Державний вищий навчальний заклад
“Українська академія банківської справи
Національного банку України”
Кафедра бухгалтерського обліку і аудиту

О.В. Козьменко, О.В. Меренкова

СТАТИСТИКА: БАНКІВСЬКИЙ ДОСВІД

Навчальний посібник

У 2 частинах

Частина 2

Для студентів економічних спеціальностей
вищих навчальних закладів

Суми
ДВНЗ “УАБС НБУ”
2009

УДК 311:336.71](075.8)
К59

Рекомендовано до друку вченою радою Державного вищого навчального закладу “Українська академія банківської справи Національного банку України”, протокол № 6 від 13.02.2009.

Рецензенти:
доктор економічних наук, професор
І.О. Школьник;

кандидат фізико-математичних наук, доцент кафедри вищої математики
Сумського державного університету
Л.В. Волкова

Відповідальний за випуск
доктор економічних наук, професор
Т.А. Васильєва

Козьменко, О. В.

К59 Статистика: банківський досвід [Текст] : навчальний посібник : у 2 ч. / О. В. Козьменко, О. В. Меренкова ; Державний вищий навчальний заклад “Українська академія банківської справи Національного банку України”. – Суми : ДВНЗ “УАБС НБУ”, 2009. – Ч. 2. – 203 с.

Навчальний посібник може використовуватися при навчанні студентів збору інформації, обробці отриманих даних, з’ясуванню закономірностей різних показників, для прогнозу досліджуваних показників і побудови найпростіших моделей. Він може бути використаний для самостійного навчання студентів та індивідуальної роботи і містить завдання та тести для самоконтролю з метою забезпечення більш ефективного опрацювання навчального матеріалу в процесі самостійної роботи.

Призначений для студентів економічних спеціальностей денної форми навчання, аспірантів економічних спеціальностей, викладачів, співробітників банків і науковців.

УДК 311:336.71](075.8)

ЗМІСТ

Тема 6. РОЗВИТОК СТАТИСТИКИ ЯК НАУКИ. ВІДОКРЕМЛЕННЯ ЕКОНОМЕТРІЇ. ЕКОНОМЕТРИЧНІ МЕТОДИ ТА МОДЕЛІ 4

1. Вибіркове спостереження як джерело статистичної інформації 6

2. Статистичне вивчення взаємозв'язку економічних показників. Економетричні моделі та методи 19

Тема 7. ДОСЛІДЖЕННЯ ІНТЕНСИВНОСТІ ДИНАМІКИ ТА ТЕНДЕНЦІЙ РОЗВИТКУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ЯВИЩ 138

1. Поняття та види динамічних рядів 140

2. Характеристики інтенсивності динаміки 143

3. Аналіз тенденцій розвитку 148

Тема 8. ЕКОНОМІЧНІ ІНДЕКСИ 161

1. Поняття та класифікація індексів 163

2. Особливості та методологія побудови індивідуальних і зведених індексів 165

3. Сутність середніх індексів 169

4. Індекси просторово-територіального співвідношення 171

5. Особливості індексів Ласпейреса, Пааше та Фішера 172

Додаткова інформація до теми 172

ПРЕДМЕТНИЙ ПОКАЖЧИК 182

СПИСОК ЛІТЕРАТУРИ 186

Тема 6. РОЗВИТОК СТАТИСТИКИ ЯК НАУКИ. ВІДОКРЕМЛЕННЯ ЕКОНОМЕТРІЇ. ЕКОНОМЕТРИЧНІ МЕТОДИ ТА МОДЕЛІ

1. ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ ЯК ДЖЕРЕЛО СТАТИСТИЧНОЇ ІНФОРМАЦІЇ.

   Способи формування та визначення достатнього обсягу вибіркової сукупності. Оцінка точності вибіркових даних. Статистична перевірка гіпотез.

2. СТАТИСТИЧНЕ ВИВЧЕННЯ ВЗАЄМОЗВ'ЯЗКУ ЕКОНОМІЧНИХ ПОКАЗНИКІВ. ЕКОНОМЕТРИЧНІ МОДЕЛІ ТА МЕТОДИ.

   Теоретичні основи кореляційно-регресійного аналізу.
   Функціональні та стохастичні зв'язки, парна регресія, умови Гаусса-Маркова відносно випадкового члена, метод найменших квадратів (МНК), оцінка точності точкових оцінок, аналіз залишків, коефіцієнт кореляції, t-статистика. Прогнозування. Лінійні та нелінійні залежності.

   Багатофакторна регресія.
   Частинні та парні коефіцієнти кореляції, коефіцієнт детермінації, статистика Дарбіна-Уотсона, дослідження на гетероскедастичність.

   Лінійна та нелінійна моделі.
   Функція Кобба-Дугласа. Поняття еластичності.

   Непараметричні методи взаємозв'язків.

   Теорія ігор.
   Метод Монте-Карло. Договірні простори Неша, ящик Еджворта, теорія ув'язненого.

Терміни

Вибіркова сукупність
Генеральна сукупність

Індивідуальний відбір
Власне випадковий
Механічний
Стратифікований
Серійний (гніздовий)

Помилка вибірки
Помилка реєстрації
Помилка репрезентативності

Парна регресія
Множинна регресія
Метод найменших квадратів (МНК)

Коефіцієнт кореляції
Коефіцієнт детермінації
t-статистика
Статистика Дарбіна-Уотсона
Тест рангової кореляції Спірмена
Тест Голдфелда-Квандта
Тест Глейзера

Фіктивні змінні
Модель з лагом

Гомоскедастичність
Гетероскедастичність
F-критерій Фішера
Статистика Дарбіна-Уотсона

Частинні коефіцієнти кореляції
Парні коефіцієнти кореляції

Метод Монте-Карло
Договірні простори Неша
Ящик Еджворта
Теорія ув’язненого

Функція Кобба-Дугласа
Еластичність

Перелік всіх термінів, наведених у даній темі,
див. у “Предметному покажчику”.

1. ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ ЯК ДЖЕРЕЛО СТАТИСТИЧНОЇ ІНФОРМАЦІЇ

Статистична методологія дослідження масових явищ розрізняє, як відомо, два види спостереження залежно від повноти охоплення об'єкта: суцільне і несуцільне. Різновидом несуцільного спостереження є вибіркове, що в умовах розвитку ринкових відносин знаходить більш широке застосування. *Вибірковий метод* застосовується в тих випадках, коли проведення суцільного спостереження неможливе або економічно недоцільне. Вибіркове спостереження використовують також для перевірки результатів суцільного спостереження.

Вибіркове спостереження	несуцільне спостереження, при якому статистичному обстеженню (спостереженню) піддаються одиниці досліджуваної сукупності, відібрані випадковим способом;
вибіркова сукупність	та частина одиниць, що відібрані для спостереження;
генеральна сукупність	вся сукупність одиниць, з яких робиться відбір.

Генеральна і вибіркова сукупності

Нехай вимагається вивчити сукупність однорідних об'єктів щодо деякої якісної або кількісної ознаки, що характеризує ці об'єкти. Наприклад, якщо є група банків, то якісною ознакою може служити надійність банку, а кількісною – обсяг прибутку.

Іноді проводять суцільне обстеження, тобто обстежують кожний з об'єктів сукупності щодо ознаки, якою цікавляться. На практиці, проте, суцільне обстеження застосовують порівняно рідко. Наприклад, якщо сукупність містить дуже велику кількість об'єктів, то провести суцільне обстеження фізично неможливо. Якщо обстеження об'єкта пов'язане з його знищенням або вимагає великих матеріальних витрат, то проводити суцільне обстеження практично немає сенсу. У таких випадках випадково відбирають зі всієї сукупності обмежену кількість об'єктів і їх вивчають.

Вибірковою сукупністю, або просто вибіркою називають сукупність випадково відібраних об'єктів. *Генеральною сукупністю* називають сукупність об'єктів, з яких здійснюється вибірка.

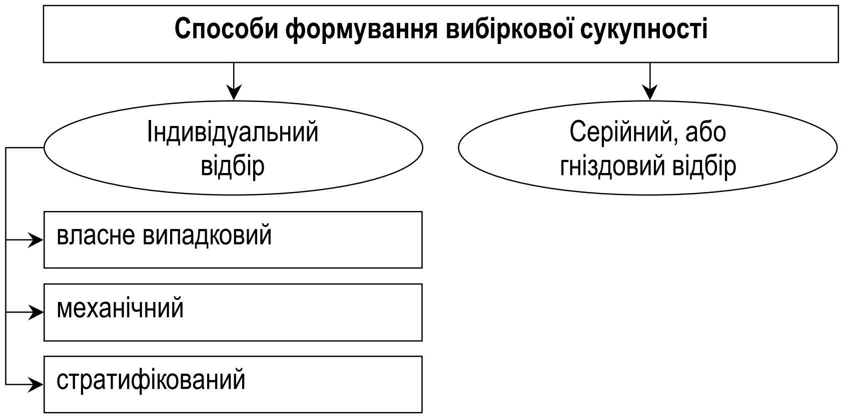
Обсяг сукупності (вибіркової або генеральної)	кількість об'єктів цієї сукупності. Наприклад, якщо із 173 банків відібрано для обстеження 10, то обсяг генеральної сукупності $N=173,$ а обсяг вибірки $n=10.$

Зауваження. Часто генеральна сукупність містить кінцеву кількість об’єктів. Проте якщо це число достатньо велике, то іноді з метою спрощення обчислень або для полегшення теоретичних висновків допускають, що генеральна сукупність складається з безлічі об’єктів. Таке допущення виправдовується тим, що збільшення обсягу генеральної сукупності (досить великого обсягу) практично не позначається на результатах обробки даних вибірки.

▷ **ПРИКЛАД** вибіркової та генеральної сукупностей.

Генеральна сукупність – всі банки України.
Вибіркова сукупність:

- декілька крупних банків з усіх, які належать до крупних банків;
- банки, розташовані на певній території. ■



Повторна і безповторна вибірки. Репрезентативна вибірка

При формуванні вибірки можна діяти двома способами: після того, як об'єкт відібраний і над ним здійснене спостереження, він може бути повернений або не повернений в генеральну сукупність. Відповідно до вищевикладеного вибірки підрозділяють на повторні та безповторні.

Повторна вибірка	вибірка, при якій відібраний об'єкт (перед відбором наступного) повертається в генеральну сукупність.
Безповторна вибірка	вибірка, при якій відібраний об'єкт в генеральну сукупність не повертається.

На практиці звичайно користуються безповторним випадковим відбором.

Для того, щоб за даними вибірки можна було досить впевнено скласти думку про ознаку генеральної сукупності, що цікавить, необхідно, щоб об'єкти вибірки правильно її представляли. Інакше кажучи, вибірка повинна правильно представляти пропорції генеральної сукупності. Цю вимогу коротко формулюють так: вибірка повинна бути репрезентативною.

Через закон великих чисел можна стверджувати, що вибірка буде репрезентативною, якщо її здійснити випадково: кожен об'єкт вибірки відібраний випадково з генеральної сукупності, якщо всі об'єкти мають однакову імовірність потрапити у вибірку.

Якщо обсяг генеральної сукупності достатньо великий, а вибірка становить лише незначну частину цієї сукупності, то відмінність між повторною і безповторною вибірками зникає; у граничному випадку, коли розглядається нескінченна генеральна сукупність, а вибірка має кінцевий обсяг, ця відмінність зникає.

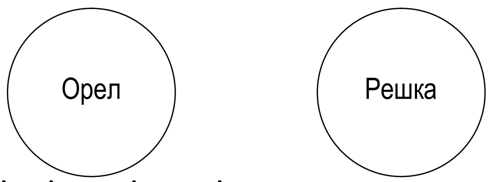
Статистичний розподіл вибірки

Нехай з генеральної сукупності відібрана вибірка, причому x_1 спостерігалося $n_1, n_2, ..., n_k$ разів, $x_2 - n_{k+1}, n_{k+2}, ..., n_{k+j}$ разів, $x_k - ... n_{k+i}$ разів і $\sum n_i = n$ – обсяг вибірки.

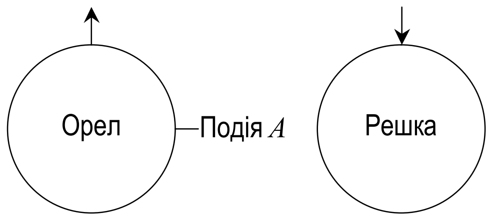
Варіанти	значення, що спостерігаються, x_i.
Варіаційний ряд	послідовність варіант, записаних у зростаючому порядку.
Частоти	кількість спостережень.
Відносні частоти	відношення частот до обсягу вибірки $n_i / n = W_i$.

▷ **ПРИКЛАД** визначення у серії іспитів частоти події A.

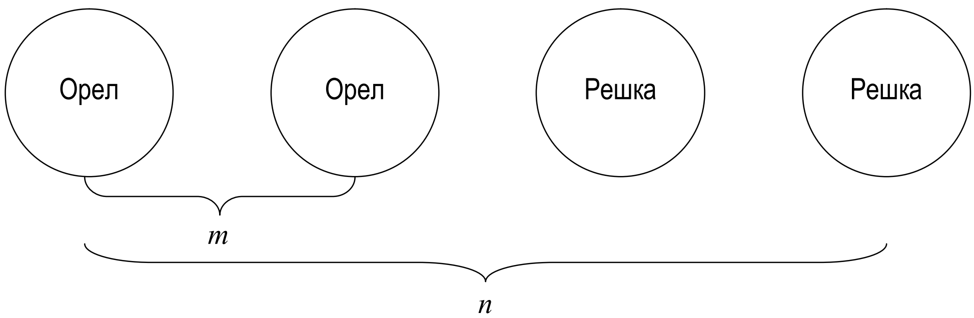
Монета має два боки – орел і решку.



Проводимо серію іспитів – підкидаємо монету n разів.



Розглянемо подію A – випав орел. Нехай A відбулося m разів.



Тоді $W(A)=\dfrac{m}{n}$. ■

Статистичний розподіл вибірки — перелік варіант і відповідних їм частот або відносних частот. Статистичний розподіл можна задати також у вигляді послідовності інтервалів і відповідних їм частот (як частоту, відповідну інтервалу, беруть суму частот, що потрапили в цей інтервал).

Помітимо, що в теорії імовірності під *розподілом* мають на увазі відповідність між можливими значеннями випадкової величини та їх імовірністю, а в математичній статистиці – відповідність між варіантами, які спостерігаються, і їх частотами або відносними частотами.

Визначення функції розподілу — Нехай x – дійсне число. Імовірність події, яка полягає в тому, що X набуде значення менше x, тобто імовірність події $X < x$, позначимо через $F(x)$. Якщо x змінюється, то змінюється і $F(x)$, тобто $F(x)$ – функція від x.

Функцією розподілу називають функцію $F(x)$, яка визначає імовірність того, що випадкова величина X у результаті випробування набуде значення менше x, тобто

$$F(x)=P(X<x).$$

Властивості функції розподілу

Властивість 1. Значення функції розподілу належить відрізку $[0,1]$:

$$0 \le F(x) \le 1.$$

Властивість 2. $F(x)$ – неспадаюча функція, тобто

$$F(x_2) \ge F(x_1), \text{ якщо } x_2 > x_1.$$

Властивість 3. Якщо можливі значення випадкової величини належать інтервалу $(a,\ b)$, то:

1) $F(x)=0$ при $x \le a$; 2) $F(x)=1$ при $x \ge b$.

Емпірична функція розподілу

Нехай відомий статистичний розподіл частот кількісної ознаки X. Введемо позначення: n_x – кількість спостережень, при яких спостерігалося значення ознаки, менше x; n – загальна кількість спостережень (обсяг вибірки). Ясно, що відносна частота події $X < x$ дорівнює n_x/n. Якщо x змінюється, то змінюється і відносна частота, тобто відносна частота n_x/n є функцією від x. Оскільки ця функція знаходиться емпіричним шляхом, то її називають емпіричною.

Емпірична функція розподілу

функцією розподілу вибірки називають функцію $F\cdot(x)$, що визначає для кожного значення x відносну частоту події $X < x$.

Отже, за визначенням

$$F\cdot(x) = n_x/n, \qquad (6.1)$$

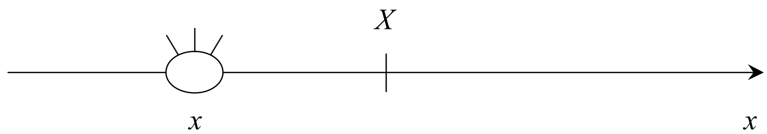
де n_x – кількість варіант, менших x;
n – обсяг вибірки.

Теоретична функція розподілу

на відміну від емпіричної функції розподілу вибірки, функція розподілу $F(x)$ генеральної сукупності.

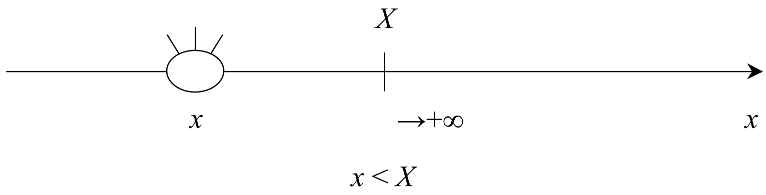
▷ **ПРИКЛАД** функції розподілу.

Кидаємо сніжок (x) на пряму лінію (X).

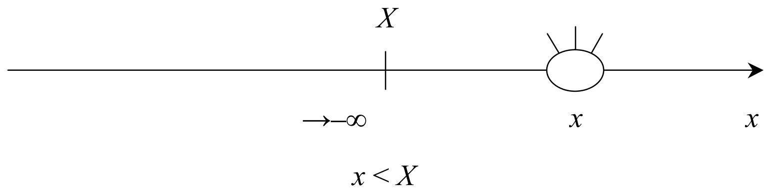


$F(x) = P(x \le X)$. Функція $F(x)$ – імовірність того, що x набуває значення менше X.

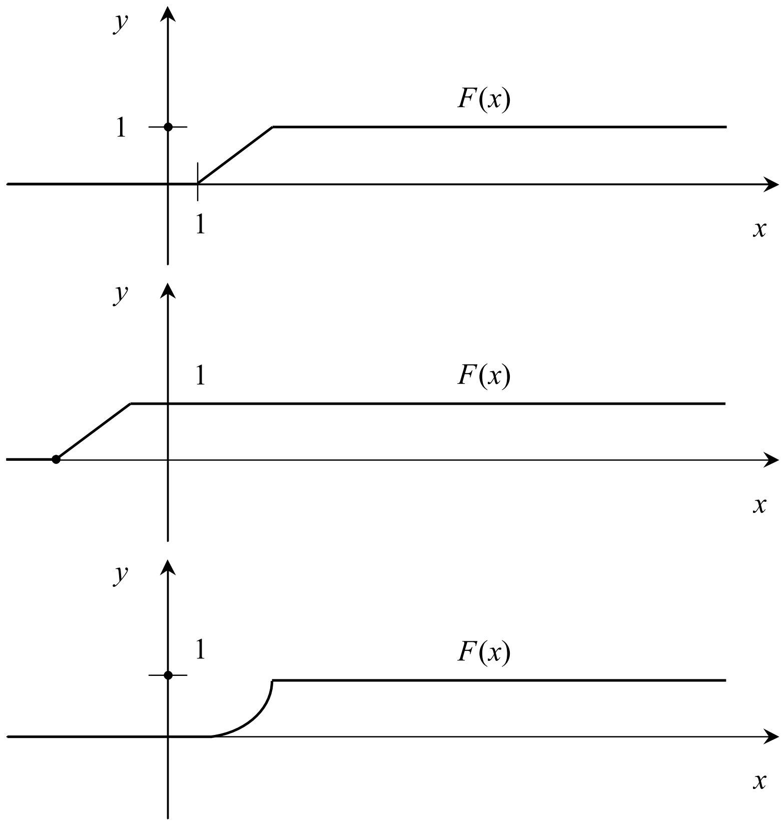
Якщо $X \to -\infty$, то $F(x) = 0.$



Якщо $X \to +\infty$, то $F(x) = 1.$



Приклади $F(x)$



■

Відмінність між емпіричною і теоретичною функціями полягає у тому, що теоретична функція $F(x)$ визначає імовірність події $X < x,$ а емпірична функція $F \cdot (x)$ визначає відносну частоту цієї ж події. З теореми Бернуллі випливає, що відносна частота події $X < x,$

тобто $F\cdot(x)$ *прагне* по імовірності до імовірності $F(x)$ *цієї* події. Іншими словами, при великих *n числа* $F\cdot(x)$ *і* $F(x)$ *мало* відрізняються одне від іншого в тому сенсі, що $\lim\limits_{n\to\infty} P\left[\left|F(x)-F\cdot(x)\right|<\varepsilon\right] = 1\ (\varepsilon>0)$. Вже звідси випливає доцільність використання емпіричної функції розподілу вибірки для наближеного представлення теоретичної (інтегральної) функції розподілу генеральної сукупності.

Такий висновок підтверджується і тим, що $F\cdot(x)$ володіє всіма властивостями $F(x)$. *Дійсно, з визначення функції* $F\cdot(x)$ випливають такі її властивості:

1) значення емпіричної функції належать відрізку [0, 1];
2) $F\cdot(x)$ – неспадаюча функція;
3) якщо x_1 – найменша варіанта, то $F\cdot(x)=0$ при $x\le x_1$, якщо x_k – *найбільша* варіанта, то $F\cdot(x)=1$ при $x>x_k$.

Отже, емпірична функція розподілу вибірки служить для оцінки теоретичної функції розподілу генеральної сукупності.

Визначення щільності розподілу

Щільністю розподілу імовірностей неперервної випадкової величини X називають функцію $f(x)$ – першу похідну від функції розподілу $F(x)$:

$$f(x)=F'(x).$$

З цього визначення випливає, що функція розподілу є первісною для щільності розподілу.

Властивості щільності розподілу

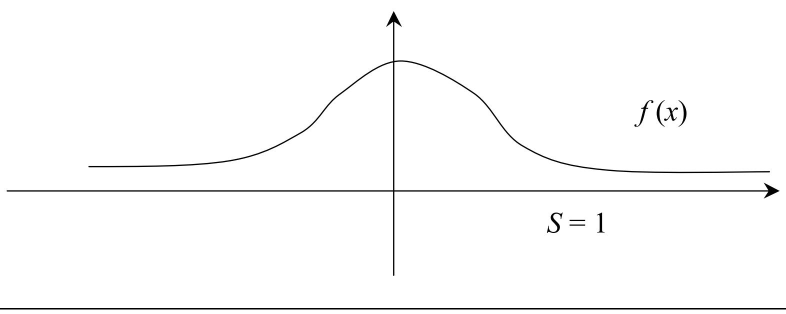
Властивість 1. Щільність розподілу – невід’ємна функція:

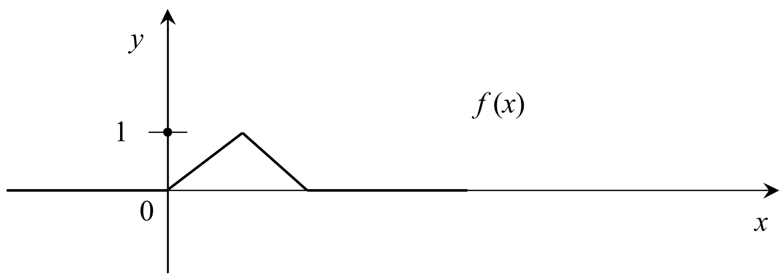
$$f(x)\ge 0.$$

Властивість 2. Невласний інтеграл від щільності розподілу в межах від $-\infty$ до ∞ дорівнює одиниці:

$$\int_{-\infty}^{\infty} f(x)\,dx=1.$$

▷ **ПРИКЛАД** $F(x)$.





■

Власне випадковий відбір	здійснюється за допомогою жеребкування або за таблицею випадкових чисел. У першому випадку всім елементам генеральної сукупності привласнюється порядковий номер. У другому випадку виробляється вибір випадкових чисел (зі спеціальних таблиць), що утворять порядкові номери для відбору. Процес формування випадкових чисел і визначення номера одиниці, що відбирається, продовжується доти, поки не буде отриманий заданий обсяг вибіркової сукупності.
Механічний відбір	відбирається кожний (n/N)- й елемент генеральної сукупності.
Стратифікований відбір	здійснюється відбір одиниць з неоднорідної сукупності. У цьому випадку генеральну сукупність попередньо розбивають на однорідні групи за допомогою типологічного угруповання, після чого роблять відбір одиниць з кожної групи у вибіркову сукупність випадковим або механічним способом.
Серійний, або гніздовий відбір	відбір, при якому в порядку випадкової або механічної вибірки вибирають не одиниці, а визначені райони, серії (гнізда), усередині яких виробляється суцільне спостереження.

▷ **ПРИКЛАД** способів відбору.

Власне випадковий – відбір банків з генеральної сукупності у вибіркову випадковим чином.

Механічний – відбір кожного 20-го банку з генеральної сукупності у вибіркову.

Стратифікований – банки генеральної сукупності розбиваються на 4 групи за обсягами активів, з кожної групи випадковим чином обираються банки у вибіркову сукупність.

Серійний – банки генеральної сукупності розбиваються на 10 груп за кількістю відділень, випадковим чином обираються 3 групи, де проводиться суцільне обстеження. ■

Оцінка точності вибіркових даних. Статистична перевірка гіпотез

Помилка вибірки	різниця між показниками вибіркової та генеральної сукупностей;
помилки реєстрації	виникають через неправильні або неточні зведення;
помилки репрезентативності	виникають через неправильний, тенденційний добір одиниць.

Таблиця 6.1

Умовні позначки

Показник	Генеральна сукупність	Вибіркова сукупність
Обсяг сукупності	N	n
Кількість елементів, що мають досліджувану ознаку	M	m
Частка одиниць, що мають досліджувану ознаку	$p = \frac{M}{N}$	$\omega = \frac{m}{n}$
Середній розмір ознаки	$\bar{x} = \frac{\sum x_i}{N}$	$\tilde{x} = \frac{\sum x_i}{n}$
Дисперсія кількісної ознаки	$\sigma_{\bar{x}}^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{N}$	$\sigma_{\tilde{x}}^2 = \frac{\sum\left(x_i - \tilde{x}\right)^2}{n}$
Дисперсія частки	$\sigma_p^2 = p\left(1-p\right)$	$\sigma_\omega^2 = \omega\left(1-\omega\right)$
Гранична помилка вибірки: • для частки $\left(\Delta\omega\right)$; • для середнього розміру ознаки $\left(\Delta\tilde{x}\right)$	$\Delta\omega = \left\|p-\omega\right\| = t\mu$ $\Delta\tilde{x} = \left\|\bar{x}-\tilde{x}\right\| = t\mu$ де t – довірчий коефіцієнт; μ – середня помилка вибірки	

Середня помилка вибірки дорівнює квадратичному відхиленню, діленому на квадратний корінь з чисельності вибірки:

$$\mu_x = \frac{\delta_x}{\sqrt{n}} = \sqrt{\frac{\delta_x{}^2}{n}} \text{ – для середньої;} \tag{6.2}$$

$$\mu_ч = \sqrt{\frac{\rho\left(1-\rho\right)}{n}} \text{ – для частки.} \tag{6.3}$$

Граничні помилки вибірки визначаються за формулами, поданими в табл. 6.2-6.3.

Таблиця 6.2

Формули для розрахунку граничної помилки вибірки

Способи відбору	Повторний	
	для середньої	для частки
Власне випадкова, механічна	$\Delta_x = t\sqrt{\frac{\sigma^2}{n}}$	$\Delta_\omega = t\sqrt{\frac{\omega(1-\omega)}{n}}$
Стратифікована (при пропорційному обсязі груп у вибірці)	$\Delta_x = t\sqrt{\frac{\overline{\sigma_i}^2}{n}}$	$\Delta_\omega = t\sqrt{\frac{\overline{\omega_i(1-\omega_i)}}{n}}$
Серійна	$\Delta_x = t\sqrt{\frac{\delta_x^{\ 2}}{r}}$	$\Delta_\omega = t\sqrt{\frac{\delta_\omega^{\ 2}}{r}}$
	Безповторний	
	для середньої	для частки
Власне випадкова, механічна	$\Delta_x = t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$	$\Delta_\omega = t\sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)}$
Стратифікована (при пропорційному обсязі груп у вибірці)	$\Delta_x = t\sqrt{\frac{\overline{\sigma_i}^2}{n}\left(1-\frac{n}{N}\right)}$	$\Delta_\omega = t\sqrt{\frac{\overline{\omega_i(1-\omega_i)}}{n}\left(1-\frac{n}{N}\right)}$
Серійна	$\Delta_x = t\sqrt{\frac{\delta_x^{\ 2}}{r}\left(\frac{R-r}{R-1}\right)}$	$\Delta_\omega = t\sqrt{\frac{\delta_\omega^{\ 2}}{r}\left(\frac{R-r}{R-1}\right)}$

Таблиця 6.3

Формули для визначення необхідного обсягу вибірки

Види вибіркового спостереження	Повторний	Безповторний
Випадковий, механічний відбір: а) для визначення середнього розміру ознаки; б) для визначення частки ознаки	$n = \frac{t^2\sigma^2}{\Delta_x^{\ 2}}$ $n = \frac{t^2\omega(1-\omega)}{\Delta_\omega^2}$	$n = \frac{t^2\sigma^2 N}{N\Delta_x^{\ 2} + t^2\sigma^2}$ $n = \frac{t^2\omega(1-\omega)N}{N\Delta_\omega^2 + t^2\omega(1-\omega)}$
Типова: а) для визначення середнього розміру ознаки; б) для визначення частки ознаки	$n = \frac{t^2\overline{\sigma}^2}{\Delta_x^{\ 2}}$ $n = \frac{t^2\overline{\omega(1-\omega)}}{\Delta_\omega^2}$	$n = \frac{t^2\overline{\sigma^2}N}{N\Delta_x^{\ 2} + t^2\overline{\sigma^2}}$ $n = \frac{t^2\overline{\omega(1-\omega)}N}{N\Delta_\omega^2 + t^2\overline{\omega(1-\omega)}}$
Серійна: а) для визначення середнього розміру ознаки; б) для визначення частки ознаки	$r = \frac{t^2\delta^2}{\Delta_x^2}$ $r = \frac{t^2\omega_r(1-\omega_r)}{\Delta_\omega^2}$	$n = \frac{t^2\delta^2 R}{R\Delta_x^2 + t^2\delta^2}$ $r = \frac{t^2\omega_r(1-\omega_r)R}{R\Delta_\omega^2 + t^2\omega_r(1-\omega_r)}$

Дискретна випадкова змінна

Випадкова змінна	будь-яка змінна, значення якої не може бути точно передбачене.
Дискретна випадкова змінна	випадкова величина, що має певний набір можливих значень.

▷ **ПРИКЛАД** дискретної випадкової змінної.

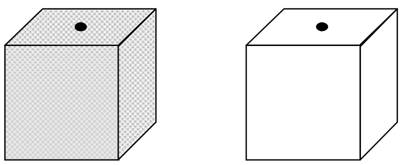
Сума очок, що випали при киданні двох гральних кісточок. Припустимо, що одна з них чорна, а інша – біла. Якщо їх кинути, то можливі 36 елементарних результатів експерименту, оскільки на чорній кісточці може випасти будь-яке число від 1 до 6 і те ж саме – на білій. Випадкова змінна, визначена як їх сума, яку ми позначимо через x, може набувати лише одного з 11 числових значень – від 2 до 12. Взаємозв'язок між результатами експерименту і значеннями випадкової величини в даному випадку показаний у табл. 6.4.

Таблиця 6.4

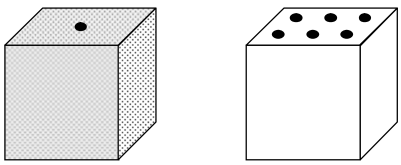
Результати в прикладі з двома гральними кісточками

Біла	Чорна					
	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11

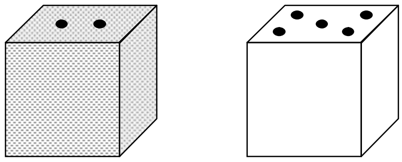
Припустивши, що кісточки "правильні", ми можемо скористатися табл. 6.4 для визначення імовірності кожного значення x. Оскільки на кісточках є 36 різних комбінацій, кожен результат має імовірність 1/36.



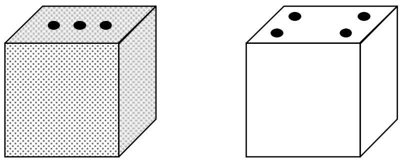
Лише одна з можливих комбінацій {чорна = 1, біла = 1} дає суму, що дорівнює 2, тому імовірність $x = 2$ дорівнює $1/36$. Щоб одержати суму $x = 7$, нам потрібно поєднання {чорна = 1, біла = 6},



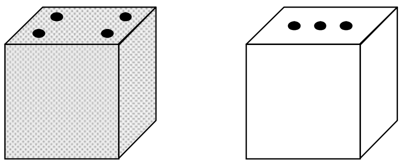
або {чорна = 2, біла = 5},



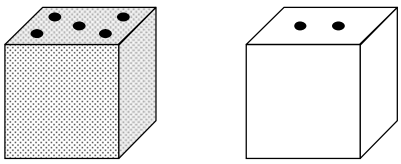
або {чорна = 3, біла = 4},



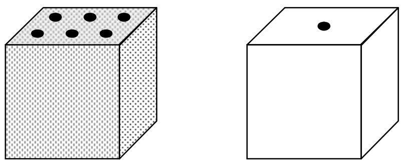
або {чорна = 4, біла = 3},



або {чорна = 5, біла = 2},



або {чорна = 6, біла = 1}.



У даному випадку нас влаштують 6 можливих результатів, і тому імовірність отримання 7 дорівнює 6/36. Вся ця імовірність наведена в табл. 6.5. Якщо всі їх скласти, то отримаємо рівно 1. Це буде так, оскільки з імовірністю 100 % дана сума набуде одного із значень від 2 до 12.

Сукупність всіх можливих значень випадкової змінної описується *генеральною сукупністю*, з якої витягуються ці значення. У нашому випадку генеральна сукупність – це набір чисел від 2 до 12.

Таблиця 6.5

Значення x	2	3	4	5	6	7	8	9	10	11	12
Імовірність	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Проілюструємо наші міркування на рис. 6.1.

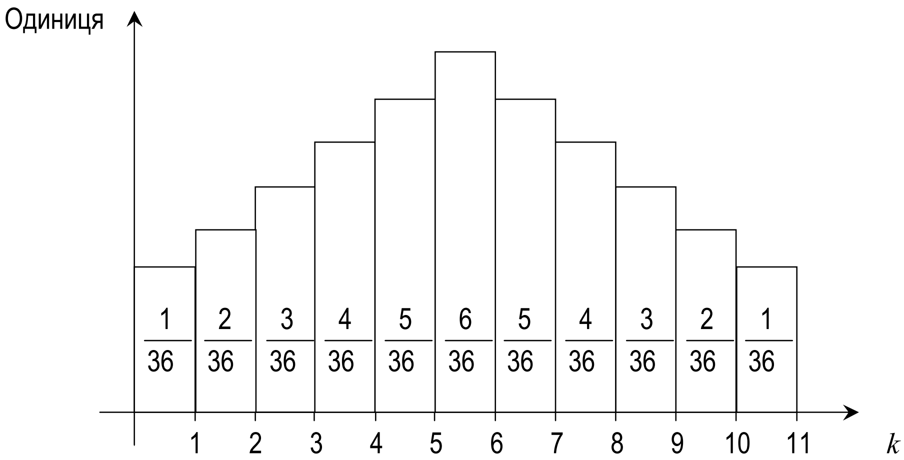


Рис. 6.1. Дискретна імовірність
(приклад з двома гральними кісточками) ■

Способи оцінювання та оцінки

Процедура оцінювання завжди однакова. Береться вибірка з *n* спостережень, і за допомогою відповідної формули розраховується оцінка потрібної характеристики.

Спосіб оцінювання	загальне правило або формула.
Значення оцінки	конкретне число, яке змінюється від вибірки до вибірки.

У табл. 6.6 наведені формули оцінювання для двох найважливіших характеристик генеральної сукупності. *Вибіркове середнє* $\bar{x}$ *звичайно дає оцінку для математичного очікування, а формула в табл. 6.6 – оцінку дисперсії генеральної сукупності*.

Таблиця 6.6

Характеристика генеральної сукупності	Формула оцінювання
Середнє, μ	$\bar{x} = \frac{1}{n}\sum x_i$
Дисперсія, σ^2	$s^2 = \frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2$

Одержана *оцінка* являє собою окремий випадок випадкової змінної. Причина тут у тому, що поєднання значень x у вибірці випадкове, оскільки x – *випадкова змінна, отже, випадковою величиною є і функція набору її значень.*

Незміщеність	якщо математичне очікування оцінки дорівнює відповідній характеристиці генеральної сукупності, то оцінка називається *незміщеною*. Якщо це не так, то оцінка називається *зміщеною*, і різниця між її математичним очікуванням і відповідною теоретичною характеристикою генеральної сукупності називається *зсувом*.
Ефективність	ми хотіли б, щоб наша оцінка з максимально можливою імовірністю давала б близьке значення до теоретичної характеристики, що означає бажання одержати функцію щільності імовірності, якомога більш "густішу" навколо істинного значення. Один із способів виразити цю вимогу – отримати наскільки можна малу дисперсію. Зауваження – ефективність оцінок можна порівнювати лише тоді, коли вони використовують одну і ту ж саму інформацію, наприклад, один і той же набір спостережень декількох випадкових змінних. Якщо одна з оцінок використовує в 10 разів більше інформації, ніж інша, вона цілком може мати меншу дисперсію, але було б неправильно вважати її ефективнішою.
Спроможність	якщо межа оцінки за імовірністю дорівнює істинному значенню характеристики генеральної сукупності, то ця оцінка називається спроможною. *Іншими словами, спроможною називається така оцінка, яка дає точне значення для великої вибірки незалежно від конкретних спостережень, які містяться в ній.*

2. СТАТИСТИЧНЕ ВИВЧЕННЯ ВЗАЄМОЗВ'ЯЗКУ ЕКОНОМІЧНИХ ПОКАЗНИКІВ. ЕКОНОМЕТРИЧНІ МОДЕЛІ ТА МЕТОДИ

Усі економічні явища і процеси взаємопов'язані і взаємообумовлені. Це пов'язано з тим, що будь-які явища є наслідком дії певної множини факторів (результативною ознакою) і, водночас, – причиною інших явищ (факторною ознакою).

Факторна ознака	причини дії певної множини факторів і умови їх виникнення.
Результативна ознака	наслідок (результат) дії певної множини факторів.

Теоретичні основи кореляційно-регресійного аналізу

Функціональні та стохастичні зв'язки. При аналізі зібраного реального статистичного матеріалу ми найчастіше маємо дискретні значення показника, який нас цікавить.

▷ **ПРИКЛАД** статистичного матеріалу.

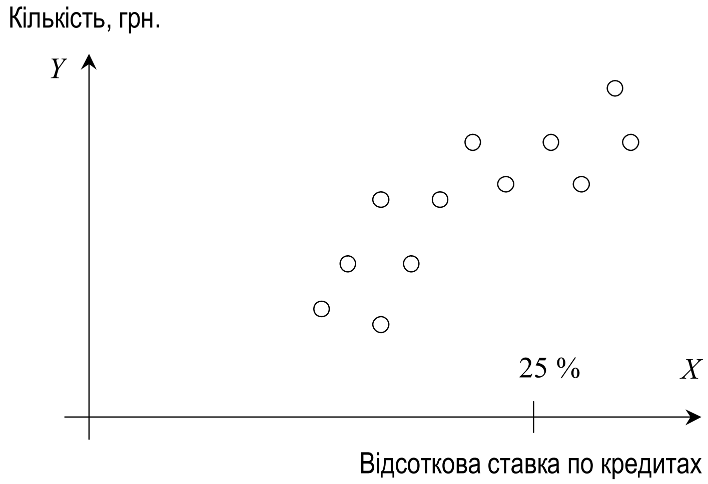
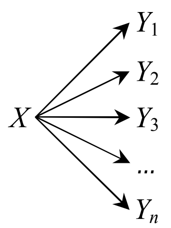
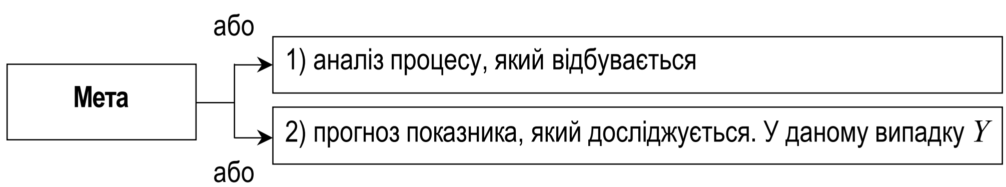


Рис. 6.2. Залежність обсягу кредитів від відсоткової ставки, грн. ■

Стохастичні залежності — X є випадковою величиною і одному значенню X може відповідати декілька значень Y.

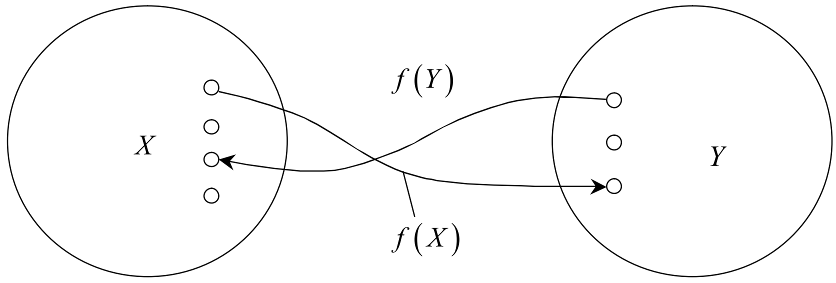


У нас може бути 2 мети:



При реалізації цих варіантів ми використовуватимемо математичний апарат (математичний аналіз). Ми наближатимемо поле точок функцією, але функціональна залежність – це взаємно однозначна відповідність множин X і Y області визначення множини значень.

І це є основною складністю.



Коректним є основний постулат стохастичної школи (вперше сформульований англійським вченим С. Джевонсом (1835-1882), що зв'язав теорію імовірності з індуктивним методом): дані про вірогідні події, здобуті шляхом спостереження, можуть бути тільки приблизними. Як писав Косинський: "...все решения теории вероятностей в пределе (при бесконечном числе наблюдений) достоверны. Но ... произвести бесконечного числа наблюдений мы не можем" (Указ. соч. С. 1). Звідси важливе значення має імовірна помилка. У сукупності випадкові причини, які викликають цю помилку, підпорядковані деяким законам, які він сформулював так: 1) оскільки немає причини, внаслідок якої помилка швидше відбулася б в один, ніж в інший бік, то позитивні та негативні помилки, однакові за величиною, рівноімовірні; 2); величина помилки не може перевершити деякої межі; 3) помилка може набувати будь-яких значень в інтервалі між нулем і межею помилки; 4) за відсутності яких-небудь навмисних перекручувань спостереження із збільшенням помилки імовірність її зменшується, із зменшенням – збільшується. Таким чином, за Косинським, розмір імовірної помилки залежить від кількості спостережень і ступеня точності окремого спостереження.

Парна регресія

1. Нехай є дві випадкові величини X і Y із заданим сумісним розподілом імовірностей.

2. Для кожного можливого значення $X = x$, яке можна зафіксувати, існуватиме деякий *розподіл* випадкової величини Y. З деякою часткою умовності це можна подати так, як показано на рис. 6.3.

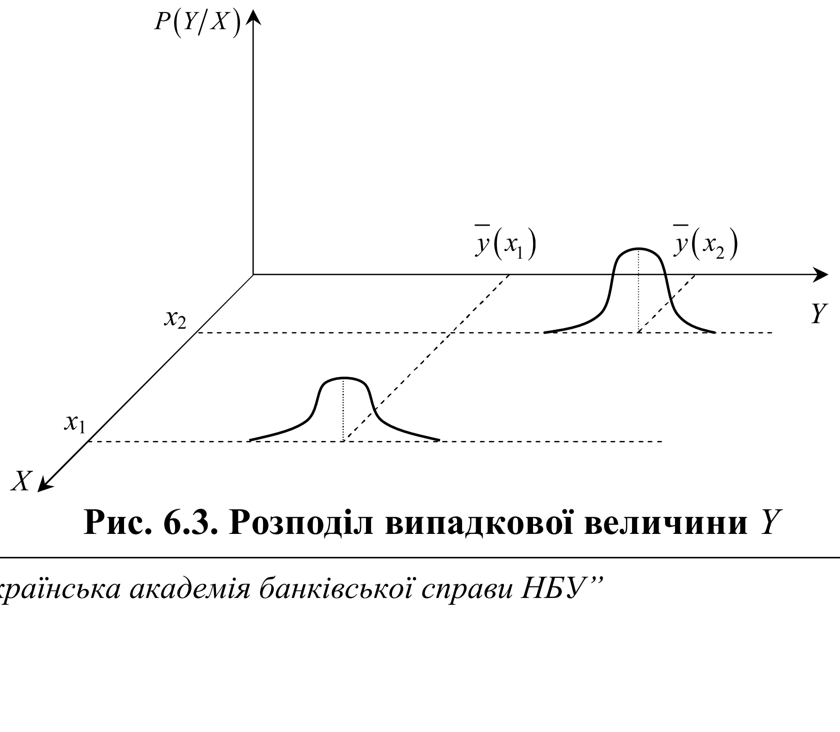


Рис. 6.3. Розподіл випадкової величини Y

3. Для кожного $X = x$ можна визначити умовне математичне очікування величини Y:

$$\bar{y}(x) = M\big(y(X = x)\big). \qquad (6.4)$$

4. Крива регресії якщо x змінюється, то точка $\big((x, \bar{y}(x)\big)$ описує деяку криву, деяку безліч точок на площині (XY). Ця крива графічно описує залежність між величиною X і математичним очікуванням величини Y (середнім значенням). Ця крива називається кривою регресії (рис. 6.4).

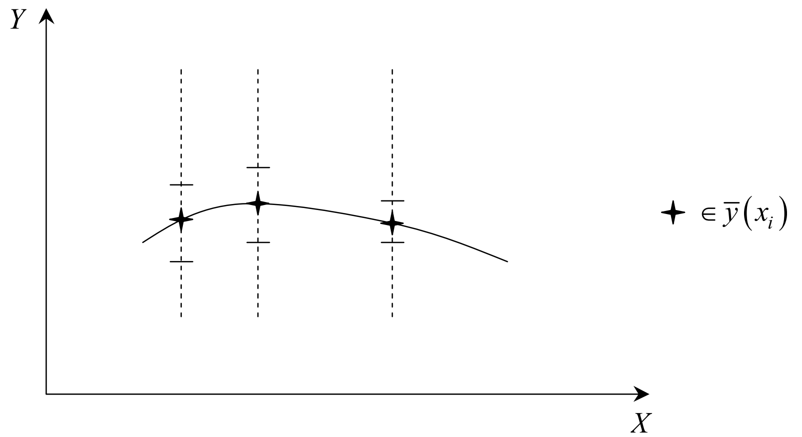


Рис. 6.4. Крива регресії

Основне припущення – існує зв’язок між X і Y.

5. Функція регресії функція, яка описує залежність між X і математичним очікуванням (середнім значенням) величини Y, називається *функцією регресії*

$$\bar{y}(x) = g(x).$$

6. Функція регресії Y на X має таку важливу *властивість*: у класі Q всіх дійсних функцій $F(x)$, які могли б виражати (за пропозицією) залежність між X і Y, функція регресії дає найкраще представлення величини Y з погляду критерію оптимальності, що виражається методом найменших квадратів, *тобто зі всіх функцій тільки для функції регресії $g(x)$ забезпечується мінімальне значення математичного очікування квадрата відхилень Y від даної функції $g(x)$*;

$$\min_{F(x)\in Q} M\big[Y - F(x)\big]^2 = M\big[Y - g(x)\big]^2. \qquad (6.5)$$
$$F(x) \in Q.$$

Впровадження у статистику теорії помилок вимірювання, починаючи з Кетле, привело до розробки *теорії оцінювання.* На початку ХХ ст. завдяки працям російського статистика А.А. Чупрова розповсюдилося трактування вибіркових статистичних характеристик, що підсилило увагу до методологічних проблем статистичного оцінювання (точкового та інтервального). К. Пірсон бачив основу їх рішення в понятті апостеріорної або зворотної імовірності та теоремі Т. Байєса (1763), згідно з якою, якщо за даними про імовірність події можна міркувати про його частість, то і за даними про частість події можна судити про його імовірність. Цю точку зору поділяв Ф.І. Еджворт. На основі методу найменших квадратів і методу моментів Пірсон отримав оцінки основних характеристик ряду розподілу, коефіцієнта кореляції. Всі вони ґрунтувалися на припущенні про нормальність розподілу вибіркових характеристик. Він досліджував розподіл χ^2, запропонував *критерій* χ^2 (1900), який вважав лише "критерієм добротності вирівнювання".

Еджворт розробив методи оцінки істотності відмінностей двох середніх величин (для великих вибірок), запропонував формулу середньої квадратичної помилки коефіцієнта регресії. Процедура оцінювання у нього тісно пов'язана з групуванням даних, виділенням однорідних об'єктів, підпорядкованих нормальному розподілу.

У 1908 р. англієць Вільям Госсет (1876-1937), що опублікував свою працю під псевдонімом "Стьюдент", висвітливши один з найважливіших статистичних розподілів – t*-розподіл (розподіл Стьюдента).* Було доведено, що імовірність помилки вибіркової середньої (або частки, частоти) залежить не тільки від величини відхилення від генеральної середньої (або імовірності), але і від обсягу вибірки. Це стало підставою для вирішення проблем малої вибірки.

Наступний крок в розвитку теорії оцінювання був зроблений Р.А. Фішером. Він відкинув теорію апостеріорної імовірності та теорему Бейєса і запропонував новий метод оцінювання – *максимальної правдоподібності* (1925). Фішер строго розмежовував *параметри* – невідомі характеристики генеральної сукупності і *статистики* – характеристики, що спостерігаються. Він ввів поняття спроможних, ефективних, достатніх статистик, які стали фундаментальними поняттями математичної статистики. Фішер показав, що у випадках малих вибірок і при значеннях коефіцієнтів кореляції, близьких до 1 (за абсолютною величиною), розподіл коефіцієнта кореляції не може вважатися нормальним, і запропонував спеціальні методи для визначення того, що кореляція, яка спостерігається, істотно відрізняється від деякого теоретичного значення і що дві кореляції, що спостерігаються, істотно відрізняються одна від одної. Рішення цих задач ґрунтувалося на перетворенні коефіцієнта кореляції: переході від величини r до величини $z=\frac{1}{2}\left[\ln(1+r)-\ln(1-r)\right]$, яка має нормальний розподіл із середньою квадратичною помилкою $\sigma_z = 1/\sqrt{n-3}$. Перехід до перетворення Фішера дозволив оцінювати частинні і множинні коефіцієнти кореляції.

Нейман ввів поняття *найкращої незміщеної оцінки,* підкресливши, що серед різних можливих розрахунків найбільшою точністю володітиме той, який включає мінімальний показник варіації. Він розвивав теорію інтервального оцінювання (1937).

У розробку ефективних оцінок при заданих функціях втрат і інших проблем оцінювання значний внесок зробив шведський математик Гарольд Крамер (1893).

У період 1925-1935 рр. сформувався новий напрям математичної статистики – *теорія випробування статистичних гіпотез.*

Умови Гаусса-Маркова

1. $M[e_i] = 0$ для всіх спостережень, тобто математичне сподівання випадкового члена у будь-якому спостереженні має дорівнювати нулю. Іноді випадковий член буде додатним, іноді від'ємним, але не повинен мати систематичного зміщення в жодному з двох можливих напрямків

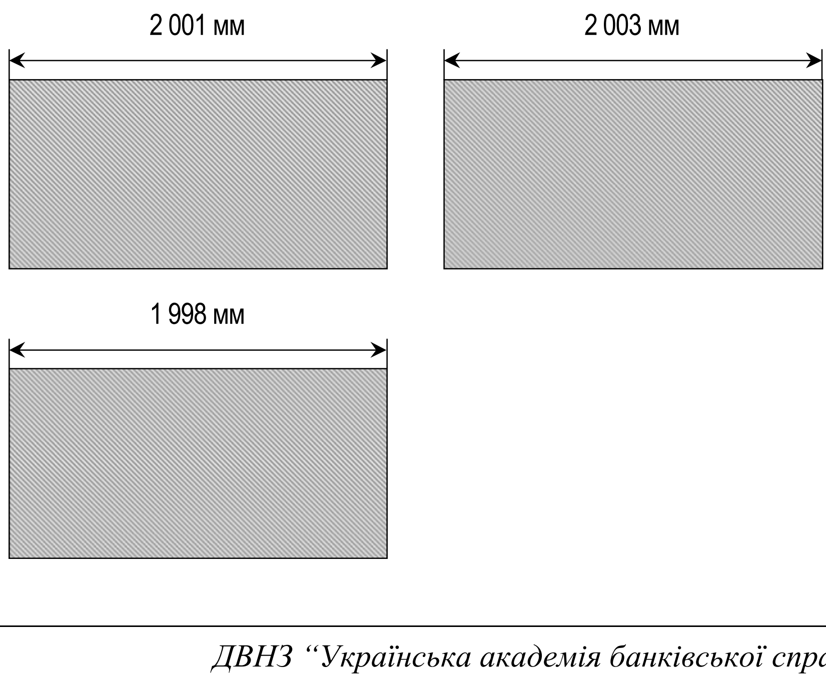
2. $D[e_i] \to \min$ для усіх спостережень, тобто дисперсія випадкового члена має бути постійною для всіх спостережень. Іноді випадковий член буде додатним, іноді від'ємним, проте не повинно бути апріорної причини для того, щоб він породжував більшу помилку в одних спостереженнях, ніж в інших

3. Відсутність систематичного зв'язку між значеннями випадкового члена у будь-яких двох спостереженнях. Наприклад, якщо випадковий член великий і додатний в одному спостереженні, це не повинно обумовлювати систематичну тенденцію до того, що він буде великим і додатним у наступному спостереженні (або великим і від'ємним, або малим і додатним, або малим і від'ємним). Випадкові члени мають бути абсолютно незалежними один від одного

4. Випадковий член має бути розподілений незалежно від пояснюючих змінних

▷ **ПРИКЛАД** виконання умов Гауса-Маркова. Розглянемо довжину столів у банку. Банк придбав 3 столи. За накладною $l = 2\,000$ мм.

Проведемо експеримент – візьмемо сантиметр і заміряємо довжину столів.



$l_{теор} = 2\,000, l_{пр1} = 2\,001, l_{пр2} = 2\,003, l_{пр3} = 1\,998.$

$e_i = Y_T - Y_{сп}; e_1 = 2\,001 - 2\,000 = 1, e_2 = 2\,003 - 2\,000 = 3, e_3 = 1\,998 - 2\,000 = -2.$

$M(e_i) = 0, D(e_i) \to \min.$ ■

Метод найменших квадратів

Прості лінійні регресійні моделі встановлюють лінійну залежність між двома змінними, наприклад, обсягом резервів банку та складом кредитного портфеля; обсягом витрат банку та обсягом депозитів, зміною рейтингу банку залежно від часу та ін.

Залежна та незалежна змінні	при цьому одна із змінних вважається залежною змінною (y) та розглядається як функція від незалежної змінної (x).

У загальному вигляді проста вибіркова регресійна модель запишеться так:

$$y = a_0 + a_1 x + e, \tag{6.6}$$

де y – вектор спостережень за залежною змінною; $y = \{y_1, y_2, ..., y_n\}$;

x – вектор спостережень за незалежною змінною; $x = \{x_1, x_2, ..., x_n\}$;

$a_0,\ a_1$ – невідомі параметри регресійної моделі;

e – вектор випадкових величин (помилок), $e = \{e_1, e_2, ..., e_n\}$.

Лінійна регресійна модель	регресійна модель називається лінійною, якщо вона лінійна за своїми параметрами. Отже, модель (6.6) є лінійною регресійною моделлю, її ще можна трактувати як пряму на площині, де a_0 – перетин з віссю ординат, а a_1 – нахил (звичайно, якщо абстрагуватися від випадкової величини e).

Щоб мати явний вид залежності, необхідно знайти (оцінити) невідомі параметри a_0, a_1 цієї моделі. Як це зробити? Яким критерієм краще користуватися? Щоб відповісти на ці запитання, розглянемо спочатку приклад.

▷ **ПРИКЛАД** регресійної моделі. Відділ економічного аналізу комерційного банку оцінює ефективність кредитного відділу. Для такої оцінки вони мають досвід роботи у 5 географічних зонах з майже однаковими умовами (потенційні клієнти, ставлення до даного банку та ін.). У цих зонах вони зафіксували протягом однакового періоду обсяги виданих кредитів і витрати банку, пов'язані з рекламною кампанією (млн. грн.). Дані наведені в табл. 6.7.

Таблиця 6.7

I	y_i	x_i
1	25	5
2	30	6
3	35	9
4	45	12
5	65	18

Реальні спостереження y зобразимо точками у системі координат (X, Y) (рис. 6.5).

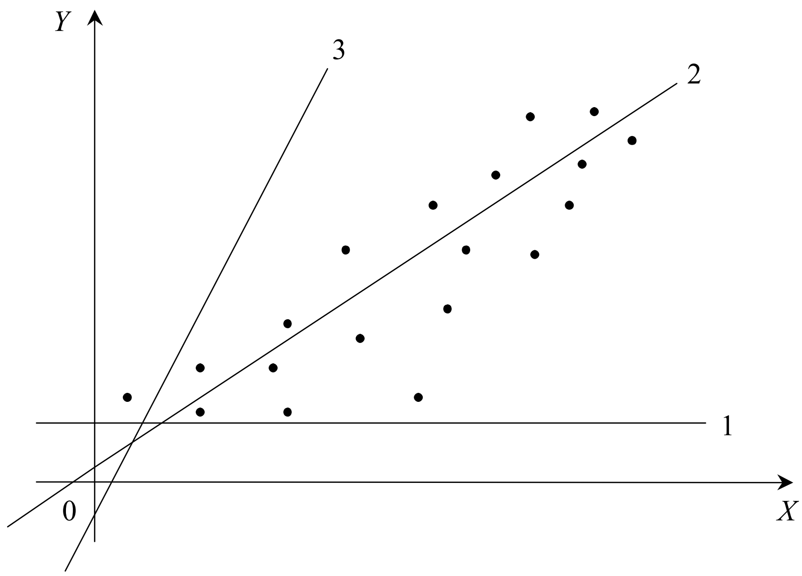


Рис. 6.5. Залежність між обсягами виданих банком кредитів і витратами на рекламу ■

Яка пряма більше відповідає? Інтуїтивно ми обираємо 2. Запишемо це у математичних термінах. Що таке математичні терміни? Це координати точок.

Візуально можна припустити, що між даними є лінійна залежність, тобто їх можна апроксимувати прямою лінією.

Взагалі існує необмежена кількість прямих $y = a_0 + a_1 x,$ які можна провести через множину спостережуваних точок. Яку ж із них вибрати?

Щоб це визначити, потрібно мати у розпорядженні певний критерій, що дозволяв би вибрати з множини можливих прямих "найкращу" з точки зору даного критерію. Найпоширенішим є критерій мінімізації суми квадратів відхилень. На рис. 6.5, наприклад, пряма (1), як і інші, розташована таким чином, що деякі точки знаходяться вище, деякі нижче цієї прямої, на основі чого можна встановити відхилення (помилки) щодо цієї прямої:

$$e_i = y_i - \hat{y}_i = y_i - a_o - a_1 x_i, \; i = \overline{1, n}, \tag{6.7}$$

де $\hat{y}_i$ – i-та точка на прямій, яка відповідає значенню x_i (рис. 6.6).

Відхилення, або помилки, ще інколи називають залишками. Логічно, що треба проводити пряму таким чином, щоб сума квадратів помилок була мінімальною. У цьому і полягає критерій найменших квадратів: невідомі параметри a_0 та a_1 визначаються таким чином, щоб мінімізувати $\sum_{i=1}^{n} e_i^2$. Справді, за критерієм маємо

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i \right)^2 = f\left(a_0, a_1\right) \to \min. \tag{6.8}$$

Це функція двох змінних a_0 та a_1.

Знайдемо мінімум функції двох змінних.

Теорема (необхідна умова екстремуму) нехай функція $f\left(x_1, ..., x_n\right)$ визначена в деякому околі точки $x^0 = \left(x_1^0, ..., x_n^0\right) \in R^n$. Припустимо, що x^0 – точка екстремуму функції f. Нехай в точці x^0 існують частинні похідні функції f за змінною x_i. Тоді

$$\frac{\partial f}{\partial x_i}\left(x^0\right) = 0.$$

Теорема (достатня умова екстремуму) нехай функція $f\left(x_1, ..., \cdots x_n\right)$ визначена та має неперервні частинні похідні другого порядку в деякому околі стаціонарної точки $x^0 = \left(x_1^0, ..., x_n^0\right)$. Тоді для того, щоб точка x^0 була точкою строгого максимуму (строгого мінімуму) функції f, достатньо, щоб функція f була строго випуклою вверх (вниз) в околі точки x^0.

Визначимо значення a_0 та a_1, які мінімізують вираз (6.8). Мінімум функції (6.8) досягається за необхідних умов, коли перші похідні дорівнюють нулю, тобто:

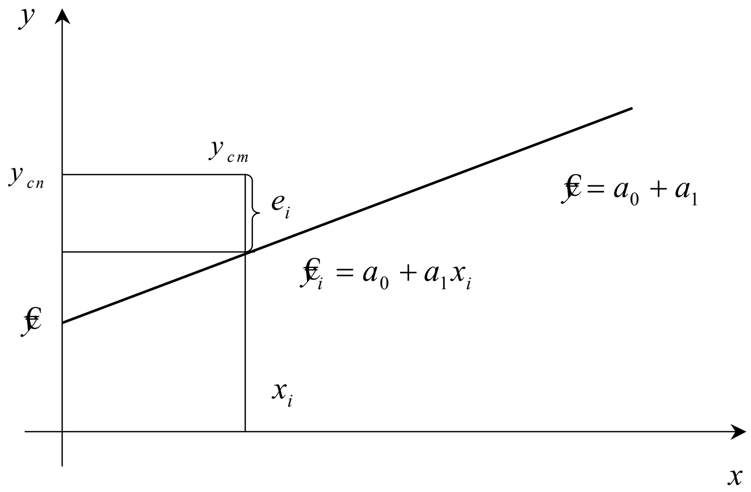


Рис. 6.6. Відхилення теоретичних значень від фактичних

$$\frac{\partial\left(\sum_{i=1}^{n} e_i^2\right)}{\partial a_0} = \frac{\partial f\left(a_0, a_1\right)}{\partial a_0} = 0; \tag{6.9}$$

$$\frac{\partial\left(\sum_{i=1}^{n}\left(y_i - a_0 - a_1 x_i\right)^2\right)}{\partial a_1} = \frac{\partial f\left(a_0, a_1\right)}{\partial a_1} = 0.$$

$$\begin{cases} \dfrac{\partial\left(\sum_{i=1}^{n} e_i^2\right)}{\partial a_0} = -2\sum_{i=1}^{n}\left(y_i - a_0 - a_1 x_i\right) = 0 \\ \dfrac{\partial\left(\sum_{i=1}^{n} e_i^2\right)}{\partial a_1} = -2\sum_{i=1}^{n} x_i\left(y_i - a_0 - a_1 x_i\right) = 0. \end{cases} \tag{6.10}$$

Вирішимо систему лінійних алгебраїчних рівнянь за допомогою теореми Кронекера-Капеллі.

Теорема Кро́некера-Капе́ллі	критерій сумісності системи лінійних алгебраїчних рівнянь: система лінійних алгебраїчних рівнянь сумісна тоді і тільки тоді, коли ранг її основної матриці дорівнює рангу її розширеної матриці.

Отримаємо систему лінійних рівнянь:

$$\begin{cases} \sum_{i=1}^{n} y_i = na_0 + a_1 \sum_{i=1}^{n} x_i; \\ \sum_{i=1}^{n} y_i x_i = a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2, \end{cases} \tag{6.11}$$

яка називається нормальною.

Розв'язок (6.11) щодо нахилу прямої (невідома a_1) дає

$$a_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}. \tag{6.12}$$

З метою спрощення виразу для a_1 чисельник і знаменник виразу (6.12) помножимо на $1/n$. Отримаємо:

$$a_1 = \frac{\frac{1}{n}\left(\sum_{i=1}^{n} x_i y_i\right) - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}, \tag{6.13}$$

де $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i;\ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$

Додаток

Коефіцієнт коваріації	$\operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$
Дисперсія величини x	$\operatorname{var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$

Отже, кут нахилу прямої регресії можна встановити за формулою (6.12).

Для визначення параметра a_0 повернемося до (6.10). Маємо:

$$\frac{\partial\left(\sum_{i=1}^{n} e_i^2\right)}{\partial a_0} = \sum_{i=1}^{n} \frac{\partial}{\partial a_0}\left[\left(y_i - a_0 - a_1 x_i\right)^2\right] = \sum_{i=1}^{n}\left(y_i - a_0 - a_1 x_i\right) = 0. \quad (6.14)$$

Вираз (6.14) дає нам, по-перше, підтвердження того, що сума помилок дорівнює нулю. Справді,

$$y_i - a_0 - a_1 x_i = e_i \Rightarrow \sum_{i=1}^{n} e_i; \quad (6.15)$$

по-друге, розділивши (6.15) на n, маємо вираз для визначення a_0:

$$\frac{1}{n}\sum_{i=1}^{n} y_i - a_0 - a_1 \frac{1}{n}\sum_{i=1}^{n} x_i = 0 \Rightarrow a_0 = \overline{y} - a_1 \overline{x}. \quad (6.16)$$

Таким чином, ми знайшли формули для визначення невідомих параметрів a_0 та a_1 і можемо записати в явному вигляді регресію y від x, в якій параметри обчислені за методом найменших квадратів, її інколи називають регресією найменших квадратів y від x. Маємо:

$$\hat{y} = a_0 + a_1 x \quad (6.17)$$

або

$$y = \hat{y} + e = a_0 + a_1 x + e. \quad (6.18)$$

▷ **ПРИКЛАД** ілюстрації побудови рівняння регресії.

Для ілюстрації цих викладок повернемося до нашого прикладу про дослідження ефективності витрат на рекламу. Проведені попередні розрахунки подамо у вигляді табл. 6.8.

Для обчислення невідомих параметрів a_0 та a_1 необхідно послідовно здійснити такі розрахунки:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{50}{5} = 10; \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{200}{5} = 40.$$

Таблиця 6.8

Дослідження ефективності витрат на рекламу

I	y_i	x_i	x_i^2	$x_i y_i$
1	25	5	25	125
2	30	6	36	180
3	35	9	81	315
4	45	12	144	540
5	65	18	324	1170
2	200	50	610	2330
$\sum / n$	40	10	122	466

$$\operatorname{var}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{610}{5};$$

$$\operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\,\overline{y} = \frac{2\,330}{5} - 400 = 66;$$

$$a_1 = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)} = \frac{66}{22} = 3;\ a_0 = \bar{y} - a_1\bar{x} = 40 - 3\cdot 10 = 10.$$

Знаючи параметри a_0 та a_1, отриману пряму запишемо у вигляді: $\hat{y} = 3x + 10$. ■

Кетле вважав, що можна виміряти будь-яку властивість людини і виділити в ній детерміновану компоненту та індивідуальну (випадкову): $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, де y_{ijk} – значення ознаки y для k-ї одиниці, що входить до i-ї групи за однією ознакою та j-ї – за іншою; характеристика типу “середня величина” для ij-ї групи. Він вважав, що в групових середніх μ_{ij} можна виділити загальну причину μ, яка відображає вплив суспільства, і змінні причини (α_i, β_j), що відображають властивості даного типу (групи): $\mu_{ij} = \mu + \alpha_i + \beta_j$. Кетле надавав дуже великого значення кількості спостережень, вважаючи, що чим їх більше, тим ближча істина.

7. Дійсну функцію регресії визначити важко. Вчиняють так: припускають, що функція регресії належить до якого-небудь наперед вибраного класу функцій. При цьому ставиться задача вже з вибраного класу функцій виділити таку, для якої математичне очікування квадрата відхилень від Y було б мінімальним.

Функція середньої квадратичної регресії — функція $\varphi(x)$, для якої математичне очікування квадрата відхилень від Y було б мінімальним.

8. Доведено, що якщо випадкові величини X і Y розподілені за двовимірним *нормальним законом,* то

Випадкові величини X і Y розподілені за двовимірним *нормальним законом* ⇒
- функція регресії – лінійна
- крива регресії – пряма лінія
- лінія регресії *збігається з кривою середньої квадратичної регресії*

Останнє дає важливий практичний результат, а саме: те, що можна визначити параметри функції регресії, користуючись *методом найменших квадратів як обчислювальною процедурою.*

9. Лінійна функція регресії

$$\varphi(x) = \beta_0 + \beta_1 x.$$

Параметри функції на основі МНК обчислюються за формулами:

$$\beta_0 = MY - \rho\frac{\delta_y}{\delta_x}MX.$$

$$\beta_1 = \rho\frac{\delta_y}{\delta_x},$$

де ρ – коефіцієнт кореляції;

δ_x, δ_y – стандартні відхилення X, Y.

10. Випадкова величина Y

$$Y = \varphi(x) + \varepsilon,$$

де ε – залишок, випадкова величина.

Оскільки $MY = \varphi(x)$, то $M\varepsilon = 0$.

11. Згідно з процедурою МНК шукається мінімум виразу

$$M\left[y - \varphi(x)\right]^2. \tag{6.19}$$

Цей вираз є дисперсією залишку:

$$M\left[Y - \varphi(x)\right]^2 = M(\varepsilon)^2 = \delta^2{}_{ост}. \tag{6.20}$$

Очевидно, що чим краще крива $\varphi(x)$ апроксимує величину *Y*, тим $\delta^2{}_{ост}$ менше. Доведено, що $\delta^2{}_{ост} = \delta^2{}_y\left(1-\rho^2\right)$ або $\delta^2{}_{ост} = \delta^2{}_y - \rho^2\delta_y$, або

$$\rho^2 = \frac{\rho^2{}_y - \rho^2{}_{ост}}{\rho^2{}_y} \quad 1-\frac{\rho^2{}_{ост}}{\rho^2{}_y}. \tag{6.21}$$

Коефіцієнт ρ^2 є показником точності апроксимації величини *Y* функцією $\varphi(x)$. Якщо $\rho = 1$, то $\rho^2{}_{ост} = 0$, а це означає, що залежність між *Y* і *X* функціональна, усі *Y* повністю визначаються величиною *X*.

Якщо $\rho = 0$, то залишкова дисперсія дорівнює дисперсії самої величини *Y*. У цьому випадку говорять, що *X* і *Y* некорельовані. Лінія регресії на графіку буде паралельною осі *OX*.

12. Якщо випадкові величини розподілені за іншим законом, що відрізняється від нормального, то лінії регресії – деякі криві.

13. Оскільки на практиці надається обмежена кількість даних, деяка вибірка із сукупності іноді з відомим, частіше з невідомим законом розподілу, то на основі цих даних будують *емпіричну функцію регресії*, параметри якої розглядаються як *оцінки* параметрів істинної регресії.

Задачу оцінки параметрів теоретичної функції регресії за вибірковими даними вирішує *регресійний аналіз.*

Основні задачі та поняття регресійного аналізу

Регресійний аналіз

1. Регресійний аналіз – розділ математичної статистики, що об'єднує методи дослідження регресійної залежності між величинами на основі статистичних даних.
2. Регресійний аналіз вирішує такі завдання:
   - вибір моделі регресії;
   - оцінка параметрів моделі;
   - перевірка статистичних гіпотез про модель.
3. Практичні методи регресійного аналізу засновані на цілому ряді теоретичних передумов, ступінь виконання яких припускає успішність застосування регресійного аналізу і коректність висновків.
4. Важлива перевага регресійного аналізу – можливість використання його апарату не тільки у випадку, якщо всі дані представлені випадковими величинами.

Основним завданням регресійного аналізу є побудова регресійної моделі.

Регресійна модель вираз вигляду $\hat{y} = f(x)$, де x – факторні ознаки, або фактори. *Регресійна модель* описує як у середньому залежить якась результативна ознака y від впливаючих на неї факторіальних ознак x.

Факторні ознаки факторна ознака може бути одна, тоді розглядається однофакторна модель регресії, або декілька, тоді розглядається багатофакторна модель регресії.

Модель регресії може описувати як *лінійну* залежність результативної ознаки Y від факторів, так і *нелінійну* залежність.

Лінійна однофакторна модель

$$\hat{y} = a_0 + a_1 x. \quad (6.22)$$

Лінійна багатофакторна модель

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + ... + a_k x_k. \quad (6.23)$$

При цьому необхідно мати на увазі, що регресійна модель може бути лінійною щодо параметрів або нелінійною щодо факторіальних ознак. Наведені моделі лінійні щодо параметрів і факторіальних ознак.

Регресійні моделі можуть бути представлені поліномами різних ступенів:

$$\hat{y} = a_0 + a_1 x + a_2 x^2 + ..., \quad (6.24)$$

показниковими функціями

$$\hat{y} = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} ... \quad (6.25)$$

Лінеаризація зведення функції, нелінійної щодо параметрів щодо лінійного вигляду.

Функцію $y = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2}$ можна привести до лінійного вигляду щодо параметрів a_1 і a_2 шляхом логарифмування.

$$\lg y = \lg a_0 + a_1 \lg x_1 + a_2 \lg x_2.$$

Економічна інтерпретація параметрів моделі

Оскільки коефіцієнти при факторних ознаках у рівняннях (6.22)-(6.23) показують, як змінюється значення результативної ознаки Y, якщо величина даного фактора зростає на одиницю.

Можна обчислити також показники, звані *коефіцієнтами еластичності*, які визначаються таким чином:

$$K_{эi} = a_i \frac{\overline{x}_i}{\overline{y}},$$

і показує, на скільки відсотків зміниться середнє значення Y при зміні даного i-го фактора на 1 %.

У рівнянні регресії (6.25) коефіцієнти регресії a_i відразу інтерпретуються як коефіцієнти еластичності.

У 30-ті роки XX ст. *закон великих чисел* вважався центральним положенням, теоретичною основою статистики. У кінці 40-х – на початку 50-х років XX ст. виникла протилежна точка зору, яка зводилася до визнання дії закону великих чисел лише у вибірковому спостереженні, висловлювалася навіть думка про повне заперечення його значення для соціально-економічної статистики. Йосип Саулович Пасхавер (1907-1980) вважав, що закон великих чисел виконує в соціально-економічній статистиці підсобну, допоміжну роль, оскільки він обумовлений лише формою прояву певних причин, а не їх суттю. Федір Давидович Ліфшиц (1897-1975) підкреслював об’єктивність закону великих чисел і його значення для пізнання статистичних закономірностей. Важливим було внесене Б.І. Карпенком розмежування функцій закону великих чисел як закону теорії імовірності та закону статистики. Кореляційний і регресійний аналіз були першими методами багатовимірного статистичного аналізу (БСА) в радянській статистиці. Вперше після видання монографії А.А. Чупрова “Основные проблемы теории корреляции” (1925) з’явилася вітчизняна робота, присвячена *кореляційному аналізу*, – книга Якова Ілліча Лукомського (1906-1961) “Теория корреляции и ее применение к анализу производства” (1961). Ця книга дотепер залишається однією з кращих щодо застосування кореляційного аналізу, щоправда в ній розглянуте застосування кореляцій в основному в технологічній, а не в соціально-економічній статистиці. Широке впровадження цього методу в практику стримувалося відсутністю електронно-обчислювальної техніки. Тому в книзі Лукомського спеціально приділялась увага способам рішення системи лінійних рівнянь для оцінки параметрів множинної регресії без застосування обчислювальної техніки.

Означення еластичності

Еластичність попиту та пропозиції [255]

Еластичність міра чутливості функціонально пов’язаних величин. Вона визначається як співвідношення процентних змін залежної та незалежної змінних

$$E = \lim_{\Delta x \to 0} \frac{\Delta x}{\Delta y} \cdot \frac{\Delta y}{\Delta x} \cdot 100\,\%.$$

Види еластичності щодо попиту

- еластичність попиту за ціною
- перехресна еластичність попиту
- еластичність попиту за доходом

Еластичність лінійної функції попиту непостійна. Кожна лінійна крива попиту має два відрізки: верхній, в межах якого попит є еластичним, і нижній, в межах якого попит стає нееластичним, вони розмежовуються точкою одиничної еластичності (рис. 6.7). Для нелінійної функції попиту ця закономірність може виконуватись, а може й не виконуватись.

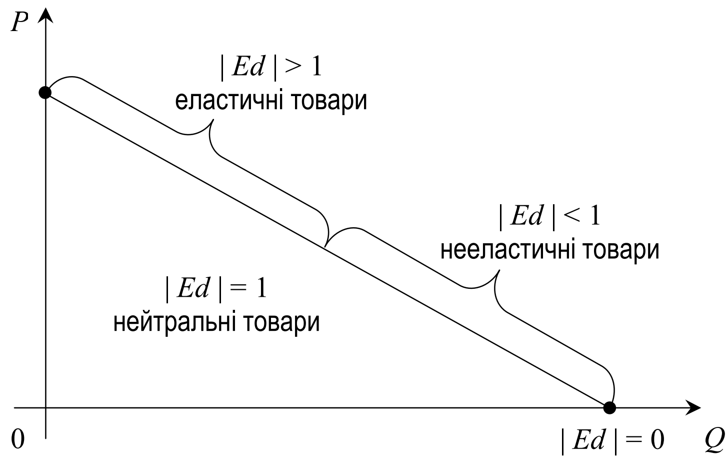


Рис. 6.7. Еластичність лінійної функції попиту

Еластичність пропозиції характеризує чутливість продавців (виробників) до зміни ціни на продукцію.

Для пропозиції, як і для попиту, розрізняють декілька випадків еластичності:

Види еластичності

- еластична пропозиція
- нееластична пропозиція
- пропозиція з одиничною еластичністю

▷ **ПРИКЛАД** еластичних, нейтральних і нееластичних товарів.

Еластичні	продукти харчування не першої необхідності.
Нейтральні	сірники, лампочки.
Нееластичні	товари першої необхідності (хліб, молоко та ін.), тобто споживчий кошик. ■

Залежну величину Y можна подати у вигляді лінійної моделі таким чином:

$$y_i = a_0 + \sum_{j=1}^{k} a_j x_j + e_i, \qquad (6.26)$$

де a_0 і a_i – оцінки параметрів регресії β_0 і β_i.

e_i – залишок, який можна розглядати як оцінку ε_i.

Для отримання чисельних значень оцінок у регресивному аналізі найчастіше застосовується метод найменших квадратів. Метод найменших квадратів у вибраному класі функцій дозволяє знайти функцію $f(x)$, для якої забезпечується

$$\min \sum_{i} \left[y_i - f(x) \right]^2. \qquad (6.27)$$

Звичний метод найменших квадратів є його модифікацією і припускає, що функція $f(x)$ лінійна щодо параметрів.

Задача регресійного аналізу	знаходження таких чисельних значень оцінок параметрів регресії, які б не суперечили основним припущенням про властивості теоретичної регресії.

Отримання таких оцінок методом найменших квадратів припускає виконання *ряду передумов*.

У регресійному аналізі передбачається

- *результативна ознака* – обов'язково *випадкова величина* з нормальним або іншим законом розподілу, а *факторні ознаки* можуть бути як *випадковими*, так і *невипадковими змінними* величинами
- важливою умовою є відсутність лінійної залежності між ознаками факторів (відсутність явища мультиколінеарності)
- найголовніші вимоги в регресійному аналізі ставляться до залишків

Виконання цих умов практично більшою чи меншою мірою здійснено.

При виконанні цих вимог оцінки, одержані за МНК, відповідають вимогам *незміщеності, спроможності та ефективності.*

Якщо ж виконується вимога нормальності розподілу залишків, то незміщеність оцінок зберігається для рівнянь регресії, лінійних щодо параметрів.

Виконання основних вимог при побудові регресійної моделі дає можливість одержати статистично надійні оцінки параметрів регресії, застосовувати математико-статистичний апарат для перевірки тих або інших статистичних гіпотез.

Обчислення за методом найменших квадратів оцінки параметрів регресії є *точковими оцінками* теоретичної регресії.

У регресійному аналізі розв'язуються такі задачі *статистичного аналізу моделі.*

Задачі статистичного аналізу моделі

- а) оцінка точності точкових оцінок, коефіцієнтів регресії
- б) визначення статистичної значущості оцінок
- в) побудова довірчих інтервалів оцінок
- г) перевірка адекватності моделі

а) оцінка точності точкових оцінок коефіцієнтів регресії

здійснюється на основі визначення середньоквадратичної помилки коефіцієнта регресії Sa_i:

$$Sa_i = \sqrt{\frac{\sum (y - \hat{y})^2}{n-k-1} b_{ij}}, \quad i = j - 1, \qquad (6.28)$$

де $\sum (y - \hat{y})^2$ – сума квадратів відхилень фактичних даних від знайдених за регресією (сума квадратів залишків);

$n - k - 1$ – число ступенів свободи;

b_{ij} – діагональний елемент матриці, оберненої до матриці системи нормальних рівнянь.

Ця помилка Sa_i показує, на скільки в середньому відрізняється одержана оцінка параметра регресії a_i від теоретичного значення β_i;

б) для перевірки статистичної значущості одержаних оцінок параметрів рівняння регресії

перевіряють гіпотезу $H_0 / \beta_i = 0$ за допомогою *t*-статистики. Розраховують здійснити перевірку на основі даних $t_{расч} = \frac{a_i}{Sa_i}$, порівнюють *t*-розрахунок з табличним значенням *t*-розподілу Стьюдента при заданому числі ступенів свободи $(n-k-1)$. Якщо $\left|t_{расч}\right| \geq t_{\frac{\alpha}{2}(n-k-1)}$, то коефіцієнт a_i статистично значущий із заданим рівнем довірчої ймовірності, якщо $\left|t_{расч}\right| \langle t_{\frac{\alpha}{2}(n-k-1)}$, то незначущим.

в) довірчі інтервали для коефіцієнтів теоретичного рівняння регресії

довірчі інтервали для коефіцієнтів теоретичного рівняння регресії β_i із заданою довірчою імовірністю $(1-\alpha)$ розраховуються за формулою:

$$a_i - t_{\frac{\alpha}{2}(n-k-1)} Sa_i \leq \beta_i \leq a_i + t_{\frac{\alpha}{2}(n-k-1)} Sa_i. \qquad (6.29)$$

Аналіз регресивної моделі

г) кінцевою стадією статистичного аналізу регресивної моделі є перевірка її адекватності. Розраховується коефіцієнт детермінації.

Коефіцієнт детермінації

$$R^2 = 1 - \frac{\sum\left(y - \hat{y}\right)^2}{\sum\left(y - \bar{y}\right)^2}. \qquad (6.30)$$

Перевіряється його статистична значущість. Коефіцієнт детермінації показує, наскільки вибрана модель краща, ніж $y = \bar{y}$.

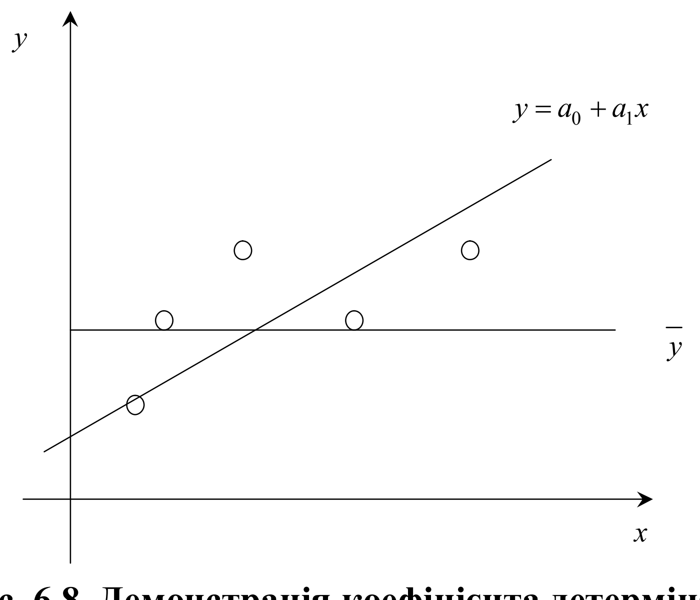


Рис. 6.8. Демонстрація коефіцієнта детермінації

Перевірка значущості проводиться за F-критерієм, значення якого розраховується за формулою

$$F_{расч} = \frac{R^2(n-k-1)}{(1-R^2)\cdot k}. \qquad (6.31)$$

За таблицею F-розподілу за даним числом ступеней свободи і заданим рівнем значущості знаходиться значення $F_{\alpha};\ k,(n-k-1)$.

Якщо

$$F_{расч} \geq F_{\alpha;\ k,(n-k-1)},$$

то R^2 статистично значущий, і модель адекватна.

Якщо

$$F_{расч} < F_{крит},$$

то R^2 незначущий.

10. *Вибір форми* моделі одна з найважливіших задач регресивного аналізу.

Вибір форми моделі	добір рівняння регресії здійснюється із залученням різних методів і критеріїв. Основою насамперед є економічна теорія, що встановлює причинно-наслідкові зв'язки між економічними показниками, характер взаємодії показників. Також враховують практичний досвід аналогічних розробок у схожих сферах. Важливим критерієм є розробленість і простота аналітичного апарату дослідження і облік можливостей ЕОМ, забезпеченість програмними засобами. Розробляються насамперед моделі, що дозволяють коректно використовувати математичні методи і допускають простоту, змістовну економічну інтерпретацію. Найпоширенішою з цієї точки зору є лінійна відносно параметрів модель регресії. Це *загальні критерії*.

Використовуються різні частинні кількісні критерії, в основі яких лежить принцип мінімальної розбіжності величин результативної ознаки, одержаних на основі моделі, з фактичними значеннями за ретроспективний період.

Цим принципом керуються і при графічному аналізі розташування точок діаграми розсіяння на площині та добору кривої (з такого аналізу починається будь-яке дослідження).

Кількісні критерії

1) $\sum\left[y-\hat{y}_{(k)}\right]^2$, де $\hat{y}_{(k)}$ – модель регресії. Модель, при якій забезпечується мінімум суми квадратів відхилень фактичних даних від модельних, вважатиметься найкращою;

2) коефіцієнт Тейла:

$$U=\frac{\sqrt{\frac{\sum(y-\hat{y})^2}{n}}}{\sqrt{\frac{\sum y^2}{n}}+\sqrt{\frac{\sum \hat{y}^2}{n}}}, \tag{6.32}$$

чим ближче до нуля, тим краще;

3) коефіцієнт детермінації (критерій, за допомогою якого також вимірюється щільність зв’язку між двома або більше показниками та перевіряється адекватність (відповідність) побудованої регресійної моделі реальній дійсності. Тобто дається відповідь на запитання, чи справді зміна значення y лінійно залежить саме від зміни значення x, а не відбувається під впливом різних випадкових факторів)

$$R^2_{(k)}=1-\frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}\cdot\frac{n-1}{n-k-1}, \tag{6.33}$$

$\frac{n-1}{n-k-1}$ – поправка на число ступеней свободи, чим $R^2_{(k)}$ більше, тим краще підібрана модель.

Висновки. Вибір форми залежності є дуже важливим. Існує два економетричні підходи.

Економетричні підходи

1 – беруть значущі фактори з точки зору дослідника і будується модель (тобто знаходять усі коефіцієнти). Якщо модель не достатньо описує вхідні дані, додаємо фактори, і процедуру повторюємо.

2 – включаємо усі можливі або вищезгадані максимально пояснюючі змінні та проводимо процедуру оцінки і робимо аналіз “показник – фактор”, незначущі фактори з моделі вилучаємо. Хоча відразу цього робити не варто, а слід розглядати ці фактори з лагом. Тоді вони з незначущих перетворюються у значущі. Або у нелінійних багатомірних моделях включати ці фактори лінійно.

Але, як уже згадувалося, це дуже складний дослідницький підхід, де відіграє роль ситуація, досвід, вірогідність зібраної інформації. Тому підходів дуже багато, і вони різноманітні.

Побудова регресійної моделі

Процес побудови регресійної моделі можна подати у вигляді декількох етапів, на кожному з яких розв’язуються свої задачі.

Етапи побудови регресійної моделі

Етап 1
Визначення мети дослідження.
Збір вихідних даних

Етап 2
Аналіз даних і відбір факторів

Етап 3
Оцінка параметрів багатофакторної лінійної моделі регресії

Етап 4
Перевірка якості моделі

Етап 1. Визначення мети дослідження. Збір початкових даних

Мета ставиться задача дослідити залежність деякої економічної змінної від впливаючих на неї факторів і виразити це в кількісному вигляді. Це можна зробити, побудувавши модель регресії та оцінивши її параметри на основі деякого набору початкових даних про значення аналізованої економічної змінної та відібраних факторів. Аналіз проводитиметься на основі лінійної багатофакторної моделі регресії.

Збір початкових даних

(див. тему 2 “Статистичне спостереження”)

Збір початкових даних

- спосіб отримання даних
- кількість необхідних для побудови моделі даних за кожною змінною (довжина ряду)

А. Вихідні дані, які дослідник збирає з метою проведення аналізу і побудови моделі, можна об’єднати в три основні групи за способом отримання.

1 спосіб. Дані, одержані шляхом випадкової вибірки з деякої n-мірної сукупності випадкових величин.

▷ **ПРИКЛАД.** Нехай є дані про який-небудь банк А, для якого визначений деякий набір кількісних показників. Залежно від деяких умов (рівня активів, М1, М2, М3 і т.д.) ці показники можуть варіювати в тих або інших межах. Величина цих показників впливає на рейтинг банку, що визначається при здійсненні угод, і обсяг обороту. У кожному випадку дані про рейтинг банку, відповідні показники і обсяг обороту реєструються окремо. Нехай є достатньо велика кількість таких реєстраційних звітностей.

Дослідник поставив перед собою задачі:
1) з'ясувати, як впливають значення показників на зміну рейтингу банку;
2) побудувати модель, за допомогою якої можна було б робити деякі попередні розрахунки за рейтингом банку.

Дослідник може проаналізувати всі наявні реєстраційні звітності, а може вибрати, наприклад, *випадковим чином m* кількість звітностей. І в тому, і в іншому випадку він матиме в своєму розпорядженні деяку вибірку.

Рейтинг банку розглядатиметься як випадковий, залежна змінна, показники і обсяг обороту – як випадкові змінні, що є незалежними змінними факторами. ■

2 спосіб. Дані про фактори одержані шляхом проведення "економічного експерименту".

▷ **ПРИКЛАД** аналізу ефективності функціонування банку. Значення показників, що визначаються ефективністю, відомі наперед, задані; ефективність – величина випадкова.

Разом з невипадковими факторами можуть бути використані в тій же моделі і фактори, представлені випадковими величинами. Вимагається, щоб "експеримент" проводився в одних умовах, наприклад, на одному і тому ж ринку. ■

3 спосіб. *Дані для моделі є статистичними часовими рядами економічних показників, вони за своєю природою можуть бути представлені як випадковими величинами, так і заданими. Впорядкованість у часі не дозволяє розглядати як випадкову вибірку часові ряди, представлені випадковими величинами. Крім того, багато часових рядів мають тренд.*

Спосіб подання початкових даних для побудови моделі має істотне значення для методів їх обробки і подальшого використовування.

Далі будуть розглянуті два різні способи аналізу даних і побудова моделі регресії у випадку, якщо всі початкові дані представлені *випадковими величинами, у тому числі і часовими рядами, і у випадку, якщо серед чинників є невипадкові, задані показники.*

У першому випадку процедура регресійного аналізу істотно доповнюється кореляційним аналізом, в другому – використовується один з методів покрокової регресії, який може бути застосований і в першому випадку.

Кількість *необхідних даних* (довжина ряду) пов'язана з розмірністю моделі, числом параметрів моделі. Звичайно вважається, що довжина рядів даних повинна перевищувати число шуканих параметрів у регресійній моделі у *6-8 разів.*

Так, якщо модель регресії з вільним членом має 4 фактори, то число шуканих параметрів дорівнює 5, в цьому випадку довжина рядів даних повинна містити не менше 30 показників.

Підсумок 1 етапу. Відповідно до поставленої мети дослідження на основі якісного аналізу процесу:

1) формується блок початкових даних у вигляді таблиці:

№ пор.	y	x_1	x_2	x_3	…	x_k
1	y_1	x_{11}	x_{12}	x_{13}	…	x_{1k}
2	y_2	x_{21}	x_{22}	x_{23}	…	x_{2k}
…	…	…	…	…	…	…
n	y_n	x_{n1}	x_{n2}	x_{n3}	…	x_{nk}

2) визначається метод відбору і аналізу факторів для побудови моделі.

Етап 2. Аналіз даних і відбір факторів

Мета необхідно визначити набір факторів, які будуть включені в модель регресії.

Три описані вище способи отримання даних визначають методи їх аналізу для відбору факторів, що включаються в модель. Необхідно одразу відзначити, що ця процедура має ітеративний характер, і у міру поглиблення аналізу набір факторів може переглядатися.

1. Якщо всі аналізовані показники є випадковими величинами, одержаними методом випадкової вибірки, в цьому випадку для відбору факторів застосовують методи *кореляційного аналізу*.

2. Якщо випадкові величини представлені часовими рядами, в цьому випадку методи кореляційного аналізу для оцінки ступеня щільності зв'язку застосовуються не до самих даних, а до перетворених.

Аналіз здійснюється таким чином:

1. У часовому ряді для кожного показника визначається тренд
2. Знаходяться відхилення від тренда
3. Кореляційний аналіз застосовується до відхилень від трендів
4. Висновки розповсюджуються на показники

У побудові моделі беруть участь показники у вигляді часових рядів, а не відхилення, якщо це не обмежено спеціальними умовами.

3. При всіх трьох способах отримання даних для відбору факторів можна застосовувати кроковий метод послідовного виключення.

Б. Застосування кореляційного аналізу для відбору факторів.

Дані представлені *випадковими величинами*, що мають нормальний закон розподілу і отримані *методом випадкової вибірки. У цьому випадку лінія регресії є прямою лінією, зв'язок між змінними лінійний, і коефіцієнт кореляції можна використовувати як показник щільності зв'язку між змінними.*

Якщо ж невідомий характер розподілу випадкових величин або якщо перевірити гіпотезу про нормальність не видається можливим, то для правомірності використання в аналізі коефіцієнта кореляції необхідно перевірити гіпотезу про лінійність зв'язку.

Перевірка гіпотези про лінійність зв'язку

1) візуальна оцінка за графіком.

Будується графік з координатними осями $\left(Y, X_i\right)$, $\left(X_i, X_j\right)$ і аналізується поле розсіяння.

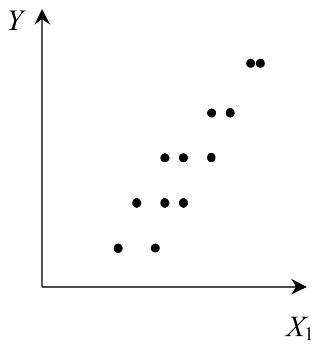


Рис. 6.9

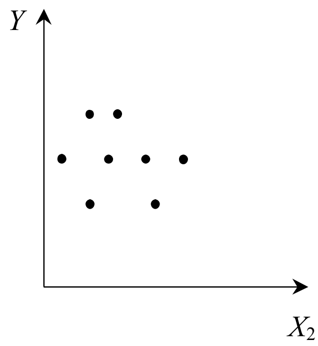


Рис. 6.10

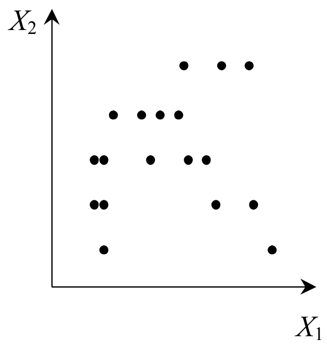


Рис. 6.11

Рис. 6.9 свідчить про наявність позитивного лінійного зв'язку між змінними Y і X_1, рис. 6.10 дає підстави більше говорити про відсутність кореляційного зв'язку, а рис. 6.11 – про наявність нелінійного кореляційного зв'язку.

Графічний аналіз дає деякі підстави для висновку про те, що тільки в першому і в другому випадках про наявність і ступінь щільності кореляційного зв'язку можна буде судити за величиною лінійного коефіцієнта кореляції. Причому в другому випадку він може бути близьким до нуля;

2) *статистична перевірка гіпотези про лінійність зв'язку.*

Розраховується коефіцієнт кореляції r_{xy}. Перевіряється гіпотеза. Якщо гіпотеза відхиляється, то вважається, що зв'язок між змінними y і x лінійний. Перевірка гіпотези здійснюється за допомогою t-статистики (див. далі перевірку гіпотези про значущість лінійного коефіцієнта кореляції).

Застосування кореляційного аналізу припускає такі операції:

- Розрахунок лінійних парних коефіцієнтів кореляції
- Перевірка статистичної значущості парних коефіцієнтів кореляції
- Розрахунок частинних коефіцієнтів кореляції
- Перевірка статистичної значущості частинних коефіцієнтів кореляції

1. *Розрахунок парного коефіцієнта* кореляції між змінними X і Y.

Парний коефіцієнт кореляції

$$r_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{\left(\overline{x^2} - \overline{x}^2\right)\left(\overline{y^2} - \overline{y}^2\right)}}. \qquad (6.34)$$

Якщо кореляційний зв'язок між змінними позитивний, тобто із зростанням значень однієї змінної зростає середнє значення іншої, то коефіцієнт кореляції матиме знак "+". Якщо коефіцієнт кореляції має знак "–", це означає те, що між змінними існує негативний кореляційний зв'язок, тобто із зростанням значень однієї змінної середнє значення іншої зменшується.

Парні коефіцієнти кореляції розраховуються між всіма наявними змінними. Потім з них складається кореляційна матриця

$$K = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ r_{21} & r_{22} & \cdots & r_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ r_{k1} & r_{k2} & \cdots & r_{kk} \end{bmatrix}. \qquad (6.35)$$

оскільки, $r_{ii} = 1в$, $r_{ik} = r_{ki}$, то ця матриця симетрична щодо головної діагоналі

$$K = \begin{bmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{12} & 1 & \dots & r_{2k} \\ . & . & . & . \\ r_{1k} & r_{2k} & \dots & 1 \end{bmatrix} \qquad (6.36)$$

2. *Перевірка статистичної значущості* парних коефіцієнтів кореляції.

Після розрахунку всіх парних вибіркових коефіцієнтів кореляції необхідно перевірити, чи достатня величина для того, щоб зробити статистично обґрунтований висновок про наявність кореляційного зв'язку між кожною парою досліджуваних змінних (зважаючи на ці змінні в цілому в генеральній сукупності).

За наявності початкових припущень про закон розподілу досліджуваних змінних і чисельності вибірки (двомірний закон розподілу змінних у генеральній сукупності вважається нормальним, а обсяг вибірки – будь-яким) перевіряється гіпотеза $H_0 / \rho = 0$.

***t*-критерій Стьюдента**

для перевірки значущості вибіркового коефіцієнта кореляції розраховується *t*-статистика:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \qquad (6.37)$$

що має розподіл Стьюдента із $(n-2)$ ступенями свободи.

Для перевірки нульової гіпотези знаходять за таблицями розподілу Стьюдента при фіксованому рівні значущості та числі ступенів свободи критичне значення $t_{\frac{\alpha}{2}}(n-2)$.

Якщо $\left|t_{расч}\right| \geq t_{\frac{\alpha}{2}}, n-2,$ то нульову гіпотезу про відсутність лінійної залежності між змінними X і Y слід відкинути, як таку, що не узгоджується з наявними даними. Коефіцієнт кореляції статистично значущий. Якщо $\left|t_{расч}\right| < t_{\frac{\alpha}{2}}, n-2,$ то немає підстав відкидати нульову гіпотезу про некорельованості змінних X і Y.

▷ **ПРИКЛАД** розрахунку критерію Стьюдента.

Нехай парний коефіцієнт кореляції між змінними при чисельності вибірки $n = 21$ дорівнює $r_{xy} = 0,8.$ Знак r_{xy} вказує на наявність позитивного кореляційного зв'язку.

Для перевірки статистичної значущості даного коефіцієнта обчислимо *t*-статистику:

$$t_{расч} = \frac{0,849\sqrt{21-2}}{\sqrt{1-0,721}} = 7,02.$$

Табличне значення *t*-статистики при рівні значущості $\alpha = 0{,}05$ і числі ступеней свободи 19 дорівнює $t_{0,025;\,19} = 2{,}09$ $\left|t_{расч}\right| < t_{табл}$, коефіцієнт кореляції значущий з рівнем довірчої імовірності 0,95. ■

Відбір факторів на основі аналізу кореляційної матриці

На основі наведених розрахунків можна вже здійснити відбір факторів для включення в модель регресії як незалежних змінних.

Аналіз кореляційної матриці

1. Насамперед на основі кореляційної матриці відбираються всі статистично значущі коефіцієнти кореляції як між змінними, так і між незалежними змінними $\left(r_{x_i x_j}\right)$

2. Якщо між незалежними змінними тісний кореляційний зв'язок, то виникає явище мультиколінеарності (в подальшому буде показано, чому воно не бажане).
У багатьох практичних дослідженнях вважають, що якщо $\left|r_{x_i x_j}\right| \ge 0{,}8$, то є мультиколінеарність.
У цьому випадку пропонують з двох змінних $x_i u x_j$ $x,$ для яких $\left|r_{x_i x_j}\right| \ge 0{,}8$, вибирати ту змінну, для якої виконується умова $\max\left[r_{yx_i}, r_{yx_j}\right]$. Одна із змінних входить до числа істотних факторів, а друга не включається в модель

3. Далі зі всіх змінних, для яких не виконується умова 2, тобто $\left|r_{xx_{ji}}\right| \le 0{,}8$, відбирати в модель ті змінні, яким відповідають статистично значущі коефіцієнти кореляції

4. Виникає питання: скільки істотних змінних повинно бути включено в модель? Тут треба пригадати про кількісні межі, пов'язані з розмірністю початкових даних

Можуть виникнути спірні випадки: відразу декілька змінних можуть мати достатньо вдалі і приблизно однакові значення коефіцієнтів r_{yx_i}, а всі їх включати може бути вже недоцільно з міркувань розмірності. У цьому випадку можуть надати перевагу тій або іншій змінній з міркувань змістовного характеру або використати додаткову інформацію про взаємозв'язок змінних, який дає аналіз приватних коефіцієнтів кореляції.

Аналіз частинних коефіцієнтів кореляції

Оскільки частинний коефіцієнт кореляції дозволяє оцінити ступінь щільності зв'язку між ознаками без урахування впливу інших факторів на ці ознаки та їх взаємозв'язки, то маючи частинні коефіцієнти кореляції, можна одержати додаткову інформацію та взаємозв'язки досліджуваних показників з метою вибору факторів, що мають бути включенні в модель.

Аналіз частинних коефіцієнтів кореляції припускає рішення таких завдань:

1. Розрахунок частинних коефіцієнтів кореляції.
2. Визначення статистичної значущості приватних коефіцієнтів кореляції.

Частинні коефіцієнти кореляції розрахунок частинних коефіцієнтів кореляції між двома якими-небудь ознаками з безлічі $(k+1)$ ознак розраховується за формулою:

$$r_{12\cdot3\ldots k+1}=-\frac{D_{12}}{\sqrt{D_{11}D_{22}}},\qquad(6.38)$$

де $r_{12\cdot3\ldots(k+1)}$ – частинний коефіцієнт кореляції між ознаками 1 і 2 при фіксованому значенні решти $\left[(k+1)-2\right]$ ознак;

D_{ik} – доповнення алгебри для елемента r_{ik} у визначнику кореляційної матриці k.

Якщо розглядаються тільки три ознаки $y,\ x_1,\ x_2$, то частинні коефіцієнти кореляції легше розраховувати за такою формулою (яка легко виводиться з формули (6.35)).

$$r_{12\cdot3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{\left(1-r_{13}^2\right)\left(1-r_{23}^2\right)}}.\qquad(6.39)$$

Перевірка статистичної значущості частинних коефіцієнтів кореляції відбувається аналогічно парним коефіцієнтам кореляції за *t*-статистикою, значення якої розраховується за формулою:

$$t_{расч}=\frac{r_{част}\sqrt{n-2-p}}{\sqrt{1-r_{част}^2}},\qquad(6.40)$$

де p – число виключених факторів. Для перевірки нульової гіпотези про дорівнювання частинного коефіцієнта кореляції в генеральній сукупності нулю знаходять за таблицями розподілу Стьюдента при фіксованому рівні значущості α і числі ступенів свободи $(n-2-p)$ критичне значення $t_{\frac{\alpha}{2},\, n-2-p}$.

Якщо $\left|t_{расч}\right| \ge t_{\frac{\alpha}{2},\, n-2-p}$, то частинний коефіцієнт кореляції вважається значущим, якщо $\left|t_{расч}\right| < t_{табл}$, то незначущим.

Коефіцієнти частинної кореляції дають додаткову інформацію про ступінь щільності зв'язку між досліджуваними показниками:

1. У разі наявності колінеарності між факторами в модель включають один з факторів, той, у якого частинний коефіцієнт кореляції з результативною ознакою Y більше.
2. Серед факторів, для яких $\left|r_{x_i x_j}\right| < 0,8,$ вибирають для розрахунку в модель насамперед ті, для яких парний і частинний коефіцієнти щільності зв'язку з Y більше.

▷ **ПРИКЛАД.**

Нехай маємо три ознаки y, x_1, x_2. Кореляційна матриця парних коефіцієнтів кореляції:

$$k = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0,849 & 0,91 \\ 0,849 & 1 & 0,78 \\ 0,91 & 0,78 & 1 \end{vmatrix} \qquad (6.41)$$

Частинні коефіцієнти кореляції:

$$r_{12\cdot 3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}} \quad \frac{0,13389}{\sqrt{0,0629}} \quad 0,5. \qquad (6.42)$$

$$r_{13\cdot 2} = \frac{r_{13} - r_{12} r_{32}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{32}^2\right)}} = \frac{0,2512}{\sqrt{0,108}} \quad 0,7.$$

Перевірка статистичної значущості частинних коефіцієнтів показує, що вони значущі.

$$t_{расч(r_{12\cdot 3})} = 2,82. \; t_{расч(r_{13\cdot 2})} = 5,29, \; t_{\frac{\alpha}{2},\, 18} = 2,1.$$

Оскільки факторні ознаки можна вважати колінеарними, в модель слід включити один з факторів, але через те, що стоїть проблема вибору, то з двох факторів x_1 і x_2 треба надати переваги фактора, оскільки у нього і парний, і частинний коефіцієнти більше. ■

Підсумок 2 етапу. *На основі якісного і кількісного аналізу відібрані фактори, які за результатами аналізу передбачається включити в модель регресії.*

Етап 3. Оцінка параметрів багатофакторної лінійної моделі регресії

Мета	знаходження чисельних значень параметрів моделі. Чисельні значення параметрів моделі регресії знаходяться шляхом рішення системи нормальних рівнянь, яка складається для моделі на підставі вимоги методу найменших квадратів.

Згідно з МНК для шуканої моделі регресії має виконуватися вимога:

$$S = \sum \left[y - f\left(x_1, x_2, ... x_k\right) \right]^2 \to \min. \tag{6.43}$$

Якщо $f\left(x_i\right) = a_0 + a_1 x_1 + ... + a_k x_k$, то $S = \sum \left(y - a_0 - \sum_j a_j x_j \right)^2$, де a_0, a_j – шукані параметри, S є функцією від параметрів a_i.

Після знаходження частинних похідних функції S за кожним з параметрів a_i, і прирівнявши їх до нуля (виконавши необхідну умову екстремуму функції), одержуємо систему з $k+1$ рівнянь.

$$\begin{cases} \dfrac{\partial S}{\partial a_0} = \sum 2\left(y - a_0 - a_1 x_1 - ... - a_k x_k\right)\left(-1\right) = 0 \\ \dfrac{\partial S}{\partial a_1} = \sum 2\left(y - a_0 - a_1 x_1 - ... - a_k x_k\right)\left(-x_1\right) = 0 \\ .... \\ \dfrac{\partial S}{\partial a_k} = \sum 2\left(y - a_0 - a_1 x_1 - ... - a_k x_k\right)\left(-x_k\right) = 0 \end{cases} \tag{6.44}$$

Після нескладних перетворень отримуємо систему вигляду:

$$\begin{cases} na_0 + a_1\sum x_1 + a_2\sum x_2 + ... + a_k\sum x_k = \sum y \\ a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_1x_2 + ... + a_k\sum x_1x_k = \sum x_1y \\ .... \\ a_0\sum x_k + a_1\sum x_1x_k + a_2\sum x_2x_k + ... + a_k\sum x_k^2 = \sum x_ky \end{cases} \quad (6.45)$$

У результаті рішення до системи нормальних рівнянь входять чисельні значення параметрів моделі регресії $a_0, a_1, a_2, ... a_k$. Вони є оцінками теоретичної регресії. Знайдені значення параметрів підставляють у модель:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + ... + a_kx_k. \quad (6.46)$$

Підсумок 3 етапу. Побудована модель регресії, що включає ті фактори, які були відібрані на другому етапі. Параметри моделі мають конкретні числові значення. Можна перевірити точність розрахунків, зіставивши підсумки 2 і 3 етапів таким чином. Знаки коефіцієнтів кореляції між y та x_i, і знаки параметрів моделі при факторах x_i повинні бути однаковими.

Етап 4. Перевірка якості моделі.
Перевірка умов Гаусса-Маркова для багатомірної моделі.
Друга умова Гаусса-Маркова.
Гомо- та гетероскедастичність $D(X) \to \min$

Гомоскедастичність – однаковий розкид.

Залишками в регресійній моделі називається послідовний ряд чисел e_i, одержаний як різниця між фактичними значеннями випадкової величини Y і значеннями $\hat{Y}_i$, одержаними на основі моделі регресії з чисельними параметрами шляхом підстановки в рівняння чисельних значень факторіальних ознак, тобто $e_i = y_i - \hat{y}_i$, $i = 1, 2 ... n$.

Графічно залишки можна подати таким чином (рис. 6.12):

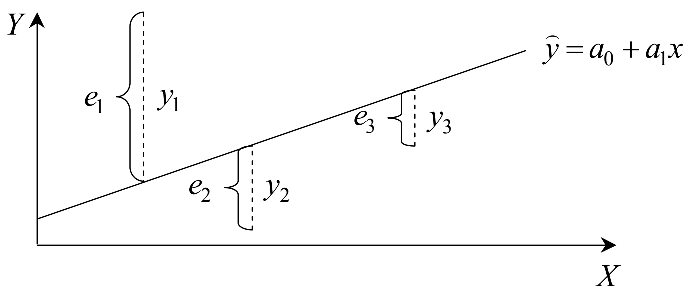


Рис. 6.12. Лінія регресії та залишки

Залишки можна пояснити як ту частину варіації ознаки, яку не можна пояснити за допомогою регресійної моделі. Це та частина варіації, яка пояснюється впливом не тих факторів, які включені в модель, а неврахованих.

Якщо вплив неврахованих факторів на *Y* досить сильний і постійний, то це впливає на залишки, які можуть бути корельовані з величиною *Y*, залишки можуть бути автокорельовані тощо.

Таким чином, за величиною залишків і їх властивістю можна судити про якість моделі, повноту набору включених факторів.

Більше того, від властивостей залишків залежать подальші можливості використання апарату регресійного аналізу.

На подальших етапах регресійного аналізу (для вирішення задач оцінки статистичної надійності коефіцієнтів регресії та оцінки адекватності моделі) використовується апарат імовірності, заснований на припущенні про незалежність нормально розподілених випадкових величин. Як такі випадкові величини розглядаються залишки.

При проведенні регресійного аналізу робляться такі припущення щодо залишків (вище їх вже перераховували):

Припущення щодо залишків

- → приймаються незалежними випадковими величинами
- → мають нульові середні та однакові кінцеві дисперсії
- → підпорядковані нормальному розподілу

Таким чином, якщо обирається модель "правильно", залишки повинні мати тенденцію до підтвердження цих припущень або, щонайменше, не суперечити їм. Після дослідження залишків ми повинні дійти таких висновків:

1) припущення порушені;
2) припущення, мабуть, не порушені.

При цьому треба мати на увазі, що результати пропозиції це означає тільки тоді, якщо на основі даних немає підстав для висновку про неправильність.

Способи дослідження залишків

1. Графічний спосіб дослідження залишків.

▷ **ПРИКЛАД.** Нехай маємо таку послідовність залишків:

№ пор.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
e_i	–2	7	3	–1	–8	1	–2	–3	4	–4	–7	6	8	–2

Проведемо графічний аналіз залишків.

Розглянемо графік часової послідовності залишків $\left(e_i, t_i\right)$. Припустимо, що залишки розташовані саме в такій послідовності в часі, як вони подані в таблиці.

Тоді графік залишків матиме такий вигляд:

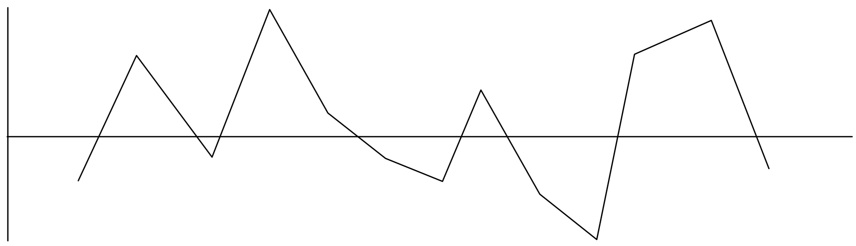


Рис. 6.13. Графік залишків

Якщо за даним графіком виявляється, що всі залишки розташовані в певній "смузі" типу

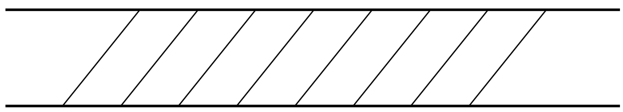


Рис. 6.14. Зображення залишків

то це можна розглядати як ознаку того, що ефект часу не впливає на залишки, що і вони "незалежні" від часу (у них немає тренда), дисперсія однакова. Якщо ж смуга матиме, наприклад, такий вигляд:

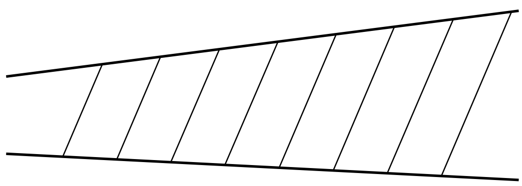
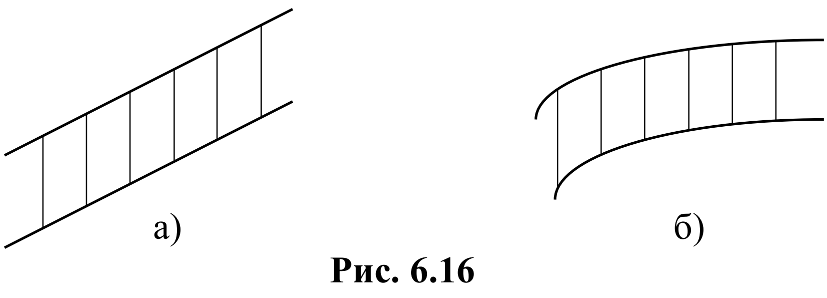


Рис. 6.15

це означає, що дисперсія не постійна в часі. У цьому випадку пропонують використовувати зважений метод найменших квадратів. Якщо "смуга" залишків має вигляд:



а) б)

Рис. 6.16

тут залишки містять “тренд”. У моделі регресії в цих випадках слід ввести параметр t як факторної ознаки, у випадку а) – лінійний член $\left(a_{k+1}t\right)$, у випадку б) – квадратний член $\left(a_{k+1}t+a_{k+2}t^2\right)$.

У нашому випадку “смугу” можна вважати паралельною осі t.

Викиди

Серед залишків іноді зустрічається залишок, який за абсолютною величиною значно перевищує решту залишків. Такі залишки називаються викидами.

Викид показує точку, яка не зовсім типова щодо решти даних. У практичних дослідженнях рекомендують відкидати точку викиду і дане спостереження не розглядати, якщо аналіз викиду дає можливість зазначити, що він викликаний випадковим збігом обставин. Після виключення спостережень, що викликали викид, *знову* аналізують дані і будують модель без цих спостережень.

У будь-якому випадку треба аналізувати точки викиду, оскільки вони можуть дати нову специфічну інформацію, бути викликані незвичайним поєднанням умов. При прогнозуванні цю інформацію треба враховувати.

2. Аналітичні способи дослідження залишків.

Для аналізу залишків можна використовувати деякі аналітичні методи, аналітичні критерії. Вони полягають у тому, що якщо залишки випадкові і незалежні, то у ряді залишків не може бути серій з більшою кількістю плюсів, що йдуть підряд, або мінусів, а власне таких серій з однаковими підряд знаками не повинно бути мало.

Зокрема, при рівні значущості $0{,}05<\gamma<0{,}0975$ кількісний вираз цього правила має вигляд:

$$S(n)\rangle\left[\frac{1}{3}(2n-1)-1{,}96\sqrt{\frac{16n-29}{90}}\right],\ l(n)<l_0(n), \qquad (6.47)$$

де n – число членів у ряду;

$S(n)$ – число серій;

$l(n)$ – максимальна довжина серії.

При цьому, якщо $n\le 26$, то $l_0(n)=5$, якщо $26<n\le 153$, то $l_0(n)=6$. Якщо хоча б одна нерівність порушена, то гіпотези про випадковість відкидають.

Перевірка наявності гетероскедастичності методами критерію Спірмена, тесту Голдфелда-Квандта, тесту Глейсера

Тест рангової кореляції Спірмена

При виконанні тесту рангової кореляції Спірмена передбачається, що дисперсія випадкового члена або збільшуватиметься, або зменшуватиметься у міру збільшення x, і тому в регресії, що оцінюється за допомогою МНК, абсолютні величини залишків і значення x будуть корельовані. Дані за x і залишки упорядковуються, і коефіцієнт рангової кореляції визначається як

$$r_{x,e} = 1 - \frac{6\sum D_i^2}{n\left(n^2 - 1\right)}, \qquad (6.48)$$

де D_i – різниця між рангом x і рангом e.

Якщо припустити, що коефіцієнт кореляції для генеральної сукупності дорівнює нулю, то коефіцієнт рангової кореляції має нормальний розподіл з математичним очікуванням 0 і дисперсією $1/(n-1)$ у великих вибірках. Отже, відповідна тестова статистика дорівнює $r_{x,e}\sqrt{n-1}$, і при використанні двостороннього критерію нульова гіпотеза про відсутність гетероскедастичності буде відхилена при рівні значущості 5 %, якщо вона перевищить 1,96, і при рівні значущості 1 %, якщо вона перевищить 2,58. Якщо в моделі регресії є більше однієї пояснюючої змінної, то перевірка гіпотези може виконуватися з використанням будь-якої з них.

Тест Голдфелда-Квандта

Імовірно, найпопулярнішим формальним критерієм є критерій, запропонований С. Голдфелдом і Р. Квандтом. При проведенні перевірки за цим критерієм передбачається, що стандартне відхилення (σ_i) розподілу імовірностей u_i, пропорційне значенню x у цьому спостереженні. Передбачається також, що випадковий член розподілений нормально і не схильний до автокореляції.

Усі n спостережень у вибірці упорядковуються за величиною x, після чого оцінюються окремі регресії для перших n' для останніх n' спостережень; середні $(n-2n')$ спостережень відкидаються. Якщо припущення щодо природи гетероскедастичності правильне, то дисперсія u в останніх n' спостереженнях буде більшою, ніж в перших n', і це буде відображено в сумі квадратів залишків у двох вказаних “частинних” регресіях. Позначаючи суми квадратів залишків у регресіях для перших k останніх l спостережень відповідно через $\sum_{i=1}^{k} e_i = F1$ і $\sum_{i=k+l}^{m} e_i = F2$, розрахуємо відношення $F2/F1$, яке має F-розподіл з $(n' - k - 1)$ і $(n' - k - 1)$ ступенями свободи, де k – число пояснюючих змінних у регресійному рівнянні. Потужність критерію залежить від вибору n' щодо n. Ґрунтуючись на результатах деяких проведених ними експериментів, С. Голдфелд і Р. Квандт стверджують, що n' повинне бути порядку 11, коли $n = 30$, і порядку 22, коли $n = 60$. Якщо в моделі є більше однієї пояснюючої змінної, то спостереження повинні упорядковуватися за тією з них, яка, як передбачається, пов’язана з σ_i, і n' повинне бути більше ніж $k+1$ (де k – число пояснюючих змінних). Метод Голдфелда-Квандта може також використовуватися для перевірки на гетероскедастичність при припущенні, що σ_i обернено пропорційна до x_i. При цьому використовується та ж процедура, що і описана вище, але тестовою статистикою тепер є показник $F1/F2$, який знов має F-розподіл з $(n'-k-1)$ ступенями свободи.

Тест Глейзера

Тест Глейзера дозволяє більш ретельно розглянути характер гетероскедастичності. Ми відкидаємо припущення про те, що σ_i пропорційна x_i і хочемо перевірити, чи може бути більш відповідною яка-небудь інша функціональна форма, наприклад,

$$\sigma_i = \alpha + \beta x_i^{\gamma}. \tag{6.49}$$

Щоб використовувати даний метод, слід оцінити регресійну залежність y від x за допомогою звичного МНК, а потім обчислити абсолютні величини залишків $|e_i|$ за функцією $\sigma_i = \alpha + \beta x_i^{\gamma}$ для даного значення y. Можна побудувати декілька таких функцій, змінюючи значення y. У кожному випадку нульова гіпотеза про відсутність гетероскедастичності буде відхилена, якщо оцінка β значно відхилюється від нуля. Якщо при оцінюванні більш ніж однієї функції отримуємо значущу оцінку β, то орієнтиром при визначенні характеру гетероскедастичності може служити найкраща з них.

Перевірка якості моделі. Третя умова Гаусса-Маркова щодо залишків. Критерій Дарбіна-Уотсона

Раніше йшла мова про те, що для можливості застосування статистичних критеріїв для оцінки в регресивному аналізі необхідна відсутність автокореляції в залишках.

Найчастіше досліджується випадок для залежності першого порядку: $l_i = \rho_1 l_{i-1} + \beta_i$, де β_i також є випадковою змінною. Звичайно перевіряється справедливість однієї з таких двох альтернативних гіпотез: нульової і альтернативної (про існування автокореляції залишків першого порядку). Перевірку здійснюють за допомогою критерію Дарбіна-Уотсона на основі показника:

$$DW = \frac{\sum_{i=2}^{n} (l_i - l_{i-1})^2}{\sum_{i=1}^{n} l_i^2}, \tag{6.50}$$

значення статистики затабульовані при різних рівнях значущості (0,05; 0,01).

Критерій Дарбіна-Уотсона

Таблиця 6.9

Критерій Дарбіна-Уотсона

5 % рівень значущості

Число спостережень	Число факторів у моделі									
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	–	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96

Спосіб користування таблицею пояснюється на рис. 6.17.

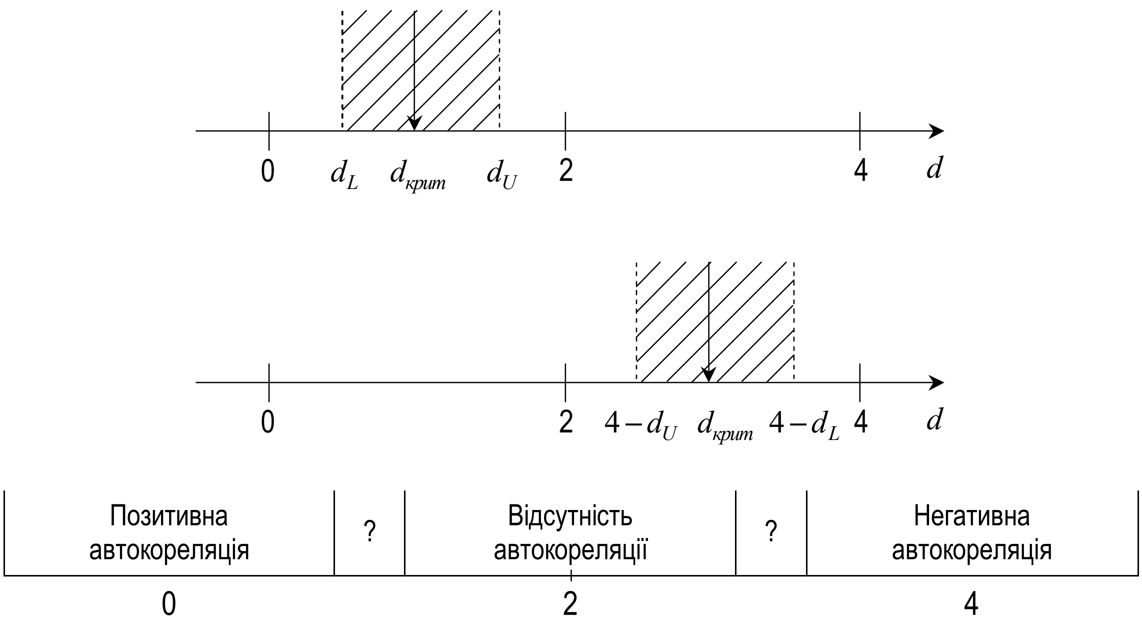


Рис. 6.17. Області перевірки гіпотези за критерієм Дарбіна-Уотсона

За таблицею відповідно до обсягу вибірки n і числа включених у модель факторів k знаходяться значення d_l і d_4. Потім розраховуються значення $(4-d_l)$ і $(4-d_4)$.

Критерій Дарбіна-Уотсона

За даними моделі розраховуються значення показника DW:

$$DW = \frac{\sum_{i=1}^{n}\left(l_i - l_{i-1}\right)^2}{\sum_{i=1}^{n} l_i^2}. \qquad (6.51)$$

Потім розглядається, в який інтервал потрапляє здобуте значення DW. Якщо:

1) $0 \le DW \le d_l$ – в залишках автокореляція позитивна;
2) $4 - d_l < DW \le 4$ – в залишках автокореляція негативна;
3) $d_4 \le DW \le 4 - d_4$ – автокореляція в залишках відсутня;
4) $4 - d_4 < DW \le 4 - d_l,\ d_l < DW < d_4$ – область невизначеності, на основі наявних даних не можна встановити однозначну відповідь.

Причина існування автокореляції може полягати в помилці специфікації, коли вплив деякого істотного фактора або групи факторів виявляється включеним до складу залишку l_i. Отже, дану проблему можна розв'язати, виявивши даний фактор і включивши його в модель. Заздалегідь треба перевірити, чи дійсно існує залежність між цим фактором і перевірити знову автокореляцію залишків.

$\triangleright$ **ПРИКЛАД.** У нашому прикладі $DW = 1{,}98$.

№ пор.	l_i	$l_i - l_{i-1}$	l_i^2	$\left(l_i - l_{i-1}\right)$
1	–2		4	
2	7	9	49	81
3	3	–4	9	16
4	–1	–4	1	16
5	–8	–7	64	49
6	1	9	1	81
7	–2	–3	4	9
8	–3	–1	9	1
9	4	7	16	49
10	–4	–8	16	64
11	7	–3	49	9
12	6	13	36	169
13	8	2	64	4
14	–2	–10	4	100
		326		648

$$DW = \frac{648}{326} = 1{,}98.$$

Число спостережень дорівнює $W = 14$, нехай число факторів у моделі $k = 2$, за таблицею знаходимо d_l і d_4:

$$d_l = 0{,}95;\ d_4 = 1{,}54,\ \text{тобто } 1{,}54 \le DW \le 2{,}46.$$

Дане значення DW свідчить про відсутність автокореляції залишків в аналізованій моделі регресії. ■

Аналіз залишків може привести до таких висновків:

1) залишки задовольняють основним вимогам регресійного аналізу, можна переходити до наступного етапу;
2) залишки не задовільняють основним вимогам регресійного аналізу, необхідно повернутися до дослідження специфікації моделі на 1 і 2 етап.

В. *Перевірка статистичних гіпотез щодо властивостей моделі.*

Після процедури аналізу залишків можна перейти до перевірки статистичних гіпотез щодо властивостей регресійної моделі.

1. Перевірка гіпотези про окремі коефіцієнти регресії.

На основі побудованої за вибірковими даними регресійної моделі можна перевірити гіпотезу про величину коефіцієнта регресії генеральної сукупності. Найчастіше розглядається гіпотеза про рівність коефіцієнта регресії d_i нулю, тобто H_0: $d_i = 0$.

Перевірка цієї гіпотези здійснюється за допомогою t-статистики. На основі даних регресійної моделі розраховується показник

$$t = \frac{a_i}{S_{a_i}}, \tag{6.52}$$

де a_i – оцінка коефіцієнта регресії $\left(i = 0, 1 \ldots k\right)$;

S_{a_i} – оцінка стандартної помилки коефіцієнта регресії a_i в моделі;

$$S_{a_i} = \sqrt{\frac{\sum_{i=1}^{n} l_i}{n-k-1} b_j j},\ i = j - 1, \tag{6.53}$$

де l_i – залишки;

n – число спостережень;

k – число включених у модель факторів;
$b_j j$ – діагональний елемент зворотної матриці системи нормальних рівнянь.

Відношення $t = \dfrac{a_i}{S_{a_i}}$ при виконанні передумов регресійного аналізу (зокрема, передумов щодо залишків), є випадковою величиною, що характеризується t-розподілом. Тому обчислені значення t можна порівняти з табличними значеннями t-розподілу для різних рівнів значущості α і для різних ступенів свободи. Якщо $\left|t_{расч}\right| \geq t_{табл}$, то нульову гіпотезу H_0: $d_i = 0$: при обраному рівні значущості можна відкинути як таку, що не узгоджується з даними спостережень.

Якщо $\left|t_{расч}\right| < t_{тадл}\left(\dfrac{\alpha}{2}, n-k-1\right)$, то гіпотеза приймається.

Якщо встановлено, що коефіцієнт регресії a_i істотно не відрізняється від нуля, то це означає, що відповідний цьому коефіцієнту фактор x_i *не робить статистично значущого внеску в рівняння регресії, при цьому кажуть, що коефіцієнт регресії статистично незначущий. Інакше його називають значущим.*

Перевірку статистичної значущості здійснюють для всіх коефіцієнтів моделі регресії. Якщо ті або інші коефіцієнти моделі статистично незначущі, то відповідні їм фактори можна виключити з моделі, вони істотно не впливають на величину Y.

Г. *Перевірка якості моделі в цілому.*

Для перевірки якості моделі в цілому, оцінки того, наскільки добре дана модель описує фактичні дані (варіацію результативної ознаки Y), використовується дисперсійний аналіз, на основі якого виводиться коефіцієнт детермінації.

Відомо, що загальна варіація ознаки Y $\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$ розкладається на дві складові:

$$\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2 + \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2, \tag{6.54}$$

де $\sum_{i=1}^{n}\left(y - \bar{y}\right)^2$ – загальна варіація ознаки Y, яку треба пояснити за допомогою моделі регресії;

$\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)$ – варіація ознаки Y, що пояснюється регресією;

$\sum_{i=1}^{n}\left(y_i - \hat{y}\right)^2$ – залишкова, нез’ясовна варіація.

Можна наочно подати графічно це співвідношення таким чином:

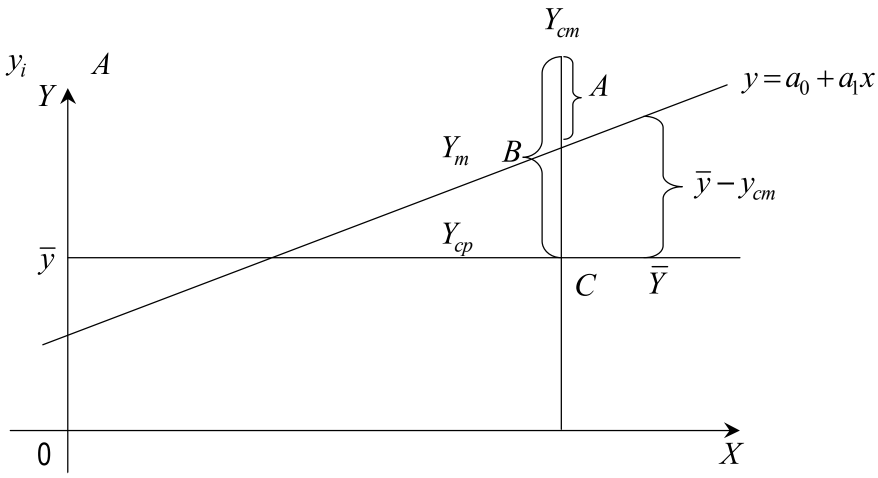


Рис. 6.18. Пояснення до формули (6.51):

- відрізок $AC = \left(y_i - \bar{y}\right)$;
- відрізок $AB = \left(y_i - \hat{y}_i\right)$;
- відрізок $BC = \left(\hat{y}_i - \bar{y}\right)$.

Чим більше частина варіації, що пояснюється регресією, тим краще підібрана модель, або тим менше залишкова варіація. Показник $R^2 = \frac{\sum\left(\hat{y} - \bar{y}\right)^2}{\sum\left(y - \bar{y}\right)^2}$, що показує, якою часткою загальної варіації є пояснена регресією варіація, служить показником якості моделі та називається коефіцієнтом детермінації.

R^2 можна розраховувати і за іншою формулою (6.52):

$$R^2 = \frac{\sum\left(\hat{y} - \bar{y}\right)^2}{\sum\left(y - \bar{y}\right)^2} = \frac{\sum\left(y - \bar{y}\right)^2 - \sum\left(y - \hat{y}\right)^2}{\sum\left(y - \bar{y}\right)^2} = 1 - \frac{\sum\left(y - \hat{y}\right)^2}{\sum\left(y - \bar{y}\right)^2}. \quad (6.55)$$

Статистична значущість коефіцієнта детермінації перевіряється за F-критерієм, розрахованим за даними спостережень за формулою

$F_{расч} = \dfrac{R^2(n-k-1)}{(1-R^2)k}$. Розрахункове значення F порівнюється з табличним за таблицями розподілу Фішера. Якщо $F_{расч} \geq F_{табл}, j, k, n-k-1$, то R^2 статистично значущий і модель адекватна.

Модель з лагом

Автокореляція з лаговою залежною змінною. Маємо випадки, коли X впливає на Y (ми це знаємо). Наприклад, банк знизив процентну ставку за кредитами, але в цей день клієнти не прийшли для отримання кредиту. Вони прийшли через день, тобто існує певний лаг між залежною та незалежною змінними.

Припустимо, що є модель, в якій залежна змінна, узята з лагом в один період, використовується як одна з пояснюючих змінних. У цьому випадку вплив автокореляції, мабуть, зробить оцінки за звичайним МНК неспроможними.

Наприклад, припустимо, що модель має вигляд:

$$y_t = \alpha + \beta_1 x_t + \beta_2 y_{t-1} + u_t \tag{6.56}$$

і припустимо, що випадковий член u_t схильний до дії автокореляції першого порядку:

$$u_t = \rho u_{t-1} + \varepsilon_r. \tag{6.57}$$

Тоді рівняння $y_t = \alpha + \beta_1 x_t + \beta_2 y_{t-1} + u_t$ може бути переписане

$$y_t = \alpha + \beta_1 x_t + \beta_2 y_{t-1} + \rho u_{t-1} + \varepsilon_r. \tag{6.58}$$

Разом з тим y_{t-1} залежить від u_{t-1}, оскільки, якщо співвідношення $y_t = \alpha + \beta_1 x_t + \beta_2 y_{t-1} + u_t$ правильне для t, то воно справедливе і для $(t-1)$:

$$y_{t-1} = \alpha + \beta_1 x_{t-1} + \beta_2 y_{t-2} + u_{t-1}. \tag{6.59}$$

Отже, є систематичний зв'язок між однією з пояснюючих змінних у рівнянні $y_t = \alpha + \beta_1 x_t + \beta_2 y_{t-1} + \rho u_{t-1} + \varepsilon_r$ і першим компонентом випадкового члена. Четверту умову Гаусса-Маркова не задоволено, і оцінки будуть зміщеними навіть у великих вибірках.

Виявлення автокореляції в моделі з лаговою залежною змінною

Як відзначили в своїй первинній статті Дж. Дарбін і Дж. Уотсон, d-статистика Дарбіна-Уотсона непридатна у випадку, коли рівняння регресії включає лагову залежну змінну. У такому разі можна використовувати h-статистику Дарбіна, яка також обчислюється на основі залишків. Вона визначається як

$$h = \hat{\rho}\sqrt{\frac{n}{1-nVar(b)}}, \tag{6.60}$$

де $\hat{\rho}$ – оцінка ρ в автокореляції першого порядку $u_t = \rho u_{t-1} + \varepsilon_r$;

$Var(b)$ – оцінена дисперсія коефіцієнта при лаговій залежній змінній;

n – число спостережень у вибірці.

Приблизна оцінка ρ випливає з виразу $(1-0,5d)$, де d – звичайна статистика Дарбіна-Уотсона і $Var(b)$ – квадрат стандартної помилки b. Тому h можна обчислити на основі звичайних результатів оцінювання регресії.

У великих вибірках h розподіляється як $N(0,1)$, тобто як нормальна змінна із середнім значенням 0 і дисперсією, що дорівнює одиниці за нульовою гіпотезою відсутності автокореляції. Отже, гіпотеза відсутності автокореляції може бути відхилена при рівні значущості 5 %, якщо абсолютне значення h *більше ніж 1,96, і при рівні 1 %, якщо воно більше ніж 2,58, при застосуванні двостороннього критерію і великій вибірці.*

Основна проблема, пов’язана з використанням цього тесту, полягає в неможливості обчислення h в тому випадку, якщо $nVar(b)$ більше одиниці.

Введення фіктивних змінних

Часто трапляється так, що окремі фактори, які ви хотіли б ввести в регресійну модель, є якісними за своєю природою, і отже, не вимірюються в числовій шкалі.

▷ **ПРИКЛАД** введення фіктивних змінних.

1. Досліджується залежність між тривалістю одержаної освіти і доходом. У вибірці представлені особи як чоловічої, так і жіночої статі. Потрібно з’ясувати, чи обумовлює стать відмінність у результатах.
2. Досліджуються фактори, що визначають інфляцію, і в деякі роки періоду спостережень уряд проводив політику регулювання доходів. Потрібно перевірити, чи мало цей який-небудь вплив на досліджувану залежність.

Фіктивна змінна У кожному з цих прикладів одним з можливих рішень було б оцінювання окремих регресій для двох вказаних категорій з подальшим з'ясуванням, чи розрізняються одержані коефіцієнти. Інший можливий підхід до рішення полягає в оцінюванні єдиної регресії з використанням всієї сукупності спостережень і вимірюванням ступеня впливу якісного чинника за допомогою введення так званої *фіктивної змінної*. Другий підхід має дві важливі переваги: по-перше, є простий спосіб перевірки, чи є дія якісного фактора значущою; по-друге, за умови виконання певних припущень регресійні оцінки виявляються більш ефективними.

Ілюстрація використання фіктивної змінної

Ми проілюструємо метод використання фіктивних змінних на прикладі регресійного аналізу основних факторів, що впливають на обсяги кредитів. Візьмемо до уваги факт, що обсяги кредитів у клієнтів банку, які беруть кредити не вперше, більші ніж у клієнтів, що беруть кредит вперше. Хоча найбільша частина дисперсії обсягу кредиту обумовлена доходом, сімейним станом, кредитною історією та іншими факторами.

Як відправну точку візьмемо модель:

$$y = \alpha + \beta x + u, \tag{6.61}$$

де y – обсяг кредиту у грн.;

x – дохід клієнта.

Це тільки відправна точка. Далі ми досліджуватимемо дію якісного фактора: чи брав даний клієнт кредит до цього, чи ні. Це можна змоделювати за допомогою двох рівнянь:

$$\begin{aligned} y &= \alpha + \beta x + u, \\ y &= \alpha' + \beta x + u, \end{aligned} \tag{6.62}$$

де перше рівняння стосується кредитів, що були взяті клієнтами вперше, а друге – всіх інших.

Помітимо, що ці два рівняння записані з одним і тим же коефіцієнтом при x, але з різними вільними членами. Ми припускаємо, що той факт, чи кредит взятий клієнтом вперше, чи ні, впливає на обсяг кредиту.

Еквівалентним способом запису моделі було б зберегти рівняння для клієнтів, що беруть кредит вперше, і записати інше рівняння у вигляді:

$$y = \alpha + \delta + \beta x + u. \tag{6.63}$$

Обсяг кредиту, що не виданий клієнту не вперше (α'), розділений тут на дві складові: обсяг кредиту, отриманого клієнтом вперше (α), і додатковий обсяг кредиту, обумовлений тим, що кредит береться не вперше (δ). Ця модель ілюструється на рис. 6.19. Дві прямі лінії показують залежність між обсягом кредиту і доходом клієнта без урахування випадкового фактора.

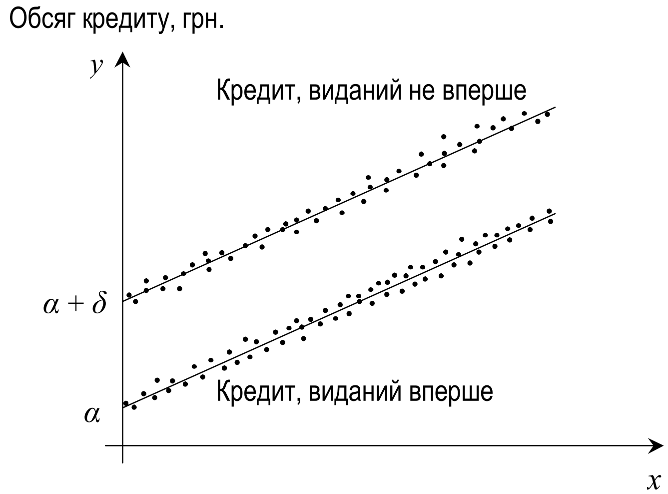


Рис. 6.19. Залежність кредиту, виданого вперше, від кредиту, виданого не вперше

Лінія регресії для кредиту, який видається не вперше, така ж, як для першого кредиту, з тією відмінністю, що вона зсунута вгору на величину δ. Нашою метою є оцінка цього параметра зрушення, і ми одержимо її за допомогою введення так званої фіктивної змінної. Перепишемо модель у вигляді:

$$y = \alpha + D\delta + \beta x + u,$$

де D – фіктивна змінна, тобто штучно введена змінна, яка набуває значення 0, якщо спостереження стосується першого кредиту, і значення 1, якщо воно стосується не першого кредиту.

Ми бачимо, що ситуація визначається тим, що відбувається при D, котре дорівнює нулю або одиниці. Якщо кредит видається клієнту вперше, то D береться рівним нулю і рівняння спрощується до вигляду $y = \alpha + \beta x + u$. Якщо кредит видається не вперше, то D береться рівним одиниці, і рівняння записується у вигляді $y = \alpha + \delta + \beta x + u$. ■

2. Мультиколінеарність.

Явище мультиколінеарності виникає тоді, коли між факторними ознаками існують майже точні лінійні залежності.

Це призводить до того, що матриця системи нормальних рівнянь є погано обумовленою (її визначник буде близьким до нуля). Це відповідно вплине на результати тих розрахунків у регресійному аналізі, які будуть зв'язані з використанням визначника:

1) можуть бути великими за абсолютною величиною оцінки коефіцієнтів регресії в моделі, оскільки при їх розрахунку використовується визначник;
2) більшими будуть середньоквадратичні помилки коефіцієнтів регресії, зокрема за рахунок величини b_{jj} (оскільки при її визначенні використовується визначник – у зворотній матриці), коефіцієнт регресії буде статистично незначущим;
3) величини оцінок коефіцієнтів регресії стають нестійкими, вони істотно змінюються при незначній зміні початкових даних.

Це істотно ускладнює інтерпретацію параметрів моделі. За наявності мультиколінеарності факторів використовують інші методи обробки початкових даних, наприклад, метод головних компонент.

На практиці часто один з мультиколінеарних факторів включають у модель, а інший – ні.

Деякі дослідники пропонують такий прийом: будується регресія між цими факторами, один розглядається як результативна ознака, а інший – як факторна. В основну модель включається не сам фактор (узятий як результативний), а залишки, тобто $\hat{x}_p = b_0 + b_1 x_k$, залишки – $u_i = x_p - \left(b_0 + b_1 x_k\right)$, вони і включаються в загальну модель регресії замість фактора x_p.

Підсумок 4 етапу. *З використанням статистичних критеріїв вироблена оцінка якості моделі. Якщо за якими-небудь критеріями дається незадовільна оцінка, то ставиться питання про перегляд всієї моделі та проведення аналогічного аналізу з новим набором змінних або побудову моделі за іншою формою.*

Способи побудови моделей регресії

Розглянемо коротко деякі способи побудови моделей регресії, що використовують різні статистичні показники якості моделі.

А. *Метод "всіх можливих регресій".*

Нехай є k змінних факторів, які за припущеннями змістовного характеру передбачається включити до моделі регресії.

Метод “всіх можливих регресій” припускає спочатку побудову моделей регресії, в які входить різна кількість змінних у різних поєднаннях.

Наприклад, якщо є 4 змінні фактори: x_1, x_2, x_3, x_4, будуються 4 серії моделей.

Серія 1. Включає всі моделі, в яких міститься один із змінних факторів x_i. Таких моделей буде 4. *Наприклад,*
$\hat{y}_{x_1} = a_0 + a_1 x_1$, $\hat{y}_{x_2} = a'_0 + a'_1 x_2$ і т.д.

Серія 2. Вона включає всі моделі, що містять по дві змінні, таких моделей буде $C_4^2 = 6$

Серія 3. Містить всі моделі, що мають по три фактори, таких моделей буде $C_4^3 = 4$

Серія 4. Містить одну модель, в яку входять всі 4 фактори.
Всього таких наборів буде $\left(2^k - 1\right)$.
Кожну модель перевіряють за критерієм величини коефіцієнта детермінації, скоректованим на число ступенів свободи.

$$\tilde{R}_k^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-k-1},$$

де n – число спостережень;
k – число факторів у даній моделі;
R^2 – нескоректований коефіцієнт детермінації.

Та модель, якій відповідає максимальне значення $\tilde{R}_k^2$, вважається кращою. Якщо майже однакове більше значення коефіцієнта $\tilde{R}_k^2$ забезпечується різними моделями, то може надатися перевага моделі з меншим числом включених факторів, оскільки при прогнозуванні потрібно менше інформації

При великому числі факторів цей метод виявляється дуже громіздким.

Б. *Покрокові процедури формування наборів факторів.*

1. Метод включення (послідовного приєднання)

можна подати у вигляді таких кроків:

1) розраховуються коефіцієнти кореляції між результативною ознакою Y і кожною з факторних. На основі одержаних коефіцієнтів кореляції складається кореляційна матриця k.

Як перша змінна, що включається в модель, вибирається фактор, який має найвищий коефіцієнт кореляції з Y, нехай це буде x_1.

Для моделі $\hat{y}_{(1)} = f(x_1)$ розраховується t-статистика, перевіряється значущість моделі за критеріями R^2 і F;

2) розраховуються частинні коефіцієнти кореляції між всіма змінними факторами, що не увійшли до моделі, та Y при виключенні впливу виділеної змінної x_1: $r_{yx_i - x_1}, i = 1, \dots k-1$. Як наступний фактор, який включається в модель, вибирається факторна змінна, що має найвищий частинний коефіцієнт кореляції (нехай це буде x_2).

Нова модель містить дві змінні $\hat{y}_{(2)} = f(x_1, x_2)$.

Для моделі розраховується t-статистика, коефіцієнт детермінації, перевіряється значущість моделі за F-критерієм;

3) розраховуються частинні коефіцієнти кореляції з Y для всіх факторних змінних, що не увійшли до моделі на попередньому кроці при виключенні впливу вже включених змінних x_1 і x_2, тобто $r_{yx_i x_1 x_2}$. Змінна, що має найбільшу величину частинного коефіцієнта кореляції, включається в модель.

Модель містить вже три фактори. Для одержаної моделі також розраховується t-статистика, перевіряється значущість моделі. Процес продовжується доти, поки включення змінної покращує якість моделі. Одночасно перевіряється вплив включеної змінної на коефіцієнти регресії, оцінюється їх значущість.

Якщо додавання фактора в модель не змінює істотно коефіцієнти регресії в моделі, але збільшує R^2, то він вважається *корисним; якщо додавання фактора в модель радикально змінює всі коефіцієнти регресії, але* R^2 залишається без помітного поліпшення, то фактор вважається *шкідливим*.

2. Метод виключення (послідовного видалення)

розглянемо цей метод для випадку, коли серед змінних факторів є невипадкові величини, в цьому випадку не можна використовувати кореляційний аналіз. Відбір факторів здійснюється на основі статистики. Будується регресійна модель, що включає всі факторні змінні. Для коефіцієнтів регресії розраховується $t_{расч}$, порівнюється з $t_{табл}$ при заданому рівні значущості та числі ступенів свободи $(n-k-1)$. Якщо $t_{расч} < t_{табл}$, то даний фактор з моделі виключається. Якщо відразу для декількох факторів ця умова виконується, то виключається тільки один фактор, для якого $t_{расч} = \min$. Здійснюється перерахунок рівняння регресії з урахуванням змінних, що залишаються, і аналіз моделі повторюється доти, поки в сторонній моделі для всіх включених факторів не виконуватиметься умова $t_{расч} \geq t_{табл} \frac{\alpha}{2}(n-k-1)$. Для всіх моделей розраховується R^2 і F-статистика.

Оцінка придатності моделі

Після того, як модель побудована і здійснена її оцінка за статистичними критеріями, для практичного використання моделі необхідно перевірити її якість за рядом інших ознак.

1. Правдоподібність рівняння – розглядається, чи дійсно змінні фактори, що входять в остаточний варіант моделі, є найважливішими визначальними для формування результативного фактора із змістовної точки зору, точки зору теорії даного явища. Але чи виявилися пропущеними які-небудь з теоретичної точки зору важливі фактори? Якщо пропущені, то чому? Не були включені в початковий список фактори або не було можливості зібрати дані спостережень? Або відсіяні в процесі побудови моделі? Останнє важливе у разі виявлення взаємозалежності, мультиколінеарності факторів, коли одну з взаємозв'язаних змінних включали в модель, а другу – ні.

2. Правдоподібність коефіцієнтів регресії – якщо змінні фактори, включені в рівняння регресії, незалежні, тобто не виявлено явища мультиколінеарності, то коефіцієнти a_i можна використовувати для оцінки "внеску" i-ї змінної у зміну залежної змінної. Якщо ж у рівняння включені взаємозалежні змінні фактори, то така інтерпретація коефіцієнтів регресії не має сенсу.

Далі необхідно перевірити коректність коефіцієнтів регресії з погляду відповідності знаку перед даним коефіцієнтом загальному характеру впливу фактора на залежну змінну.

Наприклад, якщо y – величина доходу від операційної діяльності банку, а x_i – обсяг операцій, то знак коефіцієнта a_i повинен бути позитивним.

3. Стабільність коефіцієнтів регресії – якщо дані для побудови моделі регресії одержані в результаті спостережень, проведених протягом деякого періоду часу, то можна перевірити, чи стійкі коефіцієнти регресії в часі. Перевірку можна здійснити різними способами. Наприклад, розділити весь період спостережень на декілька частин. Для кожної частини даних побудувати рівняння і порівняти коефіцієнти. Якщо коефіцієнти істотно відзрізнятимуться, матимуть певну тенденцію в зміні, то говорити про їх стабільність не можна і використання моделі в колишньому вигляді є проблематичним. Виправити ситуацію можна, використовуючи виявлений "дрейф" коефіцієнтів регресії. У цьому випадку розрахунки по моделі будуються таким чином. Нехай є дані за показниками $y_0, x_1, x_2, \ldots x_m$ за n періодів часу. Вибирається період k від початкової точки, для якого будується рівняння регресії на основі перших k спостережень.

Потім цей період спостереження зміщується на одиницю часу і знову будується рівняння регресії на основі k спостережень з $t = 2$ тощо. Отже, $n - k + 1$ рівнянь і відповідних значень коефіцієнтів регресії a_i.

Кожен коефіцієнт a_i може бути поданий у вигляді функції від часу $a_i(t) = f_i(t)$. Модель регресії матиме вигляд $\hat{y} = a_0(t) = a_1(t)x_1 + a_2(t)x_2 + \ldots + a_m(t)x_m$. При використанні такої моделі для прогнозування спочатку здійснюється прогноз коефіцієнтів моделі на основі рівнянь $a_i(t) = f_i(t)$, потім прогноз значень факторів. Підстановка в рівняння регресії прогнозних значень коефіцієнтів і факторів дає прогнозну оцінку залежності змінних.

Побудова прогнозу на основі моделі регресії

Прогнозування економічних показників на основі моделі регресії по суті є екстраполяцією тих основних властивостей процесу, які описуються моделлю і засновані на аналізі наявних у минулому статистичних даних. При цьому виходять з припущення, що в майбутньому збережеться не тільки загальний напрям розвитку, але і структура, і форма взаємозв'язків.

Побудова прогнозу пердбачає ряд кроків.

1. Визначення значень незалежних змінних (факторів) на період прогнозу $\overline{x}_{пр} = \left(x_1^{пр}, x_2^{пр} \ldots\right)$.
Для їх розрахунку можуть будуватися спеціальні моделі, вони можуть визначатися і експертним шляхом

2. Визначення точкової прогнозної оцінки $\hat{y}_{пр}$ залежної змінної, яка визначається шляхом підстановки в рівняння прогнозних значень змінних факторів

3. Визначення довірчих інтервалів прогнозу $\hat{y}_{пр}$.
Довірчий інтервал прогнозної оцінки визначається за формулою:

$$y = \hat{y}_{пр} \pm t_\alpha S_{пр},$$

де $\hat{y}_{пр}$ – прогнозна точкова оцінка;
t_α – параметр t-розподілів при заданому рівні довірчої імовірності $(1-\alpha)$ і числі ступенів свободи, $(n-k-1)$,

$$S_{пр} = S\sqrt{1 + x_p' a^{-1} x_p},$$

$$S = \sqrt{\frac{\sum\left(y - \hat{y}\right)^2}{n-k-1}}.$$

$$x_p' = \left(1, x_1^{пр}, x_2^{пр}, \ldots\right) \quad x_p = \begin{pmatrix} 1 \\ x_1^{пр} \\ x_2^{пр} \\ \ldots \end{pmatrix}$$

A^{-1} – зворотна матриця матриці системи нормальних рівнянь

Функція Кобба-Дугласа

У 1927 р. Пол Дуглас, економіст за освітою, визначив, що якщо нанести на одну і ту ж діаграм графіки логарифмів показників реального обсягу випуску (Y), капітальних витрат (K) і витрат праці (L), то відстані від точок графіка показників випуску до точок графіків показників витрат праці та капіталу становимуть постійну пропорцію. Потім він звернувся до математика Чарльза Коббу з проханням знайти математичну залежність, що має таку особливість, і Кобб запропонував таку функцію:

$$Y = AK^{\alpha}L^{1-\alpha}. \tag{6.64}$$

Ця функція була запропонована приблизно 30 роками раніше Філіпом Уікстидом, як було вказано Ч. Коббом і П. Дугласом, але вони були першими, хто використовував для її побудови емпіричні дані індексів реального обсягу виробництва капітальних витрат і реальних витрат праці (промисловість США, 1899-1922 рр.).

Автори не описують, яким чином вони насправді підібрали функцію, але імовірно вони використовували початкову форму регресійного аналізу, оскільки посилалися на “теорію найменших квадратів”. За їх оцінкою $\alpha = 1/4$.

Непараметричні методи взаємозв’язків

Методи, за допомогою яких можна виміряти зв’язок між явищами, не використовуючи при цьому кількісні значення ознаки, параметри розподілу – *непараметричні*. Визначають розрахунок коефіцієнтів асоціації, контингенції, взаємної спряженості Чупрова, рангової кореляції Спірмена. Два перші застосовуються для оцінки взаємозв’язку між двома альтернативними ознаками.

Коефіцієнт контингенції

$$K_k = \frac{ad - bc}{\sqrt{(a+d)(d+c)(a+c)(b+d)}}. \tag{6.65}$$

Коефіцієнт асоціації

$$K_a = \frac{ad - bc}{ad + bc}. \tag{6.66}$$

Коефіцієнт контингенції повинний бути менше коефіцієнта асоціації. Зв’язок вважається підтвердженим, якщо $K_k \geq 0{,}3$; $K_a \geq 0{,}5$.

При аналізі таблиць взаємної спряженості, які складаються з 4 клітин, використовується відношення перехресних шансів, що характеризує ступінь відносного ризику фактора *x*, а результат *y*:

$$W = \frac{ad}{bc}. \tag{6.67}$$

Коефіцієнт взаємної спряженості Пірсона і Чупрова застосовується при оцінці зв'язку між явищами, не використовуючи при цьому кількісних значень ознаки.

$$C = \sqrt{\frac{\Phi^2}{1+\Phi^2}};\ K = \sqrt{\frac{\Phi^2}{\sqrt{(K_1-1)(K_2-1)}}}, \qquad (6.68)$$

де Φ^2 – це показник середньої квадратичної спряженості, визначений шляхом вирахування одиниці із суми відносин квадратів частот кожної клітини кореляційної таблиці до добутку частот відповідного стовпця і рядка.

$$\Phi^2 = \sum_{ij} \frac{f_{ij}^2}{f_i f_j} - 1;\ f_i \quad \sum_j \neq f_{ij};\ f_j \quad \sum_i \neq f_{ij}. \qquad (6.69)$$

$ДО_1$ і $ДО_2$ – число груп за кожною з ознак.

Теорія ігор

Метод Монте-Карло

Значне місце в економетрії займає *теорія випадкових процесів,* у розробку якої великий внесок зробив Дж. фон Нейман. У цій сфері широко вживається *метод Монте-Карло* (1949). Він заснований на кібернетичній ідеї "чорного ящика" і полягає у тому, що досліджуваний процес моделюється шляхом багаторазових повторень його випадкових реалізацій. Цей метод призначений для тих випадків, коли побудова аналітичної моделі важка або неможлива, наприклад, при рішенні задач теорії масового обслуговування.

Експеримент за методом Монте-Карло ніхто точно не знає, чому *експеримент за методом Монте-Карло* називається саме так. Можливо ця назва якось стосується відомого казино як символу дії законів випадковості. Основне поняття буде пояснене за допомогою аналогії. Припустимо, що дві свині навчені знаходити трюфелі, – дикорослі земляні гриби, що зустрічаються у Франції та Італії і вважаються делікатесом. Вони дорого коштують, оскільки їх важко знайти, і гарна свиня, навчена пошуку трюфелів, коштує дорого. Проблема полягає в тому, щоб визначити, наскільки добре кожна свиня шукає трюфелі. Вона може знаходити їх час від часу, але можливо також, що велику кількість трюфелів вона пропускає. У разі дійсної зацікавленості ви могли б вибрати ділянку землі, закопати трюфелі в декількох місцях, відпустити свиней і прослідкувати, скільки грибів кожна з них знайде. За допомогою такого контрольованого експерименту можна було б безпосередньо оцінити ступінь успішності пошуку.

Як це стосується регресійного аналізу (модель парної регресії $y = \alpha_0 + \alpha_1 x + u$, рівняння регресії $\hat{y} = a_0 + a_1 x$? Проблема у тому, що ми ніколи не знаємо істинних значень α_0, α_1 (інакше навіщо б ми використовували регресійний аналіз для їх оцінки?). Тому ми не можемо сказати, гарні або погані оцінки дає наш метод. Експеримент за методом

Монте-Карло – це штучний контрольований експеримент, що дає можливість такої перевірки. Простий можливий експеримент за методом Монте-Карло складається з трьох частин, по-перше:

1) вибираються істинні значення α_0, α_1;
2) у кожному спостереженні вибирається значення x;
3) використовується деякий процес генерації випадкових чисел (або береться послідовність з таблиці випадкових чисел) для отримання значень випадкового фактора u в кожному із спостережень.

По-друге, в кожному спостереженні генерується значення y з використанням співвідношення $y = \alpha_0 + \alpha_1 x + u$ і значень α_0, α_1, x, u.

По-третє, застосовується регресійний аналіз для оцінювання параметрів a_0, a_1 з використанням тільки набутих вказаним чином значень y і даних для x. При цьому ви можете бачити, чи є a_0, a_1 гарними оцінками α_0, α_1, і це дозволить відчути придатність методу побудови регресії.

На перших двох етапах проводиться підготовка до застосування регресійного методу. Ми повністю контролюємо модель, яку створюємо, і знаємо істинні значення параметрів, тому що самі їх визначили. На третьому етапі ми визначаємо, чи може поставлена нами задача бути розв'язана за допомогою методу регресії, тобто чи можуть бути одержані гарні оцінки для α_0, α_1 при використовуванні тільки даних про y і x. Відзначимо, що проблема виникає унаслідок включення випадкового фактора в процес отримання y. Якщо б цей фактор був відсутній, то точки, відповідні значенням кожного спостереження, лежали б точно на прямій $y = \alpha_0 + \alpha_1 x + u$ і точні значення α_0, α_1 можна було б дуже просто визначити за значеннями y і x.

Особливості в застосуванні методу Монте-Карло в економічних проблемах	застосування методу статистичних випробувань – методу Монте-Карло для аналізу економічних ситуацій і процесів досить прозоре та привабливе. По-перше, основні економічні проблеми зводяться до моделей математичного програмування (пошук екстремуму функцій багатьох змінних в області допустимих планів), і якщо модель не лінійна, потребують значних затрат часу або на наближення до лінійних моделей, або пошук та застосування чисельного методу рішення. По-друге, будь-яка економічна ситуація, навіть та, що зводиться до лінійних моделей, має в собі елементи невизначеності чи випадковості. Ці обставини приводять до цензу, щоб не заглиблюючись у складні економічні проблеми, використати випадковість проти випадковості.
Алгоритм застосування методу Монте-Карло	генеруються випадкові управляючі змінні, проводиться перевірка всіхобмежень, що складають область допустимих рішень, і якщо всі вони виконуються, шукається значення критерію ефективності або функції цілі, що вносяться до комірок пам'яті. Проводячи

статистичні випробування задане число разів, ми обираємо зі значень функції найбільше чи найменше значення залежно від напрямку екстремуму. Проводиться алгоритм і програма самонавчання, що слідкує за напрямом екстремуму та звужує область пошуку.

Нехай є деяка функція відносно параметрів

$$\varphi\left(x_1, x_2, \ldots, x_n\right), \qquad (6.70)$$

$$x_1 \in X. \qquad (6.71)$$

Область X складається із системи обмежень, що її описують. Важливо знати границі виміру цих невідомих

$$a_i \le x_i \le A_i, \quad i = 1, \ldots, n. \qquad (6.72)$$

Необхідно в області X знайти x^* оптимальну точку, що надає значення функції $\max(\min)$.

Узагальнений алгоритм застосування методу Монте-Карло для вирішення економічних ситуацій і процесів

1. Беруться випадковим чином значення невідомих, і перевіряються з проміжків (1.3) всі системи обмежень (1.2). Якщо хоча б одна нерівність не виконується, ця точка відкидається.

Зауваження. Якщо генеруємо випадкове число з відрізка (0,1), то x змінюється (1.3) $a_i \le x_i \le A_i$, $x = a_i + \left(A_i - a_i\right) \cdot r, \ 0 < r < 1$

2. Перевіряємо одержані x_i на всі обмеження, що входять в постановку задачі. Якщо не проходять всі обмеження, переходимо до пункту 1. Якщо всі обмеження виконуються, переходимо до пункту 3

3. Допустиме рішення x_i підставляємо у функцію φ і знаходимо φ_i

4. Перевіряємо, чи це є першим рішенням, чи ні. Якщо перше рішення, його запам'ятовують, не тільки φ_i, а і (φ_i, x_i). Якщо рішення не перше, то одержане значення функції звіряємо з тим значенням, що маємо на даний момент. Якщо краще того, що було, то його запам'ятовують замість старого. Якщо ні – відкидають і переходять до пункту 1

Процес закінчується або при заданому числі точок іспитів n, або машинному часі випробувань.

Цікавим є застосування методу Монте-Карло в економічних задачах теорії ігор. Пошук рішення в мішаних стратегіях зводить ситуацію до подібної проблеми. Класичні або традиційні методи розв’язання задач теорії ігор – це зведення задачі теорії ігор до задачі лінійного програмування або застосування методів наближень. Хоча математична модель задачі для двох гравців у мішаних стратегіях може бути представлена як така, що не містить в собі наявної функції мети:

$$\begin{gathered}
a_{11}p_1 + a_{21}p_2 + \ldots + a_{m1}p_m \ge \upsilon, \\
a_{12}p_1 + a_{22}p_2 + \ldots + a_{m2}p_m \ge \upsilon, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
a_{1n}p_1 + a_{2n}p_2 + \ldots + a_{mn}p_m \ge \upsilon, \\
\sum_{i=1}^{m} p_i = 1,
\end{gathered} \tag{6.73}$$

де $A = \left(a_{ij}\right)$ – платіжна матриця гри;

p_i – ймовірність застосування гравцем відповідних чистих стратегій;

υ – цілі гри.

Метод Монте-Карло можна застосувати безпосередньо до моделі (6.70), знаючи верхню та нижню ціну гри $a \le \upsilon \le \beta$.

Основні поняття теорії ігор

Одна з характерних рис будь-якого суспільного, соціально-економічного явища полягає в множинності, різнобічності інтересів і у наявності сторін, що виражають ці інтереси. Класичними прикладами тут є подані нижче ситуації.

Більш складні подібні ситуації виникають, якщо є об’єднання або коаліції осіб, що беруть участь у зіткненні інтересів.

Будь-яка адекватна математична модель соціально-економічного явища повинна відображати властиві різноманітні інтереси сторін, тобто описувати.

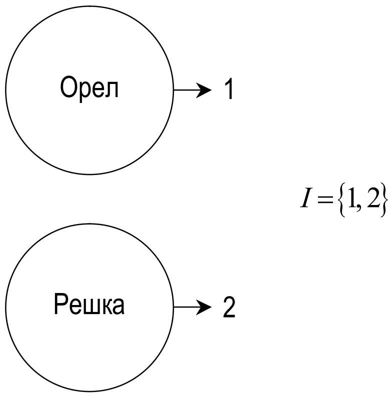
Математична модель соціально-економічного явища

- а) безліч зацікавлених сторін (ми називатимемо їх *гравцями*)
- б) можливі дії кожної із сторін, названі також *стратегіями,* або *ходами*
- в) інтереси сторін, представлені *функціями виграшу* (*платежу*) для кожного з гравців

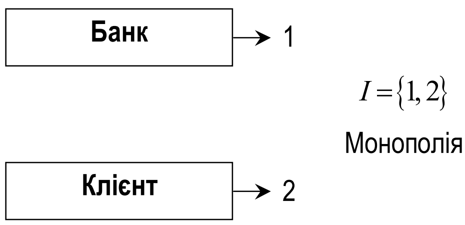
Гра формалізація змістовного опису конфлікту, що відображає різноманітні інтереси сторін, є його математичною моделлю, яку називають *грою*.

Теорія ігор вперше була систематично викладена Дж. фон Нейманом і О. Моргинштерном у 1944 р., хоча окремі результати були опубліковані ще в 20-х роках XX ст. Нейман і Моргинштерн написали оригінальну книгу, яка містила головним чином економічні приклади, оскільки економічному конфлікту найлегше надати чисельної форми. Під час Другої світової війни і відразу після неї теорією ігор серйозно зацікавилися військові, які побачили в ній апарат для дослідження стратегічних рішень. Потім велика увага знову стала приділятися економічним проблемам. Зараз ведеться велика робота, направлена на розширення сфери застосування теорії ігор.

Формальне представлення ігор безліч всіх гравців, що позначається I, у разі кінцевого їх числа може задаватися простим переліком гравців. Наприклад, $I = \{1, 2\}$ при грі в орлянку,



$I = \{\text{Банк, Клієнт}\}$ – у ситуації монополії,



$I = \{1, 2 \ldots n\}$ – у разі аналізу результатів голосування в Парламенті. Безліч стратегій гравця і позначимо через X_i. При грі в орлянку кожен гравець має в своєму розпорядженні дві стратегії: $X_i = \{\text{Орел, Решка}\}$; кожен учасник голосування має вибір на безліч стратегій $\{\text{За, Проти}\}$. У разі взаємодії на ринку як Банк, так і Клієнт можуть призначати деяку не негативну процентну ставку, тобто безліч стратегій кожного з них $X_i : P_i > 0$.

У кожній партії гравець вибирає деяку свою стратегію $x_i \ X_i$, внаслідок чого складається набір стратегій $x = \{x_1, x_2, ..., x_n\}$, що називається *ситуацією*. Так, ситуацію в Парламенті описує список {За, За, Проти, За...}, одержаний у результаті проведеного голосування.

Виграш — зацікавленість гравців у ситуаціях виявляється у тому, що кожному гравцю i у кожній ситуації x приписується число, що виражає ступінь задоволення його інтересів у даній ситуації. Це число називається *виграшем* гравця i і позначається через $h_i(x)$.

Функція виграшу — відповідність між набором ситуацій і виграшем гравця i називається *функцією виграшу (платіжною функцією)* цього гравця H_i.

Матриця виграшів — у разі кінцевої гри двох осіб функції виграшу кожного з гравців зручно подавати у вигляді *матриці виграшів*, де рядки – стратегії одного гравця, стовпці – стратегії іншого гравця, а в клітинах матриці вказуються виграші кожного з гравців в кожній із ситуацій, що утворюються.

▷ **ПРИКЛАД** гри в орлянку.

Наприклад, у разі гри в орлянку кожний з гравців має по дві стратегії, іменовані Орел і Решка. Якщо гравці вибирають однакові стратегії, тобто у випадках, якщо обидва говорять "Орел" або обидва говорять "Решка", 1-й гравець виграє 1 рубль, а другий гравець програє 1 рубль. У ситуаціях, коли обидва гравці вибирають різні стратегії, 1-й гравець програє 1 рубль, а 2-й гравець відповідно цей 1 рубль виграє.

У результаті матриця виграшів 1-го гравця H_1 виглядає таким чином:

		Стратегії 2-го гравця	
		Орел	Решка
Стратегії 1-го гравця	Орел	1	–1
	Решка	–1	1

Відповідно матриця виграшів 2-го гравця H_2 має вигляд:

		Стратегії 2-го гравця	
		орел	Решка
Стратегії 1-го гравця	Орел	–1	1
	Решка	1	–1

■

Дилема ув'язненого — зміст гри такий: два злочинці чекають вироку суду за скоєний злочин. Адвокат конфіденційно пропонує кожному із злочинців полегшити його долю (і навіть звільнити!), якщо він зізнається і дасть свідчення проти спільника, якому загрожує потрапити у в'язницю за скоєний злочин на 10 років. Якщо жоден з них не зізнається, то обом загрожує вирок на певний термін (скажімо, 1 рік) за звинуваченням у незначному злочині. Якщо зізнаються обидва злочинці, то з урахуванням щиросердного зізнання їм обом загрожує потрапити у в'язницю на 5 років. Кожен ув'язнений має на вибір 2 стратегії: не зізнаватися або зізнаватися, виказавши при цьому спільника. У результаті можна одержати таку матрицю "виграшів" для обох гравців:

		Стратегія 2-го гравця	
		зізнався	не зізнався
Стратегія 1-го гравця	зізнався	(5,5)	(0,10)
	не зізнався	(10,0)	(1,1)

З безлічі своїх змішаних стратегій Гравець 1, який прагне досягти найбільшого з гарантованих виграшів, вибирає вектор імовірності x так, щоб одержати максимум мінімальних значень очікуваних виграшів, тобто вирішує задачу

$$v_1 = \max_x \min_y xHy'. \tag{6.74}$$

Аналогічно метою Гравця 2 є досягнення мінімуму максимальних значень своїх програшів, тобто він вирішує задачу

$$v_1 = \min_y \max_x xHy'. \tag{6.75}$$

Фундаментальним результатом теорії ігор є так звана теорема про мінімакс.

Теорема про мінімакс сформульовані задачі для Гравця 1 і Гравця 2 завжди мають рішення для будь-якої матриці виграшів H і, крім того,

$$v_1 = v_2 = v.$$

Як і для цілком визначених ігор, стратегія x^* Гравця 1 називається *максимінною стратегією*, стратегія Гравця 2 y^* – *мінімаксною стратегією*, значення v – *ціною гри*; у разі, коли $v = 0,$ гра називається *справедливою.*

Кооперативна гра гра з ненульовою сумою, в якій гравцям дозволяється обговорювати перед грою свої стратегії та домовлятися про сумісні дії, тобто гравці можуть утворювати коаліції.

Один з підходів до рішення некооперативних ігор полягає у визначенні *точок рівноваги* гри.

Точка рівноваги за Нешем у загальному випадку пари стратегій X, Y для Гравця 1 і Гравця 2 називається точкою рівноваги за Нешем, якщо жодному з гравців невигідно відхилятися від своєї стратегії поодинці, тобто, якщо $H_1\left(X, Y^*\right) \le H_1\left(X^*, Y^*\right)$ – для будь-яких X і

$$H_2\left(X^*, Y\right) \le H_2\left(X^*, Y^*\right) \text{ – для будь-яких } Y.$$

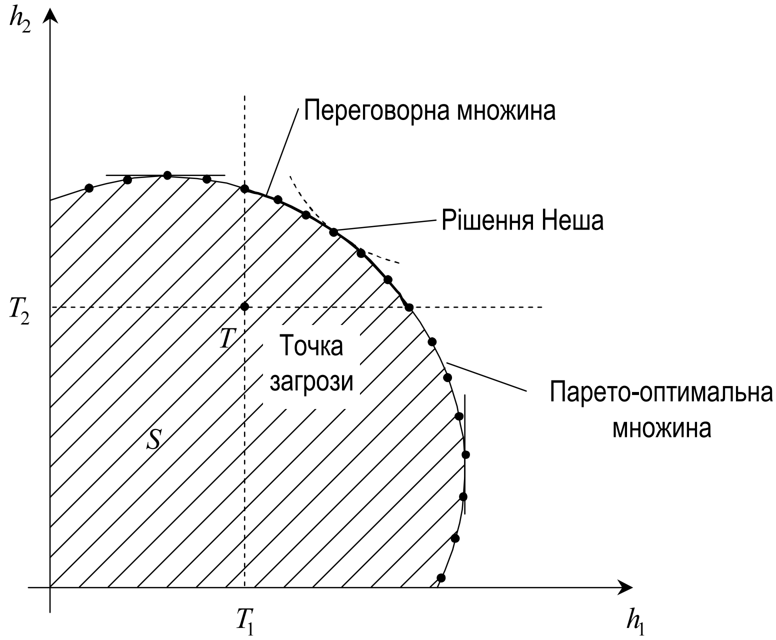


Рис. 6.20. Безліч можливих виграшів, Парето-оптимальна множина, точка загрози, переговорна множина і рішення Неша в кооперативній грі 2 осіб

Точка загрози у разі гри двох осіб передбачається, що два гравці не можуть впливати один на одного, поки не дійдуть деякої угоди. Таким чином, гра визначається як множина S у просторі змінних h_1 і h_2, являючи собою загальні виграші; крім того, задані два числа T_1, T_2 – визначальні величини виграшу, які кожний з гравців може одержати, не вступаючи в коаліцію зі своїм партнером. Звичайно припускають, що множина S є замкнутою, опуклою і обмеженою зверху. Точка T з координатами (T_1, T_2) називається *точкою загрози* (рис. 6.20).

Парето-оптимальні рішення на безлічі можливих виграшів виділяється безліч Парето-оптимальних рішень, тобто безліч точок, що належать S, для яких збільшення виграшу одного з гравців можливе тільки за рахунок зменшення виграшу його партнера. Очевидно, безліч таких точок утворює північно-східна межа множини S.

Переговорна множина	всі точки Парето-оптимальної множини, що знаходяться одночасно вище і правіше за точку загрози T, утворюють так звану *переговорну множину*. Очевидно, що гравцям немає сенсу домовлятися щодо рішень, які не належать переговорній множині, або тому, що положення одного з гравців може бути поліпшене при збереженні положення його партнера і можна домовлятися про вигідніші рішення, або тому, що принаймні для одного з гравців втрачає сенс вступати в коаліцію з своїм партнером – не гірших результатів він може досягти і один.
Точка рішення Неша	отже, на переговорній множині виділяється *точка рішення Неша* N, в якій досягається максимум перевищення виграшів кожного з гравців над платежами, які можуть бути одержані без вступу до коаліції: $$\max\left(h_1 - T_1\right)\left(h_2 - T_2\right).$$

▷ **ПРИКЛАД** сімейної суперечки.

Згідно з умовами цієї гри сімейна пара – Чоловік і Дружина кожен вечір вирішують проблему: як їм провести своє дозвілля. У містечку, де вони живуть, є два види розваг: Балет і Футбол. У кожного з подружжя є своє улюблене видовище: Дружина надає перевагу Балету, Чоловік – Футболу. Проте подружжя настільки прихильне один до одного, що відвідини улюбленої розваги порізно приносять їм зовсім не таке задоволення, як присутність на них удвох, тобто якщо Дружина йде увечері на Балет з Чоловіком, вона одержує максимум задоволення (скажімо, 4 одиниці); Чоловіку не подобається Балет, але присутність на ньому з Дружиною скрашує обтяжливе проведення часу (Чоловік одержує 1 од. задоволення). Історія повторюється, але навпаки, коли Дружина йде з Чоловіком на Футбол, Чоловік одержує 4 од. задоволення від гри улюбленої команди і присутності коханої Дружини; Дружина одержує 1 од. задоволення, провівши вечір з Чоловіком на Футболі. Насправді Чоловік може сходити на Футбол, а Дружина – на Балет поодинці, але відсутність чоловіка знижує задоволення від улюблених видовищ – кожний з них одержує по 2 од. задоволення. І, нарешті, вечір буде проведений зовсім без користі (тобто подружжя одержить по 0 од. задоволення), якщо Чоловік переглядатиме Балет, у той час як Дружина на стадіоні дивитиметься Футбол.

У результаті матриця виграшів описаної гри виглядає таким чином:

		Жінка	
		Балет	Футбол
Чоловік	Балет	(4,1)	(0,0)
	Футбол	(2,2)	(1,4)

Можна показати, що якщо подружжя дотримуватиметься різних неузгоджено змішаних стратегій, безліч можливих виграшів утворює в системі координат значень виграшів подружжя h_1, h_2 трикутник *ABC* з вершинами в точках $(0,\ 0)$, $(1,\ 4)$, $(4,\ 1)$ (рис. 6.21).

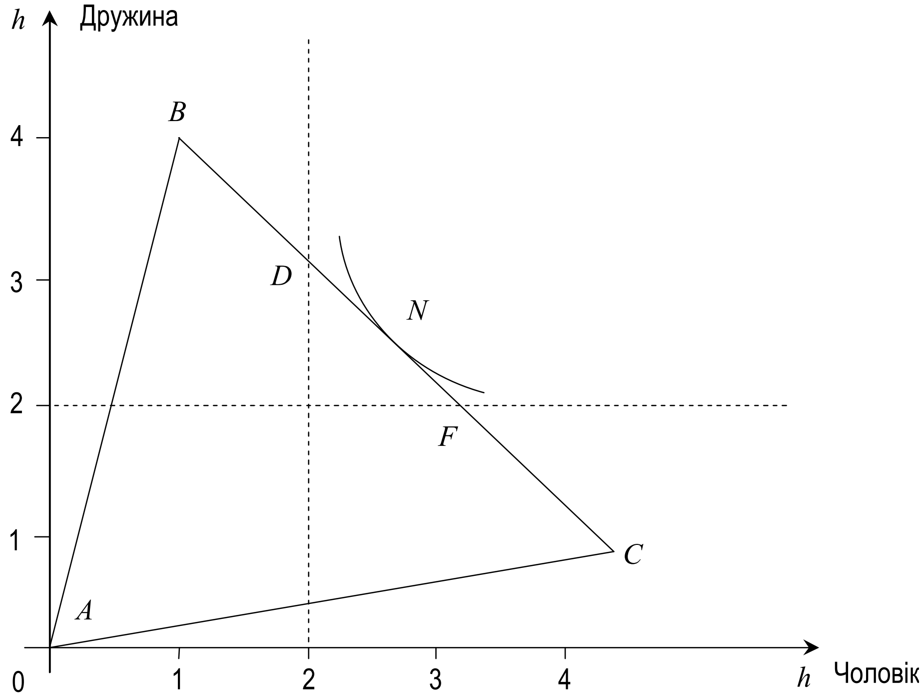


Рис. 6.21. Рішення кооперативної гри "Сімейна суперечка"

Лінія *BC* є безліччю Парето-оптимальних рішень: уздовж цієї лінії зростання задоволення, одержане Дружиною, можливе тільки за рахунок зниження задоволення Чоловіка. Точка *T* з координатами (2, 2) є точкою загрози в цій грі, а "загроза", наприклад з боку Дружини, може звучати буквально таким чином: "Замість того, щоб більше 2/3 свого вільного часу згаяти на Футболі, я ходитиму на Балет (з Чоловіком або без нього – не важливо) – нічого не втрачу". Аналогічно може звучати "загроза" Чоловіка.

У результаті переговорна множина, утворювана точкою загрози *T*, представлена лінією *DE* на Парето-оптимальній безлічі рішень *BC* (див. рис. 6.21). На лінії *DE* Чоловік і Дружина можуть домовлятися, як часто вони буватимуть разом на одному з видовищ; але при цьому, щоб уникнути взаємних загроз, жодній з розваг вони не повинні приділяти більше своїх вільних вечорів.

Рішення Неша, коли максимальний добуток приростів задоволення Чоловіка і Дружини в порівнянні із задоволенням від незалежних відвідин Футболу і Балету, представлене точкою N на рис. 6.21 – подружжя домовляється половину свого вільного часу проводити разом на Балеті, іншу половину – на Футболі. ■

Застосування апарату теорії ігор для аналізу проблем мікроекономіки

Ящик Еджворта в ящику Еджворта (рис. 6.22) довжина горизонтальної осі, відповідної першому товару, дорівнює загальній кількості цього товару X_1, довжина вертикальної осі – загальній кількості товару X_2. Виділений простір є безліччю всіх можливих розподілів наявних товарів між двома гравцями. Нижній лівий кут вважається початком координат для 1-го Гравця, верхній правий кут – початком координат для 2-го Гравця. На виділеному просторі представлені також безліч кривих байдужості (ліній рівня функцій виграшу), що належать кожному з гравців. При цьому точка початкового розподілу товарів має координати $\left(X^1{}_1, X^1{}_2\right)$ у системі відліку 1-го Гравця (і, відповідно, $\left(X^2{}_1, X^2{}_2\right)$ у системі відліку 2-го Гравця).

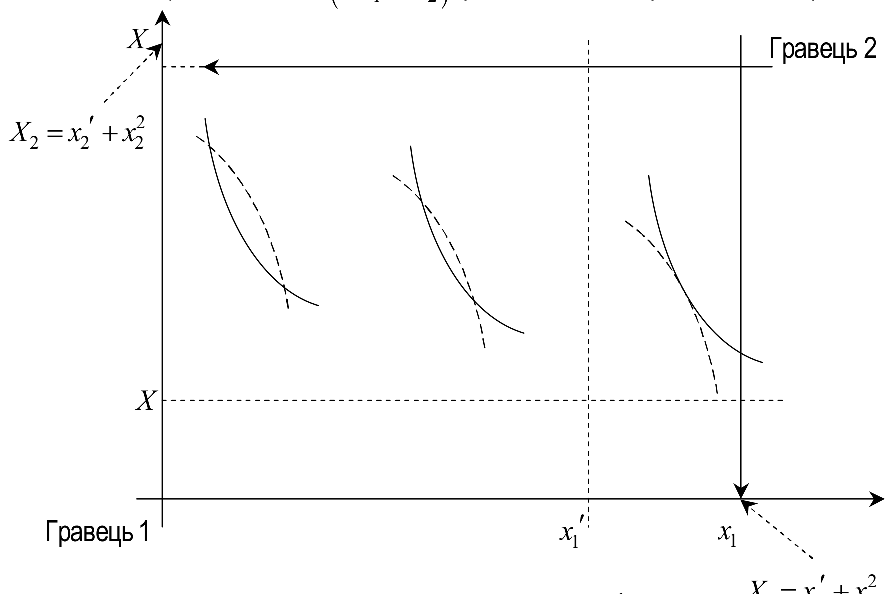


Рис. 6.22. Ящик Еджворта[1]

Розглянемо спершу проблему ефективного розподілу товарів між гравцями. Єдиною вимогою до розподілу, яку ми можемо поставити на початковому етапі аналізу, є вимога Парето-оптимальності. Нагадаємо, що розподіл називається Парето-оптимальним, якщо положення жодного з гравців не можна покращити, не погіршуючи при цьому становища його партнера.

Безліч Парето-оптимальних розподілів може бути наочно представлена за допомогою ящика Еджворта. У разі участі 2-х гравців Парето-оптимальне рішення може бути знайдене за допомогою фіксації рівня корисності одного з гравців (скажімо, Гравця 2) і пошуку максимуму функції корисності іншого гравця.

У термінах ящика Еджворта це означає, що необхідно знайти таку точку на фіксованій кривій байдужості Гравця 2, в якій Гравець 1 одержує максимум своєї функції корисності. Очевидно, що такою точкою є точка, де криві байдужості торкаються одна однієї, оскільки інакше

[1] Тут і далі суцільними лініями позначені криві байдужості 1-го Гравця, пунктирними – криві байдужості 2-го Гравця.

Гравець 1 може, просуваючись уздовж фіксованої лінії рівня Гравця 2 всередину, збільшити значення своєї функції корисності (рис. 6.23). Можна показати, що безліч Парето-оптимальних розподілів у ящику Еджворта буде безліччю всіх точок, в яких криві байдужості Гравця 1 і Гравця 2 торкаються одна одної (рис. 6.24).

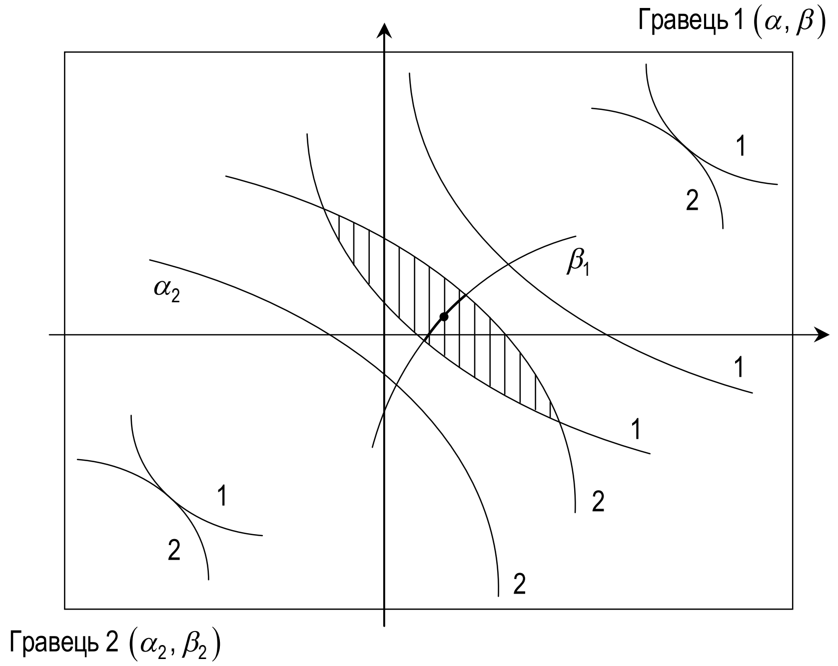


Рис. 6.23. Парето-оптимальний розподіл і ящик Еджворта

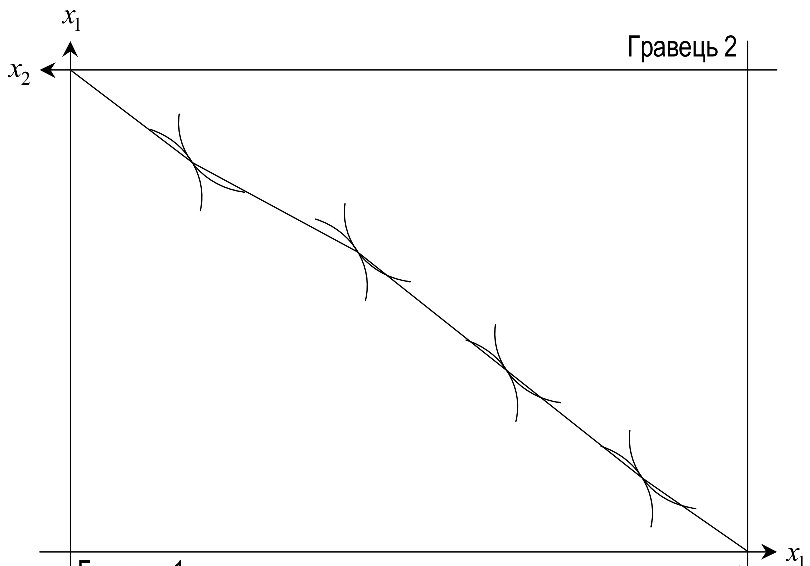


Рис. 6.24. Безліч Парето-оптимальних розподілів (контрактна множина) для ящика Еджворта

Безліч Парето-оптимальних розподілів у просторі товарів називається *контрактною* множиною, оскільки гравцям у загальному випадку має сенс домовлятися між собою саме на цьому наборі ефективних розподілів.

Полн. собр. соч. – Т.4. – С. 52.

Широкое применение в социальных науках находит раздел МСА – *многомерное шкалирование* (для экономного описания данных – понижения размерности, их модельного представления). Развитие этого метода связано прежде всего с именами американских ученых Р.В. Хемминга, Л. Гуттмана, Л. Терстоуна.

Для решения задачи разделения совокупности наблюдений, когда требуется приписать некоторый новый объект к той или иной совокупности на основе "обучающей выборки", разработан *метод дискриминантного анализа.*

Развитие МСА (многомерного статистического анализа) во многом способствовало созданию новой науки – *эконометрии.* Ее предметом является изучение количественной стороны экономических явлений и процессов средствами математического и статистического анализа. Термин "эконометрия" был предложен в 1910 г. польским ученым Павлом Цьомпой, а введен в науку норвежцем Рагнаром Фришем. В 1930 г. было создано Международное эконометрическое общество с центром в Йельском университете (США), в 1933 г. стал издаваться журнал "Эконометрика". Хотя эконометрия как особая дисциплина появилась сравнительно недавно, корни ее уходят глубоко в историю математической формализации экономики и тех отделов математической статистики, которые применяются в анализе экономических данных. Так как первые попытки измерения в экономике принадлежат политическим арифметикам, то можно назвать эконометрию политической арифметикой XX в. Одно из ведущих направлений эконометрии – *построение эконометрических моделей*, задача которых состоит в проверке экономических теорий на фактическом (эмпирическом) материале при помощи методов математической статистики.

Общий метод для оценивания коэффициентов нерекурсивных систем предложен голландским эконометриком Генри Тейлом – *двухшаговый метод 5 наименьших квадратов* (1950). Дальнейшее развитие этот метод получил при разработке А. Цельнером и Г. Тейлом *трехшагового метода наименьших квадратов* (1957).

Плошко Б.Г., Елисеева И.И. История статистики: учеб. пособие. – М.: Финансы и статистика, 1990. – С. 184–185.

После Первой мировой войны развернулись работы по прогнозу динамики отдельных экономических показателей и построению *экономического барометра*, основанного на том, что в динамике различных сторон экономики существуют такие показатели, которые в своих изменениях опережают другие с определенным лагом, а потому могут служить предвестниками их изменений. При этом не исследовалась политэкономическая сущность закономерностей, а изучалась лишь связь и последовательность динамики нескольких показателей.

Первый конъюнктурный барометр построили в начале XX в. американские статистики – Комитет экономических исследований при Гарвардском университете под руководством У.М. Персонса и У.К. Митчелла.

Последователи Персонса и Митчелла продолжают основываться на чистом эмпиризме, полагая, что только индукция без обобщений, без теорий – залог объективного изучения экономических процессов.

Учеником Митчелла и приверженцем эмпирико-статистического подхода является Милтон Фридмен (1912) – глава монетарного направления в современной западной экономике, лауреат Нобелевской премии. *Монетарная теория* основана на идее достижения равенства между массой наличных денег, скоростью их обращения, суммой банковских депозитов и скоростью их движения с одной стороны, и товарной массой – с другой.

Вестник статистики. – 1919. – № 8–12.

Работы А.А. Чупрова “Основные задачи стохастической теории статистики” и “Основные проблемы теории корреляции”.

Чаянов писал: “...нужно, чтобы они чувствовали, знали, сжились с тем, что дело крестьянской кооперации – их крестьянское дело, чтобы дело это тоже было действительно мощным социальным движением, а не предприятием только”.

Статистический метод в научном исследовании: сб. статей. – М., 1925. – С. 196.

Советской статистикой были восприняты основные положения материалистической диалектики. Философская интерпретация категорий статистики была впервые дана в статьях М.Н. Смит: “Сравнительная роль статистического и аналитического метода в научном исследовании” и “Экономика и статистика”.

Смит писала: “Экономист, изучающий эволюцию хозяйственной жизни, все время имеет дело не с устойчивостью, а с изменяемостью больших масс”.

В годы войны применялся особый способ получения статистических данных – *срочные переписи.*

Вестник статистики. – 1950. – № 2. – С. 18.

Сессия ВАСХНИЛ, состоявшаяся в августе 1948 г. и объявившая лженаучными исследования механизма наследственности, имеющего вероятностный характер, искореняла статистический метод из биологической науки и объявляла генетику лженаукой. Выдвинутый сессией тезис “наука – враг случайностей” стал тормозом исследований тех закономерностей природы и общества, которые постигаются при статистическом изучении массовых явлений.

Немчинов В.С. От редактора: Ученые записки по статистике АН СССР. – М., 1955. – Т. 1. – С. 5.

По инициативе В.С. Немчинова и при активном участии Т.В. Рябушкина, Г.И. Бакланова, Л.Е. Минца, Ф.Д. Лифшица было организовано новое издание – *“Ученые записки по статистике АН СССР”* (Т. I, 1955), в котором освещались самые острые проблемы статистики, развивалась свободная творческая дискуссия по определению статистической закономерности, значению закона больших чисел, индексному методу анализа экономических явлений, балансовому методу, корреляционному анализу, нормативной статистике, измерению производительности труда и т.д.

Развитию математико-статистических методов способствовало возобновление изданий работ зарубежных авторов (после почти двадцатилетнего перерыва). В 1958 г. вышли знаменитая книга Р.А. Фишера “Статистические методы для исследователей”, учебник Ф. Миллса “Статистические методы”, в 1960 – фундаментальный курс Дж. Эдни Юла и Мориса Дж. Кендэла “Теория статистики”.

Яккока Л. Карьера менеджера / пер. с англ.; при участии У. Новака; общ. ред. и вступ. ст. С.Ю. Медведева. – М.: Прогресс, 1991. – 384 с.

В то же время наши исследователи рынка подтвердили, что перспектива возникновения в новом десятилетии спроса на автомобили со стороны молодежи, а следовательно, и на автомобили, отвечающие вкусам молодежи, имеет под собой реальную основу, коренящуюся в демографической статистике. На национальный рынок вот-вот должны были вторгнуться миллионы юношей, появившихся на свет в период послевоенного бума рождаемости. В течение 60-х годов численность возрастной группы от 20 до 24 лет должна была возрасти более чем на 50 процентов. Более того, на долю группы в возрасте от 18 до 34 лет приходилась по крайней мере половина гигантского прироста продаж автомобилей, прогнозируемого для всего автомобильного рынка на ближайшее десятилетие.

К этому исследователи делали весьма предположительное, но интересное добавление. Ожидалось не только беспрецедентное увеличение общей численности людей молодого возраста, но также и более высокий уровень их образования по сравнению с предыдущими поколениями. Мы уже знали, что люди с высшим образованием чаще покупают автомобили, чем люди менее образованные, а наши прогнозы показали, что к 1970 году число студентов колледжей и университетов должно возрасти вдвое.

Любой автомобиль, который мог бы понравиться этим молодым покупателям, должен был обладать тремя главными свойствами: элегантным внешним видом, высокими эксплуатационными качествами и низкой ценой.

Задолго до выпуска "Мустанга" мы уже развернули исследование рыночного спроса. Один из наших завершающих тестов вселил в нас особенно большие надежды. Мы пригласили в демонстрационный зал дизайна специально подобранную группу из 52 семейных пар, проживавших в районе Детройта. Каждая из этих пар уже владела автомобилем стандартного размера и имела средний доход, то есть не являлась первоочередным потенциальным покупателем второй машины. Мы вводили их небольшими группами в нашу дизайнерскую мастерскую, показывали им опытный образец "Мустанга", а их отзывы записывали на пленку.

Выяснилось, что на пары, принадлежавшие к слою "белых воротничков", большое впечатление произвел внешний облик машины, а пары из среды "синих воротничков" усмотрели в ней символ высокого социального статуса и престижа. Когда мы попросили их высказать предположение о цене автомобиля, почти все они назвали цифру, превышавшую как минимум на тысячу долларов намеченную нами цену. Когда же мы спросили, купят ли они "Мустанг", большинство ответило отрицательно. Одни объяснили это тем, что она слишком дорога, другие – тем, что она слишком мала или что с ней слишком сложно будет обращаться.

Однако, когда мы сообщили им настоящую цену автомобиля, возникла забавная ситуация. Большинство заявило: "К черту мои возражения, готов купить!" Внезапно все их отговорки испарились. Выдвигались всякого рода новые соображения, почему данная модель в целом весьма практична. Один из приглашенных заявил: "Когда я припаркую этот автомобиль возле моего дома, все соседи подумают, что мне удалось устроиться на высокооплачиваемую работу". Другой сказал: "Автомобиль выглядит необычно, а цену на него вы установили, как на обыкновенный".

Еластичність попиту та пропозиції [255].

Розрізняють такі випадки цінової еластичності попиту.

Попит еластичний, якщо однопроцентна зміна ціни призводить до більшої процентної зміни обсягу попиту.

Попит нееластичний, коли однопроцентна зміна ціни спричиняє менш ніж однопроцентну зміну обсягу попиту.

Попит з одиничною еластичністю спостерігається, коли однопроцентна зміна ціни приводе до однопроцентної зміни обсягу попиту.

Існують також граничні випадки еластичності. Абсолютно еластичний попит спостерігається, коли споживачі купують товар у необмеженій кількості, але лише за однією ціною. Найменше зростання ціни зменшує попит до нуля, а будь-яке зниження ціни приведе до безмежного його зростання. Крива попиту є горизонтальною лінією.

Абсолютно нееластичний попит спостерігається, коли покупці зовсім нечутливі до зміни ціни, незалежно від її рівня попит виникає на одну й ту саму кількість товару. Крива попиту має вигляд вертикальної лінії.

За неціновими чинниками попиту розрізняють перехресну еластичність попиту та еластичність попиту за доходом. Обидва показники вимірюють, на скільки процентів зміститься крива попиту під впливом даного нецінового чинника.

Перехресна еластичність попиту – це процентна зміна обсягу попиту на один товар при зміні на 1 % ціни іншого товару.

Для товарів-субститутів перехресна еластичність попиту додатна, тому що при зростанні ціни одного товару обсяг його продажу зменшується, а попит на товар-замінник зростає.

Для товарів-компліментаріїв перехресна еластичність попиту від'ємна, оскільки зростання ціни одного товару призводить до зменшення обсягу попиту і на цей товар, і на товар-доповнювач.

У випадку, коли два товари є незалежними у споживанні, перехресна еластичність попиту дорівнює нулю.

Еластичність попиту за доходом – це процентна зміна обсягу попиту, викликана однопроцентною зміною доходу.

ПРАКТИЧНІ ЗАВДАННЯ

▷ **ПРИКЛАД.**

Завдання 1. *Задача для однофакторної моделі.*

Дано	Статистичні дані – показники депозитів домашніх господарств $X(i)$ за місяцями 2007-2008 рр. (табл. 6.10)
Побудувати	1. Графік тренда змінної $X(i)$. 2. Обрати форму однофакторної моделі. 3. Оцінити усі параметри моделі. 4. Визначити зони надійності при рівні значущості $a = 95$

Таблиця 6.10

Депозити домашніх господарств в Україні за період з 2007 по червень 2008 р. (залишки коштів на кінець періоду, млн. грн.)

Період	Усього
2007 р.	
Січень	111 270
Лютий	115 503
Березень	119 199
Квітень	121 035
Травень	123 312
Червень	129 209
Липень	133 929
Серпень	139 260
Вересень	143 783
Жовтень	148 608
Листопад	157 055
Грудень	167 239
2008 р.	
Січень	171 326
Лютий	177 223
Березень	182 856
Квітень	189 707
Травень	190 599
Червень	196 893

Завдання 2. *Багатофакторна модель.*

Дано	Статистичні дані – показники депозитів домашніх господарств $X(i)$ за місяцями 2007-2008 рр. залежно від строку (табл. 6.11): x_1 – іпотечні кредити строком до 1 року; x_2 – іпотечні кредити строком від 1 до 5 років; x_3 – іпотечні кредити на придбання, будівництво та реконструкцію нерухомості; y – усього іпотечних кредитів
Знайти	1. Обрати форму багатофакторної моделі. 2. Оцінити усі параметри моделі. 3. Визначити зони надійності при рівні значущості $a = 95$. 4. Оцінити коефіцієнти детермінації, автокореляції та перевірити показники на мультиколінеарність між факторами

Таблиця 6.11

Кредити, надані домашнім господарствам, за цільовим спрямуванням і строками погашення за місяцями 2007-2008 рр. (залишки коштів на кінець періоду, млн. грн.)

Період	Усього	Кредити на поточні потреби				Іпотечні кредити									Інші кредити			
		усього	у тому числі за строками			усього	у тому числі за строками			з них на придбання, будівництво та реконструкцію нерухомості	у тому числі за строками				усього	у тому числі за строками		
			до 1 року	від 1 до 5 років	більше 5 років		до 1 року	від 1 до 5 років	більше 5 років		до 1 року	від 1 до 5 років	від 5 до 10 років	більше 10 років		до 1 року	від 1 до 5 років	більше 5 років
2007																		
Січень	83 489	56 789	12 626	44 162	…	21 105	325	20 780	…	20 988	323	20 666	…	…	5 596	195	5 401	…
Лютий	87 181	59 633	12 887	29 496	17 251	25 744	376	2 360	23 008	21 586	342	1 985	4 728	14 532	1 803	183	719	901
Березень	93 262	63 695	13 503	30 522	19 670	27 607	433	4 524	22 650	23 172	392	2 967	4 872	14 941	1 960	217	831	913
Квітень	99 339	67 736	14 058	31 661	22 017	29 625	466	3 539	25 619	24 759	423	3 031	5 161	16 144	1 978	213	740	1 025
Травень	105 060	71 694	14 714	32 663	24 317	31 222	494	3 731	26 997	26 047	447	3 198	5 434	16 967	2 145	186	826	1 133
Червень	111 819	76 293	15 097	34 091	27 105	33 205	524	3 766	28 916	27 754	479	3 212	5 913	18 150	2 321	208	864	1 249
Липень	120 244	82 655	17 084	35 763	29 809	35 049	547	3 950	30 553	29 437	487	3 397	6 196	19 357	2 539	190	941	1 408
Серпень	127 244	87 501	17 539	37 425	32 536	36 916	519	4 142	32 254	31 229	488	3 522	6 461	20 758	2 827	332	983	1 512
Вересень	134 011	92 189	17 891	38 575	35 724	38 887	540	4 223	34 124	33 118	504	3 577	6 724	22 314	2 934	292	1 036	1 606

Продовж. табл. 6.11

Період	Усього	Кредити на поточні потреби				Іпотечні кредити									Інші кредити			
		усього	у тому числі за строками			усього	у тому числі за строками			з них на придбання, будівництво та реконструкцію нерухомості	у тому числі за строками				усього	у тому числі за строками		
			до 1 року	від 1 до 5 років	більше 5 років		до 1 року	від 1 до 5 років	більше 5 років		до 1 року	від 1 до 5 років	від 5 до 10 років	більше 10 років		до 1 року	від 1 до 5 років	більше 5 років
Жовтень	141 712	97 513	18 684	40 444	38 385	41 077	547	4 409	36 121	35 352	510	3 735	7 007	24 100	3 123	282	1 066	1 775
Листопад	150 108	103 171	19 435	42 008	41 729	43 586	538	4 501	38 547	37 821	499	3 749	7 425	26 149	3 351	235	1 128	1 988
Грудень	160 386	110 121	19 990	44 593	45 538	46 626	534	4 568	41 523	40 826	503	3 828	7 875	28 619	3 640	208	1 272	2 160
2008																		
Січень	164 775	113 355	20 884	45 332	47 138	47 445	573	4 526	42 345	41 889	524	3 808	8 034	29 523	3 976	206	1 324	2 446
Лютий	174 234	120 350	22 773	47 056	50 521	50 103	685	4 618	44 800	44 221	605	3 862	8 328	31 426	3 782	155	1 309	2 317
Березень	183 580	126 593	23 757	48 914	53 922	53 023	825	4 785	47 413	47 173	746	4 013	8 785	33 630	3 964	137	1 390	2 438
Квітень	191 899	132 109	24 826	50 123	57 160	55 294	833	4 775	49 686	49 894	738	3 952	9 284	35 920	4 497	139	1 409	2 949
Травень	193 546	133 933	25 894	50 077	57 962	55 157	845	4 615	49 698	50 011	749	3 791	9 120	36 352	4 455	148	1 377	2 930
Червень	198 650	137 291	25 429	51 063	60 800	56 746	841	4 581	51 323	51 658	763	3 779	9 376	37 739	4 613	149	1 401	3 063

Етапи виконання

Завдання 1. *Однофакторна модель.*

1. Наведені дані за 18 місяців є динамічним (часовим) рядом.

Кожен статистичний показник – рівень даного ряду характеризує кількість деякої ознаки економічного процесу за певний проміжок часу – місяць. Такий ряд динаміки називають інтервальним.

Зображений графічно шуканий ряд динаміки.

По осі абсцис відкладаємо час – місяці, а по осі ординат – значення показника $X(i)$. Масштаби для осей беремо зручні для побудови:

По OX: 1 см – 1 місяць.

По OY: 1 см – 1 млн. грн.

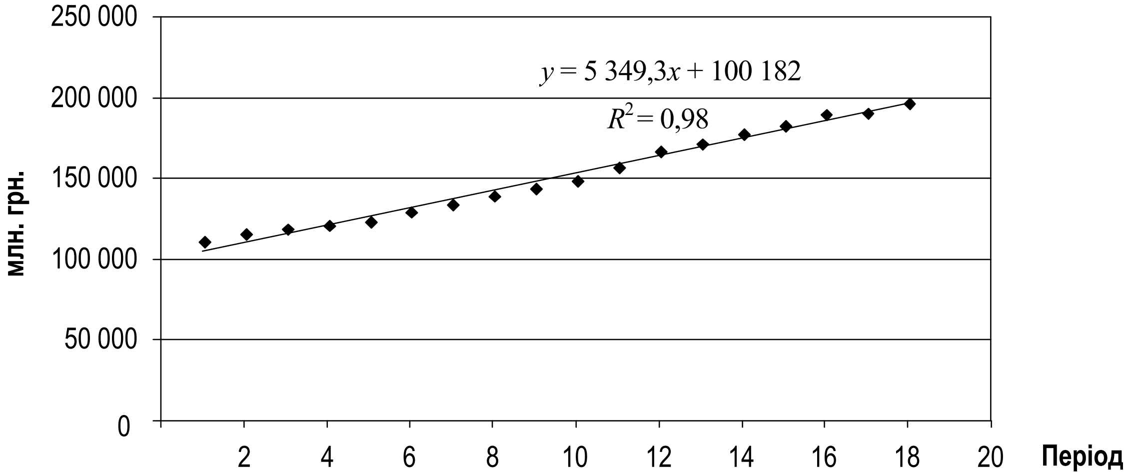
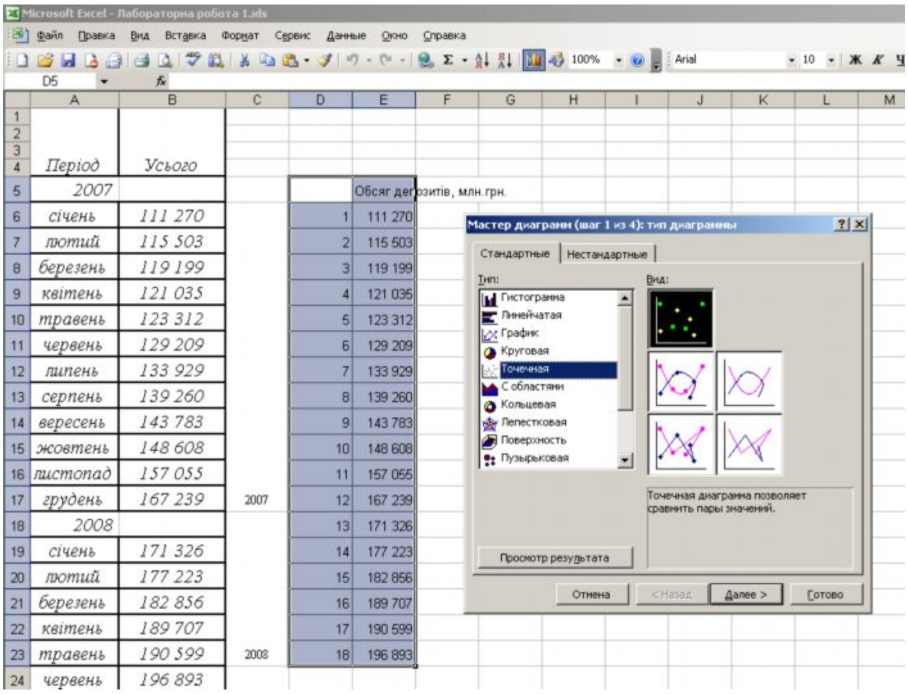


Рис. 6.25. Депозити домашніх господарств в Україні за період з 2007 по червень 2008 р.

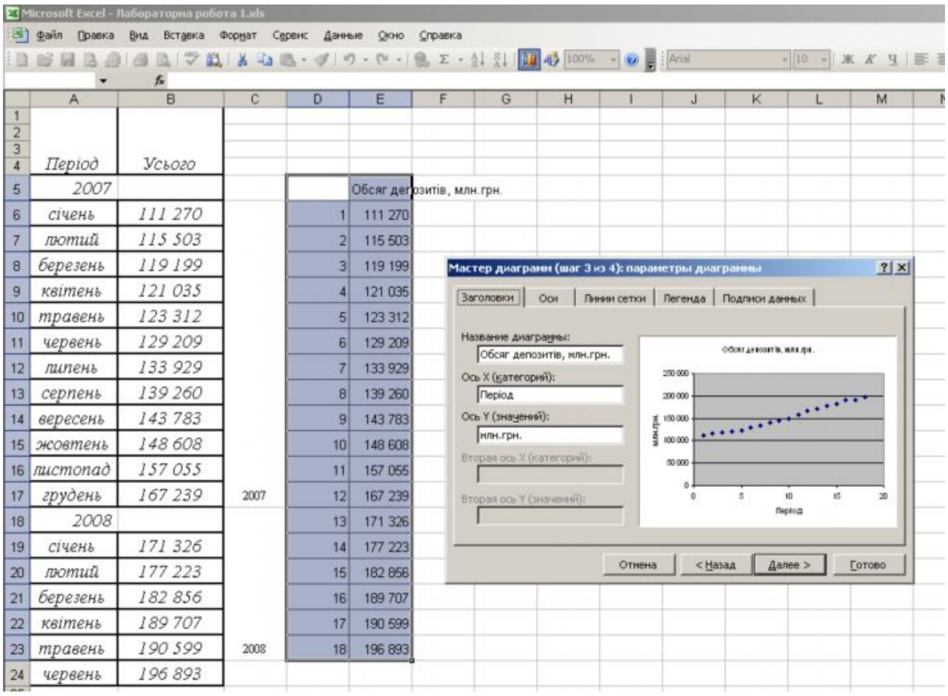
Зобразимо графічно ряд динаміки та лінію тренда засобами Microsoft Exel.

1. Виділіть стовбці для побудови діаграми, розташувавши спочатку факторну ознаку, потім результативну. Клацніть на кнопці **Мастер диаграмм**, яка розташована на стандартній панелі інструментів, або виберіть команду **Вставка/Диаграмма**.

2. На першому кроці виберіть тип діаграми **График** і клацніть на кнопці **Далее**.



На другому кроці роботи засобу **Мастер диаграмм** перевірте правильність посилань на комірки базової лінії. Клацніть на кнопці **Далее**. На третьому кроці виберіть параметри графіка, що включає в себе як лінії, так і маркери. Клацніть на кнопці **Далее**.



На останньому, четвертому, кроці роботи майстра необхідно визначити місце розташування діаграми: на окремому або тому ж листі. Клацніть на кнопці **Готово**.

3. Розрахунок ковзної середньої.

Рівні шуканого динамічного ряду зростають в більшості своїй і можна говорити лише про загальні тенденції розвитку даного явища: зростання ознаки.

Виявлення основної тенденції розвитку (тренда) називається також вирівнюванням часового ряду.

Один з найпростіших методів вирівнювання – метод ковзної середньої, який полягає у тому, що обчислюється середній рівень з певного числа перших за порядком рівнів ряду, потім – середній рівень з такою ж кількістю рівнів, починаючи з другого, далі – починаючи з третього і т.д. Таким чином, при розрахунках середнього рівня немов "ковзають" по ряду динаміки.

Алгоритм розрахунку ковзної середньої

1) визначити інтервал згладжування, тобто число вхідних у нього рівнів $m(m<n)$.

   У нашому випадку інтервал згладжування візьмемо, наприклад, інтервал, що містить 3 рівні.

2) обчислити середнє значення рівнів, що утворюють інтервал згладжування, яке одночасно є згладжувальним значенням рівня, що знаходиться в центрі інтервалу згладжування.

   У нашому випадку середню по інтервалу, що містить 3 рівні, шукаємо за формулою:

$$\tilde{X}=\frac{X(i-1)+X(i)+X(i+1)}{3}; \qquad (6.76)$$

3) зсунути інтервал згладжування на одну точку вправо, потім за формулою (6.73) обчислити згладжене значення, знову здійснити зрушення і т.д.

   У результаті послідовного застосування наведеної ітеративної процедури вийде $n-(m-1)$ нових згладжених рівнів, тобто ми отримаємо $12-(3-1)=10$ нових рівнів.

У колонці 3 табл. 6.12 одержали вирівняний динамічний ряд за методом ковзної середньої, який показує, що показник X щомісячно зростав.

Таблиця 6.12

Місяць	**Усього**	**Ковзні тримісячні середні**	**Абсолютний ланцюговий приріст**
2007 р.			
Січень	111 270	–	–
Лютий	115 503	115 324	4 233
Березень	119 199	118 579	3 696
Квітень	121 035	121 182	1 836
Травень	123 312	124 519	2 277
Червень	129 209	128 817	5 897
Липень	133 929	134 133	4 720
Серпень	139 260	138 991	5 331
Вересень	143 783	143 884	4 523
Жовтень	148 608	149 815	4 825
Листопад	157 055	157 634	8 447
Грудень	167 239	165 207	10 184
2008 р.			
Січень	171 326	171 929	4 087
Лютий	177 223	177 135	5 897
Березень	182 856	183 262	5 633
Квітень	189 707	187 721	6 851
Травень	190 599	192 400	892
Червень	196 893	–	–

Метод простої ковзної середньої прийнятний, якщо графічне зображення ряду динаміки нагадує пряму лінію.

Метод ковзної середньої є лише емпіричним прийомом попереднього аналізу.

Для того, щоб представити кількісну однофакторну модель, яка виражає загальну тенденцію зміни рівнів динамічного ряду в часі, скористаємося аналітичним вирівнюванням.

Розрахуємо спочатку абсолютні ланцюгові прирости за формулою:

$$_{\Delta}X_{yi} = X_i - X_{i-1}. \qquad (6.77)$$

Як видно з колонки 4 таблиці 6.12, ланцюгові прирости не мають тенденцію ні до збільшення, ні до зниження.

Тому обираємо як аналітичну залежність $f(t)$ лінійну однофакторну модель:

$$f(t) = a_0 + a_1 \cdot t. \tag{6.78}$$

***2.* Парна регресія** характеризує зв'язок між двома ознаками: результативною і факторною. Визначити тип рівняння можна, досліджуючи залежність графічно. Проте існують методи, що дозволяють виявити рівняння зв'язку, не вдаючись до графічного зображення. Якщо результативна і факторна ознаки зростають однаково, приблизно в арифметичній прогресії, то це свідчить про те, що зв'язок між ними лінійний, а при зворотному зв'язку – гіперболічний. Якщо результативна ознака збільшується в арифметичній прогресії, а факторна – значно швидше, то використовується параболічна або степенева регресія.

Оцінка параметрів рівнянь регресії (a_0, a_1) здійснюється методом найменших квадратів, в основі якого лежить припущення про незалежність спостережень сукупності, що досліджується.

Основний принцип методу найменших квадратів розглянемо на такому прикладі: вважатимемо, що дві величини (два показники) T і X взаємозв'язані між собою, причому X знаходиться в деякій залежності від T. Отже, X буде залежною, а T – незалежною величинами. Припустивши, що залежність між ознаками має лінійний характер, рівняння регресії має вигляд:

$$f(t) = a_0 + a_1 t, \tag{6.79}$$

де $a_0; a_1$ – параметри рівняння регресії.

Суть методу найменших квадратів полягає в знаходженні параметрів моделі a_0, a_1, при яких мінімізується сума квадратів відхилень емпіричних (фактичних) значень результативної ознаки від теоретичних, одержаних за вибраним рівнянням регресії

$$S = \sum_{1}^{n} \left(X_t - f(t) \right)^2 \to \min. \tag{6.80}$$

Для прямої залежності:

$$S = \sum \left(x_i - a_0 - a_1 t_i \right)^2 \to \min. \tag{6.81}$$

Розглядаючи S як функцію параметрів a_0 і a_1 та проводячи математичні перетворення (диференціювання), одержуємо:

$$\begin{cases} \dfrac{dS}{da_0} = \sum 2\left(a_0 + a_1 t_i - x_i\right) = 0 \\ \dfrac{dS}{da_1} = \sum 2\left(a_0 + a_1 t_i - x_i\right) t_i = 0 \end{cases} \tag{6.82}$$

Звідки система нормальних рівнянь для знаходження параметрів лінійної парної регресії методом найменших квадратів має такий вигляд:

$$\begin{cases} a_0 \cdot n + a_1 \cdot \sum_1^n t_i = \sum_1^n x_i, \\ a_0 \cdot \sum_1^n t_2 + a_1 \cdot \sum_1^n t_i^2 = \sum_1^n x_i \cdot t_i. \end{cases} \tag{6.83}$$

де n – обсяг досліджуваної сукупності (число одиниць спостережень).

Пошук параметрів можна спростити, якщо скористатися методом відліку від умовного нуля.

За умовний нуль візьмемо значення t, яке знаходиться посередині ряду, але оскільки членів ряду парне число, то члени верхньої половини ряду (від середини) позначимо числами –1, –3, –5, –7, –9, –11, –13, –15, –17, а члени нижньої половини – числами +1, +3, +5, +7, +9, +11, +13, +15, +17.

Тоді $\sum_1^n t_i = 0,$ і система рівнянь перетвориться таким чином:

$$\begin{cases} a_0 n = \sum_1^n x_i \\ a_1 \cdot \sum_1^n t_i^2 = \sum_1^n x_i \cdot t_i \end{cases} \Rightarrow \begin{cases} a_0 = \dfrac{\sum_1^n x_i}{n} = \bar{x} \\ a_1 = \dfrac{\sum_1^n x_i \cdot t_i}{\sum_1^n t_i^2}. \end{cases} \tag{6.84}$$

Параметр a_0 – це постійна величина в рівнянні регресії. Економічного змісту a_0 найчастіше не має, але в ряді випадків може трактуватися як початкове значення X; a_1 – параметр, який характеризує вплив, що здійснює зміна факторної ознаки (t) на результативну ознаку (x). Він показує на скільки одиниць у середньому зміниться (x) при зміні (t) на одну одиницю. Якщо $a_1 > 0$, то зв'язок позитивний, якщо значення a_1 є від'ємною величиною, то збільшення (t) на одиницю спричиняє зменшення (x) у середньому на a_1.

У випадку, якщо залежність між факторною та результативною ознаками має нелінійний характер, необхідно рівняння залежності перетворити до лінійного вигляду, логарифмуючи праву і ліву частини рівняння та вводячи нові позначення змінних. Після знаходження параметрів нового рівняння необхідно повернутись до початкових позначень і записати рівняння залежності нелінійного виду.

Складемо розрахункову табл. 6.13.

Використовуючи дані останнього рядка "Всього" колонок 2, 4, 5, за формулою (6.81) визначаємо параметри прямої:

$$a_0 = 100\,182{,}2.$$
$$a_1 = 5\,349{,}2.$$

Записуємо рівняння прямої ряду динаміки:

$$\hat{x}_t = 100\,182{,}2 + 5\,349{,}2 \cdot t.$$

Це і є однофакторна модель (один чинник t – час) рівняння тренда, що відображає розвиток явища в часі.

Використовуючи це рівняння, розрахуємо для кожного місяця теоретичне значення $\hat{x}_t$ (параметр t набуває значень, записаних у колонці 3, від –17 до 17), наприклад,

$$\hat{x}_t = 100\,182{,}2 + (-17) \cdot 5\,349{,}2 = 9\,244{,}6.$$

Правильність розрахунку рівнів ряду динаміки, що вирівнюється, може бути перевірена таким чином.

Сума значень емпіричного ряду повинна збігатися із сумою обчислених рівнів ряду, що вирівнюється, тобто $\sum_1^n x_i = \sum_1^n \hat{x}_t$ дійсно збіються (див. останній рядок колонок 2 і 6 табл. 6.13).

Параметри одержаного рівняння тренда можна інтерпретувати таким чином:

$a_0 = 100\,182{,}2$ – середній рівень ряду;

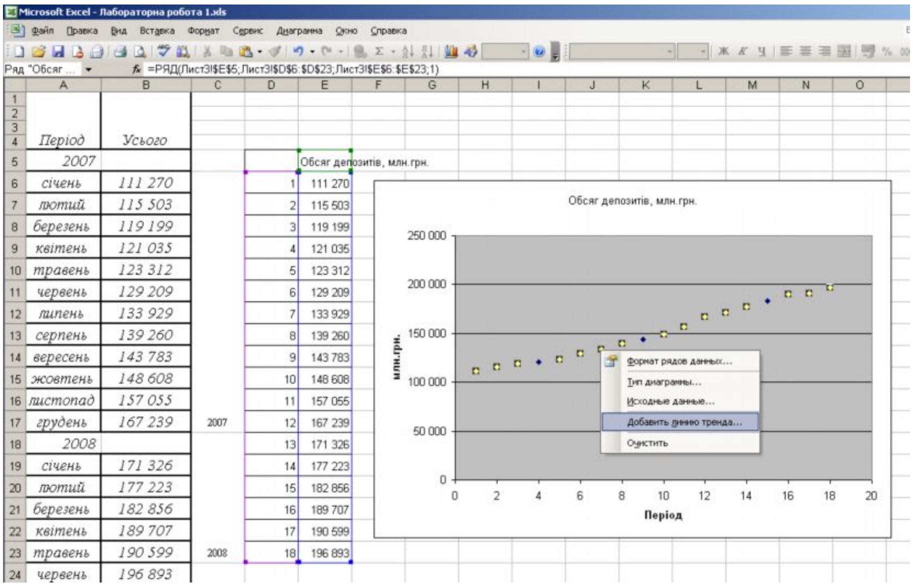
$a_1 = 5\,349{,}2$ – показник сили зв'язку, тобто щомісячно показник x зростає на 5 349,2.

Таблиця 6.13

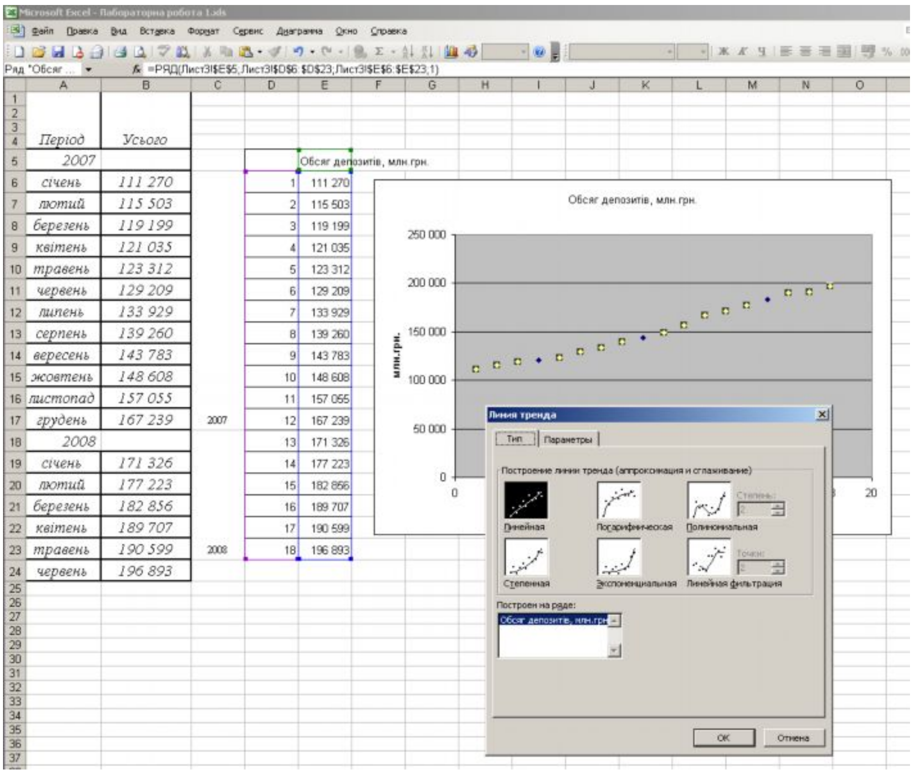
Місяць t	**Показник** x_i	**Умовне позначення періодів** t_i	$x_i t_i$	t_i^2	**Вирівняні рівні ряду** $\hat{x}_t$	$x_i - \hat{x}_t$	$(x_i - \hat{x}_t)^2$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	111 270	–17	–1 891 590	289	9 244,640867	102 025	10 409 173 906	–39 730	1578499387
2	115 503	–15	–1 732 545	225	19 943,18369	95 560	9131678492	–35 497	1260060674
3	119 199	–13	–1 549 587	169	30 641,72652	88 557	7 842 390 686	–31 801	1 011 324 802
4	121 035	–11	–1 331 385	121	41 340,26935	79 695	6 351 250 093	–29 965	897921201,8
5	123 312	–9	–1109 808	81	52 038,81218	71 273	5079867302	–27 688	766 643 802,8
6	129 209	–7	–904 463	49	62737,35501	66 472	4 418 479 588	–21 791	474 862 208,4
7	133 929	–5	–669 645	25	73435,89783	60 493	3 659 415 410	–17 071	291 430 421,8
8	139 260	–3	–417 780	9	84 134,44066	55 126	3 038 827 292	–11 740	137 835 426,8
9	143 783	–1	–143 783	1	94 832,98349	48 950	2 396 104 117	–7 217	52 089 900,44
10	148 608	1	148 608	1	105 531,5263	43 076	1 855 582 585	–2 392	5 723 258,778
11	157 055	3	471 165	9	116 230,0691	40 825	1 666 674 979	6 055	36 658 988,44
12	167 239	5	836 195	25	126 928,612	40 310	1 624 927 383	16 239	263 694 295,1
13	171 326	7	1 199 282	49	137 627,1548	33 699	1 135 612 168	20 326	413 132 725,4
14	177 223	9	1 595 007	81	148 325,6976	28 897	835 054 084,5	26 223	687 628 247,1
15	182 856	11	2 011 416	121	159024,2405	23 832	567 952 763,1	31 856	1 014 783 499
16	189 707	13	2 466 191	169	169722,7833	19 984	399 368 917,8	38 707	1 498 206 044
17	190 599	15	2 858 985	225	180 421,3261	10 178	103 585 045,8	39 599	1 568 054 402
18	196 893	17	3 347 181	289	191 119,8689	5 773	33 329 042,27	45 893	2 106 136 854
Всього	**2 718 006**	**0**	**5 183 444**	**1 938**	**1 803 281**	**914 725**	**60 549 273 856**	**0**	**14 064 686 138**

Узявши координати двох крайніх точок, побудуємо лінію ряду, що вирівнюється (трендовую лінію).

Для побудови лінії тренда засобами MS Exel необхідно виклика-ти контекстне меню (клацнути праву клавішу миші) та обрати пункт **Добавить линию тренда**.



На вкладці **Тип** можна обрати вид тренда (лінійний, логариф-мічний, поліноміальний, степеневий, експоненційний).



3. Використовуючи дані стовпця 8 табл. 6.13, розрахуємо середню квадратичну помилку лінійного рівняння тренда:

$$S_{€} = \sqrt{\frac{\sum_{i}^{n}\left(x_i - €_i\right)^2}{n-m}}, \qquad (6.85)$$

де n – число рівнів ряду (n = 18);
m – кількість параметрів у рівнянні тренда (для прямої m = 18), тоді

$$S_{€} = \sqrt{\frac{200\ 861\ 731}{18-2}} = 3\ 543{,}14.$$

Загальна дисперсія результативної ознаки визначається:

$$\sigma_{об}^2 = \frac{\sum_{1}^{n}\left(x_i - \overline{x}\right)^2}{n}, \qquad (6.86)$$

де – середнє $\left(\overline{x} = 151\,000\right)$.

Використовуючи дані стовпця 10 табл. 6.12, знайдемо:

$$\sigma_{ош}^2 = \frac{14\ 064\ 686\ 138}{18} = 781\,371\,452{,}11.$$

$$\sigma_{ош} = \sqrt{781\,371\,452} = 27\,953{,}02.$$

Дисперсія помилок:

$$\sigma_{ош}^2 = \frac{\sum_{1}^{n}\left(x_i - €_i\right)^2}{n}. \qquad (6.87)$$

$$\sigma_{ош}^2 = \frac{200\ 861\ 731}{18} = 11\,158\,985{,}04.$$

Загальна дисперсія складається з двох частин: дисперсії, яка пояснює регресію, і дисперсії помилок (або дисперсії випадкової величини):

$$\sigma_{об}^2 = \sigma_{ош}^2 + \sigma_{регр}^2, \qquad (6.88)$$

тоді

$$\sigma^2_{регр} = \sigma^2_{об} - \sigma^2_{ош} \quad 781\,371\,452{,}11 - 11\,158\,985{,}04 = 77\,021\,2467{,}1.$$

Частина дисперсії, яка пояснює регресію, називається коефіцієнтом детермінації R^2:

$$R^2 = \frac{\sigma^2_{регр}}{\sigma^2_{об}}. \tag{6.89}$$

Маємо:

$$R^2 = \frac{770\,212\,467{,}1}{781\,371\,452} = 0{,}9.$$

Коефіцієнт детермінації є мірою щільності зв'язку між факторною і результативною ознаками.

Чим ближче R^2 до 1, тим більш сильний зв'язок між x і t.

Коефіцієнт детермінації використовується як критерій адекватності моделі за допомогою F-критерію Фішера:

$$F_{расч} = \frac{0{,}985718719(18-2)}{(1-0{,}97429369)(2-1)} = 1\,104{,}3.$$

За таблицями F-розподілу Фішера знаходимо F-критичне значення при заданому рівні значущості $(\alpha = 0{,}95)$ і з $(1;\,18-2) = (1;\,16)$ ступенями свободи (для простої лінійної регресії):

$$F_{кр} = 4{,}49.$$

Оскільки $F_{расч} = 1104{,}3 > F_{кр} = 4{,}49,$ то можна зробити висновок про адекватність моделі.

Таким чином, вид взаємозв'язку був вибраний правильно.

4. Побудуємо довірчі інтервали для параметрів.

Розрахуємо середні квадратичні відхилення для параметрів за формулами:

$$S_{a_0} = \sigma^2_{ош} \cdot \frac{\sum_1^n t_i^2}{n\sum_1^n \left(t_i - \bar{t}\right)^2}. \tag{6.90}$$

Оскільки $\bar{t} = 0,$ то $\sum_1^n \left(t - \bar{t}\right)^2 = \sum_1^n t_i^2,$ тоді

$$S_{a_0} = \sigma^2_{ош} \cdot \frac{1}{n}; \;\; S_{a_1} = \sigma^2_{ош} \cdot \frac{1}{\sum_1^n t_i^2};$$

тоді

$$S_{a_0} = 1\,742{,}4.$$
$$S_{a_1} = 160{,}9.$$

Значення *t*-статистики знаходимо за статистичними таблицями *t*-розподілу Стьюдента при рівні значущості $1 - \alpha = 0{,}95$ і ступеня свободи $n - 2 = 18 - 2 = 16$:

$$t_{\frac{\alpha}{2};\, n-2} = t_{0{,}025;\,16} = 2{,}1.$$

Тоді згідно з формулою:

$$\hat{a}_i - S_{a_i} \cdot t_{\frac{\alpha}{2};\, n-2} \le a_i \le \hat{a}_i + S_{a_i} \cdot t_{\frac{\alpha}{2};\, n-2} \tag{6.91}$$

маємо такі довірчі інтервали для параметрів:

1) $100\,182{,}2549 - 1\,742{,}383524 \cdot 2{,}1199 \le a_0 \le 100\,182{,}2549 + 1\,742{,}383524 \cdot 2{,}1199;$

$$96\,488{,}5 \le a_0 \le 103\,875{,}9;$$

2) $5\,349{,}2 - 160{,}9 \cdot 2{,}1 \le a_1 \le 5\,349{,}2 + 160{,}9 \cdot 2{,}1;$

$$5\,008{,}03 \le a_1 \le 5\,690{,}5.$$

5. Перевіримо показник *X* на автокореляцію.

Автокореляція – це кореляційна залежність між послідовними значеннями рівнів ознаки.

Коефіцієнт автокореляції розраховуємо за формулою:

$$r_a = \frac{\overline{x_t \cdot x_{t-1}} - \overline{x_t} \cdot \overline{x_{t-1}}}{\sigma_{x_t} \cdot \sigma_{x_{t-1}}}, \qquad (6.92),$$

де x_{t-1} – ряд, зсунутий на одну одиницю часу щодо первинного положення,

$$\left.\begin{aligned} &\overline{x_t \cdot x_{t-1}} = \frac{1}{n}\sum x_t \cdot x_{t-1}; \\ &\overline{x_t} = \frac{1}{n}\sum x_t; \\ &\overline{x_{t-1}} = \frac{1}{n-1}\sum x_{t-1}; \\ &\overline{x_t^2} = \frac{1}{n}\sum x_t^2; \\ &\overline{x_{t-1}^2} = \frac{1}{n-1}\sum x_{t-1}^2. \end{aligned}\right\} \text{ оцінки середніх}$$

$$\left.\begin{aligned} &\sigma^2_{x_t} = \overline{x_t^2} - (\overline{x_t})^2 \\ &\sigma^2_{x_{t-1}} = \overline{x_{t-1}^2} - \left(\overline{x_{t-1}}\right)^2. \end{aligned}\right\} \text{ оцінки дисперсій}$$

Для зручності складемо розрахункову таблицю 6.14:

Тоді

$$\sigma^2_{x_t} = 23\,582\,472\,119 - \left(151\,000\right)^2 = 781\,371\,452{,}1.$$

$$\sigma^2_{x_{t-1}} = 22\,689\,273\,217 - \left(148\,301\right)^2 = 696\,156\,404{,}8.$$

$$\sigma_{x_t} = \sqrt{781\,371\,452{,}1} = 27\,953{,}02.$$

$$\sigma_{x_{t-1}} = \sqrt{696\,156\,404{,}8} = 26\,384{,}78.$$

Таблиця 6.14

Місяць	x_t	x_{t-1}	x_t^2	x_{t-1}^2	$x_t \cdot x_{t-1}$	$x_t + x_{t-1}$	$(x_t + x_{t-1})^2$
1	111 270	–	12 381 012 900	–	–	111 270	12 381 012 900
2	115 503	111 270	13 340 943 009	12 381012 900	12 852 018 810	226 773	51 425 993 529
3	119 199	115 503	14 208 401 601	13 340 943 009	13 767 842 097	234 702	55 085 028 804
4	121 035	119 199	14 649 471 225	14 208 401 601	14 427 250 965	240 234	57 712 374 756
5	123 312	121 035	15 205 849 344	14 649 471 225	14 925 067 920	244 347	59 705 456 409
6	129 209	123 312	16 694 965 681	15 205 849 344	15 933 020 208	252 521	63 766 855 441
7	133 929	129 209	17 936 977 041	16 694 965 681	17 304 832 161	263 138	69 241 607 044
8	139 260	133 929	19 393 347 600	17 936 977 041	18 650 952 540	273 189	74 632 229 721
9	143 783	139 260	20 673 551 089	19 393 347 600	20 023 220 580	283 043	80 113 339 849
10	148 608	143 783	22 084 337 664	20 673 551 089	21 367 304 064	292 391	85 492 496 881
11	157 055	148 608	24 666 273 025	22 084 337 664	23 339 629 440	305 663	93 429 869 569
12	167 239	157 055	27 968 883 121	24 666 273 025	26 265 721 145	324 294	1,05167E+11
13	171 326	167 239	29 352 598 276	27 968 883 121	28 652 388 914	338 565	1,14626E+11
14	177 223	171 326	31 407 991 729	29 352 598 276	30 362 907 698	348 549	1,21486E+11
15	182 856	177 223	33 436 316 736	31 407 991 729	3 2406 288 888	360 079	1,29657E+11
16	189 707	182 856	35 988 745 849	33 436 316 736	34 689 063 192	372 563	1,38803E+11
17	190 599	189 707	36 327 978 801	35 988 745 849	36 157 964 493	380 306	1,44633E+11
18	196 893	190 599	38 766 853 449	36 327 978 801	37 527 608 907	387 492	1,5015E+11
разом	2718006	2521113	4,24484E+11	3,85718E+11	3,98653E+11	–	1,60751E+12
середня	151 000	148 301	23 582 472 119	22 689 273 217	23 450 181 295	–	89 306 017 049

Для контролю правильності обчислень розраховуємо ще $\sum\left(x_t + x_{t-1}\right)^2$ і використовуємо тотожність:

$$\sum\left(x_t + x_{t-1}\right)^2 = \sum x_t^2 + 2\sum x_t \cdot x_{t-1} + \sum x_{t-1}^2.$$

Дійсно,

$$1{,}6 \cdot 10^{12} = 4{,}2 \cdot 10^{11} + 2 \cdot 3{,}9 \cdot 10^{11} + 3{,}8 \cdot 10^{11};$$
$$1{,}6 \cdot 10^{12} \equiv 1{,}6 \cdot 10^{12}.$$

Тепер за формулою (6.89) знаходимо значення коефіцієнта автокореляції: $r_a = 0{,}9,$ оскільки $r_a > 0{,}5,$ то автокореляція існує.

6. Критерій Дарбіна-Уотсона – найвідоміший тест перевірки моделі на наявність автокореляції в залишках, який розраховується за формулою:

$$d = DW = \frac{\sum\left(\ell_t - \ell_{t-1}\right)^2}{\sum \ell_t^2}, \tag{6.93}$$

де $\ell_t = x_i - \hat{x}_t.$

$$d = DW = \frac{200\ 861\ 731}{23\ 582\ 472\ 119} = 0{,}008.$$

Коефіцієнт автокореляції першого порядку і статистика Дарбіна-Уотсона $\left(d = DW\right)$ тісно зв'язані співвідношенням:

$$d \approx 2 \cdot \left(1 - r_a\right).$$

За таблицею значень *d-статистики* Дарбіна-Уотсона при рівні значущості $\alpha = 0{,}95\,\%,$ числі спостережень n = 18, числі незалежних змінних у рівнянні регресії k = 1 знаходимо два критичні значення: нижнє d_L – межа для визнання позитивної автокореляції залишків і верхнє d_U – межа визнання її відсутності.

Маємо $d_L = 1{,}16;\ d_U = 1{,}39,$ тоді $4 - d_U = 4 - 1{,}39 = 2{,}61.$

Таким чином, інтервал (1,39; 2,61) є областю ухвалення нульової гіпотези H_0: – автокореляція відсутня.

Оскільки розрахункове значення d-статистики Дарбіна-Уотсона

$$d = 0{,}008517416 < d_L = 1{,}16,$$

то гіпотеза H_0 відкидається, є позитивна кореляція.

Таблиця 6.15

Області ухвалення і неприйняття нульової гіпотези за методом Дарбіна-Уотсона

Значення статистики DW	Висновок
$4 - d_L < DW < 4$	Гіпотеза відкидається, є негативна кореляція
$4 - d_U < DW < 4 - d_L$	Невизначеність, не приймається і не відкидається
$2 < DW < 4 - d_U$. $d_U < DW < 2$	Приймається гіпотеза H_0
$d_L < DW < d_U$	Невизначеність, не приймається і не відкидається
$0 < DW < d_L$	Гіпотеза відкидається, є позитивна кореляція

7. Застосування прогнозування припускає, що закономірність розвитку, що діяла у минулому (усередині ряду динаміки), збережеться і в прогнозованому майбутньому, тобто прогноз заснований на екстраполяції.

Найпоширенішим методом прогнозування вважають аналітичний вираз тренда.

При цьому для виходу за межі досліджуваного періоду достатньо продовжити значення незалежної змінної часу *t*.

У нашому випадку t = 19; 21; 23 для наступних трьох місяців.

Маємо:

$$\hat{x}_t = 100\,182{,}2 + 5\,349{,}2 \cdot t.$$
$$\hat{x}_{19} = 100\,182{,}2 + 5\,349{,}2 \cdot 19 = 201\,817{,}7.$$
$$\hat{x}_{21} = 100\,182{,}2 + 5\,349{,}2 \cdot 21 = 212\,516{,}9.$$
$$\hat{x}_{23} = 100\,182{,}2 + 5\,349{,}2 \cdot 23 = 223\,215{,}4.$$

Будь-який статистичний прогноз має наближений характер, тому доцільно визначити довірчі інтервали прогнозу:

$$\hat{x}_t \pm t_\alpha \cdot \frac{\sigma_{\bar{x}_t}}{\sqrt{n}}, \tag{6.94}$$

де $\sigma_{\bar{x}_t}$ – середня квадратична помилка тренда;

$\hat{x}_t$ – розрахункове значення рівня;

t_α – довірча величина.

Значення t-статистики ми вже знаходили:

$$t_{(0,025;16)} = 2,1.$$

Середню квадратичну помилку тренда ми знаходили в другому пункті рішення:

$$\sigma_{\bar{x}} = S_\epsilon = 3\,543,14.$$

Тоді

$$t_\alpha \cdot \frac{\sigma_{\bar{x}_t}}{\sqrt{n}} \quad 2,1 \cdot \frac{3\,543,14}{\sqrt{16}} = 1\,877,7.$$

Таким чином, прогноз для наступних трьох місяців з імовірністю 95 %

$$201\,817,7 \pm 1\,877,7;$$
$$212\,516,9 \pm 1\,877,7;$$
$$223\,215,4 \pm 1\,877,7.$$

Виникнення цих відхилень пояснюється такими причинами:

1) вибрана для прогнозування крива не є єдино можливою для опису тенденції. Можна підібрати таку криву, яка дасть більш точні результати;
2) прогноз здійснюється на основі обмеженого кількості початкових даних. Крім того, кожен початковий рівень володіє ще випадковою

компонентою. Тому і крива, за якою здійснюється екстраполяція, міститиме випадкову компоненту;
3) тенденція характеризує лише рух середнього рівня ряду динаміки, тому окремі спостереження від нього відхиляються. Якщо такі відхилення спостерігалися у минулому, то вони спостерігатимуться і в майбутньому.

Алгоритм використання інструмента РЕГРЕССИЯ в MS Exel

Для проведення прогнозу скористаємося інструментом **Регрессия**, який поставляється разом з надбудовою **Пакет анализа**. Припустимо, наприклад, що ви вже побудували для свого набору даних точкову діаграму і додали до неї лінію тренда, що дозволило вам поверхнево оцінити наявні дані. Тепер ви хочете провести детальне дослідження та одержати більш повну і точну інформацію. Щоб провести регресійний аналіз з використанням можливостей надбудови **Пакет анализа**, виконайте ряд дій.

1. Вибираємо команду **Сервис / Анализ данных**.

2. У діалоговому вікні, що відкрилося, **Анализ данных** у списку **Инструменты анализа** вибираємо пункт **Регрессия** і клацаємо на кнопці **OK**. На екрані відображається діалогове вікно **Регрессия**.

3. Визначте значення *X* і *Y*.

У полі **Входной интервал** *Y* вкажіть посилання на діапазон комірок, в яких міститься набір залежних значень. Потім перейдіть до поля **Входной интервал** *X* і вкажіть посилання на діапазон комірок, в яких міститься набір незалежних значень.

Кожний з цих діапазонів повинен бути одиничним стовпцем із значеннями. Наприклад, якщо Ви хочете використовувати інструмент **Регрессия** для вивчення дії реклами на обсяг продажів, введіть в полі **Входной интервал** *X* (кількість виходів реклами), в полі **Входной интервал** *Y* (обсяг продажів). Якщо вказані комірки містять підписи даних, встановіть прапорець **Метки**.

4. Встановіть прапорець **Константа-ноль** (за потреби).

Встановіть цей прапорець, якщо лінія регресії повинна перетинати вісь значень у нульовій точці. Іншими словами, якщо необхідно, щоб нульовому значенню незалежної величини відповідало нульове значення залежної величини.

5. Вкажіть, чи потрібно при проведенні регресійного аналізу враховувати рівень надійності.

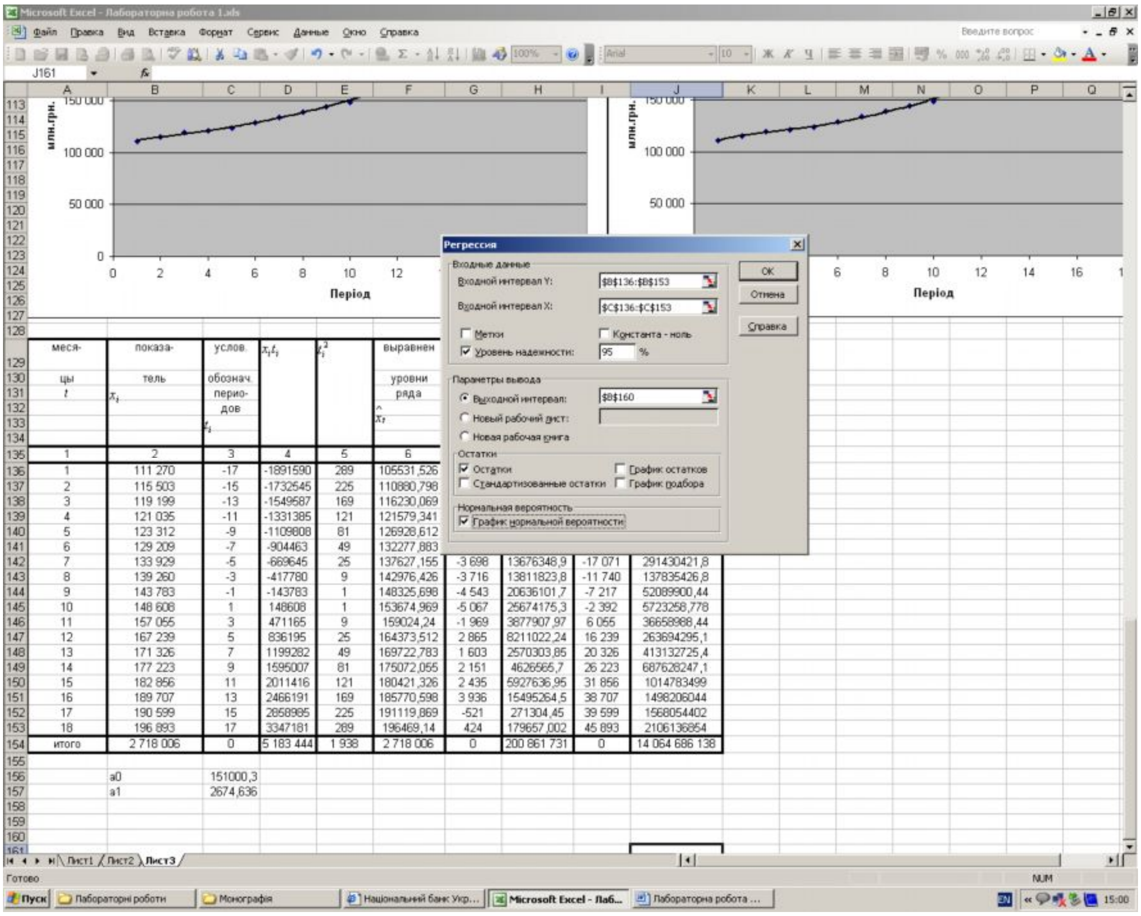
Для цього встановіть прапорець **Уровень надежности** і в розташованому справа текстовому полі задайте значення цього рівня.

6. Вкажіть місце, куди повинні бути поміщені результати проведеного регресійного аналізу.

Використовуйте перемикачі групи **Параметры выводов**, щоб вказати *Excel*, де повинен бути розміщений звіт про одержані результати. Наприклад, щоб розташувати цей звіт на тому ж листі, де розташовані початкові дані, виберіть перемикач **Выходной Интервал** і в полі праворуч від нього вкажіть адресу комірок, які повинні містити результати проведеного дослідження. Щоб розташувати звіт в якомунебудь іншому місці, виберіть один з двох інших перемикачів.

7. Визначте, які саме значення повинні бути обчислені. Використовуйте прапорці групи **Остатки**, щоб визначити, яка інформація про залишки повинна бути включена в звіт про проведений регресійний аналіз.

Також встановіть прапорець **График нормальний вероятности**, щоб додати в звіт інформацію про залишки, які відповідають нормальній вірогідності, і відобразити графік нормальної вірогідності.



8. Клацніть на кнопці **ОК**.

Excel виконає всі необхідні обчислення і відобразить звіт.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,992834
R-квадрат	0,985719
Нормированный R	0,984826
Стандартная оши	3543,142
Наблюдения	18

Дисперсионный анализ

	df	SS	MS	F	ачимость F
Регрессия	1	1,39E+10	1,39E+10	1104,3477	3,42E-16
Остаток	16	2,01E+08	12553858		
Итого	17	1,41E+10			

	Коэффициент	дартная о	статисти	P-Значение	ижние 95%	Верхние 95%	жние 95,0	Верхние 95,0%
Y-пересечение	151000,3	835,1267	180,8113	6,4394E-28	149229,9	152770,723	149229,9	152770,7228
Переменная X 1	2674,636	80,4844	33,23173	3,4199E-16	2504,016	2845,25502	2504,016	2845,255017

ВЫВОД ОСТАТКА

Наблюдение	дсказанно	Остатки
1	105531,5	5738,474
2	110880,8	4622,202
3	116230,1	2968,931
4	121579,3	-544,341
5	126928,6	-3616,61
6	132277,9	-3068,88
7	137627,2	-3698,15
8	142976,4	-3716,43
9	148325,7	-4542,7
10	153675	-5066,97
11	159024,2	-1969,24
12	164373,5	2865,488
13	169722,8	1603,217
14	175072,1	2150,945
15	180421,3	2434,674
16	185770,6	3936,402
17	191119,9	-520,869
18	196469,1	423,8596

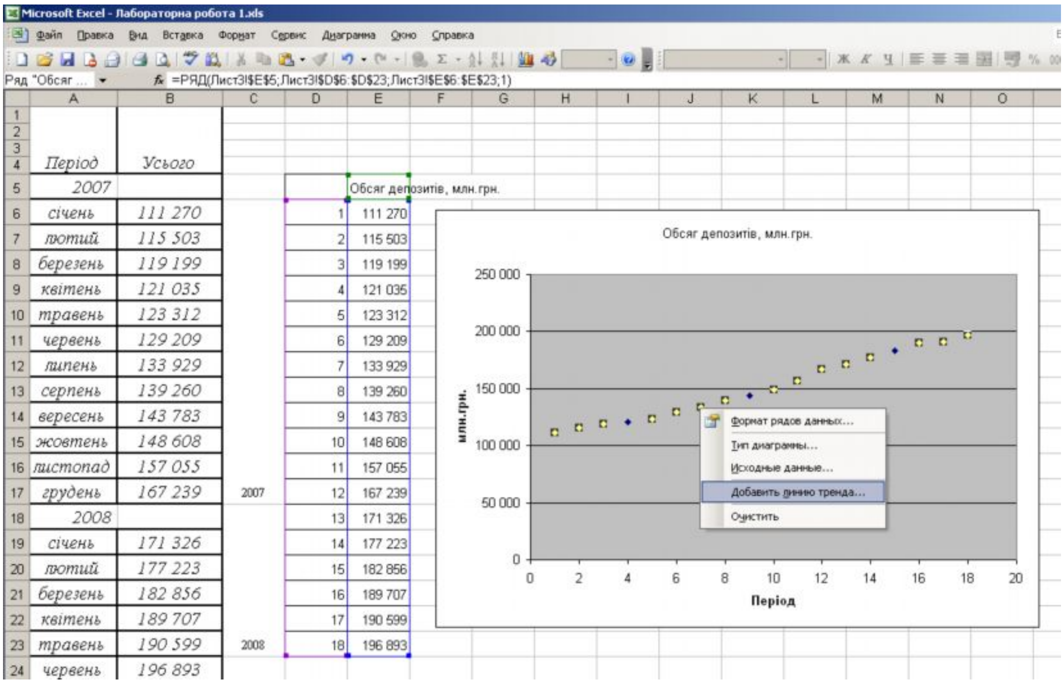
ВЫВОД ВЕРОЯТНОСТИ

Персентиль	Y
2,77777778	111270
8,33333333	115503
13,8888889	119199
19,4444444	121035
25	123312
30,5555556	129209
36,1111111	133929
41,6666667	139260
47,2222222	143783
52,7777778	148608
58,3333333	157055
63,8888889	167239
69,4444444	171326
75	177223
80,5555556	182856
86,1111111	189707
91,6666667	190599
97,2222222	196893

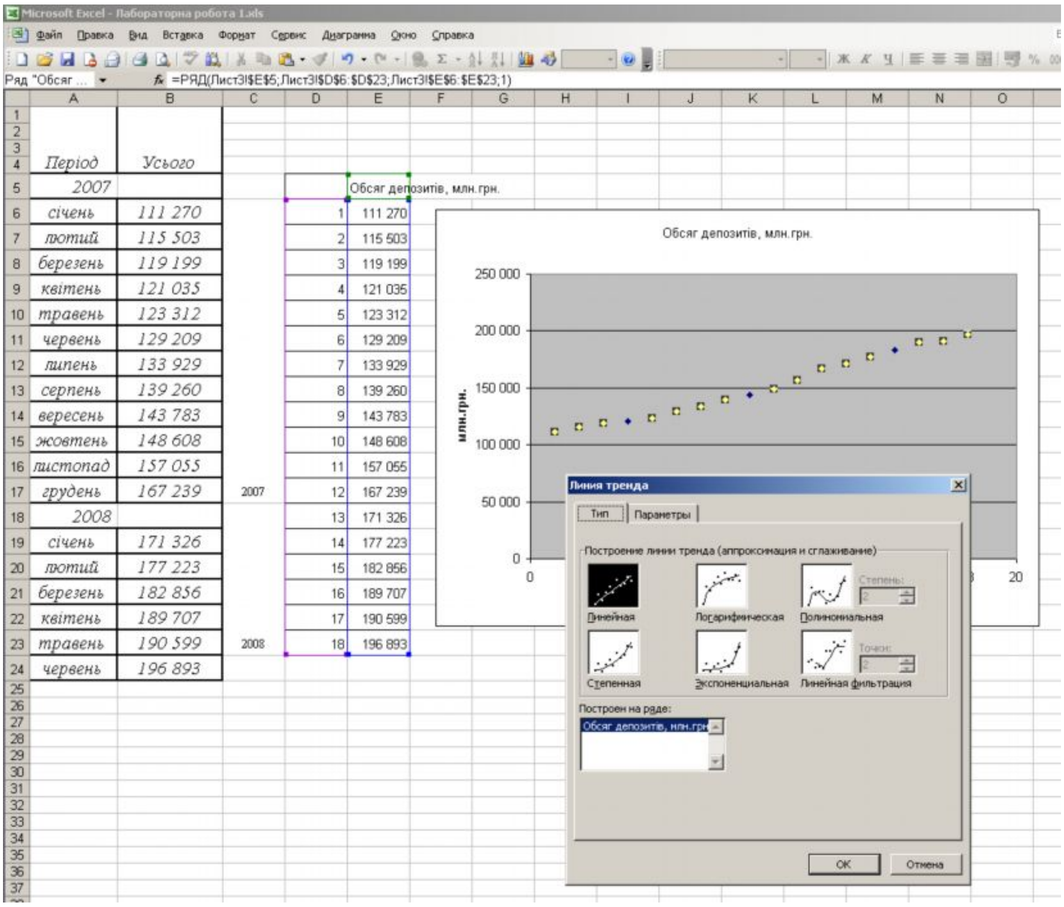
Цей звіт містить значення деяких ключових статистичних регресійних показників, включаючи значення *R*-квадрата, значення стандартної помилки і кількість наглядів. Далі – дані дисперсійного аналізу, включаючи інформацію про кількість ступенів свободи, суми квадратів. Нижче подані дані про побудовану лінію регресії, включаючи показники коефіцієнтів, стандартної помилки, *t*-статистики, показники вірогідності, а також деякі дані про незалежну змінну, яка в даному прикладі є кількістю виходів рекламних роликів.

Побудова лінії тренда в MS Exel

Для побудови лінії тренда необхідно викликати контекстне меню (клацнути правою клавішою миші) та вибрати пункт **Добавить линию тренда**.



На вкладці Тип можна обрати вид тренда (лінійний, логарифмічний, поліноміальний, степеневий, експоненційний).



Завдання 2. *Багатофакторна модель.*

1. Вивчення зв'язку між трьома і більше зв'язаними між собою ознаками має назву множинної (багатофакторної) регресії.

У нашому випадку залежну змінну – усього іпотечних кредитів (тис. грн.), позначимо y.

Пояснюючі змінні:

x_1 – іпотечні кредити строком до 1 року;

x_2 – іпотечні кредити строком від 1 до 5 років;

x_3 – іпотечні кредити на придбання, будівництво та реконструкцію нерухомості;

Число спостережень $n = 9$.

Вимагається побудувати статистичну модель, що виражає залежність результативної ознаки від чотирьох пояснюючих змінних – факторів і проаналізувати побудовану модель.

Загальна множинна лінійна регресійна модель може бути записана у вигляді:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_m x_m + E, \tag{6.95}$$

де $x_1, x_2, \ldots, x_m$ – незалежні змінні (фактори);

$a_0, a_1, \ldots, a_m$ – параметри моделі;

E – випадкова величина.

У нашому випадку маємо модель множинної регресії з 3 факторами:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3. \tag{6.96}$$

Багатофакторна модель (6.92) може бути записана у вигляді:

$$Y = XA + E, \tag{6.97}$$

де $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ – векторний стовпець розмірності;

$A = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix}$ – вектор-стовбець розмірності $\left((p+1)\cdot 1\right)$ невідомих параметрів рівняння;

X – матриця спостережень розмірності $n \cdot (p+1)$;

E – вектор-стовпець розмірності $n \cdot 1$ випадкових величин – помилок.

Вектор невідомих параметрів ми знаходимо методом найменших квадратів (МНК), мінімізуючи суму квадратів залишків:

$$\sum_{i=1}^{n} \ell_i^2 = \sum_{i=1}^{n} (y_1 - a_0 - a_1 x_{1i} - a_2 x_{2i} - \ldots - a_p x_{pi})^2. \tag{6.98}$$

Після знаходження частинних похідних і прирівняння їх до нуля, після відповідних перетворень ми одержимо систему з $(p+1)$ невідомим:

$$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_{1i} + a_2 \sum_{i=1}^{n} x_{2i} + \ldots + a_p \sum_{i=1}^{n} x_{pi} = \sum_{i=1}^{n} y_1 \\ a_0 \sum_{i=1}^{n} x_{1i} + a_1 \sum_{i=1}^{n} x_{1i}^2 + a_2 \sum_{i=1}^{n} x_{1i} x_{2i} + \ldots + a_p \sum_{i=1}^{n} x_{1i} x_{pi} = \sum_{i=1}^{n} x_{1i} y_1 \\ a_0 \sum_{i=1}^{n} x_{2i} + a_1 \sum_{i=1}^{n} x_{2i} x_{1i} + a_2 \sum_{i=1}^{n} x_{2i}^2 + \ldots + a_p \sum_{i=1}^{n} x_{2i} x_{pi} = \sum_{i=1}^{n} x_{2i} y_1 \\ a_0 \sum_{i=1}^{n} x_{pi} + a_1 \sum_{i=1}^{n} x_{pi} x_{1i} + a_2 \sum_{i=1}^{n} x_{pi} x_{2i} + \ldots + a_p \sum_{i=1}^{n} x_{pi}^2 = \sum_{i=1}^{n} x_{pi} y_1 \end{cases} \tag{6.99}$$

У нашому випадку ми маємо:

$$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_1 + a_2 \sum_{i=1}^{n} x_2 + \ldots + a_3 \sum_{i=1}^{n} x_3 = \sum_{i=1}^{n} y \\ a_0 \sum_{i=1}^{n} x_1 + a_1 \sum_{i=1}^{n} x_1^2 + a_2 \sum_{i=1}^{n} x_1 x_2 + \ldots + a_3 \sum_{i=1}^{n} x_1 x_3 = \sum_{i=1}^{n} x_1 y \\ a_0 \sum_{i=1}^{n} x_2 + a_1 \sum_{i=1}^{n} x_2 x_1 + a_2 \sum_{i=1}^{n} x_2^2 + \ldots + a_3 \sum_{i=1}^{n} x_2 x_3 = \sum_{i=1}^{n} x_2 y \\ a_0 \sum_{i=1}^{n} x_3 + a_1 \sum_{i=1}^{n} x_3 x_1 + a_2 \sum_{i=1}^{n} x_3 x_2 + \ldots + a_3 \sum_{i=1}^{n} x_3^2 = \sum_{i=1}^{n} x_3 y \end{cases} \tag{6.100}$$

У матричному вигляді (оскільки $p = 3$, $n = 9$):

$$\begin{pmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 \\ \sum x_1 & \sum x_1^2 & \sum x_2 x_1 & \sum x_3 x_1 \\ \sum x_2 & \sum x_2 x_1 & \sum x_2^2 & \sum x_2 x_3 \\ \sum x_3 & \sum x_3 x_1 & \sum x_3 x_2 & \sum x_3^2 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & \dots & x_{19} \\ x_{21} & x_{22} & \dots & x_{29} \\ x_{31} & x_{32} & \dots & x_{39} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \quad (6.101)$$

У скороченому вигляді можна записати:

$$X'XA = X'Y. \quad (6.102)$$

Звідси одержуємо рівняння для знаходження невідомих параметрів у матричному вигляді:

$$A = \left(X'X\right)^{-1} X'Y, \quad (6.103)$$

де $\left(X'X\right) = \begin{pmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 & \sum x_1 x_3 \\ \sum x_2 & \sum x_2 x_1 & \sum x_2^2 & \sum x_2 x_3 \\ \sum x_3 & \sum x_3 x_1 & \sum x_{32} & \sum x_3^2 \end{pmatrix}$

$$X'Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & \dots & x_{19} \\ \dots & \dots & \dots & \dots \\ x_{31} & x_{32} & \dots & x_{39} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \\ \sum x_3 y \end{pmatrix}$$

З таблиці 6.15 одержуємо, що

$$X'X = \begin{pmatrix} 9{,}00 & 6221{,}74 & 41377{,}49 & 398845{,}72 \\ 6221{,}74 & 4467803{,}45 & 28709567{,}58 & 281992011{,}20 \\ 41377{,}49 & 28709567{,}58 & 190351154{,}94 & 1837782795{,}03 \\ 398845{,}72 & 281992011{,}20 & 1837782795{,}03 & 17941575238{,}10 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 449056{,}07 \\ 316489875{,}21 \\ 2068600037{,}56 \\ 20157138073{,}89 \end{pmatrix}.$$

Значення для матриць беремо з таблиці 6.16 з рядка "Сума" і стовпців, а також стовпців $YX_1, YX_2, YX_3,$ враховуючи, що $n = 9$.

Потім знаходимо зворотну матрицю $(X'X)^{-1}$ і множимо її на матрицю $(X'Y)$.

У результаті одержуємо вектор-стовпець із знайденими параметрами регресійної моделі:

$$A = \begin{pmatrix} -1\,856{,}5 \\ -0{,}3 \\ 2{,}2 \\ 0{,}9 \end{pmatrix} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}.$$

Тепер можемо записати лінійну модель множинної регресії:

$$Y = -1\,856{,}5591 - 0{,}3949 \cdot x_1 + 2{,}2695 \cdot x_2 + 0{,}9385 \cdot x_3.$$

2. Коефіцієнт детермінації R^2 – так називають квадрат R – коефіцієнта множинної кореляції. Його шукаємо за формулою:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_{іррас}\right)^2}{\sum_{i=1}^{n}\left(y_i - y_{ср}\right)^2} = 1 - \frac{\sum_{i=1}^{n}\ell_i^2}{\sum_{i=1}^{n}\left(y_i - y_{ср}\right)^2}. \qquad (6.104)$$

$$R = \sqrt{R^2}.$$

Використовуючи таблицю 6.16, рядок "Сума" стовпців ℓ_i^2 і $\left(y_i - y_{ср}\right)^{\wedge 2}$

$R^2 = 0{,}9 > 0{,}8,$ отже, рівняння регресії достовірне. Коефіцієнт детермінації показує, що 95,95 % варіації залежної ознаки (Y) пояснюється включеними в модель факторами.

R = 0,9 – чим ближче до 1 коефіцієнт множинної кореляції, тим більше сильний зв'язок між Y і безліччю X.

Для подальших розрахунків скористаємося таблицею 6.16.

Таблиця 6.16

№ пор.	y	x_1	x_2	x_3	yx_1	yx_2	yx_3	y^2
1	**41 077**	**547**	**4 409**	**35 352**	22 463 982,6	181 111 030,7	1 452 146 320	1 687 332 006
2	**43 586**	**538**	**4 501**	**37 821**	23 467 103,3	196 178 843,3	1 648 465 111	1 899 747 852
3	**46 626**	**534**	**4 568**	**40 826**	24 915 046,8	212 983 288,5	1 903 517 772	2 173 955 527
4	**47 445**	**573**	**4 526**	**41 889**	27 195 300,9	214 756 622,6	1 987 428 130	2 250 999 368
5	**50 103**	**685**	**4 618**	**44 221**	34 342 781,6	231 370 327,6	2 215 611 219	2 510 300 488
6	**53 023**	**825**	**4 785**	**47 173**	4 372 2736	253 701 052,6	2 501 248 282	2 811 393 778
7	**55 294**	**833**	**4 775**	**49 894**	46 055 389,5	264 014 155,4	2 758 861 753	3 057 429 311
8	**55 157**	**845**	**4 615**	**50 011**	46 590 182,9	254 544 372,9	2 758 493 249	3 042 338 223
9	**56 746**	**841**	**4 581**	**51 658**	47 737 351,6	259 940 343,9	2 931 366 238	3 220 055 742
Сума	449 056	6 222	41 377	398 846	316 489 875	2 068 600 038	20 157 138 074	22 653 552 296

Продовж. табл. 6.16

№ пор.	x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3
1	299 070,1	194 39 686,63	1 249 741 562	2 411 188,21	19 332 881,48	155 867 200,9
2	289 883,2	20 258 550,9	1 430 419 947	2 423 347,49	20 363 071,36	170 229 948,3
3	285 543,8	20 866 057,57	1 666 722 185	2 440 937,05	21 815 641,47	186 488 393,9
4	328 558,2	20 488 858,24	1 754 718 649	2 594 568,02	24 010 982,3	189 610 605,3
5	469 834,8	21 325 028,12	1 955 516 121	3 165 318,52	30 311 212,77	204 209 295,3
6	679 975,1	22 894 062,23	2 225 317 214	3 945 553,35	38 899 359,83	225 713 426,2
7	693 752,4	22 798 065,67	2 489 450 253	3 976 960,23	41 557 936,36	238 232 345,3
8	713 479,2	21 297 052,81	2 501 130 528	3 898 077,08	42 243 398,23	230 795 816,5
9	707 706,6	20 983 792,77	2 668 558 779	3 853 617,63	43 457 527,39	236 635 763,2
Сума	4 467 803	190 351 155	17 941 575 238	28 709 568	281 992 011	1 837 782 795

Продовж. табл. 6.16

№ пор.	$\hat{y}$	$y-\hat{y}$	$(y_i-\hat{y}_i)(y_{i-1}-\hat{y}_{i-1})$	$(y-\hat{y})^2$
1	41 111,32	–34	–	1 167,917
2	43 640,58	–54	1 862,039	2 968,696
3	46 614,03	12	–635,821	136,177
4	47 502,98	–58	–680,085	3 396,432
5	49 854,61	248	–14 470	61 647,28
6	52 948,85	74	18 304,92	5 435,276
7	55 476,54	–183	–13 455,6	33 310,65
8	55 218,81	–61	11 208,92	3 771,764
9	56 688,35	57	–3 511,92	3 269,975
Сума	449 056	–0	–1 378	115 104

3. Проведемо відсівання неістотних факторів.

Для цього розрахуємо коефіцієнти парної кореляції за формулою:

$$r_{yx_i} = \frac{\sum yx_i - \frac{\sum y \cdot \sum x_i}{n}}{\sqrt{\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)\left(\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right)}}. \qquad (6.105)$$

Дані беремо з таблиці 6.16 рядка "Сума" відповідних стовпців:

$$r_{yx_1} = \frac{316\,489\,875 - \frac{449\,056 \cdot 6\,222}{9}}{\sqrt{\left(22\,653\,552\,296 - \frac{449\,056^2}{9}\right)\left(4\,467\,803 - \frac{6\,222^2}{9}\right)}} = 0{,}9.$$

Аналогічно

$$r_{yx_2} = 0{,}7.$$

$$r_{yx_3} = 0{,}9.$$

Набуті значення коефіцієнтів парної кореляції більше за модулем 0,3, отже, всі три фактори слід включити в модель, що розробляється.

4. Перевірка показників на мультиколінеарність.

Під мультиколінеарністю мають на увазі наявність висококореляційного зв'язку між двома факторами.

Для цього визначаємо парні коефіцієнти кореляції за формулами (6.106)-(6.108):

$$r_{x_1x_2} = \frac{\sum x_1x_2 - \frac{\sum x_1 \cdot \sum x_2}{n}}{\sqrt{\left(\sum x_1^2 - \frac{(\sum x_1)^2}{n}\right)\left(\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n}\right)}}; \qquad (6.106)$$

$$r_{x_1x_3} = \frac{\sum x_1x_3 - \frac{\sum x_1 \cdot \sum x_3}{n}}{\sqrt{\left(\sum x_1^2 - \frac{(\sum x_1)^2}{n}\right)\left(\sum x_3^2 - \frac{\left(\sum x_3\right)^2}{n}\right)}}; \qquad (6.107)$$

$$r_{x_2x_3} = \frac{\sum x_2x_3 - \frac{\sum x_2 \cdot \sum x_3}{n}}{\sqrt{\left(\sum x_2^2 - \frac{(\sum x_2)^2}{n}\right)\left(\sum x_3^2 - \frac{(\sum x_3)^2}{n}\right)}}. \qquad (6.108)$$

Маємо:

$$r_{x_1x_2} = \frac{28\ 709\ 568 - \frac{6\ 222 \cdot 41\ 377}{9}}{\sqrt{\left(4\ 467\ 803 - \frac{6\ 222^2}{9}\right)\left(190\ 351\ 155 - \frac{41\ 377^2}{9}\right)}} = 0{,}7.$$

Аналогічно

$$r_{x_1x_3} = 0{,}9.$$

$$r_{x_2x_3} = 0{,}7.$$

Теоретично вважається, що мультиколінеарність відсутня, якщо коефіцієнти парної кореляції менше 0,85 за модулем.

У нас коефіцієнт $r_{x_1x_3} = 0{,}940895$ більше 0,85, тому робимо висновок про наявність мультиколінеарності.

Оцінка будь-якої регресії потерпатиме від мультиколінеарності певною мірою, якщо тільки всі незалежні змінні не виявляться абсолютно некорельованими.

У нашому випадку на наявність мультиколінеарності може впливати недостатнє число спостережень. Можливо збільшення числа спостережень зменшить проблему мультиколінеарності.

5. Перевірка надійності впливу окремих факторів на результат.

Ця перевірка здійснюється за допомогою коефіцієнтів надійності:

$$m_{yx_i} = \frac{r_{yx_i} \cdot \sqrt{n}}{1 - \left(r_{yx_i}\right)^2}. \qquad (6.109)$$

$$m_{yx_1} = \frac{0{,}9421 \cdot \sqrt{9}}{1 - \left(0{,}9421\right)^2} = 25{,}1.$$

$$m_{yx_2} = \frac{0{,}751 \cdot \sqrt{9}}{1 - \left(0{,}751\right)^2} = 5{,}1.$$

$$m_{yx_3} = \frac{0{,}9992 \cdot \sqrt{9}}{1 - (0{,}9992)^2} = 1\ 916{,}66.$$

Оскільки всі коефіцієнти надійності за модулем більше 2,6, то всі дані фактори дійсно впливають на обсяг споживчих кредитів.

6. Розрахунок коефіцієнта кореляції.

Автокореляція – це кореляційна залежність між послідовними значеннями рівнів однієї і тієї ж ознаки.

Щоб виявити наявність автокореляції за часом у помилках використовують таку ідею: якщо кореляція є у помилок *E*, то вона присутня і в залишках, одержуваних після застосування методу МНК.

Лінійний коефіцієнт автокореляції першого порядку обчислюється за формулою:

$$r_1 = \frac{\sum \ell_i \cdot \ell_{i-1}}{\sqrt{\sum \ell_i^2 \cdot \sum \ell_{i-1}^2}}, \tag{6.110}$$

де $\ell_i = y_t - y_{расч}$ – залишки;

$y_{расч}$ – розрахункові значення y за рівнянням регресії;

ℓ_{i-1} – залишки, зсунуті на 1 крок.

Тоді (за даними табл. 6.16):

$$r_1 = \frac{-1378}{\sqrt{115\ 104 \cdot 113\ 936}} = -0{,}99.$$

У тому випадку, якщо коефіцієнт автокореляції першого порядку менше 0,5, то автокореляція відсутня, а якщо більше 0,5, то автокореляція присутня.

У нашому випадку $|r_1| > 0{,}5$, тому ми можемо стверджувати, що автокореляція присутня.

Перевіряючи гіпотезу про існування лінійної автокореляції першого порядку, виконують перевірочну процедуру, засновану на обчисленні *d*-статистики Дарбіна-Уотсона:

$$d = DW = \frac{\sum(\ell_i - \ell_{i-1})^2}{\sum \ell_i^2}, \tag{6.111}$$

d – зважена сума квадратів різниць послідовних залишків.

Маємо:

$$d = \frac{228\,525}{115\,104} = 1{,}99.$$

За таблицею значень d-статистики Дарбіна-Уотсона: рівень значущості $\alpha = 0{,}05\,\%$, при числі спостережень n = 9, числі незалежних змінних у рівнянні регресії R = 3 знаходимо два критичні значення: нижнє $d_L = 0{,}455$ (межа для визнання позитивної автокореляції залишків) і верхнє $d_U = 2{,}128$ (межа визнання її відсутності).

Тоді $4 - d_U = 1{,}872$; $4 - d_L = 3{,}545$.

Обчислене значення d = 1,98538 потрапляє в інтервал $4 - d_U < d < 4 - d_L$.

За таблицею 6.9 із завдання 1 одержуємо, що d потрапляє в область невизначеності (у разі передбачуваної негативної кореляції), тому гіпотеза про відсутність автокореляції не приймається і не відкидається.

7. Визначимо приватні коефіцієнти множинної кореляції для зміни щільності зв'язку між y і факторами x_1, x_2, x_3 за формулами (6.112-6.114):

$$R_{y(x_1;\,x_2)} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}}; \tag{6.112}$$

$$R_{y(x_1;\,x_3)} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_3}^2 - 2r_{yx_1} \cdot r_{yx_3} \cdot r_{x_1x_3}}{1 - r_{x_1x_3}^2}}; \tag{6.113}$$

$$R_{y(x_2;\,x_3)} = \sqrt{\frac{r_{yx_2}^2 + r_{yx_3}^2 - 2r_{yx_2} \cdot r_{yx_3} \cdot r_{x_2x_3}}{1 - r_{x_1x_3}^2}}. \tag{6.114}$$

Маємо:

$$R_{y(x_1;\,x_2)} = \sqrt{\frac{0{,}9421^2 + 0{,}751^3 - 2 \cdot 0{,}9421 \cdot 0{,}751 \cdot 0{,}748758}{1 - (0{,}748758)^2}} = 0{,}94.$$

Аналогічно

$$R_{y(x_1;\,x_3)} = 0{,}99. \; R_{y(x_2;\,x_3)} = 0{,}99.$$

8. Перевірка достовірності одержаної моделі здійснюється:

1) за допомогою розрахунку теоретичних значень за одержаним рівнянням $y_{расч}$.

   Якщо $\left.\begin{array}{l} \sum y_{расч} = 449\ 056 \\ \sum y = 449\ 056 \end{array}\right\}$, то рівняння достовірне;

2) за допомогою розрахунку коефіцієнта детермінації:

$$R^2 = 0{,}9995 > 0{,}8,$$

отже, рівняння достовірне і може бути використане для прогнозу.

Знайдемо скоректований коефіцієнт R^2 за формулою:

$$\overline{R}^2 = 1 - \left(1 - R^2\right) \cdot \frac{n-1}{n-m-1}, \qquad (6.115)$$

де n – число спостережень;
m – число незалежних змінних.

Маємо:

$$\overline{R}^2 = 1 - \left(1 - 0{,}9995\right) \cdot \frac{9-1}{9-3-1} = 0{,}99.$$

При невеликому числі спостережень величина, як правило, завищується.

9. Обчислимо стандартну помилку регресії:

$$S = \sqrt{\frac{\sum \ell_i^2}{n-2}}. \qquad (6.116)$$

$$S = \sqrt{\frac{115\ 104}{9-2}} = 151{,}73.$$

Знайдемо відношення стандартної помилки до середнього значення залежної змінної:

$$V = \frac{S}{\overline{y}} = \frac{151{,}73}{49\ 895} = 0{,}003, \text{ або } 0{,}304\ \%.$$

Це відношення служить критерієм прогнозних якостей оціненої регресійної моделі.

Алгоритм використання інструмента КОРРЕЛЯЦИЯ

Інструмент аналізу **Корреляция** (який також поставляється разом з надбудовою **Пакет анализа**) використовується для оцінки ступеня залежності між двома наборами даних. Щоб застосувати інструмент **Корреляция**, виконайте ряд дій.

1. Виберіть команду **Сервис / Анализ данных**.

2. У діалоговому вікні **Анализ данных**, що відкрилося, у списку **Инструменты анализа** оберіть **пункт Корреляция** і клацніть на кнопці **ОК**.

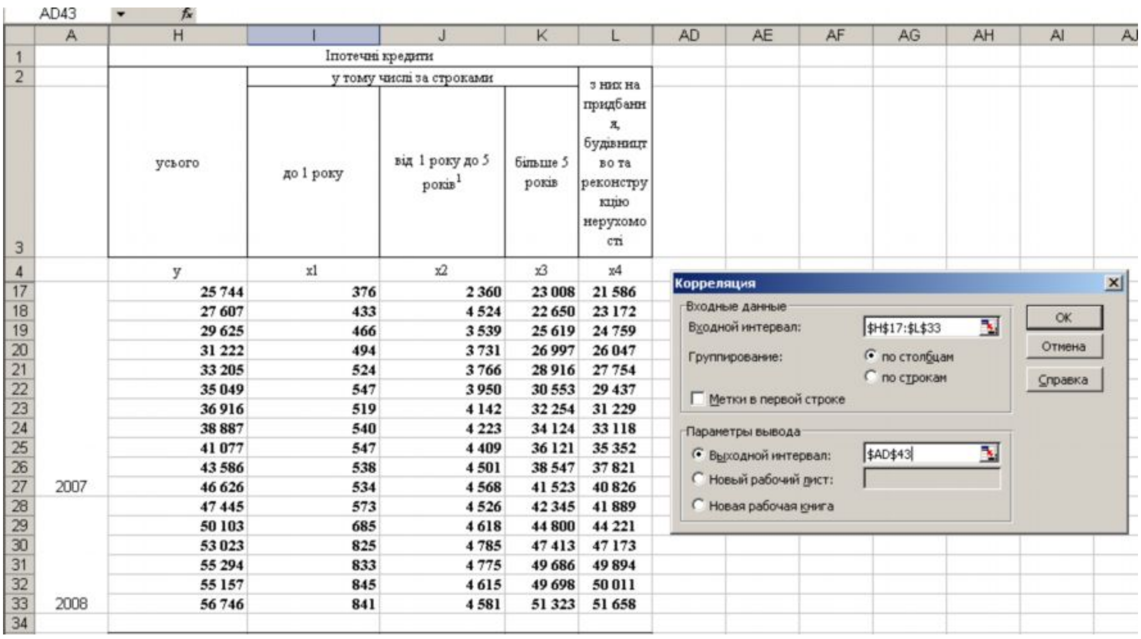
Excel відобразить на екрані діалогове вікно **Корреляция**.

3. Визначте значення *X* і *У*, які повинні бути проаналізовані.

У полі **Входной интервал** вкажіть посилання на комірки. Якщо цей діапазон містить підписи даних, встановіть прапорець **Метки в** першому рядку. Перевірте, чи правильно вибраний перемикач **Группирование**, що визначає спосіб організації даних у виділеному діапазоні комірок.

4. Вкажіть місце, куди повинні бути поміщені результати обчислень, що проводяться.

Використовуйте перемикачі групи **Параметры выводов**, щоб вказати *Excel*, де повинен бути розміщений звіт про одержані результати. Наприклад, щоб помістити цей звіт на тому ж листі, де розташовані початкові дані, виберіть перемикач **Выходной интервал** і в полі праворуч від нього вкажіть адресу комірок, які повинні містити обчислені значення. Щоб розташувати звіт в якому-небудь іншому місці, виберіть один з двох інших перемикачів.



	A	H	I	J	K	L
1		Іпотечні кредити				
2		усього	у тому числі за строками			з них на придбання, будівництво та реконструкцію нерухомості
3			до 1 року	від 1 року до 5 років[1]	більше 5 років	
4		y	x1	x2	x3	x4
17		25 744	376	2 360	23 008	21 586
18		27 607	433	4 524	22 650	23 172
19		29 625	466	3 539	25 619	24 759
20		31 222	494	3 731	26 997	26 047
21		33 205	524	3 766	28 916	27 754
22		35 049	547	3 950	30 553	29 437
23		36 916	519	4 142	32 254	31 229
24		38 887	540	4 223	34 124	33 118
25		41 077	547	4 409	36 121	35 352
26		43 586	538	4 501	38 547	37 821
27	2007	46 626	534	4 568	41 523	40 826
28		47 445	573	4 526	42 345	41 889
29		50 103	685	4 618	44 800	44 221
30		53 023	825	4 785	47 413	47 173
31		55 294	833	4 775	49 686	49 894
32		55 157	845	4 615	49 698	50 011
33	2008	56 746	841	4 581	51 323	51 658
34						

5. Клацніть на кнопці ОК.

Excel обчислить коефіцієнт кореляції для вказаних вами даних і розташувати його в заданому місці. Значення цього коефіцієнта, наприклад, дорівнює числу 0,91802024, що означає, що майже на 92 % обсяги депозитів залежать від кількості виходів реклами.

	Столбец 1	*Столбец 2*	*Столбец 3*	*Столбец 4*	*Столбец 5*
Столбец 1	1				
Столбец 2	0,91802024	1			
Столбец 3	0,76095654	0,66251471	1		
Столбец 4	0,99919175	0,91708731	0,73456426	1	
Столбец 5	0,99906638	0,9238193	0,74934722	0,99883724	1

Побудова моделі множинної регресії

Вхідні дані для побудови моделі множинної регресії лінійного виду та у вигляді функції Кобба-Дугласа наведені нижче.

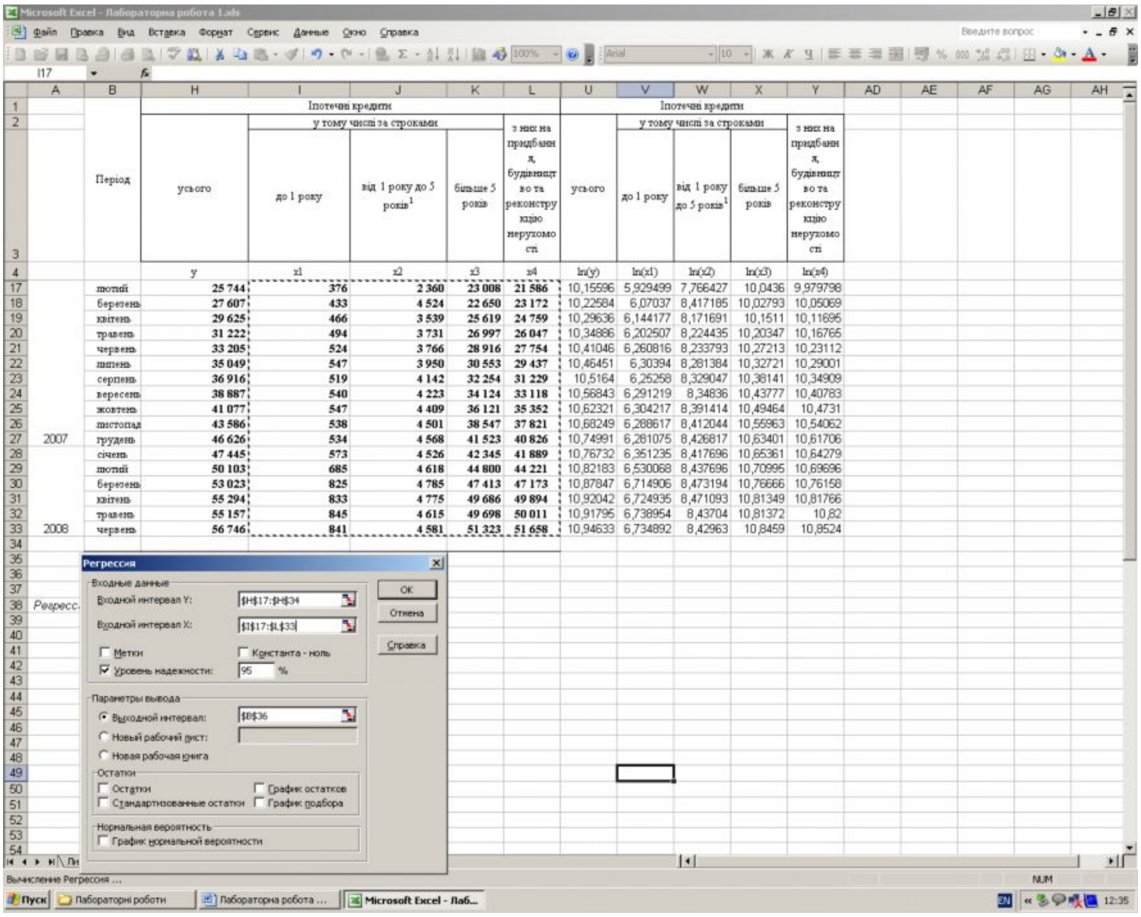
	Період	Іпотечні кредити					Іпотечні кредити				
			у тому числі за строками					у тому числі за строками			
		усього	до 1 року	від 1 року до 5 років[1]	більше 5 років	з них на придбання, будівництво та реконструкцію нерухомості	усього	до 1 року	від 1 року до 5 років[1]	більше 5 років	з них на придбання, будівництво та реконструкцію нерухомості
		y	x1	x2	x3	x4	ln(y)	ln(x1)	ln(x2)	ln(x3)	ln(x4)
	лютий	**25 744**	**376**	**2 360**	**23 008**	**21 586**	10,15596	5,929499	7,766427	10,0436	9,979798
	березень	**27 607**	**433**	**4 524**	**22 650**	**23 172**	10,22584	6,07037	8,417185	10,02793	10,05069
	квітень	**29 625**	**466**	**3 539**	**25 619**	**24 759**	10,29636	6,144177	8,171691	10,1511	10,11695
	травень	**31 222**	**494**	**3 731**	**26 997**	**26 047**	10,34886	6,202507	8,224435	10,20347	10,16765
	червень	**33 205**	**524**	**3 766**	**28 916**	**27 754**	10,41046	6,260816	8,233793	10,27213	10,23112
	липень	**35 049**	**547**	**3 950**	**30 553**	**29 437**	10,46451	6,30394	8,281384	10,32721	10,29001
	серпень	**36 916**	**519**	**4 142**	**32 254**	**31 229**	10,5164	6,25258	8,329047	10,38141	10,34909
	вересень	**38 887**	**540**	**4 223**	**34 124**	**33 118**	10,56843	6,291219	8,34836	10,43777	10,40783
	жовтень	**41 077**	**547**	**4 409**	**36 121**	**35 352**	10,62321	6,304217	8,391414	10,49464	10,4731
	листопад	**43 586**	**538**	**4 501**	**38 547**	**37 821**	10,68249	6,288617	8,412044	10,55963	10,54062
2007	грудень	**46 626**	**534**	**4 568**	**41 523**	**40 826**	10,74991	6,281075	8,426817	10,63401	10,61706
	січень	**47 445**	**573**	**4 526**	**42 345**	**41 889**	10,76732	6,351235	8,417696	10,65361	10,64279
	лютий	**50 103**	**685**	**4 618**	**44 800**	**44 221**	10,82183	6,530068	8,437696	10,70995	10,69696
	березень	**53 023**	**825**	**4 785**	**47 413**	**47 173**	10,87847	6,714906	8,473194	10,76666	10,76158
	квітень	**55 294**	**833**	**4 775**	**49 686**	**49 894**	10,92042	6,724935	8,471093	10,81349	10,81766
	травень	**55 157**	**845**	**4 615**	**49 698**	**50 011**	10,91795	6,738954	8,43704	10,81372	10,82
2008	червень	**56 746**	**841**	**4 581**	**51 323**	**51 658**	10,94633	6,734892	8,42963	10,8459	10,8524

1. Обираємо команду **Сервис** / **Анализ данных**.

2. У діалоговому вікні **Анализ данных**, що відкрилося, у списку **Инструменты анализа** вибираємо пункт **Регрессия** і клацаємо на кнопці **ОК**. На екрані відображається діалогове вікно **Регрессия**.

3. Визначте значення *X* і *Y*.

У полі **Входной интервал** *У* вкажіть посилання на діапазон комірок, в яких міститься набір залежних значень (обсяг продажів). Потім перейдіть до поля **Входной интервал** *X*. Переконайтеся, наприклад, що дані про розміри торгової площі та кількості персоналу розташовуються в сусідніх колонках. Позначте блок, що складається з цих обох колонок, як значення *X*. У цьому полягає єдина відмінність введення даних для лінійного і множинного регресійного аналізів. Потім аналіз проводиться аналогічно попередньому, але коли з'являються його результати, ми бачимо два коефіцієнти при X.



Отримаємо такі результати:

- лінійна модель

$$y = -0{,}0005 + 0{,}99x_1 + x_2 + 0{,}99x_3 + 1{,}7 \cdot 10^7 x_4;$$

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	1
R-квадрат	1
Нормированный R-к	1
Стандартная ошибк	0,000641771
Наблюдения	17

Дисперсионный анализ

	df	SS	MS	F	Значимость F
Регрессия	4	1,71E+09	427034006,4	1,03682E+15	4,10784E-87
Остаток	12	4,94E-06	4,1187E-07		
Итого	16	1,71E+09			

	Коэффициенты	Зартная о	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	-0,000564043	0,00193	-0,292263766	0,775076503	-0,00476896	0,00364087	-0,004768956	0,00364087
Переменная X 1	0,999995474	3,04E-06	329094,1727	4,17405E-61	0,999988853	1,000002095	0,999988853	1,000002095
Переменная X 2	1,000000285	4,62E-07	2163996,453	6,38735E-71	0,999999278	1,000001292	0,999999278	1,000001292
Переменная X 3	0,999999883	4,14E-07	2413421,219	1,72505E-71	0,99999898	1,000000785	0,99999898	1,000000785
Переменная X 4	1,71214E-07	4,29E-07	0,398635525	0,697163904	-7,6458E-07	1,10701E-06	-7,64585E-07	1,10701E-06

- модель у вигляді функції Кобба-Дугласа

$$y = e^{0,5} x_1^{\ 0,01} x_2^{\ 0,09} x_3^{\ 0,6} x_4^{\ 0,2}.$$

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,999997
R-квадрат	0,999994
Нормированный R-квадрат	0,999991
Стандартная ошибка	0,000755
Наблюдения	17

Дисперсионный анализ

	df	SS	MS	F	Значимость F
Регрессия	4	1,06244345	0,265610863	465886,7359	4,99027E-31
Остаток	12	6,84143E-06	5,70119E-07		
Итого	16	1,062450291			

	Коэффициент	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	рхние 95,0%
Y-пересечение	0,569353	0,015491	36,75378012	1,0553E-13	0,535600818	0,603104797	0,535600818	0,603105
Переменная X 1	0,014276	0,002037997	7,005003518	1,42434E-05	0,00983576	0,018716586	0,00983576	0,018717
Переменная X 2	0,097637	0,00221116	44,1566773	1,18467E-14	0,092819795	0,102455204	0,092819795	0,102455
Переменная X 3	0,665531	0,015065878	44,17474358	1,1789E-14	0,632705554	0,698357009	0,632705554	0,698357
Переменная X 4	0,206308	0,015569578	13,25071681	1,59044E-08	0,172384875	0,240231268	0,172384875	0,240231

6. Для оцінки суттєвості лінійного коефіцієнта кореляції використаємо критерій Стьюдента, що розраховується за формулою:

$$t = r\sqrt{\frac{n-2}{1-r^2}}. \tag{6.117}$$

Проведемо порівняння розрахованого значення критерію Стьюдента з його критичним значенням. Критичне значення знаходимо за таблицями розподілу Стьюдента для заданого рівня ймовірності та числа ступенів свободи $v = n - 2$, де n – число спостережень.

За умови, якщо $t_{факт} > t_{кр}$, лінійний коефіцієнт кореляції можна вважати статистично значущим. Для заданих умов $t_{крит}\left(0,05;\, 8\right) = 2,307.$

Оцінка суттєвості рівняння регресії проводиться за критерієм Фішера:

$$F = \frac{r^2}{1-r^2} \cdot (n-2).$$

Якщо розрахункове значення критерію Фішера перевищує його критичне значення, модель можна вважати адекватною.

$$F_{крит}\left(0,05; 1; 8\right) = 5,32.$$

Оскільки коефіцієнт детермінації для обох рівнянь суттєво не відрізняється і приблизно дорівнює 1, F-критерій Фішера показує адекватність обох рівнянь, а t-критерій Стьюдента підтверджує значущість лише коефіцієнтів при x_2, x_3, x_4 для лінійної моделі та значущість усіх коефіцієнтів для моделі у вигляді функції Кобба-Дугласа, тому більш прийнятною вважається модель у вигляді функції Кобба-Дугласа. ■

ЗАВДАННЯ ДЛЯ САМОСТІЙНОГО ОПРАЦЮВАННЯ

Завдання 1. На базі статистичних даних (показників кредитів, наданих домашнім господарствам $X(i)$ за місяцями 2007-2008 рр. залежно від строку) за варіантами (табл. 6.17) побудувати графік тренда змінної $X(i)$, обрати форму однофакторної моделі, оцінити усі її параметри, визначити зони надійності при рівні значущості $a = 95$.

Завдання 2. На базі статистичних даних (показників кредитів, наданих домашнім господарствам $X(i)$ за місяцями 2007-2008 рр. залежно від строку) за варіантами (табл. 6.17) обрати форму багатофакторної моделі, оцінити усі її параметри, визначити зони надійності при рівні значущості $a = 95$. Оцінити коефіцієнти детермінації, автокореляції та перевірити показники на мультиколінеарність між факторами.

Таблиця 6.17

Варіанти завдань

№ варіанта	Завдання 1 (№ графи табл. 6.11)	Завдання 2 (№ графи табл. 6.11)	
		результативна ознака	факторна ознака
1	3	7	8, 9, 10
2	4	11	12, 13, 14
3	5	16	17, 18, 19
4	6	7	8, 9, 10
5	7	11	12, 13, 14
6	8	16	17, 18, 19
7	9	7	8, 9, 10
8	10	11	12, 13, 14
9	11	16	17, 18, 19
10	12	7	8, 9, 10

ХРЕСТОМАТІЯ

Аллен Р.Дж. Экономические индексы. – М., 1980.

С целью обоснования построения индексов американский экономист и статистик Ирвинг Фишер (1867-1947) разработал серию тестов (1925), которым должны удовлетворять формулы индексов. Тем самым была заложена *тестовая теория индексов*, которая является основной в зарубежной статистике. И. Фишер выдвинул две основные идеи: с помощью тестов выявляются погрешности индексных формул, тесты позволяют построить новые формулы индексов, лишенные погрешностей и наиболее пригодные для измерения динамики цен и количеств. Основные тесты, предложенные И. Фишером.

1. Тест обратимости во времени – индексы, исчисленные в прямом и обратном направлениях, должны быть взаимно обратными числами. Например, если формула показывает, что уровень цен в отчетном периоде по сравнению с базисным повысился в два раза, то она должна также отражать, что в базисном периоде цены были вполовину ниже, чем в отчетном, т.е. $I_{rs} \cdot I_{sr} = 1$, где r и s – сравниваемые периоды.
2. Циркулярный тест, или тест кружного испытания – произведение последовательных цепных индексов должно равняться соответствующему базисному индексу: $I_{rs} \cdot I_{st} = I_{rt}$.
3. Тест обратимости по факторам – переход от индекса одного из факторов-сомножителей к индексу другого фактора-сомножителя может быть выполнен путем перестановки их мест в исходном индексе; произведение индексов этих двух факторов-сомножителей должно равняться индексу произведения этих двух факторов:

$$\frac{\sum_{i=1}^{n} p_{1i}q_{1i}}{\sum_{i=1}^{n} p_{0i}q_{1i}} \cdot \frac{\sum_{i=1}^{n} p_{0i}q_{1i}}{\sum_{i=1}^{n} p_{0i}q_{0i}} = \frac{\sum_{i=1}^{n} p_{1i}q_{1i}}{\sum_{i=1}^{n} p_{0i}q_{0i}}$$

или

$$\frac{\sum_{i=1}^{n} p_{1i}q_{0i}}{\sum_{i=1}^{n} p_{0i}q_{0i}} \cdot \frac{\sum_{i=1}^{n} p_{0i}q_{1i}}{\sum_{i=1}^{n} p_{1i}q_{0i}} = \frac{\sum_{i=1}^{n} p_{1i}q_{1i}}{\sum_{i=1}^{n} p_{0i}q_{0i}}.$$

И. Фишер считал, что тесты позволяют отобрать и сконструировать “идеальный” индекс, отвечающий серии тестов. Однако выяснилось, что с помощью системы тестов невозможно создать индекс, одновременно отвечающий требованиям тестов обратимости во времени и циркулярности; первый и второй тесты, логически равноправные, носят альтернативный характер.

В дальнейшем американские экономисты Т.Л. Келли (1923) и Б. Мюджетт (1951) интерпретировали индексы с позиции теории выборки. Совокупность индивидуальных индексов рассматривалась ими как выборочная совокупность, а сводный индекс – как выборочная средняя. При этом предлагалось рассчитывать вероятностную ошибку сводного индекса. Однако эта теория игнорирует то обстоятельство, что совокупность товаров-представителей, по которой рассчитываются индексы цен, вряд ли можно рассматривать как выборочную, так как принципы ее формирования не имеют ничего общего со случайной выборкой.

Стохастическая теория индексов не имела такого распространения, как тестовая. Знаменитый экономист Джон Мейнард Кейнс (1883-1946) назвал ее “в корне ложной”.

ПРАКТИЧНІ ЗАВДАННЯ

Завдання 1. На основі наведених розрахунків визначити, як факторна ознака впливає на результативну. Визначити:

1) залежність за допомогою лінійної функції, її параметри та пояснити їх зміст;
2) оцінити напрям і щільність зв’язку за допомогою лінійного коефіцієнта кореляції;
3) перевірити зв’язок на суттєвість з рівнем ймовірності 0,95;
4) довести рівність показників щільності зв’язку;
5) побудувати графік залежності з нанесенням лінії тренда.

Задача 1

Таблиця 6.18

Залежність між вартістю квартири та площею

№	Вартість квартири, ум. од.	Площа квартири, м^2
1	28 000	64
2	24 000	61
3	28 000	62
4	34 000	62
5	32 000	70
6	27 000	61
7	25 000	45
8	24 000	62
9	31 000	62
10	25 000	60

Задача 2

Таблиця 6.19

Залежність між обсягом банківських операцій і стажем роботи працівника банку

№ працівника	Обсяг банківських операцій, ум. од.	Стаж, років
1	40,4	6,0
2	40,6	5,0
3	39,5	6,0
4	39,0	4,0
5	38,9	4,5
6	38,6	3,0
7	39,2	5,0
8	39,1	5,5
9	39,9	6,0
10	39,6	5,4

Завдання 2. На основі даних про розміри вкладів в ощадні установи міським і сільським населенням визначити:

1) чи існує зв'язок між місцем проживання вкладника та розміром його вкладу;
2) чи існує зв'язок між часткою вкладу розміром 25 000 і більше грн. і місцем проживання вкладника.

Таблиця 6.20

Розміри вкладів в ощадні установи міським і сільським населенням, грн.

№ вкладника	Розмір вкладу групи населення, грн.	
	міське (X_1)	сільське (X_2)
1	18 000	7 000
2	25 000	25 000
3	16 000	19 000
4	10 000	2 000
5	12 000	4 000
6	25 000	–

Продовж. табл. 6.20

№ вкладника	Розмір вкладу групи населення, грн.	
	міське (X_1)	сільське (X_2)
7	10 000	–
8	26 000	–
9	15 000	–
10	20 000	–

Завдання 3. З метою визначення середнього віку працівників банку проведено безповторний 15 % відбір. З ймовірністю 0,99 визначити: середній вік працівників банку у вибірковій сукупності та довірчі межі для генеральної середньої.

Таблиця 6.21

Розподіл працівників банку за віком

Вік, років	20-30	30-40	40-50	50-60
Кількість працівників, чол.	8	29	12	3

Завдання 4. Рада директорів банку запланувала інвестувати прибуток за 7 пріоритетними напрямками, беручи до уваги очікувану прибутковість і ступінь ризику кожного проекту. За допомогою коефіцієнта рангової кореляції перевірити істотність зв'язку з ймовірністю 0,95. Зробити висновки.

Таблиця 6.22

Інвестування прибутку банку за напрямками

Напрям інвестування	Очікувана прибутковість, ум. од.	Ступінь ризику
A	20,3	2
B	30,6	5
C	18,5	1
D	32,8	6
E	21,3	3
F	44,5	9
G	40,8	8

Завдання 5. Рада директорів банку запланувала інвестувати прибуток за 7 пріоритетними напрямками, беручи до уваги очікувану прибутковість і ступінь ризику кожного проекту. За допомогою коефіцієнта рангової кореляції перевірити істотність зв'язку з ймовірністю 0,95. Зробити висновки.

Таблиця 6.23

Інвестування прибутку банку за напрямками

Напрям інвестування	Очікувана прибутковість, ум. од.	Ступінь ризику
A	20,3	2
B	30,6	5
C	18,5	1
D	32,8	6
E	21,3	3
F	44,5	9
G	40,8	8

Завдання 6. Якість засвоєння студентами лекційного матеріалу залежить від форми організації занять. Розрахувати коефіцієнти контингенції, асоціації та відношення перехресних шансів, зробити висновки за розрахунками.

Таблиця 6.24

Залежність якості засвоєння студентами лекційного матеріалу від форми організації занять

Форма організації лекції	Якість засвоєння матеріалу		
	висока	задовільна	всього
З наочною демонстрацією	45	31	
Традиційна	20	100	
Всього			

КОНТРОЛЬНІ ПИТАННЯ

1. Яке спостереження називається вибірковим?
2. У чому переваги вибіркового спостереження перед суцільним?
3. Чому при вибірковому спостереженні неминучі помилки і як вони класифікуються?
4. Які умови правильного відбору одиниць сукупності при вибірковому спостереженні?
5. Як виробляються власне випадковий, механічний, стратифікований і серійний відбори?

6. У чому відмінність повторної та безповторної вибірки?
7. Чим є середня помилка вибірки (для середньої та частки)?
8. За якими розрахунковими формулами знаходять середні помилки вибірки (для середньої та частки) при повторному і безповторних відборах?
9. Що характеризує гранична помилка вибірки і за якими формулами вона обчислюється (для середньої та частки)?
10. Якими способами здійснюється розповсюдження результатів вибіркового спостереження на всю сукупність?
11. Навіщо і як обчислюються граничні статистичні помилки вибірки (для середньої та частки)?
12. За якими формулами визначається необхідна чисельність вибірки, що забезпечує з визначеною вірогідністю задану точність спостереження?

ТЕСТИ

1. Вибірковий метод застосовується в тих випадках, коли:
   а) проведення суцільного спостереження неможливе;
   б) проведення суцільного спостереження економічно недоцільне;
   в) всі варіанти правильні.

2. Назвати вид відбору, який припускає, що кожна одиниця, що потрапила у вибірку, або серія повертається в генеральну сукупність і має шанс удруге потрапити у вибірку:
   а) повторний;
   б) безповторний;
   в) індивідуальний;
   г) всі варіанти неправильні.

3. Емпірична функція розподілу:
   а) функція $F(x)$, що визначає для кожного значення x відносну частоту події $X < x$;
   б) функція розподілу $F(x)$ генеральної сукупності;
   в) всі варіанти правильні.

4. Назвати спосіб відбору, при якому здійснюється добір одиниць з неоднорідної сукупності. У цьому випадку генеральну сукупність попередньо розбивають на однорідні групи за допомогою типологічного угруповання, після чого роблять добір одиниць з кожної групи у вибіркову сукупність випадковим або механічним способом:
   а) власне випадковий;
   б) механічний;
   в) стратифікований;
   г) всі варіанти неправильні.

5. Вкажіть властивості функції розподілу:
   а) значення емпіричної функції належать відрізку [0, 1];
   б) $F(x)$ – неспадаюча функція;
   в) якщо x_1 – найменша варіанта, то $F(x)=0$ при $x \leq x_1$, якщо x_k – *найбільша* варіанта, то $F(x)=1$ при $x > x_k$;
   г) всі варіанти правильні.

6. Задачі статистичного аналізу моделі:
   а) оцінка точності точкових оцінок, коефіцієнтів регресії;
   б) визначення статистичної значущості оцінок;
   в) побудова довірчих інтервалів оцінок;
   г) перевірка адекватності моделі;
   ґ) всі варіанти правильні.

7. Коефіцієнт детермінації $R^2 = 1 - \frac{\sum (y-\hat{y})^2}{\sum (y-\bar{y})^2}$ використовується для:
   а) перевірки адекватності регресійної моделі;
   б) перевірки значущості регресійної моделі;
   в) перевірки значущості коефіцієнтів регресійної моделі;
   г) всі варіанти правильні.

8. До етапів побудови регресійної моделі включають:
   а) визначення мети дослідження, збір вихідних даних;
   б) аналіз даних і відбір факторів;
   в) оцінка параметрів регресійної моделі;
   г) перевірка якості моделі;
   ґ) всі перераховані варіанти правильні.

9. Вказати умови Гаусса-Маркова для однофакторної моделі:
   а) $D[e_i] \to \min$ – для усіх спостережень, тобто дисперсія випадкового члена має бути постійною для всіх спостережень;
   б) $M[e_i]=0$ – для всіх спостережень, тобто математичне сподівання випадкового члена у будь-якому спостереженні має дорівнювати нулю;
   в) систематичний зв’язок між значеннями випадкового члена у будь-яких двох спостереженнях;
   г) випадковий член має бути розподілений незалежно від пояснюючих змінних;
   ґ) всі варіанти правильні.

10. Міра чутливості функціонально пов'язаних величин визначається як співвідношення процентних змін залежної та незалежної змінних $E = \lim\limits_{\Delta x \to 0} \frac{\Delta x}{\Delta y} \cdot \frac{\Delta y}{\Delta x} \cdot 100\ \%$:
   а) еластичність;
   б) коефіцієнт коваріації;
   в) коефіцієнт детермінації;
   г) всі варіанти неправильні.

11. Коефіцієнти частинної кореляції дають додаткову інформацію про ступінь щільності зв'язку між досліджуваними показниками:
   а) у разі наявності колінеарності між факторами в модель включають один з факторів, той, у якого частинний коефіцієнт кореляції з результативною ознакою Y більше;
   б) серед факторів, для яких $\left| r_{x_i x_j} \right| < 0{,}8,$ вибирають для розрахунку в модель, насамперед ті, для яких парний і частинний коефіцієнти щільності зв'язку з Y більше;
   в) всі варіанти правильні.

12. Наявності гетероскедастичності можна перевірити:
   а) за допомогою тесту рангової кореляції Спірмена;
   б) за допомогою тесту Голдфелда-Квандта;
   в) за допомогою тесту Глейзера;
   г) всі варіанти правильні.

13. Якщо критерій Дарбіна-Уотсона $0 \le DW \le d_l$, то:
   а) в залишках є автокореляція (позитивна);
   б) в залишках є автокореляція (негативна);
   в) автокореляція в залишках відсутня;
   г) область невизначеності на основі наявних даних не можна встановити однозначну відповідь.

14. Зв'язок вважається підтвердженим, якщо:
   а) коефіцієнт контингенції менше коефіцієнта асоціації;
   б) $K_k \ge 0{,}3;\ K_a \ge 0{,}5;$
   в) $K_k \le 0{,}5;\ K_a \ge 0{,}3;$
   г) правильні відповіді а) і б);
   ґ) правильні відповіді а) і в).

Тема 7. ДОСЛІДЖЕННЯ ІНТЕНСИВНОСТІ ДИНАМІКИ ТА ТЕНДЕНЦІЙ РОЗВИТКУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ЯВИЩ

1. ПОНЯТТЯ ТА ВИДИ ДИНАМІЧНИХ РЯДІВ.

   Динамічний ряд, його компоненти, види динамічних рядів. Порівнянність рівнів ряду та змикання динамічних рядів. Характеристика причин непорівнянності рівнів динамічного ряду.

2. ХАРАКТЕРИСТИКИ ІНТЕНСИВНОСТІ ДИНАМІКИ.

   Показники динамічних змін: ланцюгові та базисні. Середні показники динамічних змін: сутність, застосування, розрахунок.

3. АНАЛІЗ ТЕНДЕНЦІЙ РОЗВИТКУ.

   Коливання та їх вплив на динамічний ряд. Поняття тенденції розвитку, види тенденцій та їх перевірка. Характеристика методів перевірки ряду на наявність тренда. Методи вирівнювання. Дослідження сезонних коливань. Прогнозування та інтерполяція. Характеристика методів прогнозування.

Терміни

Динамічний ряд
Абсолютний приріст
Коефіцієнт зростання

Темп зростання
Темп приросту
Абсолютне значення 1 % приросту

Абсолютне прискорення
Відносне прискорення
Середній абсолютний приріст

Середній темп зростання
Середній темп приросту

Тренд
Циклічні коливання
Сезонні коливання

Тенденція середнього рівня
Тенденція дисперсії
Тенденція автокореляції
Вирівнювання

Інтерполяція
Екстраполяція

Перелік всіх термінів, наведених у даній темі,
див. у “Предметному покажчику”.

1. ПОНЯТТЯ ТА ВИДИ ДИНАМІЧНИХ РЯДІВ

"...правила и образцы, которым последуя при обрабатывании каждого особенного Статистического предмета, можно бы было сделать полезнейшее изображение в рассуждении Государственного Хозяйства и Политики"

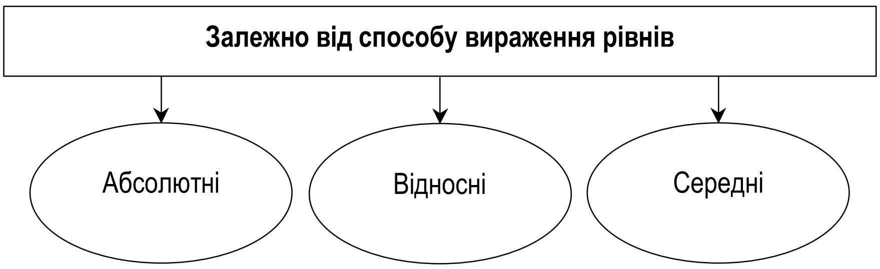
К.Ф. Герман [20, с. 25]

Динамічний ряд ряд значень статистичного показника, що змінюється в часі, розташованих у хронологічному порядку.

У результаті розрахунків Арсеньєв отримав такий динамічний ряд кількості населення Росії [20, с. 51-52]:

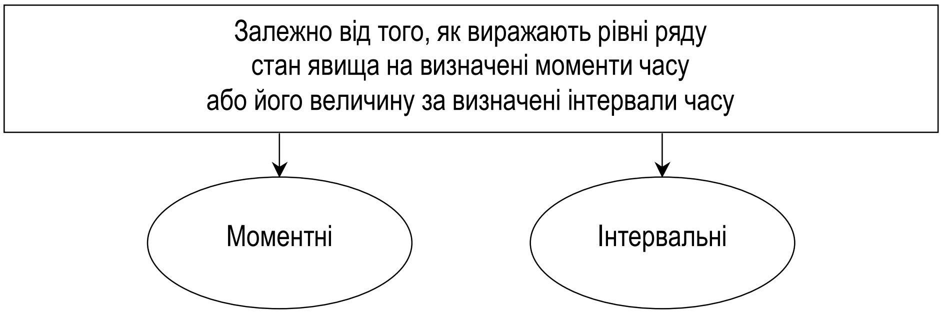
1722 р. – 14 млн. чол.
1742 р. – 16 млн. чол.
1762 р. – 19 млн. чол.
1782 р. – 28 млн. чол.
1796 р. – 36 млн. чол.
1812 р. – 45 млн. чол.
1816 р. – менше 45 млн. чол.

Він вважав причиною зменшення кількості населення війну 1812 р. приблизно в 1 млн. чол. Розрахунками вченого досі користуються історики.



▷ **ПРИКЛАД** видів динамічних рядів залежно від способу вираження рівнів.

Абсолютні	обсяги депозитів фізичних та юридичних осіб за місяцями поточного року.
Відносні	темпи зростання обсягів депозитів фізичних та юридичних осіб за місяцями поточного року.
Середні	середні плинні обсяги депозитів фізичних та юридичних осіб за місяцями поточного року. ■



▷ **ПРИКЛАД** видів динамічних рядів залежно від того, як виражають рівні ряду стан явища на визначені моменти часу або його величину за визначені інтервали часу.

Моментні	обсяги кредитів, наданих фізичним та юридичним особам на початок місяця.
Інтервальні	обсяги кредитів, наданих фізичним та юридичним особам за II квартал 2008 року. ■

▷ **ПРИКЛАД** видів динамічних рядів залежно від відстані між рівнями.

З рівновіддаленими рівнями	обсяги депозитів фізичних та юридичних осіб за місяцями поточного року.
З нерівновіддаленими рівнями	обсяги депозитів фізичних та юридичних осіб за період з 2002 по 2007 рік та січень-червень 2008 року. ■

Залежно від наявності основної тенденції досліджуваного процесу

- Стаціонарні
- Нестаціонарні

Найважливішою умовою правильної побудови ряду динаміки є порівнянність усіх рівнів, що входять у нього.

Причини непорівнянності рівнів ряду

- зміни одиниць виміру або одиниць рахунку
- методологія обліку або розрахунку показників
- умовою порівнянності рівнів ряду динаміки є періодизація динаміки (процес виділення однорідних етапів розвитку рядів динаміки)
- важливо також, щоб у ряді динаміки інтервали або моменти, за якими визначені рівні, мали однаковий економічний зміст
- умовою порівнянності рівнів інтервального ряду є наявність рівних інтервалів
- рівні ряду динаміки можуть виявитися непорівнянними за кількістю охоплюваних об'єктів унаслідок переходу ряду об'єктів з одного підпорядкування в інше
- непорівнянність рівнів ряду може виникнути внаслідок змін територіальних границь областей, районів тощо

Для того щоб привести рівні ряду динаміки до порівнянного вигляду, іноді доводиться вдаватися до прийому, що називається *"змикання рядів динаміки"*.

Змикання рядів динаміки — об'єднання в один ряд (більш довгий) двох або декількох рядів динаміки, рівні яких обчислені за різною методологією або різними територіальними границями.

Етапи змикання рядів динаміки

- спочатку приводять до порівнянного вигляду абсолютні величини, а потім відносні
- усі показники приводять до бази, до загальної підстави

2. ХАРАКТЕРИСТИКИ ІНТЕНСИВНОСТІ ДИНАМІКИ

Абсолютний приріст характеризує розмір збільшення або зменшення рівня ряду за визначений проміжок часу. Він дорівнює різниці двох порівнюваних рівнів і виражає абсолютну швидкість зростання.

$$\Delta_{y_i} = Y_i - Y_{i-1}. \quad (7.1)$$

▷ **ПРИКЛАД** розрахунку ланцюгового абсолютного приросту.

Таблиця 7.1

Обсяги депозитів юридичних осіб, млн. грн.

Рік	Усього	Абсолютний приріст	
		розрахунок	значення
2002	190 350	–	–
2003	330 115	= 330 115 – 190 350	139 765
2004	563 502	= 563 502 – 330 115	233 387
2005	740 778	= 740 778 – 563 502	177 276
2006	988 860	= 988 860 – 740 778	248 082

■

Коефіцієнт зростання показник інтенсивності зміни рівнів ряду показує, у скільки разів даний рівень ряду більше базисного рівня (якщо цей коефіцієнт більше одиниці) або яку частину базисного рівня становить рівень поточного періоду за деякий проміжок часу

$$R_p = \frac{Y_i}{Y_{i-1}}. \quad (7.2)$$

▷ **ПРИКЛАД** розрахунку ланцюгового коефіцієнта зростання.

Таблиця 7.2

Обсяги депозитів юридичних осіб, млн. грн.

Рік	Усього	Коефіцієнт зростання	
		розрахунок	значення
2002	190 350	–	–
2003	330 115	= 330 115 / 190 350	1,73
2004	563 502	= 563 502 / 330 115	1,71
2005	740 778	= 740 778 / 563 502	1,31
2006	988 860	= 988 860 / 740 778	1,33

■

Темп зростання показує, у скільки разів рівень ряду більше (менше) рівня, взятого за базу порівняння, однак виражається у відсотках.

$$T_{\text{р}} = \frac{Y_i}{Y_{i-1}} \cdot 100\,\%. \qquad (7.3)$$

▷ **ПРИКЛАД** розрахунку ланцюгового темпу зростання.

Таблиця 7.3

Обсяги депозитів юридичних осіб

Рік	Усього, млн. грн.	Темп зростання, %	
		розрахунок	значення
2002	190 350	–	–
2003	330 115	= 330 115 / 190 350 · 100	173,43
2004	563 502	= 563 502 / 330 115 · 100	170,70
2005	740 778	= 740 778 / 563 502 · 100	131,46
2006	988 860	= 988 860 / 740 778 · 100	133,49

■

Темп приросту характеризує відносну швидкість зміни рівня ряду в одиницю часу. Показує, на яку частку (або відсоток) рівень даного періоду або моменту часу більше або менше базисного рівня.

$$T_{\text{р}} = \frac{\Delta}{Y_{i-1}} \cdot 100\,\% = \frac{Y_i - Y_{i-1}}{Y_{i-1}} \cdot 100\,\% = T_p - 100\,\%. \qquad (7.4)$$

▷ **ПРИКЛАД** розрахунку ланцюгового темпу приросту.

Таблиця 7.4

Обсяги депозитів юридичних осіб

Рік	Усього, млн. грн.	Темп приросту, %	
		розрахунок	значення
2002	190 350	–	–
2003	330 115	= 330 115 / 190 350 · 100 – 100	73,43
2004	563 502	= 563 502 / 330 115 · 100 – 100	70,70
2005	740 778	= 740 778 / 563 502 · 100 – 100	31,46
2006	988 860	= 988 860 / 740 778 · 100 – 100	33,49

■

Абсолютне значення 1 % приросту являє собою одну соту частину базисного рівня й, у той же час, – відношення абсолютного приросту до відповідного темпу зростання:

$$|\%| = \frac{\Delta_{i/i-1}}{T_{\text{p}.i/i-1(\%)}} = \frac{Y_i - Y_{i-1}}{\frac{Y_i - Y_{i-1}}{Y_{i-1}} \cdot 100\,\%} = \frac{Y_{i-1}}{100}. \qquad (7.5)$$

▷ **ПРИКЛАД** розрахунку абсолютного значення 1 % приросту.

Таблиця 7.5

Обсяги депозитів юридичних осіб, млн. грн.

Рік	Усього	Абсолютне значення 1 % приросту	
		розрахунок	значення
2002	190 350	–	–
2003	330 115	= 190 350 / 100	1 904
2004	563 502	= 330 115 / 100	3 301
2005	740 778	= 563 502 / 100	5 635
2006	988 860	= 740 778 / 100	7 408

■

Абсолютне прискорення різниця між поточним і попереднім абсолютними приростами показує, на скільки дана швидкість більше (менше) попередньої.

$$\Delta' = \Delta_{y_i} - \Delta_{y_{i-1}}. \qquad (7.6)$$

▷ **ПРИКЛАД** розрахунку абсолютного прискорення.

Таблиця 7.6

Обсяги депозитів юридичних осіб, млн. грн.

Рік	Усього	Абсолютний приріст	Абсолютне прискорення	
		значення	розрахунок	значення
2002	190 350	–	–	–
2003	330 115	139 765	–	–
2004	563 502	233 387	= 233 387 – 139 765	93 622
2005	740 778	177 276	= 177 276 – 233 387	–56 111
2006	988 860	248 082	= 248 082 – 177 276	70 806

■

Відносне прискорення — це відношення прискорення до абсолютного приросту, взятого за базу, тобто відносне прискорення є темпом приросту абсолютного приросту. Обчислюється лише в тому випадку, коли абсолютний приріст, взятий за базу порівняння, число позитивне:

$$\frac{\Delta'}{\Delta y_i} = \frac{\text{абсолютне прискорення}}{\text{абсолютний приріст, прийнятий за базу}}. \qquad (7.7)$$

▷ **ПРИКЛАД** розрахунку відносного прискорення.

Таблиця 7.7

Обсяги депозитів юридичних осіб

Рік	Усього	Абсолютний приріст, млн. грн.	Абсолютне прискорення, млн. грн.	Відносне прискорення, %	
		значення	значення	розрахунок	значення
2002	190 350	–	–	–	–
2003	330 115	139 765	–	–	–
2004	563 502	233 387	93 622	= 93 622 / 139 765	0,67
2005	740 778	177 276	–56 111	= –56 111 / 139 765	–0,40
2006	988 860	248 082	70 806	= 70 806 / 139 765	0,51

■

Середній абсолютний приріст дає можливість установити, наскільки в середньому за одиницю часу повинен збільшуватися рівень ряду (в абсолютному вираженні), щоб, відправляючись від початкового рівня за дане число періодів (наприклад, років), досягти кінцевого рівня

$$\overline{\Delta} = \frac{y_n - y_1}{n-1}. \qquad (7.8)$$

▷ **ПРИКЛАД** розрахунку середнього абсолютного приросту.

$$\overline{\Delta} = \frac{y_n - y_1}{n-1} = \frac{988\,860 - 190\,350}{5-1} = 199\,628 \text{ млн. грн.}$$ ■

Середній коефіцієнт зростання показує, у скільки разів у середньому за одиницю часу змінився рівень динамічного ряду протягом аналізованого періоду.

$$\overline{K_p} = \sqrt[m]{K_{2/1} K_{3/3} \ldots K_{n/n-1}}. \qquad (7.9)$$

▷ **ПРИКЛАД** розрахунку середнього коефіцієнта зростання.

$$\overline{K_p} = \sqrt[5-1]{\frac{988\,860}{190\,350}} = 1{,}51. \ ■$$

Середній темп зростання показує, у скільки разів (на скільки відсотків) у середньому за одиницю часу змінився рівень динамічного ряду.

$$\overline{T_з} = \sqrt[m]{K_{2/1}K_{3/2}......K_{n/n-1}} \cdot 100\,\% = \sqrt[n-1]{\frac{y_n}{y_1}}100. \qquad (7.10)$$

▷ **ПРИКЛАД** розрахунку середнього темпу зростання.

$$\overline{T_з} = \sqrt[n-1]{\frac{y_n}{y_1}}100 = 1{,}5097 \cdot 100\,\% = 150{,}97\,\%.$$

Середній темп приросту

$$\overline{T_{пр}} = \overline{Tp} - 100. \qquad (7.11)$$

▷ **ПРИКЛАД** розрахунку середнього темпу приросту.

$$\overline{T_{пр}} = \overline{Tp} - 100 = 150{,}97\,\% - 100\,\% = 50{,}97\,\%.$$

Середній рівень ряду динаміки

для моментного ряду обчислюється за середньою хронологічною:

- з рівновіддаленими рівнями розраховується за формулою середньої хронологічної простої

$$y_{ср} = \frac{\frac{y_1}{2} + y_2 + y_2 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}; \qquad (7.12)$$

- з нерівновіддаленими рівнями визначається за формулою середньої хронологічної зваженої

$$\bar{y} = \frac{(y_1+y_2)\cdot t_1 + (y_2+y_3)\cdot t_2 + \ldots + (y_{n-1}+y_n)\cdot t_{n-1}}{2(t_1+t_2+\ldots+t_{n-1})} = \frac{\sum(y_i+y_{i+1})\cdot t_{n-1}}{2\sum t_{n-1}}, \qquad (7.13)$$

де y – рівні рядів динаміки;

t – тривалість інтервалу часу між рівнями;

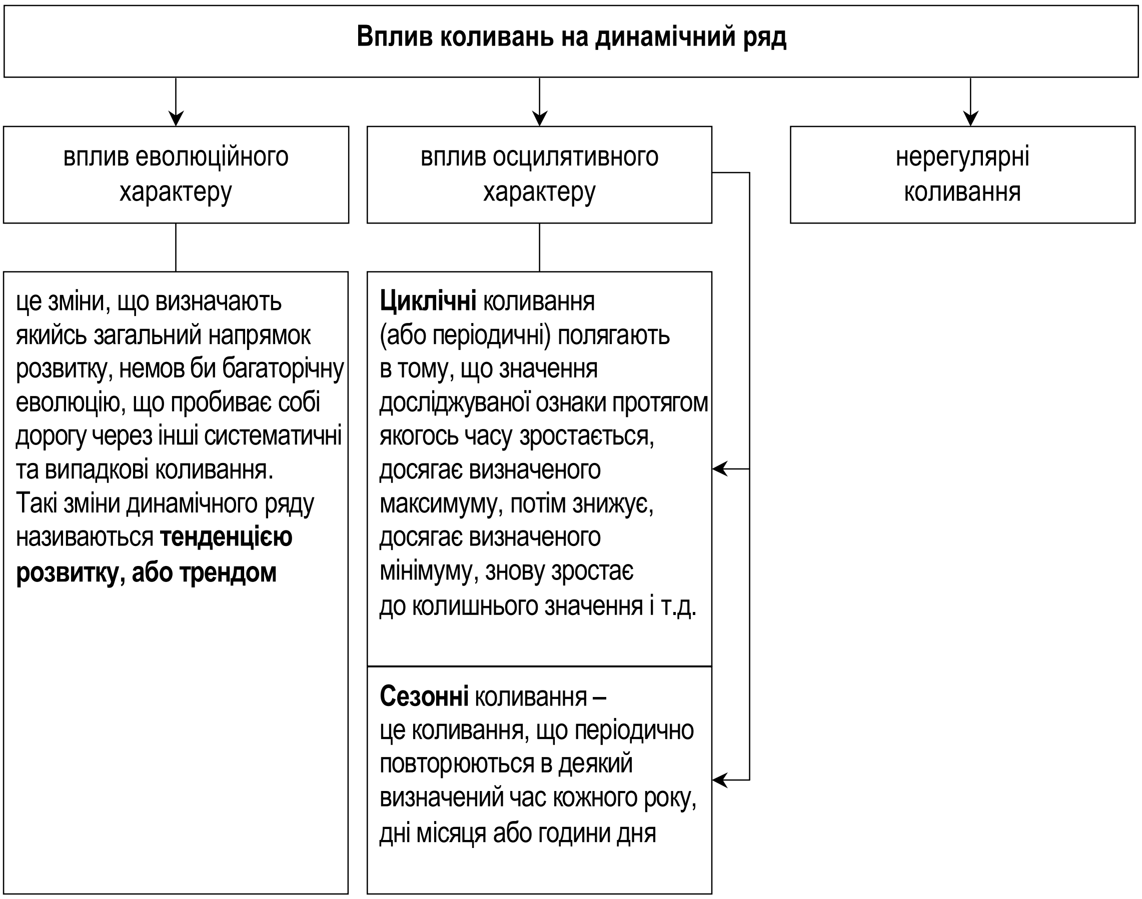
для інтервального ряду обчислюється за середньою арифметичною.

3. АНАЛІЗ ТЕНДЕНЦІЙ РОЗВИТКУ

"В сколько-нибудь демократических и свободных государствах возможна сносная правительственная статистика. У нас об этом говорить не приходится. Наша правительственная статистика плоха, нелепо раздроблена между "ведомствами", недостоверна и поздно выходит в свет"

К. Маркс, Ф. Энгельс [84, с. 392]

Ряд динаміки може зазнавати впливу факторів *еволюційного й осцилятивного характеру*, а також знаходитися під впливом факторів різного впливу.



Тренд	довгострокова компонента ряду динаміки. Вона характеризує основну тенденцію його розвитку, при цьому інші компоненти розглядаються тільки як такі, що заважають процедурі його визначення.

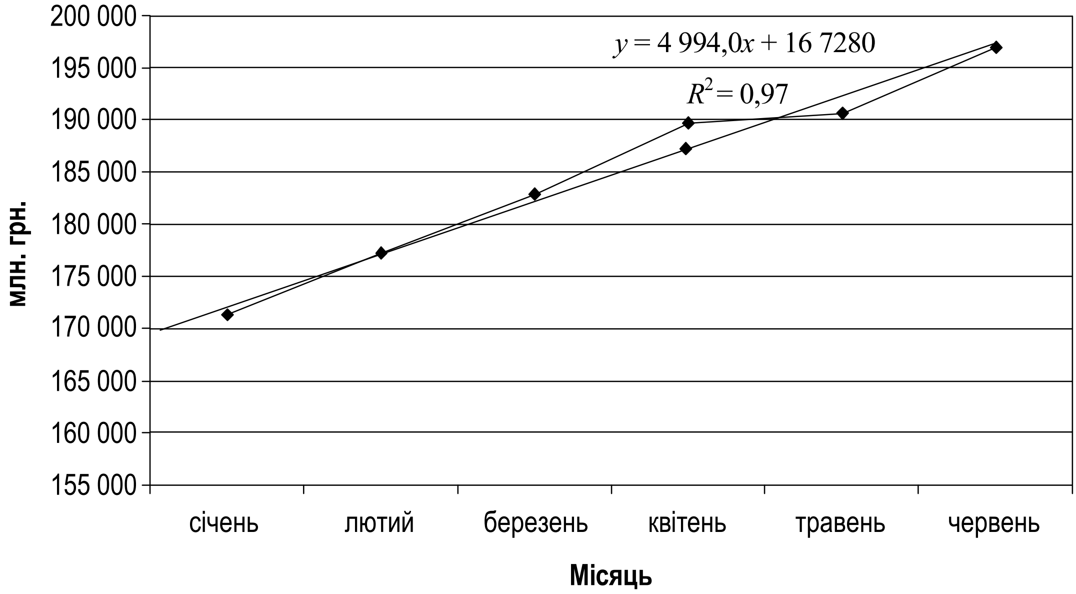
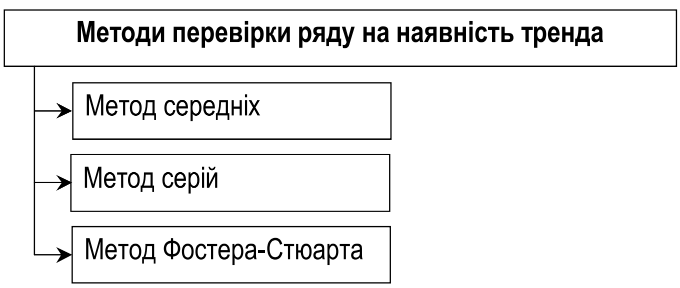


Рис. 7.1. Депозити домашніх господарств за січень-червень 2008 р.

Перш ніж перейти до виділення тренда, варто перевірити гіпотезу про те, чи існує він узагалі, відсутність тренда говорить про те, що середній рівень ряду в часі залишається незмінним.



Метод середніх Ряд розбивається звичайно на два інтервали, для кожного визначається $\overline{Y_1}, \overline{Y_2}$. Висувається гіпотеза про істотне розходження середніх.

За основу перевірки береться критерій Стьюдента. У випадку, якщо розрахункове значення виявляється більше або дорівнює табличному значенню цього критерію при заданому рівні імовірності помилки, то гіпотеза про відсутність тренда відкидається, в іншому випадку – гіпотеза приймається. У випадку рівності або при несуттєвому розходженні двох досліджуваних сукупностей розраховується відношення середніх:

$$t = \frac{\overline{Y_1} - \overline{Y_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}; \sigma = \sqrt{\frac{(n_1-1)^2\sigma_1^2 + (n_2-1)^2\sigma_2^2}{n_1+n_2-2}}; \sigma_{1\ \text{або}\ 2}^{2} = \frac{\sum\left(y_i - \overline{y_{1\ \text{або}\ 2}}\right)^2}{(n-1)}, \quad (7.14)$$

де t – критерій Стьюдента;

$\overline{Y_1}, \overline{Y_2}$ – середнє для першої та другої половин ряду динаміки;

σ – середньоквадратичне відхилення різниці середніх;

n_1, n_2 – число спостережень у частинах ряду.

Після того як установлена наявність тенденції в ряді динаміки, відбувається її опис за допомогою методів згладжування (вирівнювання).

Вирівнювання визначення основної тенденції розвитку, що виявляється в часі, досліджуваних явищ.

Методи виділення тренда

- Укрупнення інтервалів (середня ступенева)
- Середня плинна
- Аналітичне вирівнювання

Укрупнення інтервалів (середня ступенева) ряд динаміки розподіляють на деяку досить велику кількість однакових інтервалів. Якщо середні рівні по інтервалах не дозволяють побачити тенденцію, то переходять до розрахунку рівнів за ще більші проміжки часу, збільшуючи довжину кожного інтервалу.

Середня плинна вихідні рівні ряду заміняються середніми величинами, що одержують з даного рівня і декількох симетрично навколишніх. Ціле число рівнів, за якими розраховується середнє значення, називають інтервалом згладжування. Найчастіше беруть непарні інтервали (3,5 і т.д. точок). При непарному згладжуванні отримане середнє арифметичне закріплюють за серединою розрахункового інтервалу. Недолік цієї методики – в умовному визначенні згладжених рівнів для точок на початку і в кінці ряду.

▷ **ПРИКЛАД** розрахунку середньої плинної.

Таблиця 7.8

Обсяги депозитів юридичних осіб, млн. грн.

Рік	Усього	Середня плинна	
		розрахунок	значення
2002	190 350	–	–
2003	330 115	= (190 350 + 330 115 + 563 502) / 3	361 322
2004	563 502	= (330 115 + 563 502 + 740 778) / 3	544 798
2005	740 778	= (563 502 + 740 778 + 988 860) / 3	764 380
2006	988 860	–	–

■

Аналітичне вирівнювання визначення основної тенденції розвитку, що виявляється в часі, досліджуваного явища (тобто розвиток явища немов залежить тільки від часу). Відхилення конкретних рівнянь ряду від рівнів, що відповідають загальній тенденції, пояснюється дією факторів, що виявляються випадково або циклічно.

Трендовая модель:

$$Y_t = f(t) + E_t, \qquad (7.15)$$

де $f(t)$ – рівень, обумовлений тенденцією розвитку;

E_t – випадкове і циклічне відхилення від тенденції.

Екстраполяція продовження виявленої тенденції за межі ряду (коли закономірність розвитку, що діє в минулому (усередині ряду динаміки), зберігається й у прогнозованому майбутньому);

перспективна екстраполяція, проведена на майбутнє.

ретроспективна екстраполяція, проведена на минуле.

Інтерполяція визначення відсутнього рівня усередині відомого інтервалу.

Доугерти К. Введение в эконометрику: учебник / англ. – М.: Инфра-М, 1997. – 402 с.

Метод Фостера-Стюарта, кроме определения наличия тенденции явления, позволяет обнаружить тренд дисперсии уровней ряда динамики, что важно знать при анализе и прогнозировании экономических явлений. Расчет состоит из следующих этапов.

1. Сравнивается каждый уровень ряда со всеми предыдущими, при этом:

если $y_i > y_{i-1}; y_{i-2} \ldots y_i$, то $U_i = 1; e_i = 0$;
при $y_i < y_{i-1}; y_{i-2} \ldots y_i$, то $U_i = 0; e_i = 1$.

2. Вычисляются значения величин S и d:

$$S = \sum_{i=2}^{n} S_i\, ; d = \sum_{i=2}^{n} d_i\, ,$$

где $S_i = U_i + e_i; d_i = U_i - e_i$.

Нетрудно заметить, что величина S может принимать значения $0 \le S \le n-1$, причем S = 0, когда все уровни ряда равны между собой, и $S = n-1$, когда ряд динамики монотонно убывает или возрастает. Показатель S характеризует тенденцию изменения дисперсии ряда динамики.

Показатель d имеет нижний предел, равный $(n-1)$, и верхний – $(n-1)$. В первом случае ряд является монотонно убывающим, во втором – монотонно возрастающим. Кроме того, показатель d может быть равен нулю:

- если все уровни ряда равны между собой, тогда $\sum U_i = \sum e_i$. (Данное условие выполняется для ряда, который в первой половине является монотонно убывающим, а во второй – монотонно возрастающим);
- если уровни подъема и спада чередуются, причем каждое следующее значение уровня подъема (спада) должно быть больше (меньше) всех последующих.

Перечисленные случаи, при которых показатель d = 0, представляют лишь теоретический интерес, и вероятность их использования при проведении практических расчетов крайне незначительна. Показатель d характеризует изменение тенденций в среднем.

Оба показателя S и d асимптотически нормальны и имеют независимые распределения.

3. Проверяется с использованием t-критерия Стьюдента гипотеза о том, можно ли считать случайными разности $S-\mu$ и $d-0$:

$$t_s = \frac{S-\mu}{\sigma_1};\ t_d = \frac{d-0}{\sigma_2},$$

где μ – среднее значение величины S, определенное для ряда, в котором уровни расположены случайным образом;

σ_1 – стандартная ошибка величины S;

σ_2 – стандартная ошибка величины d.

4. Сравниваются расчетные значения t_s и t_d с табличным при заданном уровне значимости. Если $t_s < t_{табл}$ и $t_в < t_{табл}$, то гипотеза об отсутствии тренда в средней и дисперсии подтверждается.

Теория статистики: учебник / под ред. проф. Р.А. Шмойловой. – М.: Финансы и статистика, 1996. – 464 с.

Метод взвешенной скользящей средней. Взвешенная скользящая средняя отличается от простой скользящей средней тем, что уровни, входящие в интервал усреднения, суммируются с различными весами. Это связано с тем, что аппроксимация сглаживаемого ряда динамики в пределах интервала сглаживания осуществляется с использованием уровней, рассчитанных по полиному $\overline{y_i} = a_0 + a_i \cdot i + a_2 \cdot i^2 + \ldots$ (здесь i – порядковый номер уровня в интервале сглаживания). Полином первого порядка $\overline{y_i} = a_0 + a_i \cdot i$ есть уравнение прямой, следовательно, метод простой скользящей средней является частным случаем метода взвешенной скользящей средней. Коэффициенты полиномов находятся по способу наименьших квадратов.

На первом этапе сглаживания определяются интервал сглаживания и порядок аппроксимирующего полинома – параболы. Считается, что при использовании полиномов высоких степеней и при меньших размерах интервалов сглаживание ряда динамики будет более “гибким”.

Центральная ордината параболы принимается за сглаженное значение соответствующего фактическим данным уровня. Поскольку отсчет времени в пределах интервала сглаживания производится от его середины, т.е. $(t = i)$ $i = \ldots,-2,\ -1,\ 0,\ 1,\ 2,\ldots$, то сглаженное значение уровня равно параметру a_0 подобранной параболы и является соответствующей скользящей средней. Поэтому для сглаживания нет необходимости прибегать к процедуре подбора системы парабол, так как величину a_0 можно получить как взвешенную среднюю из m уровней.

Например, если в интервал сглаживания входят пять последовательных уровней ряда со сдвигом во времени на один шаг, а выравнивание проводится по полиному второго порядка, то коэффициенты полинома находятся из условия

$$\sum_{t=1}^{\sigma}(y - a_0 - a_1 \cdot i - a_2 \cdot i^2) \to \min.$$

Учитывая, что для нечетных $k \sum t^k = 0$, приходим к системе:

$$\begin{cases} \sum y - 5a_0 - a_2 \sum t^2 = 0 \\ \sum yt - a_1 \sum t^2 = 0 \\ \sum yt^2 - a_0 \sum t^2 - a_2 \sum t^4 = 0. \end{cases}$$

Для определения a_0 необходимо найти значения $\sum_{-p}^{p} i^2$ и $\sum_{-p}^{p} i^4$. Так как интервал сглаживания равен $m = 5$, тогда $\sum_{-p}^{p} i^2 = 10$ и $\sum_{-p}^{p} i^4 = 34$.

Нормальные уравнения, определяющие a_0 и a_2, в этом случае записываются так:

$$\begin{cases} 5a_0 + 10a_2 = \sum_{-p}^{p} yi \\ 10a_0 + 34a_2 = \sum_{-p}^{p} yi^2. \end{cases}$$

Решение этой системы относительно a_0 может быть представлено следующим образом:

$$a_0 = \frac{34\sum y - 10\sum y^2}{5 \cdot 34 - 10^2} = \frac{17\sum y - 5\sum yi^2}{35} =$$

$$= \frac{1}{35}(-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{y+2}).$$

Аналогичным путем получим выражения и для других интервалов сглаживания по параболе второго и третьего порядка. Так, например, для

$$\begin{cases} m=7: \overline{y_t} = \frac{1}{21}(-2y_{t-3} + 3y_{t-2} + 6y_{t-1} + 7y_t + 6y_{t+1} + 3y_{t+2} - 2y_{t+3}) \\ m=9: \overline{y_t} = \frac{1}{231}(-21y_{t-4} + 14y_{t-3} + 39y_{t-2} + 5y_{t-1} + 59y_t + 54y_{t+1} + 39y_{t+2} + 14y_{t+3} - 21y_{t+4}). \end{cases}$$

Согласно приведенным формулам веса симметричны относительно центрального уровня (y_t), и их сумма с учетом общего множителя, вынесенного за скобки, равна единице.

Выравнивание ряда динамики с помощью метода конечных разностей

Этот метод заключается в следующем. Пусть ряд динамики y_t описывается полиномом p-й степени. Для полинома p-й степени вычислим первые разности:

$$\Delta_t^{(1)} = y_{t+1} - y_t,$$

вторые разности:

$$\Delta_t^{(2)} = \Delta_{t+1}^{(1)} - \Delta_t^{(1)}$$

и т.д. Общая формула p-й разности:

$$\Delta_t^{(p)} = y_p + py_{p-1} + \frac{p(p-1)}{2!}y_{p-2} - \frac{p(p-1)(p-2)}{3!}y_{p-3} + ... + (-1)^p y_0.$$

Любой член y_i (i = 0, 1, 2, 3... n) ряда динамики можно выразить через начальный уровень ряда y_0 и конечные разности:

$$y_1 = y_0 + \Delta_0^{(1)}; y_2 = y_0 + \Delta_0^{(1)} + \Delta_1^{(1)},$$

но $\Delta_1^{(1)} = \Delta_0^{(1)} + \Delta_0^{(2)}$, поэтому $y_2 = y_0 + 2\Delta_0^{(1)} + \Delta_0^{(2)}$ и т.д.

Отсюда получаем:

$$\overline{y_i} = y_0 + t\Delta_0^{(1)} + \frac{t(t-1)}{2!}\Delta_0^{(2)} + \frac{t(t-1)(t-2)}{3!}\Delta_0^{(3)} + ... + \Delta_0^{(t)}.$$

Если первые разности не равны, но варьируют с незначительными отклонениями друг от друга, а средняя арифметическая вторых разностей настолько мала, что ею можно пренебречь, то первые разности можно считать практически равными.

Окончательная формула для расчета уровней ряда динамики при равных или почти равных первых разностях:

$$\overline{y_t} = \overline{y} + \overline{\Delta}^{(1)} \cdot t^1.$$

Если анализируя вторые разности, мы придем к выводу, что они практически равны, то, вычисляя коэффициенты параболы второго порядка, получаем тренд ряда динамики:

$$\overline{y_t} = \overline{y} - \frac{n^2 - 1}{24}\overline{\Delta}^{(2)} + \overline{\Delta}^{(1)} \cdot t^1 + \frac{\overline{\Delta}^{(2)}}{2} \cdot t^2,$$

где $\overline{y_t}$ – выравненное значение ряда динамики;
$\overline{y}$ – средний уровень ряда динамики;
$\overline{\Delta}^{(1)}$ – средняя арифметическая первых разностей;
$\overline{\Delta}^{(2)}$ – средняя арифметическая вторых разностей;
n – число уровней;
t^1 – независимая переменная (время).

ПРАКТИЧНІ ЗАВДАННЯ

Завдання 1. На основі наведених даних у таблиці 7.9 провести розрахунки ланцюгових тібазисних абсолютних приростів, коефіцієнтів зростання, темпів приросту, абсолютного значення одного відсотка приросту, а також середнього абсолютного приросту, середнього коефіцієнта зростання, середнього темпу приросту та середнього рівня ряду.

Таблиця 7.9

Обсяги депозитів домашніх господарств у розрізі строків погашення, млн. грн.

Рік	Усього	у тому числі за строками		
		на вимогу	до 1 року	від 1 до 2 років
2002	19 699	5 379	7 146	7 174
2003	33 115	7 925	11 545	13 644
2004	42 502	9 832	14 986	17 684
2005	74 778	18 660	19 025	37 093
2006	108 860	25 940	22 853	60 066

Завдання 2. На основі наведених даних у таблиці 7.10 провести розрахунки ланцюгових і базисних абсолютних приростів, коефіцієнтів зростання, темпів приросту, абсолютного значення одного відсотка приросту, а також середнього абсолютного приросту, середнього коефіцієнта зростання, середнього темпу приросту та середнього рівня ряду.

Таблиця 7.10

Обсяги кредитів, наданих домашнім господарствам у розрізі строків погашення за місяцями 2007 року, млн. грн.

Місяць	Усього	Кредити на поточні потреби			
		усього	у тому числі за строками		
			до 1 року	від 1 до 5 років	більше 5 років
Січень	83 489	56 789	12 626	44 162	…
Лютий	87 181	59 633	12 887	29 496	17 251
Березень	93 262	63 695	13 503	30 522	19 670
Квітень	99 339	67 736	14 058	31 661	22 017
Травень	105 060	71 694	14 714	32 663	24 317
Червень	111 819	76 293	15 097	34 091	27 105
Липень	120 244	82 655	17 084	35 763	29 809
Серпень	127 244	87 501	17 539	37 425	32 536
Вересень	134 011	92 189	17 891	38 575	35 724
Жовтень	141 712	97 513	18 684	40 444	38 385
Листопад	150 108	103 171	19 435	42 008	41 729
Грудень	160 386	110 121	19 990	44 593	45 538

Завдання 3. Маємо дані про основні засоби комерційного банку у фактичних цінах за 2000, 2003 та 2006 роки. Визначити середньорічні темпи приросту за вказані періоди, знайти відсутні рівні ряду методами середнього абсолютного приросту та середнього коефіцієнта зростання.

Таблиця 7.11

Основні засоби банку "Хрещатик" у фактичних цінах за 2000, 2003 та 2006 роки

Основні засоби, млрд. грн.	Рік		
	2000	2003	2006
	850	896	987

Таблиця 7.12

**Основні засоби “Альфа-банку”
у фактичних цінах за 2000, 2003 та 2006 роки**

Основні засоби, млрд. грн.	Рік		
	2000	2003	2006
	1 500	1 655	1 847

КОНТРОЛЬНІ ПИТАННЯ

1. Дати визначення ряду динаміки. З яких елементів він складається і яке їх значення?
2. Які існують види рядів динаміки?
3. Які причини виникнення непорівнянності динамічних рядів?
4. Що характеризують показники абсолютного приросту і як вони обчислюються?
5. Що є темпом зростання? Як він обчислюється?
6. Який існує взаємозв’язок між послідовними ланцюговими коефіцієнтами зростання і базисним коефіцієнтом зростання за відповідний період? Яке практичне застосування цього взаємозв’язку?
7. Що показує абсолютне значення одного відсотка приросту і як воно обчислюється?
8. Як розраховується середній абсолютний приріст? Його економічне значення.
9. За якою формулою обчислюється середній темп зростання? Його економічне значення.
10. Як обчислюється середній темп приросту? Його економічне значення.
11. Якими найпоширенішими статистичними методами здійснюється вивчення тренда у рядах динаміки?
12. Як здійснюється згладжування рядів динаміки способом ковзаючої (рухомої) середньої? Переваги і недоліки цього методу.
13. У чому суть методу аналітичного вирівнювання динамічних рядів?
14. Що таке екстраполяція рядів динаміки?

ТЕСТИ

1. Ряд значень статистичного показника, що змінюється в часі, розташованих у хронологічному порядку:
   а) динамічний ряд;
   в) варіаційний ряд;
   г) тренд.

2. Вказати види динамічних рядів залежно від способу вираження рівнів:
   а) абсолютні;
   б) моментні;
   в) середні;
   г) з рівновіддаленими рівнями.

3. Причини непорівнянності рівнів ряду:
   а) зміни одиниць виміру або одиниць рахунку;
   б) умовою порівнянності рівнів ряду динаміки є періодизація динаміки (процес виділення однорідних етапів розвитку рядів динаміки);
   в) рівні ряду динаміки можуть виявитися непорівнянними за колом охоплюваних об'єктів унаслідок переходу ряду об'єктів з одного підпорядкування в інше;
   г) всі відповіді правильні.

4. Який показник характеризує розмір збільшення або зменшення рівня ряду за визначений проміжок часу:
   а) абсолютний приріст;
   б) темп зростання;
   в) абсолютне прискорення;
   г) всі відповіді неправильні.

5. Абсолютне значення 1 % приросту:
   а) показник інтенсивності зміни рівнів ряду показує, у скільки разів даний рівень ряду більше базисного рівня (якщо цей коефіцієнт більше одиниці) або яку частину базисного рівня становить рівень поточного періоду за деякий проміжок часу;
   б) являє собою одну соту частину базисного рівня й, у той же час, – відношення абсолютного приросту до відповідного темпу зростання;
   в) різниця між поточним і попереднім абсолютними приростами показує, на скільки дана швидкість більше (менше) попередньої.

6. Який показник відображує відношення прискорення до абсолютного приросту, прийнятого за базу:
   а) коефіцієнт зростання;
   б) темп приросту;
   в) відносне прискорення;
   г) середній темп зростання.

7. Середній рівень моментного ряду динаміки обчислюється за:
   а) середньою хронологічною;
   б) середньою арифметичною;
   в) середньою геометричною;
   г) всі відповіді неправильні.

8. Сезонні коливання:
   а) полягають у тому, що значення досліджуваної ознаки протягом якогось часу зростає, досягає визначеного максимуму, потім знижується, досягає визначеного мінімуму, знову зростає до колишнього значення і т.д.;
   б) коливання, що періодично повторюються в деякий визначений час кожного року, дні місяця або години дня;
   в) всі відповіді неправильні.

9. Продовження виявленої тенденції за межі ряду (коли закономірність розвитку, що діє в минулому (усередині ряду динаміки), зберігається й у прогнозованому майбутньому):
   а) екстраполяція;
   б) інтерполяція.

Тема 8. ЕКОНОМІЧНІ ІНДЕКСИ

1. ПОНЯТТЯ ТА КЛАСИФІКАЦІЯ ІНДЕКСІВ.

   Індексний метод у проведенні статистичного аналізу. Класифікація індексів. Індекси кількісних показників. Індекси якісних показників. Базисні та ланцюгові індекси.

2. ОСОБЛИВОСТІ ТА МЕТОДОЛОГІЯ ПОБУДОВИ ІНДИВІДУАЛЬНИХ І ЗВЕДЕНИХ ІНДЕКСІВ.

   Методологія побудови індексів. Характеристика основних індивідуальних і загальних індексів. Взаємозв'язок індексів.

3. СУТНІСТЬ СЕРЕДНІХ ІНДЕКСІВ.
   Середньозважені індекси. Індекси середнього рівня інтенсивного показника.

4. ІНДЕКСИ ПРОСТОРОВО-ТЕРИТОРІАЛЬНОГО СПІВВІДНОШЕННЯ.

   Побудова територіальних індексів. Методи побудови територіальних індексів (метод стандартних вагів).

5. ОСОБЛИВОСТІ ІНДЕКСІВ ЛАСПЕЙРЕСА, ПААШЕ ТА ФІШЕРА.

   Індекси Ласпейреса, Пааше та Фішера: розрахунки та властивості.

   ДОДАТКОВА ІНФОРМАЦІЯ ДО ТЕМИ.

   Індекс-дефлятор у системі національних рахунків. Застосування факторного аналізу при дослідженні явищ і процесів.

Терміни

Індекс
Індивідуальні індекси
Зведені (загальні) індекси

Індекс фізичного обсягу продукції
Індивідуальний індекс цін
Індивідуальний індекс собівартості продукції

Індекс кількості продукції, зробленої за одиницю часу
Індекс витрат часу на виробництво одиниці продукції (трудомісткість)

Індивідуальний індекс вартості продукції (товарообігу)
Індивідуальний індекс кількості працівників

Індексована величина
Ваги індексу

Індекс вартості продукції (товарообігу)
Індекс фізичного обсягу продукції
Індекс цін

Індекс змінного складу
Індекс постійного складу
Індекс структурних зрушень

Факторний аналіз
Індекс-дефлятор
Індекс Ласпейреса
Індекс Пааше
Індекс Фішера

Перелік всіх термінів, наведених у даній темі, див. у "Предметному покажчику".

1. ПОНЯТТЯ ТА КЛАСИФІКАЦІЯ ІНДЕКСІВ

“Проблема построения индексового числа относится больше к экономической теории, чем к статистической технике”

Рагнар Фриш [173, с. 1]

Індекс — відносний показник, що виражає співвідношення величин будь-якого явища в часі, у просторі або порівняння фактичних даних з будь-яким еталоном (план, прогноз, норматив тощо);

величина, що показує, у скільки разів рівень досліджуваного явища в даних умовах відрізняється від рівня того ж явища в інших умовах.

Індексові числа, як зазначено автором роботи [4, с. 53], настільки ж економічні, як і статистична конструкція. Виникає питання: чи можна визначити індексові числа як поняття економічної теорії? Це зробити можливо в одній галузі економічної теорії та для одного індексу. Індекс є мірою зміни цін індивідуального споживача при максимізації корисності в умовах незмінної схеми преференційності згідно з теорією споживчого вибору. Коротко поняття економічного індексу можна визначити терміном “індекс ціни при константній корисності”. Індекс ціни тут кореспондує індексу фізичного обсягу шляхом дефляції фактичних змін вартості індексом ціни при константній корисності. Цей кореспондуючий індекс фізичного обсягу є індексом реального доходу або реального споживання.

Задачі, які вирішуються за допомогою економічних індексів

- вимір динаміки соціально-економічного явища за два і більше періоди часу
- вимір динаміки середнього економічного показника
- вимір співвідношення показників за різними регіонами
- визначення ступеня впливу змін значень одних показників на динаміку інших
- перерахування значення макроекономічних показників з фактичних цін у порівнянні

Теорія споживчого вибору і, відповідно, індекс цін константної корисності строго обмежуються індивідуальним споживачем з фіксованою схемою преференції. Схема преференцій для одного товару має змінюватися протягом часу, тому це проблема індексових чисел. Найбільше тут критикується бажання визначити та застосувати індексові числа до груп індивідів. Для утворення тут економічних понять ми повинні вірити в існування групових і середніх схем преференцій, які допускають порівняння між індивідуальними корисностями.

Найбільше критикується спроба визначити та застосувати індексові числа до груп індивідів. Для утворення тут економічних понять ми маємо вірити в існування групових або середніх схем преференцій, які допускають порівняння між індивідуальними корисностями.

Визначення індексу константної корисності є достатньо зрозумілим, воно дає тільки схему преференцій індивідуального споживача або систему опуклих поверхонь преференції в n-мірному товарному просторі. На одній поверхні преференції комбінації споживчих товарів залишають споживача з рівною корисністю або з рівним задоволенням потреб. Він референтний до вибору між бюджетними обсягами, що лежать на поверхні преференції. Так, позначаючи символами p_0 та p_1 дві ситуації цін, а символом q_0 – фактичні покупки в ситуації 0. Спочатку ми визначимо поверхню преференції, на якій лежить обсяг q_0 (при ціні p_0), а потім переходимо до визначення того бюджетного обсягу $\overline{q_1}$, яке має бути придбаним за ціною p_1 і утримувати споживача на тій же поверхні преференційності. Дамо визначення.

Індекс ціни константної корисності $I_{01}(q_0) = \frac{\sum p_1 \overline{q_1}}{\sum p_0 q_0}$ як зміна витрат, що залишаються на одній поверхні преференційності (позначена символом q_0). Позначення $I_{01}(q_0)$ вказує, що індекс ціни є функцією рівня q_0 преференції споживача. За різними рівнями реального доходу індекс ціни $I_{01}(q_0)$ змінюється. Індекс константної корисності залежить від обраного рівня константної корисності.

Схема преференцій індивідуального споживача на практиці не спостерігається за звітними даними про ціни та фізичні обсяги. Ми можемо знайти бюджетний набір $\overline{q_1}$, ідентичний початковому бюджету q_0 при зміні цін від $\overline{p_0}$ до p_1. Таким чином, форма $I_{01}(q_0)$ майже не є формулою індексового числа. Вона дає лише основу для міркування, наскільки ми наблизились до "істинного" індексу; вона показує мету, яку ми ставимо.

Таблиця 8.1

Класифікація індексів

Ознака	Вид	Приклад
Ступінь охоплення явища	Індивідуальні	Зміни обсягу формування окремих складових кредитного портфеля комерційного банку (короткострокові – один обсяг, довгострокові – інший), а також відсоткова ставка кредиту окремого банку
	Зведені	Зміни обсягу формування всього кредитного портфеля комерційного банку
	Групові, або субіндекси	Індекси обсягу виданих кредитів за окремими напрямками призначення, індекси відсоткових ставок за групами строкових і безстрокових депозитів
База порівняння	Динамічні	
	Територіальні	
Види вагів	З постійними вагами	
	Зі змінними вагами	
Форма побудови	Агрегатні	
	Середні	

Продовж. табл. 8.1

Ознака	Вид	Приклад
Характер об'єкта дослідження	Кількісних (об'ємних) показників	
	Якісних показників	
Об'єкт дослідження	Продуктивності праці	
	Собівартості	
	Фізичного обсягу продукції	
	Вартості продукції та ін.	
Склад явища	Постійного (фіксованого) складу	
	Перемінного складу	
Період обрахунку	Річні	
	Квартальні	
	Місячні	
	Тижневі	

2. ОСОБЛИВОСТІ ТА МЕТОДОЛОГІЯ ПОБУДОВИ ІНДИВІДУАЛЬНИХ І ЗВЕДЕНИХ ІНДЕКСІВ

"Я предлагаю определить индексовое число как число, приспособленное для того, чтобы своими вариациями указывать увеличение или уменьшение величины, не допускающей точного измерения"

Эджворт [160, с. 379]

"Проблема индексовых чисел возникает всегда, когда мы нуждаемся в количественном выражении *комплекса,* который создается из индивидуальных измерений, для которых не существует *общей физической* единицы. Желание объединить такие измерения и тот факт, что это не может быть сделано применением только физических или технических принципов сравнения, составляют сущность проблемы индексового числа, и все трудности сосредоточиваются здесь" [183, с. 41].

Таблиця 8.2

Методологія побудови індексів

Умовне позначення	Пояснення
i	Індивідуальні індекси
I	Загальні індекси
q	Кількість (обсяг) будь-якого продукту в натуральному вираженні
p	Ціна одиниці товару
z	Собівартість одиниці продукції

Продовж. табл. 8.2

Умовне позначення	Пояснення
t	Витрати часу на виробництво одиниці продукції
w	Вироблення продукції у вартісному вираженні на одного працівника або за одиницю часу
v	Вироблення продукції в натуральному вираженні на одного працівника або за одиницю часу
T	Загальні витрати часу (tq) або кількість працівників
pq	Вартість продукції або товарообіг
zq	Витрати виробництва
0	Базисний
1	Звітний період

Індивідуальні індекси визначаються у результаті порівняння однотоварних явищ і являють собою відносні величини динаміки, виконання плану, порівняння.

Індекс фізичного обсягу продукції

$$i_q = \frac{q_1}{q_0}. \qquad (8.1)$$

Цей індекс показує, у скільки разів зріс (зменшився) випуск якого-небудь одного товару у звітному періоді в порівнянні з базисним, або скільки відсотків становить зростання (зниження) випуску товару. У знаменнику може бути не тільки кількість продукції, зробленої в будь-якому попередньому періоді, але і планове значення $(q_{пл})$, нормативне $(q_н)$ або еталонне значення, взяте за базу порівняння (q_e).

Індивідуальний індекс цін

$$i_p = \frac{p_1}{p_0}. \qquad (8.2)$$

Цей індекс характеризує зміна ціни одного визначеного товару в поточному періоді у порівнянні з базисним.

Індивідуальний індекс собівартості продукції

$$i_z = \frac{z_1}{z_0}. \qquad (8.3)$$

Цей індекс показує зміну собівартості одиниці продукції в поточному періоді у порівнянні з базисним.

Індекс кількості продукції, зробленої за одиницю часу

$$i_v = \frac{v_1}{v_0} = \frac{q_1}{T_1} : \frac{q_0}{T_0}. \qquad (8.4)$$

Показує, у скільки разів виробництво продукції в одиницю часу (на одного працівника) зросло (знизилося) у поточному періоді в порівнянні з базисним.

Індекс витрат часу на виробництво одиниці продукції (трудомісткість)

$$i_t = \frac{t_0}{t_1}. \qquad (8.5)$$

Показує, у скільки разів трудомісткість виробництва продукції в базисному періоді вище (нижче), ніж у звітному.

Ідивідуальний індекс вироблення продукції у вартісному вираженні на одного працівника

$$i_w = \frac{w_1}{w_0} = \frac{q_1 p}{T_1} : \frac{q_0 p}{T_0}. \tag{8.6}$$

Індивідуальний індекс вартості продукції (товарообігу)

$$i_{pq} = \frac{p_1 q_1}{p_0 q_0} = i_p \cdot i_q. \tag{8.7}$$

Показує, у скільки разів змінилася вартість будь-якого товару в поточному періоді у порівнянні з базисним або скільки відсотків становить зростання (зниження) вартості товару.

Індивідуальний індекс кількості працівників

$$i_T = \frac{T_1}{T_0} = \frac{t_1 q_1}{t_0 q_0} = \frac{i_q}{i_t}. \tag{8.8}$$

Він показує, у скільки разів змінилася кількість працівників у поточному періоді у порівнянні з базисним або скільки відсотків становить зростання (зниження) кількості працівників.

Зведені, або загальні індекси розраховуються для виміру динаміки складного явища, складові частини якого безпосередньо непорівнянні.

Групові, або субіндекси якщо індекси охоплюють не всі елементи складного явища, а тільки їх частину.

У роботах Фріша розглядається типовий комплекс – це комплекс цін деякої номенклатури товарів, виражених у різних одиницях, наприклад, пенса за пінту, фунта стерлінгів за дюжину, фунта стерлінгів за тону. Вирішити проблему можна шляхом переходу від поняття загального рівня цін, яке не спостерігається, до поняття зміни в рівнях цін, що піддається спостереженню. Це вимагає визначення Еджворта-Боулі. При цьому один підхід – вимірювати зміну ціни одиничного товару числом, що виражає відношення двох цін, які спостерігаються. Приклад, коли вартість кредиту піднімається від 15 до 20 % за рік, то друга вартість дорівнює 20 / 15 · 100 = 133 % першої, і ця вартість має збільшення на 33 %. Зовсім інший підхід – дійти висновку, що загальний рівень цін деякої номенклатури операцій збільшився. Перше – це просто співвідношення, хоча часто називається індексовим числом. Друге – це те, що становить мета індексового числа. Рунет (Руїст, 1968 [92, с. 154]) зазначає, що виникає проблема, як комбінувати відносні зміни цін різних товарів в єдине індексове число, яке можна за змістом інтерпретувати як міру відносної зміни загального рівня цін.

“Мы понимаем под покупательной силой денег способность приобретать товары и услуги, на покупки которых с целью потребления данное общество индивидуумов расходует свой денежный доход и подходящее индексовое число является ее символом, иногда обозначаемым как индекс потребления. Из этого следует, что покупательная сила всегда должна быть определена относительно отдельной группы индивидуумов в данной ситуации и именно той, действительное потребление которой образует стандарт, и она не имеет ясного значения, если это отношение не дано”. Кейнс цитирует Маршалла в поддержку этих взглядов [188, с. 21, 30].

Агрегатний індекс складний відносний показник, який характеризує середню зміну соціально-економічного явища, що складається з непорівнянних елементів. Особливість цієї форми індексу полягає в тому, що в агрегатній формі безпосередньо порівнюються дві суми однойменних показників. Чисельник і знаменник агрегатного індексу являють собою суму добутків двох величин, одна з яких зміняється (*індексована величина*), а інша залишається незмінною в чисельнику і знаменнику (*вага індексу*).

Індексована величина ознака, зміна якого вивчається.

Вага індексу величина, що служить для цілей порівняння індексованих величин.

Індекс вартості продукції або товарообігу являє собою відношення вартості продукції поточного періоду до вартості продукції в базисному періоді:

$$I_{qp} = \frac{\sum q_1 p_1}{\sum q_0 p_0}. \qquad (8.9)$$

Такий індекс показує, у скільки разів зросла (зменшилася) вартість продукції (товарообігу) звітного періоду в порівнянні з базисним, або скільки відсотків становить зростання (зниження) вартості продукції.

Індекс фізичного обсягу продукції це індекс кількісного показника. У цьому індексі індексованою величиною буде кількість продукції в натуральному вираженні, а вагою – ціна.

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}, \qquad (8.10)$$

де в чисельнику дробу – умовна вартість зроблених у поточному періоді товарів у цінах базисного періоду, а в знаменнику – фактична вартість товарів, зроблених у базисному періоді. Індекс фізичного обсягу продукції показує, у скільки разів зросла (зменшилася) вартість продукції через зростання (зниження) обсягу її виробництва, або скільки відсотків становить зростання (зниження) вартості продукції в результаті зміни фізичного обсягу її виробництва.

Індекс цін індекс якісного показника. Індексованою величиною буде ціна товару, тому що цей індекс характеризує зміна цін. Вагою буде кількість зроблених товарів.

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}, \qquad (8.11)$$

де в чисельнику дробу – фактична вартість продукції поточного періоду, а в знаменнику – умовна вартість тих же товарів у цінах базисного періоду. Індекс показує, у скільки разів зросла (зменшилася) вартість продукції через зміну цін, або скільки відсотків становить зростання (зниження) вартості продукції у результаті зміни цін.

В економічній теорії індексові числа виступають парами: одні – індекси ціни, інші – фізичного обсягу. Існує тенденція розподіляти їх таким чином. Іноді один чи інший застосовуються окремо, але майже завжди на другому плані його напарник. Така пара може бути призначена для пояснення варіацій вартості агрегату тоді, коли рух агрегату витрат споживачів аналізується за двома компонентами: зміни цін і зміни реального споживання.

У статистичних роботах зустрічаються і інші помилки. Фактори помилок включають: неточності даних і, таким чином, використання апроксимацій*; відсутність* деяких даних, тобто некомплектність; *недоступність того, що визнане специфічним*, і застосування замінників. Вирішення задачі оцінки помилок вибірки залежить від схеми вибірки, що передбачена конструкцією індексу.

Із самого початку вимірюване *поняття* повинне бути визначене з достатньою точністю згідно з поставленими цілями. Потім повинна бути обрана міра (оцінювач), наскільки можливо відповідна певному поняттю. Далі особливо уважно має бути досліджений *вибірковий аспект* даних для обчислення, якщо це можливо, стандартної помилки оцінки. У результаті одна або декілька *оцінок* міри (оцінювача) повинні бути одержані і розцінені на основі застосованих апроксимацій, неповноти колекцій даних і замінників, введених при отриманні оцінок.

Добуток ланцюгових індексів дорівнює базисному:

$$i_p = \frac{p_2}{p_0} = \frac{p_2}{p_1} \cdot \frac{p_1}{p_0}. \tag{8.12}$$

3. СУТНІСТЬ СЕРЕДНІХ ІНДЕКСІВ

Числа не должны рассматриваться как необходимо точные, вплоть до приведенной конечной цифры

Р. Аллен [4]

Середня гармонійна форма індексів

розглянемо побудову загального індексу цін у вигляді середньої гармонійної, застосувавши дані про товарообіг звітного та базисного періодів і значення індивідуальних індексів цін.

Нехай є дані $\sum p_0 q_0,\ \sum \mathrm{p}_1 q_{1,}\ i_p = \frac{p_1}{p_0}$.

Вихідна інформація для обрахунку знаменника, тобто умовного товарообігу відсутня, тому його розрахунки можливо провести, застосувавши індивідуальні індекси цін і товарообіг звітного періоду, тоді $I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum p_1 q_1}{\sum \frac{p_1}{i_p} \cdot q_1}$ – середня гармонійна форма індексу. Загальний індекс цін є середньою гармонійною величиною з добутку індивідуальних індексів цін.

“Определение индекса безотносительно к количеству товаров; о гипотезе существования многочисленной группы товаров, цены которых варьируют по манере совершенного рынка, с изменениями под воздействием снабжения деньгами” (Эджворт [166, с. 233]).

Середня арифметична форма індексів для проведення розрахунків індексу фізичного обсягу у формі середньої арифметичної використаємо дані про товарообіг базисного періоду та значення індивідуального індексу товарообігу.

Нехай є дані $\sum p_0q_0$, $\sum p_1q_1$, $i_q = \frac{q_1}{q_0}$,

тоді $I_q = \frac{\sum p_0q_1}{\sum p_0q_0} = \frac{\sum p_0q_0 \cdot i_q}{\sum p_0q_0} = \frac{\sum p_0q_1}{\sum p_0q_0}$ – середня арифметична форма індексу. Загальний індекс фізичного обсягу є середньою арифметичною з індивідуальних індексів фізичного обсягу.

▷ **ПРИКЛАД** застосування середньої арифметичної форми індексів.

Середня арифметична форма індексів застосовується, якщо використовуються дані про вартість кредитного портфеля базисного періоду та значення індивідуального індексу вартості окремих складових кредитного портфеля. ■

“Возникает проблема, как комбинировать относительные изменения цен различных товаров в единое индексовое число, которое можно по смыслу интерпретировать как меру относительного изменения общего уровня цен” [204, с. 154].

Індекс змінного складу

$$I_{пер.\,с} = \overline{p_1} \div \overline{p_0} = \frac{\sum p_1q_1}{\sum q_1} \div \frac{\sum p_0q_0}{\sum q_0}. \tag{8.13}$$

Індекс змінного складу показує, у скільки разів змінилася середня ціна на товар у звітному році в порівнянні з базисним.

$$I_{пер.с} = I_{пост.с} \cdot I_{стр.сд}. \tag{8.14}$$

Індекс постійного складу

$$I_{пост.\,с} = \frac{\sum p_1q_1}{\sum q_1} \div \frac{\sum p_0q_1}{\sum q_1} = \overline{p_1} \div \overline{p_{усл}}. \tag{8.15}$$

Індекс постійного складу показує, як зміниться середня ціна на товари лише під впливом зміни індивідуальних рівнів.

Індекс структурних зрушень

$$I_{стр.\,сд} = \frac{\sum p_0q_1}{\sum q_1} \div \frac{\sum p_0q_0}{\sum q_0} = \overline{p_{усл}} \div \overline{p_0}. \tag{8.16}$$

Індекс структурних зрушень показує, у скільки разів зміниться середня ціна лише під впливом структурних змін в обсязі продажів.

▷ **ПРИКЛАД.**

Якщо розглядати вартість кредитного портфеля банку, то індекс змінного складу показує, у скільки разів змінилася середня відсоткова ставка за кредитом у звітному році в порівнянні з базисним. Індекс постійного складу показує, як зміниться середня відсоткова ставка за кредитом тільки під впливом зміни індивідуальних рівнів відсоткової ставки. Індекс структурних зрушень показує, у скільки разів зміниться середня відсоткова ставка лише під впливом структурних змін в обсязі кредитного портфеля. ■

4. ІНДЕКСИ ПРОСТОРОВО-ТЕРИТОРІАЛЬНОГО СПІВВІДНОШЕННЯ

У статистичній практиці часто виникає потреба зіставлення рівнів економічного явища в просторі: за країнами, економічними районами, областями, тобто у визначенні територіальних індексів. При побудові територіальних індексів доводиться вирішувати питання, яка вага використовувалася при їх визначенні. Наприклад, якщо стоїть задача порівняти ціни двох регіонів (*А* і *Б*), то можна побудувати два індекси:

$$I_{А/Б} = \frac{\sum p_А q_А}{\sum p_Б q_А} \tag{8.17}$$

та

$$I_{Б/А} = \frac{\sum p_Б q_Б}{\sum p_А q_Б}, \tag{8.18}$$

де $I_{А/Б}$ – індекс, в якому як база порівняння застосовуються дані по регіону *А*;

$I_{Б/А}$ – індекс, використовуваний як база порівняння даних по регіону *Б*.

Ці формули можуть дати абсолютно різне уявлення про співвідношення рівнів явища. Наприклад, при розрахунку за формулою (8.17) значення ознаки буде нижчим у регіоні *А*, а за формулою (8.18) – у регіоні *Б*.

У теорії та практиці статистики пропонуються різні методи побудови територіальних індексів, зокрема *метод стандартних вагів*.

Метод стандартних вагів	полягає у тому, що значення величини, яка індексується, зважується не за вагами якогось одного регіону, а за вагами області, економічного району, республіки, в яких знаходяться порівнювані регіони.

5. ОСОБЛИВОСТІ ІНДЕКСІВ ЛАСПЕЙРЕСА, ПААШЕ ТА ФІШЕРА

Таблиця 8.3

Формули Ласпейреса, Пааше, Фішера для розрахунку індексів цін і фізичного обсягу

Назва індексу	Формула розрахунку індексу		
	Ласпейреса (базисні ваги)	Пааше (звітні ваги)	Фішера
Індекс цін	$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}$	$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$	$I_p = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_0 q_1}}$
Індекс фізичного обсягу	$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}$	$I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1}$	$I_q = \sqrt{\frac{\sum q_1 p_0}{\sum q_0 p_0} \cdot \frac{\sum q_1 p_1}{\sum q_0 p_1}}$

ДОДАТКОВА ІНФОРМАЦІЯ ДО ТЕМИ

Перерахунок найважливіших вартісних показників системи національних рахунків (національний дохід, валовий національний продукт і т.д.) з фактичних цін у порівнювані здійснюється за допомогою індексу-дефлятора.

Індекс-дефлятор коефіцієнт, що переводить значення вартісного показника за звітний період у вартісні вимірники базисного. Індекс-дефлятор розраховується як відношення фактичної вартості продукції звітного періоду до вартості обсягу продукції, структура якого аналогічна структурі звітного року, але визначеного в цінах базисного року. В основі розрахунку індексу-дефлятора лежить формула Пааше – агрегатна формула індексу з поточними вагами. Індекс-дефлятор для ВВП у n-му році визначається за формулою:

$$I_d = \frac{\sum p_n q_n}{\sum p_0 q_n}, \qquad (8.19)$$

де I_d – індекс-дефлятор;

q_n – обсяг продукції в n-му році;

p_n, p_0 – ціни, що фактично діяли в n-му році і базисному році відповідно.

▷ **ПРИКЛАД** індексу-дефлятора валового внутрішнього продукту (ВВП) є індексом цін, вживаним для коригування номінального обсягу ВВП з урахуванням інфляції, та отримання на цій основі реального його обсягу. ■

Особливість індексу-дефлятора він не може бути використаний для порівняльної оцінки динаміки цін за два періоди. Індекси-дефлятори дають уявлення тільки про відношення вартості продукції в поточному періоді до її вартості в базисному періоді. При цьому не враховується відмінність складу і структури продукції в базисному періоді в порівнянні із звітним.

У статистичній практиці індекси-дефлятори визначаються не тільки в цілому по народному господарству, вони обчислюються по окремих регіонах, різних товарних групах, каналах реалізації споживацьких благ, галузях економіки тощо.

Факторний аналіз має на меті: установити вплив окремих факторів на зміну досліджуваного показника. Факторний аналіз товарообігу або вартості продукції

$$\Delta pq = \sum p_1 q_1 - \sum p_0 q_0$$
$$(\Delta pq)_p = \sum p_1 q_1 - \sum p_0 q_1$$
$$(\Delta pq)_q = \sum p_0 q_1 - \sum p_0 q_0. \qquad (8.20)$$

Частковий внесок визначають у тих випадках, коли вплив факторів односпрямований.

Другий спосіб факторного аналізу товарообігу застосовується лише за умови, що має сенс середня ціна

$$\Delta pq = \sum p_1 q_1 - \sum p_0 q_0.$$

1) $(\Delta pq)_{\bar{p}} = (\bar{p} - \overline{p_{усл}})\sum q_1$ – вплив індивідуального рівня цін на товарний оборот;
2) $(\Delta pq)_{\overline{p_{стр}}} = (\overline{p_{усл}} - \overline{p_0})\sum q_1$ – вплив структурних зрушень в обсязі продажів на товарний оборот;
3) $(\Delta pq)_q = (\sum q_1 - \sum q_0)\overline{p_0}$.

ХРЕСТОМАТІЯ

Аллен Р. Экономические индексы. – М.: Статистика, 1980. – 256 с.

Позиция Кейнса теперь общепринята как более реалистическая и удобная, чем позиция Эджворта. Настоящее изложение поэтому концентрируется на рассмотрении индексовых чисел в агрегатных формах и в эквивалентных, взвешенных вариантах. Все же стохастический подход и формы невзвешенных индексовых чисел, к которым он ведет, могут в отдельных случаях применяться. В частности, когда практическое построение индексовых чисел допускает в некоторых аспектах выборки построение индексового числа из его детальных составных частей.

Экономико-теоретическое основание для построения индексовых чисел может быть получено в одном важном случае – индекса цены константной полезности и корреспондирующего индекса реального потребления. Должен быть изыскан экономический базис в теории потребительского выбора индивидуума, олицетворяющего стремление к максимальной полезности при неизменной схеме преференций. Анализируются две типичные ситуации, представленные двумя точками во времени. В целях экспозиции это показано на примере из двух товаров, что делает возможным пользование диаграммой в двух измерениях.

Символы q_1 и q_2 обозначают покупаемые количества двух товаров. *Схема поверхности покупательной преференциальности* иллюстрируется рис. 8.1, содержащим два комплекта пересекающихся кривых на поверхности Oq_1q_2. Они пересекают эту поверхность так, что мы прослеживаем изменение покупок потребителя при вариациях его дохода и движения рыночных цен. Предполагается, что сбережение отсутствует, и что доход потребителя и его затраты на два товара одинаковы.

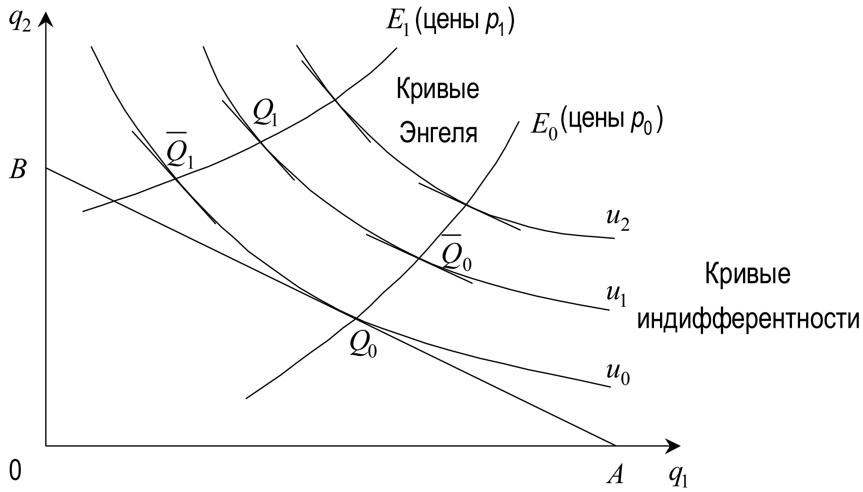


Рис. 8.1. Схема потребительской преференции

Книга Фриш [173].

Фриш называет первый из них *стохастическим подходом*, где слово “стохастический” – прилагательное, ставшее теперь общеупотребительным вместо менее точного “пробабилистический”, корреспондирующий существительному “вероятность”. Этот подход направлен на широкий объект и главным образом на общий уровень цен или ценность денег, безотносительно к какой-либо группе или к применению в каком-либо комплексе обстоятельств. Другой подход может быть назван *агрегативным.* Он имеет отношение к некоторому агрегату и к некоторой группе, определенным заранее. Например, он может иметь отношение к агрегированным расходам отдельной группы потребителей, а его целью может быть суждение о чистом доходе или о жизненном стандарте группы. Спецификация группы может быть совсем узкой (например, семьи пенсионеров-одиночек) или такой широкой, как все потребители страны.

Аллен Р. Экономические индексы. – М.: Статистика, 1980. – 256 с.

Индексовые числа – это практические конструкции, определяемые и исчисляемые главным образом для обеспечения решения практических задач. Один индекс цены, например, может быть предназначен для исчисления индексно-связанных ставок заработной платы, другой – для правильного определения пенсий для престарелых и, наконец, третий – для включения в систему уравнений краткосрочной модели экономики. Эти и другие подобные индексовые числа особенно сложны по конструкции, поскольку, как мы видели, они касаются некоторых понятий (таких, например, как общий уровень цен), не поддающихся непосредственному наблюдению.

Практика должна зависеть от теории. Можно возражать против этого, но не может быть "измерений без теории": в экономике и в социальных науках так же, как в физических науках. Многое из того, что смутно и двусмысленно в практике индексовых чисел, может быть прослежено до того момента, когда обнаруживается недостаток хорошего теоретического базиса. Это причина обширной разработки теоретического остова индексов в настоящей работе, особенно для формы агрегатного средневзвешенного индекса сначала для двух ситуаций в гл. 2, а затем для рядов индексовых чисел в гл. 4 с обильными промежуточными иллюстрациями в гл. 3. Последующее исследование, хотя и изложенное в легкой алгебраической форме с упрощающим символом 2, не является математическим. Все формулы имеют прикладной характер, и все они подкрепляются словесным описанием их содержания.

Прежде чем начать это теоретическое исследование, мы должны изложить его перспективы, давая представления о некоторых существенных практических положениях, которые имеются в виду. Эти практические замечания и указания большей частью относятся к вопросам, возникающим во всяком применении статистических методов. Статистическая практика индексовых чисел отличается не по роду, а во многих отношениях скорее по степени от всех остальных приемов прикладной статистики.

Сначала необходимо изложить более точно то, *что* должны измерять индексовые числа, рассматриваемые в специальном применении.

Это *понятие* должно быть скорее общим в формулировке и относиться к некоторой теоретической (например, экономической) модели. Оно обычно требуется для определения изменений некоторого уровня, непосредственно не измеримого, т.е. общего уровня цен. Это понятие нужно сделать несколько определеннее, может быть, относительно цен, уплачиваемых отдельной группой лиц, и, следовательно, относительно денежного агрегата расходов данной группы. Типичная задача, которую в таком случае должны решать индексовые числа, состоит в том, чтобы проследить изменение стоимости специфического бюджетного набора товаров, характерного для отдельной группы лиц.

Статистик должен выбрать *меру* или *оцениватель* понятия из возможных варьирующих альтернатив. Термин *мера* является общеупотребительным при переводе понятия в практические оценки; термин *оцениватель* статистически более технический со специальным значением при оценке выборочных данных. Мы можем легко сохранить оба термина.

Выбор меры (оценивателя) – стандартная статистическая практика. В простом случае, например, можно пытаться получить некоторую среднюю из относительных цен, такую, как приведенная за 1973 г.

Опять возникает вопрос выбора между равным взвешиванием и некоторой системой взвешивания и в случае последней – какой именно. Например, выбор между индексом Ласпейреса (базисно-взвешенным) и Пааше (текуще-взвешенным) может быть совершенно противоположным при уточнении вопроса, на который нужно получить ответ. Изменяя или пересматривая вопрос, вы можете изменить выбор меры (оценивателя).

При выбранной мере (оценивателе) мы сталкиваемся с экстраординарно широким диапазоном практических проблем получения действительных *оценок с* помощью нашей меры из данных, которыми мы можем располагать. Если мы имеем полные и точные данные, то вполне обоснованно можно пользоваться формулой выбранной меры и получить ее истинную величину в рассматриваемом случае. Однако наши данные никогда не бывают полными и точными. Всякая оценка, которую мы делаем, расходится с истинной величиной. Это обычное явление, и в действительности важнее всего различить два рода ошибок или расхождений оценок с истинной величиной: *ошибки выборки*, возникающие из-за того, что данные являются выборкой из общего числа, и *другие ошибки*, возникающие в диапазоне факторов (иных, чем свойства выборки), воздействующих на имеющиеся данные. Ошибки выборки свойственны индексовым числам так же, как другим статистическим оценкам. Часто случается, что они не устраняются надлежащим образом, так как обычно им уделяется мало внимания в схемах индексовых чисел. Другие ошибки также встречаются во всех статистических работах. Иногда они делаются эпидемическими при построении индексовых чисел. Факторы ошибок включают: неточности данных и, таким образом, использование *аппроксимаций*; отсутствие некоторых данных, т.е. *некомплектность*; недоступность того, что признано специфическим, и применение *заменителей*. Решение задачи по оценке ошибок выборки зависит от схемы выборки (или от ее отсутствия), предусмотренной конструкцией индекса; оценка других ошибок делается ad hoc, бывает делом удачи или неудачи. *Мы можем наметить следующие основные направления в построении и оценке индексовых чисел.*

[166 с. 195-343].

Индексовые числа имеют длительную историю, и Кендэл (1969) [182] дал хороший обзор их раннего периода. Классическое определение индексовых чисел простирается в прошлое – к Эджворту. В 1877-1879 гг. Эджворт был секретарем комитета Британской ассоциации, учрежденной для изучения методов измерения вариации ценности денег. В этой должности он написал три обширных меморандума, воспроизведенных в его книге (1925).

[223, с. 196].

Величиной, которую он особенно имел в виду, был общий уровень цен или же ценность (покупательная сила) денег, обратные по отношению друг к другу. То же понятие можно усмотреть в несколько более развернутой форме в определении Боули (1926).

Индексовые числа применяются для измерения изменения некоторого количества, которое мы не можем наблюдать непосредственно, но которое, как мы знаем, имеет определенное влияние на многие другие количества, которые мы можем непосредственно наблюдать и которые имеют в целом общую тенденцию к увеличению или уменьшению, причем это влияние скрывается под действием многих причин, влияющих на отдельные количества разными путями. И опять количество, не поддающееся наблюдению, которое подразумевал Боули, было экономическим понятием, таким, как ценность денег.

Существенной чертой определения является то, что в нем не делается попытки получить меру или индикатор действительного уровня, достигнутого величиной, не поддающейся измерению. Число-показатель ограничивается измерением *изменений* этой величины от одной ситуации к другой. Две сравниваемые ситуации не имеют каких-либо ограничительных условий; они могут быть двумя периодами времени (например, двумя разными годами) или двумя ситуациями в пространственном смысле (например, двумя районами страны), или двумя группами лиц (например, семьями пенсионеров из одного и двух человек). Здесь опять для удобства представления, и поскольку это наиболее обычный случай в практике, мы в тексте большей частью ссылаемся на индексовые числа во времени. Практически все, что здесь рассматривается, с надлежащими коррективами терминов и обозначений можно применить к сравнениям ситуаций другого рода.

Поскольку индексовые числа измеряют изменения, они выражаются одной избранной ситуацией, принятой за 100. Она называется *исходной базой* серии индексовых чисел. База сравнения – часто употребляемый альтернативный термин. В годовой серии, например, исходной базой служит год с уровнем 100, взятый для сравнения. Другой год может тогда оказаться с индексовым числом, скажем 126. Это означает, что в соответствии с примененным индексовым числом рассматриваемая величина во втором году составляет 126 % ее уровня в базисном году. Действительный (абсолютный) уровень не измерен ни за один год; здесь индексом дано только изменение от одного года к другому, составляющее увеличение на 26 %.

ПРАКТИЧНІ ЗАВДАННЯ

Завдання 1. На основі даних про обсяги наданих кредитів по 3 відділеннях банку розрахувати:

1) зведені індекси вартості кредиту, обсягу та відсоткової ставки;
2) індекси середньої відсоткової ставки за кредит: постійного, змінного складу та структурних зрушень;
3) провести факторний аналіз вартості кредиту двома способами (на основі використання даних про вартість кредиту, на основі використання даних про середні відсоткові ставки).

Таблиця 8.4

Надані кредити по 3 відділеннях банку

Відділення	Обсяг кредиту, тис. грн.		Відсоткова ставка, %	
	базисний	звітний	базисний	звітний
	q_0	q_1	p_0	p_1
1	250	300	12,0	13,0
2	150	150	15,0	14,0
3	800	900	10,0	9,0

Завдання 2. На основі даних таблиці розрахувати індивідуальні індекси відсоткових ставок за кожним відділенням банку, загальні індекси відсоткових ставок за формулами Ласпейреса, Пааше та Фішера. Провести факторний аналіз вартості кредиту та зробити висновки.

Завдання 3. На основі наведених даних розрахувати:
1) загальні індекси;
2) індивідуальні індекси;
3) провести факторний аналіз середнього розміру депозиту;
4) зробити висновки за кожним пунктом розрахунків.

Таблиця 8.5

Середній розмір депозиту та кількість вкладників за філіями

№ філії банку	Середній розмір депозиту, тис. грн.		Кількість вкладників, чол.	
	базисний	звітний	базисний	звітний
1	5 500	7 000	340	350
2	7 300	7 400	150	185
3	8 500	9 400	352	370

КОНТРОЛЬНІ ПИТАННЯ

1. Що називається індексом у статистиці?
2. Які задачі вирішують за допомогою індексів?
3. Що характеризують індивідуальні індекси? Навести приклади.
4. У чому суть загальних індексів?
5. Для чого необхідний розподіл на індекси об'ємних (кількісних) та якісних показників і яка система зважування в теорії індексів?
6. Як обчислюється агрегатний індекс вартості продукції (товарообігу у фактичних цінах) і що він характеризує?
7. Як обчислюється агрегатний індекс фізичного обсягу продукції (товарообігу) і що він характеризує? Написати формулу.
8. Коли виникає необхідність перетворення індексу фізичного обсягу в середній арифметичний і середній гармонійний; яким чином відбуваються такі перетворення? Показати на прикладах.
9. Як обчислюють агрегатні індекси цін (Пааше і Ласпейреса), собівартості, продуктивності праці та що вони показують? Написати їх формули.
10. Коли виникає необхідність перетворення агрегатного індексу цін у середній гармонійний і середній арифметичний, яким чином відбуваються такі перетворення? Показати на прикладі.
11. Який варіант агрегатних індексів якісних показників використовують при розрахунку індексу споживчих цін (ІСЦ) і чому?
12. Що називається індексом змінного складу, як він обчислюється і що характеризує? Написати його формулу.

13. Який індекс називається індексом постійного складу, як він обчислюється і що характеризує?
14. Що характеризує індекс структурних зрушень і як він обчислюється?
15. Який взаємозв'язок існує між індексами змінного, постійного складу і структурних зрушень?
16. Як будуються базисні та ланцюгові індекси і який між ними існує взаємозв'язок?
17. Чим є індекси з постійними і змінними вагами?
18. У чому суть і призначення індексів-дефляторів?
19. Чим є система взаємозв'язаних індексів, для чого вона застосовується?
20. У чому виражається взаємозв'язок індексів цін, фізичного обсягу і товарообігу, як практично вона використовується?
21. Яка система взаємозв'язаних індексів використовується при аналізі собівартості, фізичного обсягу і витрат у виробництві?
22. Як визначити частку впливу різних чинників на зміну результативного показника?
23. В яких випадках виробляється розкладання індексів за трьома і більше чинниками?
24. Як здійснюється розкладання абсолютного приросту по чинниках, що воно характеризує?

ТЕСТИ

1. Індекс:
   а) відносний показник, що виражає співвідношення величин якого-небудь явища в часі, у просторі або порівняння фактичних даних з будь-яким еталоном (план, прогноз, норматив тощо);
   б) величина, що показує, у скільки разів рівень досліджуваного явища в даних умовах відрізняється від рівня того ж явища в інших умовах;
   в) всі відповіді правильні.

2. Визначити види індексів за характером об'єкта дослідження:
   а) індекси кількісних (об'ємних) показників;
   б) індивідуальні індекси;
   в) індекси якісних показників;
   г) динамічні індекси.

3. Який індекс показує, у скільки разів зріс (зменшився) випуск якого-небудь одного товару у звітному періоді в порівнянні з базисним, або скільки відсотків становить зростання (зниження) випуску товару:
   а) індивідуальний індекс фізичного обсягу продукції;
   б) індивідуальний індекс цін;

в) індивідуальний індекс собівартості продукції;
г) всі відповіді неправильні.

4. Індекс кількості продукції, зробленої за одиницю часу:
а) показує, у скільки разів виробництво продукції в одиницю часу (на одного працівника) зросло (знизилося) у поточному періоді у порівнянні з базисним;
б) показує, у скільки разів трудомісткість виробництва продукції в базисному періоді вище (нижче), ніж у звітному;
в) показує, у скількох разів змінилася вартість якого-небудь товару в поточному періоді у порівнянні з базисним, або скільки відсотків становить зростання (зниження) вартості товару;
г) всі відповіді неправильні.

5. Який індекс показує, у скільки разів змінилася середня ціна на товар у звітному році в порівнянні з базисним:
а) індекс змінного складу;
б) індекс постійного складу;
в) індекс структурних зрушень.

6. За формулою $I_p = \dfrac{\sum p_1 q_0}{\sum p_0 q_0}$ розраховується:
а) індекс цін Ласпейреса;
б) індекс цін Пааше;
в) індекс-дефлятор;
г) індекс фізичного обсягу Ласпейреса.

7. Визначити коефіцієнт, що переводить значення вартісного показника за звітний період у вартісні вимірники базисного і розраховується як відношення фактичної вартості продукції звітного періоду до вартості обсягу продукції, структура якого аналогічна структурі звітного року, але визначеного в цінах базисного року:
а) індекс-дефлятор;
б) індекс Фішера;
в) індекс цін Ласпейреса;
г) всі відповіді неправильні.

8. Факторний аналіз:
а) має на меті установити вплив окремих факторів на зміну досліджуваного показника;
б) має на меті установити вплив змін факторів на збільшення досліджуваного показника;
в) всі відповіді правильні.

ПРЕДМЕТНИЙ ПОКАЖЧИК

Абсолютне значення 1 % приросту 145 (Ч. 2)
Абсолютне прискорення 145 (Ч. 2)
Абсолютний приріст 143 (Ч. 2)
Аналітичне групування 69 (Ч. 1)
Асиметрія 127 (Ч. 1)
Атрибутивний ряд 71 (Ч. 1)
Безперервне спостереження 42 (Ч. 1)
Вага індексу 168 (Ч. 2)
Варіаційний ряд 71 (Ч. I), 8 (Ч. 2)
Варіація 13, 116 (Ч. 1)
Вибіркова сукупність 6, 7 (Ч. 2)
Вирівнювання 150 (Ч. 2)
Відносне лінійне відхилення 117 (Ч. 1)
Відносне прискорення 146 (Ч. 2)
Відносний показник динаміки (*ВПД*) 103 (Ч. 1)
Відносний показник інтенсивності (*ВПІ*) 103 (Ч. I)
Відносний показник координації (*ВПК*) 103 (Ч. 1)
Відносний показник плану (*ВПП*) 103 (Ч. 1)
Відносний показник порівняння ($ВПП_р$) 103 (Ч. 1)
Відносний показник реалізації плану (*ВПРП*) 103 (Ч. 1)
Відносний показник структури (*ВПС*) 103 (Ч. 1)
Власне випадковий відбір 13 (Ч. 1)
Вторинна інформація 30 (Ч. 1)
Генеральна сукупність 6 (Ч. 2)
Гетероскедастичність 52 (Ч. 2)
Гістограма 72 (Ч. 1)
Гомоскедастичність 52 (Ч. 2)
Графічний образ 86 (Ч. 1)
Групова таблиця 114 (Ч. 2)
Групувальна ознака 68, 86 (Ч. 1)
Групування 67 (Ч. 1)
Дилема ув'язненого 78 (Ч. 2)
Динамічний ряд 95, 140 (Ч. 2)
Дисперсія 118 (Ч. 1)
Експлікація 86 (Ч. 1)
Екстраполяція 151 (Ч. 2)
Ексцес 127 (Ч. 1)
Еластичність 35 (Ч. 2)
Емпіричне кореляційне відношення 126 (Ч. 1)
Емпіричний коефіцієнт детермінації 125 (Ч. 1)

Ентропія 117 (Ч. 1)
Закономірність 13 (Ч. 1)
Зведені абсолютні показники 101 (Ч. 1)
Зведені (загальні) індекси 167 (Ч. 2)
Зведення 66 (Ч. 1)
Звітність 43, 48 (Ч. 1)
Індекс 163 (Ч. 2)
Індекс вартості продукції (товарообігу) 167, 168 (Ч. 2)
Індекс витрат часу на виробництво одиниці продукції (трудомісткість) 166 (Ч. 2)
Індекс змінного складу 170 (Ч. 2)
Індекс кількості продукції, зробленої за одиницю часу 166 (Ч. 2)
Індексована величина 168 (Ч. 2)
Індекс постійного складу 170 (Ч. 2)
Індекс структурних зрушень 170 (Ч. 2)
Індекс фізичного обсягу продукції 166 (Ч. 2)
Індекс цін 168 (Ч. 2)
Індивідуальний індекс вартості продукції (товарообігу) 166 (Ч. 2)
Індивідуальний індекс кількості працівників 166 (Ч. 2)
Індивідуальний індекс собівартості продукції 167 (Ч. 2)
Індивідуальний індекс цін 166 (Ч. 2)
Індивідуальні індекси 166 (Ч. 2)
Інтервал 68 (Ч. 1)
Інтерполяція 151 (Ч. 2)
Коефіцієнт варіації 122 (Ч. 1)
Коефіцієнт детермінації 39 (Ч. 2)
Коефіцієнт зростання 143 (Ч. 2)
Коефіцієнт осциляції 121 (Ч. 1)
Конкретний статистичний показник 99 (Ч. 1)
Крива розподілу 126 (Ч. 1)
Критична дата 39 (Ч. 1)
Кумулята 72 (Ч. 1)
Макет 75 (Ч. 1)
Медіана 106 (Ч. 1)
Метод Монте-Карло 73 (Ч. 2)
Механічний відбір 13 (Ч. 1)
Мода 106 (Ч. 1)
Модель з лагом 63 (Ч. 2)
Монографічна таблиця 80 (Ч. 1)
Несуцільне спостереження 43 (Ч. 1)
Огіва 73 (Ч. 1)

Одиниця спостереження 34 (Ч. 1)
Одиниця сукупності 11, 34 (Ч. 1)
Ознака 13 (Ч. 1)
Опитування 44 (Ч. 1)
Основа таблиці 74 (Ч. 1)
Парні коефіцієнти кореляції 46 (Ч. 2)
Первинна інформація 30 (Ч. 1)
Період спостереження 39 (Ч. 1)
Повна статистична таблиця 75 (Ч. 1)
Показник-категорія 99 (Ч. 1)
Поле графіка 86 (Ч. 1)
Полігон 71 (Ч. 1)
Помилка вибірки 14 (Ч. 2)
Помилка реєстрації 14 (Ч. 2)
Помилка репрезентативності 14 (Ч. 2)
Програма спостереження 35 (Ч. 1)
Реєстр 43, 48 (Ч. 1)
Розмах 117 (Ч. 1)
Ряд розподілу 70 (Ч. 1)
Сезонні коливання 148 (Ч. 2)
Середнє квадратичне відхилення 120 (Ч. 1)
Середнє лінійне відхилення 118 (Ч. 1)
Середній абсолютний приріст 146 (Ч. 2)
Середній темп зростання 147 (Ч. 2)
Середній темп приросту 147 (Ч. 2)
Середня арифметична 105 (Ч. 1)
Середня гармонійна 105 (Ч. 1)
Середня геометрична 105 (Ч. 1)
Середня квадратична 105 (Ч. 1)
Серійний (гніздовий) відбір 13 (Ч. 2)
Система статистичних показників 99 (Ч. 1)
Статистика 8, 13 (Ч. 1)
Статистика Дарбіна-Уотсона 254
Статистична сукупність 11, 13 (Ч. 1)
Статистичне спостереження 15, 26 (Ч. 1)
Статистичний графік 86 (Ч. 1)
Статистичний показник 12, 99 (Ч. 1)
Статистичний формуляр 36 (Ч. 1)
Стратифікований відбір 13 (Ч. 2)
Суцільне спостереження 43 (Ч. 1)
Темп зростання 144 (Ч. 2)

Темп приросту 144 (Ч. 2)
Тенденція автокореляції 149 (Ч. 2)
Тенденція дисперсії 149 (Ч. 2)
Тенденція середнього рівня 149 (Ч. 2)
Тест Гейзера 57 (Ч. 2)
Тест Голдфелда-Квандта 56 (Ч. 2)
Тест рангової кореляції Спірмена 56 (Ч. 2)
Точка рішення Неша 81 (Ч. 2)
Тренд 148 (Ч. 2)
Фіктивні змінні 65 (Ч. 2)
Функція Кобба-Дугласа 72 (Ч. 2)
Циклічні коливання 148 (Ч. 2)
Частинні коефіцієнти кореляції 49 (Ч. 2)
Щільність розподілу 72 (Ч. 1)
Ящик Еджворта 83 (Ч. 2)
t -статистика 47, 49 (Ч. 2)
F -критерій Фішера 40 (Ч. 2)

СПИСОК ЛІТЕРАТУРИ

1. Адамов, В. Е. Факторный индексный анализ (методология и проблемы) [Текст] / В. Е. Адамов. – М. : Статистика, 1977. – 199 с.
2. Айвазян, С. А. Прикладная статистика. Основы моделирования и первичной обработки данных [Текст] / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1983. – 471 с.
3. Айвазян, С. А. Прикладная статистика. Исследование зависимостей [Текст] / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1985. – 485 с.
4. Аллен, Р. Экономические индексы [Текст] / Р. Аллен. – М. : Статистика, 1980. – 256 с.
5. Альбом наглядных пособий по общей теории статистики [Текст]. – М. : Финансы и статистика, 1991. – 75 с.
6. Андерсен, Т. Статистический анализ временных рядов [Текст]. – М. : Мир, 1976. – 155 с.
7. Бакланов, Г. И. Некоторые вопросы индексного метода [Текст] / Г. И. Бакланов. – М. : Статистика, 1972. – 72 с.
8. Бешелев, С. Л. Математико-статистические методы экспертных оценок [Текст] / С. Л. Бешелев, Ф. Г. Гурвич. – М. : Статистика, 1980. – 159 с.
9. Болч, Б. Многомерные статистические методы для экономики [Текст] / Б. Болч, К. Хуань. – М. : Статистика, 1979. – 316 с.
10. Боярский. А. Я. Теоретические исследования по статистике [Текст] : сб. научных трудов / А. Я. Боярский. – М. : Статистика, 1974. – 303 с.
11. Вайну, Я. Я. Корреляция рядов динамики [Текст] / Я. Я. Вайну. – М. : Статистика, 1977. – 116 с.
12. Васильев, К. И. К вопросу о методике и технике массовых наблюдений [Текст] / К. И. Васильев. – Л., 1929. – 74 с.
13. Венецкий, И. Г. Основные математико-статистические понятия и формулы в экономическом анализе [Текст] : справочник / И. Г. Венецкий, В. И. Венецкая. – 2-е изд. – М. : Статистика, 1979. – 447 с.
14. Венецкий, И. Г. Теория вероятностей и математическая статистика [Текст] / И. Г. Венецкий, Г. С. Кильдишев. – М. : Статистика, 1975. – 264 с.
15. Вестник Русского географического общества [Текст]. – 1852. – Ч. 5. – Отд. VII.
16. Винн, Р. Введение в прикладной эконометрический анализ [Текст] / Р. Винн, К. Холден. – М. : Финансы и статистика, 1981. – 294 с.
17. Володарский, Л. М. Статистика рассказывает [Текст] / Л. М. Володарский. – М. : Молодая гвардия, 1982.
18. Выборочное наблюдение в статистике СССР [Текст] : сб. статей / под ред. А. Я. Боярского и др. – М. : Статистика, 1966.
19. Вызов, Л. А. Графические методы в статистике, планировании и учете [Текст] : пособие для экономических вузов и для самообразования / Л. А. Вызов. – М. : Госпланиздат, 1940.

20. Герман, К. Ф. Всеобщая теория статистики [Текст] / К. Ф. Герман. – Спб., 1809.
21. Герчук, Я. П. Графические методы в статистике [Текст] / Я. П. Герчук : Статистика, 1968.
22. Гозулов, А. И. Очерки истории отечественной статистики [Текст] / А. И. Гозулов. – М. : Статистика. – С. 972. –312 с.
23. Грачев, Н. Г. Классификация и показатели структуры промышленности [Текст] / Н. Г. Грачев. – М. : Изд-во Академии наук СССР, 1963. – 123 с.
24. Грачев, Н. Г. Статистические группировки [Текст] / Н. Г. Грачев. – М. : Госстатиздат, 1951. – 156 с.
25. Громыко, Г. Л. Статистика [Текст] / Г. Л. Громыко. – М. : Изд-во МГУ им. М. В. Ломоносова, 1981.
26. Гусаров, В. М. Статистические методы моделирования связи [Текст] : учеб. / В. М. Гусаров. – М. : Изд-во Всерос. заочн. фин.-экон. ин-та, 1991.
27. Демографический энциклопедический словарь [Текст]. – М., 1985.
28. Джессен, Р. Методы статистических обследований [Текст] / Р. Джессен ; под ред. Е. М. Четыркина ; пер. с англ. Ю. П. Лукашина и Я. Ш. Паппэ. – М. : Финансы и статистика, 1985.
29. Джини, К. Средние величины [Текст] / К. Джини. – М. : Статистика, 1970. – 447 с.
30. Долгущевский, Ф. Г. Общая теория статистики [Текст] / Ф. Г. Долгущевский, В. С. Козлов, М. И. Полушин, Я. М. Эрлих. – М. : Статистика, 1967. – 384 с.
31. Доугерти, К. Введение в эконометрику [Текст] / К. Доугерти ; пер. с англ. – М. : ИНФРА-М, 1997. – XIV. – 402 с.
32. Дрейпер, Н. Прикладной регрессионный анализ [Текст] / Н. Дрейпер, Г. Смит. – М. : Финансы и статистика, 1986. – Кн. 1. – 366 с.
33. Дрейпер, Н. Прикладной регрессионный анализ [Текст] / Н. Дрейпер, Г. Смит. – М. : Финансы и статистика, 1987. – Книга 2. – 351 с.
34. Дружинин, П. К. Развитие основных идей статистической науки [Текст] / П. К. Дружинин. – М. : Статистика, 1979. – 269 с.
35. Дубров, А. М. Статистические методы многомерной классификации в экономике [Текст] / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин, А. А. Френкель. – М. : МЭСИ, 1984. – 96 с.
36. Дубров, А. М. Факторный и компонентный анализ [Текст] / А. М. Дубров. – М. : МЭСИ, 1989. – 127 с.
37. Дюран, Б. Кластерный анализ [Текст] / Б. Дюран, П. Оделл. – М. : Статистика, 1977. – 128 с.
38. Эренберг, А. Анализ и интерпретация статистических данных [Текст] / А. Эренберг ; под ред. А. А. Рывкина ; пер. с англ. Б. И. Клименко. – М. : Финансы и статистика, 1951.
39. Евзекиэл. М. Методы корреляции и регрессии [Текст] / М. Евзекиэл, К. Фокс. – М. : Статистика, 1966. – 558 с.

40. Ежов, А. И. Выравнивание и вычисление рядов распределений [Текст] / А. И. Ежов. – М. : Госстатиздат, 1961.
41. Ежов, А. И. Организация государственной статистики СССР / А. И. Ежов. – М. : Госстатиздат, 1957. – 132 с.
42. Езекиэл, М. Методы анализа корреляций и регрессий линейных и криволинейных [Текст] / М. Езекиэл, К. А. Фокс. – М. : Статистика, 1966. – 558 с.
43. Елисеева, И. И. Общая теория статистики [Текст] : учеб. для вузов / И. И. Елисеева, М. М. Юзбашев. – М. : Финансы и статистика, 1995.
44. Елисеева, И. И. Статистические методы измерения связей [Текст] / И. И. Елисеева. – Л. : Изд-во ЛГУ, 1982.
45. Ефимова, М. Р. Общая теория статистики [Текст] : учеб. для вузов / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев. – М. : Инфра-М, 1996.
46. Ефимова, М. Р. Общая теория статистики [Текст] : учеб. для вузов / М. Р. Ефимова, В. М. Рябцев. – М. : Финансы и статистика, 1991.
47. Жамбю, М. Иерархический кластерный анализ и соответствия [Текст] / М. Жамбю. – М. : Финансы и статистика, 1988. – 342 с.
48. Закс, Л. Статистическое оценивание [Текст] / Л. Закс. – М. : Статистика, 1976. – 598 с.
49. Иберла, К. Факторный анализ [Текст] / К. Иберла. – М. : Статистика, 1980. – 398 с.
50. Иванов, О. В. Теория статистической группировки [Текст] / О. В. Иванов. – М., 1992. – 91 с.
51. Иващенко, Г. А. Статистическое изучение основной тенденции и взаимосвязи в рядах динамики [Текст] / Г. А. Иващенко, Г. С. Кильдишев, Р. А. Шмойлова. – Томск : Изд-во Томского ун-та, 1985. – 168 с.
52. История советской государственной статистики [Текст]. – Изд. 2-е. – М. : Статистика, 1969. – 527 с.
53. Итоги единовременных обследований по социально-экономическим вопросам, опросов общественного мнения населения Российской Федераций, обследований семейных бюджетов [Текст]. – М. : Росинформцентр Госкомстата РФ, 1994. – Вып. III.
54. Йейтс, Ф. Выборочный метод в переписях и обследованиях [Текст] / Ф. Йейтс ; под. ред. А. Г. Волкова ; пер. с англ. Е. И. Арона. – М. : Статистика, 1965.
55. Казинец, Л. С. Измерение структурных сдвигов в экономике [Текст] / Л. С. Казинец. – М. : Экономика, 1969. – 164 с.
56. Казинец, Л. С. Темпы роста и абсолютные приросты [Текст] / Л. С. Казинец. – М. : Статистика, 1975. –191 с.
57. Казинец, Л. С. Темпы роста и структурные сдвиги в экономике [Текст] / Л. С. Казинец. – М. : Экономика, 1981. – 184 с.
58. Казинец, Л. С. Теория индексов (основные вопросы) [Текст] / Л. С. Казинец. – М. : Госстатиздат, 1963. – 352 с.

59. Карамзин, Н. Изучение вариации сублинейными отклонениями [Текст] / Н. Карамзин // Вестник статистики. – 1986. – № 6. – С. 63–65.
60. Карпенко, Б. И. Развитие идей и категорий математической статистики [Текст] / Б. И. Карпенко. – М. : Наука. 1979. – 376 с.
61. Кауфман, А. А. Статистическая наука в России: теория и методология [Текст] : 1806-1917. Историко-критический очерк / А. А. Кауфман. – М., 1922.
62. Кетле А. Социальная физика [Текст] / А. Кетле. – Киев, 1911. – Т. 1. – С. 16–17.
63. Кильдишев, Г. С. Анализ временных рядов и прогнозирование [Текст] / Г. С. Кильдишев, А. А. Френкель. – М. : Статистика, 1973. – 103 с.
64. Кильдишев, Г. С. Многомерные группировки [Текст] / Г. С. Кильдишев, Ю. И. Аболенцев. – М. : Статистика, 1978. – 160 с.
65. Кимбл, Г. Как пользоваться статистикой [Текст] / Г. Кимбл. – М. : Финансы и статистика, 1982. – 292 с.
66. Ковалева, Л. Н. Многофакторное прогнозирование на основе рядов динамики [Текст] / Л. Н. Ковалева. – М. : Статистика, 1980. – 103 с.
67. Ковалевский, Г. В. Индексный метод в экономике [Текст] / Г. В. Ковалевский. – М. : Финансы и статистика, 1989. – 238 с.
68. Кокрен, У. Методы выборочного исследования [Текст] / У. Кокрен ; под ред. А. Г. Волкова ; пер. с англ. И. М. Сонина. – М. : Статистика, 1976.
69. Колмогоров, А. Предисловие к книге Г. Лебега “Общие величины” [Текст] / А. Колмогоров. – М. : Госстатиздат, 1938. – 4 с.
70. Королев, Ю. Г. Выборочный метод в социологии : учебное пособие [Текст] / Ю. Г. Королев. – М. : МЭСИ, 1975.
71. Королев, Ю. Г. Метод наименьших квадратов в социально-экономических исследованиях [Текст] / Ю. Г. Королев. – М. : Статистика, 1980. – 112 с.
72. Королев, Ю. Г. Статистическое моделирование и прогнозирование [Текст] : учебное пособие / Ю. Г. Королев, П. М. Рабинович, Р. А. Шмойлова. – М. : МЭСИ, 1985. – 103 с.
73. Крылова, Н. И. Ряды распределения [Текст] / Н. И. Крылова. – М. : МИНХ, 1972.
74. Курс лекций по общей теории статистики [Текст] / под ред. В. Е. Овсиенко. – М. : МЭСИ, 1975. – 1-я часть. – 229 с.
75. Курс лекций по общей теории статистики [Текст] / под ред. В. Е. Овсиенко – М. : МЭСИ, 1976. – 231 с.
76. Ланге, О. Теория статистики [Текст] / О. Ланге, А. Банасиньский. – М. : Статистика, 1971. – 399 с.
77. Ленин, В. И. Полн. собр. соч. [Текст] / В. И. Ленин. – Т. 17.
78. Ленин, В. И. Современная статистика [Текст] / В. И. Ленин. – М. : Статистика, 1970-1973. – Т. I-III.
79. Лившиц, Ф. Д. Статистические таблицы [Текст] / Ф. Д. Лившиц. – М. : Госстатиздат, 1958. – 139 с.

80. Лоули, Д. Факторный анализ как статистический метод [Текст] / Д. Лоули, А. Максвелл. – М. : Мир, 1967. – 144 с.
81. Малый, И. Г. Вопросы статистики в “Капитале” К. Маркса [Текст] / И. Г. Малый. – М. : Госстатиздат 1967.
82. Малый, И. Г. Вопросы статистики в трудах Ф. Энгельса [Текст] / И. Г. Малый. – М. : Статистика, 1973. – 140 с.
83. Маркс, К. Соч. [Текст] / К. Маркс, Ф. Энгельс. – Т. 13.
84. Маркс, К. Соч. [Текст] / К. Маркс, Ф. Энгельс. – Т. 23.
85. Маркс, К. Соч. [Текст] / К. Маркс, Ф. Энгельс. – Т. 25. – Ч. I.
86. Маслов, П. П. Техника работы с цифрами [Текст] / П. П. Маслов. – М. : Статистика, 1969. – 120 с.
87. Махалонобис, П. Ч. Выборочные обследования в Индии [Текст] / П. Ч. Махалонобис ; под ред. А. Я. Боярского ; пер. с англ. Л. И. Лукина. – М. : Госстатиздат, 1958.
88. Мендель, И. Д. Кластерный анализ [Текст] / И. Д. Мендель. – М. : Финансы и статистика, 1988. – 176 с.
89. Мердок, Дж. Контрольные карты [Текст] / Дж. Мердок. – М. : Финансы и статистика, 1986.
90. Замков, О. О. Математические методы в экономике [Текст] : учебник / О. О. Замков, А. В. Толстопятенко, Ю. Н. Черемных. – 2-е изд. – М. : МГУ им М. В. Ломоносова ; Дело и сервис, 1999. – 368 с.
91. Миллс, Ф. Статистические методы [Текст] / Ф. Миллс. – М. : Госстатиздат, 1958.
92. Мхитарян, В. С. Дисперсионный анализ [Текст] / В. С. Мхитарян, Л. И. Трошин. – М. : МЭСИ, 1990.
93. Надеждин, Н. И. Об источниках и употреблении статистических сведений [Текст] / Н. И. Надеждин. – Изд. 2-е. – М., 1946.
94. Ненецкий, И. Г. Вариационные ряды и их характеристики [Текст] / И. Г. Ненецкий. – М. : Статистика, 1970. – 159 с.
95. Об источниках и употреблении статистических сведений [Текст]. – Изд. 2-е. – М., 1946.
96. Общая теория статистики [Текст] / Г. С. Килъдишев, В. Е. Овсиенко, Т. В. Рабинович, Т. В. Рябушкин. – М. : Статистика, 1980.
97. Общая теория статистики [Текст] / под ред. А. Я. Боярского, Г. Л. Громыко. – 2-е изд. – М. : Изд-во Московского университета, 1985. – 376 с.
98. Общая теория статистики [Текст] : Статистическая методология в изучении коммерческой деятельности / под ред. А. А. Спирина, О. Э. Башиной. – М. : Финансы и статистика, 1994. – 296 с.
99. Общая теория статистики [Текст] : учеб. для вузов / А. Я. Боярский, Л. Л. Викторова, А. М. Гольдберг и др. – М. : Финансы и статистика, 1985.
100. Общая теория статистики [Текст] : учеб. для вузов / В. С. Козлов, Я. М. Эрлих, Ф. Г. Долгушевский, П. И. Полушин. – М. : Финансы и статистика, 1985.

101. Общая теория статистики : учебник для вузов по специальности “Статистика” [Текст] / под ред. Г. С. Кильдишева, В. Е. Овсиенко, П. М. Рабиновича, Т. В. Рябушкина. – М. : Статистика, 1980. – 423 с.
102. Общая теория статистики [Текст] : учеб. для вузов / Т. В. Рябушкин, М. Р. Ефимова, И. М. Ипатова, Н. И. Яковлева. – М. : Финансы и статистика, 1981.
103. Овсиенко, В. Е. Выбор формы средней и о некоторых ошибках, допускаемых в этом вопросе [Текст] / В. Е. Овсиенко // Вестник статистики. – 1989. – № 2. – С. 16–24.
104. Овсиенко, В. Е. Статистическое наблюдение [Текст] / В. Е. Овсиенко. – М., 1961. – 24 с.
105. Окунь, Я. Факторный анализ [Текст] / Я. Окунь. – М. : Статистика, 1974. – 200 с.
106. Основные организационные положения и календарный план работ по подготовке, проведению и автоматизированной обработке материалов микропереписи населения 1994 г. [Текст]. – М., 1993. – 24 с.
107. Очерки по истории статистики СССР [Текст] : сб. ст. – М. : Госстатиздат, 1955. – Л., 1972. – С. 1–5.
108. Пасхавер, И. С. Закон больших чисел и статистическая закономерность [Текст] / И. С. Пасхавер. – М. : Статистика, 1972. – 149 с.
109. Пасхавер, И. С. Общая теория статистики [Текст] : учеб. пособие / И. С. Пасхавер, А. Л. Яблочник. – М. : Финансы и статистика, 1983.
110. Пасхавер, И. С. Средние величины в статистике [Текст] / И. С. Пасхавер. – М. : Статистика, 1979. – 279 с.
111. Перегудов, В. Н. Теоретические вопросы индексного анализа [Текст] / В. Н. Перегудов. – М. : Госстатиздаг, 1960. – 267 с.
112. Петренко, Е. С. Социально-демографические показатели в социологических исследованиях [Текст] / Е. С. Петренко, Т. М. Ярошенко. – М. : Статистик, 1979.
113. Плошко Б. Г. Группировка и система статистических показателей [Текст] / Б. Г. Плошко. – М., 1994. – 176 с.
114. Плошко, Б. Г. Индексы [Текст] / Б. Г. Плошко. – Л. : Изд-во Ленинградского университета, 1958. – 91 с.
115. Плошко, Б. Г. История статистики [Текст] : учеб. пособие / Б. Г. Плошко, И. И. Елисеева. – М. : Финансы и статистика, 1990. – 295 с.
116. Половников, В. А. Анализ и прогнозирование транспортной работы морского флота [Текст] / В. А. Половников. – М. : Транспорт, 1983. – 222 с.
117. Пономаренко, А. Н. Система национального счетоводства: принципы построения [Текст] / А. Н. Пономаренко, Б. И. Башкатов. – М. : Экономика, 1992. – 47 с.
118. Популярный экономико-статистический словарь-справочник [Текст] / под ред. И. И. Елисеевой. – М. : Финансы и статистика, 1993. – 192 с.

119. Практикум по общей теории статистики [Текст] : учеб. пособие / Н. Н. Ряузов, Н. С. Партешко, А. И. Харламов и др. ; под ред. Н. Н. Ряузова. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 1981.
120. Птуха, М. В. Очерки по истории статистики XVII–XVIII веков [Текст] / М. В. Птуха. – М., 1945. – 359 с.
121. Птуха, М. В. Очерки по истории статистики в СССР [Текст] / М. В. Птуха. – М. : Изд-во АН СССР, 1955. – Т. 1; 1959. – Т. 2. – 476 с.
122. Рабинович, П. М. Некоторые вопросы статистического исследования структуры социально-экономических явлений [Текст] / П. М. Рабинович // Вестник статистики. – 1975. – № 10. – 18–24 с.
123. Рабинович, П. М. Общая теория статистики [Текст] : курс лекций / П. М. Рабинович. – М. : МЭСИ, 1989. – 31 с.
124. Рабочая книга социолога [Текст]. – 2-е изд., перераб. и доп. – М. : Наука, 1983.
125. Рябушкин, Б. Т. Система национальных счетов [Текст] / Б. Т. Рябушкин, Т. А. Хоменко. – М. : Финансы и статистика, 1993. – 51 с.
126. Рябушкин, Т. В. Теоретические концепции в отечественной статистике [Текст] / Т. В. Рябушкин, В. М. Симчера, Е. А. Машихин. – М. : Наука, 1986. – 310 с.
127. Ряузов, Н. Н. Общая теория статистики [Текст] : учеб. для вузов / Н. Н. Ряузов. – М. : Финансы и статистика, 1984.
128. Ряузов, Н. Н. Общая теория статистики [Текст] : учебник / Н. Н. Ряузов. – 4-е изд., перераб. и доп. – М. : Финансы и статистика, 1981.
129. Ряузов, Н. Н. Сводка и группировка статистических материалов [Текст] / Н. Н. Ряузов. – М., 1961. – 48 с.
130. Савицкий, Н. А. Земские подворные переписи (обзор методологии) [Текст] / Н. А. Савицкий. – М. : Госстатиздат, 1961. – 355 с.
131. Садовникова, Н. А. О предварительном анализе исходных данных (Методология моделирования социально-экономического потенциала) [Текст] : сб. научных трудов / Н. А. Садовникова. – М. : МЭСИ, 1990. – С. 17–20.
132. Сиповская, И. В. Выборочное наблюдение и некоторые вопросы его применения в социологических исследованиях [Текст] / И. В. Сиповская. – Тбилиси : Изд-во ТГУ, 1972.
133. Сироткина, Т. С. Основы теории статистики [Текст] : учеб. пособие / Т. С. Сироткина, А. М. Каманина. – М. : Финстатинформ, 1995.
134. Сироткина, Т. С. Статистическое моделирование и прогнозирование [Текст] / Т. С. Сироткина, А. Н. Хорин. – М. : Изд-во Всерос. заочн. фин.-экон. ин-та, 1988.
135. Сиськов, В. И. Корреляционный анализ в экономических исследованиях [Текст] / В. И. Сиськов. – М. : Статистика, 1975. – 168 с.
136. Статистика [Текст] : курс лекций для вузов / под ред. В. Г. Ионина. – М. : ИНФРА-М, 1996.

137. Статистический словарь [Текст] / под ред. М. А. Королева. – М. : Финансы и статистика, 1989. – 623 с.
138. Стоимость жизни и ее измерение [Текст]. – М. : Финансы и статистика, 1991. – 174 с.
139. Суслов, И. П. Общая теория статистики [Текст] : учебник / И. П. Суслов. – М. : Статистика, 1970.
140. Теория статистики : учеб. для вузов [Текст] / под ред. Р. А. Шмойловой. – М. : Финансы и статистика, 1996.
141. Тинтнер, Г. Введение в эконометрию [Текст] / Г. Тинтнер. – М. : Статистика, 1965. – 361 с.
142. Торвей, Р. Индексы потребительских цен [Текст] : методическое руководство / Р. Торвей. – М. : Финансы и статистика, 1993. – 248 с.
143. Успенский, Г. И. Живые цифры : избранные произведения [Текст] / Г. И. Успенский. – М. : Московский рабочий, 1949. – 720 с.
144. Фишер Р. А. Статистические методы для исследователей [Текст] / Р. А. Фишер. – М., 1958.
145. Ферстер, Э. Методы корреляционного и регрессионного анализа [Текст] / Э. Ферстер, Б. Ренц. – М. : Финансы и статистика, 1983. – 302 с.
146. Френкель, А. А. Вариационные ряды и их статистические характеристики [Текст] / А. А. Френкель, Е. В. Адамова. – М. : МЭСИ, 1986.
147. Френкель, А. А. Корреляционный и регрессионный анализ в экономических приложениях [Текст] : учебное пособие / А. А. Френкель, Е. В. Адамова. – М. : МЭСИ, 1987. – 96 с.
148. Хан, Г. Статистические модели в инженерных задачах [Текст] / Г. Хан, С. Шапиро. – М. : Мир, 1969. – 395 с.
149. Хрестоматия по истории русской статистики (история теоретических взглядов) [Текст] / сост. Н. К. Дружинин. – М. : Госстатиздат, 1963. – 294 с.
150. Четыркин, Е. М. Статистические методы прогнозирования [Текст] / Е. М. Четыркин. – М. : Статистика, 1975. – 175 с.
151. Шварц, Г. Выборочный метод [Текст] / Г. Шварц ; под ред. И. Г. Венецкого и В. М. Ивановой ; пер. с нем. Я. Ш. Паппэ. – М. : Статистика, 1978.
152. Юзбашев, М. М. Статистический анализ тенденций и колеблемости [Текст] / М. М. Юзбашев, А. И. Манелля. – М. : Финансы и статистика, 1983.
153. Юл, Э. Дж. Теория статистики [Текст] / Э. Дж. Юл, М. Дж. Кендэл. – М., 1960.
154. Яхот, О. О. Статистическая закономерность в социологическом анализе [Текст] / О. О. Яхот. – М. : Знание, 1964. – 48 с.
155. Barten, A. P. Family Composition, Prices and Expenditure Patterns [Text] : In: Econometric Analysis for National Economic Planning. Ed. P. E. Hart, Colston Papers, 16, Bristol, 1964.

156. Central Statistical Office. Method of Construction nnd Calculation of the Index of Retail Prices [Text] // Studies in Official Statistics, JN<? 6, 4th. ed., HMSO 1967.
157. Bow1ey, A. L. Relations between the Accuracy of an Average and that of its Constituent Parts [Text] // Journal of the Royal Statistical Society. – 1897. – № 60. – P. 855– 866.
158. Brown, J. A. C. and Deaton A. Models of Consumer Behaviour: A Survey [Text] // Economic Journal, 82, 1972, 1145–1236.
159. Brown, R. L., Cow1ey A. H. and Durbin J. Seasonal Adjustment of Unemployment Series. Studies in Official Statistics [Text] // Research Series, Ns 4, HMSO, 1971.
160. Carter, C. F., Rcddawa y W. B. and Stonc Richard. The Measurement of Production Movements [Text]. – Cambridge, 1948.
161. Central Statistical Office. The Index of Industrial Production [Text] // Studies in Official Statistics, Jfe 17, HMSO, 1970.
162. Cost of Living Advisory Committee. Report [Text] // Cmnd. 3677, HMSO, 1968.
163. Court, A. T. Hedonic Price Indexes with Automotive Examples [Text] : In: S. L. Homer et al. The Dynamics of Automobile Demand. – New York, 1939.
164. Afriat, S. N. The Theory of International Comparisons of Real Income and Prices [Text] : In: International Comparisons of Prices and Output, ed. D. J. Daly. – New York, 1972.
165. Durbin, J. and Murphy M. J. Seasonal Adjustment Based on a Mixed Additive-multiplicative Model [Text] // Journal of the Royal Statistical Society, 1975 (forthcoming).
166. Edgeworth, F. Y. The Plurality of Index Numbers [Text] // Economic Journal, 35. 1925, 379–388.
167. Fe11egi, I. P. and Sunter A. B. Balance between Different Sources of Survey Errors [Text] // Bulletin of International Statistical Institute, 45, III, 1973, 334–354.
168. Fisher, F. M. and She11 K. Tastes and Quality Change in the Pure Theory of the True Cost-of-Living Index [Text] : In: Value, Capital and Growth: Papers in Honour of Sir John Hicks. Ed. J. N. Wolfe, Edinburgh, 1968.
169. Fisher, F. M. The Economic Theory of Price Indices [Text]. – New York, 1972.
170. Fisher, Irving. The Making of Index Numbers [Text]. – Boston, 1922.
171. Fow1er, R. F. Further Problems of Index Number Construction. Studies in Official Statistics, Research Series, ,N° 5, HMSO, 1973.
172. Fow1er, R. F. An Ambiguity in the Terminology of Index Number Construction [Text] // Journal of the Royal Statistical Society, A, 137, 1974, 75–88.
173. Frisch, Ragnar. The Problem of Index Numbers [Text] // Econornetrica, 4, 1936, 1–38.

174. Gardner, J. W., Brown R. L. and FrancombeK- Historical Series of the Index of Industrial Production [Text] // Economic Trends, 223, 1972.
175. Geary, R. C. A Note on a Constant-utility Index of the Cost of Living [Text] // Review of Economic Studies, 18, 1950, 64–66.
176. Ha1sbury, Lord (chairman). Report of the Committee of Enquiry on Decimal Currency [Text] // Cmnd. 2145, HMSO, 1963.
177. Hansen, M. H., Hurwitz W. N. and Bershad M. A. Measurement Errors in Censuses and Surveys [Text] // Bulletin of International Statistical Institute, 38, II, 1959, 359–374.
178. Hofsten, E. Witchcraft and Index Numbers [Text] // Malayan Economic Review, 1, 1956, 6–14.
179. Ironmonger, D. S. New Commodities and Consumer Behaviour, unpublished thesis [Text] 1961 and Cambridge, 1972.
180. Jabine, T. B. and Tepping B, J. Controlling the Quality of Occupation and Industry Data [Text] // Bulletin of International Statistical Institute, 45, III, 1973, 360–388.
181. Jazairi, N. T. An Empirical Study of the Conventional and Statistical Theories of Index Numbers [Text] // Oxford Institute of Economics and Statistics Bulletin, 33, 1971,181–195.
182. Kendall, M. G. The Early History of Index Numbers [Text] // International Statistical Review, 37, 1969, 1–12.
183. Keynes, J. M. A Treatise on Money [Text]. – London, 1930. – Vol. 1.
184. K1oek, T. and de Wit G. M. Best Linear and Best Linear Unbiased Index Numbers [Text] // Eccnometrica, 29, 1961, 602–616.
185. Konus, A. A. The Problem of the True Index of the Cost of Living. Institute of Economic Conjuncture [Text] // Economic Bulletin, n. 9–10, Moscow, 1924.
186. Leser, C. E. V. Forms of Engel Functions [Text] // Econometrica, 31, 1963, 694–703.
187. Marks, Peter and Stuart Alan. An Arithmetic Version of the Financial Times Industrial Ordinary Share Index [Text] // Journal Institute of Actuaries, 97, 1971, 297–324.
188. Marshall, Alfred. Money, Credit and Commerce [Text]. – London, 1923.
189. Maunder, W. F. (ed.). Bibliography of Index Numbers [Text] // International Statistical Institute, 1970.
190. McKenzie, L. Demand Functions without a Utility Index [Text] // Review of Economic Studies, 24, 1957, 185–189.
191. Mitchell, B. R. Abstract of British Historical Statistics [Text]. – Cambridge, 1962.
192. Mudgett, B. D. Index Numbers [Text]. – New York, 1951.
193. Mue11bauer, J. Household Production Theory, Quality and the Hedonic Technique and The Cost of Living and Taste and Quality Change, Birkbcck College [Text] // Discussion Papers. – London, 1973.

194. Muc11bauer, J. Prices and Inequality. The United Kingdom Experience [Text] // Economic Journal, 84, 1974 a, 32–55.
195. A11en, R. G. D. and Bow1ey A. L. Family Expenditure [Text]. – London, 1935.
196. Mue1bauer, J. Household Composition, Engel Curves and Welfare Comparisons between Households [Text] // European Economic Review, 5, 1974b, 103–122.
197. Nicho1son, J. L. The Measurement of Quality Changes [Text] // Economic Journal, 77, 19G7, 5.12-530.
198. Paasche, H. Uber die Preisentwicklung der letzten Jahre [Text] // Jahrbiicher fur Nationalokonomie und Statistik, 23, 1874, 168,
199. Prais, S. J. and Houthakker H. S. The Analysis of Family Budgets [Text]. – Cambridge, 1955.
200. Reddaway, W. B. Movements in the Real Product of the United Kingdom, 1946–1949 [Text] // Journal ol the Royal Statistical Society, A, 113, 1950, 435–455.
201. Rich, C. D. Ti.c Rationale of the Use of the Geometric Average as an Investment Index [Text] // Journal Institute of Actuaries, 74, 1948, 338–339.
202. Banerji, K. S. Precision in the Construction of Cost of Living Index Numbers. Sankhya, 21, 1959, 393–400.
203. Roy, R. Les Index econcmiques [Text] // Revue d’Eeonomic Politiquc, 41, 1927, 1251 – 1291.
204. Ruist, Erik. Index Numbers, Theoretical Aspects [Text] // International Encyclopaedia of the Social Sciences, vol. 7, New York, 1908.
205. Saucrbcck, A. Prices of Commodities and the Precious Metals [Text] // Journal of the Royal Statistical Society, 49, 1886, 581.
206. Stig1er, G. J. (chairman). Report on the Price Statistics of the Federal Government [Text] // National Bureau of Economic Research, General Series, Ni 73. – New York, 1961.
207. Stone, Richard. Linear Expenditure Systems and Demand Analysis [Text] // Economic Journal, 64, 1954,511–527.
208. Stone, Richard. Quantity and Price Indexes in National Accounts [Text] // O. E. E. C., Paris, 1956.
209. Stuva1, G. A New Index Number Formula [Text] // Econometrica, 25, 1957, 123–131.
210. Thei1, H. Best Linear Index Numbers of Prices and Quantities [Text] // Econometrica, 28, 1960, 464–480.
211. Ulmer, M. J. The Economic Theory of Cost of Living Index Numbers [Text]. – New York, 1949.
212. Wa1sh, C. H. The Measurement of General Exchange Value [Text]. – New York, 1901.
213. A11en, R. G. D. Index Numbers of Volume and Price [Text] : In: International Trade Statistics, ed. R. G. D. Allen and J. Edward Eiy. – New York., 1953.
214. A1Ien, R. G. D. Macro-economic Theory [Text]. – London, 1968.

215. Adc1man, Irma and Gri1iches Z. On an Index of Quality Change [Text] // Journal of the American Statistical Association, 56, 1961 535–548.
216. Ade1inan, Irma. A New Approach to the Construction of Index Numbers [Text] // Review of Economics and Statistics, 40, 1958, 240–249.
217. Allen, R. G. D. Price Index Numbers [Text] // International Statistical Review, 31, 1963, 281–301.
218. Allen, R. G. D. Sampling for Current Economic Statistics [Text] // Journal of the Royal Statistical Society, A, 127, 1964, 76–88.
219. Banerji, K. S. A Unified Statistical Approach to the Index Number Problem [Text] // Econometrics, 29, 1961, 596–601.
220. Banerji, K. S. Best Linear Unbiased Index Numbers and Index Numbers Obtained through a Factorial Approach [Text] // Econometrica, 31, 1963, 712–718.
221. Bortkiewicz, L. von Zweck. Struktur einer Preisindexzahi [Text] // Nordisk Statistisk Tidskrift, 1, 1922 and 3, 1924.
222. Bow1cy, A. L. The Measurement of the Accuracy of an Average [Text] // Journal of the Royal Statistical Society, 75, 1912, 77–88.
223. Bow1ey, A. L. Elements of Statistics [Text]. – 5th. ed. – London, 1926.
224. Bow1ey, A. L. Wages and Income since 1860 [Text]. – Cambridge, 1937.
225. Beckerman, W. An Introduction to National Income Analysis [Text]. – London, 1968.
226. Beckerman, W. and Bacon P. The International Distribution of Incomes [Text] : In: Unfashionable Economics. Ed. Paul Streeten, London, 1970.
227. Cochran, W. G. Sampling Techniques [Text]. – 2nd ed. – New York, 1962.
228. Craig, J. On the Elementary Treatment of Index Numbers [Text] // Journal of the Royal Statistical Society, C, 18, 1969, 141–152.
229. Central Statistical Office [Text] : National Accounts Statistics, Sources and Methods. 2nd ed., HMSO, 1968.
230. Dorfman, R., Samnelson P. A. and So1ow R. M. Linear Programming and Economic Analysis [Text]. – New York, 1958.
231. Edgeworth, F. Y. Papers Relating to Political Economy [Text]. – London, 1925. – Vol. I.
232. Enge1, E. Die Productions – und Consumptions – Verhaltnisse des Konigreichs Snciisen [Text] // Reprinted in Bulletin de 1'Institut International de Statisti-que, 9, Appendix, 1857, 1895.
233. Frisch, Ragnar. Some Basic Principles of Price of Living Measurements [Text] // Econornetrica, 22, 1954, 407–421.
234. Fowler, R. F. Some Problems of Index Number Construction [Text] // Studies in Official Statistics, Research Series, N° 3, HMSO, 1970.
235. Gri1iches, Z. Hedonic Price Indexes for Automobiles [Text] // In: Stig1er, 1961 and reprinted in: Gri1iches, 1971.
236. Gri1iches, Z. Price Indexes and Quality Change [Text]. – Cambridge, Mass, 1971.
237. Haber1er, G. von. Der Sinn. Der Indexzahlen [Text]. – Tubingen, 1927.

238. Hicks, J. R. Value and Capital [Text]. – 2nd ed. – Oxford, 1946.
239. Houthakker, H. S. Compensated Changes in Quantities and Qualities Consumed [Text] // Review of Economic Studies, 19, 1952, 155–164.
240. Hicks, J.R. A revision of Demand Theory [Text]. – Oxford, 1956.
241. Hofsten, E. von. Price Indices and Quality Changes [Text]. – Stockholm, 1952.
242. Ichimura, S. A Critical Note on the Definition of Related Goods [Text] // Review of Economic Studies, 18, 1951, 179–183.
243. K1ein, L. R. and Rubin H. Constant Utility Index of the Cost of Living [Text] // Review of Economic Studies, 15, 1948, 84–87.
244. Keynes, J. M. A Treatise on Probability [Text]. – London, 1921.
245. Lancaster, K. A New Approach to Consumer Theory [Text] // Journal of Political Economy, 74, 1966, 132–157.
246. Laspeyres, E. Hamburger Warenpreise 1850–1863 [Text] // Jahrbuchcr fur' Nationalokonomie and Statistik, 3, 1864, 81, 209. ,
247. Moroney, M. J. Facts from Figures [Text] / M. J. Moroney. – London, 1951.
248. Muth, R. F. Household Production and Consumer Demand Functions [Text] Econometrica, 34, 1966, 699–708.
249. Peston, M. Unemployment. Why We Need a New Measurement [Text] Lloyds. Bank Review, 104, 1972, 1–7.
250. Richter, M. K. Invariants Axioms and Economic Indexes [Text] // Econometrica. 34, 1S66, 739–755.
251. Retail Prices Index Advisory Committee [Text] Proposals for Retail Prices Indices for Regions, Cmrid. 4719, HMSO, 1971.
252. Divisia, F. L'Indice monetaire et la theorie de la monnaie [Text] Revue d'Eeo-nomie Politiquc, 39, 1925, 980–1008.
253. Wood, G. H. Real Wages and the Standard of Comfort since 1860 [Text] // Journal of the Royal Statistical Society, 72, 1909.
254. Yaari, M. E. Linear Algebra for Social Sciences [Text] Enylevood Cliffs, N. J., 1971.
255. http://www.vuzlib.net/mikro/3.htm.
256. www.megamir.com.ua.

РОЗГОРНУТИЙ ЗМІСТ

Тема 1. СТАТИСТИКА ЯК НАУКА. ЗАДАЧІ ТА МЕТОДИ

1. КОНЦЕПЦІЇ СТАТИСТИКИ ЯК НАУКИ.
   Поняття "статистика".
   Історія статистики.
   Концепції статистичного дослідження.
2. СУТЬ, ПРИНЦИПИ ТА ФУНКЦІЇ СТАТИСТИКИ.
   Предмет та задачі статистики, особливості предмета статистики. Характеристика основних статистичних категорій.
3. ОСОБЛИВОСТІ СТАТИСТИЧНОЇ МЕТОДОЛОГІЇ.
   Статистична методологія, характеристика методів на відповідність стадіям статистичного дослідження.
   Рівні статистики.

   ДОДАТКОВА ІНФОРМАЦІЯ ДО ТЕМИ. ВИКОРИСТАННЯ СТАТИСТИКИ В ЕКОНОМІЧНІЙ ДІЯЛЬНОСТІ ТА РЕГУЛЮВАННЯ ДЕРЖАВНОЇ ЗВІТНОСТІ.
   Законодавче регулювання державної статистичної діяльності.
   Структура органів державної статистики, їх функції, права та обов'язки.
   Нормативна регламентація статистичного спостереження та його видів.

Тема 2. СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ

1. СУТЬ ТА ЕТАПИ СТАТИСТИЧНОГО СПОСТЕРЕЖЕННЯ.
   Зміст і необхідність статистичного спостереження.
   Етапи проведення статистичного спостереження.
   Первинна і вторинна інформація.
2. МЕТОДОЛОГІЯ СТАТИСТИЧНОГО СПОСТЕРЕЖЕННЯ.
   Програмно-методологічні питання статистичного спостереження.
   Організаційні питання статистичного спостереження.
3. СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ.
   Основні види, форми та способи статистичного спостереження.
4. ВЕРИФІКАЦІЯ СТАТИСТИЧНИХ ДАНИХ.
   Точність спостереження.
   Характеристика помилок спостереження.

   ДОДАТКОВА ІНФОРМАЦІЯ ДО ТЕМИ.
   Перепис – як приклад спеціально організованого спостереження.
   Реєстр як форма спостереження.

Тема 3. ОБРОБКА СТАТИСТИЧНИХ ДАНИХ

1. ГРУПУВАННЯ СТАТИСТИЧНИХ ДАНИХ.
   Зведення як основа систематизації та узагальнення статистичних даних.
   Метод групування як основа наукової обробки даних.
   Види статистичних групувань.
   Додаткові питання: поняття про ряд розподілу, їх види, графічне зображення.
2. СТАТИСТИЧНІ ТАБЛИЦІ.
   Поняття статистичної таблиці.
   Особливості побудови.
   Види таблиць за структурною розробкою підмета. Види таблиць за структурною розробкою присудка.
   Додаткові питання: таблиці-матриці та таблиці спряженості як різновиди статистичних таблиць.
3. ВИКОРИСТАННЯ ГРАФІЧНИХ МЕТОДІВ ДЛЯ ЗОБРАЖЕННЯ СТАТИСТИЧНИХ ДАНИХ.
   Поняття та основні елементи статистичного графіка.
   Особливості побудови.
   Класифікація графіків за способом побудови.
   Діаграми порівняння, структури та динаміки як види статистичних графіків за задачами.
   ДОДАТКОВА ІНФОРМАЦІЯ ДО ТЕМИ.
   Статистичні карти.

Тема 4. СТАТИСТИЧНІ ПОКАЗНИКИ

1. ПОНЯТТЯ СТАТИСТИЧНИХ ПОКАЗНИКІВ.
   Статистичний показник, система статистичних показників, конкретний статистичний показник, показник-категорія.
   Класифікація статистичних показників.
2. СУТНІСТЬ І ЗНАЧЕННЯ АБСОЛЮТНИХ ПОКАЗНИКІВ.
   Індивідуальні абсолютні показники, зведені абсолютні показники.
   Одиниці виміру абсолютних величин.
3. ОСОБЛИВОСТІ ВІДНОСНИХ ПОКАЗНИКІВ.
   Сутність відносних показників.
   Відносний показник динаміки (*ВПД*), відносний показник плану (*ВПП*), відносний показник реалізації плану (*ВПРП*), відносний показник структури (*ВПС*), відносний показник координації (*ВПК*), відносний показник інтенсивності (*ВПІ*), відносний показник порівняння (*ВППр*).

4. СУТНІСТЬ І ВИДИ СЕРЕДНІХ ПОКАЗНИКІВ.
   Принципи застосування середніх величин.
   Застосування та розрахунок середніх ступеневих.
   Методика обрахунку моди та медіани як середніх структурних.
   ДОДАТКОВА ІНФОРМАЦІЯ ДО ТЕМИ.
   Властивості середніх величин.

Тема 5. АНАЛІТИЧНА СТАТИСТИКА: ПОКАЗНИКИ ВАРІАЦІЇ, АНАЛІЗ ЧАСТОТНИХ РОЗПОДІЛІВ

1. ПОНЯТТЯ ТА ЗНАЧЕННЯ ВАРІАЦІЇ.
   Сутність варіації.
   Міри варіації, варіація альтернативної ознаки, ентропія розподілу.
   Абсолютні та відносні показники варіації.
   Дисперсія та правило додавання дисперсій.
2. СУТНІСТЬ І ВИДИ РЯДІВ РОЗПОДІЛУ.
   Форми розподілу, теоретичний розподіл в аналізі варіаційних рядів.
   ДОДАТКОВА ІНФОРМАЦІЯ ДО ТЕМИ.
   Використання інтегралу Пуассона при аналізі варіаційних рядів.

Тема 6. РОЗВИТОК СТАТИСТИКИ ЯК НАУКИ. ВІДОКРЕМЛЕННЯ ЕКОНОМЕТРІЇ. ЕКОНОМЕТРИЧНІ МЕТОДИ ТА МОДЕЛІ

1. ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ ЯК ДЖЕРЕЛО СТАТИСТИЧНОЇ ІНФОРМАЦІЇ.
   Способи формування та визначення достатнього обсягу вибіркової сукупності.
   Оцінка точності вибіркових даних.
   Статистична перевірка гіпотез.
2. СТАТИСТИЧНЕ ВИВЧЕННЯ ВЗАЄМОЗВ'ЯЗКУ ЕКОНОМІЧНИХ ПОКАЗНИКІВ. ЕКОНОМЕТРИЧНІ МОДЕЛІ ТА МЕТОДИ.
   Основні задачі кореляційно-регресійного аналізу.
   Функціональні та стохастичні зв'язки, метод найменших квадратів (МНК), умови Гаусса-Маркова відносно випадкового члена, парна регресія, аналіз залишків, коефіцієнт кореляції, *t*-статистика.
   Прогнозування.
   Лінійні та нелінійні залежності.

Багатофакторна регресія.

Частинні та парні коефіцієнти кореляції, коефіцієнт детермінації, статистика Дарбіна-Уотсона, дослідження на гетероскедастичність.

Лінійна та нелінійна моделі.

Функція Кобба-Дугласа.

Поняття еластичності.

Теорія корисності, мотивація поведінки споживача.

Побудова функції корисності.

Криві переваг І байдужості.

Непараметричні методи взаємозв'язків.

Теорія ігор.

Договірні простори Неша, ящик Еджворта, теорія ув'язненого.

ДОДАТКОВА ІНФОРМАЦІЯ ДО ТЕМИ.

Тема 7. ДОСЛІДЖЕННЯ ІНТЕНСИВНОСТІ ДИНАМІКИ ТА ТЕНДЕНЦІЙ РОЗВИТКУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ЯВИЩ

1. ПОНЯТТЯ ТА ВИДИ ДИНАМІЧНИХ РЯДІВ.
   Динамічний ряд, його компоненти, види динамічних рядів.
   Порівнянність рівнів ряду та змикання динамічних рядів.
   Характеристика причин непорівнянності рівнів динамічного ряду.
2. ХАРАКТЕРИСТИКИ ІНТЕНСИВНОСТІ ДИНАМІКИ.
   Показники динамічних змін: ланцюгові та базисні.
   Середні показники динамічних змін: сутність, застосування, розрахунок.
3. АНАЛІЗ ТЕНДЕНЦІЙ РОЗВИТКУ.
   Коливання та їх вплив на динамічний ряд.
   Поняття тенденції розвитку, види тенденцій та їх перевірка.
   Характеристика методів перевірки ряду на наявність тренда.
   Методи вирівнювання.
   Дослідження сезонних коливань.
   Прогнозування та інтерполяція.
   Характеристика методів прогнозування.

Тема 8. ЕКОНОМІЧНІ ІНДЕКСИ

1. ПОНЯТТЯ ТА КЛАСИФІКАЦІЯ ІНДЕКСІВ.
   Індексний метод у проведенні статистичного аналізу.
   Класифікація індексів.
   Індекси кількісних показників.
   Індекси якісних показників.
   Базисні та ланцюгові індекси.

2. ОСОБЛИВОСТІ ТА МЕТОДОЛОГІЯ ПОБУДОВИ ІНДИВІДУАЛЬНИХ І ЗВЕДЕНИХ ІНДЕКСІВ.

   Методологія побудови індексів.

   Характеристика основних індивідуальних і загальних індексів.

   Взаємозв'язок індексів.

3. СУТНІСТЬ СЕРЕДНІХ ІНДЕКСІВ.

   Середньозважені індекси.

   Індекси середнього рівня інтенсивного показника.

4. ІНДЕКСИ ПРОСТОРОВО-ТЕРИТОРІАЛЬНОГО СПІВВІДНОШЕННЯ.

   Побудова територіальних індексів.

   Методи побудови територіальних індексів (метод стандартних вагів).

5. ОСОБЛИВОСТІ ІНДЕКСІВ ЛАСПЕЙРЕСА, ПААШЕ ТА ФІШЕРА.

   Індекси Ласпейреса, Пааше та Фішера: розрахунки та властивості.

   ДОДАТКОВА ІНФОРМАЦІЯ ДО ТЕМИ.

   Індекс-дефлятор у системі національних рахунків.

   Застосування факторного аналізу при дослідженні явищ і процесів.

Навчальне видання

Козьменко Ольга Володимирівна
Меренкова Ольга Віталіївна

СТАТИСТИКА: БАНКІВСЬКИЙ ДОСВІД

Навчальний посібник

У 2 частинах

Частина 2

Редактор *І.О. Кругляк*

Комп'ютерна верстка *В.А. Івакін*

Підписано до друку 23.12.2009. Формат 60х90/16. Гарнітура Times.
Обл.-вид. арк. 8,23. Умов. друк. арк. 12,75. Тираж 90 пр. Зам. № 862

Державний вищий навчальний заклад
"Українська академія банківської справи Національного банку України"
40030, м. Суми, вул. Петропавлівська, 57

Свідоцтво про внесення до Державного реєстру видавців, виготівників
і розповсюджувачів видавничої продукції: серія ДК, № 3160 від 10.04.2008

Надруковано на обладнанні Державного вищого навчального закладу
"Українська академія банківської справи Національного банку України"
40030, м. Суми, вул. Петропавлівська, 57