягає в тому, що вони є важливим елементом процесу управління на всіх його рівнях. З розвитком ринкових відносин ця роль статистичних показників зростає. Підвищується значення й *контрольної функції* статистичних показників, насамперед – за виконанням договірних умов.

Статистичні показники як відображення об'єктивної реальності тісно пов'язані між собою, тому їх розглядають не ізольовано один від одного, а в певному взаємозв'язку. Наприклад, для характеристики діяльності промислового підприємства потрібно розглядати кілька показників, які перебуваючи у певному взаємозв'язку, утворюють систему статистичних показників.

4.2. Абсолютні статистичні величини, їх види та одиниці виміру

Абсолютна статистична величина — це характеристика явища, що виражає чисельність сукупності або її частин, а також розміри, обсяги та рівні первинних ознак.

Абсолютні статистичні величини широко використовуються у статистичному обліку та аналізі, управлінні, плануванні, менеджменті, маркетингу тощо. Вони поділяються на індивідуальні, групові та загальні (сумарні). Індивідуальна абсолютна величини відносяться до окремої одиниці сукупності. Групова та загальна відповідно характеризує окрему групу одиниць або всю сукупність разом, їх одержують шляхом сумування індивідуальних величин. Загальна абсолютна величина має назву «обсяг ознаки».

Абсолютні величини — це іменовані числа. Вони виражаються у *натуральних, трудових та вартісних вимірниках*.

Натуральними або фізичними називають такі одиниці виміру, які відтворюють фізичні властивості одиниць сукупності (вагу, площу, довжину, потужність тощо). Складна одиниця виміру утворюється за допомогою двох простих, наприклад, тонно-кілометр, кВт-год. У разі потреби використовуються умовно-натуральні одиниці виміру. Тоді рівень ознаки обчислюється у одиницях певного еталону. Перехід від натуральних до умовно-

натуральних одиниць виміру здійснюється за допомогою перехідних коефіцієнтів. Наведемо приклад розрахунку перехідних коефіцієнтів та обчислення кількості двигунів за потужностями:

Потужність двигуна, кВт	Кількість, шт.	Перехідний коефіцієнт	Кількість в умовно-натуральних одиницях
10	100	1,0	100
15	50	1,5	75
20	200	2,0	400
Разом:	350	-	575

Якщо перший вид двигунів прийнято за еталон, то перехідні коефіцієнти визначаються так:

Кількість двигунів в умовно-натуральних одиницях становить:

$$N_1 = 100 \times 1,0 = 100; N_2 = 50 \times 1,5 = 75; N_3 = 200 \times 2,0 = 400.$$

Отже, загальна кількість двигунів дорівнює 350 шт., або 575 у. о. в перерахунку на двигуни потужністю 10 кВт.

Трудові одиниці виміру (людино-день та людино-година) використовуються для характеристики робочого часу або витрат праці. Вартісні вимірники характеризують розмір явища у грошовому виразі. Вони є найбільш універсальними, але їх головним недоліком є залежність від зміни цін та курсів валют.

4.3. Відносні величини. Види відносних величин, техніка їх обчислення та форми виразу

Відносними величинами називають такі показники, які виражають кількісне співвідношення між ознаками, що характеризують досліджувані явища та процеси.

Відносна величина використовується в тому випадку, коли треба охарактеризувати:

- в скільки разів одна ознака більша або менша за іншу;
- яку частину становить одна ознака відносно іншої;
- скільки одиниць однієї ознаки припадає на 1, 1000, 10000 одиниць іншої ознаки.

Відносну величину завжди одержують шляхом ділення, тому її можна подати у вигляді дроби: $ВВ = \frac{А}{Б}$.

Показник, що знаходиться у чисельнику (А), називається порівнюваною величиною, а той, що стоїть у знаменнику (Б) — базою порівняння. В залежності від розмірів обох величин використовуються різні форми виразу відносних величин, що дозволяє одержати відносну величину в зручному для сприйняття та інтерпретації вигляді.

У тому випадку, коли $А > Б$ або $А < Б$, використовується коефіцієнтна форма виразу: $ВВ = \frac{А}{Б}$.

З формули видно, що результат ділення залишається без змін, при цьому його прийнято визначати з точністю до 0,001. Наприклад, на денній формі навчання є 2500 студентів, а на вечірній — 700, тоді $ВВ = \frac{2500}{700} = 3,6$.

Отже, на денній формі навчання нараховується у 3,6 разів більше студентів, ніж на вечірній.

Якщо $А < Б$ або $А < Б$, використовується процентна форма виразу (%):

$$ВВ = \frac{А}{Б} \cdot 100.$$

Отже, для одержання цієї форми виразу результат ділення збільшується у 100 разів. Наприклад, у вересні реалізовано продукції на 12500 грн, а у серпні – на 17400 грн, тоді $ВВ = \frac{12500}{17400} \cdot 100 = 71,8\%$.

Таким чином, у вересні обсяг реалізації продукції становив 71,8% порівняно із серпнем.

У тому випадку, коли А значно менше Б, застосовують промільну форму виразу (‰ — проміле): $ВВ = \frac{12500}{17400} \cdot 1000$.

В цьому випадку результат ділення збільшується у 1000 разів.

Наприклад, на підприємстві працюють 7830 робітників та 420 службовців тоді $ВВ = \frac{420}{7830} \cdot 1000 = 54 \text{ ‰}$.

Отже, на 1000 робітників припадає 54 службовці.

У деяких випадках використовується продецемільна форма виразу відносних величин ($^0/_{000}$ — продецеми́ле): $ВВ = \frac{А}{Б} \cdot 10000$.

Для одержання цієї форми виразу результат ділення збільшується у 10000 разів.

Наприклад, в області проживає 1610 тис. чол. населення, а працює 34,5 тис. Лікарів тоді $ВВ = \frac{34,5}{1610} \cdot 10000 = 214 ^0/_{000}$.

Отже, на 10000 жителів області припадає 214 лікарів.

При характеристиці рівня захворюваності населення та в інших окремих випадках застосовується просантимільна форма виразу (просантимі́ле – $0/0000$):

$$ВВ = \frac{А}{Б} \cdot 100000.$$

4.4. Взаємозв'язок між окремими видами відносних величин

Між окремими видами відносних величин існує взаємозв'язок. В залежності від суті та методики розрахунку розрізняють сім видів відносних величин: планового завдання, виконання плану, динаміки, структури, координації, порівняння та інтенсивності.

Відносна величина планового завдання показує, у скільки разів або на скільки процентів запланований рівень показника більший чи менший фактично досягнутого рівня:

$$ВВ_{пз} = \frac{П_1}{Ф_0}, \text{ або } ВВ_{пз} = \frac{П_1}{Ф_0} \cdot 100,$$

де $П_1$ та $Ф_0$ — відповідно планове значення показника у наступному та фактичне значення у попередньому періоді.

Наприклад, у II кварталі прибуток становив 13450 грн., а у III кварталі планується одержати 14000 грн. прибутку, тоді:

$$ВВ_{пз} = \frac{14000}{13450} = 1,041, \text{ або } 104,1\%.$$

Таким чином, у III кварталі порівняно з другим планувалося збільшити прибуток на 4,1%.

Відносна величина виконання плану характеризує, у скільки разів або на скільки процентів фактичне значення показника більше або менше запланованого:

$$ВВ_{еп} = \frac{\Phi_1}{\Pi_1}, \text{ або } ВВ_{еп} = \frac{\Phi_1}{\Pi_1} \cdot 100.$$

Наприклад, у III кварталі при плановому прибутку 14000 грн фактичний прибуток становив 12170 грн. Тоді:

$$ВВ_{еп} = \frac{12170}{14000} \cdot 100 = 86,9\%.$$

У тому випадку, коли планується не абсолютне значення показника, а його збільшення або зменшення у процентах, відносна величина виконання плану визначається як відношення фактичної зміни у процентах до планової. Наприклад, по підприємству планувалося у серпні збільшити обсяг реалізації продукції на 4%, а фактично він зріс лише на 0,7%. Тоді,

$$ВВ_{еп} = \frac{100 + 0,7}{100 + 4,0} \cdot 100 = 96,8\%.$$

Таким чином, план по реалізації продукції у серпні недовиконаний на 3,2% ($96,8 - 100 = -3,2$).

Відносна величина динаміки характеризує зміну показника у часі і визначається як відношення значення у наступному періоді до величини у попередньому періоді: $ВВ_{д} = \frac{\Phi_1}{\Phi_0}$ або $ВВ_{д} = \frac{\Phi_1}{\Phi_0} \cdot 100$.

Наприклад, якщо у III кварталі фактичний прибуток становив 12170 грн., а у II кварталі — 13450 грн. відносна величина динаміки становить:

$$ВВ_{д} = \frac{12170}{13450} = 0,905, \text{ або } 90,5\%.$$

Отже, у III кварталі порівняно з II обсяг прибутку зменшився на 9,5% ($90,5 - 100 = -9,5$).

Між названими видами відносних величин, якщо вони підраховані за одними даними, існує взаємозв'язок:

$$ВВ_{д} = ВВ_{пз} \cdot ВВ_{еп} = \frac{\Pi_1}{\Phi_0} \cdot \frac{\Phi_1}{\Pi_1} = \frac{\Phi_1}{\Phi_0}.$$

$$\text{Звідси: } ВВ_{пз} = \frac{ВВ_{д}}{ВВ_{еп}}, \text{ або: } ВВ_{еп} = \frac{ВВ_{д}}{ВВ_{пз}}.$$

Відносна величина структури характеризує співвідношення частини та цілого. Вона показує, яку частину або скільки процентів становить частина від загального підсумку. Якщо ця відносна величина визначається у вигляді коефіцієнту $ВВ_c$, вона називається часткою, а якщо у процентах — питомою вагою, тобто $ВВ_c = \frac{\text{Частина}}{\text{Ціле (сума частин)}}$

Наприклад, в області за рік народилося 8216 хлопчиків та 8203 дівчинки. Тоді $ВВ_c = \frac{8416}{8416+8203} = \frac{8416}{16619} = 0,506$, або 50,6%; $ВВ_c = \frac{8203}{16619} = 0,494$, або 49,4%.

Слід мати на увазі, що сума відносних величин структури становить 1 або 100%.

Відносна величина координації показує співвідношення між окремими частинами одного цілого, при цьому одна частина приймається за базу порівняння. Вона може визначатися на 100, 1000 або 10000 одиниць знаменника.

4.5. Суть середніх величин та умови їх використання

Важливе значення у статистиці має такий узагальнюючий показник як середня величина (середня зарплата, середня врожайність, середній відсоток виконання плану тощо). Вона узагальнює характеристику однорідних елементів масових явищ, які виражаються різними числовими значеннями залежно від конкретних умов.

Середня величина – це узагальнююча міра варіюючої ознаки, що характеризує її рівень у розрахунках на одиницю сукупності. Тільки за допомогою середньої можна охарактеризувати сукупність за кількісною варіаційною ознакою.

Середні величини використовують для порівняння показників двох і більше об'єктів (порівняння урожайності окремих культур по господарствах області, порівняння цін на деякі товари, що реалізуються на ринках даного регіону, тощо).

Середні величини використовують при вивченні взаємозв'язків між явищами та зміни рівнів явищ у часі. Їх використовують для проведення факторного аналізу явищ з метою виявлення невикористаних резервів. Вони використовуються у плануванні і прогнозуванні для економіки в цілому та її окремих галузей.

Умовами використання середніх величин є наступні:

- 1) індивідуальні величини повинні відноситися до якісно однорідної сукупності та кількість їх має бути достатньо великою;
- 2) розподіл сукупності на якісно однорідні групи, тобто поєднання методу середніх величин із методом групування;
- 3) застосування загальних та групових середніх при умові обчислення їх із якісно однорідної сукупності.

Багатогранність суспільних явищ обумовлює виняткову важливість використання середніх величин в економіко-статистичних дослідженнях. Вони є активним засобом управління, планування і прогнозування економіки держави.

4.6. Види середніх величин, їх обчислення. Багатовимірні середні.

Залежно від характеру усереднюваної ознаки і наявної вихідної інформації в статистиці застосовуються різні види середніх величин, серед яких найбільше використовуються такі: *середня арифметична, середня гармонічна, середня геометрична і середня квадратична.*

Кожну середню можна визначити як просту, коли значення варіант спостерігаються тільки один раз або однаково кількість разів, і як зважену, коли значення варіант повторюється різну кількість разів.

Вибір того чи іншого виду середньої визначається цілями і завданнями дослідження і наявною інформацією.

Загальною умовою правильного обчислення усіх видів середніх є збереження незмінним загального обсягу варіюючої ознаки при заміні індивідуальних значень ознак їхньою середньою. Так, середня арифметична застосовується тоді, коли обсяг варіюючої ознаки утворюється як сума окремих

варіант; середня гармонічна – коли обсяг варіюючої ознаки утворюється як сума обернених значень окремих варіант; середня геометрична – коли обсяг варіюючої ознаки утворюється як добуток окремих варіант; середня квадратична – коли обсяг варіюючої ознаки утворюється як сума квадратів окремих варіант.

Розглянемо перелічені вище види середніх докладніше.

Середня арифметична – найпоширеніший вид середньої. Середня арифметична проста являє собою частку від ділення суми індивідуальних значень ознаки на їх загальне число, її обчислюють за формулою:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum x_i}{n}, \quad (4.1)$$

де x_i – окремі варіанти ознаки X ($i = 1, 2, 3, \dots, n$);

n – кількість варіант.

Якщо в сукупності варіанти зустрічаються неоднакову кількість разів, то їх об'єднують в групи і переходять до обчислення середньої арифметичної зваженої за формулою:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_n f_n}{n} = \frac{\sum x_i f_i}{\sum f_i}, \quad (4.2)$$

де f_i – частоти (ваги), які показують скільки разів зустрічаються значення ознаки в сукупності.

Приклад 1. Нехай задано дані про заробітну плату робітників і число робітників, яку отримують дану зарплату (табл. 4.1).

Таблиця 4.1

Дані для розрахунку середньої заробітної плати робітників

Табельний номер робітника	Заробітна плата одного робітника, грн (ознака X)	Число робітників, чол. (частота f)
1	2300	10
2	2500	17
3	2800	40
4	3000	18
5	3400	15
Разом:	-	100

Використовуючи формулу (4.2), розрахуємо середню заробітну плату одного робітника:

$$\bar{x} = \frac{2300 \cdot 10 + 2500 \cdot 17 + 2800 \cdot 40 + 3000 \cdot 18 + 3400 \cdot 15}{100} = 2825 \text{ (грн)}.$$

Бувають випадки коли значення величин задаються не конкретними числами, а за допомогою інтервалів (інтервальним рядом розподілу). В таких випадках знаходять середину кожного інтервалу і після цього обчислюють середнє значення за формулою (4.2).

Приклад 2. Нехай задано дані інтервального ряду розподілу (табл. 4.2).

Таблиця 4.2

Дані для розрахунку середньої ваги деталей

Вага виробів, г (ознака X)	Число виробів, шт. (частота f)	Середина інтервалу (x_i)	Вага всіх деталей, г ($x \cdot f$)
до 100	5	97,5	487,5
100-105	19	102,5	1947,5
105-110	52	107,54	5590,0
110-115	18	112,5	2025,0
більше 115	6	117,5	705,0
Разом:	100	-	10755,0

В нашому випадку дані табл. 4.2 свідчать про те, що є два інтервали з нечітко вираженими межами – перший та останній є інтервали з нечітко вираженими межами (перший інтервал "до" і останній – "понад"), тому для визначення серединного значення потрібно встановити умовні межі цих інтервалів. Звичайно в цих випадках вирішують так: для першого інтервалу беруть величину другого інтервалу, а для останнього – величину передостаннього інтервалу.

Використовуючи дані таблиці 4.2 обчислюємо середню вагу деталі за формулою (4.2):

$$\bar{x} = \frac{10755,0}{100} = 107,55 \text{ г.}$$

Зауважимо, що на практиці при обчисленні середньої арифметичної можна використовувати такі її властивості:

- 1) Якщо всі варіанти збільшити (зменшити) на одне й те ж число, то й середня арифметична збільшиться (зменшиться) на те саме число;
- 2) Якщо всі варіанта збільшити (зменшити) в k разів, то й середня арифметична збільшиться (зменшиться) також в k разів;
- 3) Якщо всі частоти (ваги) поділити (помножити) на яке-небудь число, що відмінне від нуля, то середня арифметична від цього не зміниться.

На практиці у статистиці бувають випадки, коли середню величину потрібно обчислювати за формулою *середньої гармонічної*. Це необхідно робити тоді, коли підсумовуванню підлягають не самі варіанти, а обернені до них числа. *Середня гармонічна проста* обчислюється за формулою:

$$\bar{x} = \frac{n}{\sum \frac{1}{x_i}}, \quad (4.3)$$

де x_i – окремі варіанти ознаки X ($i = 1, 2, 3, \dots, n$);

n – кількість варіант.

Приклад 3. Три тракторні агрегати при обробітку ґрунту умовної оранки одного гектара затратили часу: перший – 2,2 години, другий – 2,4 години, третій – 2,6 години. Використовуючи формулу (4.3), визначимо середні затрати часу на обробіток одного гектара умовної оранки тракторним агрегатом:

$$\bar{x} = \frac{n}{\sum \frac{1}{x_i}} = \frac{3}{\frac{1}{2,2} + \frac{1}{2,4} + \frac{1}{2,6}} = \frac{3}{1,2558} = 2,38 \text{ год.}$$

Середню гармонічну зважену застосовують в тих випадках, коли є дані про індивідуальні значення ознаки в загальній сукупності і загальний обсяг сукупності, але в готовому вигляді немає частот. Таку середню величину обчислюють за формулою:

$$\bar{x} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}}, \quad (4.4)$$

де $\sum \frac{w_i}{x_i}$ – сума добутків обернених значень ознаки на їхні частоти, тобто $w_i = x_i \cdot f_i$, звідси $f_i = \frac{w_i}{x_i}$.

Приклад 4. Дані про заробітну плату робітників заводу в розрізі цехів і фонд заробітної плати представлені в табл. 4.3.

Таблиця 4.3

Дані для розрахунку середньої зарплати робітників

Номер цеху	Середня зарплата одного робітника, грн (x)	Фонд заробітної плати, грн. (w)
1	2900	174000
2	3200	160000
3	3500	140000

Розрахуємо середню заробітну плату одного робітника за формулою (4.4), використовуючи дані табл. 4.3:

$$\bar{x} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} = \frac{174000 + 160000 + 140000}{\frac{174000}{2900} + \frac{160000}{3200} + \frac{140000}{3500}} = \frac{474000}{150} = 3160 \text{ грн.}$$

Середню геометричну застосовують, коли загальний обсяг явища є не сума, а добуток значень ознаки. Ця середня використовується здебільшого для розрахунку середніх коефіцієнтів (темтів) зростання і приросту при вивченні динаміки явищ та обчислюється за формулою:

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots x_n} . \quad (4.5)$$

Приклад 5. За даними про посівну площу цукрових буряків у господарстві за 5 років знайти середній коефіцієнт зростання площі посіву цукрових буряків за 2010 - 2014 рр. (табл. 4.4).

Таблиця 4.4

Дані для розрахунку середнього коефіцієнта зростання площі посіву цукрових буряків

Рік	Площа посіву цукрових буряків, га	Коефіцієнт зростання
2010	250	-
2011	275	1,1000
2012	290	1,0545
2013	310	1,0690
2014	320	1,0323

За формулою (4.5) маємо:

$$\bar{x} = \sqrt[4]{x_1 \cdot x_2 \cdot x_3 \cdot x_4} = \sqrt[4]{1,1000 \cdot 1,0545 \cdot 1,0690 \cdot 1,0323} = 1,0637$$

або $\bar{x} = 106,37\%$.

Отже, середній коефіцієнт зростання посівної площі цукрових буряків у господарстві за 2010 - 2014 рр. становив 1,0637. Інакше кажучи, посівна площа цукрових буряків у господарстві щорічно збільшувалась в середньому на 6,37%.

Середня квадратична використовується переважно для розрахунку показників варіації (коливання) ознаки – дисперсії і середнього квадратичного відхилення, які обчислюються на основі квадратів відхилень індивідуальних значень ознаки від їхньої середньої арифметичної. Крім того, вона застосовується для узагальнення ознак, виражених лінійними мірами яких-небудь площ (при обчисленні середніх діаметрів стовбурів дерев, кошиків, листків, клубнів тощо).

Проста середня квадратична обчислюється за формулою:

$$\bar{x} = \sqrt{\frac{\sum x_i^2}{n}}. \quad (4.6)$$

Зважена середня квадратична обчислюється за формулою:

$$\bar{x} = \sqrt{\frac{\sum x_i^2 \cdot f_i}{\sum f_i}}. \quad (4.7)$$

Наприклад, є дані про розмір діаметрів стовбурів трьох яблунь: 17; 22; 19 см. Потрібно обчислити середній розмір діаметра стовбура яблуні. Оскільки вихідні дані представлені у вигляді квадратних функцій, середній розмір діаметра стовбура яблуні визначимо за формулою (4.6) середньої квадратичної простої:

$$\bar{x} = \sqrt{\frac{17^2 + 22^2 + 19^2}{3}} = 19,44 \text{ см.}$$

Якби у наведеному прикладі окремі значення діаметра стовбура повторювались неоднакове число разів, то середній розмір діаметра стовбура слід було б розраховувати за формулою середньої квадратичної зваженої.

Особливим видом середніх величин є середня багатовимірна, яка являє собою середню величину кількох ознак для однієї одиниці сукупності. Оскільки неможливо розрахувати середню величину за абсолютними значеннями різних ознак (різномірних, виражених у різних одиницях виміру), то багатовимірна середня визначається з відносних величин (часток, процентів тощо), як правило, з відношень абсолютних значень для одиниці сукупності до середніх значень цих ознак.

Середня багатовимірна – похідна величина, розраховується для статистичної сукупності чисельністю N одиниць з порядковими номерами i ($i = 1, 2, 3, \dots, k$), які володіють k ознаками x_i з порядковими номерами i ($i = 1, 2, 3, \dots, k$).

Таким чином, спочатку обчислюють відношення p_i значень кожної ознаки x_i кожної величини сукупності до її середнього значення, а після цього визначають середню з цих відношень для кожної одиниці сукупності p_i , яку і називають багатовимірною середньою, тобто

$$\bar{x} = \frac{\sum p_i}{k} \quad (4.8)$$

Багатовимірні середні дають узагальнену характеристику кожної одиниці сукупності за кількома ознаками одночасно. При цьому значущість ознаки для багатовимірної оцінки одиниці сукупності вважається однаковою, що економічно неточно.

Середню багатовимірну використовують для аналізу господарської діяльності підприємств, зокрема при визначенні ефективності використання виробничого потенціалу (землі, трудових ресурсів, виробничих фондів) тощо.

Питання для повторення і самоконтролю

1. Що таке статистичний показник і для чого він використовується?
2. Які статистичні показники розрізняють за способом їх обчислення?

3. Які показники розрізняють за ознакою часу?
4. Які показники розрізняють за способом виконання своїх функцій, суттю досліджуваних явищ та ступенем їх агрегування?
5. В чому полягає пізнавальна, управлінська та контрольна функції статистичних показників?
6. Що таке абсолютні статистичні величини і якими вимірниками вони виражаються?
7. Що таке відносні статистичні величини і для чого вони використовуються?
8. Що таке відсоткова, промільна, продецимільна і просантимільна форми виразів відносних статистичних величин?
9. Які існують форми взаємозв'язку між відносними величинами і що вони означають?
10. Що таке середня статистична величина і для чого її використовують?
11. Які існують умови використання середніх статистичних величин?
12. Які Ви знаєте основні види середніх величин?
13. Що таке середня арифметична величина і за якими формулами обчислюються проста і зважена середні?
14. Як на практиці обчислюється середня арифметична у випадку коли значення величини задаються інтервальним рядом розподілу?
15. Що таке середня гармонічна (проста і зважена) і за якими формулами вони обчислюються?
16. Коли використовується середня геометрична і за якою формулою вона обчислюється?
17. Коли використовується середня квадратична і за якими формулами обчислюються проста і зважена середні квадратичні?
18. Що таке середня багатовимірна, як її обчислюють і для чого використовують?

Лекція 5. Аналіз рядів розподілу

План

- 5.1. Ряди розподілу, їх види і правила побудови. Закономірність розподілу.
- 5.2. Характеристики центра розподілу.
- 5.3. Характеристики варіації.
- 5.4. Характеристика форми розподілу.
- 5.5. Види та взаємозв'язок дисперсій.

5.1. Ряди розподілу, їх види і правила побудови. Закономірність розподілу

Маючи в розпорядженні дані статистичного спостереження, що характеризують те чи інше явище, перш за все необхідно їх впорядкувати, тобто надати характер системності. Така систематизація проводиться у вигляді статистичного ряду розподілу.

Статистичний ряд розподілу – це впорядковані статистичні сукупності (табл. 5.1, 5.2).

Таблиця 5.1

Загальний вигляд дискретного ряду розподілу

Варіанти, (x_i)	x_1	x_2	x_3	...	x_k	
Частоти, (f_i)	f_1	f_2	f_3	...	f_k	$\sum f_i$

Таблиця 5.2

Загальний вигляд інтервального ряду розподілу

Інтервали	$x_1 - x_2$	$x_2 - x_3$	$x_3 - x_4$...	$x_{k-1} - x_k$	
Частоти, (f_i)	f_1	f_2	f_3	...	f_k	$\sum f_i$

Найпростішим видом статистичного ряду розподілу є ранжирований ряд, тобто ряд чисел, що знаходяться в порядку зростання чи спадання варіюючої ознаки. Такий ряд не дозволяє судити про закономірності, закладені в

розподілених даних: біля якої величини групується більшість показників; які є відхилення від цієї величини; яка загальна картина розподілу. З цією метою групують дані, показуючи, як часто зустрічаються окремі спостереження в загальному їх числі (рис. 5.1).

Розподіл одиниць сукупності за ознаками, що не мають кількісного виразу, називається *атрибутивним рядом* (наприклад, розподіл підприємств за їх виробничим напрямом).

Ряди розподілу одиниць сукупності за ознаками, що мають кількісний вираз, називаються *варіаційними рядами*. У таких рядах значення ознаки (варіанти) знаходяться в порядку зростання чи спадання

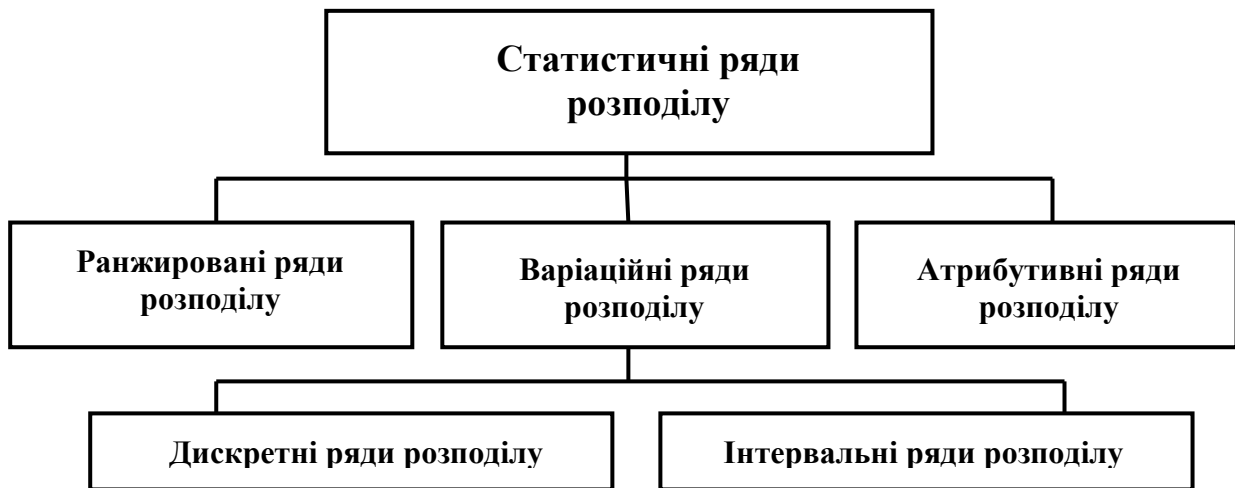


Рис. 5.1. Схема статистичних рядів розподілу

У *варіаційному ряді розподілу* розрізняють два елементи: варіанта і частота. *Варіанта* – це окреме значення груповальної ознаки, *частота* – число, яке показує, скільки разів зустрічається кожна варіанта. Саме у співвідношенні варіант і частот виявляється *закономірність розподілу*. Характерною властивістю правильно складеного варіаційного ряду розподілу є така: сума усіх частот ряду дорівнює загальному обсягу сукупності, а сума усіх частостей ряду дорівнює одиниці.

У математичній статистиці обчислюється ще один елемент варіаційного ряду – *частість*. Остання визначається, як відношення частоти випадків даного інтервалу до загальної суми частот. Частість визначається в частках одиниці, відсотках (%) в промілі (‰).

Таким чином, варіаційний ряд розподілу – це такий ряд, у якому варіанти розташовані в порядку зростання або спадання, вказані їх частоти або частоти. Варіаційні ряди бувають дискретні та інтервальні (неперервні).

Дискретні варіаційні ряди – це такі ряди розподілу, в яких варіанта як величина кількісної ознаки може приймати тільки певне значення. Варіанти різняться між собою на одну чи кілька одиниць (табл. 5.1).

Так, кількість вироблених деталей за зміну конкретним робітником може виражатися тільки одним певним числом (6, 10, 12 і т.д.). Прикладом дискретного варіаційного ряду може бути розподіл працівників за кількістю вироблених деталей (табл. 5.3).

Таблиця 5.3

Приклад дискретного ряду розподілу

Вироблено за зміну деталей, шт. (x_i)	6	7	8	9	10	11	12
Кількість робітників, чол. (n_i)	16	10	8	10	12	16	3

Інтервальні (неперервні) варіаційні ряди – такі ряди розподілу (табл. 5.2), в яких значення варіанти дано у вигляді інтервалів, тобто значення ознак можуть відрізнятися одне від одного на скільки завгодно малу величину. При побудові варіаційного ряду неперервної ознаки неможливо вказати кожне значення варіанти, тому сукупність розподіляється за інтервалами. Останні можуть бути рівні і нерівні. Для кожного з них вказуються частоти або частоти (табл. 5.4).

В інтервальних рядах розподілу з нерівними інтервалами обчислюють такі математичні характеристики, як щільність розподілу і відносна щільність розподілу на даному інтервалі. Перша характеристика визначається відношенням частоти до величини того ж інтервалу, друга – відношенням частоти до величини того ж інтервалу. Для наведеного вище прикладу щільність розподілу на першому інтервалі становитиме $3:5 = 0,6$, а відносна щільність на цьому інтервалі – $7,5:5 = 1,5\%$.

Приклад інтервального ряду розподілу

Чисельність працюючих, чол. (x_i)	Кількість цехів (f_i)	% до підсумку
20-25	3	7,5
25-30	9	22,5
30-35	16	40,0
35-40	8	20,0
40-45	4	10,0
Всього:	40	100,0

Додатковою характеристикою варіаційних рядів є кумулятивна частота (частка), що являє собою результат послідовного об'єднання груп і підсумовування відповідних їм частот (часток). Кумулятивна частота (частка) характеризує обсяг сукупності зі значеннями варіант, які не перевищують x_i (табл. 5.5).

5.2. Характеристики центра розподілу

Центром тяжіння будь-якої статистичної сукупності є типовий рівень ознаки, узагальнююча характеристика всього розмаїття її індивідуальних значень. Такою характеристикою є *середня величина* \bar{x} .

За даними ряду розподілу середня обчислюється як арифметична зважена:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_n f_n}{n} = \frac{\sum x_i f_i}{\sum f_i}, \quad (5.1)$$

де f_i – частоти (ваги), які показують скільки разів зустрічаються значення ознаки в сукупності.

В інтервальних рядах як варіанту використовують середину інтервалу.

Окрім типового рівня важливе значення має домінанта, тобто найбільш поширене значення ознаки. Таке значення називають *модой* (M_o).

Мода (M_o) – це та варіанта, яка частіше за все повторюється у ряді розподілу.

В дискретному ряді – це варіанта, якій відповідає найбільша частота.

Характеристикою центра розподілу вважається також *медіана*.

Медіана (Me) - це варіанта, яка ділить ранжируваний ряд на дві рівні за чисельністю частини.

Якщо непарне число варіант записати в порядку збільшення або зменшення, то центральна з них буде медіаною. При парній кількості варіант, медіана розраховується як середня арифметична двох центральних варіант.

Нехай, наприклад, задано дискретний ряд розподілу:

Ріст студентів, см	165	167	170	173	176	178	180	182	185	187	189	191	195
Число студентів, чол.	8	13	24	30	38	47	55	31	16	9	7	3	2

Очевидно, що в цьому прикладі модою є студент, який має ріст 180 см, оскільки цьому значенню варіанти відповідає найбільше число студентів (55 чол.). Посередині цього варіаційного ряду знаходиться число 180, тому воно визначає також і медіану.

В інтервальному ряді легко знаходиться лише модальний інтервал, а сама мода обчислюється за формулою:

$$M_0 = x_{mo} + h \cdot \frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)}, \quad (5.2)$$

де x_{mo} – нижня межа модального інтервалу;

h – розмір модального інтервалу;

f_1 – перед модальна частота;

f_2 – модальна частота;

f_3 – післямодальна частота.

В інтервальному ряді розподілу аналогічно визначається медіанний інтервал. Конкретне значення медіани обчислюється за формулою:

$$M_e = x_{me} + h \cdot \frac{\frac{\sum f_i}{2} - S_{me-1}}{f_{me}}, \quad (5.3)$$

де x_{me} – нижня межа медіанного інтервалу;

h – розмір медіанного інтервалу;

$\sum f_i$ – сума частот;

$\frac{\sum f_i}{2}$ – порядковий номер медіани;

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

f_{me} – частота медіанного інтервалу.

Мода і медіана – це особливий вид середніх величин. На відміну від абстрактної середньої арифметичної, ці характеристики завжди співпадають з конкретними варіантами.

Покажемо обчислення моди і медіани на конкретному прикладі розподілу робітників за заробітною платою (табл. 5.5).

Таблиця 5.5

Розподіл робітників за заробітною платою

Заробітна плата, грн	Число робітників, чол.	Кумулятивні частоти
1800-1900	5	5
2000-2100	10	15
2100-2200	61	76
2200-2300	105	181
2300-2400	130	311
2400-2500	109	420
2500-2600	62	482
2600-2700	11	493
2700-2800	7	500
Разом:	500	-

В нашому прикладі мода знаходиться в інтервалі від 2300 до 2400 грн, тому що йому відповідає найбільша частота (130 чол.). Цей інтервал називається *модальним*. Використовуючи формулу (5.2), одержимо:

$$M_0 = 2300 + 100 \cdot \frac{130-105}{(130-105)+(130-109)} = 2354,35 \text{ грн.}$$

Цей показник означає, що найбільше робітників мали заробітну плату 2354,35 грн.

Щоб визначити медіану інтервального варіаційного ряду, спочатку, за допомогою кумулятивних частот, потрібно знайти інтервал, що містить медіану. Медіанному інтервалу відповідає перша з кумулятивних частот, яка перевищує півсуму частот всього обсягу сукупності, тобто число $\frac{\sum f_i}{2} = 250$.

Отже, медіана знаходиться в інтервалі від 2300 до 2400 грн. За формулою (5.3) маємо:

$$M_e = 2300 + 100 \cdot \frac{250 - 181}{130} = 2353,08 \text{ грн.}$$

Це означає, що половина робітників отримує заробітну плату, меншу за 2353,08 грн, а друга половина – більшу.

Окрім моди і медіани, в аналізі закономірностей розподілу використовуються також квантілі та децилі. *Квантілі* — це варіанти, які поділяють обсяги сукупності на чотири рівні частини, *децилі* — на десять рівних частин. Ці характеристики визначаються на основі кумулятивних частот (часток) за аналогією з медіаною, яка є другим квантилем або п'ятим децилем.

5.3. Характеристики варіації.

В одних сукупностях індивідуальні значення ознаки щільно групуються навколо центра розподілу, в інших — значно відхиляються. Чим менші відхилення, тим однорідніша сукупність, а отже, тим більш надійні й типові характеристики центра розподілу, передусім середня величина. Вимірювання ступеня коливання ознаки, її варіації — невід'ємна складова аналізу закономірностей розподілу.

Для вимірювання та оцінювання варіації використовуються абсолютні та відносні характеристики. До абсолютних належать: розмах варіації, середнє лінійне та середнє квадратичне відхилення, дисперсії; відносні характеристики подаються низкою коефіцієнтів.

Розмах варіації – це різниця між найбільшим та якнайменшим значеннями ознаки, тобто $R = x_{max} - x_{min}$.

В інтервальному ряді розподілу розмах варіації визначають як різницю між верхньою межею останнього інтервалу й нижньою межею першого або як різницю між середніми значеннями цих інтервалів.

Перевагою варіаційного розмаху є простота його обчислення й тлумачення. Проте, коли частоти крайніх варіант надто малі, варіаційний розмах неадекватно характеризує варіацію. У таких випадках використовують квартильні або децильні розмахи. Квартильний розмах охоплює 50% обсягу сукупності та обчислюється за формулою: $R_Q = Q_c - Q_1$.

Інші абсолютні характеристики варіації базуються на відхиленнях індивідуальних значень ознаки від середньої величини.

Оскільки $\sum(x_i - \bar{x}) \cdot f_i = 0$, то при розрахунку такого роду показників використовують або модулі, або квадрати відхилень. В результаті маємо наступні характеристики варіації: середнє лінійне \bar{l} , середнє квадратичне s відхилення та дисперсію σ^2 (див. таблицю 5.6).

Таблиця 5.6

Розрахунок узагальнюючих характеристик варіації

Показник	Середнє відхилення	
	лінійне	квадратичне
За незгрупованими даними	$\bar{l} = \frac{\sum x - \bar{x} }{n}$	$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$
За згрупованими даними	$\bar{l} = \frac{\sum x_i - \bar{x} \cdot f_i}{\sum f_i}$	$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2 \cdot f_i}{\sum f_i}}$

Очевидно, що дисперсія – це середній квадрат відхилень. Середнє лінійне \bar{l} та середнє квадратичне відхилення σ є безпосередніми мірами варіації. Це іменовані числа (в одиницях вимірювання ознаки), за змістом вони ідентичні, проте завдяки математичним властивостям $\sigma > \bar{l}$. Коли обсяг сукупності досить великий і розподіл ознаки, що варіює, наближається до нормального, то $\sigma = 1,25 \bar{l}$, а $R = 6\sigma$. Значення ознаки в межах $\bar{x} \pm \sigma$ мають 68,3% обсягу

сукупності, у межах $\bar{x} \pm 2\sigma$ — 95,4%, у межах $\bar{x} \pm 3\sigma$ — 99,7%. Це відоме «правило трьох сигм».

Розглянуті абсолютні характеристики варіації мають одиниці вимірювання ознаки. При порівнянні варіації різних ознак або варіації однієї ознаки в різних сукупностях використовують відносні характеристики — коефіцієнти варіації.

Коефіцієнти варіації розраховують за формулами:

$$\text{- лінійний: } V_{\bar{l}} = \frac{\bar{l}}{\bar{x}}; \quad (5.4)$$

$$\text{- квадратичний: } V_{\sigma} = \frac{\sigma}{\bar{x}}; \quad (5.5)$$

$$\text{- осциляції: } V_R = \frac{R}{\bar{x}}; \quad (5.6)$$

$$\text{- кuartильний: } V_Q = \frac{R_Q}{2M_e}. \quad (5.7)$$

Для оцінки однорідності сукупності та порівняння варіацій найчастіше використовують квадратичний коефіцієнт варіації. В економічних розрахунках вважають, що сукупність є однорідною, а середня — типовою, якщо коефіцієнт варіації не перевищує 10-15%.

5.4. Характеристики форми розподілу

Аналіз закономірностей розподілу передбачає оцінювання ступеня *однорідності сукупності, асиметрії та ексцесу розподілу.*

Однорідність сукупності — передумова використання інших статистичних методів (середніх величин, регресійного аналізу тощо).

В однорідних сукупностях розподіли одновершинні (одномодальні). Багатовершинність свідчить про неоднорідний склад сукупності, про різнотиповість окремих складових. Критерієм однорідності сукупності вважається квадратичний коефіцієнт варіації,

Серед одновершинних розподілів є симетричні та асиметричні (скошені), гостро- та плосковершинні. У симетричному розподілі рівновіддалені від центра значення ознаки мають однакові частоти, в асиметричному — вершина розподілу зміщена. Напрямо асиметрії протилежний напряму зсуву вершини. Якщо вершина зміщена вліво, то це правостороння асиметрія та навпаки.

В симетричному розподілі характеристики центру мають однакові значення $\bar{x} = M_o = M_e$; при правосторонній асиметрії – $\bar{x} > M_o > M_e$, при лівосторонній – $\bar{x} < M_o < M_e$.

Чим більша асиметрія, тим більше відхилення $\bar{x} - M_o$. Очевидно, найпростішою мірою асиметрії є відносне відхилення $A = \frac{\bar{x} - M_o}{\sigma}$, яке характеризує напрям і міру скошеності в середині розподілу; при правосторонній асиметрії $A > 0$, при лівосторонній — $A < 0$.

Іншою властивістю одновершинних розподілів є ступінь зосередженості елементів сукупності навколо центра розподілу. Цю властивість називають *ексцесом розподілу*.

Асиметрія та ексцес — дві пов'язані з варіацією властивості форми розподілу. Комплексне їх оцінювання виконується на базі моментів розподілу.

Момент розподілу – це середня арифметична k -го ступеню відхилення $(x-a)$. В загальному вигляді момент розподілу розраховується за формулою:

$$m_k = \frac{\sum (x_i - a)^k}{n} \quad (5.8)$$

або:

$$m_k = \frac{\sum (x_i - a)^k \cdot f_i}{\sum f_i} \quad (5.9)$$

де m_k – момент k -го порядку;

x_i – варіанти ряду;

f_i – частоти ряду;

n – обсяг сукупності;

k та a – задані числа.

Залежно від величини a моменти розділяють на початкові при $a=0$, центральні при $a = \bar{x}$ та умовні при $a = x_0$, де x_0 – деяка варіанта ряду, звичайно близька до його середини. Ступінь k визначає порядок моменту.

Початковий момент k -го порядку виражається формулою:

$$M_k = \frac{\sum x_i^k \cdot f_i}{\sum f_i} \quad (5.10)$$

Центральний момент k -го порядку виражається формулою:

$$m_k = \frac{\sum (x_i - \bar{x})^k \cdot f_i}{\sum f_i}. \quad (5.11)$$

Умовний момент k -го порядку виражається формулою:

$$m_k^* = \frac{\sum (x_i - x_0)^k \cdot f_i}{\sum f_i}. \quad (5.12)$$

Очевидно, що початковий момент 1-го порядку є середня арифметична, 2-го – середній квадрат значень ознаки. Центральний момент 2-го порядку характеризує дисперсію.

Центральні моменти 3-го і 4-го порядків характеризують відповідно асиметрію та ексцес. У симетричному розподілі $m_3 = 0$. Чим більша скошеність ряду, тим більше значення величини. Для того щоб характеристика скошеності не залежала від масштабу вимірювання ознаки, для порівняння ступеня асиметрії різних розподілів використовується стандартизований момент $A_s = m_3/s_3$ – коефіцієнт асиметрії, який, на відміну від коефіцієнта скошеності, залежить від крайніх значень ознаки. При правосторонній асиметрії коефіцієнт $A_s > 0$, при лівобічній $A_s < 0$. Тому правосторонню асиметрію називають позитивною (додатною), а лівобічну – негативною (від'ємною). Вважається, що при $A_s < 0,25$ асиметрія низька, якщо A_s не перевищує 0,5, – середня та при $A_s > 0,5$ – висока.

Для вимірювання ексцесу використовують стандартизований момент 4-го порядку $E = \frac{m_4}{s_4}$. При симетричному, близькому до нормального розподілі $E = 3$, при гостровершинному розподілі $E > 3$, при плосковершинному $E < 3$.

Розрахунок центральних моментів m_3 і m_4 за даними інтервального ряду розподілу доцільно проводити за формулами:

$$m_3 = \frac{\sum \left(\frac{x_i - \bar{x}}{h}\right)^3 \cdot f_i}{\sum f_i} \cdot h^3; \quad (5.13)$$

$$m_4 = \frac{\sum \left(\frac{x_i - \bar{x}}{h}\right)^4 \cdot f_i}{\sum f_i} \cdot h^4. \quad (5.14)$$

де h – довжина інтервалу або будь-яке число;

f – частота або частість інтервалу.

Аналіз закономірностей розподілу можна поглибити, якщо описати його певною функцією.

5.5. Види та взаємозв'язок дисперсій

Дисперсія, або середній квадрат відхилення $D = \sigma^2$ посідає особливе місце у статистичному аналізі соціально-економічних явищ. Вона є невіддільним і важливим елементом інших статистичних методів, зокрема, дисперсійного аналізу.

Якщо сукупність розбита на групи за однією ознакою, то для неї можна обчислити такі види дисперсій: загальну, групові (часткові), середню з групових і міжгрупову.

Загальна дисперсія обчислюється як середнє арифметичне з квадратів відхилень кожного значення ознаки від їх загальної середньої величини. Така дисперсія характеризує варіацію досліджуваної ознаки за рахунок впливу всіх факторів. Її обчислюють за формулами:

– для незгрупованих даних (дисперсія проста):

$$\sigma_3^2 = \frac{\sum(x_i - \bar{x}_3)^2}{n}, \quad (5.15)$$

де n – обсяг сукупності;

\bar{x}_3 – загальне середнє значення ознаки усієї сукупності;

– для згрупованих даних (дисперсія зважена):

$$\sigma_3^2 = \frac{\sum(x_i - \bar{x}_3)^2 \cdot f_i}{\sum f_i}. \quad (5.16)$$

Групова (часткова) дисперсія визначається як середня арифметична з квадратів відхилень кожного значення в групі від групової середньої. Групові дисперсії обчислюються за формулами:

– для незгрупованих даних (дисперсія проста):

$$\sigma_i^2 = \frac{\sum(x_i - \bar{x}_i)^2}{n}, \quad (5.17)$$

– для згрупованих даних (дисперсія зважена):

$$\sigma_i^2 = \frac{\sum(x_i - \bar{x}_i)^2 \cdot f_i}{\sum f_i}. \quad (5.18)$$

де n – обсяг сукупності в групі;

x_i – індивідуальні значення групових ознак;

\bar{x}_i – середнє значення ознак i – ої групи;

f_i – групові частоти.

Групова (часткова) дисперсія вимірює варіацію ознаки лише за рахунок чинників, які діють усередині групи, тобто всіх інших чинників, крім покладеного в основу групування.

Середню внутрішньогрупову (залишкову) дисперсію визначають як середню арифметичну групових дисперсій:

$$\overline{\sigma_i^2} = \frac{\sum \sigma_i^2}{n} \text{ – проста; } \overline{\sigma_i^2} = \frac{\sum \sigma_i^2 \cdot f_i}{\sum f_i} \text{ – зважена.}$$

Міжгрупова дисперсія визначається як середня арифметична з квадратів відхилень групових середніх від загальної середньої за формулами:

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x}_3)^2}{n} \text{ – проста;} \quad (5.19)$$

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x}_3)^2 \cdot f_i}{\sum f_i} \text{ – зважена,} \quad (5.20)$$

де \bar{x}_i – групові середні; \bar{x}_3 – загальна середня.

Міжгрупова дисперсія відображає варіацію досліджуваної ознаки під впливом чинника, покладеного в основу групування.

Засобами математичної статистики встановлено, що між загальною дисперсією, середньою з групових дисперсій і міжгруповою дисперсією існує зв'язок:

$$\sigma_3^2 = \overline{\sigma_i^2} + \delta^2. \quad (5.21)$$

Ця рівність в статистиці називається *правилом додавання дисперсій*. За допомогою цього правила, знаючи два види дисперсій, завжди можна визначити невідомий третій вид:

$$\overline{\sigma_i^2} = \sigma_3^2 - \delta^2; \text{ або } \delta^2 = \sigma_3^2 - \overline{\sigma_i^2}.$$

Правило додавання дисперсій використовують при проведенні вибіркового спостереження, для спрощеного обчислення дисперсій громіздкого варіаційного ряду, вимірювання сили зв'язку між явищами тощо.

За допомогою відношення міжгрупової дисперсії до загальної дисперсії визначається показник, який називається *коефіцієнтом детермінації*, і показує, яка частка всієї варіації ознаки, обумовлена ознакою, покладеною в основу групування, тобто:

$$R^2 = \frac{\delta^2}{\sigma_3^2}. \quad (5.22)$$

Квадратний корінь з коефіцієнта детермінації називається *емпіричним кореляційним відношенням*, тобто: $\eta = \sqrt{\frac{\delta^2}{\sigma_3^2}}$. (5.23)

Питання для повторення і самоконтролю

1. Що таке статистичний ряд розподілу і які існують типи статистичних рядів розподілу?
2. Що таке варіаційний ряд розподілу і в чому проявляється закономірність розподілу?
3. Що таке частість розподілу і якими одиницями вона виражається?
4. Які ряди розподілу називаються дискретними та інтервальними і як виконується їх побудова?
5. Що таке кумулятивна частота варіаційного ряду розподілу?
6. Що таке середня арифметична, мода і медіана і як ці характеристики визначаються для дискретного ряду розподілу?
7. Як визначається мода і медіана для інтервального ряду розподілу?
8. Що називається розмахом варіації та як обчислюються середнє лінійне і середнє квадратичне відхилення?
9. Які відносні характеристики варіації Ви знаєте та як вони обчислюються?
10. Які сукупності називаються однорідними і що характерно для їхніх графіків розподілів?
11. Що таке асиметрія та ексцес і як вони використовуються для характеристики форми розподілу?
12. Для чого використовують моменти розподілу і як обчислюються початковий, центральний та умовний моменти k -го порядку?

13. Як виконується розрахунок центральних моментів третього і четвертого порядків за даними інтервального ряду розподілу?
14. Для чого використовується дисперсія при аналізі варіаційного ряду розподілу?
15. Що таке загальна дисперсія і як вона обчислюється?
16. Що таке групова (часткова) дисперсія і як вона обчислюється?
17. Що таке міжгрупова дисперсія і як вона обчислюється?
18. Який зв'язок існує між загальною, середньою з групових та між-груповою дисперсією?
19. Яка роль коефіцієнта детермінації і як він обчислюється?
20. Що таке емпіричне кореляційне відношення і як воно обчислюється?

Лекція 6. Аналіз концентрації, диференціації та подібності розподілу

План

- 6.1. Поняття коефіцієнтів концентрації і локалізації, їх обчислення та використання.
- 6.2. Коефіцієнт подібності структур двох сукупностей.
- 6.3. Інтенсивність структурних зрушень.

6.1. Поняття коефіцієнтів концентрації і локалізації, їх обчислення та використання

Важливою у статистичному аналізі є характеристика нерівномірності розподілу певної ознаки між окремими складовими сукупності, а також оцінка концентрації значень ознаки в окремих її частинах (наприклад, розподіл майна чи доходів між окремими групами населення, кількості зайнятих між окремими галузями промисловості, площі сільськогосподарських угідь між окремими агрогосподарствами).

Так, наведені в табл. 6.1 дані про розподіл промислових підприємств регіону за вартістю основних виробничих фондів і за обсягами спожитої електроенергії

свідчать про нерівномірне споживання електроенергії. До першої групи належить 20% підприємств, а частка спожитої електроенергії становить 4%. Натомість шоста група містить 3% підприємств, які споживають 46% електроенергії. На відхиленнях часток двох розподілів — за кількістю елементів сукупності d_j і обсягом значень ознаки D_j — ґрунтується оцінювання концентрації.

Таблиця 6.1

До розрахунку коефіцієнта концентрації

Вартість основних виробничих фондів, грн.	У % до підсумку		Модуль відхилення часток $\frac{1}{100} d_j - D_j $
	Кількість підприємств d_j	Спожито 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
Разом:	100	100	1,28

Якщо розподіл значень ознаки в сукупності рівномірний, то частки однакові, тоді $d_j = D_j$, відхилення часток свідчать про певну концентрацію.

Верхня межа суми відхилень $\sum |d_j - D_j| = 2$, а тому коефіцієнт концентрації обчислюється як півсума модулів відхилень:

$$K = \frac{1}{2} \sum_{j=1}^m |d_j - D_j|. \quad (6.1)$$

Значення коефіцієнта коливаються в межах від нуля (рівномірний розподіл) до одиниці (повна концентрація). Чим більший ступінь концентрації, тим більше значення коефіцієнта K . У нашому прикладі $K = 1,28 : 2 = 0,64$, що свідчить про високий ступінь концентрації споживання електроенергії у промисловості регіону.

Коефіцієнти концентрації широко використовуються в регіональному аналізі для оцінювання рівномірності територіального розподілу виробничих потужностей, фінансових ресурсів тощо. За кожним регіоном визначається також коефіцієнт

локалізації, який характеризує співвідношення часток та обчислюється за формулою:

$$L_j = \frac{D_j}{d_j} 100 \quad (6.2)$$

За даними табл. 6.2 коефіцієнти локалізації свідчать про нерівномірність купівлі (продажу) на душу населення і певною мірою про варіацію життєвого рівня населення різних регіонів.

Таблиця 6.2

Коефіцієнти територіальної локалізації

Регіон	У % до підсумку		Коефіцієнти локалізації L_j , %
	Чисельність населення d_j	Обсяг товарообороту D_j	
А	30	34	113
В	50	42	84
С	20	24	120
Разом:	100	100	-

Порівняння структур на основі відхилень часток доцільне в рядах з нерівними інтервалами, а, особливо, в атрибутивних рядах.

6.2. Коефіцієнт подібності структур двох сукупностей

За аналогією з коефіцієнтом концентрації обчислюється *коефіцієнт подібності (схожості) структур* двох сукупностей:

$$P = 1 - \frac{1}{2} \sum_1^m |d_j - d_k|. \quad (6.3)$$

Якщо структури однакові, то $P = 1$; якщо абсолютно протилежні, то $P = 0$. Чим більше схожі структури, тим більше значення P . За наведеними у табл. 6.3 даними про галузеву структуру зайнятості населення у двох країнах коефіцієнт подібності структур становить:

$$P = 1 - \frac{|36 - 25| + |24 - 42| + |40 - 33|}{2} = 1 - 0,18 = 0,82,$$

тобто розподіл зайнятих за галузями економіки відхиляється в середньому на 18 пунктів.

Таблиця 6.3

Галузева структура зайнятості населення

Країна	Структура зайнятих, %		
	Сільське господарство	Промисловість та будівництво	Сфера послуг
А	36	24	40
В	25	42	33

6.3. Інтенсивність структурних зрушень

Структура будь-якої статистичної сукупності динамічна. Змінюються склад і технічний рівень виробничих фондів, вікова й професійна структура робітників, склад і якість залучених до виробництва природних ресурсів, асортимент і якість продукції, що виробляється, структура споживчого бюджету тощо. Зміна часток окремих складових сукупності свідчить про структурні зрушення. Так, за даними табл. 6.4 структура спожитого в регіоні палива (у перерахунку на умовне) змінилася: зменшились частки газу та мазуту, зросли частки вугілля та інших видів палива. *Інтенсивність структурних зрушень* оцінюється за допомогою середнього лінійного \bar{l}_d або середнього квадратичного σ_d відхилень часток:

$$\bar{l}_d = \frac{\sum_1^m |d_{j1} - d_{j0}|}{m}; \quad (6.4)$$

$$\sigma_d = \sqrt{\frac{\sum_1^m (d_{j1} - d_{j0})^2}{m}}, \quad (6.5)$$

де d_{j0} та d_{j1} — частки відповідно базисного та поточного періоду; m — число складових сукупності.

Структура і структурні зрушення споживання палива по роках

Вид палива	1995 р., d_0	2000 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
Разом:	100	100	0	32	308

Лінійний коефіцієнт структурних зрушень становить $\bar{l}_d = \frac{32}{4} = 8$, тобто частки окремих видів палива змінилися в середньому на 8 п. п. Завдяки своїм математичним властивостям квадратичний коефіцієнт структурних зрушень дещо більший — $\sigma_d = \sqrt{\frac{308}{4}} = 8,8$ п. п.

Питання для повторення і самоконтролю

1. В чому полягає суть коефіцієнта концентрації, як він обчислюється і які може приймати значення?
2. Що характеризує коефіцієнт локалізації, як він обчислюється і в яких одиницях виражається?
3. Що характеризує коефіцієнт подібності і як він обчислюється?
4. Якими параметрами оцінюється інтенсивність структурних зрушень?

Лекції 7, 8. Статистичні методи вимірювання взаємозв'язків

План

7.1. Види взаємозв'язків. Основні поняття кореляційного та регресійного аналізу.

- 7.2. Поняття про дисперсійний аналіз.
- 7.3. Метод аналітичного групування і графічного зображення.
- 7.4. Парна лінійна кореляція і парна регресія.
- 7.5. Поняття про багатофакторну регресію.
- 7.6. Нелінійна кореляція.
- 7.7. Непараметричні методи вивчення зв'язків.
- 7.8. Використання кореляційно-регресійного аналізу в соціально-економічних дослідженнях.

7.1. Види взаємозв'язків

Одним із найзагальніших законів об'єктивного світу є закон загального зв'язку і залежності між явищами суспільного життя. Усі явища суспільного життя існують не ізольовано, а у нерозривному взаємозв'язку, тобто залежать одне від одного, тому вивчення будь-якого явища буде неповним, якщо не досліджені його зв'язки з іншими явищами і процесами. Статистичне дослідження взаємозв'язків дає можливість виявити не тільки наявність і напрямок зв'язку, але дозволяє кількісно оцінити і виразити його аналітично.

Визначення зв'язків між явищами дає змогу перейти від констатації фактів до пояснення і використання їх на практиці. Так, при вивченні урожайності сільськогосподарських культур можна визначити кількісні характеристики впливу багатьох факторів на урожайність. Це дозволяє виявити резерви зростання урожайності, встановити ступінь залежності їх як від об'єктивних причин, так і від умов діяльності сільськогосподарських підприємств. Визначення взаємозв'язків дозволяє проводити науково обґрунтовані прогнози.

Зв'язки між явищами, окремими їх ознаками досить різноманітні, однак у будь-якому випадку одні ознаки виступають як фактори, що впливають на інші і зумовлюють їх зміну, інші ж – результати дії цих факторів. Одні із них є причиною, інші наслідком.

Якщо перші прийнято називати ознаками-факторами, або факторними (причинними) ознаками, то другі — результативними (наслідковими) ознаками.

Різноманітність зв'язків, в яких перебувають явища, зумовлює необхідність їх класифікації, зведення зв'язків до певних типів, форм за їх істотними рисами і властивостями.

В основу класифікації зв'язків у статистиці покладено відмінність і подібність зв'язків за такими їх особливостями, як ступінь тісноти, спрямованість, аналітичне вираження, одиничність або множинність факторів. Відповідно до цього розрізняють зв'язки *функціональні і кореляційні, прямі та обернені, прямолінійні, криволінійні, однофакторні та багатфакторні*.

За статистичною природою зв'язки поділяють на *функціональні і статистичні*.

При функціональному зв'язку *кожному* можливному значенню факторної ознаки X відповідає *чітко визначене значення результативної ознаки* Y , тобто функціональні зв'язки характеризуються повною відповідністю між причиною і наслідком, факторною і результативною ознаками. Така залежність притаманна фізичним, хімічним явищам тощо. У соціально-економічних процесах – це найчастіше зв'язок складових елементів розрахункових формул відповідних показників, наприклад, залежність валового збору від урожайності сільськогосподарської культури, кількості внесених добрив, розміру посівної площі тощо.

На відміну від функціональних, статистичні зв'язки неоднозначні. *Статистичні зв'язки* проявляються як узгодженість варіації двох чи більше ознак. У зв'язку між ознаками X і Y , що характеризують досліджуване явище чи процес, кожному значенню ознаки X (факторна ознака) відповідає не одне, а певна множина значень ознаки Y (результативна ознака), які утворюють, так званий, умовний розподіл. *В цьому випадку кожному значенню ознаки X ставиться у відповідність також одне число \bar{y}_x , яке є середнім арифметичним множини значень ознаки Y . Такий зв'язок між ознаками називається кореляційним*.

За напрямком дії (спрямованістю) розрізняють зв'язок *прямий та обернений*. *Прямий* — це такий зв'язок, при якому зі збільшенням або

зменшенням значень факторної ознаки відповідно збільшується або зменшується значення результативної ознаки, тобто факторна і результативна ознаки змінюються в одному напрямку.

Прикладом прямого зв'язку може бути зв'язок між фондоозброєністю і продуктивністю праці, між собівартістю продукції і рівнем рентабельності.

Оберненим зв'язком називають такий, при якому значення результативної ознаки змінюється в протилежному напрямку відносно зміни значення факторної ознаки. Прикладом такого зв'язку може бути продуктивність праці і собівартість продукції: підвищення продуктивності праці призводить до зниження собівартості продукції.

За формою аналітичного вираження в загальній класифікації виділяють *зв'язки прямолінійні та криволінійні*. Якщо певний зв'язок явищ можна точно або наближено зобразити рівнянням будь-якої прямої лінії, то його називають *лінійним (прямолінійним) зв'язком*, а якщо рівнянням будь-якої кривої лінії (параболи, гіперболи і т.п.) — *нелінійним (криволінійним)*.

Аналітичним рівнянням можна описувати лише функціональні зв'язки. Кореляційні зв'язки описуються рівнянням наближено, *таке рівняння називається рівнянням регресії Y на X* (можливий також обернений зв'язок, який виражається рівнянням регресії X на Y) і має загальний вигляд $\overline{y_x} = f(x)$.

Однак, навіть наближене описування за допомогою рівнянь регресії кореляційних зв'язків суспільних явищ дає можливість отримати результати, цілком придатні для наукових і практичних потреб.

7.2. Поняття про дисперсійний аналіз

В епоху розвитку ринкової економіки використання методів математичної статистики в економічних дослідженнях стає нагальною необхідністю. Як і інші ймовірно-статистичні методи, він набагато розширює можливості економістів в аналізі виробництва й значно підвищує рівень наукових досліджень.

Головне призначення дисперсійного аналізу – *статистично виявити вплив різних факторів на мінливість ознаки, що вивчається*. Особливий інтерес становить використання методу в аналізі економічних процесів та явищ, коли мінливість результативної ознаки зумовлена одночасно дією кількох факторів із неоднаковою силою впливу. Зокрема, це спостерігається при аналізі результативних синтетичних показників економічної ефективності виробництва. Найбільш ефективний тут одночасний дисперсійний аналіз усіх відібраних факторів – багатофакторний аналіз.

За допомогою дисперсійного аналізу досліджують вплив однієї або декількох незалежних змінних (факторів) на одну залежну змінну (результативну ознаку) у випадку однофакторного аналізу або на декілька залежних змінних (багатофакторний аналіз).

Незалежні змінні в цьому випадку приймають лише дискретні значення.

Однофакторний дисперсійний аналіз використовується у тих випадках, коли є в розпорядженні три або більше незалежних масивів даних (груп), отриманих з однієї загальної сукупності шляхом зміни якого-небудь фактора, для якого, з якихось причин, немає кількісних вимірів.

Для цих груп припускають, що вони мають різні внутрішньогрупові середні й однакові групові дисперсії. Тому необхідно відповісти на питання: спричинив цей фактор істотний вплив на зміну групових середніх залежної змінної чи така зміна (розсіювання значень) є наслідком випадків, викликаних невеликими обсягами груп? Іншими словами, якщо групи належать до однієї і тієї ж загальної сукупності, то розсіювання даних між групами повинно бути не більшим, ніж розсіювання даних усередині цих груп.

Дисперсійний аналіз виконується при допущенні, що вихідна величина є випадковою і підпорядковується нормальному закону розподілу ймовірностей. Основна ідея дисперсійного аналізу полягає у дослідженні відмінності між середніми значеннями окремих груп спостережень. Кожна група характеризується певним рівнем або діапазоном фактора. Може здаватися дивним, що процедура порівняння середніх називається дисперсійним аналізом.

Насправді це пов'язано з тим, що при дослідженні статистичної значущості різниці між середніми значеннями груп ми порівнюємо оцінки дисперсій.

В основі дисперсійного аналізу лежить представлення загальної дисперсії як суми двох дисперсій – внутрішньогрупової дисперсії та дисперсії, підрахованої без урахування належності до якоїсь групи. Якщо фактор є впливовим, то дисперсія спостережень в окремій групі повинна бути значно меншою, ніж дисперсія всієї сукупності. Причина цього, очевидно, полягає в суттєвій різниці між середніми значеннями у групах.

Розглянемо такий набір даних (табл. 7.1).

Аналіз результатів табл. 7.1 показує, що середні двох груп суттєво відрізняються (2 і 6 відповідно). Сума квадратів відхилень у кожній групі однакова і дорівнює 2. Додаючи їх, отримуємо суму квадратів відхилень від середніх у групах:

$$D^{**} = 2 + 2 = 4.$$

Таблиця 7.1

До розрахунку показників дисперсійного аналізу

	Номер групи	
	Група 1	Група 2
Спостереження 1	2	7
Спостереження 2	1	5
Спостереження 3	3	6
Середнє	$(2+1+3)/3 = 2$	$(7+5+6)/3 = 6$
Сума квадратів відхилень у групах	$(2 - 2)^2 + (1 - 2)^2 + (3 - 2)^2 = 2$	$(7 - 6)^2 + (5 - 6)^2 + (6 - 6)^2 = 2$
Загальне середнє	$(2+1+3+7+5+6)/6 = 4$	
Загальна сума квадратів відхилень	$(2 - 4)^2 + (1 - 4)^2 + (3 - 4)^2 + (7 - 4)^2 + (5 - 4)^2 + (6 - 4)^2 = 28$	

Якщо тепер не враховувати наявність окремих груп, тобто впливу фактора на результат, то загальна сума квадратів відхилень від загального середнього цих двох вибірок буде: $D = 28$.

Відомо, що основна тотожність дисперсійного аналізу має вигляд:

$$D = D^* + D^{**},$$

де D^* – сума квадратів, яка обумовлена різницею середніх значень між групами, тобто $D^* = D - D^{**} = 28 - 4 = 24$.

Перевірка значущості в дисперсійному аналізі заснована на порівнянні складової дисперсії, обумовленої міжгруповим розсіюванням, і складової дисперсії, обумовленої внутрішньогруповим розсіюванням.

Отримані складові дисперсії можна порівняти за допомогою критерію Фішера-Снедекора (F – критерію), що перевіряє, чи дійсно відношення дисперсій значно більше за 1, тобто дисперсія, що обумовлена міжгруповим розсіюванням, значно більша за складову дисперсію, обумовлену внутрішньогруповим розсіюванням.

Обчислимо критерій Фішера-Снедекора за формулою:

$$F = \frac{D^*/(m-1)}{D^{**}/(n-m)}, \quad (7.1)$$

де m – кількість груп; n – кількість даних.

Одержимо: $F = \frac{24/(2-1)}{4/(6-2)} = 24$.

Порівняємо це значення з критичним значенням розподілу Фішера, яке залежить від рівня значущості $\alpha = 0,05$ і двох значень ступенів волі $m - 1 = 1$ та $n - m = 4$. За таблицею знаходимо: $F_{кр.} = 7,71$. Оскільки $F > F_{кр.}$, робимо висновок, що наявність груп впливає на результат.

Якщо фактор не є впливовим, сума дисперсій спостережень окремих груп не відрізняється суттєво від дисперсії всієї сукупності. Якщо справджується нульова гіпотеза (рівність середніх у двох групах), то можна чекати порівняно невеликої розбіжності групових середніх через чисто випадкову мінливість. Тоді внутрішньогрупова дисперсія буде практично збігатися із загальною дисперсією, підрахованою без врахування групової приналежності. Розглянемо такий набір даних, помінявши значення спостережень у групі 2 (табл. 7.2).

До розрахунку показників дисперсійного аналізу

	Номер групи	
	Група 1	Група 2
Спостереження 1	2	3
Спостереження 2	1	2
Спостереження 3	3	1
Середнє	$(2+1+3)/3 = 2$	$(3+2+1)/3 = 2$
Сума квадратів відхилень у групах	$(2 - 2)^2 + (1 - 2)^2 + (3 - 2)^2 = 2$	$(3 - 2)^2 + (2 - 2)^2 + (1 - 2)^2 = 2$
Загальне середнє	$(2+1+3+3+2+1)/6 = 2$	
Загальна сума квадратів відхилень	$(2 - 2)^2 + (1 - 2)^2 + (3 - 2)^2 + (3 - 2)^2 + (2 - 2)^2 + (1 - 2)^2 = 4$	

Сума квадратів відхилень у кожній групі знову дорівнює 2, а їхня сума $D^{**} = 2 + 2 = 4$. Але сума квадратів, яка обумовлена різницею середніх значень між групами, дорівнює нулю, тобто $D^* = D - D^{**} = 4 - 4 = 0$.

Значення критерію Фішера згідно з формулою (7.1) теж дорівнює нулю, що дає підставу зробити висновок про відсутність впливу наявності групи.

7.3. Метод аналітичного групування і графічного зображення

Для відповіді на питання про наявність або відсутність кореляційного зв'язку та про загальний вигляд відповідного рівняння регресії у випадку парної кореляції використовують в першу чергу такі прийоми: паралельне порівняння рядів значень факторної і результативної ознак, балансовий метод, метод аналітичного групування, графічне зображення.

Елементарним способом виявлення зв'язку є паралельне порівняння ряду значень факторної ознаки і відповідних значень результативної ознаки. Значення факторної ознаки розташовують у порядку зростання і потім відслідковують напрям зміни величини результативної ознаки. В тих випадках, коли збільшення величини факторної ознаки призводить до збільшення

результативної ознаки, можна стверджувати про можливу наявність прямого кореляційного зв'язку. Якщо ж із збільшенням факторної ознаки, величина результативної ознаки має тенденцію до зменшення, то можна передбачати обернений зв'язок між ознаками.

На першому етапі аналізу кореляційного зв'язку при обґрунтуванні моделі постають два питання: вибір факторних ознак і визначення числа груп і меж інтервалів.

При визначенні числа груп і меж інтервалів слід зважати на той факт, що типовість та сталість групових середніх залежить від чисельності груп.

На практиці аналітичне групування часто виконується за принципом рівних інтервалів, що значно спрощує подальший аналіз зв'язку.

Метод аналітичного групування полягає в тому, що всі елементи сукупності групують, як правило, за факторною ознакою X і в кожній групі обчислюють середні значення \bar{y}_i , результативної ознаки Y . За одержаними значеннями, які розглядають як координати точок $(x_i; \bar{y}_i)$, будують відповідні точки у прямокутній системі координат $XO\bar{Y}_x$. Множина таких точок називається *кореляційним полем*.

За виглядом кореляційного поля можна судити про наявність чи відсутність кореляційного зв'язку між ознаками X та Y , а також зробити припущення про тип кореляційної залежності (лінійна, гіперболічна, квадратична тощо), а отже, і про загальний вигляд відповідного рівняння регресії, яке її описує.

За виглядом кореляційного поля, побудованого на рис. 7.1, а, можна зробити припущення, що кореляційна залежність близька до лінійної, оскільки ці точки розміщуються досить близько від деякої прямої.

Якщо побудовані точки сполучити плавною лінією, то одержимо наближений графік, який називається *емпіричною лінією парної регресії* (рис. 7.1, б)

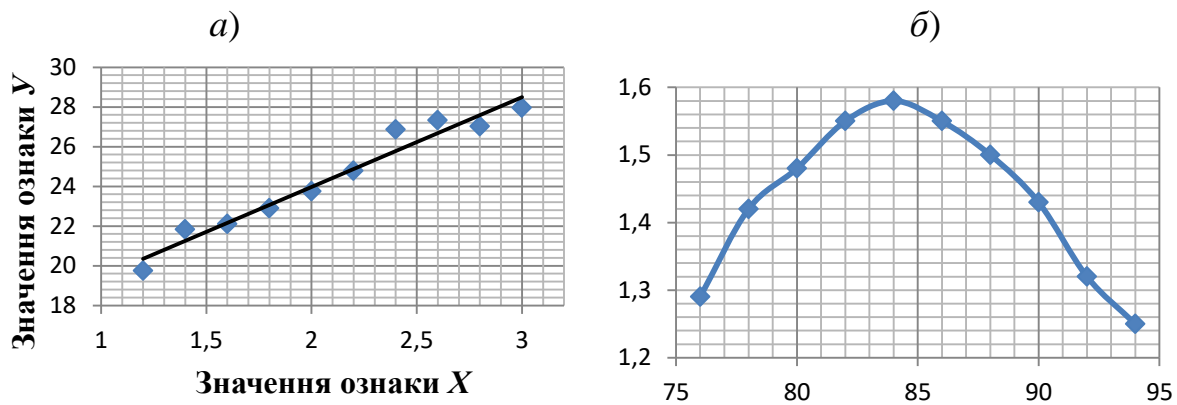


Рис. 7.1. Графічне зображення кореляційного зв'язку парної регресії:

а) лінійна кореляція; б) квадратична (параболічна) кореляція

Із побудованих графіків слідує, що в першому випадку можна припустити існування між ознаками лінійної кореляції; в другому випадку – параболічної (квадратичної), оскільки плавна лінія, що сполучає побудовані точки, є близькою до параболи.

7.4. Парна лінійна кореляція і парна регресія

При подальших дослідженнях будемо вважати, для визначеності, що між ознаками X та Y загальної сукупності існує лінійна кореляційна залежність (рис. 7.1, а), тому відповідне рівняння лінійної регресії має загальний вигляд

$$\bar{y}_x = ax + b, \quad (7.2)$$

причому коефіцієнт a , який називається *коефіцієнтом лінійної парної регресії*, вказує на те, що при зміні (збільшенні чи зменшенні) значення x на одну одиницю відповідне значення \bar{y}_x зміниться на величину a .

Наступна задача полягає в тому, щоб відшукати оптимальні значення параметрів a і b , які входять до рівняння (7.2), причому їхні значення потрібно обчислити за умови, щоб кожна точка кореляційного поля розміщувалась якомога ближче до прямої, яка є наближеним графіком лінійної кореляційної залежності. Для розв'язання задачі використовується відомий метод, який називається *методом найменших квадратів*. Його суть полягає в тому, що значення невідомих параметрів рівняння (7.2) визначають при умові, щоб сума квадратів відхилень усіх точок кореляційного поля від прямої була найменшою. В цьому смислі забезпечується оптимальність шуканих значень параметрів. Для

відшукування цих значень складають і розв'язують систему лінійних рівнянь, яка називається *нормальною* і має загальний вигляд:

$$\begin{cases} a \sum x_i^2 + b \sum x_i = \sum x_i y_i, \\ a \sum x_i + n \cdot b = \sum y_i. \end{cases} \quad (7.3)$$

Однак, систему (7.3) складають лише тоді, коли переконаються у достатній силі лінійного кореляційного зв'язку між ознаками. Силу або тісноту лінійного кореляційного зв'язку характеризує параметр, який називається *коефіцієнтом лінійної кореляції*, та обчислюється за формулою:

$$r = \frac{\sum x \cdot y - n \cdot \bar{x} \cdot \bar{y}}{n \cdot \sigma_x \cdot \sigma_y}, \quad (7.4)$$

де x і y – значення (варіанти) відповідних ознак X і Y ;

n – обсяг сукупності усіх пар значень ознак X і Y ;

σ_x і σ_y – середні квадратичні відхилення для ознак X і Y відповідно;

\bar{x} і \bar{y} – середні значення.

Можливі значення коефіцієнта r лінійної кореляції коливаються в межах від -1 до 1 , тобто $-1 \leq r \leq 1$, або $|r| \leq 1$. Крайні значення $r = \pm 1$ свідчать про те, що між ознаками X і Y існує повна кореляція, яка виродилась у *функціональну лінійну залежність*. Прийнято вважати, що при $0 < |r| < 0,3$ кореляційний зв'язок слабкий; при $0,3 < |r| < 0,7$ – зв'язок середній, а при $0,7 < |r| < 1$ – сильний. Значення r характеризують не лише тісноту, але й напрям зв'язку. Додатне значення r вказує на прямий зв'язок між ознаками, а від'ємне – на обернений.

Крім коефіцієнта r лінійної кореляції для наступних досліджень (після того як одержано рівняння регресії) використовують ще один параметр, який є мірою щільності кореляційного зв'язку і називається *коефіцієнтом детермінації* R^2 . Його обчислюють за формулою:

$$R^2 = \frac{\delta_y^2}{\sigma_y^2}, \quad (7.5)$$

де σ_y^2 і δ_y^2 – загальна і факторна (теоретична) дисперсія ознаки Y відповідно:

$$\sigma_y^2 = \frac{1}{n} \cdot \sum (y_i - \bar{y})^2; \quad (7.6)$$

$$\delta_y^2 = \frac{1}{n} \cdot \sum (\bar{y}_x - \bar{y})^2. \quad (7.7)$$

Наведена формула (7.7) для обчислення факторної дисперсії вимагає попереднього визначення (розрахунку) за допомогою одержаного рівняння регресії теоретичних значень \bar{y}_x для всіх відповідних значень x_i сукупності.

Потрібно зазначити, що значення коефіцієнта детермінації R^2 коливаються в межах від 0 до 1, причому чим ближчі вони до одиниці, тим щільніший кореляційний зв'язок між ознаками.

Даний показник є універсальним. Його можна застосовувати при будь-якій формі залежності (лінійній або нелінійній).

Приклад 1. Нехай задано два масиви чисел, які є відповідними значеннями x_i та \bar{y}_i ознак X та Y , що перебувають між собою у деякому співвідношенні:

x_i	1,2	1,4	1,6	1,8	2,0	2,2	2,4	2,6	2,8	3,0
\bar{y}_i	19,74	21,83	22,10	22,89	23,75	24,78	26,85	27,32	27,02	27,95

Необхідно: 1) графічним методом виявити можливість існування кореляційної залежності між ознаками X та Y , встановити її тип і загальний вигляд рівняння регресії;

2) оцінити силу кореляційного зв'язку між ознаками та зробити відповідний висновок;

3) у випадку, коли сила кореляційного зв'язку є достатньою, знайти невідомі параметри рівняння регресії Y на X методом найменших квадратів. Обчислити коефіцієнт детермінації R^2 та зробити висновок.

Р о з в' я з а н н я .

1. Використовуючи задані масиви значень ознак X і Y , побудуємо кореляційне поле (рис. 7.2).

За розміщенням точок на площині можемо зробити припущення, що між ознаками X і Y існує кореляційна залежність, яка близька до лінійної.

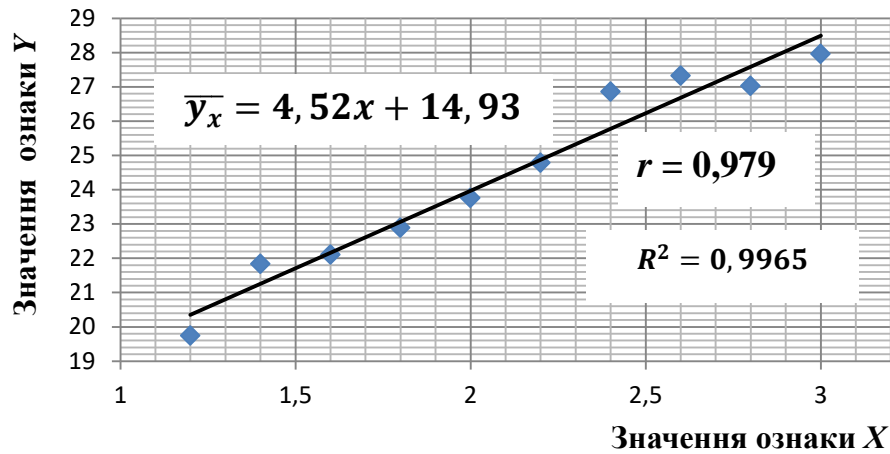


Рис. 7.2. Графічне зображення залежності (приклад 1)

Щоб переконатись у силі (тісноті) лінійного кореляційного зв'язку між ознаками обчислимо коефіцієнт лінійної кореляції за формулою. Для цього попередньо знайдемо значення показників \bar{x} , \bar{y} , σ_x і σ_y за формулами (4.1) і (5.15) та обчислимо суму $\sum x \cdot y$:

$$\bar{x} = \frac{1,2 + 1,4 + 1,6 + 1,8 + 2,0 + 2,2 + 2,4 + 2,6 + 2,8 + 3,0}{10} = 2,10;$$

$$\bar{y} = \frac{19,74 + 21,83 + 23,43 + 22,89 + 23,75 + 24,78 + 26,85 + 27,32 + 25,98 + 27,95}{10} = 24,45;$$

$$\begin{aligned} \sigma_x^2 &= \frac{1}{10} \cdot ((1,2 - 2,1)^2 + (1,4 - 2,1)^2 + (1,6 - 2,1)^2 + (1,8 - 2,1)^2 + \\ &(2,0 - 2,1)^2 + (2,2 - 2,1)^2 + (2,4 - 2,1)^2 + (2,6 - 2,1)^2 + (2,8 - 2,1)^2 + \\ &(3,0 - 2,1)^2) = \frac{1}{10} \cdot (0,81 + 0,49 + 0,25 + 0,09 + 0,01 + 0,01 + 0,09 + 0,25 + \\ &0,49 + 0,81) = 0,33. \end{aligned}$$

$$\text{Тоді } \sigma_x = \sqrt{0,33} = 0,57.$$

$$\begin{aligned} \sigma_y^2 &= \frac{1}{10} ((19,74 - 24,45)^2 + (21,83 - 24,45)^2 + (22,10 - 24,45)^2 + \\ &(22,89 - 24,45)^2 + (23,75 - 24,45)^2 + (24,78 - 24,45)^2 + (26,85 - 24,45)^2 + \\ &(27,32 - 24,45)^2 + (27,02 - 24,45)^2 + (27,95 - 24,45)^2) = \frac{1}{10} \cdot (22,1841 + \\ &0,3844 + 5,5225 + 2,4336 + 0,4900 + 0,1089 + 5,7600 + 8,2369 + 6,6049 + \\ &12,2500) = 6,3975; \end{aligned}$$

$$\text{Тоді } \sigma_y = \sqrt{6,3975} = 2,53.$$

$$\sum x \cdot y = 1,2 \cdot 19,74 + 1,4 \cdot 21,83 + 1,6 \cdot 22,10 + 1,8 \cdot 22,89 + 2,0 \cdot 23,75 + 2,2 \cdot 24,78 + 2,4 \cdot 26,85 + 2,6 \cdot 27,32 + 2,8 \cdot 27,02 + 3,0 \cdot 27,95 = 527,806.$$

Використовуючи формулу (7.4), одержимо:

$$r = \frac{527,806 - 10 \cdot 2,10 \cdot 24,45}{10 \cdot 0,57 \cdot 2,53} = \mathbf{0,979}.$$

Значення коефіцієнта лінійної кореляції близьке до одиниці, що свідчить про дуже тісний лінійний зв'язок між ознаками.

Рівняння регресії, що описує цей зв'язок, шукаємо у вигляді (7.2). Для зручності обчислення коефіцієнтів системи (7.3) складаємо допоміжну таблицю (табл. 7.3).

Таблиця 7.3

Розрахунок коефіцієнтів нормальної системи

i	x_i	\bar{y}_i	x_i^2	$x_i \cdot y_i$
1	1,2	19,74	1,44	23,688
2	1,4	21,83	1,96	30,562
3	1,6	22,10	2,56	35,36
4	1,8	22,89	3,24	41,202
5	2,0	23,75	4,00	47,50
6	2,2	24,78	4,84	54,516
7	2,4	26,85	5,76	64,44
8	2,6	27,32	6,76	71,032
9	2,8	27,02	7,84	75,656
10	3,0	27,95	9,00	83,85
Сума:	21,0	244,23	47,40	527,806

З урахуванням отриманих даних, що записані в останньому рядку таблиці, записуємо конкретний вигляд системи (7.3):

$$\begin{cases} 47,40a + 21,0b = 527,806, \\ 21,0a + 10b = 244,23. \end{cases}$$

Для розв'язання системи методом підстановки поділимо обидві частини другого рівняння на 10 і після цього виразимо з нього b :

$$b = 24,423 - 2,1a.$$

Підставивши одержаний вираз замість b в перше рівняння системи, визначаємо $a = 4,52$. Тоді $b = 14,93$ і шукане рівняння регресії має вигляд:

$$\bar{y}_x = 4,52x + 14,93.$$

Переконаємось у щільності кореляційного зв'язку за допомогою коефіцієнта детермінації R^2 .

Загальна дисперсія нам відома, тобто $\sigma_y^2 = 6,3975$. Обчислимо факторну (теоретичну) дисперсію δ_y^2 . Для цього розрахуємо за допомогою одержаного лінійного рівняння регресії необхідні значення \bar{y}_x для відповідних значень x : $\bar{y}_{1,2} = 20,354$; $\bar{y}_{1,4} = 21,258$; $\bar{y}_{1,6} = 21,05$; $\bar{y}_{1,8} = 23,066$; $\bar{y}_{2,0} = 23,97$; $\bar{y}_{2,2} = 24,874$; $\bar{y}_{2,4} = 25,778$; $\bar{y}_{2,6} = 26,682$; $\bar{y}_{2,8} = 27,586$; $\bar{y}_{3,0} = 28,49$.

За формулою (7.7) маємо:

$$\begin{aligned} \delta_y^2 = & \frac{1}{10} ((20,354 - 24,45)^2 + (21,258 - 24,45)^2 + (21,05 - 24,45)^2 + \\ & (23,066 - 24,45)^2 + (23,97 - 24,45)^2 + (24,874 - 24,45)^2 + (25,778 - \\ & 24,45)^2 + (26,682 - 24,45)^2 + (27,586 - 24,45)^2 + (28,49 - 24,45)^2) = \\ & \frac{1}{10} \cdot (16,7772 + 10,1889 + 11,56 + 1,9155 + 0,2304 + 0,1798 + 1,7636 \\ & + 4,9818 + 9,8345 + 16,3216) = 6,3753. \end{aligned}$$

$$\text{За формулою (7.5) одержимо: } R^2 = \frac{6,3753}{6,3975} = 0,9965.$$

Близькість коефіцієнта детермінації до одиниці підтверджує високу щільність розглянутого лінійного кореляційного зв'язку між ознаками та значний вплив факторної ознаки X на результативну ознаку Y . Із одержаного рівняння слідує, що зміна x на одну одиницю призводить до збільшення \bar{y}_x на 4,52 одиниць.

Зауваження.

1. Якщо у процесі розв'язання задачі відомі показники \bar{x} , \bar{y} , r , σ_x і σ_y , то рівняння лінійної регресії шукають у вигляді

$$\bar{y}_x = \bar{y} + r \cdot \frac{\sigma_y}{\sigma_x} (x - \bar{x}),$$

яке рівносильне рівнянню (7.2), оскільки теоретично виведене з нього шляхом тотожних перетворень.

2. Процес обчислень значно спрощується), якщо скористатись комп'ютерною програмою *Excel*, до якої належать стандартні підпрограми *Коррел* та *Линейн* (все зводиться лише до набору масивів, для яких потрібно обчислити коефіцієнт лінійної кореляції та знайти невідомі параметри рівняння лінійної регресії).

7.5. Поняття про багатofакторну регресію

Якщо при статистичному дослідженні беруть до уваги залежність результативної ознаки Y лише від однієї факторної ознаки X , то такий зв'язок між ознаками називається *однофакторним*; якщо ж вивчається залежність результативної ознаки Y від багатьох факторних ознак $X_1, X_2, X_3, \dots, X_n$, то – *багатofакторним*. Відповідно, в першому випадку розглядають *парну*, а в другому – *множинну кореляцію*. Рівняння регресії, що відображає множинну кореляцію, має загальний вигляд $\bar{y}_x = f(x_1, x_2, x_3, \dots, x_m)$.

При теоретичному обґрунтуванні моделі і виборі факторних ознак слід враховувати тісноту кореляційного зв'язку між ознаками. При наявності зв'язку, який близький до функціонального має місце явище, яке називається *мультиколінеарністю*, тому без його врахування формально одержане багатofакторне рівняння регресії в цьому випадку може бути хибним, а отже, оцінки його параметрів будуть ненадійними. Для оцінювання наявності мультиколінеарності між ознаками достатньо обчислити відповідні коефіцієнти кореляції. Якщо коефіцієнт кореляції двох факторних ознак близький до одиниці, то одну з них треба виключити. На цьому етапі важливо не тільки вибрати фактори, але й розкрити структуру взаємозв'язку між ними.

Проблема обґрунтування загального вигляду багатofакторного рівняння регресії є надто складною. Аналіз парних зв'язків непридатний, тому що фактори взаємозв'язані, а визначити зв'язок між \bar{y}_x і окремим фактором x_i при фіксованих значеннях інших факторних ознак також дуже складно. Тому на практиці найчастіше використовують *багатofакторні лінійні рівняння* і рівняння, що зводяться до лінійного вигляду відповідними перетвореннями, тобто:

$$\bar{y}_x = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_mx_m. \quad (7.8)$$

Багатofакторне рівняння (7.8) має серйозну перевагу перед лінійним однофакторним рівнянням (7.2), оскільки воно виражає в комплексі залежність результативної ознаки \bar{y}_x від m факторних ознак: за його допомогою можна в'яснити значущість впливу кожного окремого фактора x_i на величину результативної ознаки. Наприклад, можна стверджувати, що при зміні фактора x_1 на одну одиницю (при фіксованих значеннях інших факторів) результативна ознака \bar{y}_x зміниться на a_1 одиниць (збільшиться, якщо $a_1 > 0$, і зменшиться у протилежному випадку).

Обчислення значень невідомих коефіцієнтів рівняння (7.8) є дуже громіздким процесом, тому його виконують за допомогою комп'ютерних програм.

7.6. Нелінійна кореляція

У практиці соціально-економічних досліджень, крім лінійної, використовують також і нелінійну кореляцію, яка виражається нелінійними рівняннями регресії: гіперболічну, параболічну (квадратичну), степеневу та деякі інші. Рівняння нелінійної регресії мають загальний вигляд:

$$\bar{y}_x = \frac{a}{x} + b - \text{рівняння гіперболічної регресії};$$

$$\bar{y}_x = ax^2 + bx + c - \text{рівняння параболічної (квадратичної) регресії};$$

$$\bar{y}_x = ax^m - \text{рівняння степеневої регресії}.$$

Рівняння гіперболічної регресії легко зводиться до лінійного рівняння шляхом заміни змінної за формулою $\frac{1}{x} = z$. Тоді це рівняння набере вигляду $\bar{y}_z = az + b$ і його параметри a і b шукають методом найменших квадратів за схемою, розглянутою раніше. Для цього значення x_i факторної ознаки X

попередньо потрібно перерахувати за формулою $\frac{1}{x} = z$ і перейти в рівнянні до нової змінної z . Після обчислення значень параметрів a і b їх необхідно підставити у загальне рівняння $\bar{y}_x = \frac{a}{x} + b$.

Зупинимось детально на випадку параболічної (квадратичної) кореляційної залежності, яка графічно представлена емпіричною лінією регресії на рис. 7.1, б.

Для обчислення значень невідомих коефіцієнтів a, b і c кореляційного рівняння регресії використовують нормальну систему рівнянь, яка одержана способом найменших квадратів і в цьому випадку має загальний вигляд:

$$\begin{cases} a \sum x_i^4 + b \sum x_i^3 + c \sum x_i^2 = \sum \bar{y}_i \cdot x_i^2, \\ a \sum x_i^3 + b \sum x_i^2 + c \sum x_i = \sum \bar{y}_i \cdot x_i, \\ a \sum x_i^2 + b \sum x_i + n \cdot c = \sum \bar{y}_i. \end{cases} \quad (7.9)$$

Сила або тіснота нелінійного кореляційного зв'язку оцінюється за допомогою кореляційного відношення η_{yx} або η_{xy} .

Кореляційне відношення Y до X визначається як квадратний корінь з коефіцієнта детермінації, тобто:

$$\eta_{yx} = \sqrt{R^2} = \sqrt{\frac{\sum(\bar{y}_x - \bar{y})^2}{\sum(y_i - \bar{y})^2}}. \quad (7.10)$$

Кореляційне відношення має такі властивості:

1. Кореляційне відношення задовольняє подвійну нерівність $0 \leq \eta \leq 1$.
2. Якщо $\eta = 0$, то між ознаками X і Y кореляційного зв'язку немає.
3. Якщо $\eta = 1$, то ознака Y пов'язана з ознакою X функціональною залежністю.
4. Кореляційне відношення задовольняє умову: $\eta \geq |r|$.
5. Якщо $\eta = |r|$, то між ознаками X і Y існує точна лінійна кореляційна залежність, тобто всі точки, що визначаються відповідними парами варіант, лежать на прямій лінії регресії, знайденій способом найменших квадратів.

Із властивостей кореляційного відношення слідує, що при наближенні значень η до одиниці зв'язок між ознаками X і Y стає все сильнішим, переходячи у функціональний при $\eta = 1$.

На підставі сказаного вище можемо зробити висновок, що вибіркоче кореляційне відношення є мірою сили будь-якого кореляційного зв'язку, в тому числі і лінійного.

Приклад 2. Нехай задано два масиви чисел, які є відповідними значеннями x_i та \bar{y}_i ознак X та Y , що перебувають між собою у деякому співвідношенні:

x_i	5	6	7	8	9	10	11	12	13	14
\bar{y}_i	3,91	6,39	7,80	9,06	9,80	10	9,58	8,7	7,86	6,09

Необхідно: 1) графічним методом виявити можливість існування кореляційної залежності між ознаками X та Y , встановити її тип і загальний вигляд рівняння регресії;

2) знайти невідомі параметри рівняння регресії Y на X методом найменших квадратів;

3) оцінити силу кореляційного зв'язку між ознаками та зробити відповідний висновок.

Р о з в' я з а н н я .

1. Використовуючи задані масиви значень ознак X і Y , побудуємо відповідні точки кореляційного поля та емпіричну лінію регресії Y на X (рис. 7.3).

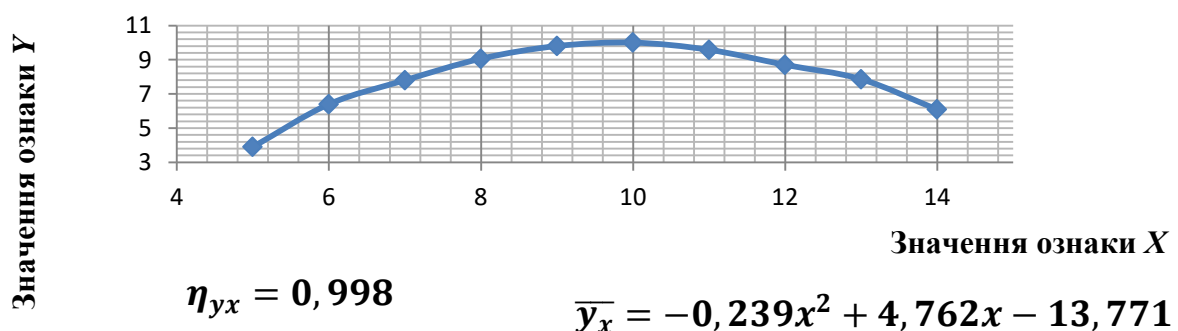


Рис. 7.3. Графічне зображення параболічного кореляційного зв'язку (приклад 2)

За розміщенням точок на площині та за виглядом емпіричної лінії регресії приходимо до висновку, що між ознаками існує параболічна (квадратична) залежність.

2. Рівняння параболічної (квадратичної) залежності шукаємо у вигляді

$$\bar{y}_x = ax^2 + bx + c, \quad (7.11)$$

невідомі коефіцієнти якого обчислимо за допомогою нормальної системи рівнянь (7.9). Для зручності їх обчислень складемо допоміжну таблицю 7.4.

Таблиця 7.4

Розрахунок коефіцієнтів нормальної системи

i	x_i	\bar{y}_i	x_i^2	x_i^3	x_i^4	$x_i \bar{y}_i$	$\bar{y}_i x_i^2$
1	5	3,90	25	125	625	19,5	97,5
2	6	6,40	36	216	1296	38,4	230,4
3	7	7,80	49	343	2401	54,6	382,2
4	8	9,06	64	512	4096	72,48	579,84
5	9	9,80	81	729	6561	88,2	793,8
6	10	10	100	1000	10000	100	1000
7	11	9,58	121	1331	14641	105,38	1159,18
8	12	8,7	144	1728	20736	104,4	1252,8
9	13	7,86	169	2197	28561	102,18	1328,34
10	14	6,09	196	2744	38416	85,26	1193,64
Сума:	95	79,19	985	10925	127333	770,4	8017,7

Використовуючи обчислені суми, що записані в останньому рядку таблиці 7.4, запишемо систему (7.9) у конкретному вигляді:

$$\begin{cases} 127333 a + 10925 b + 985 c = 8017,7, \\ 10925 a + 985 b + 95 c = 770,4, \\ 985 a + 95 b + 10 c = 79,19. \end{cases}$$

Розв'язавши цю систему трьох лінійних рівнянь будь-яким відомим методом (методом Гаусса, методом Крамера, за допомогою елементарних прийомів), одержимо шукані значення параметрів рівняння регресії (7.11):

$$a = -0,239; \quad b = 4,762; \quad c = -13,771.$$

Отже, шукане рівняння параболічної регресії має вигляд:

$$\bar{y}_x = -0,239x^2 + 4,762x - 13,771.$$

Графічне зображення розглянутого параболічного кореляційного зв'язку представлено на рис. 7.3.

3. Щоб переконатись у силі (тісноті) квадратичного кореляційного зв'язку, що описується одержаним рівнянням, обчислимо значення кореляційного відношення η . Для цього, використовуючи умову задачі, попередньо обчислимо показник \bar{y} , за допомогою одержаного рівняння регресії обчислимо теоретичні значення \bar{y}_x та суми квадратів відхилень $\sum(\bar{y}_x - \bar{y})^2$ і $\sum(y_i - \bar{y})^2$.

Середнє арифметичне: $\bar{y} = \frac{1}{10} \cdot (3,90 + 6,40 + 7,80 + 9,06 + 9,80 + 10 + 9,58 + 8,70 + 7,86 + 6,09) = 7,92$.

Теоретичні значення ознаки Y :

$$\bar{y}_5 = -0,239 \cdot 5^2 + 4,762 \cdot 5 - 13,771 = 4,06;$$

$$\bar{y}_6 = -0,239 \cdot 6^2 + 4,762 \cdot 6 - 13,771 = 6,20;$$

$$\bar{y}_7 = -0,239 \cdot 7^2 + 4,762 \cdot 7 - 13,771 = 7,85;$$

$$\bar{y}_8 = -0,239 \cdot 8^2 + 4,762 \cdot 8 - 13,771 = 9,03;$$

$$\bar{y}_9 = -0,239 \cdot 9^2 + 4,762 \cdot 9 - 13,771 = 9,73;$$

$$\bar{y}_{10} = -0,239 \cdot 10^2 + 4,762 \cdot 10 - 13,771 = 9,95;$$

$$\bar{y}_{11} = -0,239 \cdot 11^2 + 4,762 \cdot 11 - 13,771 = 9,69;$$

$$\bar{y}_{12} = -0,239 \cdot 12^2 + 4,762 \cdot 12 - 13,771 = 8,96;$$

$$\bar{y}_{13} = -0,239 \cdot 13^2 + 4,762 \cdot 13 - 13,771 = 7,74;$$

$$\bar{y}_{14} = -0,239 \cdot 14^2 + 4,762 \cdot 14 - 13,771 = 6,05.$$

Квадрати відхилень:

$$\begin{aligned} \sum(\bar{y}_x - \bar{y})^2 &= (4,06 - 7,92)^2 + (6,20 - 7,92)^2 + (7,85 - 7,92)^2 + \\ &+ (9,03 - 7,92)^2 + (9,73 - 7,92)^2 + (9,95 - 7,92)^2 + \\ &(9,69 - 7,92)^2 + (8,96 - 7,92)^2 + (7,74 - 7,92)^2 + (6,05 - 7,92)^2 = \\ &= 14,8996 + 2,9584 + 0,0049 + 1,2321 + 3,2761 + 4,1209 + 3,1329 + 1,0816 \\ &+ 0,0324 + 3,4969 = 33,2358. \end{aligned}$$

$$\sum(y_i - \bar{y})^2 = (3,90 - 7,92)^2 + (6,40 - 7,92)^2 + (7,80 - 7,92)^2 +$$

$$\begin{aligned}
& +(9,06 - 7,92)^2 + (9,80 - 7,92)^2 + (10 - 7,92)^2 + \\
& (9,58 - 7,92)^2 + (8,7 - 7,92)^2 + (7,86 - 7,92)^2 + (6,09 - 7,92)^2 = \\
& = 16,1604 + 2,3104 + 0,0144 + 1,2996 + 3,5344 + 4,3264 + 2,7556 + 0,6084 \\
& + 0,0036 + 3,3489 = 34,3621.
\end{aligned}$$

За формулою (7.10) одержимо:

$$\eta_{yx} = \sqrt{\frac{33,2358}{34,3621}} = 0,983.$$

Значення емпіричного кореляційного відношення близьке до одиниці, що свідчить про високу щільність точок кореляційного поля відносно лінії регресії, яка описується одержаним рівнянням. А з цього слідує, що одержане рівняння регресії досить точно відображає заданий кореляційний зв'язок між ознаками X та Y .

7.7. Непараметричні методи вивчення зв'язків

Розглянуті вище методи вимірювання взаємозв'язків між ознаками називають *параметричними*, оскільки вони базуються на використанні середніх величин і дисперсій, які є основними параметрами розподілу. Зрозуміло, що параметричні методи не можна застосовувати, якщо ознаки не піддаються кількісному виміру (є атрибутивними) або не виконується припущення про нормальний розподіл результативної ознаки (як кількісної так і якісної) для сукупностей незначного обсягу. В таких випадках застосовують *непараметричні методи дослідження взаємозв'язків*, які:

- не вимагають числового вираження значень ознак;
- не вимагають обчислення параметрів розподілу;
- не вимагають інформації про розподіл ознак в сукупності.

Але непараметричні методи забезпечують лише оцінку щільності зв'язку та перевірку його істотності і не дають змогу представити зв'язок аналітично.

В основі обчислення щільності зв'язку між атрибутивними ознаками лежить побудова таблиць співзалежності (взаємного спряження), у яких

представлені комбінаційні розподіли сукупностей за факторною ознакою – по рядках, та результативною – по графах. Найбільш поширеними є таблиці розмірності 2 x 2 (рис. 7.1).

Ознака	A	He A	$\sum B$
B	a	b	a + b
He B	c	d	c + d
$\sum A$	a + c	b + d	a + b + c + d

Рис. 7.1. Схема таблиці для розрахунку коефіцієнтів асоціації та контингенції

Для вимірювання щільності зв'язку між двома альтернативними ознаками використовують коефіцієнт асоціації K_A та коефіцієнт контингенції K_K :

$$K_A = \frac{ad - bc}{ad + bc}; \quad (7.12)$$

$$K_K = \frac{ad - bc}{\sqrt{(a + b)(b + d)(a + c)(c + d)}}. \quad (7.13)$$

Зв'язок вважається підтвердженим, якщо $K_A \geq 0,5$ або $K_K \geq 0,3$. Наприклад, проаналізуємо успішність студентів залежно від статі, виділивши дві групи: студенти, що склали іспит, та студенти, що не склали іспит (табл. 7.5)

Таблиця 7.5

До розрахунку коефіцієнтів асоціації та контингенції

Стать	Кількість студентів		Разом:
	Склали	Не склали	
Дівчата	a = 25	b = 2	a + b = 27
Хлопці	c = 20	d = 3	c + d = 23
Разом	a + c = 45	b + d = 5	50

Використовуючи дані табл. 7.5, обчислимо коефіцієнти асоціації та контингенції за формулами (7.12) і (7.13):

$$K_A = \frac{25 \cdot 3 - 2 \cdot 20}{25 \cdot 3 + 2 \cdot 20} = 0,3043;$$

$$K_K = \frac{25 \cdot 3 - 2 \cdot 20}{\sqrt{27 \cdot 5 \cdot 45 \cdot 23}} = \frac{35}{373,7981} = 0,0936.$$

Оскільки $K_A = 0,3043 < 0,5$ і $K_K = 0,0936$, то між статтю та успішністю студентів зв'язок надто незначний, практично відсутній. Отриманий висновок можна вважати справедливим, тому, що істотними факторами успішності є не стать, а здібності, зацікавленість, організованість, відвідування занять, кількість годин самостійної роботи тощо.

Коли кожна з якісних ознак складається більше ніж із двох груп (неквадратні таблиці), то для визначення щільності зв'язку застосовують коефіцієнт взаємної спряженості Пірсона – Чупрова (C), який теж приймає значення від 0 до 1.

Якщо одна із взаємопов'язаних ознак має кількісний вираз, а друга – альтернативний, то показником щільності є бісеріальний коефіцієнт кореляції (бісерія означає дві серії).

Наприклад, залежність рівня доходів (кількісна ознака) від рівня освіти (атрибутивна).

Оцінка всіх непараметричних показників здійснюється через t -критерій Стьюдента.

Для обробки даних соціологічних досліджень (анкет), медичних обстежень, рейтингів, експертних оцінок тощо, тобто там, де ознаки вимірюються за допомогою номінальної та порядкової шкали (наприклад, “стать”, “соціоекономічний статус сім'ї”, “діагноз” тощо) часто застосовують методи рангової кореляції.

В основу непараметричних методів рангової кореляції покладено принцип нумерації значень статистичного ряду.

Кожній одиниці сукупності надається порядковий номер за величиною значення окремої ознаки – ранг (натуральне число 1, 2, 3, ...). Ранжування, тобто процедура упорядкування об'єктів вивчення на основі віддавання переваг, проводиться за кожною ознакою окремо. При ранжуванні значень факторної і результативної ознак слід використовувати один принцип – або від менших значень до більших, або від більших до менших! Кількість рангів дорівнює обсягу сукупності. Зі збільшенням обсягу ступінь “розпізнаваності”

елементів зменшується, тому рангові оцінки щільності зв'язку доцільно використовувати для сукупностей невеликого обсягу.

До рангових оцінок щільності належать *коефіцієнт кореляції рангів Спірмена* (ρ) – базується на основі різниці рангів (d) факторної і результативної ознак для кожної одиниці сукупності та *Кендала* (τ). Ці коефіцієнти мають ті самі властивості, що і лінійний коефіцієнт кореляції – змінюються в тих же межах, оцінюють щільність зв'язку та вказують його напрям. Зв'язок між ознаками можна визнати статистично істотним, якщо значення коефіцієнтів рангової кореляції Спірмена і Кендала більше за 0,5.

Наприклад, експерти оцінили технічний та фінансовий стан семи підприємств видавничої галузі в балах (табл. 7.6) за певними критеріями: 1-й ранг найменшому значенню ознаки, останній – найбільшому (умовні дані).

Таблиця 7.6

До розрахунку коефіцієнта кореляції рангів Спірмена

№ підприємства	Експертні оцінки		Ранги		$d = R_x - R_y$	d^2
	Технічний стан (x)	Фінансовий стан (y)	R_x	R_y		
1	27	26	1	2	-1	1
2	30	30	2	1	1	1
3	38	30	6	4	2	4
4	36	32	5	5	0	0
5	33	28	3	3	0	0
6	42	37	7	7	0	0
7	35	33	4	6	-2	4
Разом:	-	-			-	10

Коефіцієнт кореляції рангів Спірмена буде:

$$P = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 10}{7 \cdot 48} = 0,8214.$$

Оскільки $P = 0,8214 > 0,5$, то це свідчить про наявність прямого зв'язку між технічним і фінансовим станом підприємств видавничої спрямованості.

7.8. Використання кореляційно-регресійного аналізу в соціально-економічних дослідженнях

Соціально-економічні дослідження проводяться з метою виявлення закономірностей, що мають місце в різних сферах людської діяльності. Оскільки такі закономірності досить часто неочевидні, то завдання статистики полягає в тому, щоб на основі зібраного і систематизованого статистичного матеріалу зробити узагальнюючі висновки про поведінку, закономірності і тенденції розвитку того чи іншого явища.

Якщо засобами кореляційно-регресійного аналізу вдається дослідити явище або процес і внаслідок цього описати отримані закономірності за допомогою рівнянь регресії, то такі рівняння є важливим інструментом у передбачуванні закономірностей та їх використанні при плануванні і прогнозуванні соціально-економічних процесів. На основі одержаного рівняння регресії можна простежити закономірності і тенденції, що відбувались у минулому і які можна спрогнозувати на найближчу перспективу. Для цього використовують *методи інтерполяції та екстраполяції*.

Метод інтерполяції дозволяє теоретично доповнити результати проведених дослідів на основі рівняння регресії, тобто за допомогою одержаного рівняння регресії можна обчислити деякі проміжні значення результативної ознаки відносно факторної, якщо такі значення відомі в деяких вузлових точках. Наприклад, при закладанні польових дослідів під час періоду вегетації проводяться заміри тих чи інших характеристик досліджуваних рослин та умов вегетації через певні інтервали (частіше всього через декілька діб). Якщо дослідника цікавлять проміжні результати, які не були охоплені дослідом, то їх можна теоретично розрахувати за допомогою одержаного рівняння регресії для будь-якої доби вегетаційного періоду.

Метод екстраполяції дає можливість прогнозувати виявлену закономірність, що мала місце у минулому, на найближче майбутнє, що надзвичайно важливо для планування розвитку соціально-економічних процесів. Такий прогноз також відбувається на підставі рівнянь регресії, що

описують досліджувану закономірність. При прогнозуванні керуються принципом: якщо закономірність чи тенденція, які були виявлені в минулому, збережуться і надалі, то можна теоретично розрахувати необхідні показники процесу на перспективу, яка може бути найближчою, а при великомасштабних дослідженнях – на більш віддалену.

Наведемо приклад прогнозування народжуваності дітей на сільських територіях Хмельницької області в залежності від деяких факторів, виконаного за допомогою одержаних лінійних рівнянь регресії, маючи на увазі середні базисні дані факторних і результативного показників з розрахунку на 10000 осіб населення ($\hat{Y} = 10,80; X_1 = 30,73; X_2 = 11,47; X_3 = 256,54$) станом на кінець 2012 року (табл. 7.7).

Таблиця 7.7

Прогнозування динаміки народжуваності дітей на сільських територіях Хмельницької області

Рівняння регресії, що виражають залежність народжуваності дітей від факторних показників	Факторні показники	Прогнозовані значення народжуваності дітей при зростанні факторних показників на:				
		5%	10%	15%	20%	25%
$\hat{Y} = 0,7341X_1 - 11,7570$	Забезпеченість населення житлом X_1 (м ² /особу)	12,14	13,28	14,41	15,55	16,69
$\hat{Y} = 1,4881X_2 - 5,7776$	Забезпеченість дітей дошкільного віку освітніми закладами X_2 (на 10 тис. осіб населення)	12,14	13,00	13,85	14,70	15,56
$\hat{Y} = 0,0571X_3 - 3,8482$	Забезпеченість дітей дошкільного віку місцями у дошкільних освітніх закладах X_3 (на 10 тис. осіб населення)	11,53	12,27	13,00	13,73	14,46

Прогнозовані результати таблиці 7.7 показують, що найбільший позитивний вплив на народжуваність дітей має забезпеченість населення

житлом, тому ця проблема, поряд з іншими, вимагає першочергового вирішення.

Питання для повторення і самоконтролю

1. З якою метою досліджуються статистичні взаємозв'язки між явищами?
2. Які ознаки сукупності називаються факторними і яка результативною?
3. Які зв'язки між ознаками називаються функціональними, статистичними і кореляційними?
4. Які зв'язки між ознаками називаються прямими та оберненими, прямолінійними та криволінійними, однофакторними і багатофакторними?
5. В чому полягає головне призначення дисперсійного аналізу?
6. Яку основну тотожність між дисперсіями покладено в основу дисперсійного аналізу?
7. Як оцінюється вплив групових сукупностей на загальну сукупність за допомогою критерію Фішера-Снедекора?
8. В чому полягає суть методу аналітичних групувань і для чого він використовується?
9. Як графічно зображується кореляційна залежність? Що таке кореляційне поле та емпірична лінія регресії?
10. Що таке парна лінійна кореляція і парна лінійна регресія? Який загальний вигляд має рівняння регресії в цьому випадку?
11. В чому полягає суть методу найменших квадратів?
12. Як обчислюється коефіцієнт лінійної кореляції і для чого він використовується?
13. Що характеризує коефіцієнт детермінації і як він обчислюється?
14. Який загальний вигляд має багатофакторне лінійне рівняння та який зміст коефіцієнтів, що входять до нього?
15. Яка кореляція називається нелінійною та які її види Ви знаєте?
16. Який вигляд має гіперболічне рівняння регресії та як можна обчислити його невідомі коефіцієнти?

17. Яким рівнянням описується параболічна (квадратична) кореляція та як обчислюються його коефіцієнти?

18. Що таке кореляційне відношення, які воно має властивості та як обчислюється?

19. Які методи дослідження взаємозв'язків називаються непараметричними і які в них переваги перед параметричними методами?

20. Як обчислюються коефіцієнти асоціації та контингенції і яка їхня роль?

21. Як обчислюється коефіцієнт кореляції рангів Спірмена і для чого він використовується?

22. Яке значення має кореляційно-регресійний аналіз в соціально-економічних дослідженнях?

Лекція 9. Аналіз інтенсивності динаміки

План

9.1. Динамічний ряд як база аналізу і прогнозування соціально-економічного розвитку. Види рядів динаміки.

9.2. Статистичні характеристики динамічних рядів, їх взаємозв'язок.

9.3. Методи обчислення середнього рівня динамічного ряду.

9.4. Економічна суть та техніка розрахунку середніх значень основних характеристик рядів динаміки.

9.1. Динамічний ряд як база аналізу і прогнозування соціально-економічного розвитку. Види рядів динаміки

Для проведення аналізу соціально-економічних явищ та прогнозування їх розвитку на перспективу необхідно систематизувати досліджуваний статистичний матеріал, тобто створити необхідну базу для вивчення динаміки цих явищ.

Динамікою в статистиці прийнято називати процес розвитку, руху соціально-економічних явищ у часі. Для відображення і аналізу динаміки будують динамічні (хронологічні, часові) ряди. Дослідження динаміки дає

змогу охарактеризувати процес розвитку явищ, розкрити основні шляхи, тенденції і темпи цього розвитку.

Ряд динаміки – це низка статистичних показників, які характеризують зміну суспільних явищ у часі.

Кожний ряд динаміки складається з двох обов'язкових елементів: 1) періодів часу (t); 2) рівнів (y).

Показниками часу в рядах динаміки можуть бути:

- певні дати (моменти) часу;
- окремі періоди (роки, квартали, місяці, декади, доба).

Рівнем ряду динаміки називають статистичний показник, який характеризує величину суспільного явища на даний момент або за певний період часу. Вони відображають кількісну оцінку (міру) розвитку досліджуваного суспільного явища.

Залежно від характеру рівнів ряду розрізняють два види рядів динаміки: *інтервальні та моментні.*

Моментними називають ряди динаміки, які характеризують розмір явища станом на певний момент часу.

За допомогою моментних рядів динаміки характеризується найчастіше стан умов і факторів виробництва.

Інтервальними називають ряди динаміки, які характеризують розмір явищ за певний період часу.

За допомогою інтервальних динамічних рядів характеризуються підсумки виробничого процесу (обсяги виробленої продукції, виконаних робіт, затрат праці, кількості внесених добрив тощо).

За кількістю показників, що змінюються, ряди динаміки бувають *одновимірні та багатовимірні.*

Якщо ряд динаміки характеризує зміну одного показника, то його називають *одномірним*, двох і більше – *багатомірним*. *Багатомірні ряди бувають двох видів: паралельні ряди та ряди взаємозв'язаних показників.*

Паралельні характеризують динаміку або одного показника щодо різних об'єктів, або різних показників щодо одного об'єкта. Зв'язок між показниками багатомірного динамічного ряду може бути функціональним або кореляційним.

За повнотою часу динамічні ряди поділяються на *повні й неповні*.

У повних динамічних рядах дати або періоди ідуть один за одним з рівними інтервалами.

У неповних динамічних рядах у послідовності показників спостерігають нерівні часові інтервали.

За способом вираження рівнів динамічні ряди поділяються на *ряди абсолютних, відносних та середніх величин*.

У динамічних рядах потрібно дотримуватися порівнянності всіх рядів між собою за територією, методикою розрахунку показників, періодом, або моментом часу, об'єктом і одиницею спостереження, ступенем охоплення одиниць досліджуваної сукупності, одиницями вимірювання.

Непорівнянність даних, що виникає внаслідок адміністративно-територіальних змін, часто виявляється в статистичній практиці. Це зумовлено тим, що межі територій господарств, районів, областей протягом досліджуваного періоду змінюються внаслідок приєднання до них нових територій або від'єднання окремих частин їхньої території. Для зведення даних до порівнянного виду необхідно виконати перерахунок даних за попередні роки (до зміни території) з урахуванням нових меж.

Найбільш суттєвою вимогою при побудові ряду динаміки є єдина методика обчислення рівня за кожний з періодів, що розглядається. Завдяки цьому забезпечується порівнянність статистичних показників за змістом.

Наприклад, при вивченні динаміки урожайності сільськогосподарських культур показник урожайності повинен відноситись до однієї і тієї ж посівної площі (весняної продуктивної, фактично зібраної тощо). При дослідженні динаміки вартісних показників обсягу продукції необхідно усунути вплив зміни цін. На практиці для вирішення цього завдання кількість продукції, виробленої в різні періоди, оцінюють в цінах одного періоду, які називають фіксованими,

або порівнянними. Якщо ряд динаміки подано узагальнюючими показниками в умовно-натуральних одиницях вимірювання, коефіцієнти сумірності для всіх рівнів мають бути єдиними.

Порівнянність рівнів ряду динаміки за періодом, або моментом спостереження, означає, що:

1) всі показники характеризують явище або за певний період часу, або на певний момент часу. В зв'язку з цим неправомірно порівнювати, наприклад, середньорічне число тракторів з числом тракторів на початок або кінець року;

2) в інтервальних динамічних рядах рівні повинні відноситись до однакових періодів часу, а в моментних – повинні бути, як правило, рівні відрізки часу між моментами (датами) спостереження. Крім того, не можна поєднувати в одному ряду динаміки періоди і моменти часу.

Порівнянність за об'єктом спостереження означає, що всі рівні ряду динаміки відносяться до одного і того самого об'єкта спостереження. Наприклад, при дослідженні динаміки продуктивності корів об'єктом спостереження можуть бути державні, колективні, фермерські господарства, особисті підсобні господарства населення або всі категорії в цілому. Для одержання порівнянної в динаміці продуктивності корів показник повинен розраховуватись по одній і тій самій категорії господарства або по їх сукупності.

Порівнянність за одиницями спостереження передбачає, що всі рівні одержані по одних і тих самих одиницях спостереження. Одиницями спостереження можуть бути окремі підприємства або їхні підрозділи. Наприклад, при вивченні динаміки урожайності сільськогосподарських культур показник урожайності повинен визначитися по одних і тих самих сільськогосподарських підприємствах, фермах тощо.

Крім перелічених вимог, без урахування яких неможливо побудувати ряд динаміки, потрібно дотримуватися одних і тих самих одиниць вимірювання. Так, якщо дані про валовий збір за одні роки наводяться в тонах, а за інші – в

центнерах, то необхідно перерахувати весь ряд в одні і ті самі одиниці вимірювання.

Науково обґрунтоване формування рядів динаміки вимагає також виділення строго однорідних періодів (етапів) у розвитку досліджуваних соціально-економічних явищ, тому що всебічного аналізу динамічних процесів можна досягти лише в межах однорідних періодів. Періодизацію динамічних рядів слід проводити на основі глибокого теоретичного аналізу основних процесів і законів, які визначають розвиток досліджуваного явища.

Особливості періодичного ряду по відношенню до моментного:

1. Рівні інтервального ряду можна додавати, отримуючи при цьому новий ряд, але за довший відрізок часу. Цю особливість використовують в статистичній звітності.

2. Рівні інтервального ряду залежать від величини періоду або відрізка часу.

9.2. Статистичні характеристики динамічних рядів, їх взаємозв'язок

Для опису рядів динаміки визначають абсолютні та відносні характеристики динаміки: *абсолютний приріст та абсолютне значення 1% приросту; темп зростання та темп приросту.*

Обчислення характеристик ґрунтується на співставленні рівнів ряду.

Якщо базою при цьому є попередній рівень, то така характеристика називається *змінною*, а самі характеристики – *ланцюговими*. Якщо ж за базу обраний початковий рівень U_0 , то базу називають *постійною*, а характеристики – *базисними*.

Ланцюгові характеристики: $\Delta_t = y_t - y_{t-1}$; , базисні: $\Delta_t = y_t - y_0$.

Абсолютний приріст характеризує розмір збільшення (зменшення) рівня ряду за певний період. Обчислюється як різниця між поточним і базисним рівнями, щоб знайти на скільки одиниць підвищився чи зменшився рівень щодо базисного за певний період часу.

Абсолютний приріст виражає абсолютну швидкість зміни рівнів ряду динаміки. Для вичерпної та глибокої характеристики явища абсолютні величини доповнюють відносними.

Абсолютний приріст може мати додатний або від'ємний знак. Якщо наступний рівень ряду динаміки більший за попередній, то абсолютний приріст буде мати знак «+», якщо менше – знак «-».

Темп (коефіцієнт) зростання K_t обчислюється як відношення зіставлюваного рівня з рівнем, прийнятим за базу зіставлення, і показує, в скільки разів (процентів) порівнюваний рівень більший чи менший від базисного.

$$\text{Ланцюгові: } K_t = \frac{y_t}{y_{t-1}}; \quad \text{базисні: } K_t = \frac{y_t}{y_0}.$$

Темп зростання може бути виражений у вигляді коефіцієнтів або процентів. Темп зростання, виражений у процентах, називають *відсотком зростання*.

Величина темпу зростання буде більшою від одиниці, якщо рівень щодо бази порівняння зростає, і меншою за одиницю, якщо рівень щодо бази порівняння зменшується.

Між ланцюговими і базисними темпами зростання існує певний взаємозв'язок:

а) добуток ланцюгових темпів зростання дорівнює базисному темпу зростання за відповідний період;

б) частка від ділення двох сусідніх базисних коефіцієнтів зростання дорівнює відповідному ланцюговому коефіцієнту зростання.

Взаємозв'язок ланцюгових та базисних темпів зростання використовують для переходу від одних темпів зростання до інших у тих випадках, коли невідомі абсолютні рівні ряду динаміки.

Поряд з темпами зростання відносна зміна явища у часі може бути також охарактеризована за допомогою *темлів приросту* (T).

Темп приросту – це відношення абсолютного приросту до абсолютного попереднього або початкового рівня і показує, на скільки відсотків порівнюваний рівень більший або менший від рівня, взятого за базу порівняння:

а) ланцюговий: $T = \frac{\Delta_t}{y_{i-1}}$; б) базисний: $T = \frac{\Delta_t}{y_0}$.

Темп приросту, як і абсолютний приріст, може бути як додатним, так і від'ємним числом (відповідно при зростанні і зниженні рівня) і виражається у вигляді коефіцієнтів або відсотків. На практиці темпи приросту найчастіше виражаються у формі відсотків.

Вони показують, на скільки відсотків збільшився або зменшився поточний рівень порівняно з базисним, взятим за 100%.

Між темпом зростання і темпом приросту існує зв'язок:

$$T_i = K_i - 1 \text{ або } T_i = (K_i \times 100\%) - 100\%,$$

тобто темп приросту завжди на одиницю менший від відповідного темпу зростання, вираженого у формі коефіцієнта, або на 100%, якщо його виражено у відсотках.

Абсолютне значення одного відсотка приросту $A\%$ можна знайти шляхом ділення абсолютного приросту на темп приросту за один і той самий період. Абсолютне значення одного відсотка приросту можна обчислити простішим способом — діленням початкового рівня на 100, оскільки за 100 % завжди беруть базисний рівень, то 1 % буде в 100 разів менший від базисного рівня: $A\% = \frac{\Delta_t}{T}$, або $A\% = \frac{y_{t-1}}{100}$.

Цей показник дає змогу визначити вагомість кожного відсотка приросту, і те, яка абсолютна величина приховується за цим відсотком.

Розрахунок цього показника має економічний зміст тільки на ланцюговій основі, оскільки на базисній основі по всіх часових відрізках буде отримано одне і те саме значення показника – сота частина початкового (першого) рівня.

Приклад. Нехай задано ряд динаміки, що характеризує валовий збір картоплі в області за 2009-2014 роки:

Роки	2009	2010	2011	2012	2013	2014
Валовий збір картоплі, млн. т. (y)	20	23	27	28	30	33

Необхідно розрахувати показники динаміки валового збору.

Розв'язання. Всі обчислення виконаємо за наведеними вище формулами і їх результати занесемо до таблиці (табл. 9.1).

Таблиця 9.1

Розрахунок показників валового збору картоплі в області за 2009-2014 рр.

Роки	Валовий збір картоплі, млн. т. (y)	Абсолютний приріст, млн. грн		Темп росту		Темп приросту		Абсолютне значення 1% приросту, млн. т.
		Базисний	Ланцюговий	Базисний	Ланцюговий	Базисний	Ланцюговий	
2009	20	-	-	1,000	-	-	-	-
2010	23	3	3	1,150	1,150	0,150	0,150	0,20
2011	27	7	4	1,350	1,174	0,350	0,174	0,23
2012	28	8	1	1,400	1,037	0,400	0,037	0,27
2013	30	10	2	1,500	1,074	0,500	0,074	0,28
2014	33	13	3	1,650	1,100	0,650	0,100	0,30

Ряди динаміки задаються не лише за допомогою таблиць, а зображуються також і графіками. В нашому прикладі такий графік має вигляд, показаний на рис. 9.1.

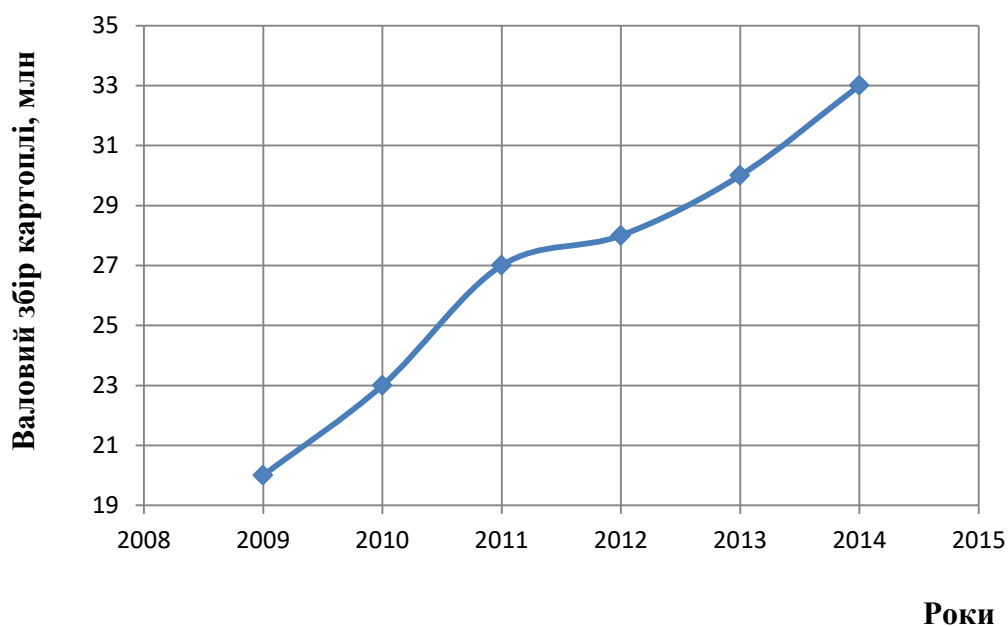


Рис. 9.1. Динаміка валового збору картоплі в області за 2009-2014 рр.

9.3. Методи обчислення середнього рівня динамічного ряду

Динамічні ряди складаються з багатьох варіаційних рівнів, а тому, як будь-яка статистична сукупність, вони потребують деяких узагальнюючих характеристик. Для цього обчислюють середні показники: *середні рівні ряду, середні абсолютні прирости, середні темпи зростання і приросту.*

Методи обчислення середнього рівня інтервального і моментного рядів динаміки залежать від їхнього вигляду.

В інтервальних рядах з рівними інтервалами *середній рівень ряду обчислюють за формулою середньої арифметичної простої*:

$$\bar{y} = \frac{\sum y_i}{n}, \quad (9.1)$$

де y_i – рівні ряду; n – кількість рівнів ряду.

Якщо інтервальний ряд динаміки має нерівновіддалені один від одного рівні, то *середній рівень розраховується за формулою середньої арифметичної зваженої*:

$$\bar{y} = \frac{\sum y_i t}{t}, \quad (9.2)$$

де t – відрізок часу, протягом якого зберігалось дане значення рівня y .

У моментному ряду з рівними відрізками часу *середній рівень обчислюється як середня хронологічна*:

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + \dots + y_{n-1} + \frac{y_n}{2}}{n-1}. \quad (9.3)$$

У знаменнику від числа рівнів віднімають одиницю, оскільки в чисельнику серед доданків першій і останній рівні беруть в половинному розмірі.

Якщо окремі періоди інтервального ряду динаміки мають різну довжину, то для визначення середнього рівня використовують середню арифметичну зважену:

$$\bar{y} = \frac{\sum y_i t}{\sum t}, \quad (9.4)$$

де y_i – рівні ряду.

Середній абсолютний приріст визначають як середню арифметичну просту з ланцюгових абсолютних приростів за певні періоди, знаходять, на скільки одиниць у середньому змінився рівень порівняно з попереднім.

$$\bar{\Delta} = \frac{\sum \Delta_t}{n}. \quad (9.5)$$

Для узагальнюючої характеристики темпів зростання за ряд років обчислюють *середній темп (коефіцієнт) зростання \bar{K}* . Він показує, в скільки

разів у середньому кожен даний рівень ряду більший (або менший) від попереднього рівня. Для динамічних рядів з рівними проміжками між датами середній темп зростання обчислюється за формулою середньої геометричної:

$$\bar{K} = \sqrt[n]{K_1 \cdot K_2 \cdot K_3 \cdot \dots \cdot K_n}, \quad (9.6)$$

де K_i – коефіцієнти зростання за окремі періоди часу;

n – кількість ланцюгових темпів зростання.

9.4. Економічна суть та техніка розрахунку середніх значень основних характеристик рядів динаміки

Використовуючи формули попереднього пункту, наведемо приклади розрахунку середніх значень основних характеристик рядів динаміки, які мають конкретний економічний зміст.

Приклад 1. Нехай задано чисельність працівників підприємства у січні місяці 2016 року, чол.:

На 1.01	На 6.01	На 15.01	На 21.01	На 29.01	На 1.02
1210	1243	1236	1248	1238	1238

Необхідно обчислити середньоспискову чисельність працівників підприємства в січні місяці.

Р о з в ‘ я з а н н я. Тут ми маємо інтервальний динамічний ряд, у якому періоди часу неоднакові, тому для розрахунку використаємо середню арифметичну зважену, яка обчислюється за формулою (9.2):

$$\bar{y} = \frac{1210 \cdot 5 + 1243 \cdot 9 + 1236 \cdot 6 + 1248 \cdot 8 + 1238 \cdot 3}{31} = 1237 \text{ чол.}$$

Приклад 2. Нехай парк тракторів у фермерських господарствах району характеризується таким динамічним рядом:

Дані на початок місяця	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.09	1.10	1.11	1.12	1.01
Кількість тракторів, шт.	622	640	643	640	664	670	682	733	753	768	800	826	888

Необхідно визначити середнє число тракторів за кожний квартал, перше і друге півріччя і за рік в цілому.

Розв'язання. Для розв'язання задачі обчислимо у всіх випадках значення середньої хронологічної за формулою (9.3):

$$\overline{y}_{1\text{кв}} = \frac{\frac{622}{2} + 640 + 643 + \frac{640}{2}}{3} = 638 \text{ шт.};$$

$$\overline{y}_{2\text{кв}} = \frac{\frac{640}{2} + 664 + 670 + \frac{668}{2}}{3} = 665 \text{ шт.};$$

$$\overline{y}_{3\text{кв}} = \frac{\frac{682}{2} + 733 + 753 + \frac{768}{2}}{3} = 737 \text{ шт.};$$

$$\overline{y}_{4\text{кв}} = \frac{\frac{768}{2} + 800 + 826 + \frac{888}{2}}{3} = 818 \text{ шт.};$$

$$\overline{y}_{1\text{пів}} = \frac{\overline{y}_{1\text{кв}} + \overline{y}_{2\text{кв}}}{2} = 651,5 \text{ шт.};$$

$$\overline{y}_{2\text{пів}} = \frac{\overline{y}_{3\text{кв}} + \overline{y}_{4\text{кв}}}{2} = 777,5 \text{ шт.};$$

$$\overline{y}_p = \frac{\overline{y}_{1\text{пів}} + \overline{y}_{2\text{пів}}}{2} = \frac{651,5 + 777,5}{2} = 714,5 \text{ шт.}$$

Приклад 3. Нехай задано ряд динаміки, що характеризує валовий збір картоплі в області за 2009-2014 роки (приклад в пункті 9.2). Обчислити середньорічний абсолютний приріст валового збору картоплі за вказаний період часу.

Розв'язання. За формулою (9.5) маємо:

$$\overline{\Delta}_y = \frac{3+4+1+2+3}{5} = 2,6 \text{ млн т.}$$

$$\text{або: } \overline{\Delta}_y = \frac{y_n - y_1}{n-1} = \frac{33-20}{6-1} = 2,6 \text{ млн т.}$$

Приклад 4. Обчислити середній темп приросту валового збору картоплі в області за 2009-2014 роки за даними прикладу в п. 9.2.

Розв'язання. Використовуючи результати розв'язання прикладу в пункті 9.3, за формулою (9.6) одержимо:

$$\begin{aligned} \overline{K} &= \sqrt[n]{K_1 \cdot K_2 \cdot K_3 \cdot \dots \cdot K_n} = \sqrt[5]{0,150 \cdot 0,174 \cdot 0,037 \cdot 0,074 \cdot 0,100} = \\ &= \sqrt[5]{0,00000714618} = 0,093. \end{aligned}$$

Питання для повторення і самоконтролю

1. Що характеризує собою ряд динаміки?
2. Які ряди динаміки називають моментними і які інтервальними?

3. Що означає порівнянність динамічних рядів за періодом або моментом спостереження?
4. Чим відрізняється моментний ряд від інтервального? Чому рівні моментного ряду не можна додавати?
5. Які характеристики динамічного ряду називаються базисними і які ланцюговими та як їх обчислювати?
6. Що таке абсолютний приріст, темп зростання і темп приросту та як вони обчислюються?
7. Який взаємозв'язок існує між базисними і ланцюговими темпами зростання?
8. Який зв'язок існує між темпом зростання і темпом приросту?
9. Що таке абсолютна величина одного відсотка приросту та який її економічний зміст?
10. Як обчислюється середній рівень ряду в інтервальних динамічних рядах?
11. Як обчислити середній рівень моментного ряду?
12. Як визначити середній абсолютний приріст і середній темп зростання?
13. У чому полягає відмінність базисного і ланцюгового способів обчислення показників динаміки?

Лекція 10. Аналіз тенденцій розвитку та коливань

План

- 10.1. Виявлення тенденцій розвитку явищ. Трендова компонента, її аналіз.
- 10.2. Періодична компонента. Статистичний аналіз сезонних коливань.
- 10.3. Факторний аналіз рядів динаміки.
- 10.4. Інтерполяція та екстраполяція. Прогнозування на основі рядів динаміки.

10.1. Виявлення тенденцій розвитку явищ. Трендова компонента, її аналіз

Одним із основних завдань аналізу рядів динаміки є *виявлення основної тенденції (тренду) розвитку соціально – економічних явищ*. Під загальною тенденцією динамічного ряду розуміють тенденцію до зростання, зниження або стабілізації рівня будь-якого суспільного явища.

Виявлення тенденції в динамічних рядах дає змогу оцінити характер розвитку досліджуваного явища, визначити ефективність факторів, що формують основну тенденцію, передбачити рівні досліджуваного явища на перспективу.

Виявлення основної тенденції зміни рівнів динамічного ряду передбачає її кількісне вираження, деякою мірою, вільною від випадкових причин. Це досягається шляхом абстрагування від індивідуальних, випадкових змін ознаки. Виявлення основної тенденції розвитку називається у статистиці також *вирівнюванням часового ряду*, а прийоми виявлення основної тенденції – *прийомами вирівнювання*. Вирівнювання дає змогу охарактеризувати особливості зміни у часі даного динамічного ряду в найбільш загальному вигляді як функцію часу, передбачаючи, що через час можна виразити вплив основних факторів.

В таких випадках для виявлення основної тенденції потрібно вдаватися до спеціальних прийомів обробки динамічних рядів.

До таких прийомів відносяться *укрупнення періодів, згладжування ряду динаміки способом ковзної середньої, вирівнювання ряду динаміки за середнім абсолютним приростом, середнім коефіцієнтом зростання і способом найменших квадратів (аналітичне вирівнювання рядів динаміки)*.

Одним із найпростіших прийомів виявлення тенденції розвитку є прийом *укрупнення періодів*. Суть його полягає в тому, що абсолютні або середні рівні ряду динаміки за короткі інтервали (рік, місяць, декаду, день тощо), що зазнають випадкових коливань, змінюють узагальнюючим, зазвичай середнім, значенням за триваліший період (триріччя, п'ятиріччя тощо).

По суті спосіб укрупнення періодів являє собою типологічне групування рівнів ряду динаміки, тому при його застосуванні необхідно дотримуватись наукових основ побудови статистичних групувань.

При укрупненні періодів дуже важливо науково обґрунтовано і правильно виділити періоди часу для укрупнення. Періоди, що їх виділяють, мають бути однорідними в якісному відношенні і досить тривалими за часом, щоб відбулося погашення випадкових коливань явища.

При укрупненні періодів число членів динамічного ряду дуже скорочується. Цей істотний недолік значною мірою усувається при використанні прийому вирівнювання динамічних рядів способом *ковзних середніх*.

Цей спосіб також ґрунтується на укрупненні періодів. Суть розрахунку ковзних середніх полягає в тому, що склад періоду безперервно і постійно змінюється – відбувається зсув на одну дату при збереженні постійного інтервалу періоду (триріччя, п'ятиріччя тощо).

Ковзна середня – це середня укрупнених періодів, створених послідовним виключенням кожного початкового рівня інтервалу і заміни його черговим наступним рівнем ряду. Таким чином, відбувається ніби ковзання періоду і отриманої середньої по динамічному ряду. Наприклад, при згладжуванні по триріччях

$$\bar{y}_1 = \frac{y_0 + y_1 + y_2}{3}; \quad \bar{y}_2 = \frac{y_1 + y_2 + y_3}{3}; \quad \bar{y}_3 = \frac{y_2 + y_3 + y_4}{3} \quad \text{і т. д.}$$

Цей прийом, як і попередній, ґрунтується на відомому теоретичному положенні про те, що в середніх величинах взаємно погашаються випадкові відхилення і виявляється типове, закономірне.

При виявленні тенденції прийомом ковзних середніх, так само, як і при використанні прийому укрупнення періодів, одним з важливих питань є питання щодо тривалості періодів. Інтервал має бути досить великим і забезпечити взаємне погашення випадкових відхилень рівнів. Якщо в розвитку явища має місце циклічність (періодичність), то інтервал ковзання слід брати

рівним тривалості циклу. Чим довший інтервал ковзання, тим більшою мірою вирівнюється ряд в результаті усереднення вихідних рівнів.

В основі застосування способу вирівнювання ряду динаміки за *середнім абсолютним приростом* лежить припущення, що кожен наступний рівень змінюється порівняно з попереднім приблизно на однакову величину, що дорівнює середньому абсолютному приросту.

Рівняння, що відображує тенденцію розвитку явища за цим способом вирівнювання ряду динаміки, є лінійним і має вигляд:

$$\tilde{y}_t = y_0 + \bar{A}t \quad (10.1)$$

де \tilde{y}_t - вирівняні рівні ряду динаміки;

y_0 - початковий рівень ряду динаміки;

\bar{A} - середній абсолютний приріст;

t – порядковий номер дати ($t = 1, 2, 3, \dots, n$).

Водночас необхідно відмітити, що теоретична пряма лінія, яка вирівнює ряд динаміки, цілком залежить тільки від двох крайніх значень рівнів ряду динаміки (початкового і кінцевого), які можуть суттєво змінюватись під впливом випадкових коливань. Відповідно тенденція, яка дійсно має місце в досліджуваному явищі, буде спотворена. В зв'язку з цим прийом вирівнювання рядів динаміки за середнім абсолютним приростом доцільно використовувати для рядів, що мають стабільні щорічні абсолютні прирости. Практично цей прийом використовується в динамічних рядах, які охоплюють нетривалий період часу, протягом якого не відбувається суттєвих якісних змін у рівнях факторів, що визначають тенденцію, і ступенях їх впливу на досліджувану ознаку.

Вирівнювання ряду динаміки за *середнім коефіцієнтом зростання* застосовується в таких випадках, коли в досліджуваному ряду кожен наступний рівень змінюється порівняно з попереднім приблизно в одну й ту саму кількість разів, що дорівнює величині середнього коефіцієнту зростання, тобто коли

фактори, що визначають основну тенденцію, зумовлюють від періоду до періоду однакові коефіцієнти зростання досліджуваного явища.

Вирівняні значення рівнів ряду динаміки визначають за формулою:

$$\tilde{y}_t = y_0 \cdot \bar{k}^t, \quad (10.2)$$

де \bar{k} - середній коефіцієнт зростання.

Досконалішим і точнішим прийомом вирівнювання рядів динаміки, який враховує всі рівні вихідного ряду, є аналітичне вирівнювання за *способом найменших квадратів*.

Вирівнювання за цим способом ґрунтується на припущенні, що зміни досліджуваного ряду динаміки можуть бути наближено виражені певним математичним рівнянням (апроксимуючою функцією), за яким і визначають вирівняні рівні динамічного ряду. Іншими словами, рівні ряду динаміки розглядаються як функція часу $\tilde{y}_t = f(t)$, де \tilde{y}_t - рівні динамічного ряду, визначені за відповідним рівнянням регресії на момент часу t .

Аналітичне вирівнювання можна провести з використанням різних типів функцій: прямої лінії, параболи другого порядку, показникової кривої (експоненти), гіперболи тощо.

Рівняння, що виражає рівні ряду динаміки як деяку функцію часу t , називають трендом. Поняття про рівняння тенденції було введене в статистику англійським вченим Гукером у 1902 р. Саме він запропонував називати таке рівняння *трендом* (*the trend*).

Суть аналітичного вирівнювання динамічних рядів полягає в тому, що фактичні рівні ряду замінюються рядом рівнів, які змінюються плавно (теоретичними рівнями), обчисленими за допомогою рівняння певної кривої, вибраної в припущенні, що вона найточніше відображає загальну тенденцію зміни досліджуваного соціально – економічного явища у часі.

Підбір найбільш придатної функції є важливим і відповідальним завданням, від якого в остаточному підсумку залежать результати вирівнювання. В основі вирішення його має бути змістовний аналітичний

аналіз істотності досліджуваного явища і законів його розвитку. Треба підібрати таку криву, яка б максимально близько наближалась до фактичних рівнів. Добитися цього можна за умови, що сума квадратів відхилень фактичних рівнів (y) від розрахованих за рівнянням (\tilde{y}_t) буде мінімальною $(y - \tilde{y}_t)^2 = \min$.

Наближено обґрунтувати рівняння, що відображує основну тенденцію, можна за допомогою побудови кореляційного поля та емпіричної лінії регресії (див. лекції 7,8).

10.2. Періодична компонента. Статистичний аналіз сезонних коливань

У практиці дослідження динамічних рядів часто доводиться мати справу з аналізом періодичних сезонних коливань рівнів рядів.

Сезонними коливаннями називають *періодичні внутрішньорічні коливання*, зумовлені зміною пори року. Такі коливання спостерігаються в багатьох галузях народного господарства.

Сезонні коливання у статистиці вимірюють за допомогою розрахунку спеціальних показників – *індексів сезонності*. В сукупності ці індекси утворюють *сезонну хвилю*.

Індекс сезонності – це процентне відношення однойменних місячних (квартальних) фактичних рівнів динамічних рядів до їх середньорічних або вирівняних рядів.

Розрахуємо, наприклад, індекс сезонності (сезонну хвилю) реалізації побутових холодильників торговельними підприємствами за 2012-2014 рр. *методом простих середніх*.

Індекс сезонності методом простої середньої обчислюється за формулою:

$$I_s = \frac{\bar{y}_1}{\bar{y}_3} \cdot 100, \quad (10.3)$$

де I_s – індекс сезонності; \bar{y}_t – середні місячні або квартальні рівні; \bar{y}_3 – загальна середня (місячна або квартальна).

Розрахунок сезонної хвилі представлено в табл. 10.1.

Таблиця 10.1

Розрахунок сезонної хвилі реалізації побутових холодильників торговими підприємствами за 2012-2014 рр., шт.

Квартал	Роки			Разом	В середньому \bar{y}_t	Сезонна хвиля, % (I_s)
	2012	2013	2014			
I	1942	2126	2505	6573	2191,00	82,1
II	2957	2704	3704	9365	3121,67	117,0
III	2504	3291	3834	9629	3209,67	120,3
IV	2194	1745	2513	6452	2150,67	80,6
Разом:	9597	9866	12556	32019	$\bar{y}_3 = 2668,25$	400,0

Судячи з табл. 10.1, реалізація побутових холодильників суттєво падає в першому і четвертому кварталах і різко зростає в другому і третьому кварталах року.

При стабільній тенденції в ряді динаміки, в якому внутрішньорічні коливання ознаки відбуваються навколо деякого постійного рівня, показники сезонності визначають як процентне відношення рівнів за кожний місяць до середньомісячного рівня за рік.

Однак місячні рівні за один рік можуть бути нетиповими через вплив випадкових причин. Тому на практиці індекси сезонності визначають за місячними даними за кілька років (три роки і більше). В цьому разі для кожного місяця встановлюють середню величину рівня за кілька років (наприклад, три роки), далі з них розраховують середньомісячний рівень для всього ряду. Після цього кожен середньомісячний рівень порівнюють з середньомісячним річним рівнем за кілька років, а знайдений результат перемножують на сто процентів.

Як характеристики сезонності можуть бути використані показники варіації: середнє квадратичне відхилення і коефіцієнт варіації. Чим менша

величина кожного з цих показників, тим меншою є сезонність досліджуваного явища.

10.3. Факторний аналіз рядів динаміки

Важливе місце у вивченні динаміки соціально-економічних явищ належить *факторному аналізу*, метою якого є дослідження впливу окремих факторів на кількісні і якісні зміни явища в часі. В аналізі динаміки суспільного виробництва важливо насамперед оцінити залежність результативних показників від комплексу економічних і природних факторів.

Для здійснення факторного аналізу рядів динаміки статистика використовує ряд методів і прийомів, серед яких найбільш поширені такі: *приведення рядів динаміки до однієї основи*, *порівняння кількох паралельних рядів результативних і факторних показників* (одного результативного і одного факторного показників, одного результативного і кількох факторних показників, одного факторного і кількох результативних показників), укрупнення періодів, розчленування досліджуваної сукупності на якісно однорідні групи і підгрупи, тобто побудова простих і комбінаційних групувань, застосування дисперсійного і кореляційного методів аналізу та ін.

Застосування деяких з перелічених прийомів факторного аналізу рядів динаміки розглянемо на прикладах. В табл. 10.2 подано динаміку виходу валової продукції сільського господарства та основних факторів у фермерському господарстві за 2005-2010 рр.

Дані таблиці 10.2 свідчать про те, що збільшення виходу валової продукції на 100 га сільськогосподарських угідь (результативний показник) в динаміці у фермерському господарстві супроводжувалося закономірним зростанням факторів інтенсивності сільськогосподарського виробництва (фондозабезпеченості, енергозабезпеченості і кількості внесених мінеральних добрив – факторні показники).

Динаміка виходу валової продукції сільського господарства та основних факторів виробництва у фермерському господарстві

Рік	В розрахунку на 100 га сільськогосподарських угідь			
	Вартість валової продукції сільського господарства, тис. грн	Вартість основних виробничих фондів сільськогосподарського призначення (фондозабезпеченість), тис. грн	Енергетична потужність (енергозабезпеченість), кінських сил	Кількість внесених мінеральних добрив, ц діючої речовини
2005	66,7	103,4	246	87,0
2006	71,4	112,5	257	85,1
2007	75,0	120,4	275	94,4
2008	88,9	135,8	291	113,5
2009	84,3	146,7	327	110,8
2010	90,9	159,0	362	134,6

Для проведення порівняльного аналізу чотирьох рядів динаміки і виявлення більш чіткої залежності результативного показника від факторів інтенсивності виробництва здійснено перетворення вихідних рядів динаміки.

Потреба перетворення рядів динаміки зумовлена тим, що різнойменні показники наведених чотирьох паралельних рядів безпосередньо непорівнянні між собою. Крім того, вони виражені в різних одиницях вимірювання.

Найпростішим і розповсюдженим прийомом вивчення залежності результативних показників від факторів, що визначають тенденцію, є прийом зведення рядів динаміки до однієї основи.

Щоб звести подібні ряди динаміки до порівнянного вигляду, використаємо прийом зведення їх до однієї основи. З цією метою розрахуємо базисні темпи зростання, взявши за постійну базу порівняння рівні 2005 року. Добуті дані виразимо у відсотках (табл. 10.3).

Порівняння темпів зростання виходу валової продукції і факторів виробництва за даними табл. 10.3 свідчить про випереджаючі темпи зростання факторів інтенсивності виробництва (в 1,5 раза) порівняно з темпами зростання виходу валової продукції (в 1,363, або на 36,3%), Це означає, що в даному

господарстві в динаміці вихід валової продукції на одиницю факторів мав тенденцію до зниження.

Таблиця 10.3

Динаміка виходу валової продукції сільського господарства і факторів інтенсивності виробництва у фермерському підприємстві за 2005 - 2010 рр. (в % до 2005 р.)

Рік	В розрахунку на 100 га сільськогосподарських угідь			
	Вартість валової продукції сільського господарства, тис. грн	Вартість основних виробничих фондів сільськогосподарського призначення (фондозабезпеченість), тис. грн	Енергетична потужність (енергозабезпеченість), кінських сил	Кількість внесених мінеральних добрив, ц діючої речовини
2005	100	100	100	100
2006	107,0	108,8	104,5	97,8
2007	112,4	116,4	111,8	108,5
2008	133,3	131,3	118,3	130,4
2009	126,4	141,9	132,9	127,4
2010	136,3	153,8	147,2	154,7

Зроблені за факторним аналізом рядів динаміки висновки доповнимо розрахунком коефіцієнтів випередження. Коефіцієнти випередження визначимо як відношення базисних темпів зростання за однакові відрізки часу по двох динамічних рядах.

Так, коефіцієнт випередження зростання фондозабезпеченості порівняно зі зростанням валової продукції становить 1,13 (1,538:1,363), зростання енергозабезпеченості порівняно зі зростанням валової продукції – 1,08 (1,472:1,363), зростання кількості внесених мінеральних добрив порівняно зі зростанням валової продукції – 1,13 (1,547:1,363). Отже, темпи зростання факторів інтенсивності виробництва – фондозабезпеченості, енергозабезпеченості, кількості внесених добрив – випереджали темпи зростання валової продукції сільського господарства відповідно в 1,13; 1,08 і 1,13 разів.

Важливу роль у проведенні факторного аналізу рядів динаміки має *прийом зіставлення кількох паралельних рядів*. Найширше цей прийом

використовується для вивчення залежності результативної ознаки від комплексу факторів. Покажемо це на прикладі табл. 10.4.

Таблиця 10.4

Динаміка продуктивності корів і факторів молочного скотарства у фермерському господарстві

Показник	У середньому за три роки		
	2002-2004 рр.	2005-2007 рр.	2008-2010 рр.
Надій молока на корову, кг	3452	3817	4264
Витрати кормів з розрахунку на корову, ц кормових	40,5	44,7	48,1
В тому числі концентрованих	9,4	10,2	12,3
Частка чистопородних корів, %	91	96	100
Вихід телят на 100 корів, гол.	92	95	98

Дані таблиці 10.4 свідчать про те, що підвищення продуктивності корів у динаміці зумовлено зростанням рівня і якості годівлі, удосконаленням породної структури стада, а також кращим використанням корів для одержання приплоду.

Факторний аналіз рядів динаміки може бути здійснений на основі низки прийомів виявлення основної тенденції, викладених у пункті 10.1 даної лекції, зокрема, з використанням *прийому укрупнення періодів* (табл. 10.5).

Таблиця 10.5

Динаміка урожайності зернових культур і її факторів господарствах району

Періоди	Урожайність, ц/га	Метеоумови		Економічні умови	
		Сума опадів за травень-червень, мм	Сума опадів за серпень-вересень, мм	Внесено міндобрив на 1 га зернових культур, ц діючої речовини	Вартість силових і робочих машин, тис. грн
1991-1995	21,7	129,5	104,3	0,85	46,5
1996-2000	25,9	106,2	90,8	1,05	68,7
2001-2005	30,8	117,4	101,5	1,57	76,8
2006-2010	38,5	133,7	109,9	2,15	81,5

Для проявлення залежності урожайності зернових культур від факторів інтенсифікації динамічний ряд урожайності був підрозділений на специфічні для району періоди. Це дало змогу по періодах нівелювати вплив метеорологічних умов (періоди за метеоумовами в цілому суттєво не відрізняються один від одного) і проявити вплив факторів інтенсифікації (кількості внесених мінеральних добрив і забезпеченості силовими і робочими машинами).

При факторному аналізі рядів динаміки важливе значення має розгляд показників, що безпосередньо визначають рівень досліджуваного явища. Якщо досліджуваний показник є складеним, то його необхідно розкласти на відповідні прості з наступним аналізом кожного з них. Наприклад, надій на корову можна розглядати як добуток двох показників: витрат кормів на одну корову і окупності кормів (вихід молока на 1ц кормових одиниць). Собівартість 1ц молока є, нарешті, відношення виробничих витрат на одну корову до продуктивності і т.п.

10.4. Інтерполяція та екстраполяція. Прогнозування на основі рядів динаміки

Під час аналізу рядів динаміки доводиться стикатися з такими випадками, коли в рядах відсутні дані про їхні рівні за той або інший період. Такі дані можуть бути відсутні або всередині ряду, або спочатку чи в кінці його.

Приблизне визначення відсутніх рівнів усередині одноякісного періоду, коли відомі рівні, що лежать по обидві сторони невідомого, називають *інтерполяцією ряду динаміки*. Приблизне визначення невідомих рівнів, що лежать за його межами, тобто в майбутньому (або в минулому), називають *екстраполяцією ряду динаміки*. Відповідно екстраполювання може здійснюватися як у бік майбутнього (перспективна екстраполяція), так і у бік минулого (ретроспективна екстраполяція). По суті, екстраполяція являє собою продовження ряду динаміки на основі виявленої закономірності зміни рівнів за досліджуваний відрізок часу.

Інтерполяцію (як і екстраполяцію) здійснюють виходячи з припущення, що зміни в межах періоду, що виражають закономірність розвитку, відносно стійкі, тобто що ні виявлена тенденція, ні її характер не зазнали і не зазнають суттєвих змін у тому проміжку часу, рівні якого нам невідомі.

Щоб мати досить надійні результати обчислення відсутніх рівнів, інтерполяцію та екстраполяцію слід проводити в межах однорідних періодів, яким властива одна закономірність розвитку.

Інтерполяцію та екстраполяцію ряду динаміки можна проводити різними способами. Найпростішим способом є використання середніх характеристик досліджуваного ряду динаміки: середнього абсолютного приросту (при стабільних ланцюгових абсолютних приростах) і середнього коефіцієнта зростання (при стабільних темпах зростання). Однак визначення відсутніх рівнів ряду динаміки, і особливо при екстраполяції, найчастіше пов'язують з аналітичним вирівнюванням рядів *способом найменших квадратів*, який дає точніші результати. При цьому для виходу за межі періоду, для якого знайдена залежність від часу, досить продовжити значення незалежної змінної – часу.

Дослідження динаміки суспільних явищ і виявлення основної тенденції їх розвитку в минулому дають основу для визначення їхніх майбутніх розмірів.

Велику роль в плануванні має екстраполяція, яка дає змогу прогнозувати соціально – економічні явища. Прогнозування є важливим етапом планової роботи.

Під *прогнозуванням* розуміють процес наукового виявлення важливих шляхів і результатів майбутнього розвитку соціально–економічних явищ, оцінку показників, що характеризують ці явища для більш або менш віддаленого майбутнього.

Розрізняють короткострокові прогнози (від кількох днів до одного року), середньострокові (від одного року до 5 років) і довгострокові прогнози (понад 5 років).

Застосування екстраполяції для прогнозування базується на припущенні, що характер динаміки, тобто певна закономірність (тенденція) зміни

досліджуваного явища, яка мала місце для певного періоду часу в минулому, збережеться на обмеженому відрізку в майбутньому. Така екстраполяція справедлива, якщо система розвивається еволюційно в досить стабільних умовах. Чим крупніша система, тим більш ймовірно збереження параметрів її зміни, звісно, на невеликий строк.

У зв'язку з цим прогнозуванню має передувати ретельний аналіз комплексу взаємопов'язаних факторів, які в майбутньому будуть визначати тенденцію розвитку досліджуваного соціально–економічного явища.

Наведемо приклад прогнозування основних показників розвитку малих підприємств на селі за допомогою одержаних рівнянь лінійного тренду, прийнявши для розрахунку $t = 14, 16, 18$ відповідно для 2014, 2016 і 2018 років (табл. 10.6).

Таблиця 10.6

**Прогнозування основних показників розвитку малих підприємств
Хмельницької області на селі**

Показники	Рівняння лінійного тренду: $\tilde{y}_t = a \cdot t + b$	Фактичні дані за 2012 рік	Прогнозовані значення показників		
			2014 р.	2016 р.	2018 р.
Кількість зайнятих працівників, осіб	$\tilde{y}_t = -748,1t + 26704,2$	17560	16230,8	14734,6	13238,4
Кількість найманих працівників, осіб	$\tilde{y}_t = -679,8t + 24831,4$	16673,8	15314,2	13954,6	12595,0
Обсяг реалізованої продукції (робіт, послуг), млн. грн.	$\tilde{y}_t = 288,9t + 231,4$	3611,1	4276,0	4853,8	5431,6
Прибуток до оподаткування, млн грн.	$\tilde{y}_t = 40,9t - 166,1$	318,2	406,5	488,3	570,1
Збиток до оподаткування, млн грн.	$\tilde{y}_t = 13,66t - 6,60$	156,8	184,6	211,0	238,2

На основі одержаних результатів середньострокового прогнозу, записаних в табл. 10.6, можемо зробити висновок про обернену залежність між кількістю зайнятих та найманих працівників області і часом, вираженим в роках: протягом найближчих років відбудеться зменшення таких працівників у малих підприємствах області. Обсяг реалізованої продукції, прибуток і збиток до оподаткування матиме позитивну динаміку.

Питання для повторення і самоконтролю

1. В чому полягає основне завдання рядів динаміки і для чого його вирішують?
2. Які Ви знаєте спеціальні прийоми обробки динамічних рядів?
3. В чому полягає суть укрупнення періодів і як його проводять?
4. В чому полягає суть способу ковзних середніх і як його проводять?
5. Як проводиться обробка динамічного ряду за середнім абсолютним приростом?
6. Як проводиться вирівнювання ряду динаміки за середнім коефіцієнтом зростання?
7. В чому полягає суть методу найменших квадратів?
8. Що таке рівняння трейду?
9. Що таке сезонні коливання і якими показниками їх вимірюють?
10. Що таке факторний аналіз та які методи і прийоми використовують для його проведення?
11. Як практично можна звести ряди динаміки до однієї порівнюваної основи?
12. Як реалізується прийом зіставлення кількох паралельних рядів динаміки?
13. Що таке інтерполяція та екстраполяція рівнів динаміки і як їх виконують?
14. В чому полягає суть прогнозування і на якій підставі воно виконується?

Лекція 11. Індексний метод

План

- 11.1. Суть індексів та їх роль у статистико-економічному аналізі.
- 11.2. Класифікація індексів. Методологічні принципи їх побудови.
- 11.3. Базисні та ланцюгові індекси, зв'язок між ними. Інтегральний індекс динаміки.
- 11.4. Індекси агрегатної форми.
- 11.5. Середньозважені індекси. Розкладання абсолютного приросту за факторами.
- 11.6. Індекси змінного і фіксованого складу та структурних зрушень, їх взаємозв'язок.
- 11.7. Індекси Ласпейреса, Пааше і Фішера. Індекс споживчих цін.

11.1. Суть індексів та їх роль у статистико-економічному аналізі

Для характеристики соціально-економічних явищ і процесів статистика широко використовує узагальнюючі показники у вигляді середніх, відносних величин та коефіцієнтів. Одним з таких узагальнюючих показників і є індекс. В широкому розумінні слово "*Index*" у перекладі з латинської означає "*показник*".

Індексом у статистиці називається відносний показник, що характеризує зміну рівня соціально-економічного явища в часі, порівняно з планом, базисним періодом або у просторі.

В статистичних дослідженнях складних соціально-економічних явищ і процесів виділяють три великі сфери застосування економічних індексів.

До першої сфери застосування індексів відносять порівняльну характеристику несумірних сукупностей в часі. Сюди входять синтетичні індекси динаміки, індекси виконання плану і територіальні індекси.

Індекси динаміки показують зміну якого-небудь складного явища в звітному періоді порівняно з базисним.

Індекс виконання плану використовують для порівняння досягнутого рівня з плановими завданнями.

Територіальні індекси застосовують для просторового порівняння рівнів урожайності, цін, продуктивності праці і т.п., в різних регіонах.

Друга сфера застосування індексів полягає в їх використанні для факторного аналізу складного явища через систему взаємозв'язаних індексів. До таких складних явищ можуть бути віднесені: вартість виробленої чи реалізованої продукції, фонд заробітної плати, валовий збір зерна тощо. Так, вартість виробленої продукції дорівнює добутку цін на кількість продукції, валовий збір зерна – добутку урожайності на посівну площу, фонд заробітної плати – добутку заробітної плати одного працівника на їх чисельність і т.п.

За допомогою *третьої сфери застосування індексів* проводять аналіз динаміки середніх величин, зміна яких піддається впливу структурних зрушень в середині досліджуваної сукупності. В зв'язку з цим, велике значення має вивчення впливу структурних зрушень на динаміку середніх показників через застосування системи взаємозв'язаних індексів змінного складу, постійного (фіксованого) складу і структурних зрушень.

11.2. Класифікація індексів. Методологічні принципи їх побудови

Всі економічні індекси статистика класифікує за трьома основними ознаками:

- а) за характером досліджуваних об'єктів;
- б) за ступенем охоплення елементів сукупності;
- в) за методикою розрахунку загальних індексів.

За характером досліджуваних об'єктів індекси ділять на *індекси об'ємних (кількісних) і якісних показників*.

До першої групи відносяться індекси фізичного обсягу продукції промисловості, сільського господарства, будівництва тощо.

До другої групи якісних показників відносять індекси цін, собівартості, урожайності і ряд інших.

За ступенем охоплення елементів сукупності індекси ділять на: а) індивідуальні; б) загальні; в) групові.

Індивідуальні індекси характеризують зміну окремих елементів складного явища. В теорії індексів показник, зміну якого характеризує індекс, називається *індексованою величиною*.

Індивідуальні індекси позначають малою латинською буквою «*i*», продукцію в натуральному виразі – через «*q*», ціну одиниці товару – через «*p*», собівартість одиниці продукції – через «*z*» і т.д. Індивідуальні індекси цих ознак визначаються за формулами:

$$\text{а) фізичного обсягу: } i_q = \frac{q_1}{q_0};$$

$$\text{б) ціни одиниці товару: } i_p = \frac{p_1}{p_0};$$

$$\text{в) собівартості одиниці продукції: } i_z = \frac{z_1}{z_0},$$

де i_q, i_p, i_z – індивідуальні індекси фізичного обсягу, ціни і собівартості одиниці продукції;

$q_1, q_0; p_1, p_0; z_1, z_0$ – фізичний обсяг, ціна, собівартість у звітному і базисному періодах.

Загальні індекси характеризують зміну сукупності в цілому і являють собою відносні числа, що визначають зміни в часі порівняно з плановим, базисним періодами або в просторі складного явища, яке складається з несумірних елементів.

Груповими або *субіндексами* називаються такі індекси, які охоплюють не всі елементи сукупності, а тільки яку-небудь частину або їх групу.

Базисні і ланцюгові індекси обчислюють в тих випадках, коли доводиться вивчати яке-небудь явище суспільного життя за ряд послідовних років.

11.3 Базисні та ланцюгові індекси, зв'язок між ними. Інтегральний індекс динаміки

За допомогою індивідуальних індексів порівнюють дані лише за два періоди. Дуже часто при аналізі того чи іншого суспільного явища доводиться оперувати даними за три і більше послідовних періодів. У цих випадках виникає питання про

вибір бази порівняння. Залежно від того, що приймається за базу порівняння, індекси можуть бути *базисними та ланцюговими*.

Базисні індекси одержують шляхом порівняння абсолютних рівнів кожного періоду з рівнем якогось періоду (звичайно початкового), прийнятого за базу порівняння.

Ланцюгові індекси одержують шляхом порівняння абсолютного рівня кожного періоду з рівнем попереднього періоду.

Між базисним і ланцюговим індивідуальними індексами існує взаємозв'язок, що дає змогу переходити від ланцюгових індексів до базисних і навпаки. Послідовне перемноження ланцюгових індексів дає базисний індекс відповідного періоду. Ланцюгові індекси можна одержати з базисних шляхом ділення відповідного базисного індексу на попередній базисний індекс. Методика обчислення базисних та ланцюгових індексів наочно свідчить про те, що індивідуальні індекси нічим не відрізняються від відносних величин динаміки.

Застосування базисних або ланцюгових індивідуальних індексів залежить від мети дослідження. Якщо треба охарактеризувати динаміку явища порівняно з певним періодом, то обчислюють базисні індекси, а якщо треба виявити результати зміни кожного окремого періоду порівняно з попереднім — ланцюгові індекси.

Інша справа, якщо нам потрібно обчислити загальні індекси базисним або ланцюговим способом, тому що завжди виникає питання про використання постійних або змінних ваг, оскільки, залежно від мети дослідження, використовуються і ті і інші. Якщо для всього індексного ряду застосовують ваги на рівні одного і того самого періоду, то базисні та ланцюгові індекси мають постійні ваги, а якщо ваги змінюються від одного індексу до другого, то індекси мають змінні ваги. Для індексів із постійними вагами зберігається співвідношення між ланцюговими і базисними індексами, яке встановлено для індивідуальних індексів.

Важливе значення при дослідженні динаміки соціально-економічного процесу чи явища має інтегральний індекс динаміки, який обчислюється за формулою:

$$I = \frac{1}{n} \cdot \sum_{i=1}^n \frac{q_{ik}}{q_{iп}}, \quad (11.1)$$

де: $q_{iп}$ – показник на початок i -го періоду; q_{ik} – показник на кінець i -го періоду; n – число періодів. Індекс динаміки, обчислений за формулою (11.1), характеризує середньостатистичну динаміку досліджуваного процесу, оскільки він охоплює всі індивідуальні ланцюгові індекси ряду динаміки, обчислені як відношення кожного наступного рівня ряду до попереднього. Незважаючи на можливі коливання динаміки ряду протягом певного періоду часу (наприклад, протягом декількох років), інтегральний індекс дозволяє визначити загальну тенденцію динамічного ряду за цей період з урахуванням усіх індивідуальних ланцюгових індексів.

Проілюструємо на прикладі малих підприємств Хмельницької області використання інтегральних індексів для характеристики їх динаміки та деяких видів діяльності протягом 2007-2012 рр. (табл. 11.1).

Таблиця 11.1

Деякі показники динаміки та діяльності малих підприємств Хмельницької області

Показники	Тип регіону	Р о к и						Інтегральні індекси динаміки	2012 р. у % до 2007 р.
		2007	2008	2009	2010	2011	2012		
Кількість малих підприємств на 10 тис. осіб зайнятого населення, одиниць	Область	57	51	55	44	44	47	0,9683	82,5
	Райони	28	24	28	22	29	34	1,0600	121,4
Кількість зайнятих працівників, осіб	Область	54220	52148	50049	48046	46246	42900	0,9543	85,2
	Райони	21843	20450	19917	18468	19343	17560	0,9585	79,1
Обсяг реалізованої продукції (робіт, послуг), млн грн	Область	5390,6	5271,4	5631,3	6080,6	7816,2	8472,7	1,0991	157,2
	Райони	1893,3	2279,6	2188,3	2283,5	2921,8	3414,3	1,1311	180,3
Прибуток до оподаткування, млн грн	Область	254,6	345,5	242,8	312,5	684,1	716,4	1,3167	281,4
	Райони	142,5	193,5	135,9	175,0	382,0	308,2	1,2675	216,3

В останній правій графі табл. 11.1 записані значення індивідуальних базових індексів динаміки, які характеризують у відсотках рівень показників 2012 року по відношенню до 2007 року, прийнятого за базовий. За їх допомогою можна зробити висновки про серйозне зменшення кількості малих підприємств і зайнятих працівників області в цілому у 2012 році порівняно з 2007 роком. Однак така характеристика є неповною, оскільки не виявляє загальної тенденції динамічного ряду протягом усіх років.

Інтегральний індекс першого показника (кількість малих підприємств області на 10 тис. осіб зайнятого населення, одиниць) розрахований, згідно з формулою (11.1), таким чином:

$$I_1 = \frac{1}{5} \left(\frac{51}{57} + \frac{55}{51} + \frac{44}{55} + \frac{44}{44} + \frac{47}{44} \right) = 0,9683.$$

Аналогічно обчислюються інтегральні індекси динаміки інших показників, їхні значення записані у передостанній графі табл. 11.1. Значення інтегрального індексу $I_1 = 0,9683$ вказує, що протягом розглянутого періоду відбувалось зменшення кількості малих підприємств в області у середньому на $100 - 96,83 = 3,18\%$ щороку; значення $1,2675$ інтегрального індексу засвідчує зростання прибутку до оподаткування в районах області у середньому на $26,75\%$ щороку, незважаючи на те, що цей показник протягом усього періоду коливався в більшу чи меншу сторону (у 2012 році прийняв значно менше значення, ніж у 2011 році). Зрозуміло, що інтегральний індекс тим точніше описує загальну динаміку, чим меншою є амплітуда таких коливань і чим більший період часу він охоплює.

11.4. Індекси агрегатної форми

В залежності від методології обчислення, загальні і групові індекси діляться на агрегатні і середні з індивідуальних індексів.

Агрегатні індекси є основною формою економічних індексів, а середні із індивідуальних індексів – похідними, отриманими в результаті перетворення агрегатних індексів.

Агрегатний індекс є основною формою економічного індексу. Його назва пішла від латинського слів «*aggrego*» – приєдную. Чисельник і знаменник цього індексу являє собою агрегат, набір різнорідних елементів.

Отже, *агрегатним індексом* в статистиці називається *загальний індекс*, який є відношенням сум добутоків індексованих (зіставлюваних) величин порівнюваних періодів до їх ваг.

При побудові формул агрегатних індексів використовують наступне правило: *якщо індексована величина – якісний показник, який знаходять шляхом ділення (ціна, собівартість, урожайність тощо), ваги беруться звітного періоду, а якщо індексована величина – кількісний показник, який можна підсумувати (фізичний обсяг продукції, чисельність працівників, посівна площа), ваги беруться базисного періоду.*

Покажемо застосування цього правила при побудові формул агрегатних індексів.

Загальний індекс цін визначається за формулою:

$$I_P = \frac{\sum p_1 q_1}{\sum p_0 q_0} \quad (11.2)$$

Цей індекс показує, як змінилися ціни на всі досліджувані товари в звітному періоді порівняно з базисним.

Загальний індекс фізичного обсягу обчислюється за формулою:

$$I_q = \frac{\sum p_0 q_1}{\sum p_0 q_0} \quad (11.3)$$

Даний індекс показує зміну кількості виробленої або реалізованої продукції в звітному періоді порівняно з базисним.

Загальний індекс обсягу товарообороту показує зміну виробництва або реалізації продукції в звітному періоді порівняно з базисним у фактичних цінах:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \quad (11.4)$$

Ці індекси взаємозв'язані залежністю:

$$I_P = I_q \times I_{pq} \quad (11.5)$$

11.5. Середньозважені індекси. Розкладання абсолютного приросту за факторами

В деяких випадках загальні індекси обчислюють як середні перетворені з відповідних агрегатних індексів.

Перетворюють агрегатний індекс в середній з індивідуальних індексів, підставляючи у його чисельник або знаменник замість індексованого показника його вираз, виведений з формули індивідуального індексу. Якщо таку заміну роблять у чисельнику, то агрегатний індекс перетворюється у середній арифметичний, а якщо у знаменнику – в середній гармонічний.

Перетворимо агрегатний індекс фізичного обсягу, який обчислюється за формулою (11.2), в середній арифметичний. Для цього врахуємо, що $i_q = \frac{q_1}{q_0}$,

звідки $q_1 = q_0 \cdot i_q$.

З урахуванням одержаного виразу формула (11.2) матиме вигляд:

$$I_q = \frac{\sum p_0 q_0 i_q}{\sum p_0 q_0} . \quad (11.6)$$

За одержаною формулою (11.5) обчислюється *середній арифметичний індекс* фізичного обсягу продукції.

Перетворимо загальний індекс цін, що визначається за формулою (11.2),

у *середній гармонічний*. Оскільки $i_p = \frac{p_1}{p_0}$, то звідси: $p_0 = \frac{p_1}{i_p}$ і формула (11.2)

набере вигляду:

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}} . \quad (11.7)$$

Середні арифметичні і гармонічні індекси повинні співпадати за своєю величиною з відповідними агрегатними індексами.

Вибір форми індексу залежить від поставленого завдання дослідження і від наявності даних, необхідних для обчислення того чи іншого індексу.

Абсолютна сума економії або перевитрат (абсолютний приріст) від зміни цін визначається як різниця між чисельником і знаменником загального індексу цін (див. формулу 1.1): $\Delta pq(p) = \sum p_1 q_1 - \sum p_0 q_1$.

Абсолютна сума економії або перевитрат (абсолютний приріст) від зміни кількості виробленої або реалізованої продукції визначається як різниця між чисельником і знаменником загального індексу фізичного обсягу (див. формулу (11.2): $\Delta pq(q) = \sum p_0 q_1 - \sum p_0 q_0$.

Загальний приріст обсягу товарообігу: $\Delta pq = \sum p_1 q_1 - \sum p_0 q_0$, або з урахуванням абсолютних приростів встановлюємо взаємозв'язок: $\Delta pq = \Delta pq(p) + \Delta pq(q)$.

На основі індексів середніх величин виконують *факторний аналіз* середньої величини складного явища.

Абсолютний приріст середньої собівартості:

а) загальний:
$$\Delta \bar{Z} = \frac{\sum z_1 q_1}{\sum q_1} - \frac{\sum z_0 q_0}{\sum q_0};$$

б) за рахунок зміни собівартості окремих видів продукції:

$$\Delta \bar{Z}(z) = \frac{\sum z_1 q_1}{\sum q_1} - \frac{\sum z_0 q_1}{\sum q_1};$$

в) за рахунок структурних змін у кількості продукції:

$$\Delta \bar{Z}(cmp) = \frac{\sum z_0 q_1}{\sum q_1} - \frac{\sum z_0 q_0}{\sum q_0}.$$

Перевірка: $\Delta \bar{Z} = \Delta \bar{Z}(z) + \Delta \bar{Z}(cmp)$.

Абсолютний приріст складного явища (грошових витрат на виробництво продукції):

а) загальний:

$$\Delta Zq = \sum Z_1 q_1 - \sum Z_0 q_0;$$

б) за рахунок зміни собівартості окремих видів продукції:

$$\Delta Zq(z) = \Delta \bar{Z}(z) \cdot \sum q_1;$$

в) за рахунок структурних змін у кількості продукції:

$$\Delta Zq(cmp) = \Delta \bar{Z}(cmp) \cdot \sum q_1;$$

г) за рахунок зміни кількості виробленої продукції:

$$\Delta Zq(q) = \left(\sum q_1 - \sum q_0 \right) \cdot \bar{Z}_0.$$

$$\text{Перевірка: } \Delta Zq = \Delta Zq(z) + \Delta Zq(\text{стр}) + \Delta Zq(q).$$

Методику обчислення та економічний зміст середньозважених індексів покажемо на прикладі аналізу реалізації продукції в господарстві за два періоди (табл.11.2).

Використовуючи індексний метод аналізу і вихідні дані табл. 11.2, встановимо загальну зміну вартісного обсягу продукції за два періоди і вплив на його зміну фізичного обсягу продукції і цін.

Таблиця 11.2

Дані для розрахунку індексів вартісного та фізичного обсягів продукції і цін

Продукція	Вихідні дані				Розрахункові дані				
	Кількість продукції, тис. ц		Ціна реалізації 1 ц, грн		Індивідуальні індекси		Виручка від реалізації, тис. грн		
	базисний період	звітний період	базисний період	звітний період	обсягу продукції	цін	базисний період	звітний період	умовна
	q_0	q_1	p_0	p_1	$i_0 = \frac{q_1}{q_0}$	$i_1 = \frac{p_1}{p_0}$	$q_0 p_0$	$q_1 p_1$	$q_1 p_0$
Зерно	25,6	30,7	16,83	17,66	1,1992	1,0493	430,8	542,2	516,7
Картопля	18,7	15,0	23,29	22,41	0,8021	0,9622	435,5	336,2	394,4
Молоко	22,5	26,4	21,24	23,08	1,1733	1,0866	477,9	609,3	560,7
М'ясо	1,9	1,7	159,37	163,12	0,8947	1,0235	302,8	277,3	270,9
Разом:	-	-	-	-	-	-	1647,0	1765,0	1697,7

Індивідуальні індекси (коефіцієнти зростання) показують, що у звітному періоді порівняно з базисним фізичний обсяг реалізації зерна та молока збільшився відповідно на 19,92 і 17,33%, а картоплі і м'яса скоротився на 19,79 і 10,53%, ціни реалізації по зерну, молоку і м'ясу збільшились відповідно на 4,93%, 8,66 та 2,35%, а по картоплі ціна зменшилась на 3,78%.

Визначимо відносну зміну вартісного обсягу реалізованої продукції по всіх видах, для чого розрахуємо індекс вартісного обсягу продукції:

$$I_p = \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{1765}{1647} = 1,0716, \quad \text{або } 107,16\%.$$

Цей індекс показує, що в динаміці внаслідок дії двох факторів (обсягу продукції і цін) вартісний обсяг збільшився на 7,16%.

Різниця між чисельником і знаменником індексу характеризує абсолютний приріст вартісного обсягу продукції за два періоди:

$$\Delta_{qp} = \sum q_1 p_1 - \sum q_0 p_0 = 1765 - 1647 = 118 \text{ тис. грн}$$

Щоб визначити зміну вартісного обсягу продукції за рахунок зміни тільки фізичного обсягу продукції, обчислимо індекс фізичного обсягу продукції, в якому продукція звітного і базисного періодів оцінена в єдиних цінах:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{1697,7}{1647,0} = 1,0308 \text{ або } 103,08\%.$$

Отже, вартісний обсяг реалізованої продукції у звітному періоді порівняно з базисним внаслідок зростання фізичного обсягу продукції збільшився на 3,08%.

Абсолютний приріст вартісного обсягу реалізованої продукції за рахунок зростання фізичного обсягу становить:

$$\Delta_q = \sum q_1 p_0 - \sum q_0 p_0 = 1697,7 - 1647,0 = 50,7 \text{ тис. грн.}$$

Вплив зміни цін реалізації на зміну вартісного обсягу продукції встановимо за допомогою індексу цін:

$$I_{pq} = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{1765,0}{1697,7} = 1,0396 \text{ або } 103,96\%.$$

Отже, вартісний обсяг реалізованої продукції внаслідок зміни цін збільшився на 3,96%.

Абсолютний приріст вартісного обсягу продукції внаслідок зростання цін реалізації становить: $\Delta_p = 1765,0 - 1697,7 = 67,3$ тис. грн.

Отже, вартісний обсяг реалізованої продукції по господарству за два періоди збільшився на 7,16%, в тому числі внаслідок зростання фізичного обсягу продукції на 3,08% і збільшення цін реалізації на 3,96%.

Обчислені агрегатні індекси показують середню зміну вартісного обсягу продукції, а тому аналіз одержаних результатів доповнимо аналізом по окремих видах продукції.

Загальне збільшення фізичного обсягу продукції в середньому на 3,08% пояснюється більш високим зростанням обсягів реалізації зерна на 19,92% і

молока на 17,33% порівняно із скороченням обсягів реалізації картоплі на 19,79% і м'яса на 10,53%.

Аналіз індивідуальних індексів цін по окремих видах продукції показує, що, не дивлячись на зниження цін реалізації по картоплі на 3,78%, в середньому ціни підвищились на 3,96% внаслідок їх зростання по зерну на 4,93%, молоку – на 8,66% і м'ясу – на 2,35%.

11.6. Індеси змінного і фіксованого складу та структурних зрушень, їх взаємозв'язок.

Для якісних показників, таких як середня ціна, собівартість, урожайність та інших по однойменній продукції, але віднесеної до різних об'єктів, обчислюють загальні індекси змінного, постійного (фіксованого) складу і структурних зрушень.

Індекс, який характеризує спільний вплив обох чинників, називається *індексом змінного складу* і визначається за формулою:

$$I_{з.с.} = \frac{\sum z_1 q_1}{\sum q_1} \cdot \frac{\sum z_0 q_0}{\sum q_0}. \quad (11.8)$$

На величину індексу собівартості змінного складу впливають зміни рівнів собівартості і зміни в структурі (її складі). Щоб виявити роль кожного чинника в загальній динаміці середньої, потрібно індекс змінного складу розкласти на два індекси-співмножники, кожний з яких відображає вплив тільки одного чинника.

Перший індекс, який характеризує вплив тільки індексованої величини (в якому змінюється лише собівартість), називається *індексом постійного (фіксованого) складу*. Він обчислюється за формулою:

$$I_{пост} = \frac{\sum z_1 q_1}{\sum q_1} \cdot \frac{\sum z_0 q_1}{\sum q_1}. \quad (11.9)$$

Другий індекс показує, як змінюється середній рівень (середня собівартість) тільки за рахунок зміни структури явища (структури продукції). Він називається *індексом структурних зрушень* і визначається за формулою:

$$I_{стр} = \frac{\sum z_0 q_1}{\sum q_1} \cdot \frac{\sum z_0 q_0}{\sum q_0}. \quad (11.10)$$

Ці індекси взаємозв'язані співвідношенням: $I_{з.с.} = I_{ПОСТ} \times I_{СТР.}$

11.7. Індекси Ласпейреса, Пааше і Фішера. Індекс споживчих цін

Наведений вище спосіб обчислення агрегатних індексів фізичного обсягу продукції і цін відображає практику вітчизняної статистики. В багатьох країнах індекси фізичного обсягу продукції і цін також обчислюються аналогічним чином. Разом з тим, у міжнародній статистиці для розрахунків агрегатних індексів фізичного обсягу продукції і цін застосовуються й інші форми індексів, пов'язаних з вибором сумірників (ваг) індексів.

При обчисленні вищенаведених індексів використані ваги базисного і звітного періодів: у індексі фізичного обсягу продукції – ваги базисного періоду, а у індексі цін – ваги звітного періоду. Проте зазначений вибір ваг при побудові агрегатних індексів фізичного обсягу продукції і цін не можна вважати обов'язковим в усіх випадках. В статистиці багато задач можуть і повинні розв'язуватися по-різному, залежно від конкретної мети і особливостей дослідження.

У зв'язку з цим у статистичному аналізі використовуються дві форми індексів фізичного обсягу продукції і цін з базисними і звітними вагами – відповідно індекси Ласпейреса і Пааше:

Назва індексу	Агрегатні індекси:	
	фізичного обсягу продукції	цін
Індекс з базисними вагами (формула Ласпейреса)	$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}$	$I_p = \frac{\sum q_0 p_1}{\sum q_0 p_0}$
Індекс з вагами звітного періоду (формула Пааше)	$I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1}$	$I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0}$
«Ідеальна» формула Фішера	$I_\Phi = \sqrt{\frac{\sum q_1 p_0}{\sum q_0 p_0} \cdot \frac{\sum q_1 p_1}{\sum q_0 p_1}}$	$I_\Phi = \sqrt{\frac{\sum q_0 p_1}{\sum q_0 p_0} \cdot \frac{\sum q_1 p_1}{\sum q_1 p_0}}$

Автором формули індексу фізичного обсягу продукції з базисними вагами є німецький економіст Е. Ласпейрас, запропонованої ним у 1864 р.

Агрегатний індекс цін з вагами звітнього періоду був запропонований німецьким статистиком Г. Пааше в 1874 р.

Індексованою величиною обох індексів цін Пааше і Ласпейреса є ціни. Вагами ж у індексу цін Пааше виступає кількість продукції звітнього періоду, а в індексі цін Ласпейреса – кількість продукції базисного періоду.

Ці індекси не ідентичні. Значення індексів цін Пааше і Ласпейреса для одних і тих самих даних не збігаються, тому що мають різний економічний зміст.

Індекс Пааше характеризує зміни цін у звітньому періоді порівняно з базисним по продукції (товарах і послугах), реалізованій у звітньому періоді, і фактичну економію (перевитрати) від зміни цін, тобто індекс цін Пааше показує, на скільки продукція у звітньому періоді стала дорожчою (дешевшою), ніж у базисному.

Індекс Ласпейреса показує, на скільки змінились ціни у звітньому періоді порівняно з базисним, але по тій продукції, яка була реалізована в базисному періоді, та економію (перевитрати), які можна було б одержати від зміни цін, тобто умовну економію (перевитрати). Інакше кажучи, індекс цін Ласпейреса показує у скільки разів продукція базисного періоду подорожчала (стала дешевшою) через зміну цін на неї у звітньому періоді.

Як правило, індекс цін, розрахований за формулою Пааше, дещо занижує, а за формулою Ласпейреса – завищує темпи інфляції. Ця систематична залежність двох індексів відома в статистиці як ефект Гершенкрона.

До переходу до ринкової економіки вітчизняна статистика віддавала перевагу індексу цін Пааше. В умовах високої інфляції зважування вагами звітнього періоду (індекс цін Пааше) потребує щорічного (щоквартального, щомісячного) перерахунку інформації для формування системи ваг, що пов'язано з великими затратами праці, матеріальних і трудових ресурсів, тому, починаючи з 90-х років ХХ ст., органи державної статистики України визначають зміну загального рівня цін на товари і послуги за формулою Ласпейреса, якій віддається перевага і в зарубіжній статистиці. При обчисленні

індексу цін за формулою Ласпейреса ваги на рівні базисного періоду залишаються незмінними протягом деякого проміжку часу.

Проте мірою віддалення від базисного року товарний кошик (товари-представники, за якими визначають індекс цін) за видами, кількістю і якістю все менше відповідає структурі і складу обсягу продукції поточного року. Тому склад товарного кошика, а отже, і система ваг мають періодично переглядатися. Особливо це важливо в період різких змін в економіці країни.

Якщо підходити до принципів побудови індексів з формально-математичних позицій, то, орієнтуючись на принцип елімінування впливу інших факторів, крім досліджуваного, можливо при обчисленні індексів фізичного обсягу продукції і цін спиратися на ваги базисного періоду (формула Ласпейреса) або ж на ваги звітного періоду (формула Пааше). Орієнтуючись на ці два варіанти американський економіст І.Фішер запропонував розраховувати середню геометричну з двох агрегатних індексів, назвавши її "ідеальною формулою".

"Ідеальний" індекс цін І.Фішера являє собою середню геометричну з добутку двох агрегатних індексів цін Ласпейреса і Пааше:

$$I_{\Phi} = \sqrt{\frac{\sum q_0 p_1}{\sum q_0 p_0} \cdot \frac{\sum q_1 p_1}{\sum q_1 p_0}}. \quad (11.11)$$

Формула, запропонована Фішером, може бути використана і для визначення індексу фізичного обсягу продукції:

$$I_{\Phi} = \sqrt{\frac{\sum q_1 p_0}{\sum q_0 p_0} \cdot \frac{\sum q_1 p_1}{\sum q_0 p_1}}. \quad (11.12)$$

Геометрична формула індексів має принциповий недолік: вона позбавлена конкретного економічного змісту. Так, на відміну від агрегатного індексу Ласпейреса або Пааше різниця між чисельником і знаменником не покаже ніякої реальної економії (або втрат) через зміну цін або фізичного обсягу продукції.

Індекс Фішера в силу складності розрахунку і труднощів економічної інтерпретації на практиці використовується досить рідко, найчастіше – при

обчисленні індексів цін за тривалий період часу для згладжування тенденцій в структурі і складі обсягу продукції, в яких відбуваються значні зміни.

В умовах ринкової економіки особливе місце серед індексів якісних показників належить індексам цін. Основним призначенням індексів цін є оцінка динаміки на товари і послуги. Крім того, індекс цін виконує роль вимірювача інфляції при макроекономічних дослідженнях, використовується при коригуванні законодавчо встановленого мінімуму оплати праці, встановленні ставок податків, перерахунку основних макроекономічних показників системи національних рахунків (валового внутрішнього продукту, національного доходу тощо) з фактичних цін у порівнянні.

Одним з найважливіших показників в статистиці цін, які широко використовуються в економічній і соціальній політиці держави, є *індекс споживчих цін (ІСЦ)*, який інколи називають індексом вартості життя. Він використовується для оцінки інфляції, індексації доходів, перегляду урядових соціальних програм, слугує основою для підвищення мінімального розміру заробітної плати, відображає реальну купівельну спроможність грошей, які має у своєму розпорядженні населення.

Індекс споживчих цін характеризує зміну в часі загального рівня цін на товари і послуги, придбані населенням для невиробничого споживання.

Визначення розрахункової формули зведеного ІСЦ пов'язане із задачею вибору між вагами поточного і базисного періодів. Зручність розрахунку індексу цін за формулою Ласпейреса зробили її основною формулою для розрахунку індексу споживчих цін тобто визначається відношення вартості споживчого кошика базисного періоду в поточних цінах до її вартості в базисному періоді.

Питання для повторення і самоконтролю

1. Що характеризує індекс у статистиці і які Ви знаєте сфери його застосування?
2. За якими ознаками класифікуються індекси?
3. Які індекси називаються індивідуальними, загальними та груповими?

4. Які індекси називаються базисними і ланцюговими та як вони обчислюються?
5. Що таке інтегральний індекс динаміки, коли використовується і як він обчислюється?
6. Які індекси називаються агрегатними?
7. Як обчислюються загальні індекси цін і фізичного обсягу? Як вони зв'язані між собою?
8. Які індекси називаються середньозваженими і як вони утворюються?
9. Як обчислюються абсолютні прирости від зміни цін та зміни кількості виробленої або реалізованої продукції?
10. Як виконується факторний аналіз середньої величини складного явища?
11. Які залежності використовують для перевірки правильності обчислення абсолютних приростів складного явища?
12. Що таке індекс змінного складу та як він обчислюється?
13. Що таке індекс постійного складу та як він обчислюється?
14. Що таке індекс структурних зрушень та як він обчислюється?
15. Що характеризує індекс Ласпейреса та як він обчислюється?
16. Що характеризує індекс Пааше та як він обчислюється?
17. Який вигляд має «ідеальна» формула Фішера і який вона виражає зв'язок з індексами Ласпейреса і Пааше?
18. Що таке індекс споживчих цін?

Лекція 12. Вибірковий метод

План

- 12.1. Суть, значення і переваги вибіркового спостереження, способи формування вибірок.
- 12.2. Обчислення похибок вибірки та визначення меж інтервалу для середньої величини і частки.
- 12.3. Визначення необхідного обсягу вибірки.
- 12.4. Способи поширення вибірових даних на генеральну сукупність.

12.5. Застосування вибірових спостережень у соціально-економічних дослідженнях.

12.1. Суть, значення і переваги вибірового спостереження, способи формування вибірок

Вибіркове спостереження – це вид несучільного спостереження, при якому обстежуються не всі елементи сукупності, а лише певним чином дібрана її частина. Сукупність, з якої вибирають елементи для обстеження, називається *генеральною сукупністю*, а сукупність, яку безпосередньо обстежують, – *вибірковою сукупністю* або *вибіркою*. Якщо до вибірки входить n елементів, то число n називається *обсягом вибірки*. *Кінцева мета вибірового спостереження* – поширення його статистичних характеристик на генеральну сукупність. Іншими словами вибіркова сукупність *репрезентує* (представляє) генеральну сукупність.

Переваги вибірового спостереження полягають в тому, що воно забезпечує, у порівнянні з суцільним спостереженням, економію трудових, фінансових ресурсів та часу. При обстеженні частини генеральної сукупності зменшуються помилки реєстрації, але можлива похибка репрезентативності. При вивченні деяких соціально-економічних явищ єдине можливе спостереження – це вибіркове. Наприклад, перевірка якості, яка пов'язана із руйнуванням продукції, перевірка чистоти та вологості зерна, наукове вивчення біологічних ресурсів тощо.

Вибіркове спостереження буде мати оптимальний результат лише тоді, коли воно буде організоване і проведене згідно з науковими принципами теорії вибірового методу. Такими принципами є:

- забезпечення випадковості відбору одиниць, коли кожна одиниця сукупності, що вивчається, має однакову можливість потрапити до вибірки;
- забезпечення достатньої кількості одиниць, що відбираються, тобто чим більше одиниць обстежено, тим точніше отримана вибірка представляє генеральну сукупність.

Найважливішою умовою при формуванні вибірки є забезпечення її *репрезентативності*, тобто при будь-якому способі відбору елементів із генеральної сукупності одержана вибірка повинна якомога точніше її представляти.

Керуючись першим принципом теорії вибіркового методу, відбір одиниць з генеральної сукупності здійснюють різними способами. Визначальним в цій ситуації буде характер генеральної сукупності. Перша передумова забезпечення репрезентативності результатів вибіркового спостереження — наукове обґрунтування і правильний вибір основи вибірки. В статистиці розглядають: просту випадкову, механічну, типову, серійну, комбіновану, багатоступінчасту, багатофазну, просякнуту вибірки.

Класичним способом формування вибіркової сукупності є *простий випадковий відбір*, в основі якого покладено жереб або таблиці випадкових чисел.

Механічна вибірка: вся генеральна сукупність механічно (без наявності суттєвої ознаки) підрозділяється на частки, і з кожної частки відбирається одна одиниця. Наприклад, треба скласти вибірку сукупність з 300 одиниць, чисельність генеральної сукупності — 3000 одиниць, тоді в вибірку повинна потрапити кожна десята одиниця.

Типова вибірка: вся генеральна сукупність підрозділяється на якісні й однорідні за суттєвою ознакою групи, а потім з цих груп проводять випадковий відбір. Наприклад, бюджетне або соціологічне дослідження сімей робітників і службовців.

Серійна вибірка: у випадковій послідовності відбирають не одиниці, а групи одиниць, які треба обстежити, а потім ці групи підлягають суцільному спостереженню. Наприклад, відбір випадкової частини продукції, виготовленої на підприємстві.

Комбінована вибірка: комбінація декількох видів вибірки. Використовується тоді, коли треба з найменшими трудовими та грошовими витратами мати найбільше представництво генеральної сукупності у вибірковій.

Багатоступінчаста вибірка: вибіркова сукупність формується по ступенях відбору. З одиниць сукупності початкового ступеня проводять поступовий відбір необхідної кількості одиниць на наступних ступенях відбору, а на останньому — формується вибіркова сукупність, одиниці якої підлягають обстеженню. Наприклад, вивчення рівня доходів сімей робітників у країні, області, місті, районі місті, підприємстві.

Багатофазна вибірка: на основі початкового формування вибіркової сукупності утворюється підвибірка, а з неї — наступна підвибірка. Наприклад, вивчаючи забезпеченість підприємств міста основними фондами, спочатку вивчають їх наявність, потім номенклатуру, зношування тощо.

Просякнута вибірка — це дві або більше незалежних вибірки з однієї і тієї ж генеральної сукупності, які утворені одним і тим же видом та способом. Наприклад, вивчення діяльності підприємства різними групами спеціалістів.

12.2. Обчислення помилок вибірки та визначення меж інтервалу для середньої величини і частки

Проводячи вибірконе спостереження за середнім розміром тієї чи іншої ознаки одиниць сукупності, або питомою вагою визначених одиниць, виникають розбіжності з існуючою ситуацією, що призводить до виникнення *помилки реєстрації та репрезентативності*.

Розпізнають *середню та граничну помилки вибірки*. Під середньою помилкою розуміють розбіжність між середньою вибірковою та середньою генеральною сукупністю ($\tilde{x} - \bar{x}$), яке не перевищує $\pm \sigma$.

Середня помилка вибіркової середньої має вигляд:

1) при повторному способі відбору:

$$\mu = \sqrt{\frac{\sigma_{\text{в}}^2}{n}}, \quad (12.1)$$

де μ — середня помилка вибіркової середньої;

$\sigma_{\text{в}}^2$ — дисперсія випадкової сукупності;

n — обсяг вибірки;

б) при безповторному способі відбору:

$$\mu = \sqrt{\frac{\sigma_g^2}{n} \times \left(1 - \frac{n}{N}\right)}, \quad (12.2)$$

де N — чисельність генеральної сукупності.

Гранична помилка вибірки — це максимальна помилка при заданій ймовірності її появи.

Гранична помилка вибірки розраховується як:

$$\Delta_x = t \times \mu, \quad (12.3)$$

де t — *коефіцієнт довіри*, який залежить від значення ймовірності.

Значення t при заданій ймовірності p знаходиться за таблицею значень функції Лапласа, що відображає залежність між t та ймовірністю p .

Наприклад, якщо $p = 0,3413$, то $t = 1$; якщо $p = 0,4772$, то $t = 2$; якщо $p = 0,4987$, то $t = 3$.

Гранична помилка вибірки дозволяє визначити граничні значення (межі) характеристик генеральної сукупності при заданій ймовірності та їх довірчі інтервали.

Довірчий інтервал для середньої:

$$\tilde{x} - \Delta_x \leq \bar{x} \leq \tilde{x} + \Delta_x. \quad (12.4)$$

Довірчий інтервал для частки:

$$\tilde{d} - \Delta_d \leq \bar{d} \leq \tilde{d} + \Delta_d \quad (12.5)$$

Якщо проводять вибірконе спостереження і при цьому розглядається альтернативна ознака, то випадкова помилка вибірки для її частини визначається за теоремою Я. Бернуллі:

$$\mu = \sqrt{\frac{qp}{n}}. \quad (12.6)$$

Граничні помилки вибірки і можливі межі альтернативної ознаки визначаються аналогічно за наведеними раніше формулами.

12.3. Визначення необхідного обсягу вибірки

У процесі проектування вибіркових спостережень визначають *мінімально достатній обсяг вибірки*, при якому вибіркові оцінки репрезентували б основні властивості генеральної сукупності. Занадто великий обсяг вибірки потребує

зайвих витрат, а занадто малий призведе до збільшення помилки репрезентативності. Оптимальний обсяг вибірки розраховують:

- для повторного відбору:

$$n = \frac{t^2 \sigma^2}{\Delta^2}; \quad (12.7)$$

- для безповторного:

$$n = \frac{t^2 \sigma^2 N}{\Delta^2 N + t^2 \sigma^2}. \quad (12.8)$$

Якщо граничну помилку вибірки замінити коефіцієнтом варіації помилок, то:

$$n = \frac{V_x^2 t^2}{V_\Delta^2}, \quad (12.9)$$

де V_x - коефіцієнт варіації ознаки X .

При вивченні альтернативної ознаки (частки p) обсяг вибірки визначається як:

$$n = \frac{t^2 pq}{\Delta^2}; \quad (12.10)$$

Визначення обсягу вибірки передбачає своєчасне обґрунтування допустимого розміру граничної помилки. Певні труднощі викликає відсутність даних про варіацію ознаки. Їх долають, використовуючи наближені оцінки варіації, обчислені за наслідками аналогічних досліджень або попередніх пробних спостережень. Для альтернативної ознаки використовують максимальне значення дисперсії $2\sigma = 0,25$ ($2\sigma = pq = p(1-p) = 0,5 \cdot 0,5 = 0,25$).

Наприклад, необхідно визначити чисельність вибірки, яка б забезпечила оцінку потенційної плинності працівників з точністю до 2% при ймовірності 0,954. Дані аналогічних досліджень свідчать, що коефіцієнт варіації плинності складає 30%.

Необхідний обсяг вибірки визначають за формулою (12.9). Відносна гранична помилка складає 2%, або 0,02; $t = 2$. Таким чином:

$$n = \frac{(0,3)^2 2^2}{(0,02)^2} = 900,$$

тобто передбачену точність вибіркової оцінки при ймовірності 0,954 забезпечить вибірка сукупність, обсяг якої 900 елементів.

Часто одночасно вивчають декілька ознак. Якщо бажаний ступінь точності визначати для кожної ознаки, то в результаті буде отриманий ряд значень обсягу вибірки. Для їх узгодження використовують або максимальний обсяг n (тоді ряд ознак оцінюється «надточно»), або обсяг головної ознаки.

12.4. Способи поширення вибірових даних на генеральну сукупність

Кінцева мета будь-якого вибіркового спостереження — поширення його характеристик на генеральну сукупність. Існують два способи розповсюдження даних вибіркового спостереження: 1) спосіб прямого перерахування; 2) спосіб коефіцієнтів.

Спосіб прямого перерахування застосовують в тому випадку, коли на основі вибірки розраховують об'ємні показники генеральної сукупності, використовуючи для цього вибіркові середню або частку. В першому випадку середній розмір ознаки, визначений в результаті вибіркового спостереження, множиться на кількість одиниць генеральної сукупності.

Встановивши, наприклад, в результаті вибіркового спостереження продуктивності тваринництва, яке перебуває в особистій власності, що середній річний надій молока на одну корову в області становить 2500 кг, і знаючи, що всього в області в особистій власності є 20000 корів, можемо отримати величину валового надою молока: $2500 \times 20000 = 50\,000\,000$ кг або 50 тис. т. Якщо при цьому відомо, що середня помилка вибірки з певною ймовірністю дорівнює ± 20 кг, а отже, генеральна середня з тією ж ймовірністю коливається в межах від 2480 до 2520 кг, то загальний валовий надій молока з врахуванням помилки вибірки буде коливатися від 49,6 тис. т до 50, 4 тис. т.

Способом прямого перерахування можна поширювати дані вибіркового спостереження також на основі співвідношення чисельності вибіркової і генеральної сукупностей. Так, наприклад, при 2 %-му вибіркового обстеженні якості продукції у вибірку потрапило 100 одиниць, з яких 20 одиниць

виявились неякісними. Помилка вибірки склала ± 2 одиниці. Використовуючи співвідношення чисельності вибіркової і генеральної сукупностей ($2\% = 1/50$), можна встановити, що число бракованих одиниць продукції для всієї партії, яка складається з 5000 одиниць, буде складати 1000 ± 100 ($20 \times 50 \pm 2 \times 50$), тобто коливатиметься в межах від 900 до 1100 одиниць.

Спосіб поправочних коефіцієнтів застосовується в тих випадках, коли вибіркоче спостереження проводиться з метою перевірки і уточнення результатів суцільного спостереження. В даному випадку, співставляючи дані вибіркового спостереження із суцільним, вираховують поправочний коефіцієнт, який використовують для внесення поправок у матеріали суцільного спостереження. Так, по закінченні переписів худоби проводять 10%-не контрольне вибіркоче обстеження худоби, яка знаходиться в особистій власності населення, і якщо в результаті контрольного обходу виявляється недооблік, тоді дані перепису коригуються на процент недообліку.

Припустимо, що в результаті перепису худоби, яка знаходиться в особистій власності населення району є 15000 свиней, в тому числі в населених пунктах, де проводиться 10 % – на вибірка, – 1200 свиней. В результаті контрольних обходів в цих же населених пунктах було обчислено 1215 свиней. Тобто, при переписі було недообліковано 15 голів свиней, що складає 1,25 % (15 від 1200). За допомогою цього поправочного коефіцієнта уточнюють матеріали перепису суцільного спостереження. В даному випадку по району всього недообліковано $0,0125 \times 15000 = 187$ голів свиней.

Отже, загальна кількість свиней, які знаходяться в особистій власності населення району з поправкою на недооблік, складає 1387 голів ($1200 + 187$).

Вибіркове спостереження в даний час знаходить все більше застосування, а тому багато питань теорії і практики вибіркового спостереження вимагають подальшого його детального вивчення і вдосконалення.

12.5. Застосування вибіркових спостережень у соціально-економічних дослідженнях

Вибіркове спостереження — один з найбільш широко застосовуваних видів спостереження. При проведенні вибіркового спостереження обстежують

не усі одиниці досліджуваного об'єкта, а лише деяку, так чи інакше відібрану частину цих одиниць. Спостереження організоване таким чином, що ця частина відібраних одиниць у зменшеному масштабі представляє всю сукупність, тобто є репрезентативною. Система правил відбору одиниць і способів характеристики досліджуваної сукупності становить зміст вибіркового методу.

Практика застосування вибірових обстежень досить різноманітна. Це обстеження домогосподарств, маркетингові дослідження, аудиторські перевірки великих фірм, вивчення суспільної думки тощо. При обстеженні невеликої частини генеральної сукупності зменшуються помилки реєстрації, можна розширити й деталізувати програму обстеження, вибіркоче обстеження забезпечує економію матеріальних, трудових, фінансових ресурсів тощо.

При вивченні певного кола соціально-економічних явищ вибіркоче обстеження є єдино можливим. Це стосується, насамперед, перевірки якості продукції (жирності молока, чистоти й вологості насіння, міцності пряжі тощо). Іноді вибіркоче обстеження проводять разом із суцільним. Наприклад, при переписі населення кожна четверта одиниця обстеження дає більш детальну інформацію. Крім того, вибіркочий метод використовується для прискорення обробки матеріалів суцільного спостереження й перевірки правильності даних переписів і одноразових обстежень.

Об'єктивною гарантією того, що вибірка репрезентативна, є суворе дотримання наукових принципів організації й проведення спостереження, насамперед, об'єктивного підходу до вибору елементів, що підлягають обстеженню.

На підставі результатів проведеної вибірки виконується аналіз досліджуваної генеральної сукупності, вивчаються взаємозв'язки між ознаками, що її характеризують, тенденції і закономірності, що мали місце в минулому. За допомогою отриманих узагальнюючих показників генеральної сукупності та аналітичних рівнянь регресії і трендів, що описують її закономірності, виконується прогнозування розвитку досліджуваного соціально-економічного явища на перспективу. Внаслідок цього складаються оптимальні плани

розвитку економіки, галузі чи підприємства, вносяться корективи в намічені раніше програми у зв'язку з появою нових тенденцій та запровадженням новітніх технологій.

Питання для повторення і самоконтролю

1. Що таке вибіркове спостереження, генеральна сукупність і вибірка?
2. В чому полягають наукові принципи теорії вибіркового методу?
3. Що таке репрезентативність вибірки і як вона досягається?
4. Які види вибірок використовуються у статистиці і в чому їхня суть?
5. Що таке середня помилка вибірки і як вона розраховується?
6. Що таке гранична помилка вибірки і як вона обчислюється?
7. Як визначаються довірчі інтервали для середньої і частки генеральної сукупності при заданій ймовірності?
8. Як визначається оптимальний обсяг вибірки при повторному і без повторному відборах?
9. В чому полягає суть способу прямого перерахування при поширенні вибірових даних на генеральну сукупність?
10. В чому полягає суть способу поправочних коефіцієнтів при поширенні вибірових даних на генеральну сукупність?
11. Як використовуються вибірккові спостереження у соціально-економічних дослідженнях.

Лекція 13. Подання статистичних даних

План

- 13.1. Поняття статистичної таблиці, її основні елементи.
- 13.2. Види статистичних таблиць, правила їх побудови.
- 13.3. Аналіз таблиць.
- 13.4. Означення та основні елементи статистичного графіка.
- 13.5. Види статистичних графіків.
- 13.6. Проблема вибору необхідних видів графіків. Аналіз графіків.

13.1. Поняття статистичної таблиці, її основні елементи

Результати статистичного зведення і групування, як правило, оформляються у вигляді статистичних таблиць.

Статистичні таблиці – це форма систематизованого, раціонального і наочного викладення статистичних даних про явища і процеси суспільного життя. Не всяка таблиця статистична. Таблиця множення, опитувальний лист соціологічного обстеження та ін. можуть носити табличну форму, але не є статистичними таблицями.

Таблична форма – це розташування числової інформації, при якій число розташовується на перетині чітко сформульованого заголовка по вертикальному стовпчику, який називається графою, і назви по відповідній горизонтальній смужі – рядку.

Значення статистичних таблиць полягає в тому, що вони дають змогу охопити матеріали статистичного зведення в цілому. Статистична таблиця по суті є системою думок про досліджуваний об'єкт, що викладається за допомогою чисел – об'єктивних статистичних показників.

Пріоритетом статистичних таблиць є виразність, наочність і компактність. Змістом статистичної таблиці є та сукупність відомостей, яка викладена в системі показників. Являючись підсумком статистичного спостереження, зведення і групування і частково аналізу, таблиці мають велике пізнавальне, наукове і практичне значення.

За зовнішнім виглядом статистична таблиця являє собою ряд горизонтальних і вертикальних ліній, які, перетинаючись по горизонталі, утворюють рядки, а по вертикалі – графи (стовпчики, колонки), які в сукупності утворюють макет таблиці.

Статистична таблиця містить три види заголовків: 1) загальний – відображає зміст всієї таблиці (до якої території і до якого часу відносяться дані), розташовується над макетом таблиці по центру і є зовнішнім заголовком;

2) верхні – характеризують зміст граф (заголовки присудка); 3) бічні (заголовки присудка) – зміст рядків. Вони є внутрішніми заголовками.

В утворенні всередині таблиці графо-клітини записується відповідна інформація. Кожна клітина має свій певний якісний зміст і кількісну міру, властиву статистичному показнику. Складену таблицю, але не заповнену числами прийнято називати макетом таблиці (рис. 13.1).

Таблиця (вказати номер)

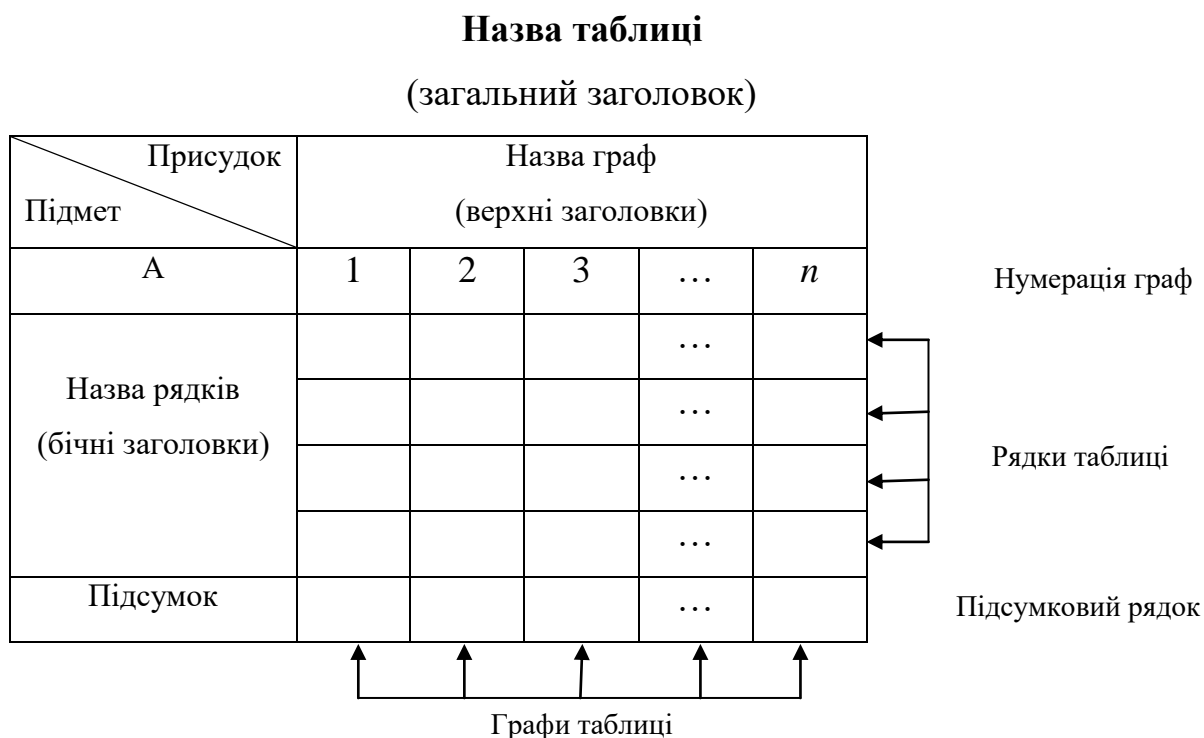


Рис. 13.1. Макет статистичної таблиці

В назвах заголовків виражається зміст і форма статистичних показників. Розробка макета статистичної таблиці є найважливішим етапом побудови статистичної таблиці, що багато в чому визначає її якість.

По суті статистична таблиця являє собою статистичне речення, яке має підмет і присудок.

Підметом таблиці є одиниці статистичної сукупності або їх групи, які підлягають характеристиці і вивченню.

Присудок таблиці — це цифрові дані, що характеризують підмет. Наприклад, в табл. 3.3, 3.4 і 3.5 підметом таблиці є групи господарств, а присудком – показники, які характеризують ці господарства (урожайність, якість ґрунту, собівартість та ін.).

Звичайно складові частини досліджуваного об'єкта, що утворюють підмет, розташовують в лівій частині таблиці, а показники, що складають присудок, розміщують справа. Але буває і обернене розташування підмета і присудка таблиці, що зумовлюється метою дослідження, характером вихідної інформації.

13.2. Види статистичних таблиць, правила їх побудови

В процесі економічних досліджень застосовуються різні види статистичних таблиць. Вони відрізняються різним числом одиниць і об'єктів, що характеризуються в них, формою підмета і присудка тощо.

Залежно від побудови (розробки) підмета розрізняють три види статистичних таблиць: *прості, групові та комбінаційні*.

Простими називають такі статистичні таблиці, в підметі яких міститься простий перелік будь-яких об'єктів, територіальних підрозділів або хронологічних дат. Відповідно таблиці можуть бути названі простими переліковними, територіальними або хронологічними.

Груповими називають статистичні таблиці, в яких статистичний підмет складається з груп, виділених за будь-якою однією суттєвою ознакою, а присудок містить ряд ознак, які характеризують зазначені групи.

Комбінаційними називають статистичні таблиці, в яких підмет являє собою комбінацію, сполучення двох або кількох ознак, а в присудку наводяться ознаки, що характеризують виділені групи і підгрупи. Комбінаційні таблиці отримують внаслідок комбінаційних групувань (див. табл. 3.6, 3.7). Комбінаційні таблиці мають дуже велике аналітичне значення. Вони дають змогу за допомогою комбінування різних групувальних ознак найбільш правильно охарактеризувати вплив окремих факторів на результативні показники. Вирівнюючи сукупність в певних межах за однією ознакою і диференціюючи за другою і т. д., комбінаційні таблиці дають змогу не тільки встановити наявність зв'язку, але й виміряти ступінь цього зв'язку.

Залежно від завдання дослідження і характеру інформації *присудок* статистичних таблиць буває *простим і складним*. Показники присудка при

простій розробці застосовуються послідовно один за другим. Розподіляючи показники на групи за однією або кількома ознаками в певному сполученні, одержують складний присудок.

При розробці і заповненні макетів таблиць необхідно строго дотримуватися правил їх побудови.

1. Статистичні таблиці не повинні бути надмірно громіздкими і ускладненими, вони повинні полегшувати, а не утруднювати їх аналіз. В зв'язку з цим по можливості таблицю слід скласти невеликою за розміром, легко доступною для огляду. Інколи доцільно замість однієї великої таблиці побудувати декілька зв'язаних між собою, послідовно розташованих таблиць.

2. Всі таблиці можуть бути пронумеровані арабськими цифрами. Номер таблиці вказують перед її заголовком. При цьому знак № не пишуть.

3. Кожна таблиця повинна мати загальний заголовок, в якому коротко і ясно відображається основний зміст таблиці, вказано до якої території і до якого періоду або моменту часу відносяться дані, що наведені в ній. Вимога точності, чіткості та ясності відноситься і до заголовків рядків і граф.

4. Показники таблиці обов'язково повинні супроводжуватись одиницями вимірювання. Якщо для всіх показників використовується одна одиниця вимірювання, то її пишуть в кінці заголовка таблиці, а якщо їх кілька – в кінці рядків або граф. Одиниця вимірювання відокремлюється від назви показника комою.

5. Слова в таблиці пишуться повністю. Можна використовувати тільки загальноприйняті скорочення.

6. Таблиці, як правило, мають бути замкненими, тобто мати підсумкові результати (в цілому, по групах і підгрупах).

7. При заповненні таблиць потрібно використовувати такі умовні позначення: при відсутності явища пишеться прочерк (-), якщо ж немає інформації про явище, ставиться три крапки (...) або пишеться "немає відомостей", в тих випадках, коли клітинка не підлягає заповненню в зв'язку з відсутністю осмисленого змісту – ставиться знак множення (x). При наявності

інформації по досліджуваному явищу, числові значення якого складають величину менше критичної в таблиці точності, прийнято записувати 0,0.

8. Однаковий ступінь точності, обов'язкова для всіх чисел, забезпечується дотриманням правил їх заокруглення. Всі значення однойменних показників мають бути записані з однаковим ступенем точності (до цілих, до 0,1, до 0,01 і т.д.).

9. Коли одна величина перевищує другу багатократно, то отримані показники динаміки краще виражати не в процентах, а в разях (коефіцієнтах). Наприклад, замість 288 % слід написати "в 2,9 разів більше". В аналітичних таблицях значність абсолютних цифр має бути найменшою. Тому великі числа необхідно заокруглювати до тисяч, мільйонів і т.д. Наприклад, замість числа 1200000 грн. краще написати 1,2 млн грн.

10. Якщо в таблиці поряд із звітними даними наводяться відомості розрахункового порядку, то про це слід зробити відповідне застереження. По можливості ці пояснення краще зробити в самій таблиці або в заголовку до неї.

11. Якщо є потреба до таблиці можуть бути застосовані і виноска. Примітки даються у вигляді необхідності додаткових пояснень змісту окремих показників таблиці. У виносках звичайно вказують джерела одержаних у таблиці відомостей.

13.3. Аналіз таблиць

Аналіз і читання даних статистичних таблиць має велике науково-пізнавальне і практичне значення. Перед тим як приступити до аналізу даних таблиць, слід ознайомитися з її назвою, заголовками граф і рядків і з'ясувати їх суть, цифрові дані необхідно починати читати з підсумків і тільки після цього переходити до аналізу даних окремих рядків і граф, тобто до оцінки частин досліджуваного об'єкта, вивчаючи при цьому спочатку важливі, а потім вже і решту елементів таблиці. Аналіз таблиць полягає в аналітичному осмисленні і тлумаченні табличних даних і він спрямований на виявлення взаємозв'язків і взаємозалежностей між ознаками. Аналіз даних в статистичних таблицях часто

вимагає виконання додаткових розрахунків різних відносних, середніх величин, додаткових індексів, коефіцієнтів та інших показників.

13.4. Означення та основні елементи статистичного графіка

Поряд з таблицями для характеристики результатів статистичного зведення та обробки масових даних широко застосовують статистичні графіки.

Статистичним графіком називають наочне масштабне зображення статистичних даних за допомогою геометричних ліній, точок, знаків, фігур, географічних картосхем та інших графічних засобів.

Графічний метод настільки міцно ввійшов в арсенал засобів наукового узагальнення і в методику наукових досліджень, що сучасну науку неможливо собі уявити без його застосування. Особливо велика роль цього методу в статистичних дослідженнях, де вивчаються складні взаємозв'язки, тенденції, закономірності соціально-економічних явищ і процесів в динаміці і просторі.

Застосування графічного методу у вивченні масових соціально-економічних явищ досить різнопланове. Так, графіки застосовують для характеристики змін суспільних явищ і процесів у часі, вивчення структури явищ, порівняння, контролю виконання плану, дослідження взаємозв'язків між результативними і факторними ознаками, зображення розміщення явищ у просторі, ступеня розповсюженості по території тих чи інших явищ, міжнародних порівнянь і зіставлень та в інших випадках.

Особливо велика роль графіків у пропаганді передового досвіду, прогресивних тенденцій, закономірностей, нових технологій, наукових досягнень тощо.

В ряді випадків графіки є незамінним засобом аналізу, дослідження і виявлення взаємозв'язків, закономірностей і тенденцій суспільних явищ (наприклад, в кореляційному аналізі і динамічних рядах).

Статистичні графіки застосовуються для того, щоб зробити статистичні матеріали наочними, доступними, зрозумілими і такими, які б сприяли кращому їх аналізу. Завдання полягає в тому, щоб у кожному випадку вибрати

найкраще графічне зображення, яке б відповідало характеру величин і повніше розкривало їх зміст.

Основні елементи графіка такі: *поле графіка, геометричні знаки, просторові орієнтири, масштаб, експлікація графіка.*

Поле графіка – простір, в якому розміщуються геометричні знаки, що утворюють графік. Він характеризується форматом і співвідношенням сторін.

Розмір графіка повинен відповідати його призначенню. Для демонстрації на лекції або докладі застосовуються графіки великих форматів, для ілюстрації наукового звіту або для розміщення в книзі, статті, курсовому проєкті (роботі) – невеликі графіки.

Суттєвим фактором забезпечення найкращого зорового сприйняття відображуваних статистичних даних є вибір пропорцій співвідношення сторін графіка. Співвідношення сторін графіка визначається законами геометричної гармонії і вимогами забезпечення неспотвореного зорового сприйняття графічного образу. Як показує практика побудови і аналізу графіків найбільш зручні формати з співвідношенням сторін (ординат і абсцис) від 1:1,3 до 1:1,5. Водночас це не заперечує можливості застосування квадратної форми графіків, яка в окремих випадках є дуже зручною формою відображення статистичних даних.

Геометричні знаки – це сукупність геометричних чи інших графічних знаків, за допомогою яких відображаються статистичні дані і створюється графічний образ. Це точки, прямі і криві лінії та їх відрізки, площини (кола, квадрати тощо), об'ємні фігури (куби, кулі тощо), геометричні фігури (знаки – символи, зображення предметів). Геометричні знаки становлять основу графіка, його мову. Залежно від типу геометричних знаків графіки поділяють на *лінійні, точкові, стовпчикові, стрічкові, квадратні, кругові, секторні, фігурні та ін.*

Важливим моментом побудови графіків є вибір графічного знака. Вибір графічного знака визначається характером вихідної інформації, а також основною метою, яка закладена в даний графік. Вдалиий його вибір сприяє його максимальному досягненню мети графіка і найбільш виразному зображенню

статистичних даних. Так, наприклад, якщо метою дослідження є вивчення обсягу будь-якого виду продукції в динаміці, то вихідні дані можуть бути зображені як у вигляді стовпчикової, кругової, квадратної та інших діаграм, так і за допомогою лінійної діаграми. Для відображення обсягу виробництва продукції доцільно використати площинну (стовпчикову, кругову, квадратну та ін.) діаграму, а для відображення динаміки – лінійну діаграму.

Просторові орієнтири визначають розташування геометричних знаків у полі графіка. Вони задаються у вигляді координатних сіток (діаграми) або контурних ліній (картограми). Координатна сітка створюється перетином ліній, які проходять через поділки горизонтальної та вертикальної шкали. Для побудови графіка, як правило, використовується система прямокутних (декартових) координат, зокрема, права верхня частина координатного поля, але нерідко зустрічаються графіки, які побудовані за принципом полярних координат (кругові, секторні, радіальні та інші діаграми). Криволінійні контурні лінії застосовують в статистичних картах (картограмах, картодіаграмах) як засіб просторової орієнтації.

На горизонтальній шкалі (вісь абсцис) прямокутних діаграм, як правило, відкладають незалежні змінні (часові відрізки, періоди, об'єкти тощо), на вертикальній (вісь ординат) – залежні змінні (наприклад, значення результативних показників).

На координатній сітці графіка обов'язково повинна бути вказана основна горизонтальна (нульова) лінія (вісь абсцис). Для наочності її виділяють потовщеною лінією. Якщо рівні відображуваних явищ такі, що основна частина координатної сітки залишається невикористаною, то на шкалі робиться розрив, який виключає непотрібну частину сітки, але з обов'язковою вказівкою нульової лінії. Це дасть змогу рівномірніше заповнити поле графіка.

Одним з основних елементів графіка є масштабна шкала графіка, тобто лінія, окремі крапки чи риски, якої можуть бути прочитані як певні числа.

Масштабні шкали, як правило, розміщуються зліва і знизу графіка. На шкалах повинен розміщуватися весь діапазон зображуваних цифрових даних,

звичайно з деякими заокругленнями. Якщо, наприклад, максимальна величина урожайності в досліджуваній сукупності становить 48,5 ц/га, то очевидно, що на шкалі мають бути поділки, що містять 49 або 50 ц/га. Тому останнє число на шкалі має дещо перевищувати максимальний рівень ознаки.

Масштабна шкала складається з трьох елементів: 1) лінії, які є носієм чи опорою шкали; 2) поділок або позначок шкали (точки або риски, які розміщені в певному порядку на носії шкали); 3) цифрові позначення чисел, що відповідають певним точкам або рискам.

Носіями шкали можуть бути пряма лінія (осі координат) або крива лінія (коло, дуга).

Довжину відрізка між двома сусідніми поділками називають графічним інтервалом, а різницю між числовими значеннями цих поділок – числовим інтервалом.

Масштабні шкали можуть бути прямолінійними, криволінійними, неперервними, перервними, рівномірними і нерівномірними.

Прямолінійними називають шкали, в яких пряма лінія поділена на сантиметри і міліметри, криволінійними, в яких крива лінія (коло) поділена на 3600 частин.

Неперервна шкала застосовується для величин, що безперервно змінюються, (всім точкам відповідає певне число, а усі проміжні значення можуть бути інтерпольовані). Перервна шкала – шкала з величинами, проміжне значення яких не інтерпольується (наприклад, якщо поділки шкали представлені річними даними, то точка між двома роками нічого не означає, так як масштаб не передбачав місячних даних).

Рівномірною (арифметичною) називається шкала, в якій рівним відрізкам (поділкам) на шкалі відповідають рівні числові значення. В рівномірній шкалі графічні матеріали пропорційні абсолютним розмірам статистичних показників. Так, якщо значення показника зростає у два рази, то відрізок, що її відображає повинен відповідно збільшуватись у два рази. Такі шкали мають переважне застосування в статистичних графіках.

Експлікація графіка – це словесне тлумачення його змісту. Вона включає назву графіка, написи вздовж масштабних шкал і змістовних значень застосовуваних геометричних знаків.

Графіки можуть супроводжуватися умовними позначеннями, що розкривають зміст застосованих геометричних знаків. Пояснення до вертикальних і горизонтальних шкал повинні розкрити зміст показників, що відображаються, одиниці їх вимірювання. Статистичний графік – це знакова модель, без експлікації його неможливо прочитати і зрозуміти, тобто перенести значення з формалізованої системи характеристики дійсності на саму дійсність.

13.5. Види статистичних графіків

Статистичні графіки відрізняються великою різноманітністю. Залежно від способу побудови їх можна поділити на дві великі групи: 1) діаграми і 2) статистичні карти.

Діаграми – це умовне зображення числових величин та їх співвідношень за допомогою геометричних знаків. Термін "діаграма" тотожний терміну "статистичний графік". Діаграми є найбільш розповсюдженим видом графіків. Виділяють такі основні види діаграм: *лінійні, стовпчикові, стрічкові, секторні, радіальні, трикутні, фігурні, знак Варзара та ін.*

Залежно від кола розв'язуваних завдань усі діаграми можна поділити на *діаграми порівняння, структури та динаміки.*

Розглянемо методику і техніку побудови статистичних графіків, що найчастіше застосовуються на практиці.

Найрозповсюдженішим видом показових діаграм є *лінійні діаграми*, які використовуються здебільшого для характеристики динамічних рядів та рядів розподілу. Поряд з цим лінійні діаграми широко використовуються для вивчення взаємозв'язків між явищами, порівняння кількох показників, ходу виконання планів тощо.

Лінійні діаграми дають можливість зображати явища у вигляді ліній, які з'єднують точки, розташовані у координатному полі. Ламані лінії, що

утворюються, показують характер розвитку явища у часі або особливості його розподілу за величиною якої-небудь ознаки, або зв'язку явищ.

За способом побудови – це графіки з рівномірною (арифметичною) шкалою. При їх побудові використовують прямокутну систему координат. Розташування будь-якої точки в цій системі визначається двома параметрами – абсцисою та ординатою. Іноді поле в межах осей координат для зручності нанесення геометричних знаків та читання графіка покривається горизонтальними і вертикальними лініями, проведеними за прийнятим масштабом. Ці лінії утворюють координатну числову сітку.

На горизонтальній осі (вісь абсцис) відкладають однакові за довжиною відрізки, що відображають періоди (роки, місяці, декади, дні і т.д.). На вертикальній осі (вісь ординат) у певному масштабі наносять значення досліджуваної величини. На перетині перпендикулярів відповідних значень досліджуваної ознаки і часових дат до осей координат отримують точки.

Ламана лінія, яка з'єднує ці точки, характеризує зміну досліджуваного явища у часі (табл. 13.1, рис. 13.2).

Таблиця 13.1

Динаміка виробництва молока в агрофірмі за 2001-2010 рр.

Роки	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Валове виробництво молока, тис. ц	27,0	27,9	26,5	28,0	29,7	32,0	31,0	33,5	34,2	37,2

Зауважимо, що у переважній більшості випадків для побудови графіків використовуються стандартні комп'ютерні програми, тому основне завдання полягає в тому, щоб вибрати вид графіка, занести масиви даних з таблиці 13.1 у програму Excel, правильно вибрати поле графіка, масштаби осей координат і вказати їхні назви.

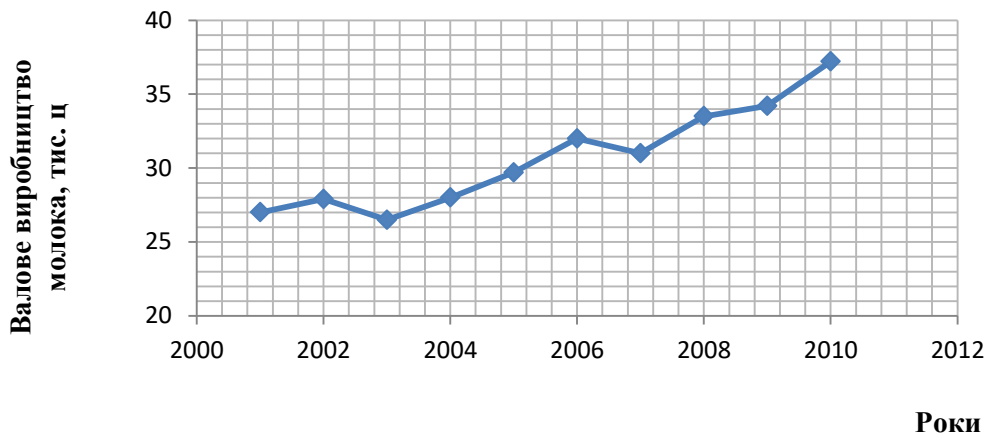


Рис. 13.2. Динаміка валового виробництва молока в агрофірмі (2001-2010 рр.)

Побудований графік показує постійне зростання виробництва молока, оскільки ламана лінія виражає цю загальну тенденцію, не дивлячись на деякі коливання.

Лінійні діаграми можуть бути побудовані з метою вивчення взаємозв'язків між двома ознаками: результативною і факторною (наприклад, між урожайністю і якістю ґрунтів). При цьому на осі абсцис відкладають значення факторної ознаки (якості ґрунтів), а на осі ординат – значення результативної ознаки (урожайності).

Лінійні діаграми зручні для зображення кількох паралельних рядів з метою їх порівняння (наприклад, динаміки продуктивності корів і рівня годівлі або інших якісно відмінних ознак). У цьому разі будують дві (при двох ознаках) або кілька шкал. Другу шкалу будують праворуч.

Особливе місце мають лінійні діаграми зі спеціальними базовими лініями. Найбільш типовими є два випадки. В першому випадку значення вертикальної шкали на початку координат приймають за 100%, тобто лінія, що виходить з цієї точки, відображає рівень базисної величини, яка дорівнює 100%. Всі значення величин, які перевищують базисну, розташовують вище цієї лінії, а значення, які менш рівня базисної величини, розташовують нижче.

У другому випадку при зображенні відхилень від середнього значення рівня (частіше у процентах) базова лінія, що характеризує середній рівень, є нульовою. Додатні відхилення (перевищення) від середнього рівня відкладають вище цієї лінії, від'ємні – нижче.

Діаграми у вигляді вертикальних стовпчиків і стрічок є найбільш простими і досить ефективними для аналізу соціально - економічних явищ видом графічного зображення.

Стовпчикові та стрічкові діаграми переважно застосовуються для порівняння різних показників у просторі і у часі, а також аналізу структури явищ.

Стовпчикові діаграми – це графіки, в яких різні величини представлені у вигляді стовпчиків однакової ширини, які розташовані один від одного на однаковій відстані або щільно. Такі діаграми називають інакше *гістограмами*. Якщо стовпчики розташовані не по вертикалі, а по горизонталі, то такі діаграми називаються *стрічковими*.

Основа порівняння в стовпчикових і стрічкових діаграмах – лінійна (одномірна). Висота стовпчиків і довжина стрічок, відповідно до прийнятого масштабу, пропорційна величині зображуваних явищ.

Рекомендується включення в діаграму масштабної шкали, яка дає змогу визначити висоту стовпчика і довжину стрічки. Шкала може співпадати з гранню першого стовпчика або стрічки або розташовуватися на окремій лінії зліва (в стовпчиковій діаграмі) або у верхній частині (в стрічковій діаграмі). Шкала, на якій встановлюється висота стовпчиків або довжина стрічок повинна бути безперервною і починатися з нуля. Написи і вказівки цифр в кінці стовпчиків (стрічок) робити не рекомендується, бо це може створити зорове подовження стовпчиків (стрічок). Якщо є потреба, то значення показників краще всього писати всередині стовпчиків (стрічок), або розташувати в один ряд над ними на рівні закінчення шкали по осі ординат.

Стовпчики (стрічки) для кращої наочності можуть бути зафарбовані суцільною фарбою, якщо стовпчик (стрічка) відображає ціле явище, або кількома фарбами, якщо зображуються порівняння різних структур явищ, кожному з яких відведена частина стовпчика (стрічки).

Стрічковою діаграмою можна зображувати те саме що й стовпчиковою. Однак вертикальні стовпчики кращі від стрічок, якщо числа виражають ідею

висоти (рівень зростання) і якщо пояснювальні написи до кожного стовпчика невеликі. Горизонтальні стрічки є наочнішими, якщо зображувані величини виражають ідею подовженості (автомобільних доріг, річок і т.п.) і якщо пояснювальний текст до них невеликий.

Побудову стовпчикової діаграми покажемо на прикладі (табл. 13.2).

Таблиця 13.2

Посівні площі сільськогосподарських культур у фермерських господарствах району за 2009 і 2010 рр., тис. га

Рік	Культури				Разом:
	Зернові	Технічні	Картопля та овоче- баштанні	Кормові	
2009	26,4	2,9	4,5	21,2	55,0
2010	33,1	4,0	3,3	19,6	60,0

Порівняємо за допомогою стовпчикової діаграми загальні посівні площі у фермерських господарствах району за 2009 і 2010 роки, яка відповідно становила 55,0 і 60,0 тис. га (рис. 13.3).

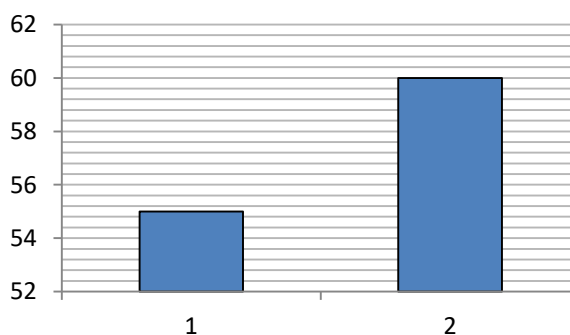


Рис. 13.3. Посівні площі у фермерських господарствах району, тис. га: 1 – 2009 р.; 2 – 2010 р.

Стовпчикова діаграма може бути використана не тільки для характеристики загального розміру, а й структури того або іншого явища. При побудові стовпчикової структурної діаграми висоту стовпчика беруть у відсотках пропорційно структурі явищ (табл. 13.3).

Структура посівних площ у фермерських господарствах району за 2009 і 2010 рр.

Культури	2009 р.			2010 р.		
	тис. га	в % до підсумку	в градусах	тис. га	в % до підсумку	в градусах
Зернові	26,4	48,0	173	33,1	55,2	199
Технічні	2,9	5,3	19	4,0	6,7	24
Картопля та овоче-баштанні	4,5	8,2	30	3,3	5,5	20
Кормові	21,2	38,5	138	19,6	32,6	117
Разом:	55,0	100,0	360	60,0	100,0	360

Щоб полегшити читання і аналіз таких діаграм, окремі складові частини розфарбовують різним кольором або штриховкою.

Використовуючи дані табл. 13.3, побудуємо плоску (а) та об'ємну (б) стовпчикову структурну діаграму (рис. 13.4).

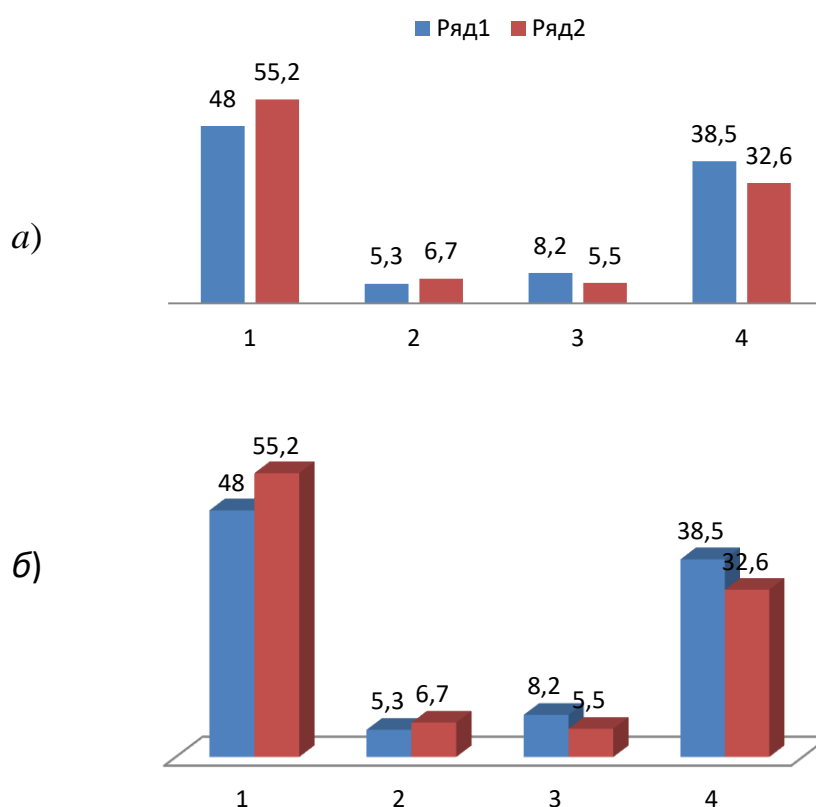


Рис. 13.4. Структура посівних площ фермерських господарств району, %:
 1 – зернові культури; 2 – технічні; 3 – картопля та овоче-баштанні; 4 – кормові.
 Ряд 1 – 2009 р.; ряд 2 – 2010 р.

Стрічкові діаграми. На відміну від стовпчикових, при побудові стрічкових діаграм прямокутники, якими зображують розмір явищ, розташовують не по вертикалі, а по горизонталі (рис. 13.5). Вимоги, що ставляться до побудови цього виду діаграм, аналогічні вимогам до стовпчикових діаграм.

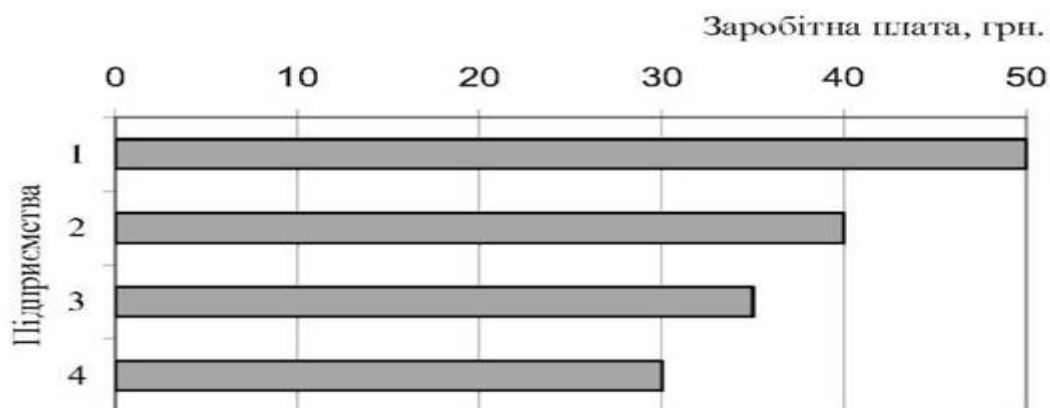


Рис. 13.5. Стрічкова діаграма денної заробітної плати на підприємствах

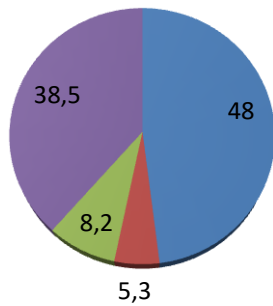
Секторна діаграма являє собою круг, розділений радіусами на окремі сектори, кожний з яких характеризує питому вагу відповідної частини в загальному обсязі зображуваної величини. Секторні діаграми використовуються переважно для характеристики структури явищ. При порівнянні різних структур загальні площі кругів беруть однаковими. Кожний сектор виділяють за кольором або штриховкою; крім того в кожному секторі нерідко дають і цифрове позначення його питомої ваги. При малому куті сектора експлікація до нього вказується стрілкою.

Секторні діаграми наочні тільки тоді, коли досліджувана сукупність поділяється не більш як на 4-5 секторів і спостерігається значна структурна диференціація. Якщо сукупність поділяється на більшу кількість секторів і структурна диференціація незначна, то для зображення структури явищ доцільно застосовувати стовпчикову або стрічкову діаграму.

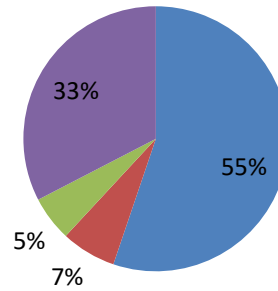
Для кращої наочності використовують також просторові (об'ємні) секторні діаграми, у яких вирізано окремі сектори. У наведеному прикладі (табл.13.3) такі діаграми – плоскі (а) та об'ємні (б), мають вигляд (рис. 13.6).

a)

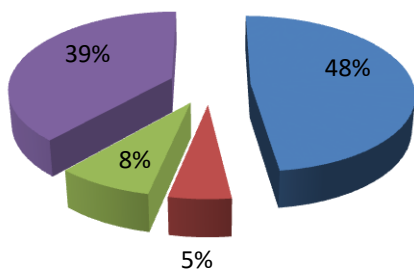
■ 1 ■ 2 ■ 3 ■ 4



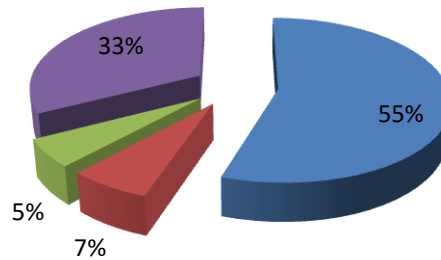
■ 1 ■ 2 ■ 3 ■ 4



б)



2009 р.



2010 р.

Рис. 13.6. Структура посівних площ у фермерських господарствах району в 2009 і 2010 рр.:

1 – зернові культури; 2 – технічні; 3 – картопля та овоче-баштанні; 4 – кормові.

Радіальні діаграми. Цей вид діаграм застосовується для графічного зображення явищ, які змінюються в замкнуті календарні строки. В основу їх побудови покладено полярну систему координат, де за вісь абсцис приймається коло, за вісь ординат – його радіуси.

Залежно від того, який зображується цикл діаграмованого явища - замкнутий чи продовжуваний (із періоду в період), – розрізняють радіальні діаграми замкнуті і спіральні. Наприклад, якщо весь цикл зміни зображуваного явища охоплює річний період, радіальну діаграму будують за формою замкнутої.

Якщо ж зміна явища вивчається впродовж циклу діаграмованого періоду (наприклад, грудень одного року сполучається з січнем другого року і т.д.), ряд динаміки зображується у вигляді суцільної кривої, яка візуально має вигляд спіралі.

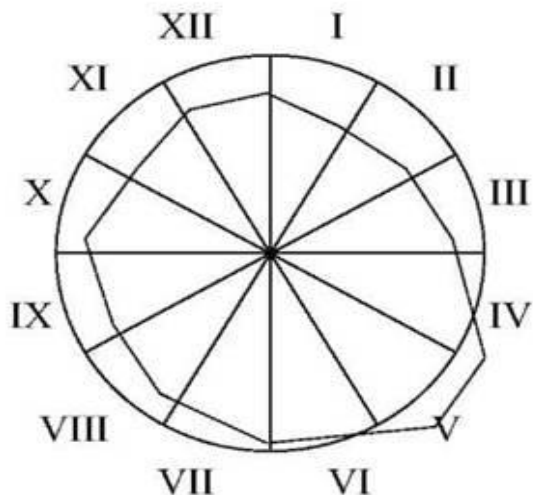


Рис. 13.7. Радіальна діаграма відпрацьованих людино-годин на підприємстві впродовж року

При побудові радіальних діаграм початком відліку (полюсом) може бути центр або окружність. Якщо за полюс прийнято центр кола, то радіальну діаграму будують у такій послідовності: коло ділять на стільки частин, скільки періодів має діаграмований цикл (наприклад, рік - 12 міс.), і будують відповідно їм радіуси (у даному випадку – 12). Періоди розміщують за годинниковою стрілкою і на кожному радіусі у масштабному вимірі відкладають відрізки (від центра кола), пропорціональні розмірам явищ. Кінці відрізків на радіусах з'єднують, у результаті чого утворюється концентрична ламана лінія. Приклад замкнутої радіальної діаграми з початком відліку від центра кола наведено на рис. 13.7.

Картограми і картодіаграми. Картограми являють собою контурну географічну карту або схему, на якій штриховкою різної густоти, крапками або фарбами різного ступеня насиченості зображена порівняльна інтенсивність будь-якого показника в межах кожної одиниці нанесеного на карту територіального поділу. На картограмах, як правило, зображують явища, що характеризуються відносними або середніми величинами (наприклад, кількість працюючих пенсіонерів у загальній кількості працюючих за регіонами, меліорованість земель у процентах до загальної площі, середня заробітна плата на підприємствах по районах області тощо).

За способом зображення діагрованих явищ розрізняють картограми крапкові і фонові.

У перших рівень явища показують за допомогою крапок, розташованих на контурній карті територіальної одиниці. Для наочності зображення щільності або частоти появи певної ознаки крапкою позначають одну або кілька одиниць сукупності.

На фонових картограмах штриховкою різної густоти або фарбою різного ступеня насиченості зображують інтенсивність будь-якого показника в межах територіальної одиниці. Один із випадків картограм зображено на рисунку 13.8.

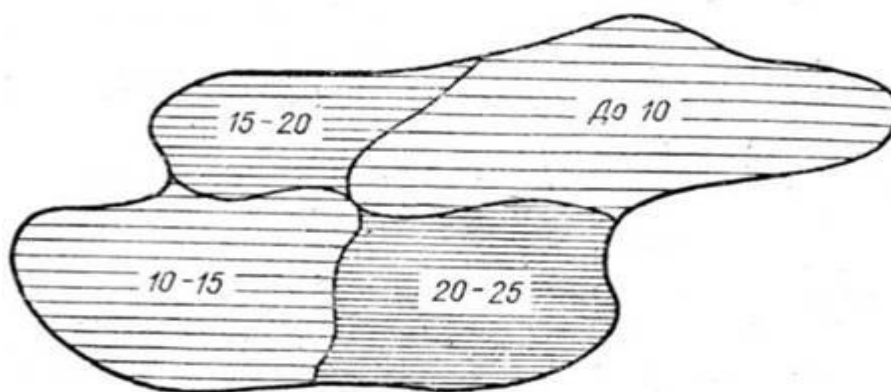


Рис. 13.8. Картограма щільності поголів'я корів на 100 га сільськогосподарських угідь у господарствах району

Якщо на контурну карту наносяться статистичні дані у вигляді діаграм, одержують картодіаграму. Яскравим її прикладом є географічна карта, на якій чисельність населення великих міст зображена у вигляді кіл різної величини.

Існують ще деякі типи статистичних графіків, які використовуються у спеціальних випадках. В разі потреби з ними можна ознайомитись у довідкових джерелах.

13.6. Проблема вибору необхідних видів графіків. Аналіз графіків

Важливим моментом побудови графіків є вибір графічного знака. Вибір графічного знака визначається характером вихідної інформації, а також основною метою, яка закладена в даний графік. Вдалиий його вибір сприяє його максимальному досягненню мети графіка і найбільш виразному зображенню статистичних даних. Так, наприклад, якщо метою дослідження є вивчення

обсягу будь-якого виду продукції в динаміці, то вихідні дані можуть бути зображені як у вигляді стовпчикової, кругової, квадратної та інших діаграм, так і за допомогою лінійної діаграми. Для відображення обсягу виробництва продукції доцільно використати площинну (стовпчикову, кругову, квадратну та ін.) діаграму, а для відображення динаміки – лінійну діаграму.

Одним з найважчих і найважливіших завдань побудови графіка є відшукання правильної його композиції, під якою розуміють поєднання всіх його елементів. Правильна композиція графіка означає: ретельний відбір з наявного цифрового статистичного матеріалу даних, що підлягають графічному зображенню; вибір виду графіка; вибір формату (розміру і співвідношення сторін) графіка; підбір масштабу та геометричних знаків і їх розміщення в полі графіка; правильне розміщення і поєднання всіх елементів графіка тощо.

Раціональне розміщення матеріалу на полі графіка створює цілісне уявлення про досліджувані явища.

Створення правильної композиції графіка повинно переслідувати головну ціль – отримати компактне, просте і логічне зображення досліджуваного явища і водночас підкреслювати ті чи інші особливості цього явища (динаміку, тенденції, зв'язки, закономірності, склад, структуру тощо)

Не менш важливим завданням композиції графіка є його художнє та естетичне оформлення. Графік повинен притягувати увагу, забезпечуючи водночас легкість його читання та засвоєння.

Щоб композиція графіка відповідала зазначеним вище вимогам, необхідно при побудові графіків виконувати певні правила.

До побудови графіків відносять ті самі вимоги, що й до побудови таблиць. Кожен графік повинен мати чітку і повну назву, що відображає зміст досліджуваного явища, час і місто показників, що наводяться.

У графіку, крім заголовка, обов'язково необхідно наводити також інший текст, в який входять назва і цифри масштабу, назва ліній, цифри, що характеризують окремі частини графіка, умовні позначення, посилання на джерела даних, одиниці вимірювання та ін. Усі пояснювальні написи і

заголовки графіка, як і в статистичній таблиці, повинні чітко, коротко і точно розкривати його зміст. Назву графіка, як правило, розміщують в нижній його частині. Написи на графіку повинні бути виконані чітко і охайно. Їх пропонується, як правило, робити горизонтально, тому що вертикальний текст менш зручний для читання. Пояснювальні написи можуть бути розміщені як на самому графіку, так і за його межами. Останній спосіб застосовують у випадках, коли не вистачає місця на полі графіка.

Масштаб на горизонтальній і вертикальній шкалах має бути оптимальним, таким що не перекручує реальне співвідношення явищ, які аналізуються. Горизонтальну шкалу (на осі абсцис) слід будувати зліва направо, а вертикальну (на осі ординат) – знизу вгору. Цифри шкали слід наносити ліворуч та знизу або вздовж осей. Якщо числові дані не включені у графік бажано їх дати окремо у формі таблиці. Нульові лінії (вертикальну та горизонтальну) рекомендується відокремлювати на графіку відмінно від усіх ліній координатної сітки. Густота координатної сітки має бути оптимальною і не ускладнювати читання графіка. У зв'язку з цим не слід перевантажувати графіки великою кількістю графічних знаків. Наприклад, на лінійних діаграмах рекомендується наносити не більше як 5 - 6 ліній; секторну діаграму не слід поділяти більше як на 4-5 секторів і т.п. Особливо не слід завантажувати графік числовою інформацією. Він потрібний для того, щоб замінювати числа, тому їх слід вписувати у графік лише у крайніх випадках (наприклад, у секторній діаграмі, де важко поглядом вловити співвідношення секторів).

Графік має бути наочним, зрозумілим, легко читатися та по можливості художньо оформленим. З цією метою лінії на графіку можуть бути зображені різним кольором або рисунком (суцільною, пунктирною, точковою, точково-пунктирною лінією).

Питання для повторення і самоконтролю

1. Яка таблиця називається статистичною і які вона має пріоритети?
2. Що таке підмет і присудок статистичної таблиці?
3. Які таблиці називаються простими, груповими і комбінованими?

4. Яких правил потрібно дотримуватись при складанні статистичних таблиць?
5. В чому полягає суть аналізу статистичної таблиці і як він проводиться?
6. Який графік називається статистичним і для чого він використовується?
7. Що таке поле і геометричні знаки статистичного графіка?
8. Що таке масштабна шкала графіка і на які типи такі шкали поділяють?
9. Що таке експлікація графіка?
10. На які групи поділяють статистичні графіки?
11. Що таке діаграми і які бувають їх види?
12. Що таке лінійна діаграма і який вона має вигляд?
13. Який вигляд мають стовпчикові та стрічкові діаграми (гістограми)?
14. Який вигляд мають секторні діаграми і коли вони використовуються?
15. Які діаграми називаються радіальними і коли вони використовуються?
16. Які умови повинні виконуватись при виборі виду графіка?

Список використаної літератури

1. ДСТУ 1.10:2005. Національна стандартизація. Правила розроблення, побудови, викладання, оформлення, ведення національних класифікаторів.
2. Курс лекцій з дисципліни «Статистика». Частина 1. Теорія статистики: В.П. Сторожук, О.В. Кустовська, Є.І. Ткач, І.М. Шост та ін.; За ред. Є.І. Ткача. – Тернопіль: Економічна думка, 2006. –224 с.
3. Лугінін О.С. Статистика: Підручник. - К.: Центр учбової літератури, 2007. - 608 с.
4. Мазуренко В.П. Теорія статистики: Навчальний посібник - К.: Видавничо-поліграфічний центр "Київський університет", 2006. - 232 с.
5. Мармоза А.Т. Практикум з теорії статистики: Навчальний посібник. - 3-е видання, виправлене. К: Ельга, Ніка - Центр, 2007. - 348 с.
6. Мармоза А.Т. Статистика: Підручник. - К: Ельга, КНТ, 2009. -896 с.

7. Матковський С.О., Гальків Л.І., Гринькевич О.С, Сорочак О.З. Статистика: Навчальний посібник - Львів.: "Новий Світ", 2009. - 430 с.
8. Общая теория статистики: Учебник / А.И. Харламов, ОЭ. Башина, В.Т., Бабурин и др. Под ред. А.А. Спирина, ОЭ. Башиной. - М.: Финансы и статистика, 1996. - 296 с.
9. Парфенцева Н. – Міжнародні статистичні класифікації в Україні: впровадження й використання. Навчальний посібник. – К.: Основи, 2000. –351 с.
10. Плошко Б.Г., Єлисеєва І.І. Історія статистики: Навч. посібник. – Москва-Ленінград: Фінанси й статистика, 1990.
11. Практикум по теории статистики: Учебное пособие / Под ред. Р.А. Шмойловой. - М.: Финансы и статистика, 1998. - 416 с.
12. Статистика. Курс лекцій /Харченко Л.П., Долженкова В.Г., Ионин В.Г. и др.; Под ред. В.Г. Ионина. - Новосибирск: Издательство НГАЭиУ. - М.: ИНФРА-М, 2000. - 310 с.
13. Статистика. Підручник / С.С. Герасименко та ін. - К.: КНЕУ, 2000. - 467 с.
14. Статистика: Учебник / Под ред. И.И. Елисеевой. - М.: ООО "ВИТРЭМ", 2002. - 448 с.
15. Тарасенко Т.О. Статистика: Навчальний посібник. - К.: Центр навчальної літератури, 2006. - 344 с.
16. Теория статистики: Учебник. Под ред. Р.А. Шмойловой. - М.: Финансы и статистика, 2001. - 560 с.
17. Ткач Є.І. Загальна теорія статистики: Підручник. - Тернопіль.: Лідер, 2004. - 388 с.

Навчальне видання

Укладач РАРОК Олександр Васильович

СТАТИСТИКА. Конспект лекцій
Навчальний посібник

Підписано до друку 22.06.2017. Формат 60x84/16. Папір офсетний.
Ум. друк. арк. 11,85. Наклад 200 прим. Зам. № 3-081.

ФОП Сисин Я.І. Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру видавців, виготівників і розповсюджувачів видавничої продукції від 18.07.2012 р. Серія ДК №4368.
32300, Хмельницька обл., м. Кам'янець-Подільський, вул. Князів Коріатовичів, 9а; а/с 111.
Тел./факс: (03849) 2-73-84; моб.: 0984253404, 0501931724, 0673808375;
e-mail: abetka2017@ukr.net, <http://abetka.in.ua>