УКРАЇНСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ ЗАЛІЗНИЧНОГО ТРАНСПОРТУ

ЕКОНОМІЧНИЙ ФАКУЛЬТЕТ

Кафедра управління державними і корпоративними фінансами

Н. М. Лисьонкова, О. А. Єрмоленко

СТАТИСТИКА

Конспект лекцій

Частина 2

Харків – 2021

Лисьонкова Н. М., Єрмоленко О. А. Статистика: Конспект лекцій: у 2 ч. – Харків: УкрДУЗТ, 2021. – Ч. 2. – 78 с.

Цей конспект лекцій підготовлено відповідно до навчальної програми з дисципліни «Статистика» і він є складовою навчально-методичного комплексу дисципліни. Матеріали конспекту лекцій розглядають сукупність усіх питань курсу «Статистика» та допомагають студентам підготуватись до іспиту.

Конспект лекцій складається з двох частин.

У першій частині конспекту висвітлено сутність статистики та її значення в аналізі суспільних явищ; розглянуто основні методи статистики: зведення, групування та подання статистичних даних і основні статистичні показники.

У другій частині детально розглянуто: статистичні методи вимірювання взаємозв’язків, показники варіації, динаміка, вплив індексного та вибіркового методу розрахунку на аналіз статистичних показників.

Конспект лекцій призначено для студентів економічних спеціальностей усіх форм навчання.

Табл. 29, бібліогр.: 16 назв.

Конспект лекцій розглянуто і рекомендовано до друку на засіданні кафедри управління державними і корпоративними фінансами 18 березня 2019 р., протокол № 9.

Рецензент

доц. М. О. Єрьоміна

ЗМІСТ

Вступ……………………………………………………………….4
Тема 1. Статистичне вивчення варіації і форми розподілу………6
Тема 2. Статистичні методи вимірювання взаємозв'язків………17
Тема 3. Статистичне вивчення динаміки…………………………26
Тема 4. Індексний метод………………………………………...39
Тема 5. Вибірковий метод…………………………………………61
Список літератури…………………………………………………78

ВСТУП

Слово «статистика» походить від латинського слова «status», що означає стан, положення явищ, або від італійського «стато»–держава. Статистика, як і інші науки, виникла з практичних потреб людей. З утворенням класового суспільства з'явилась потреба збирати відомості про країну: про її населення, засоби, які має держава. Для одержання цих даних проводились різноманітні обліково-статистичні роботи. Зокрема є відомості про перепис населення в Єгипті, проведений за 3500 років до нашої ери. У Китаї більш ніж за дві тисячі років до нашої ери проводились обчислення населення за статтю і віком, а також збиралися відомості про стан сільського господарства. У Давньому Римі велась статистика чисельності населення та майнового стану громадян. У кінці IX сторіччя проводились перші облікові операції: інвентаризація королівських маєтків, облік населення, придатного до військової служби. Спочатку роботи такого виду проводились не систематично і не були повними, вичерпними.

Регулярне проведення статистичних робіт, а також необхідність вивчення накопичених статистичних матеріалів і стали причиною виникнення науки статистики.

Статистика, як наука, зароджується в другій половині XVII сторіччя у працях видатних представників школи політичних арифметиків Уільяма Петті (1623–1687) і Джона Граунта (1620–1674).

У своїх працях («Трактат про податки і збори» 1662 р., «Слово мудрим» 1664 р., «Політична арифметика» 1676 р., «Різне про гроші» 1682 р.) Петті намагався, опираючись на кількісний розрахунок, дати відповіді на економічні і соціальні питання. Він спробував розробити програму перепису населення, широко використовував категорію середніх величин, висловлював цікаві думки з організації статистичного спостереження.

Сучасник Петті Дж. Граунт перший склав таблицю смертності, яка встановлює кількість осіб, що доживають до певного віку. Він виявив, що масовим явищам притаманні певні статистичні закономірності.

У Німеччині у цей же період (друга половина XVII сторіччя) виникла описова школа, представниками якої були Г. Конринг (1606–1681) і Г. Ахенваль (1719–1772). Г. Ахенваль, професор філософії і права у Геттенгемі, назвав цю описову науку статистикою.

Подальшого розвитку питання статистичної теорії набули в роботах цілого ряду статистиків, із яких особливо виділяються роботи видатного статистика Д. П. Журавського (1810–1856). Однією із заслуг А. Кетле є відкриття ним закономірностей масових явищ – кількісний і якісний показники і тільки їх єдність приводить до цілковитого пізнання.

До сорокових років 19 сторіччя описовий напрямок у статистиці втрачає колишнє значення і широкого розвитку набуває математичний напрямок статистики, який започаткували англійці Ф. Гальтон, К. Пірсон, В. Госсет, Р. Фішер та ін. Представники статистико-економічного напрямку основою статистики стали вважати теорію ймовірностей, яка становить одну з галузей прикладної статистики.

У рамках єдиної статистичної науки на сучасному етапі можна виділити її окремі галузі, кожна з яких має свій об’єкт дослідження, з’ясовує сутність специфічної системи показників, виробляє правила й методи їх одержання і застосування. Маючи розбіжності, кожна з галузей базується на принципах і методах загальної теорії статистики, яка може бути віднесена до першого рівня статистики.

На другому рівні виділяють статистику економічну і соціально-демографічну. Економічна статистика вивчає процеси і явища в галузі економіки: їхню структуру, пропорції, взаємозв’язки, а соціально-демографічна – населення й неекономічні (соціальні) явища й процеси в житті суспільства, комплексно характеризує різноманітні сторони соціальних умов і способи життя людей.

На третьому рівні виділяються галузі економічної і соціальної демографічної статистики.

ТЕМА 1. Статистичне вивчення варіації і форми розподілу

План

1.1 Показники варіації.
1.2 Дисперсія альтернативної ознаки.

1.1 Показники варіації

Варіацією ознаки називається її зміна в одиниць сукупності. Середні величини дають узагальнюючу характеристику сукупності за варіюючими ознаками, показують типовий для вказаних умов рівень цих ознак. Але поряд із середніми величинами велике практичне і теоретичне значення має вивчення відхилень від середніх. Від розміру і розподілу відхилень залежать типовість і надійність середніх характеристик. Для цього необхідні показники, які визначали б ступінь коливальності окремих значень ознаки від середньої. В економіці такі показники потрібні, наприклад, для характеристики ступеня ритмічності роботи підприємства. Допустимо, є два заводи. У середньому кожен з них за останній тиждень виконав план на 100 %. Але підприємства працювали по-різному. Це видно з даних, наведених у таблиці 1.1.

Таблиця 1.1

День тижня	Виконання плану, %	
	підприємство 1	підприємство 2
Понеділок	99,9	80,2
Вівторок	102,1	108,9
Середа	98,5	106,0
Четвер	100,5	92,5
П’ятниця	99,0	112,4
У середньому за тиждень	100,0	100,0

З цього прикладу видно, що хоч обидва підприємства в середньому і виконували план, але перше підприємство працювало ритмічно, а друге – неритмічно. Це показує те, що однієї середньої недостатньо, потрібно ще характеризувати співвідношення між середньою й окремими величинами, на основі яких вона обчислена.

Для вимірювання розміру варіації використовуються такі показники:

1) розмах варіації;

2) середнє лінійне відхилення;

3) середнє квадратичне відхилення;

4) дисперсія;

5) коефіцієнт варіації.

Розмах варіації R (амплітуда коливань) обчислюється як різниця між максимальним і мінімальним значенням варіюючої ознаки

$$\mathbf{R = x_{max} - x_{min}}. \tag{1.1}$$

Розмах варіації є найпростішою мірою коливальності. Величина **R** показує, у яких межах коливається розмір виучуваної ознаки. Розмах варіації як показник коливальності має істотний недолік. Цей показник показує коливальність тільки крайніх значень. Тому **R** може у ряді випадків неправильно характеризувати коливальність ознаки. Наприклад, є дані про виробіток деталей робітниками двох бригад:

1-ша бригада: 35, 40, 45 деталей; $\mathbf{\bar{x}_1 = 40}$ деталей;

$\mathbf{R = 45 - 35 = 10}$ деталей;

2-га бригада: 25, 35, 60 деталей; $\mathbf{\bar{x}_2 = 40}$ деталей;

$\mathbf{R = 60 - 35 = 35}$ деталей.

Середній виробіток робітників у двох бригадах становить 40 деталей, але якість цієї середньої різна, бо в першій бригаді різниця між індивідуальними значеннями і їх середньою незначна, а в другій бригаді – значна. Отже, розмах варіації не відображає відхилень усіх варіантів у ряду.

З огляду на те, що кожне індивідуальне значення ознаки відхиляється від середньої на визначену величину, мірою варіації може бути середнє відхилення кожного окремого варіанта від їх середньої. Такими показниками є середнє лінійне відхилення, дисперсія і середнє квадратичне відхилення.

Середнє лінійне відхилення **d** являє собою середню з абсолютних значень відхилень окремих варіантів від їх середньої.

На основі однієї з найважливіших властивостей середньої арифметичної, тобто $\sum\left(\mathbf{x}-\bar{\mathbf{x}}\right)=\mathbf{0}$, при обчисленні середнього лінійного відхилення беруться до уваги тільки його абсолютні значення, без урахування знаків («+» або « – »).

Якщо середня арифметична є простою, то середнє лінійне відхилення розраховується за формулою

$$\mathbf{d}=\frac{\sum\left|\mathbf{x}-\bar{\mathbf{x}}\right|}{\mathbf{n}}. \tag{1.2}$$

Якщо ж середня арифметична – зважена, то середнє лінійне відхилення дорівнює

$$\mathbf{d}=\frac{\sum\left|\mathbf{x}-\bar{\mathbf{x}}\right|\mathbf{f}}{\sum\mathbf{f}}. \tag{1.3}$$

Прямі дужки показують те, що відхилення беруться без урахування знаків. Але середнє лінійне відхилення як міру варіації ознаки застосовують у статистиці дуже рідко. Найчастіше відхилення від середньої підносять до квадрата і з квадратів відхилень, які будуть усі з додатним знаком, обчислюють середню величину. Отримана міра варіації називається дисперсією. Дисперсією називається середній квадрат відхилень індивідуальних значень ознаки від їх середньої величини:

$$\boldsymbol{\sigma}^2=\frac{\sum\left(\mathbf{x}-\bar{\mathbf{x}}\right)^2}{\mathbf{n}}; \tag{1.4}$$

$$\boldsymbol{\sigma}^2=\frac{\sum\left(\mathbf{x}-\bar{\mathbf{x}}\right)^2\mathbf{f}}{\sum\mathbf{f}}. \tag{1.5}$$

А корінь квадратний з дисперсії є середнє квадратичне відхилення:

$$\boldsymbol{\sigma}=\sqrt{\frac{\sum\left(\mathbf{x}-\bar{\mathbf{x}}\right)^2}{\mathbf{n}}}; \tag{1.6}$$

$$\sigma = \sqrt{\frac{\sum\left(x - \bar{x}\right)^2 f}{\sum f}}. \qquad (1.7)$$

Обчислення дисперсії і середнього квадратичного відхилення дає змогу усунути недолік середнього лінійного відхилення, тому що будь-яке число, додатне чи від’ємне, піднесене до квадрата, буде числом додатним.

Середнє квадратичне відхилення завжди виражається в тих іменованих числах, у яких виражені варіанти і середня. Воно дає абсолютну міру варіації (коливальності). Але середнього квадратичного відхилення не завжди буває достатньо для характеристики коливальності ознаки, бо воно характеризує абсолютний розмір відхилень і виражається у тих же одиницях вимірювання, що й варіанти і середня.

Наприклад, середнє квадратичне відхилення зросту хлопчиків вимірюється в сантиметрах, а вага – в кілограмах і зіставляти їх не можна.

Тому для характеристики коливальності явищ середнє квадратичне відхилення зіставляють з його середньою величиною і виражають у відсотках. Такий показник називається коефіцієнтом варіації, позначається буквою **V** і обчислюється за формулою

$$V = \frac{\sigma}{\bar{x}} \cdot 100. \qquad (1.8)$$

Коефіцієнт варіації характеризує відносну міру коливальності і виражається у відсотках. За допомогою коефіцієнта варіації можливо, наприклад, порівнювати ступінь варіації ознак в однакових рядах, але з різним рівнем середніх. Коефіцієнт варіації зручний, що дуже важливо, для порівняння варіації різних явищ.

Наприклад, є дані про розподіл робітників двох бригад за відсотком виконання норм виробітку (таблиця 1.2).

Таблиця 1.2

Виконання норм виробітку, %	1-ша бригада	110	126	92	115	101	134
	2-га бригада	107	104	100	99	105	103

$$\bar{x}_1 = 113{,}0\%; \quad \sigma_1 = 13{,}2\%; \quad V_1 = 12{,}6\%.$$
$$\bar{x}_2 = 103{,}0\%; \quad \sigma_2 = 2{,}8\%; \quad V_2 = 2{,}7\%.$$

У другій бригаді варіація робітників за відсотком виконання норм виробітку менша, оскільки $V_2 < V_1$.

Якщо необхідно порівняти варіацію декількох різних рядів, також застосовують показник варіації. Наприклад, що більше варіює: зріст хлопчиків одного віку чи їхня вага? Не можна порівнювати середнє квадратичне зросту σ_p, що дорівнює 6 см, і середнє квадратичне ваги σ_b, що дорівнює 5 кг, але, знаючи середній зріст – $\bar{x}_p$ = 120 см і середню вагу – $\bar{x}_в$ = 38 кг, можна порівнювати варіацію цих різних рядів, для чого необхідно розрахувати коефіцієнт варіації зросту і ваги. Тоді

$$V_p = \frac{6}{120} \cdot 100 = 5\%; \qquad V_в = \frac{5}{38} \cdot 100 = 13{,}2\%.$$

Отже, вага хлопчиків варіює більше, ніж їхній зріст.

Коефіцієнт варіації певною мірою є критерієм типовості середньої. Якщо коефіцієнт варіації дуже великий (перевищує 40 %), то це означає, що середня характеризує сукупність за ознакою, яка істотно змінюється в окремих одиницях. Типовість такої середньої невелика.

Розглянемо обчислення показників варіації (таблиця 1.3).

Таблиця 1.3 – Розрахунок показників варіації

x	f	$x - \bar{x}$	$(x - \bar{x})f$	$(x - \bar{x})^2$	$(x - \bar{x})^2 f$
2,5	9	-8,2	-73,8	67,24	605,16
7,5	45	-3,2	-144,0	10,24	460,80
12,5	25	+1,8	+45,0	3,24	81,00
17,5	15	+6,8	+102,0	. 46,24	693,60
22,5	6	+11,8	+70,8	139,24	835,44
Усього	100		-217,8 +217,8 0		2676,00

Середня арифметична зважена була обчислена раніше і дорівнює:

$$\bar{\mathbf{x}} = \frac{\sum \mathbf{xf}}{\sum \mathbf{f}} = \mathbf{10{,}7\ року;}$$

$$\mathbf{R} = \mathbf{x}_{\max} - \mathbf{x}_{\min} = \mathbf{22{,}5 - 2{,}5 = 20\ років;}$$

$$\mathbf{d} = \frac{\sum \left|\mathbf{x} - \bar{\mathbf{x}}\right| \mathbf{f}}{\sum \mathbf{f}} = \frac{\left|\mathbf{-217{,}8}\right| + \left|\mathbf{+217{,}8}\right|}{\mathbf{100}} = \mathbf{4{,}4\ роки;}$$

$$\boldsymbol{\sigma}^2 = \frac{\sum \left(\mathbf{x} - \bar{\mathbf{x}}\right)^2 \mathbf{f}}{\sum \mathbf{f}} = \frac{\mathbf{2676}}{\mathbf{100}} = \mathbf{26{,}76;}$$

$$\boldsymbol{\sigma} = \sqrt{\boldsymbol{\sigma}^2} = \sqrt{\mathbf{26{,}76}} = \mathbf{5{,}2\ роки;}$$

$$\mathbf{V} = \frac{\boldsymbol{\sigma}}{\bar{\mathbf{x}}} \cdot \mathbf{100} = \frac{\mathbf{5{,}2}}{\mathbf{10{,}7}} \cdot \mathbf{100 = 48{,}6\%.}$$

Таким чином, коефіцієнт варіації більший 40 %, отже, розкид окремих значень ознаки від середньої значний.

Найважливіші властивості дисперсії. Дисперсія має ряд математичних властивостей, які значно спрощують техніку її розрахунку. Розглянемо основні з них.

1 Якщо всі значення ознаки збільшити або зменшити на якесь постійне число «А», то дисперсія $\boldsymbol{\sigma}^2$ від цього не зменшиться. Після віднімання від усіх варіантів величини «А» отримаємо нові варіанти $\mathbf{x - A}$, середня з яких $\overline{\mathbf{x - A}} = \bar{\mathbf{x}} - \mathbf{A}$, тобто менше колишньої середньої на величину А (це випливає із властивостей середньої арифметичної).

Обчислимо дисперсію нового ряду варіантів

$$\boldsymbol{\sigma}_{\mathbf{н}}^2 = \frac{\sum \left[\left(\mathbf{x} - \mathbf{A}\right) - \left(\bar{\mathbf{x}} + \mathbf{A}\right) \right]^2 \mathbf{f}}{\sum \mathbf{f}} = \frac{\sum \left(\mathbf{x} - \mathbf{A} - \bar{\mathbf{x}} + \mathbf{A}\right)^2 \mathbf{f}}{\sum \mathbf{f}} = \frac{\sum \left(\mathbf{x} - \bar{\mathbf{x}}\right)^2 \mathbf{f}}{\sum \mathbf{f}} . \quad (1.9)$$

Як бачимо, дисперсія не змінилась, тобто $\boldsymbol{\sigma}_{\mathbf{н}}^2 = \boldsymbol{\sigma}_{\mathbf{o}}^2$.

Таким чином дисперсію можна обчислити не тільки за варіантами, але й за їх відхиленнями від якого-небудь постійного числа «А».

$$\sigma^2 = \sigma^2\left(x - A\right). \tag{1.10}$$

2 Якщо всі значення ознаки збільшити або зменшити в d разів, то дисперсія від цього зміниться в d^2 разів. Помножимо всі варіанти x на d, одержимо xd, середня з яких $\overline{xd} = \bar{x}d$.

Обчислимо тепер дисперсію нового ряду:

$$\sigma_{н}^2 = \frac{\sum\left(dx - d\bar{x}\right)^2 f}{\sum f} = \frac{\sum d^2\left(x - \bar{x}\right)^2 f}{\sum f}. \tag{1.11}$$

Винесемо постійний множник d за знак суми й одержимо:

$$\sigma_{н}^2 = \frac{d^2\sum\left(x - \bar{x}\right)^2 f}{\sum f} = d^2\sigma^2. \tag{1.12}$$

Дисперсія збільшилася в d^2 разів.

Отже, при обчисленні дисперсії можна всі значення ознаки розділити на постійне число d, обчислити дисперсію, а потім помножити її на це постійне число в квадраті d^2.

3 Сума квадратів відхилень індивідуальних значень ознаки x від їх середньої $\bar{x}$ менша від суми квадратів відхилень індивідуальних значень ознаки від будь-якого числа A за умови, що $A \pm \bar{x}$, тобто $\sum\left(x - \bar{x}\right)^2 < \sum\left(x - A\right)^2$ на величину $\left(\bar{x} - A\right)^2$ за умови, що $A \neq \bar{x}$.

Тоді $\sigma_{\bar{x}}^2 < \sigma_A^2$ або $\sigma_A^2 > \sigma_{\bar{x}}^2$ на конкретну величину $\left(\bar{x} - A\right)^2$.

Звідси

$$\sigma_{\bar{x}}^2 = \sigma_A^2 - \left(\bar{x} - A\right)^2 = \frac{\sum\left(x - A\right)^2 f}{\sum f} - \left(\bar{x} - A\right)^2. \tag{1.13}$$

Отже, дисперсія від середньої завжди менша від дисперсій, обчислених від будь-яких інших величин, тобто вона має властивості мінімальності.

Якщо число «А» прийняти рівним нулю, тоді дисперсія матиме вигляд

$$\boldsymbol{\sigma}^2 = \frac{\sum \mathbf{x}^2\mathbf{f}}{\sum \mathbf{f}} - \left(\frac{\sum \mathbf{xf}}{\sum \mathbf{f}}\right)^2, \text{ або } \boldsymbol{\sigma}^2 = \overline{\mathbf{x}^2} - \bar{\mathbf{x}}^2. \qquad (1.14)$$

Отже, дисперсія дорівнює різниці середнього квадрата значення ознаки і квадрата середнього значення (таблиця 1.4).

Таблиця 1.4 – Обчислення дисперсії з використанням третьої властивості

$\mathbf{x}$	$\mathbf{f}$	$\mathbf{x-A}$ $\mathbf{A=12,5}$	$\mathbf{(x-A)^2}$	$\mathbf{(x-A)^2 f}$	$\mathbf{x^2}$	$\mathbf{x^2 f}$
2,5	9	-10	100	900	6,25	56,25
7,5	45	- 5	25	1125	56,25	2531,25
12,5	25	0	0	0	156,25	3906,25
17,5	15	+ 5	25	375	306,25	4593,75
22,5	6	+10	100	600	506,25	3037,50
Усього	100			3000		14125,00

$$\boldsymbol{\sigma}^2 = \frac{\sum (\mathbf{x}-\mathbf{A})^2\mathbf{f}}{\sum \mathbf{f}} - (\bar{\mathbf{x}}-\mathbf{A})^2 = \frac{\mathbf{3000}}{\mathbf{100}} - (\mathbf{10,7}-\mathbf{12,5})^2 = \mathbf{26,76}.$$

$$\boldsymbol{\sigma}^2 = \frac{\sum \mathbf{x}^2\mathbf{f}}{\sum \mathbf{f}} - \left(\frac{\sum \mathbf{xf}}{\sum \mathbf{f}}\right)^2 = \frac{\mathbf{14125}}{\mathbf{100}} - \mathbf{10,7}^2 = \mathbf{26,76}.$$

На наведених математичних властивостях дисперсії ґрунтуються способи, які спрощують техніку обчислення самої дисперсії і середнього квадратичного відхилення.

Розрахунок середнього квадратичного відхилення являє собою трудомістку операцію. Його можна значно спростити, якщо застосувати спосіб моментів, який використовувався при обчисленні середньої арифметичної (таблиця 1.5).

Таблиця 1.5 – Обчислення дисперсії способом моментів

x	f	$x-A$ $A=7,5$	$\frac{x-A}{d}$ $d=5$	$\left(\frac{x-A}{d}\right)f$	$\left(\frac{x-A}{d}\right)^2$	$\left(\frac{x-A}{d}\right)^2 f$
2,5	9	-5	-1	-9	1	9
7,5	45	0	0	0	0	0
12,5	25	+5	+1	+25	1	25
17,5	15	+10	+2	+30	4	60
22,5	6	+15	+3	+18	9	54
Усього	100			- 9 +73 +64		148

Для розрахунку дисперсії потрібно обчислити момент другого порядку, який визначається за формулою

$$m_2 = \frac{\sum\left(\frac{x-A}{d}\right)^2 f}{\sum f},\; m_2 = \frac{148}{100} = 1,48.$$

$$m_1 = \frac{\sum\left(\frac{x-A}{d}\right) f}{\sum f},\; m_1 = \frac{64}{100} = 0,64.$$

Дисперсія, обчислена способом моментів, дорівнює

$$\sigma^2 = d^2\left(m_2 - m_1^2\right). \qquad (1.15)$$

Отже, $\sigma^2 = 5^2\left(1,48 - 0,64^2\right) = 26,76$.

А середнє квадратичне відхилення буде визначатися способом моментів за формулою

$$\sigma = d\sqrt{\left(m_2 - m_1^2\right)}. \qquad (1.16)$$

$$\sigma = 5\sqrt{\left(1,48 - 0,64^2\right)} = 5\sqrt{1,0704} = 5,2 \text{ роки}.$$

1.2 Дисперсія альтернативної ознаки

Поряд з ознаками, притаманними всім одиницям сукупності, що вивчається, статистика розглядає і такі, які одні одиниці сукупності мають, а інші – ні.

Такі ознаки називаються альтернативними. Як приклади можна назвати вищу освіту у робітників заводу, вчений ступінь у викладачів, диплом з відзнакою у тих, хто закінчив ЗВО, і т. ін.

У багатьох випадках виникає потреба виміряти варіацію альтернативної ознаки. Відсутність ознаки, яка нас цікавить, позначено через 0, її наявність – через 1, а частку одиниць, які мають цю ознаку, через $\mathbf{p}$, які не мають – $\mathbf{q}$. На основі вище викладеного одержимо статичний ряд.

x	f
1	p
0	q
Усього	0

За даними цього ряду обчислимо середні значення альтернативної ознаки і її дисперсії .Оскільки сума часток одиниць, які мають альтернативну ознаку і її не мають, дорівнює **1**, то одержимо

$$\bar{\mathbf{x}} = \frac{\mathbf{1} \cdot \mathbf{p} + \mathbf{0} \cdot \mathbf{q}}{\mathbf{p} + \mathbf{q}} = \mathbf{p}. \qquad (1.17)$$

Отже, середня альтернативної ознаки дорівнює частоті одиниць, які мають цю ознаку:

$$\boldsymbol{\sigma}_{\mathbf{a}}^{2} = \frac{(\mathbf{1} - \mathbf{p})^{2}\mathbf{p} + (\mathbf{0} - \mathbf{p})^{2}\mathbf{q}}{\mathbf{p} + \mathbf{q}} = \frac{\mathbf{q}^{2}\mathbf{p} + \mathbf{p}^{2}\mathbf{q}}{\mathbf{p} + \mathbf{q}} = \frac{\mathbf{pq}(\mathbf{p} + \mathbf{q})}{\mathbf{p} + \mathbf{q}} = \mathbf{p}(\mathbf{1} - \mathbf{p}). \qquad (1.18)$$

Тобто дисперсія альтернативної ознаки дорівнює добутку частки одиниць, які мають цю ознаку, на частку одиниць, які не її мають.

Розглянемо альтернативну ознаку і якість продукції для десяти відібраних деталей.

Перший випадок. Усі деталі стандартні. Запис буде такий:

Частка ознаки $\mathbf{p} = \dfrac{\mathbf{M}}{\mathbf{N}}$, де $\mathbf{M}$ – кількість якісних деталей (одиниць сукупності), тобто які мають цю ознаку; $\mathbf{N}$ – загальна кількість одиниць сукупності. Для цього випадку $\mathbf{p} = \dfrac{\mathbf{10}}{\mathbf{10}} = \mathbf{1}$. Отже, варіація відсутня, оскільки всі деталі мають обрану ознаку.

Другий випадок. Друга і восьма деталі браковані. Запис такий:

Тут $\mathbf{p} = \dfrac{\mathbf{8}}{\mathbf{10}} = \mathbf{0{,}8}$, $\mathbf{q} = \mathbf{1} - \mathbf{0{,}8} = \mathbf{0{,}2}$, тоді $\bar{\mathbf{x}}_{\mathbf{a}} = \mathbf{0{,}8}$, $\boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{Q}} = \mathbf{0{,}8} \cdot \mathbf{0{,}2} = \mathbf{0{,}16}$.

Варіація сукупності за обраною ознакою наявна.

Третій випадок. П’ять деталей з десяти браковані. Запис такий:

Тут $\mathbf{p} = \dfrac{\mathbf{5}}{\mathbf{10}} = \mathbf{0{,}5}$, $\mathbf{q} = \mathbf{1} - \mathbf{0{,}5} = \mathbf{0{,}5}$, тоді $\bar{\mathbf{x}}_{\mathbf{a}} = \mathbf{0{,}5}$, $\boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{Q}} = \mathbf{0{,}5} \cdot \mathbf{0{,}5} = \mathbf{0{,}25}$.

Варіація досягла максимуму, водночас дисперсія також має максимальне значення, яке дорівнює 0,25.

Питання до самоконтролю

1 Дайте визначення варіації ознаки.

2 Які показники використовують для вимірювання розміру варіації.

3 Найважливіші властивості дисперсії.

4 Особливості обчислення дисперсії, використовуючи її третю властивість.

5 Особливості обчислення дисперсії способом моментів.

6 Дисперсія альтернативної ознаки.

ТЕМА 2. Статистичні методи вимірювання взаємозв’язків

План

2.1 Загальний зв’язок явищ. Види зв’язків. Завдання статистичного вивчення зв’язку.

2.2 Кореляційний і регресійний аналізи статистичного зв’язку соціально-економічних явищ.

2.3 Показники тісноти зв’язку.

2.4 Нелінійні залежності.

2.5 П**обудова багатофакторних моделей.**

2.1 Загальний зв’язок явищ. Види зв’язків. Завдання статистичного вивчення зв’язку

Одним з найбільш загальних законів об’єктивного світу є закон зв’язку і залежності між явищами суспільного життя. Ці явища найбільш складні, оскільки вони формуються під дією численних, різноманітних і взаємопов’язаних факторів.

Усі явища суспільного життя існують не ізольовано, вони органічно пов’язані між собою, залежать одні від одних і перебувають у постійному русі і розвитку.

Розкриваючи взаємозв’язки і взаємозалежності між явищами, можна пізнати їх суть і закони розвитку. Тому вивчення взаємозв’язків є основним завданням будь-якого статистичного аналізу.

Суспільні явища або окремі їх ознаки, які впливають на інші і обумовлюють їх зміну, називаються **факторними**, а суспільні явища або окремі їх ознаки, які змінюються під впливом факторних, – **результативними**.

За характером залежності явищ розрізняють функціональні і кореляційні зв’язки.

Функціональним називається зв'язок, при якому певному значенню факторної ознаки завжди відповідає одне значення результативної ознаки. Функціональні зв’язки характеризуються певною відповідністю між причиною і наслідком.

Кореляційним називається зв'язок, при якому кожному значенню факторної ознаки відповідає декілька значень результативної ознаки. У кореляційних зв’язках між причиною і наслідком немає повної відповідності, а спостерігається лише певне співвідношення.

За напрямком розрізняють зв’язки прямі і зворотні.

Прямий зв'язок – це такий зв'язок, коли із зростанням факторної ознаки результативна також зростає.

При зворотному зв’язку із збільшенням факторної ознаки результативна зменшується або, навпаки, із зменшенням факторної ознаки результативна зростає.

За формою зв'язок поділяється на прямолінійний і криволінійний.

При **прямолінійній** кореляційній залежності рівним змінним середніх значень факторної ознаки відповідають приблизно рівні змінні середніх значень результативної ознаки.

При **криволінійній** кореляційній залежності рівним змінним середніх значень факторної ознаки відповідають нерівні змінні середніх значень результативної ознаки.

Статистичне вивчення взаємозв’язків розв’язує такі завдання:

а) визначаються форми зв’язку;

б) вимірюється тіснота (сила) зв’язку;

в) виявляється вплив окремих факторів на результативну ознаку.

Зв’язки і залежності суспільних явищ вивчаються за допомогою різних методів, які дають уявлення про їх наявність і характер. До цих методів відносять: балансовий метод, метод порівняння паралельних рядів, графічний метод, метод аналітичних групувань, індексний метод, кореляційно-регресійний аналіз та ін.

Одним з поширених методів статистичного вивчення зв’язків суспільних явищ є балансовий метод, як прийом аналізу зв’язків і пропорцій в економіці.

Статистичний баланс являє собою систему показників, яка складається із двох сум абсолютних величин, пов’язаних між собою знаком рівності:

$$\mathbf{а+б=в+г.} \tag{2.1}$$

Цю балансову ув’язку можна зобразити через балансове рівняння: залишок на початок + надходження = видатки + + залишок на кінець.

Наведена балансова рівність характеризує єдиний процес руху матеріальних ресурсів і показує взаємозв’язок і пропорції окремих елементів цього процесу.

Метод порівняння паралельних рядів полягає в тому, що отримані в результаті групування і лічильної обробки матеріали статистичного спостереження є рангованими паралельними рядами за факторною ознакою. Паралельно записуються значення результативної ознаки. Це дає можливість, порівнюючи їх, простежити співвідношення, виявити існування зв’язку і його напрямок.

Графічний метод виявлення кореляційної залежності полягає в зображенні статистичних характеристик, отриманих у результаті зведення й обробки вихідної інформації на графіку, яке наочно покаже форму зв’язку між досліджуваними ознаками та його напрямком.

Метод статистичних групувань, як прийом виявлення кореляційної залежності, належить до найважливіших прийомів дослідження взаємозв’язків. Для виявлення залежності між ознаками за допомогою цього методу матеріал статистичного спостереження групується за факторною ознакою, і для кожної групи вираховуються середні значення як факторної, так і результативної ознаки. Порівнюючи зміни середніх значень результативної ознаки в міру зміни середніх значень факторної ознаки, виявляють характер зв’язку між ними.

2.2 Кореляційний і регресійний аналізи статистичного зв’язку соціально-економічних явищ

Основним завданням кореляційного і регресійного аналізу статистичних даних є виявлення залежності між досліджуваними ознаками у вигляді певної математичної формули і встановлення за допомогою коефіцієнта кореляції порівняльної ознаки тісноти взаємозв’язку.

Кореляційний і регресійний методи аналізу розв’язують два основних завдання:

1) визначають за допомогою рівняння регресії аналітичну форму зв'язку між варіацією ознак **«x»** і **«y»**;

2) встановлюють міру тісноти зв'язку між ознаками.

У практиці економіко-статистичних досліджень часто доводиться мати справу з прямолінійною формою зв'язку, яка виражається за допомогою рівняння регресії.

Рівняння регресії характеризує зміну середнього рівняння результативної ознаки **«y»** залежно від зміни факторної ознаки **«x».**

У разі лінійної форми зв'язку рівняння регресії має вигляд

$$\mathbf{Y_x = a + bX}, \tag{2.2}$$

де **a, b** – параметри рівняння:

$$\mathbf{b = \frac{n \cdot \sum XY - \sum X \cdot \sum Y}{n \cdot \sum X^2 - \left(\sum X\right)^2}}, \tag{2.3}$$

$$\mathbf{a = \frac{\sum Y - b \cdot \sum X}{n}}. \tag{2.4}$$

Для економічної інтерпретації лінійних і нелінійних зв'язків між двома досліджуваними явищами часто використовують розраховані на основі рівнянь регресії коефіцієнти еластичності.

Коефіцієнт еластичності показує, на скільки відсотків змінюється в середньому результативна ознака **«y»** при зміні факторної ознаки **«x»** на 1 %.

Для лінійної залежності коефіцієнт еластичності визначається за формулою

$$\mathbf{E = b \cdot \frac{\overline{X}}{\overline{Y}}}. \tag{2.5}$$

2.3 Показники тісноти зв'язку

Поряд із визначенням характеру зв'язку та ефектів впливу факторів ***x*** на результат *y* важливе значення має оцінка щільності зв'язку, тобто оцінка узгодженості варіації взаємопов'язаних ознак. Якщо вплив факторної ознаки ***x*** на результативну ***y*** значний, це виявиться в закономірній зміні значень ***y*** зі зміною значень ***x***, тобто фактор ***x*** своїм впливом формує варіацію ***y***. За відсутності зв'язку варіація ***y*** не залежить від варіації ***x***.

Для оцінювання щільності зв'язку статистика використовує низку коефіцієнтів з такими спільними властивостями:

1) за відсутності будь-якого зв'язку значення коефіцієнта наближається до нуля; при функціональному зв'язку – до одиниці;

2) за наявності кореляційного зв'язку коефіцієнт виражається дробом, який за абсолютною величиною тим більший, чим щільніший зв'язок.

Серед мір щільності зв'язку найпоширенішим є **коефіцієнт кореляції** Персона, який позначається символом **r.** Оскільки сфера його використання обмежується лінійною залежністю, то і в назві фігурує слово «лінійний». Обчислення лінійного коефіцієнта кореляції **r** ґрунтується на відхиленнях значень взаємопов'язаних ознак **x** і **y** від середніх.

Коефіцієнт кореляції визначається відношенням зазначених сум:

$$\mathbf{r} = \frac{\sum\left(\mathbf{x}-\bar{\mathbf{x}}\right)\cdot\left(\mathbf{y}-\bar{\mathbf{y}}\right)}{\sqrt{\sum\left(\mathbf{x}-\bar{\mathbf{x}}\right)^{2}\cdot\sum\left(\mathbf{y}-\bar{\mathbf{y}}\right)^{2}}}. \qquad (2.6)$$

Очевидно, що в разі функціонального зв'язку фактична сума відхилень дорівнює граничній, а коефіцієнт кореляції $r=\pm 1$**;** при кореляційному зв'язку абсолютне його значення буде тим більшим, чим щільніший зв'язок.

На практиці застосовують різні модифікації наведеної формули коефіцієнта кореляції:

$$\mathbf{r} = \frac{\sum \mathbf{xy} - \mathbf{n}\bar{\mathbf{x}}\bar{\mathbf{y}}}{\sqrt{\boldsymbol{\sigma}_x^2\boldsymbol{\sigma}_y^2}}. \qquad (2.7)$$

Оцінка тісноти зв'язку проводиться за схемою:

Сила зв'язку	Величина лінійного коефіцієнта кореляції при наявності	
	прямого зв'язку	оберненого зв'язку
Слабка Середня Тісна	0,1-0,3 0,3-0,7 0,7-0,99	(-0,1)-(-0,3) (-0,3)-(-0,7) (-0,7)-(-0,99)

Відношення факторної дисперсії до загальної розглядається як міра щільності кореляційного зв'язку і називається **коефіцієнтом детермінації**:

$$\mathbf{R^2 = \frac{\delta_x^2}{\sigma_y^2}}. \tag{2.8}$$

Корінь квадратний з коефіцієнта детермінації називають **індексом кореляції** R.

На таких самих засадах ґрунтується оцінювання щільності зв'язку за даними аналітичного групування. Мірою щільності зв'язку є **кореляційне відношення**

$$\mathbf{\eta^2 = \frac{\delta^2}{\sigma^2}}, \tag{2.9}$$

де $\boldsymbol{\delta^2}$ – міжгрупова дисперсія, яка вимірює варіацію ознаки **y** під впливом фактора **x;**

$\boldsymbol{\sigma^2}$ – загальна дисперсія.

2.4 Нелінійні залежності

У практиці економічного аналізу найчастіше використовують такі нелінійні функції залежності: гіперболічну, параболічну другого порядку, напівлогарифмічну та деякі інші.

Якщо результативна ознака із збільшенням факторної ознаки зростає або спадає не безмежно, а прямує до кінцевої мети, то для її аналізу застосовують **рівняння гіперболи**

$$\mathbf{Y_x = a + b \cdot \frac{1}{x}}. \qquad (2.10)$$

Для знаходження параметрів цього рівняння способом найменших квадратів використовується система нормальних рівнянь:

$$\begin{cases} \sum \mathbf{y = n \cdot a + b} \sum \frac{\mathbf{1}}{\mathbf{x}} \\ \sum \mathbf{y} \frac{\mathbf{1}}{\mathbf{x}} = \mathbf{a} \cdot \sum \frac{\mathbf{1}}{\mathbf{x}} + \mathbf{b} \cdot \sum \frac{\mathbf{1}}{\mathbf{x^2}} \end{cases}. \qquad (2.11)$$

За способом найменших квадратів параметри гіперболи визначають за формулами:

$$\mathbf{a} = \frac{\sum \frac{\mathbf{1}}{\mathbf{x^2}} \cdot \sum \mathbf{y} - \sum \frac{\mathbf{1}}{\mathbf{x}} \cdot \sum \frac{\mathbf{y}}{\mathbf{x}}}{\mathbf{n} \cdot \sum \frac{\mathbf{1}}{\mathbf{x^2}} - \sum \frac{\mathbf{1}}{\mathbf{x}} \cdot \sum \frac{\mathbf{1}}{\mathbf{x}}}, \qquad (2.12)$$

$$\mathbf{b} = \frac{\mathbf{n} \cdot \sum \frac{\mathbf{x}}{\mathbf{y}} \cdot \sum \mathbf{y} - \sum \frac{\mathbf{1}}{\mathbf{x}} \cdot \sum \mathbf{y}}{\mathbf{n} \cdot \sum \frac{\mathbf{1}}{\mathbf{x^2}} - \sum \frac{\mathbf{1}}{\mathbf{x}} \cdot \sum \frac{\mathbf{1}}{\mathbf{x}}}. \qquad (2.13)$$

Парабола другого порядку застосовується тоді, коли із зростанням факторної ознаки відбувається нерівномірне зростання або спадання результативної ознаки. Рівняння параболи другого порядку визначається за формулою

$$\mathbf{Y_x = a_0 + a_1 x + a_2 x^2}. \qquad (2.14)$$

Параметри цього рівняння знаходять способом найменших квадратів шляхом складання і розв’язання системи нормальних рівнянь:

$$\begin{cases} \sum y = n \cdot a_0 + a_1 \sum x + a_2 \sum x^2 \\ \sum xy = a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 \\ \sum x^2 y = a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 \end{cases}. \quad (2.15)$$

Вирівнювання за напівлогарифмічною кривою проводять тоді, коли із зростанням факторної ознаки середня результативна ознака спочатку до певних меж зростає досить швидко, але пізніше темпи її зростання поступово сповільнюються.

Напівлогарифмічна функція має вигляд

$$Y_x = a + b \cdot \log x. \quad (2.16)$$

Для знаходження параметрів напівлогарифмічної функції способом найменших квадратів розв'язують систему двох рівнянь

$$\begin{cases} \sum y = n \cdot a_0 + b \cdot \sum \log x \\ \sum y \log x = a \cdot \sum \log x + b \cdot \sum (\log x)^2 \end{cases} \quad (2.17)$$

2.5 Побудова багатофакторних моделей

У багатьох випадках на результативну ознаку впливає не один, а декілька факторів. Між ними існують складні взаємозв'язки, тому їх вплив на результативну ознаку комплексний і його не можна розглядати як просту суму ізольованих впливів.

Багатофакторний кореляційно-регресійний аналіз дає змогу оцінити міру впливу на досліджуваний результативний показник кожного із введених у модель факторів при зафіксованих на середньому рівні інших факторів.

Форму зв'язку можна визначити шляхом перебору функцій різних типів, але це пов'язано з великою кількістю зайвих розрахунків. Однак, беручи до уваги, що будь-яку функцію багатьох змінних шляхом логарифмування або заміни змінних можна звести до лінійного виду

$$Y_x = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + ... + a_n x_n, \quad (2.18)$$

параметри рівняння знаходять способом найменших квадратів.

Кожен коефіцієнт рівняння показує ступінь впливу відповідного фактора на результативний показник при фіксованому положенні решти факторів, тобто як із зміною окремого фактора на одиницю змінюється результативний показник.

На основі коефіцієнтів регресії не можна судити, яка із факторних ознак найбільше впливає на результативну ознаку, оскільки коефіцієнти регресії між собою не порівнянні, бо вони вимірюються у різних одиницях.

З метою виявлення порівняльної сили впливу окремих факторів і резервів, які закладені в них, статистика вираховує часткові коефіцієнти еластичності, а також β-коефіцієнти.

Часткові коефіцієнти еластичності показують, на скільки відсотків у середньому зміниться результативна ознака із зміною на 1 % кожного фактора при фіксованому положенні інших факторів.

Для визначення факторів, у розвитку яких закладені найбільші резерви покращення досліджуваної ознаки, з урахуванням ступеня варіації факторів рівняння множинної регресії вираховують **часткові β-коефіцієнти**, які показують, на яку частину середнього квадратичного відхилення змінюється результативна ознака із зміною відповідної факторної ознаки на величину її середнього квадратичного відхилення.

Для характеристики ступеня тісноти зв’язку в множинній прямолінійній кореляції використовують **множинний коефіцієнт кореляції**

$$\mathbf{R_{YX_1X_2}} = \sqrt{\frac{\mathbf{r^2_{YX_1} + r^2_{YX_2} - 2r_{YX_1} \cdot r_{YX_2} \cdot r_{X_1X_2}}}{\mathbf{1 - r^2_{X_1X_2}}}}\,. \qquad (2.19)$$

Множинний коефіцієнт кореляції показує, яку частину загальної кореляції становлять коливання під впливом факторів $\mathbf{x_1}$, $\mathbf{x_2}$,..., $\mathbf{x_n}$, закладених у багатофакторну модель для дослідження.

Множинний коефіцієнт кореляції коливається в межах від «0» до «± 1». При **R=0** зв'язку між досліджуваними ознаками немає, а при **R=1** – він функціональний.

Питання до самоконтролю

1 У чому різниця між факторними та результативними явищами?

2 Які зв'язки розрізняють за характером залежності явищ.

3 Основне завдання кореляційного і регресійного аналізу статистичних даних.

4 Показники тісноти зв'язку та їх основні характеристики.

5 Основні нелінійні функції залежності.

6 Основи побудови багатофакторних моделей.

7 Часткові коефіцієнти еластичності та їх характеристика.

ТЕМА 3. Статистичне вивчення динаміки

План

3.1.Види динамічних рядів.

3.2 Показники аналізу динаміки і методи їх обчислення.

3.3 Розрахунок середніх показників динамічного ряду.

3.1 Види динамічних рядів

Процеси і явища суспільного життя, які вивчає статистика, перебувають у постійному русі і зміні. У процесі розвитку змінюються розміри, склад, обсяг, структура конкретних суспільних явищ. Ці зміни можна вивчати, якщо мати дані з певного кола показників за ряд моментів часу або за ряд проміжків часу, що йдуть один за одним.

Ряд розташованих у хронологічній послідовності значень статистичних показників являє собою часовий (динамічний) ряд, тобто динамічним рядом називається ряд чисел, що характеризують зміну явища в часі.

Будь-який динамічний ряд оформляється статистичною таблицею і складається з двох елементів (таблиця 3.1).

Таблиця 3.1 – Якийсь статистичний показник

Час – період часу (місяць, квартал, рік) або момент часу (число, дата)
Рівні ряду (у) – числові значення показників, узятих за певний період часу або стосовно моменту часу

Рівні можуть бути виражені абсолютними, середніми і відносними величинами. Залежно від цього ряди динаміки можуть бути рядами:

– абсолютних величин (таблиця 3.2);

– середніх величин (таблиця 3.3);

– відносних величин (таблиця 3.4).

Таблиця 3.2 – Добування вугілля в країні У

Рік	2014	2015	2016	2017	2018	2019
Вугілля, млн т	108,7	105,4	91,0	75,9	65,6	71,1

Це динамічний ряд абсолютних величин, тому що рівні виражені абсолютними величинами. Рівні такого ряду можна підсумовувати. У наведеному прикладі можна визначити загальне добування вугілля за 6 років.

Таблиця 3.3 – Середня заробітна плата робітників та службовців

Рік	2015	2016	2017	2018	2019
Заробітна плата, грн	115	123	128	134	140

У цьому динамічному ряду рівні виражені середніми величинами. Рівні такого ряду підсумовувати неприпустимо.

Таблиця 3.4 – Вантажообіг залізничного транспорту (2015 р. – 1)

Рік	2015	2016	2017	2018	2019
До 2014	1,00	1,12	0,98	0,96	0,97

Рівні поданого динамічного ряду виражені відносними величинами динаміки і підсумовувати такі рівні не можна.

Залежно від того, узяті рівні ряду за період часу чи належать до моменту часу, динамічні ряди діляться:

– на інтервальні (таблиці 3.2 – 3.4);

– моментні (таблиця 3.5).

Динамічні ряди, зображені в таблицях 3.2 – 3.4, є інтервальними, тому що їх рівні взяті за період часу.

Таблиця 3.5 – Населення N-ї області (на початок року)

Рік	2015	2016	2017	2018	2019
Населення, тис. люд	3250	3680	4020	4015	4009

У цьому динамічному ряді рівні належать до моменту часу (на 01.01) і хоча вони виражені абсолютними величинами, підсумовувати їх безглуздо. Такий динамічний ряд є моментним.

Найважливішою умовою практичної побудови рядів є порівнянність усіх статистичних показників, що до них входять. Для цього необхідно, щоб склад досліджуваної сукупності був той самий протягом усього ряду, тобто належав до однієї території, до одного кола об’єктів і був обчислений за однією методологією.

Крім того, дані динамічного ряду повинні бути виражені в тих самих одиницях виміру, а проміжки часу між значеннями ряду, якщо можна, однаковими. Для перетворення непорівняннних рядів у порівнянні вдаються до перерахунку даних, використовуючи роздільні прийоми. Розглянемо деякі з них.

Прямий перерахунок даних. Якщо динамічні ряди непорівнянні внаслідок зміни кола об’єктів обліку або територіальних меж, то для забезпечення порівнянності робиться прямий перерахунок даних за первинним матеріалом.

Змикання рядів. Цей метод застосовується в тому разі, якщо динамічні ряди непорівнянні по колу об’єктів , що охоплюються. Якщо два ряди показників, що характеризують динаміку того самого явища в нових і старих адміністративних межах, причому на один-два терміни дані в нових і старих межах по тому самому колу об’єктів, то такі динамічні ряди можна зімкнути. Наприклад, є дані про товарообіг області в старих і нових межах (таблиця 3.6).

Таблиця 3.6 – Роздрібний товарообіг N-ї області в нових і старих межах

У мільйонах гривень

Межі області	2013	2014	2015	2016	2017	2018	2019
Старі межі	310	350	370	400			
Нові межі				480	530	600	640
Порівняльний ряд	372	420	444	480	530	600	640

Для того, щоб зімкнути ці два ряди, потрібно перерахувати дані за 2013-2015 роки з урахуванням нових меж, використовуючи для цього 2016 рік, де роздрібний товарообіг відомий і в нових, і в старих межах. Обчислимо для 2016 року відношення роздрібного товарообігу в нових межах до роздрібного товарообігу в старих межах:

480/400=1,2 або 120 %.

Помноживши на цей коефіцієнт дані за попередні роки, наведені в старих межах, приведемо їх у нові межі. Тоді

$$y_{2015}=370\cdot 1{,}2=444;\ y_{2014}=350\cdot 1{,}2=420;\ y_{2013}=310\cdot 1{,}2=372\,.$$

Зімкнутий порівняльний ряд динаміки наведено в останньому рядку таблиці 3.6.

3.2 Показники аналізу динаміки і методи їх обчислення

При вивченні динаміки необхідно розв’язати цілий ряд задач. До основних задач, що виникають при вивченні динамічних рядів, відносять такі:

– характеристика інтенсивності змін від періоду до періоду або від дати до дати;

– визначення середніх показників динамічного ряду;

– виявлення основних закономірностей динаміки;

– прогноз розвитку явища на майбутнє.

У результаті порівняння рівнів ряду утворюється система абсолютних показників динаміки, до яких відносять показники, наведені у таблиці 3.7.

Таблиця 3.7

Показник	Умовне позначення
Рівень ряду	$\mathbf{y}_i$
Абсолютний приріст	Δ_y
Темп (коефіцієнт) росту	**Тр(Кр)**
Темп (коефіцієнт) приросту	**Тпр(Кр)**
Абсолютне значення одного відсотка приросту	**А%**
Коефіцієнт випередження	**Квип**

Якщо порівнянню підлягають декілька послідовних рівнів, то можливі два варіанти зіставлення:

1) кожен рівень динамічного ряду порівнюється з тим самим попереднім рівнем, взятим за базу порівняння. Як базисний рівень вибирається або початковий рівень динамічного ряду, або рівень, з якого починається якийсь новий етап розвитку явища. Такий метод розрахунку називається базисним;

2) кожен рівень ряду порівнюється з рівнем, йому попереднім. Такий метод розрахунку називається ланцюговим.

Рівнем ряду $\mathbf{y}_i$ називається величина кожного числа динамічного ряду. Усі рівні ряду характеризують його динаміку. Розрізняють початковий, кінцевий і середній рівні ряду. Початковий рівень – це величина першого члена ряду, кінцевий – останнього, середній рівень – середня з усіх значень динамічного ряду.

Абсолютний приріст Δ_y показує, на скільки більший або менший порівнювальний рівень від базисного (початкового) або попереднього рівня. Він визначається як різниця порівнювальних рівнів і характеризує абсолютну швидкість росту:

$$\Delta_y = \mathbf{y}_i - \mathbf{y}_0, \tag{3.1}$$

де Δ_y – абсолютний приріст;

$\mathbf{y}_i$ – рівень порівнювального періоду;

$\mathbf{y}_0$ – рівень базисного періоду.

$$\Delta'_{y} = \mathbf{y_i} - \mathbf{y_{i-1}}, \tag{3.2}$$

де $\mathbf{y_{i-1}}$ – рівень попереднього періоду.

Коефіцієнт росту **Кр** показує, у скільки разів досліджуваний рівень більший або менший від попереднього. Він визначається відношенням порівнювальних рівнів і характеризує відносну швидкість росту

$$\mathbf{К_р = \frac{y_i}{y_0}, К'_р = \frac{y_i}{y_{i-1}}}. \tag{3.3}$$

Якщо коефіцієнт росту виражають у відсотках, то їх називають темпами росту

$$\mathbf{Т_р = \frac{y_i}{y_0} \cdot 100, Т'_р = \frac{y_i}{y_{i-1}} \cdot 100}. \tag{3.4}$$

Темп росту **Тпр** показує, на скільки відсотків рівень цього періоду більший або менший від базисного рівня або попереднього. Він розраховується за формулою

$$\mathbf{Т_{пр} = \frac{\Delta y}{y_0} \cdot 100, Т'_{пр} = \frac{\Delta' y}{y_{i-1}} \cdot 100}. \tag{3.5}$$

Зробивши перетворення, одержимо

$$\mathbf{К_{пр} = \frac{\Delta y}{y_0} = \frac{y_i - y_0}{y_0} = \frac{y_i}{y_0} - \frac{y_0}{y_0} = К_р - 1}, \tag{3.6}$$

або

$$\mathbf{Т_{пр} = \frac{\Delta y}{y_0} \cdot 100 = \frac{y_i - y_0}{y_0} \cdot 100 = \frac{y_i}{y_0} \cdot 100 - \frac{y_0}{y_0} \cdot 100 = Т_р - 100}. \tag{3.7}$$

Отже, темп приросту можна знайти як різницю між темпом росту і 100 або між коефіцієнтом росту і 1.

Абсолютне значення одного відсотка приросту **А %**. Для того, щоб правильно оцінити отримане значення темпу приросту , його розглядають у зіставленні з показником абсолютного приросту. Результат виражається показником, який називають абсолютним значенням одного відсотка приросту. Цей показник характеризує матеріальний вміст одного відсотка приросту і визначається за формулою

$$\mathbf{А\% = \frac{\Delta' у}{Т'_{пр}(в\%)}}. \tag{3.8}$$

Якщо перетворити цю формулу, то одержимо таку:

$$\mathbf{А\% = \frac{у_i - у_{i-1}}{\frac{у_i - у_{i-1}}{у_{i-1}} * 100} = \frac{у_{i-1}}{100} = 0{,}01 \cdot у_{i-1}}. \tag{3.9}$$

Це означає, що абсолютне значення одного відсотка приросту при ланцюговому методі дорівнює одній сотій попереднього рівня ряду.

При зіставленні динаміки розвитку двох явищ можна використовувати показники, які являють собою відношення коефіцієнтів росту або приросту за однакові відрізки часу по двох або декількох динамічних рядах. Ці показники називають **коефіцієнтами випередження К$_{вип}$.**

$$\mathbf{К_{вип} = \frac{К_р^1}{К_р^2} \text{ або } К_{вип} = \frac{К_{пр}^1}{К_{пр}^2}}, \tag{3.10}$$

де $\mathbf{К_р^1, К_р^2 \text{ і } К_{пр}^1, К_{пр}^2}$ – відповідно коефіцієнти росту і приросту порівнювальних динамічних рядів.

Тоді, коли порівняння проводиться з достатньо віддаленим періодом часу, взятим за базу порівняння, розраховують так звані пункти росту, що являють собою різницю базисних темпів росту (приросту) у відсотках двох суміжних періодів.

Зведемо усі перераховані показники в таблицю 3.8.

Таблиця 3.8 – Аналітичні показники динамічного ряду

Показник	Метод розрахунку	
	базисний	ланцюговий
Абсолютний приріст	$\Delta' y = y_i - y_0$	$\Delta' y = y_i - y_{i-1}$
Темп (коефіцієнт) росту	$T_p = \frac{y_i}{y_0} \cdot 100, K_p = \frac{y_i}{y_0}$	$T'_p = \frac{y_i}{y_{i-1}} \cdot 100, K'_p = \frac{y_i}{y_{i-1}}$
Темп приросту	$T_{пр} = \frac{\Delta y}{y_0} \cdot 100$	$T'_{пр} = \frac{\Delta' y}{y_{i-1}} * 100$
Абсолютне значення 1 % приросту	***Розрахунок не має сенсу***	$A\% = \frac{\Delta' y}{T'_{пр}(в\%)} = 0{,}01 * y_{i-1}$
Коефіцієнт випередження	$K_{вип} = \frac{K_p^1}{K_p^2}, K_{вип} = \frac{K_{пр}^1}{K_{пр}^2}$	-

Розглянемо розрахунок вищевказаних показників за рядом динаміки споживання енергоресурсів у країні N (таблиця 3.9).

Таблиця 3.9 – Загальне споживання енергоресурсів у країні N

Рік	2013	2014	2015	2016	2017	2018	2019
Загальне споживання, млн т умов. од.	228,4	207,5	183,4	159,3	156,1	157,7	158,2

Для визначення показників скористаємося таблицею 3.10.

Пункти росту наведені в графі 10 таблиці 3.10. Так, для 2019 року пункт росту становитиме +0,3=-30,7-(-31,0) (дивись графу 7 таблиці 3.10).

На відміну від темпів приросту, які не можна ні підсумовувати, ні перемножувати, пункти росту можна додавати, у результаті чого одержуємо темп росту відповідного періоду порівняно з базисним періодом, тобто за даними нашого періоду сума пунктів дорівнює - 30,7, що відповідає темпу приросту рівня 2016 року порівняно з 2014 роком (-9,2-10,5-10,6-1,4+0,7+0,3) (дивись графу 10 таблиці 3.10).

Таблиця 3.10 – Розрахунок споживання енергоресурсів за роками

Рік	Споживання, млн т ум. од.	Абсолютний приріст, млн т		Темп росту, %		Темп приросту, %		Абсолютне значення 1 % приросту, тис. т	Пункт росту, %
		до 2013 року	до попер. року	до 2013 року	до попер. року	до 2013 року	до попер. року		
1	2	3	4	5	6	7	8	9	10
2013	228,4	–	–	100	–	–	–	–	–
2014	207,5	-20,9	-20,7	90,8	90,8	-9,2	9,2	–	-9,2
2015	183,4	-45	-24,1	80,3	88,4	-19,7	-11,6	–	-10,5
2016	159,3	-69,1	-24,1	69,7	86,9	-30,3	-13,1	–	-10,6
2017	156,1	-72,3	-3,2	68,3	97,9	-31,7	-2,1	–	-1,4
2018	157,7	-70,7	1,6	69	101,1	-31	1	1561	0,7
2019	158,2	-70,2	0,5	69,3	100,3	-30,7	0,3	1577	0,3

3.3 Розрахунок середніх показників динамічного ряду

Для узагальнюючої характеристики динаміки досліджуваного явища за ряд періодів визначають різноманітні середні показники (таблиця 3.11).

Таблиця 3.11

Показник	Умовне позначення
Середній рівень ряду	$\overline{\mathbf{y}}$
Середній абсолютний приріст	$\mathbf{\Delta\overline{y}}$
Середньорічний темп (коефіцієнт) росту	$\overline{\mathbf{T}}_{\mathbf{p}}(\overline{\mathbf{K}}_{\mathbf{p}})$
Середньорічний темп приросту	$\overline{\mathbf{T}}_{\mathbf{пр}}$

Метод розрахунку **середнього рівня** ряду динаміки залежить від виду часового ряду.

Щоб знайти середній рівень інтервального ряду, достатньо суму рівнів цього ряду розділити на кількість періодів, до яких вона належить, тобто за середньою арифметичною простою:

$$\mathbf{y} = \frac{\mathbf{y_1 + y_2 + y_3 + \ldots + y_n}}{\mathbf{n}} = \frac{\sum \mathbf{y}}{\mathbf{n}}, \qquad (3.11)$$

де **n** – кількість рівнів ряду.

За даними таблиці 3.9, у якій поданий інтервальний динамічний ряд, середній рівень його буде дорівнювати

$$\bar{y} = \frac{228,4 + 207,5 + 183,4 + 159,3 + 156,1 + 157,7 + 158,2}{7} =$$

$$= \frac{1250,6}{7} = 178,7 \text{ млн т} .$$

Середній рівень моментного ряду динаміки, рівні якого підсумовуванню не підлягають, так обчислювати не можна. Для моментних динамічних рядів із рівновіддаленими рівнями середній рівень розраховують за середньою хронологічною

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + y_3 + \ldots + y_{n-1} + \frac{1}{2}y_n}{n - 1}. \qquad (3.12)$$

Розглянемо для прикладу моментний ряд, що відображає зміну товарних залишків торгової бази за півріччя, тис. грн, на 1 - ше число кожного місяця:

1/I	1/II	1/III	1/IV	1/V	1/VI	1/VII
12	16	21	26	18	20	10

Тоді для наведеного прикладу середній рівень буде дорівнювати

$$\bar{y} = \frac{\frac{12}{2} + 16 + 21 + 26 + 18 + 20 + \frac{10}{2}}{7 - 1} = \frac{112}{6} = 18,7 \text{ тис.грн.}$$

Для моментних динамічних рядів із рівняннями, що різновіддалені один від одного, середній рівень визначається за середньою

$$\bar{y} = \frac{\sum yt}{\sum t}, \qquad (3.13)$$

де **t** – умовне позначення часу.

Дані для визначення середнього рівня наведені в таблиці 3.12.

Таблиця 3.12 – Дані для визначення середнього рівня

Дата	Вартість основних фондів, млн грн	Кількість місяців, протягом яких вартість не змінюється	$y_i t_i$
1/I	45	2	90
1/II	52	4	208
1/VII	64	3	192
1/X	60	3	180
Разом		12	670

Середньорічна вартість основних фондів підприємства за даними наведеного динамічного ряду становитиме 55, 8 млн грн:

$$\mathbf{\overline{y} = \frac{670}{12} = 55{,}833\ млн\ грн}.$$

Середній абсолютний приріст є середня з абсолютних приростів за рівні проміжки часу одного періоду.

Між ознаками динаміки, що обчислюються з постійною і змінною базою, є певний зв'язок. Так, сума ланцюгових абсолютних приростів дорівнює відповідному базисному абсолютному приросту. Отже, додавати є сенс тільки ланцюгові абсолютні прирости. Звідси випливає, що

$$\mathbf{\overline{\Delta y} = \frac{\Delta' y_1 + \Delta' y_2 + \ldots + \Delta' y_{n-1}}{n-1} = \frac{\sum_{n=1}^{n-1} \Delta' y_n}{n-1}},$$

де **n – 1** – кількість абсолютних приростів за період.

Оскільки $\sum \mathbf{\Delta_y}$ дорівнює різниці між останнім і першим рівнем $\mathbf{y_n - y_1}$, то середній абсолютний приріст можна знайти за формулою

$$\mathbf{\overline{\Delta y} = \frac{y_n - y_1}{n-1} = \frac{\Delta y_{n-1}}{n-1}}. \qquad (3.14)$$

Кількість абсолютних приростів менша від кількості рівнів на одиницю. За даними таблиці 3.9 знайдемо середній річний абсолютний приріст загального споживання енергоресурсів у країні N. Він становить

$$\overline{\mathbf{y}} = \frac{\mathbf{158{,}2 - 228{,}4}}{\mathbf{6}} = \frac{\mathbf{-70{,}2}}{\mathbf{6}} = \mathbf{-11{,}7\ млн\ т}.$$

Отже, протягом 2013 – 2019 років у середньому споживання енергоресурсів у країні N знижувалось на 11,7 млн т умов. од.

Середній коефіцієнт (темп) росту обчислюється за формулою середньої геометричної з показників коефіцієнтів росту за окремі періоди. За такої умови перемножувати має сенс тільки ланцюгові коефіцієнти росту:

$$\overline{\mathbf{K}}_{\mathbf{p}} = \sqrt[\mathbf{n-1}]{\mathbf{K}'_{\mathbf{p}_1} \cdot \mathbf{K}'_{\mathbf{p}_2} \ldots . \mathbf{K}'_{\mathbf{p}_{n-1}}}, \tag{3.15}$$

де $\mathbf{K}'_{\mathbf{p}_1} \cdot \mathbf{K}'_{\mathbf{p}_2} \ldots . \mathbf{K}'_{\mathbf{p}_{n-1}}$ – ланцюгові коефіцієнти росту,

n –кількість рівнів ряду.

Цю формулу можна записати інакше, якщо використовувати взаємозв’язок між ланцюговими і базисними коефіцієнтами росту. Цей взаємозв’язок полягає в такому:

1) добуток ланцюгових коефіцієнтів росту дорівнює відповідному базисному

$$\frac{\mathbf{y}_{14}}{\mathbf{y}_{13}} \cdot \frac{\mathbf{y}_{15}}{\mathbf{y}_{14}} \cdot \frac{\mathbf{y}_{16}}{\mathbf{y}_{15}} \cdot \frac{\mathbf{y}_{17}}{\mathbf{y}_{16}} \cdot \frac{\mathbf{y}_{18}}{\mathbf{y}_{17}} \cdot \frac{\mathbf{y}_{19}}{\mathbf{y}_{18}} = \frac{\mathbf{y}_{2019}}{\mathbf{y}_{2013}};$$

2) відношення наступного базисного коефіцієнта росту до попереднього дорівнює відповідному ланцюговому коефіцієнту росту

$$\frac{\mathbf{y}_{2019}}{\mathbf{y}_{2013}} : \frac{\mathbf{y}_{2018}}{\mathbf{y}_{2013}} = \frac{\mathbf{y}_{2019}}{\mathbf{y}_{2018}}.$$

Використовуючи перший взаємозв’язок коефіцієнтів росту, можна перейти до формули

$$\overline{\mathbf{K}}_{\mathbf{p}} = \sqrt[n-1]{\frac{\mathbf{y_n}}{\mathbf{y_1}}}\,. \qquad (3.16)$$

За даними прикладу таблиці 3.10 середній коефіцієнт (темп) росту буде дорівнювати

$$\overline{\mathbf{T}}_{\mathbf{p}} = \sqrt[7-1]{\mathbf{0{,}908 \cdot 0{,}884 \cdot 0{,}869 \cdot 0{,}979 \cdot 1{,}01 \cdot 1{,}003}} = \sqrt[6]{\frac{\mathbf{158{,}2}}{\mathbf{228{,}4}}} =$$

$$= \sqrt[6]{\mathbf{0{,}693}} = \mathbf{0{,}941} \text{ **або** } \mathbf{94{,}1\%}\ .$$

Це значить, що в середньому за рік динаміка споживання енергоресурсів у країні N становила 92,2 %.

Середньорічний темп приросту не можна розраховувати безпосередньо з річних темпів приросту. Спочатку потрібно знайти середньорічний темп росту, а потім одержати середньорічний темп приросту

$$\overline{\mathbf{T}}_{\mathbf{пр}} = \overline{\mathbf{T}}_{\mathbf{p}} - \mathbf{100}, \overline{\mathbf{T}}_{\mathbf{пр}} = \mathbf{94{,}2 - 100 = -5{,}8\%}.$$

Отже, за 2013 – 2019 роки споживання енергоресурсів у середньому за рік знижувалося на 5,8 %.

Питання до самоконтролю

1 Види динамічних рядів.
2 Основна характеристика динамічних рядів.
3 Основні задачі, що виникають при вивченні динамічних рядів.
4 Показники аналізу динаміки.
5 Методи обчислення показників динаміки.
6 Абсолютні показники динаміки та метод їх обчислення.
7 Розрахунок середніх показників динамічного ряду.

ТЕМА 4. Індексний метод

План

4.1 Поняття індексів.
4.2 Агрегатний індекс як вихідна форма індексу.
4.3 Середні індекси.
4 4 Індекси структурних зсувів.
4.5 Базисні та ланцюгові індекси.

4.1 Поняття індексів

Метод економічних індексів є одним з найважливіших знарядь економіко-статистичних досліджень. Індекси поряд із середніми величинами – це один з найбільш поширених видів статистичних показників. Їх широко використовують у статистичній практиці і в практиці планової роботи, характеризуючи такі найважливіші процеси, як зміни обсягу продукції, рівня продуктивності праці, собівартості, врожайності тощо.

Латинське слово індекс (index) у перекладі має значення «покажчик» («указатель» – рос.), «показник».

В економіці і статистиці цей термін вживається як у широкому, так і вузькому значенні.

У широкому значенні під індексом розуміють відносний показник, який надає характеристику співвідношення рівня соціально-економічного явища в часі порівняно з планом або в просторі. У всіх цих випадках зіставляються між собою рівні однойменних показників, які мають однаковий економічний зміст. Отже, індексами в широкому значенні є будь-які різновиди відносних величин трьох видів: динаміки, ступеня виконання плану і порівняння у просторі.

Перед безпосереднім вивченням теми розглянемо короткий перелік показників, які можуть бути досліджені за допомогою індексів, разом з їх умовними позначенями. Усі вони можуть бути розділені на три групи:

– кількісні:
  q – фізичний обсяг продукції;

– якісні:

p – ціна одиниці продукції;

z – собівартість одиниці продукції;

m – питомі витрати матеріалу;

$\boldsymbol{\omega = \frac{q}{T}}$ – рівень продуктивності праці (виробіток продукції, який припадає на одного робітника або в одиницю часу);

$\mathbf{t = \frac{1}{\omega} = \frac{T}{q}}$ – трудомісткість (витрати часу на виробництво одиниці продукції);

– складні:

$\mathbf{T = t \cdot q}$ – загальні витрати часу на виробництво продукції або кількість робітників;

$\mathbf{pq = p \cdot q}$ – вартість продукції (виробленої чи реалізованої) або товарообіг;

$\mathbf{zq = z \cdot q}$ – собівартість виробництва продукції;

$\mathbf{M = m \cdot q}$ – загальні витрати матеріалу.

Внизу справа від цих показників вказується період часу, значення за який використовується в розрахунку. Символом **«0»** позначається показник базисного періоду, символом **«1»** – звітного. Наприклад, $\mathbf{z_0}$ – собівартіть одиниці продукції базисного періоду, $\mathbf{q_1}$ – кількість продукції, виготовленої (реалізованої) у звітному періоді.

У міжнародній практиці індекси заведено позначати символами **i** і **I**.

Буквою «**i**» позначаються індивідуальні індекси, буквою «**I**» – загальні (зведені) індекси. Відповідно до ступеня охоплення явища індекси підрозділяються на індивідуальні та зведені. Залежно від економічного призначення економічні індекси (індивідуальні) бувають:

– фізичного обсягу продукції: $\mathbf{i_q = \frac{q_1}{q_0}}$;

– цін: $\mathbf{i_p = \frac{p_1}{p_0}}$;

– собівартості продукції: $\mathbf{i_z = \frac{z_1}{z_0}}$;

– трудомісткості: $\mathbf{i_t = \frac{t_1}{t_0}}$;

– продуктивності праці: $\mathbf{i_\omega = \frac{t_0}{t_1}}$;

– питомих витрат матеріалів: $\mathbf{i_m = \frac{m_1}{m_0}}$ тощо.

Ці індекси вказують, у скільки разів змінився (збільшився або зменшився) відповідний показник у звітному періоді порівняно з базисним або скільки відсотків становить його зміна (зростання або зниження) – темп росту. Якщо від значення індексу, який виражається у відсотках, відняти 100 %, то отримана величина покаже, на скільки відсотків змінився (збільшився або зменшився) відповідний показник, тобто темп приросту показника.

Припустимо, що є дані, навені в таблиці 4.1.

Таблиця 4.1 – Кількість рекламних роликів і їх собівартість

Категорія товару	Собівартість одного рекламного ролика, тис. умов. грош. од.		Кількість виготовлених роликів, од.	
	базисний період	поточний період	базисний період	поточний період
Спортивний одяг	2,0	2,3	150	142
Парфумерія	2,3	3,2	200	240

За цими даними можна визначити відносні величини динаміки (тобто індивідуальні індекси собівартості і фізичного обсягу продукції):

$$\mathbf{i_z = \frac{z_1}{z_0}}; \qquad \mathbf{i_z^{CO} = \frac{2{,}3 \cdot 10^3}{2{,}0 \cdot 10^3} = 1{,}150} \text{ або } \mathbf{115{,}0}\ \%;$$

$$\mathbf{i_z^{\Pi} = \frac{3{,}2 \cdot 10^3}{2{,}8 \cdot 10^3} = 1{,}143} \text{ або } \mathbf{114{,}3}\ \%.$$

Отже, собівартість одного рекламного ролика про спортивний одяг зросла на 15,0 % (115,0 – 100), що в абсолютному виразі становило 300 умов. грош. од. ($\mathbf{2{,}3 \cdot 10^3 - 2{,}0 \cdot 10^3 = 0{,}3}$ тис. умов. грош. од. або 300 умов. грош. од.), а собівартість рекламного ролика про парфумерію збільшилася на 14,3 % (114,3 – 100), що в абсолютному виразі становило 400 умов. грош. од. ($\mathbf{3{,}2 \cdot 10^3 - 2{,}8 \cdot 10^3 = 0{,}4}$ тис. умов. грош. од. або 400 умов. грош. од.).

$$\mathbf{i_q = \frac{q_1}{q_0}};\ \mathbf{i_q^{CO} = \frac{142}{150} = 0{,}947} \text{ або } \mathbf{94{,}7}\ \%;\ \mathbf{i_q^{\Pi} = \frac{240}{200} = 1{,}2} \text{ або } \mathbf{120{,}0}\ \%.$$

Отже, собівартість одного рекламного ролика про парфумерію знизилась на 5,3 % (94,7 – 100) або на 8 роликів (142 – 150), а кількість рекламних роликів про парфумерію збільшилась на 20,0 % (120,0 – 100), що становило 40 роликів (240 – 200).

Отже, будь-який індекс дає змогу визначити як відносну зміну показника (як різницю між відношенням двох показників і 100 %), так і відповідну до нього абсолютну зміну (як різницю між чисельником та знаменником індексу).

Отже, індивідуальними називають індекси, які дають порівняльну характеристику окремих елементів тієї чи іншої сукупності. Порядок їх побудови повністю збігається з порядком розрахунку відносної величини динаміки, тому не становить жодних складностей.

4.2 Агрегатний індекс як вихідна форма індексу

В економічних розрахунках частіше за інші використовуються загальні (агрегатні) індекси, за допомогою яких характеризують зміни сукупності в цілому. Порядок побудови цих індексів і становить основний зміст індексної методології.

Агрегатний індекс – складний показник, який характеризує середню зміну соціально-економічного явища, що складається із елементів, які безпосередньо невідповідні один одному. Тому індекс у статистиці – це, зазвичай, узагальнюючий показник порівняння суспільно-економічних явищ, які складаються з елементів, що безпосередньо не підлягають підсумовуванню.

Латинське слово «агрегат» (aggregatus) означає «той, що додається», «той, що підсумовується».

Агрегатний індекс являє собою дріб, чисельник і знаменник якого складається із суми двох множників. Один з множників – величина, яка індексується, а другий виступає як вага індексу.

Величиною, що індексується, називають ознаку, зміна якої вивчається (ціна або собівартість продукції, витрати робочого часу, кількість виробленої продукції тощо) і ім’я якої стоїть у назві індексу.

Вага індексу – величина, постійна в часі, тобто у розрахунках вона береться за той самий момент часу. Вона служить для порівняння величин, що індексуються.

При побудові агрегатних індексів необхідно розв’язати дві задачі:

1) обрати вагу індексу для подолання непідсумовування елементів, з яких складається сукупність;

2) для обраної ваги індексу правильно визначити період часу, до якого він повинен належати.

Перша задача може бути розв’язана шляхом введення в індекс будь-якого додаткового незмінного показника, який економічно тісно пов’язаний з показником, що індексується. Цей додатковий незмінний показник має назву показника порівняння («соизмерения» – рос.), або ваги агрегатного індексу.

При обиранні ваги індексу заведено дотримуватись нижченаведеного правила.

Якщо будується індекс кількісного показника, то ваги (найчастіше – якісні) беруться за базисний період. При побудові індексу якісного показника використовуються ваги (кількісні) звітного періоду.

Керуючись цим правилом, побудуємо три індекси: витрат виробництва, собівартості та фізичного обсягу продукції.

Індекс витрат виробництва (витрат на виробництво продукції) $\mathbf{I_{zq}}$ являє собою відношення витрат виробництва поточного (звітного) періоду $\sum \mathbf{z_1 q_1}$ до витрат виробництва в базисному періоді $\sum \mathbf{z_0 q_0}$ і визначається за формулою

$$\mathbf{I_{zq}} = \frac{\sum \mathbf{z_1 q_1}}{\sum \mathbf{z_0 q_0}}. \qquad (4.1)$$

Обчислимо індекс витрат на виробництво за даними таблиці 4.1:

$$\mathbf{I_{zq}} = \frac{\mathbf{2{,}3 \cdot 10^3 \cdot 142 + 3{,}2 \cdot 10^3 \cdot 240}}{\mathbf{2{,}0 \cdot 10^3 \cdot 150 + 2{,}8 \cdot 10^3 \cdot 200}} = \frac{\mathbf{1094{,}6 \cdot 10^3}}{\mathbf{860 \cdot 10^3}} = \mathbf{1{,}273} \text{ або } \mathbf{127{,}3}\ \%.$$

Отже, витрати на виробництво рекламних роликів зросли на 27,3 %, що становить в абсолютному виразі 234,6 тис. умов. грош. од. ($1094{,}6 \cdot 10^3 - 860 \cdot 10^3$).

Значення індексу витрат на виробництво залежить від двох факторів: зміни собівартості та кількості продукції (рекламних роликів), що обумовлює можливість та необхідність побудови ще двох індексів: фізичного обсягу продукції та собівартості. Наявність цих двох індексів дасть змогу виявити взаємозв’язок між змінами загальних витрат на виробництво, собівартості та фізичного обсягу продукції.

Індекс собівартості – індекс якісного показника. Він характеризує зміну собівартості, тому величиною, що індексується в цьому випадку, виступає собівартість одиниці продукції. Як вагу доцільно використати кількість продукції. Помноживши собівартість продукції на її кількість, отримаємо величину загальних витрат на виробництво, яка являє собою показник, порівнянний з іншими подібними до нього показниками, і тому підлягає підсумовуванню. Визначається індекс собівартості за такою формулою:

$$\mathbf{I_{z}} = \frac{\sum \mathbf{z_1 q_1}}{\sum \mathbf{z_0 q_1}}, \qquad (4.2)$$

де $\sum \mathbf{z_1 q_1}$ – фактичні витрати виробництва поточного періоду;

$\sum \mathbf{z_0 q_1}$ – витрати на виробництво продукції звітного періоду з базисною собівартістю.

Обчислимо індекс витрат на виробництво за даними таблиці 4.1:

$$\mathbf{I_z = \frac{2,3\cdot 10^3\cdot 142 + 3,2\cdot 10^3\cdot 240}{2,0\cdot 10^3\cdot 142 + 2,8\cdot 10^3\cdot 240} = \frac{1094,6\cdot 10^3}{956,0\cdot 10^3} = 1,145} \text{ або } \mathbf{114,5}\ \%.$$

Отже, внаслідок збільшення собівартості виготовлення рекламних роликів, витрати на їх виробництво зросли на 14,5 %, тобто на 138,6 тис. умов. грош. од. ($1094,6\cdot 10^3 - 956,0\cdot 10^3$).

Індекс фізичного обсягу продукції (у нашому прикладі – кількість виготовлених рекламних роликів) – це індекс кількісного показника. У цьому індексі змінюється (індексується) кількість продукції в натуральному виразі, а вагою виступає собівартість продукції. Помноживши непорівнянні між собою кількість різнорідної продукції на їхню собівартість, можна перейти до витрат виробництва, які вже є порівнянними між собою.

Індекс фізичного обсягу продукції визначається за такою формулою:

$$\mathbf{I_q = \frac{\sum q_1 z_0}{\sum q_0 z_0}}, \tag{4.3}$$

де $\mathbf{\sum q_1 z_0}$ – витрати на виробництво продукції звітного періоду за базисною собівартістю;

$\mathbf{\sum q_0 z_0}$ – витрати на виробництво базисного періоду.

Обчислимо індекс витрат на виробництво за даними таблиці 4.1:

$$\mathbf{I_q = \frac{2,0\cdot 10^3\cdot 142 + 2,80\cdot 10^3\cdot 240}{2,0\cdot 10^3\cdot 150 + 2,8\cdot 10^3\cdot 200} = \frac{956,0\cdot 10^3}{860,0\cdot 10^3} = 1,112} \text{ або } \mathbf{111,2}\ \%.$$

Отже, внаслідок зміни кількості виготовлених рекламних роликів витрати на їх виробництво зросли на 11,2 % або на 96,0 тис. умов. грош. од. ($956,0\cdot 10^3 - 860,0\cdot 10^3$).

Взаємозв’язок цих трьох показників дає змогу зробити перевірку правильності виконання розрахунків:

$$\mathbf{I_{zq} = I_z \cdot I_q}.$$
$$\mathbf{I_{zq} = 1{,}145 \cdot 1{,}112 = 1{,}273} \text{ або } \mathbf{127{,}3}\ \%.$$

Ці індекси також дають змогу виявити взаємозв'язок і між їх абсолютними змінами:

$$\boldsymbol{\sum z_1 q_1 - \sum z_0 q_0 = (\sum z_1 q_1 - \sum z_0 q_1) + (\sum z_0 q_1 - \sum z_0 q_0)};$$
$$\mathbf{234{,}6 \cdot 10^3 = 138{,}6 \cdot 10^3 + 96{,}0 \cdot 10^3};$$
$$\mathbf{234{,}6 \cdot 10^3 = 234{,}6 \cdot 10^3}.$$

Отже, найбільше витрати на виготовлення рекламних роликів зросли внаслідок збільшення собівартості одного рекламного ролика, що в загальному прирості становить 59,1 % $\left(\mathbf{\frac{138{,}6 \cdot 10^3}{234{,}6 \cdot 10^3} \cdot 100}\right)$ і тільки на 40,9 % $\left(\mathbf{\frac{96{,}0 \cdot 10^3}{234{,}6 \cdot 10^3} \cdot 100}\right)$ приріст витрат був обумовлений зміною кількості виготовлених рекламних роликів.

Отже, перша велика сфера застосування економічних індексів – це порівнянна характеристика сукупностей, які складаються з непорівнянних елементів.

Індексний метод застосовується в статистиці і як аналітичне знаряддя – для оцінювання ролі окремих факторів у зміні складного явища. До таких складних явищ може бути віднесено вартість виробленої продукції (товарообіг), витрати виробництва та ін.

Формули для розрахунку загальних індексів інших показників наведені в таблиці 4.2.

4.3 Середні індекси

Крім агрегатних, у статистиці також застосовуються середні (середньозважені) індекси. До їх розрахунку вдаються тоді, коли дані, які є, не дають змоги розрахувати загальний агрегатний індекс.

Середній індекс – це індекс, розрахований як середня величина з індивідуальних індексів. Агрегатний індекс є основною формою загального індексу, тому середній індекс має бути тотожним агрератному індексу. При обчисленні середніх індексів використовуються дві форми середніх: арифметична і гармонічна.

Таблиця 4.2 – Основні формули обчислення загальних індексів

Індекс	Формула індексу	Значення чисельника індексу	Значення знаменника індексу
1	2	3	4
Індекс фізичного обсягу реалізованої продукції (за ціною)	$\mathbf{I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}}$	Вартість реалізованої продукції (товарообіг) звітного періоду за цінами базисного періоду	Вартість продукції (товарообіг), реалізованої в базисному періоді
Індекс цін	$\mathbf{I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}}$	Вартість продукції (товарообіг), реалізованої у звітному періоді	Вартість реалізованої продукції (товарообіг) звітного періоду за цінами базисного періоду
Індекс вартості	$\mathbf{I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}}$	Вартість продукції (товарообіг), реалізованої у звітному періоді	Вартість продукції (товарообіг), реалізованої в базисному періоді
Індекс фізичного обсягу виготовленої продукції (за собівартістю)	$\mathbf{I_q = \frac{\sum q_1 z_0}{\sum q_0 z_0}}$	Витрати на виробництво продукції звітного періоду за базисною собівартістю	Витрати на виробництво в базисному періоді

Продовження таблиці 4.2

1	2	3	4
Індекс собівартості	$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1}$	Витрати на виробництво у звітному періоді	Витрати на виробництво продукції звітного періоду за базисною собівартістю
Індекс витрат на виробництво	$I_{zq} = \frac{\sum z_1 q_1}{\sum z_0 q_0}$	Витрати на виробництво у звітному періоді	Витрати на виробництво в базисному періоді
Індекс фізичного обсягу виготовленої продукції (за трудомісткістю)	$I_q = \frac{\sum q_1 t_0}{\sum q_0 t_0}$	Витрати часу на виробництво продукції у звітному періоді за трудомісткістю базисного періоду	Витрати часу на виробництво продукції в базисному періоді
Індекс трудомісткості	$I_t = \frac{\sum t_1 q_1}{\sum t_0 q_1}$	Витрати часу на виробництво продукції у звітному періоді	Витрати часу на виробництво продукції у звітному періоді за трудомісткістю базисного періоду
Індекс продуктивності праці (трудовий)	$I_\omega = \frac{\sum t_0 q_1}{\sum t_1 q_1}$	Витрати часу на виробництво продукції у звітному періоді за трудоміст-кістю базисного періоду	Витрати часу на виробництво продукції у звітному періоді
Індекс витрат часу на виробництво продукції	$I_{tq} = \frac{\sum t_1 q_1}{\sum t_0 q_0} = \frac{\sum T_1}{\sum T_0}$	Витрати часу на виробництво продукції у звітному періоді	Витрати часу на виробництво продукції в базисному періоді

Продовження таблиці 4.2

1	2	3	4
Індекс фізичного обсягу виготовленої продукції (за матеріало-місткістю)	$\mathbf{I_q = \frac{\sum q_1 m_0}{\sum q_0 m_0}}$	Витрати матеріалів на виробництво продукції у звітному періоді з базисною матеріало-місткістю	Витрати матеріалів на виробництво продукції в базисному періоді
Індекс питомих витрат матеріалів	$\mathbf{I_m = \frac{\sum m_1 q_1}{\sum m_0 q_1}}$	Витрати матеріалів на виробництво продукції у звітному періоді	Витрати матеріалів на виробництво продукції у звітному періоді з базисною матеріало-місткістю
Індекс загальних матеріальних витрат	$\mathbf{I_{mq} = \frac{\sum m_1 q_1}{\sum m_0 q_0}}$	Витрати матеріалів на виробництво продукції у звітному періоді	Витрати матеріалів на виробництво продукції в базисному періоді

Розглянемо порядок розрахунку серенього арифметичного індексу за даними прикладу, що наведені у таблиці 4.3.

Таблиця 4.3 – Товарообіг та відносні зміни його фізичного обсягу

Вид продукції	Товарообіг базисного періоду, млн грн	Зміна фізичного обсягу товарообігу у звітному періоді порівняно з базисним, %
1	q_0p_0	$i_q \cdot 100-100$
А	1,2	– 1,0
Б	4,0	+ 0,5
В	2,4	+ 2,3

Потрібно визначити загальну зміну товарообігу на всі види продукції, тобто необхідно обчислити такий індекс: $\mathbf{I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}}$, але у вихідних даних відомостей про рівень $\mathbf{\sum q_1 p_0}$ немає. Тому розрахунок за цією формулою неможливий. Однак дані колонки 3 таблиці 4.3 дають змогу визначити індивідуальні індекси фізичного обсягу товарообігу для кожного з видів продукції:

$$\mathbf{i_q^A = 100 - 1{,}0 = 99{,}0\%};$$

$$\mathbf{i_q^Б = 100 + 0{,}5 = 100{,}5\%};$$

$$\mathbf{i_q^В = 100 + 2{,}3 = 102{,}3\%}.$$

Для перетворення агрегатного індексу використаємо формулу індивідуального індексу

$$\mathbf{i_q = \frac{q_1}{q_0} \Rightarrow q_1 = i_q \cdot q_0}.$$

Підставимо отриману рівність у чисельник агрегатного індексу, а знаменник залишимо без змін. Тоді формула індексу фізичного обсягу товарообігу (продукції) набуде такого вигляду:

$$\mathbf{I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \Rightarrow I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0}}. \qquad (4.4)$$

Це і буде середньоарифметичний індекс. У такому вигляді індекс фізичного обсягу товарообігу виступає як середня арифметична величина з індивідуальних індексів, зважена за вартістю продукції базисного періоду в базисних цінах $\mathbf{q_0 p_0}$. Обчислимо його за даними таблиці 4.3:

$$\mathbf{I_q = \frac{0{,}99 \cdot 1{,}2 \cdot 10^6 + 1{,}005 \cdot 4{,}0 \cdot 10^6 + 1{,}023 \cdot 2{,}4 \cdot 10^6}{1{,}2 \cdot 10^6 + 4{,}0 \cdot 10^6 + 2{,}4 \cdot 10^6}} =$$

$$= \mathbf{\frac{7{,}6632 \cdot 10^6}{7{,}600 \cdot 10^6} = 1{,}008} \text{ або } \mathbf{100{,}8}\ \%.$$

Це означає, що під впливом змін фізичного обсягу продукції у звітному періоді порівняно з базисним товарообіг збільшився на 0,8 % або на 63,2 тис. грн ($7663{,}2 \cdot 10^3 - 7600{,}0 \cdot 10^3$).

Агрегатний індекс може бути перетворений не тільки в середній арифметичний, а й у середній гармонічний індекс. Розглянемо це перетворення на прикладі індексу цін. Для цього за допомогою індивідуального індексу цін $\mathbf{i_p = \frac{p_1}{p_0}}$ визначимо, чому дорівнювала ціна продукції в базисному періоді: $\mathbf{p_0 = \frac{p_1}{i_p}}$.

Використаємо цю рівність і в знаменнику агрегатного індексу зробимо відповідну заміну, а чисельник залишимо без змін. Тоді формула індексу набуде такого вигляду:

$$\mathbf{I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} \Rightarrow I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}}. \qquad (4.5)$$

У такому вигляді індекс цін виступає як середня гармонічна з індивідуальних індексів цін, зважених за сумою фактичного товарообігу звітного періоду $\mathbf{p_1 q_1}$.

Проведемо розрахунок середнього гармонічного індексу на прикладі таблиці 4.4.

Таблиця 4.4 – Обсяг реалізованої продукції та відносні зміни цін

Вид продукції	Продукція, реалізована у звітному періоді, млн грн	Зміни цін на реалізовану продукцію у звітному періоді порівняно з базисним, %
1	$q_1 p_1$	$i_p100-100$
А	5,0	+ 2,5
Б	3,0	– 0,4
В	6,0	без змін

Потрібно визначити загальну зміну цін. Спершу визначимо індивідуальні індекси цін, а потім використаємо їх для розрахунку середнього гармонічного індексу:

$$\mathbf{i_p^А = 100 + 2{,}5 = 102{,}5\%} \ ; \quad \mathbf{i_p^Б = 100 - 0{,}4 = 99{,}6\%} \ ; \quad \mathbf{i_p^В = 100\%} \ ;$$

$$\mathbf{I_p} = \frac{\mathbf{5{,}0 \cdot 10^6 + 3{,}0 \cdot 10^6 + 6{,}0 \cdot 10^6}}{\frac{\mathbf{5{,}0 \cdot 10^6}}{\mathbf{1{,}025}} + \frac{\mathbf{3{,}0 \cdot 10^6}}{\mathbf{0{,}996}} + \frac{\mathbf{6{,}0 \cdot 10^6}}{\mathbf{1{,}000}}} = \frac{\mathbf{14{,}0 \cdot 10^6}}{\mathbf{13{,}89 \cdot 10^6}} = \mathbf{1{,}008} \text{ або } \mathbf{100{,}8}\%.$$

Отже, під впливом змін цін на реалізовану продукцію у звітному періоді порівняно з базисним загальний розмір реалізації збільшився на 0,8 % або 110 тис. грн ($14\,000{,}0 \cdot 10^3 - 13\,890{,}0 \cdot 10^3$).

У таблиці 4.5 наведемо загальну схему перетворення аргегатних індексів.

Таблиця 4.5 – Схема перетворення агрегатних індексів

Індекс	Індивідуальний індекс	Перетворення індивідуального індексу	Агрегатний індекс	Середній арифметичний індекс	Середній гармонічний індекс
Фізичного обсягу продукції	$i_q = \frac{q_1}{q_0}$	$q_1 = i_q \cdot q_0$	$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}$	$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0}$	–
Цін	$i_p = \frac{p_1}{p_0}$	$p_0 = \frac{p_1}{i_p}$	$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$	–	$I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}$
Собівартості	$i_z = \frac{z_1}{z_0}$	$z_0 = \frac{z_1}{i_z}$	$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1}$	–	$I_z = \frac{\sum z_1 q_1}{\sum \frac{z_1 q_1}{i_z}}$
Продуктивності праці	$i_\omega = \frac{t_0}{t_1}$	$t_0 = i_\omega \cdot t_1$	$I_\omega = \frac{\sum t_0 q_1}{\sum t_1 q_1}$	$I_\omega = \frac{\sum i_\omega t_1 q_1}{\sum t_1 q_1}$	–

4.4 Індекси структурних зсувів

Інколи потрібно вивчити динаміку суспільних явищ, рівні яких виражені середніми величинами (середня заробітна плата, середня собівартість, середня продуктивність праці, середня врожайність тощо).

Оскільки на зміну середнього значення показника впливають два фактори, виникає задача визначити ступінь впливу кожного з факторів на загальну динаміку середньої.

Ця задача зазвичай розв'язується за допомогою індексного методу, тобто шляхом побудови системи взаємопов'язаних індексів, до якої включають три індекси: змінного складу, постійного складу, структурних зсувів.

Індексом змінного складу називають індекс, який виражає співвідношення середніх рівнів виучуваного явища, які належать до різних періодів часу. Індекс змінного складу можна розкласти на два індекси (множники). Перший множник показує, як змінюється середній рівень під впливом змін рівня самого якісного показника. Це індекс постійного (фіксованого) складу. Другий індекс показує вплив зміни структури явища і має назву індексу структурних зсувів.

Отже,

$$\mathbf{I}_{\text{змінного складу}} = \mathbf{I}_{\text{фіксованого складу}} \cdot \mathbf{I}_{\text{структурних зсувів}}. \qquad (4.6)$$

Проілюструємо особливості побудови цих індексів на прикладі індексу продуктивності праці.

Продуктивність праці $\boldsymbol{\omega}$ показує кількість продукції $\mathbf{q}$, виробленої в одиницю часу $\mathbf{T}$, або витрати робочого часу на виробництво одиниці продукції $\mathbf{t}$.

Тоді $\boldsymbol{\omega} = \frac{\mathbf{q}}{\mathbf{T}}$ – це прямий показник продуктивності, а $\mathbf{t} = \frac{\mathbf{T}}{\mathbf{q}}$ – зворотний показник. Звідси індивідуальний індекс продуктивності праці буде розраховуватись за формулою

$$\mathbf{i}_{\omega} = \frac{\omega_1}{\omega_0} = \frac{q_1}{T_1} : \frac{q_0}{T_0} = \frac{q_1}{t_1 q_1} : \frac{q_0}{t_0 q_0} = \frac{t_0}{t_1}.$$

Для побудови агрегатного індексу продуктивності праці потрібно витрати робочого часу на виробництво одиниці продукції зважити на кількість продукції, яка була вироблена у звітному періоді:

$$\mathbf{I_{\omega}} = \frac{\sum \mathbf{t_0 q_1}}{\sum \mathbf{t_1 q_1}}, \tag{4.7}$$

де $\mathbf{t_1 q_1}$ – фактичні витрати часу на виробництво всієї продукції у звітному періоді;

$\mathbf{t_0 q_1}$ – умовні витрати часу, які вказують, скільки часу необхідно було б витратити на виробництво всієї продукції у звітному періоді порівняно з базисним, тобто з базисною трудомісткістю.

Формула (4.7) описує трудовий індекс продуктивності праці. У ньому, на відміну від усіх попередніх індексів, базисне значення показника, що вивчається, $\mathbf{t_0}$ міститься в чисельнику, а не в знаменнику, тому що час і продуктивність праці перебувають у зворотній залежності. Чим менші витрати часу на виробництво одиниці продукції, тим вища (за інших рівних умов) продуктивність праці.

Тому, щоб при вирахуванні індексу продуктивності праці отримати її прямий показник, $\mathbf{t_0}$ необхідно записати в чисельник, а $\mathbf{t_1}$ – у знаменник індексу.

Однак при розрахунку індексів продуктивності праці за формулою (4.7) на практиці часто трапляються труднощі, тому що для цього необхідно мати дані про норми витрат робочого часу в базисному періоді за всією номенклатурою виробів поточного періоду. Тому на практиці в статистичній роботі широко використовується вартісний метод обчислення індексів продуктивності праці:

$$\mathbf{I_{\omega}} = \frac{\sum \mathbf{q_1 p_0}}{\sum \mathbf{T_1}} : \frac{\sum \mathbf{q_0 p_0}}{\sum \mathbf{T_0}}, \tag{4.8}$$

де $\sum \mathbf{q_1 p_0}$ і $\sum \mathbf{q_0 p_0}$ – вартість виготовленої продукції у звітному та базисному періодах у порівнянних цінах;

$\sum \mathbf{T_1}$ і $\sum \mathbf{T_0}$ – загальні витрати часу на виробництво продукції у звітному і базисному періоді.

Це індекс продуктивності праці змінного складу, який указує одночасно як ступінь зміни продуктивності праці, так і ступінь зміни структури витрат часу (або кількості робітників):

$$\mathbf{I_{\omega d} = \frac{\sum \omega_1 d_1}{\sum \omega_0 d_0}}, \tag{4.9}$$

де $\boldsymbol{\omega_1, \omega_0}$ – рівень продуктивності праці у звітному та базисному періодах;

$\mathbf{d_1}$, $\mathbf{d_0}$ – частка витрат робочого часу на виробництво кожного виду продукції.

Для того, щоб абстрагуватися від впливу структурних зсувів, потрібно вирахувати індекс продуктивності праці постійного складу, на розмір якого впливає тільки ступінь зміни продуктивності праці:

$$\mathbf{I_{\omega} = \frac{\sum \omega_1 d_1}{\sum \omega_0 d_1}}. \tag{4.10}$$

Для визначення зміни структури витрат робочого часу (або кількості робітників) необхідно індекс змінного складу поділити на індекс фіксованого (постійного) складу, тобто:

$$\mathbf{I_{\text{структурних зсувів}} = \frac{I_{\text{змінного складу}}}{I_{\text{фіксованого складу}}}}, \tag{4.11}$$

$$\mathbf{I_d = \frac{I_{\omega d}}{I_{\omega}} = \frac{\sum \omega_1 d_1}{\sum \omega_0 d_0} : \frac{\sum \omega_1 d_1}{\sum \omega_0 d_1} = \frac{\sum \omega_0 d_1}{\sum \omega_0 d_0}}. \tag{4.12}$$

Розглянемо порядок розрахунку перелічених індексів на прикладі таблиці 4.6.

Таблиця 4.6 – Обсяг продукції і кількість робітників підприємств

Підпри-ємство	2018		2019	
	Кількість виробленої продукції у порівнянних цінах, тис. грн	Кількість робітників, люд	Кількість виробленої продукції у порівнянних цінах, тис. грн	Кількість робітників, люд
1	37 500	1 500	39 750	1 500
2	2 250	300	5 400	600

Для розрахунку індексів змінного складу, постійного складу та індексу структурних зсувів складемо розрахункову таблицю 4.7.

Таблиця 4.7 – До розрахунку індексів продуктивності праці

Підприємство	2018 р			2019 р			Питома вага робітників, %		Індекс, %
	Кількість виробленої продукції у порівнянних цінах, тис. грн	Кількість робітників, люд	Виробіток на одного робітника, тис. грн	Кількість виробленої продукції у порівнянних цінах, тис. грн	Кількість робітників, люд	Виробіток на одного робітника, тис. грн	2018 р.	2019 р.	
	q_0p_0	T_0	ω_0	q_1p_0	T_1	ω_1	d_0	d_1	I_ω
1	37500	1500	25,0	39750	1500	26,5	83,3	71,4	106,0
2	2250	300	7,5	5400	600	9,0	16,7	28,6	120,0
Разом	39750	1800	22,08	45150	2100	21,50	100,0	100,0	97,4

Таким чином, на підприємстві 1 продуктивність праці у 2019 році становила 106,0 % відносно 2018 року, тобто збільшилась на 6,0 %. Аналогічно на підприємстві 2 продуктивність праці зросла у 2019 році порівняно з 2018 роком на 20,0 %. Тобто по кожному з підприємств було збільшення продуктивності праці. Однак середня продуктивність праці по обох підприємствах зменшилась на 2,6 % (97,4 – 100,0).

Розрахунки проведені без помилок. У нашому випадку маємо так званий «економічний парадокс»: за окремими елементами показник збільшується, а за всією сукупністю – навпаки, зменшується. Він підкреслює, що на зміну середнього рівня продуктивності праці впливає не тільки зміна самої продуктивності, але й зміна структури кількості робітників.

Це ж саме показує й індекс продуктивності праці змінного складу:

$$\mathbf{I_{\omega d} = \frac{45\ 150 \cdot 10^6}{2\ 100 \cdot 10^6} : \frac{39\ 750 \cdot 10^6}{1\ 800 \cdot 10^6} = \frac{21{,}50 \cdot 10^6}{22{,}08 \cdot 10^6} = 0{,}974} \text{ або } \mathbf{97{,}4}\ \%$$

або

$$\mathbf{I_{\omega d} = \frac{26{,}5 \cdot 10^3 \cdot 0{,}714 + 9{,}0 \cdot 10^3 \cdot 0{,}286}{25{,}0 \cdot 10^3 \cdot 0{,}833 + 7{,}5 \cdot 10^3 \cdot 0{,}167} = \frac{21{,}50 \cdot 10^3}{22{,}08 \cdot 10^3} = 0{,}974} \text{ або } \mathbf{97{,}4}\ \%.$$

Для визначення впливу зміни самої продуктивності праці і відокремлення від впливу структурних зсувів обчислимо індекс постійного складу:

$$\mathbf{I_{\omega} = \frac{26{,}5 \cdot 10^3 \cdot 0{,}714 + 9{,}0 \cdot 10^3 \cdot 0{,}286}{25{,}0 \cdot 10^3 \cdot 0{,}714 + 7{,}5 \cdot 10^3 \cdot 0{,}286} = \frac{21{,}50 \cdot 10^3}{19{,}995 \cdot 10^3} = 1{,}075} \text{ або } \mathbf{107{,}5}\ \%.$$

Отже, внаслідок зростання продуктивності праці окремо по кожному з підприємств у цілому по двох підприємствах вона збільшилась на 7,5 % (107,5 – 100).

Для визначення впливу на середню продуктивність праці структурних зсувів розрахуємо індекс структурних зсувів:

$$\mathbf{I_{d} = \frac{25{,}0 \cdot 10^3 \cdot 0{,}714 + 7{,}5 \cdot 10^3 \cdot 0{,}286}{25{,}0 \cdot 10^3 \cdot 0{,}833 + 7{,}5 \cdot 10^3 \cdot 0{,}167} = \frac{19{,}995 \cdot 10^3}{22{,}08 \cdot 10^3} = 0{,}906} \text{ або } \mathbf{90{,}6}\ \%$$

або

$$\mathbf{I_{\text{структурних зсувів}} = \frac{0{,}974}{1{,}075} = 0{,}906} \text{ або } \mathbf{90{,}6}\ \%.$$

Таким чином, завдяки змінам структури робітників на двох підприємствах (за рахунок збільшення у два рази кількості робітників на підприємстві з більш низькою продуктивністю праці) середня продуктивність праці зменшилась на 9,4 % (90,6 – 100), хоч власне виробіток і збільшився на 7,5 %. Саме ці дві зміни під впливом обох показників призвели до загального зменшення продуктивності праці.

Кожен з наведних індексів виконує своє економічне завдання, а саме:

– індекс змінного складу характеризує загальну зміну розмірів якісного показника (його середнього рівня);

– індекс фіксованого (постійного) складу показує, як змінився рівень якісного показника незалежно від змін у структурі сукупності;

– індекс структурних зсувів показує зміни якісного показника під впливом змін у структурі сукупності.

Так само, як і для продуктивності праці, можна будувати індекси будь-яких якісних показників, рівень яких виражений середніми величинами.

4.5 Базисні та ланцюгові індекси

Для вирахування індексів, як і будь-якої іншої відносної величини, необхідно мати дані за два періоди часу. Період, рівень якого порівнюється, має назву звітного або поточного періоду, період, з рівнем якого проводиться порівняння, має назву базисного.

Якщо замість двох надана більша кількість періодів часу, залежно від бази порівняння системи індексів бувають базисними та ланцюговими.

І базисні, і ланцюгові індекси мають визначене значення в економічному аналізі:

– базисні характеризують зміни явища за досить довгий період часу щодо величини базисного періоду;

– ланцюгові індекси використовують у разі виникнення потреби стеження за поточними змінами явищ.

Якщо базисні та ланцюгові індекси охоплюють той самий період часу, між ними є певний взаємозв’язок. Взаємозв’язок, який є між базисними та ланцюговими індексами, дає можливість розрахувати базисні індекси за даними про ланцюгові і, навпаки, ланцюгові – за даними про базисні.

Наприклад, є дані за 5 років (з 2015 по 2019 роки) про деякий показник **у**. Тоді, з одного боку, добуток ланцюгових індексів дорівнює відповідному базисному:

$$\frac{y_{16}}{y_{15}} \cdot \frac{y_{17}}{y_{16}} \cdot \frac{y_{18}}{y_{17}} \cdot \frac{y_{19}}{y_{18}} = \frac{y_{2019}}{y_{2015}},$$

де $\frac{y_{16}}{y_{15}}$, $\frac{y_{17}}{y_{16}}$, $\frac{y_{18}}{y_{17}}$, $\frac{y_{19}}{y_{18}}$ – ланцюгові індекси показника **у**;

$\frac{y_{2019}}{y_{2015}}$ – базисний індекс показника **у**.

У свою чергу відношення наступного базисного індексу до попереднього дорівнює відповідному ланцюговому:

$$\frac{y_{19}}{y_{15}} : \frac{y_{18}}{y_{15}} = \frac{y_{2019}}{y_{2018}},$$

де $\frac{y_{19}}{y_{15}}$, $\frac{y_{18}}{y_{15}}$ – базисні індекси показника **у**;

$\frac{y_{2019}}{y_{2018}}$ – ланцюговий індекс показника **у**.

Ряд послідовно побудованих індексів має назву системи індексів. Системи ланцюгових та базисних індексів можуть бути побудовані як для індивідуальних, так і для загальних індексів. У таблиці 4.8 показано порядок побудови деяких індексів. Аналогічно можна будувати системи індексів для будь-яких показників.

Таблиця 4.8 – Системи індексів

Індекс	Система індексів	
	базисних	ланцюгових
Індивідуальний індекс цін	$\frac{p_1}{p_0};\frac{p_2}{p_0};...;\frac{p_n}{p_0}$	$\frac{p_1}{p_0};\frac{p_2}{p_1};...;\frac{p_n}{p_{n-1}}$
Індивідуальний індекс собівартості	$\frac{z_1}{z_0};\frac{z_2}{z_0};...;\frac{z_n}{z_0}$	$\frac{z_1}{z_0};\frac{z_2}{z_1};...;\frac{z_n}{z_{n-1}}$
Індивідуальний індекс трудомісткості	$\frac{t_1}{t_0};\frac{t_2}{t_0};...;\frac{t_n}{t_0}$	$\frac{t_1}{t_0};\frac{t_2}{t_1};...;\frac{t_n}{t_{n-1}}$
Індивідуальний індекс фізичного обсягу продукції	$\frac{q_1}{q_0};\frac{q_2}{q_0};...;\frac{q_n}{q_0}$	$\frac{q_1}{q_0};\frac{q_2}{q_1};...;\frac{q_n}{q_{n-1}}$
Загальний індекс цін	$\frac{p_1q_1}{p_0q_1};\frac{p_2q_2}{p_0q_2};...;\frac{p_nq_n}{p_0q_n}$	$\frac{p_1q_1}{p_0q_1};\frac{p_2q_2}{p_1q_2};...;\frac{p_nq_n}{p_{n-1}q_n}$
Загальний індекс собівартості	$\frac{z_1q_1}{z_0q_1};\frac{z_2q_2}{z_0q_2};...;\frac{z_nq_n}{z_0q_n}$	$\frac{z_1q_1}{z_0q_1`};\frac{z_2q_2}{z_1q_2};...;\frac{z_nq_n}{z_{n-1}q_n}$
Загальний індекс трудомісткості	$\frac{t_1q_1}{t_0q_1};\frac{t_2q_2}{t_0q_2};...;\frac{t_nq_n}{t_0q_n}$	$\frac{t_1q_1}{t_0q_1};\frac{t_2q_2}{t_1q_2};...;\frac{t_nq_n}{t_{n-1}q_n}$
Загальний індекс фізичного обсягу продукції	$\frac{q_1p_0}{q_0p_0};\frac{q_2p_0}{q_0p_0};...;\frac{q_np_0}{q_0p_0}$ $\frac{q_1z_0}{q_0z_0};\frac{q_2z_0}{q_0z_0};...;\frac{q_nz_0}{q_0z_0}$ $\frac{q_1t_0}{q_0t_0};\frac{q_2t_0}{q_0t_0};...;\frac{q_nt_0}{q_0t_0}$	$\frac{q_1p_0}{q_0p_0};\frac{q_2p_1}{q_1p_1};...;\frac{q_np_{n-1}}{q_{n-1}p_{n-1}}$ $\frac{q_1z_0}{q_0z_0};\frac{q_2z_1}{q_1z_1};...;\frac{q_nz_{n-1}}{q_{n-1}z_{n-1}}$ $\frac{q_1t_0}{q_0t_0};\frac{q_2t_1}{q_1t_1};...;\frac{q_nt_{n-1}}{q_{n-1}t_{n-1}}$
Індекс вартості	$\frac{p_1q_1}{p_0q_0};\frac{p_2q_2}{p_0q_0};...;\frac{p_nq_n}{p_0q_0}$	$\frac{p_1q_1}{p_0q_0};\frac{p_2q_2}{p_1q_1};...;\frac{p_nq_n}{p_{n-1}q_{n-1}}$
Індекс витрат	$\frac{z_1q_1}{z_0q_0};\frac{z_2q_2}{z_0q_0};...;\frac{z_nq_n}{z_0q_0}$	$\frac{z_1q_1}{z_0q_0};\frac{z_2q_2}{z_1q_1};...;\frac{z_nq_n}{z_{n-1}q_{n-1}}$
Індекс витрат часу	$\frac{t_1q_1}{t_0q_0};\frac{t_2q_2}{t_0q_0};...;\frac{t_nq_n}{t_0q_0}$	$\frac{t_1q_1}{t_0q_0};\frac{t_2q_2}{t_1q_1};...;\frac{t_nq_n}{t_{n-1}q_{n-1}}$

Питання до самоконтролю

1 Метод економічних індексів та його характеристика.
2 Порядок побудови агрегатних індексів.
3 Застосування середніх індексів.
4 Індекси структурних зсувів.
5 Базисні та ланцюгові індекси.
6 Системи ланцюгових та базисних індексів.

ТЕМА 5. Вибірковий метод

План

5.1 Поняття про вибіркове спостереження та його значення.
5.2 Характеристика вибіркової і генеральної сукупності.
5.3 Помилки вибірки.
5.4 Методи і способи відбору.
5.5 Визначення необхідного обсягу вибірки.

5.1 Поняття про вибіркове спостереження та його значення

У багатьох випадках здійснити суцільне спостереження дуже важко або зовсім не можливо. Тоді виникає питання: чи не можна обстежити окрему частину сукупності і за цією частиною судити про всю сукупність у цілому? Як показує практика, це зробити можна і за такої умови отримані уявлення про сукупність достатньо точні і правильні. Одним з видів такого несуцільного спостереження є вибіркове спостереження. При вибірковому спостереженні дослідженню підлягає деяка частина сукупності, відібрана у певному порядку, а узагальнюючі показники, які характеризують цю досліджувану частину сукупності, поширюються на всю сукупність. Вибіркове спостереження – найбільш поширений вид несуцільного спостереження в економіко-статистичному аналізі. З ряду причин у багатьох випадках випадковому спостереженню віддається перевага перед суцільним. До цих переваг належать:

1) оперативність одержання інформації;

2) економія часу і засобів;

3) досягнення більшої точності результатів спостереження;

4) зведення до мінімуму псування або навіть знищення досліджуваних одиниць.

При вибірковому спостереженні досліджується менша кількість одиниць сукупності, що скорочує обсяг роботи. Завдяки цьому одержують результати в більш короткі терміни, скорочується час дослідження й одночасно знижується його вартість. Відомо, що перевірка якості продукції часто пов'язана з псуванням або повним знищенням.

Прикладами можуть бути випробування крила літака на міцність, електролампи на тривалість горіння, пряжі на розривання тощо. Застосування суцільного спостереження призвело б до безглуздого знищення досліджуваних одиниць, тому в таких випадках використовують вибіркове обстеження. При величезному обсязі статистичних робіт здешевлення їх вартості, прискорення термінів проведення при досягненні більшої точності результатів мають велике значення. Тому вибірковий метод набув великого поширення у статистичних дослідженнях.

5.2 Характеристика вибіркової і генеральної сукупності

У статистиці заведено називати сукупність, з якої робиться відбір одиниць для обстеження, генеральною сукупністю, а кількість одиниць, відібраних для обстеження, вибірковою сукупністю.

Генеральна і вибіркова сукупності характеризуються такими узагальнюючими показниками: середньою генеральною і вибірковою, дисперсією генеральною і вибірковою, часткою генеральною і часткою (частістю) вибірковою (таблиця 5.1).

Таблиця 5.1 – Символи основних характеристик генеральної і вибіркової сукупностей

Характеристика	Генеральна сукупність	Вибіркова сукупність
Обсяг сукупності	$\mathbf{N}$	$\mathbf{n}$
Середній розмір ознаки	$\bar{\mathbf{x}}$	$\tilde{\mathbf{x}}$
Дисперсія кількісної ознаки	$\boldsymbol{\sigma}_0^2$	$\boldsymbol{\sigma}^2$
Дисперсія частки	$\boldsymbol{\sigma}_p^2 = \mathbf{p} \cdot \mathbf{q} = \mathbf{p}(\mathbf{1} - \mathbf{p})$	$\boldsymbol{\sigma}_w = \mathbf{w}(\mathbf{1} - \mathbf{w})$
Кількість одиниць, які мають альтернативні ознаки	$\mathbf{M}$	$\mathbf{m}$
Частка одиниць, які мають альтернативні ознаки	$\mathbf{p} = \frac{\mathbf{M}}{\mathbf{N}}$	$\mathbf{w} = \frac{\mathbf{m}}{\mathbf{n}}$

Припустимо, що на заводі працює 1000 робітників, з них 980 виконують і перевиконують денну норму виробітку.

Тобто

N=1000 роб. M=980 роб.	Тоді
P=?	$\mathbf{p} = \frac{\mathbf{M}}{\mathbf{N}}$; **p=980/1000=0,98 або 98 %**.

Отже, 98 % усіх робітників заводу виконують і перевиконують денну норму виробітку.

Припустимо далі, що було проведено 10 % обстеження, тобто із генеральної сукупності відібрали для обстеження тільки 100 робітників, серед яких тільки 95 виконують і перевиконують денну норму виробітку.

Тобто

N=1000 роб. n=100 роб. m=95 роб.	Тоді $\mathbf{w} = \frac{\mathbf{m}}{\mathbf{n}}$; $\mathbf{w} = \frac{\mathbf{95}}{\mathbf{100}} = \mathbf{0.95}$ **або 95%**.
w-?	

Отже, 95 % робітників заводу виконують і перевиконують денну норму виробітку.

Але це можна стверджувати не з абсолютною достовірністю, а тільки з певною мірою ймовірності.

Питання про те, наскільки велика різниця між вибірковими і генеральними характеристиками, з якою і можна судити про цю різницю, вивчається на основі так званих граничних теорем теорії ймовірностей, які описують поведінку масових явищ.

Якщо чисельність вибірки досить велика, то вибіркові характеристики достатньо добре відтворюють генеральні характеристики. Тому вони мають важливе значення і в обґрунтуванні вибіркового спостереження.

При масових спостереженнях емпіричних частот явище підлягає закону нормального розподілу.

Нормальний розподіл у свою чергу показує, що більша частина величин зосереджується біля генеральної середньої. Близько 68,3 % чисельності вибіркових середніх не буде виходити за межі $\pm\boldsymbol{\sigma}$ генеральної середньої, 95,4 % цієї чисельності буде міститися в межах $\pm 2\ \boldsymbol{\sigma}$ і 99,7 % їх не вийде за межі $\pm 3\boldsymbol{\sigma}$.

5.3 Помилки вибірки

При проведенні вибіркового спостереження не можна одержати абсолютно точні дані, тому що обстеженню підлягає не вся сукупність, а тільки частина. Тому при проведенні вибіркового спостереження неминучі деякі огріхи, помилки. На відміну від статистичного спостереження в цілому, при проведенні якого виникають помилки реєстрації, вибірковому спостереженню властиві помилки, які називають помилками репрезентативності. Вони являють собою розмір розходження між узагальнюючими величинами вибіркової сукупності і величинами цих показників у генеральній сукупності

Помилки репрезентативності можуть бути систематичними і випадковими. Систематичні помилки репрезентативності – це відхилення, які виникають унаслідок порушення принципів випадковості відбору одиниць сукупності для спостереження.

Прикладом може бути відбір із студентів ЗВО найбільш підготовлених осіб для дослідження успішності.

Випадкові помилки репрезентативності – це відхилення, які виникають через несуцільний характер спостереження внаслідок того, що вибіркова сукупність недостатньо точно відтворює всю сукупність у цілому. Випадкові помилки можуть бути доведені до незначних розмірів, а головне, – величини і їхні межі можуть бути визначені досить точно на основі закону великих чисел і теорії ймовірностей.

Найважливішою умовою проведення вибіркового спостереження є правильний відбір одиниць сукупності, а саме:

а) строго об'єктивний вибір одиниць, при якому кожна з них отримувала б однакову можливість потрапити у вибірку,

б) достатня кількість відібраних одиниць.

При додержуванні цих умов вибірка буде репрезентативною, тобто представницькою.

Для судження про право поширення даних вибіркового спостереження на генеральну сукупність визначають величину помилок між узагальнюючими показниками вибіркової і генеральної сукупностей. Звичайно зіставляють такі показники:

1) середню вибіркової сукупності із середньою генеральною, у результаті чого одержують помилку середньої: $|\tilde{\mathbf{x}} - \bar{\mathbf{x}}| = \mathbf{\Delta_x}$;

2) частість вибіркової сукупності з генеральною часткою, що дає змогу визначити помилку частки $|\mathbf{w} - \mathbf{p}| = \mathbf{\Delta_w}$.

Вибіркові середня і частість, а також середня і частка генеральні є змінними величинами, тобто вони можуть набувати різних значень. Отже, помилки вибірки також є змінними величинами і також можуть набувати різних значень. Тому визначається середня з можливих помилок, яка позначається буквою $\boldsymbol{\mu}$(мю).

З математичних теорем закону великих чисел випливає, що середня помилка буде дорівнювати

$$\boldsymbol{\mu}_{\mathbf{x}} = \sqrt{\frac{\boldsymbol{\sigma}_0^2}{\mathbf{n}}}. \tag{5.1}$$

Якщо вибіркове спостереження застосовується для визначення частки ознаки, то середня помилка частки обчислюється за формулою

$$\mu_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}. \qquad (5.2)$$

З формули видно, що для зменшення середньої помилки вибірки, наприклад у два рази, обсяг вибірки треба збільшити в чотири рази.

При безповторному відборі середня помилка середньої дорівнює

$$\mu_x = \sqrt{\frac{\sigma_0^2}{n}\left(1-\frac{n}{N}\right)}. \qquad (5.3)$$

Помилка частки

$$\mu_p = \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)}. \qquad (5.4)$$

Множник $\left(1-\frac{n}{N}\right)$ завжди менший за одиницю, бо **n**<**N**. Тому величина середньої помилки вибірки при безповторному відборі менша ніж при повторному.

У вищенаведених формулах σ_0^2 і $p(1-p)$ – характеристики генеральної сукупності, які при вибірковому спостереженні невідомі. На практиці їх заміняють аналогічними характеристиками вибіркової сукупності. Ця дія цілком правомірна, бо основана на законі великих чисел. У математичній статистиці доведено, що співвідношення між дисперсіями генеральної і вибіркової сукупностей виражаються формулами:

$$\sigma_0^2 = \sigma^2\frac{n}{n-1}; \qquad pq = w(1-w)\frac{n}{n-1}. \qquad (5.5)$$

Отже, вибіркова дисперсія дещо менша від генеральної (на величину $\frac{\mathbf{n}}{\mathbf{n-1}}$). При досить великих n величина $\frac{\mathbf{n}}{\mathbf{n-1}}$ прямує до 1 і тому наближено можна вважати, що вибіркова і генеральна дисперсії рівні, тобто $\boldsymbol{\sigma^2 \approx \sigma_0^2}$ **і** $\mathbf{w(1\text{-}w) \approx pq}$.

Наприклад, при **n=150** величина $\frac{\mathbf{n}}{\mathbf{n-1}}$ буде дорівнювати $\frac{\mathbf{150}}{\mathbf{149}} = \mathbf{1{,}007},$ а при **n=500** $\frac{\mathbf{n}}{\mathbf{n-1}} = \frac{\mathbf{500}}{\mathbf{499}} = \mathbf{1{,}002}$ та ін.

Тому у звичайних вибірках величиною $\frac{\mathbf{n}}{\mathbf{n-1}}$ нехтують, але обов'язково враховують у малих вибірках.

З урахуванням сказаного, формули середньої помилки вибірки подано у таблиці 5.2.

Таблиця 5.2

Спосіб відбору	Формули розрахунку	
	для середньої	для частки
Повторний	$\boldsymbol{\mu_x = \sqrt{\frac{\sigma_0^2}{n}}}$	$\boldsymbol{\mu_x = \sqrt{\frac{w(1-w)}{n}}}$
Безповторний	$\boldsymbol{\mu_x = \sqrt{\frac{\sigma_0^2}{n}\left(1-\frac{n}{N}\right)}}$	$\boldsymbol{\mu_x = \sqrt{\left(1-\frac{n}{N}\right)\frac{w(1-w)}{n}}}$

Для вирішення практичних задач вибіркового обстеження середньої помилки вибірки недостатньо, тому що при обчисленні помилки конкретної вибірки фактична помилка може бути більшою чи меншою від середньої помилки вибірки. Тому користуються не середньою, а граничною помилкою вибірки.

Теоретичною основою вибіркового методу є теореми П. Я. Чебишева і А. М. Ляпунова, Я. Бернуллі і С. Пуассона Згідно з теоремою А. М. Ляпунова, при досить великій кількості незалежних спостережень у генеральній сукупності з кінцевою середньою й обмеженою дисперсією ймовірність того, що розходження між вибірковою і генеральною середньою $|\mathbf{\tilde{x} - \bar{x}}|$ не перевищать за абсолютною величиною деяку величину **tμ**, дорівнює інтегралу Лапласа. Це можна записати таким чином:

$$\mathbf{p}\left(\left|\tilde{\mathbf{x}}-\overline{\mathbf{x}}\right| \le \mathbf{t\mu}\right)=\mathbf{F(t)}, \tag{5.6}$$

де **F(t)** – нормована функція Лапласа.

Окремим випадком теореми П. Я. Чебишева є теорема Я. Бернуллі. Ця теорема розглядає помилку вибірки для альтернативної ознаки, тобто ознаки, у якої можливі тільки два виходи: наявність ознаки **(1)** і її відсутність **(0).**

Теорема Я. Бернуллі стверджує, що при досить великому обсязі вибірки у міру його збільшення ймовірність розходження між частістю **w** генеральною часткою **p** буде прямувати до одиниці.

У математичних символах теорему Я. Бернуллі можна записати так:

$$\mathbf{p}\left(\left|\mathbf{w}-\mathbf{p}\right| \le \mathbf{t\mu}\right)\to 1. \tag{5.7}$$

З огляду на те, що ймовірність розходження між частістю і генеральною часткою відповідає закону нормального розподілу, цю ймовірність можна знайти за функцією **F(t)** залежно від величини **t**, яка задається.

Тоді

$$\mathbf{p}\left(\left|\mathbf{w}-\mathbf{p}\right| \le \mathbf{t\mu}\right)=\mathbf{F(t)}. \tag{5.8}$$

Як видно з формули, з певною ймовірністю можна стверджувати, що ці відхилення не перевисять деяку величину **t μ**, яку можна назвати граничною помилкою вибірки **Δ.**

Отже, гранична помилка вибірки залежить від певної ймовірності і пов'язана із середньою помилкою таким рівнянням:

$$\mathbf{\Delta = t\mu}, \tag{5.9}$$

де **t** – коефіцієнт довіри.

Коефіцієнт довіри залежить від імовірності, з якою можна стверджувати, що гранична помилка не перевищить t-кратну середню помилку.

Щоб визначити величину ймовірності для різних значень **t**, на практиці користуються готовою таблицею значень функції **F(t)**.

Наведемо найбільш часто вживані рівні довірчої ймовірності і відповідні значення **t** для вибірок досить великого обсягу (**n≥30**).

t	1,00	1,96	2,00	2,58	3,00
F(t)	0,689	0,950	0,954	0,990	0,997

Із збільшенням **t** імовірність **p** наближається до одиниці.

Вибіркове спостереження проводиться з метою поширення результатів, отриманих за даними вибірки, на генеральну сукупність.

Знаючи вибіркову середню ($\tilde{\mathbf{x}}$) і граничну помилку вибірки $\mathbf{\Delta_w}$, можна визначити межі, у яких з цією ймовірністю буде міститися генеральна середня ($\bar{\mathbf{x}}$):

$$\mathbf{x} - \mathbf{\Delta_x} \le \bar{\mathbf{x}} \le \mathbf{x} + \mathbf{\Delta_x}. \tag{5.10}$$

Аналогічно можуть бути записані довірчі межі генеральної частки:

$$\mathbf{w} - \mathbf{\Delta_w} \le \mathbf{p} \le \mathbf{w} + \mathbf{\Delta_w}. \tag{5.11}$$

Величина довірчого інтервалу для генеральної частки залежить від величини граничної помилки вибірки $\mathbf{\Delta_x}$ або $\mathbf{\Delta_w}$. Чим більша величина граничної помилки вибірки, тим більша величина довірчого інтервалу і тим нижча точність оцінки.

5.4 Методи і способи відбору

Відбір одиниць із генеральної сукупності можна проводити по- різному. Систему організації відбору одиниць із генеральної сукупності називають способом відбору.

Залежно від того, беруть участь відібрані одиниці в подальшій вибірці чи ні, розрізняють такі методи відбору: повторний і безповторний.

Повторним називається такий метод відбору, при якому відібрана одного разу одиниця повертається назад у генеральну сукупність і знову бере участь у вибірці. При повторному відборі залишається постійною ймовірність потрапити у вибірку для всіх одиниць відбору.

Безповторним називається такий метод відбору, при якому один раз відібрана одиниця назад у сукупність одиниць, з яких відбувається відбір, не повертається.

Повторний і безповторний методи відбору застосовуються в поєднанні з іншими видами відбору. У практичній діяльності застосовуються три такі види відбору:

l) індивідуальний відбір одиниць сукупності;

2) груповий відбір групи одиниць;

3) комбінований – комбінація першого і другого видів.

Різні види відбору можуть здійснюватися різними способами. Спосіб відбору визначає конкретний механізм вибірки одиниць з генеральної сукупності. У практиці вибіркових обстежень найбільшого поширення набули такі вибірки:

1) власне-випадкова;

2) механічна;

3) типова;

4) серійна;

5) комбінована.

При власне-випадковій вибірці одиниці з генеральної сукупності вибираються навмання. У цьому випадку необхідно переконатися, що всі одиниці генеральної сукупності мають рівні шанси потрапляння у вибірку. Технічно власне-випадковий відбір проводять методом жеребкування або за таблицею випадкових чисел .

Прикладом власне-випадкового відбору можуть бути: тиражі національної лотереї, виграші грошово-речової лотереї або державних позик, коли із загальної кількості випущених білетів у випадковому порядку навмання відбирається частина номерів, на які випадає виграш. Водночас усім номерам надається рівна можливість потрапити у вибірку.

Власне-випадковий відбір може бути повторним і безповторним.

Припустимо, що при відборі 100 із 1000 робітників можна організувати жеребкування або зробити вибір навмання. Поставлено завдання установити вибірковим шляхом відсоток робітників, які мають вищу освіту. Із загального стосу анкет беруть навмання одну, відмічають на окремому аркуші факт наявності або відсутності вищої освіти у робітника і відкладають убік. Потім беруть другу анкету і так 100 анкет.

Вибіркове спостереження, організоване таким чином, називається випадковою безповторною вибіркою. Випадковою, тому що при діставанні анкет із загальної маси їх вибирають навмання. Безповторний – тому що витягнуті анкети не повертаються в загальний стіс.

Але вибірка може бути випадковою повторною тоді, коли вибрана анкета знову повертається в загальний стіс анкет, тому виникає можливість, що в одному з наступних діставань та сама анкета трапиться ще раз, тобто повторно.

Безповторний відбір називається схемою неповерненої кулі Безповторний відбір дає більш точні результати порівняно з повторним, бо при тому самому обсязі вибірки спостереження охоплює більшу кількість одиниць сукупності, що вивчається. Тому він набуває великого поширення у статистичній практиці. І тільки тоді, коли не можна провести безповторний відбір (при обстеженні пасажирів міського транспорту), застосовується повторна вибірка. Обидві схеми відбору можуть застосовуватися у поєднанні з різними видами відбору, за винятком механічного.

Механічний відбір полягає в тому, що вся сукупність розбивається на рівні за обсягом групи за випадковою, а не типовою ознакою. Потім з кожної групи береться одна одиниця. Усі одиниці сукупності, що вивчається, попередньо розміщуються у певному порядку, наприклад за алфавітом, а залежно від обсягу вибірки (наприклад, із 1000 студентів 100) механічно, через певний інтервал (1000/100=10) потрібно відібрати необхідну кількість одиниць. Якщо почати відбір, наприклад, з номера 5 за списком, то у вибірку потраплять 100 студентів, які стоять в алфавітному списку під номерами 5, 15, 25, 35 і Т.д. до 995.

Механічний відбір – це різновид власне-випадкового відбору, але він має ряд організаційних переваг, зокрема ним легше і простіше організувати перевірку правильності відбору одиниць сукупності. Механічний відбір завжди буває тільки безповторним.

Типовий відбір передбачає розподіл усієї сукупності, що вивчається, за типовою ознакою на якісно однорідні групи. Потім з кожної групи власне-випадковим або механічним способом відбирається кількість одиниць, пропорційна питомій вазі групи у всій сукупності.

Припустимо, необхідно провести типовий відбір студентів, які навчаються на чотирьох факультетах інституту. Спершу потрібно розгрупувати їх на однорідні групи по факультетах, а потім по кожному з них відібрати кількість студентів, пропорційну питомій вазі всієї кількості студентів.

Припустимо, необхідно провести типовий відбір 1000 студентів із 10000, які навчаються на чотирьох факультетах інституту. Згрупуємо їх в однорідні групи по факультетах, а потім по кожному з них відберемо кількість студентів, пропорційну питомій вазі контингенту студентів інституту по факультетах.

Типовий відбір дає більш точні результати, ніж власне-випадковий або механічний, тому що при ньому у вибірку у такій же пропорції, як і в генеральній сукупності, потрапляють представники всіх типових груп. Тому вибірка стає більш репрезентативною представницькою і, отже, більш точною (таблиця 5.3).

Таблиця 5.3 – Гранична помилка вибірки при типовому відборі

Метод відбору	Гранична помилка	
	для середньої	для частки
Повторний	$\Delta_x = t\sqrt{\frac{\overline{\sigma_i^2}}{n}}$	$\Delta_w = t\sqrt{\frac{\overline{w_i(1-w_i)}}{n}}$
Безповторний	$\Delta_x = t\sqrt{\left(1-\frac{n}{N}\right)\frac{\overline{\sigma_i^2}}{n}}$	$\Delta_w = t\sqrt{\left(1-\frac{n}{N}\right)\frac{\overline{w_i(1-w_i)}}{n}}$

З цих формул видно, що при типовому відборі помилка вибірки залежить тільки від внутрішньогрупових дисперсій, а не від загальної дисперсії як при власне-випадковому або механічному відборі. За правилом складання дисперсій між загальною, між груповою і між внутрішьогруповою дисперсіями є такий зв'язок:

$$\boldsymbol{\sigma_0^2} = \boldsymbol{\delta^2} + \boldsymbol{\bar{\sigma}^2}, \text{ звідки } \boldsymbol{\bar{\sigma}^2} < \boldsymbol{\sigma_0^2} \ . \qquad (5.12)$$

Отже, помилка вибірки при типовому відборі завжди буде меншою від помилки вибірки, проведеної випадковим і механічним відбором.

Серійний відбір. Цей спосіб відбору ручний тоді, коли одиниці сукупності об'єднані в невеликі групи або серії. Як такі серії можуть розглядатися окремі райони, області, бригади, студентські групи та інші об'єднання. Серії, гнізда відбираються власне-випадковим або механічним способом. Усередині відібраних серій проводиться суцільне спостереження.

Точність серійної вибірки залежить не від величини загальної дисперсії, а від міжсерійної (міжгрупової) дисперсії або, інакше, дисперсії групових середніх.

Серійна вибірка може проводитися як повторний і безповторний відбір. Крім того, серії можуть бути рівновеликими і нерівновеликими (таблиця 5.4).

Таблиця 5.4 – Гранична помилка вибірки при серійному відборі

Метод відбору	Гранична помилка	
	для середньої	для частки
Повторний	$\boldsymbol{\Delta_x = t\sqrt{\frac{\sigma^2}{r}}}$	$\boldsymbol{\Delta_w = t\sqrt{\frac{w(1-w)_m}{r}}}$
Безповторний	$\boldsymbol{\Delta_x = t\sqrt{\left(1-\frac{r}{R}\right)\frac{\sigma^2}{r}}}$	$\boldsymbol{\Delta_w = t\sqrt{\left(1-\frac{r}{R}\right)\frac{w(1-w)_m}{r}}}$

Як правило, міжгрупова (міжсерійна) дисперсія обчислюється за формулою

$$\boldsymbol{\delta^2 = \frac{\sum(\tilde{x}_i - \tilde{x})^2}{r}}, \qquad (5.13)$$

де $\mathbf{r}$ – кількість відібраних серій;

$\mathbf{R}$ – загальна кількість серій у генеральній сукупності;

$\tilde{\mathbf{x}}_{\mathbf{i}}$ – середня в окремих серіях;

$\tilde{\mathbf{x}}$ – загальна середня для всієї сукупності.

Комбінований відбір. У практиці статистичних обстежень, крім розглянутих вище способів відбору, застосовується і їх комбінування. Наприклад, можна комбінувати типову і серійну вибірку, коли серії відбираються в установленому порядку з декількох типових груп. Можна комбінувати серійний і власне-випадковий відбір, коли окремі одиниці відбираються всередині серії у власне-випадковому порядку.

Така комбінована вибірка може бути повторною і безповторною. Квадрат середньої помилки комбінованого відбору визначається за формулами:

– при повторному відборі:

$$\boldsymbol{\mu}^2 = \frac{\boldsymbol{\sigma}^2}{\mathbf{n}} + \frac{\boldsymbol{\delta}^2}{\mathbf{r}}; \tag{5.14}$$

– при безповторному відборі:

$$\boldsymbol{\mu}^2 = \frac{\boldsymbol{\sigma}^2}{\mathbf{n}}\left(\mathbf{1} - \frac{\mathbf{n}}{\mathbf{N}}\right) + \frac{\boldsymbol{\delta}^2}{\mathbf{r}}\left(\mathbf{1} - \frac{\mathbf{r}}{\mathbf{R}}\right). \tag{5.15}$$

5.5 Визначення необхідного обсягу вибірки

При проектуванні вибіркового спостереження виникає питання про необхідну чисельність вибірки. Вона може бути визначена з огляду на допустиму помилку при вибірковому спостереженні ймовірності, з якою потрібно гарантувати величину встановлюваної помилки, і спосіб відбору.

Необхідна чисельність вибірки визначається по-різному для вибіркового спостереження, у якому встановлюється середній розмір ознаки, і для спостереження, у якому встановлюється частка одиниць, які мають альтернативні ознаки.

Необхідна чисельність вибірки **n** визначається на основі формул граничної помилки вибірки. Так, якщо вибірка повторна, то чисельність вибірки при власне-випадковому і механічному відборі визначається за формулою

$$\Delta_x = t\sqrt{\frac{\sigma^2}{r}}.$$

Для того, щоб знайти n, піднесемо обидві частини рівняння до квадрата й одержимо $\Delta^2{}_x = \frac{t^2\sigma^2}{n}$, звідки $\Delta^2{}_x n = t^2\sigma$, а

$$n = \frac{t^2\sigma^2}{\Delta^2{}_x}. \qquad (5.16)$$

При безповторному відборі чисельність вибірки визначається за формулою

$$\Delta_x = \sqrt[t]{\left(1 - \frac{n}{N}\right)\frac{\overline{\sigma_i^2}}{n}}. \qquad (5.17)$$

Підносячи обидві частини до квадрата, одержимо

$$\Delta^2{}_x = \frac{t^2\sigma^2}{n} * \frac{N - n}{N}.$$

Після перетворення маємо:

$$\Delta^2{}_x \ nN + t^2\sigma^2 n = t^2\sigma^2 N;$$
$$n(\Delta^2{}_x N + t^2\sigma^2) = t^2\sigma^2 N.$$

Звідси

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

Приклад. На ткацькій фабриці працює 6 тис. ткаль. Для встановлення норм виробітку передбачається провести випадковий відбір ткаль. За попереднім обстеженням установлено, що середнє квадратичне відхилення денного виробітку становить 25 м. Визначте необхідну чисельність вибірки за умови, що з імовірністю 0,954 помилка вибірки не перевищить 5 м.

Обчислимо необхідну чисельність вибірки:

для повторного відбору

$$\mathbf{n = \frac{t^2\sigma^2}{\Delta^2_x},\ n = \frac{2^2 \cdot 25^2}{5^2} = 100\ ткаль;}$$

для безповторного відбору:

$$\mathbf{n = \frac{t^2\sigma^2 N}{\Delta^2_x N + t^2\sigma^2},\ n = \frac{2^2 \cdot 25^2 \cdot 6000}{5^2 \cdot 6000 + 2^2 \cdot 25^2} = 98\ ткаль}.$$

Як видно з прикладу, обсяг вибірки при безповторному відборі менший ніж при повторному за інших рівних умов.

Визначаючи необхідну чисельність вибірки, потрібно знати ступінь коливальності ознак, що вивчається, або на основі попереднього досвіду, або на основі припущення. Якщо ступінь коливальності ознаки невідомий, то його можна знайти наближено до величини припущеного розмаху варіації, бо в нормальному ряду розподілу (коли $\mathbf{\bar{x} = M_0 = M_e}$) між $\mathbf{\bar{x}}$**, d, σ** і **R** є певні співвідношення, які дають змогу судити про близькість розподілу, що вивчається, до нормального, а також обчислити за показниками варіації, що вже є, невідомі. Середнє квадратичне відхилення у цьому разі знаходимо за формулою $\mathbf{\sigma = \frac{R}{6}}$, бо з імовірністю 0,997 можна стверджувати, що розмах варіації укладається при нормальному розподілі ознаки **6σ** (крайні значення віддалені в той або інший бік від середньої на відстань **3σ**), тобто $\mathbf{R = 6\sigma}$, де $\mathbf{R = x_{мах} - x_{мін}}$. При достатньо великому обсязі сукупності між середнім квадратичним і середнім лінійним відхиленням є співвідношення

$$\mathbf{\sigma \approx 1{,}25d.}$$

Важливою умовою практичного використання цих формул є близькість фактичного розподілу до нормального. Обчислення середнього квадратичного відхилення для явно несиметричних розподілів не має сенсу.

При проведенні соціально-економічних досліджень, як правило, можна з достатньою точністю вказати максимально і мінімально можливі значення ознаки в досліджуваній сукупності.

Питання до самоконтролю

5.1 Поняття про вибіркове спостереження та його значення.
5.2 Характеристика вибіркової і генеральної сукупності.
5.3 Помилки вибірки.
5.4 Методи і способи відбору.
5.5 Власне-випадковий відбір та його характеристика.
5.6 Безповторний відбір та його характеристика.
5.7 Механічний відбір.
5.8 Типовий відбір.
5.9 Визначення необхідного обсягу вибірки.

СПИСОК ЛІТЕРАТУРИ

1 Акімов О. В. Статистика в малюнках та схемах : навч. посіб. Київ : ЦНЛ, 2007. 168 с.

2 Бек В. Л. Теорія статистики: курс лекцій : навч. посіб. Київ : ЦНЛ, 2003. 412 с.

3 Горкавий В. К. Статистика : навч. посіб. Київ : ЦНЛ, 2012. 608 с.

4 Єріна А. М., Пальян З. О. Теорія статистики: практикум. Київ : Знання, 2002. 422 с.

5 Лугінін О. Є. Статистика : навч. посіб. Київ : ЦНЛ, 2007. 608 с.

6 Матковський С. О., Марець О. Р. Теорія статистики : навч. посіб. Київ : Знання, 2010. 535 с.

7 Опря А. Т. Статистика. Математична статистика. Теорія статистики : навч. посіб. Київ : ЦНЛ, 2005. 496 с.

8 Опря А. Т. Статистика : навч. посіб. Київ : ЦНЛ, 2012. 448 с.

9 Статистика : підручник / Р. Я. Баран та ін. Чернівці : Наші книги, 2008. 240 с.

10 Статистика : підручник / С. С. Герасименко та ін. Київ : КНЕУ, 2000. 467 с.

11 Статистика: теоретичні засади і прикладні аспекти / за ред. Р. В. Фещура. Львів : Інтелект-Захід, 2003. 346 с.

12 Тарасенко І. О. Статистика : навч. посіб. Київ : ЦНЛ, 2006. 344 с.

13 Теорія статистики : підручник / Є. І. Ткач, В. П. Сторожук та ін. Тернопіль : Астон, 2004. 589 с.

14 Тринько Р. І., Тадник М. Є. Основи теоретичної і прикладної статистики : навч. посіб. Київ : Знання, 2011. 400 с.

15 Уманець Т. В. Загальна теорія статистики : навч. посіб. Київ : Знання, 2006. 294 с.

16 Штагрет А. М. Статистика : навч. посіб. Київ : ЦНЛ, 2005. 232 с.

Н. М. Лисьонкова, О. А. Єрмоленко

СТАТИСТИКА

Конспект лекцій

Частина 2

Відповідальний за випуск Лисьонкова Н. М.

Редактор Еткало О. О.

Підписано до друку 22.04.19 р.

Формат паперу 60х84 1/16. Папір писальний.

Умовн.-друк.арк. 4,5. Тираж 50. Замовлення №

Видавець та виготовлювач Український державний університет
залізничного транспорту,
61050, Харків-50, майдан Фейєрбаха, 7.
Свідоцтво суб'єкта видавничої справи ДК № 6100 від 21.03.2018 р.