Е К О Н О М Е Т Р І Я
Частина 2

Міністерство освіти і науки, молоді та спорту України
Вінницький національний технічний університет

Е К О Н О М Е Т Р І Я
Частина 2

Навчальний посібник

Вінниця
ВНТУ
2011

УДК 330.43(075)
ББК 65в6я73
Е45

Автори:
Азарова А. О., Сачанюк-Кавецька Н. В., Роїк О. М., Міронова Ю. В.

Рекомендовано до друку Вченою радою Вінницького національного технічного університету Міністерства освіти і науки, молоді та спорту України (протокол № 4 від 24.11.2011 р.)

Рецензенти:
І. Г. Лук'яненко, доктор економічний наук, професор
В. В. Зянько, доктор економічний наук, професор
О. О. Мороз, доктор економічний наук, професор

Е45 **Економетрія.** Частина 2 : навчальний посібник / [Азарова А. О., Сачанюк-Кавецька Н. В., Роїк О. М., Міронова Ю. В.] – Вінниця : ВНТУ, 2011. – 118 с.

У посібнику розглянуто фундаментальні засади економетричного моделювання. Викладено основи теорії економетричного моделювання за умов кризових явищ в економіці України; наведено основні принципи та методи проведення економетричного дослідження з використанням сучасних інформаційних технологій; викладено методологічні основи до побудови економетричних моделей, що аналізують економічні процеси та явища та дозволяють прогнозувати майбутню поведінку суб'єктів господарювання. Посібник розроблено згідно з планом кафедри та навчальною програмою дисципліни "Економетрія".

УДК 330.43(075)
ББК 65в6я73

ЗМІСТ

ВСТУП

Тенденції розвитку економічної науки за умов безперервно зростаючої складності соціально-економічних процесів, удосконалення форм господарювання, ускладнення управлінських функцій на всіх рівнях ієрархії економічних систем зумовили необхідність широкого застосування в економічних дослідженнях та у практиці управління економікою математичних методів. Застосування математичного моделювання та комп’ютерних технологій в економіці дає змогу якісно й кількісно оновлювати засоби теоретичного аналізу та методи прийняття рішень в усіх галузях економіки, забезпечує можливість підтверджувати або відкидати теоретичні та абстрактно-розрахункові припущення в нових напрямах економічної теорії, прогнозувати та регулювати економічні процеси, не здійснюючи реальних експериментів, які можуть призвести до небажаних соціально-економічних наслідків.

У даній частині навчального посібника розглянуто дві основні найбільш використовувані теми: однофакторний кореляційний аналіз і множинна регресія. У першому розділі посібника розкрито основи кореляційно-регресійного аналізу, загальні підходи до оцінювання невідомих параметрів економетричних моделей, описано методи побудови однофакторної лінійної регресії, проведено повне економетричне дослідження цієї моделі на адекватність і точність.

До істотних особливостей даного розділу слід віднести розгляд нелінійної залежності, моделювання сезонних економічних явищ і наведення чіткого алгоритму розв’язання практичної задачі однофакторного кореляційного аналізу. Друга тема присвячена особливостям побудови, дослідження та використання багатофакторних регресійних моделей.

У кожному розділі навчального посібника наведено велику кількість прикладів, різні вправи, запитання та завдання для самоконтролю. У додатках наведено необхідні для розв’язування задач математико-статистичні таблиці та математичні відомості.

Запитання і завдання для самоконтролю орієнтовані не лише на традиційну перевірку знань, а й на допомогу в заповненні прогалин, що виникли під час засвоєння матеріалу, на індивідуалізацію та інтенсифікацію навчання.

Якщо студенти навчаться моделювати різні економічні ситуації, вони стануть носіями нової ідеології економічного мислення, що ґрунтується на глибоких знаннях сучасних методів системного та економетричного аналізу і комп’ютерних технологій, які застосовуються для дослідження реальних економічних об’єктів, процесів, явищ.

ТЕМА 1 ОДНОФАКТОРНИЙ КОРЕЛЯЦІЙНИЙ АНАЛІЗ

Досить часто в практиці економічних досліджень намагаються визначити криву (поверхню), яка дає найкраще (в сенсі методу найменших квадратів) наближення вихідних даних. Відповідні методи наближення одержали назву регресійного аналізу.

Методи і моделі регресійного аналізу займають центральне місце в математичному апараті економетрії. Задачами регресійного аналізу є з'ясування форми залежності між змінними, оцінювання функції регресії, оцінювання невідомих значень (прогноз значень) залежної змінної.

1.1 Функціональна, статистична та кореляційна залежності

У природничих науках часто йде мова про функціональну залежність, коли кожному значенню однієї змінної відповідає певне значення іншої (наприклад, швидкість вільного падіння у вакуумі в залежності від часу і т.п.).

В економіці у більшості випадків відповідні показники комерційної діяльності можуть знаходитись у нижченаведених видах зв'язку.

1) *Балансовий зв'язок* показників характеризує залежності між джерелами формування ресурсів та їх використанням. Наприклад, формула товарного балансу:

$$З_п+Н=В+З_к,$$

де $З_п$ – залишок товару на початок досліджуваного періоду,
Н – надходження товару,
В – вибуття товару,
$З_к$ – залишок товару на кінець досліджуваного періоду.

2) При *компонентному зв'язку* зміна статистичного показника відбувається під впливом зміни компонент, які входять до даного показника як множники. Наприклад, виручка дорівнює добутку кількості реалізованої продукції на ціну одиниці даної продукції.

3) *Статистичний зв'язок* показників, при якому кожному значенню однієї змінної відповідає множина можливих значень іншої змінної. Такий зв'язок стає можливим через те, що на досліджуваний (залежний) показник впливає велика кількість неврахованих факторів і, до того ж, при оцінюванні значень факторів, що впливають на даний показник, можливі помилки. Прикладом статистичного зв'язку може бути залежність врожайності зернових від кількості внесених добрив.

Важливе місце в економетричних дослідженнях займає *кореляційна залежність*, при якій кожному значенню однієї змінної відповідає умовне середнє (умовне математичне сподівання) значення іншої.

Умовною середньою величиною $\overline{y_x}$ називається середнє арифметичне значення ознаки *Y*, обчислене для конкретного значення показника *X*.

Рівняння виду:

$$\overline{y_x} = f(x) \tag{1.1}$$

називається рівнянням регресії. При цьому $f(x)$ називається регресією *Y* на *X*, а графік цієї функції – лінією регресії. *Залежну* змінну *Y* називають також *функцією відгуку, такою, що результує, ендогенною змінною, результативною ознакою*, а незалежну змінну *X* – *пояснювальною, предикторною, екзогенною змінною, фактором, регрессором, факторною ознакою.*

Іноді кореляційну залежність можна подати у вигляді модельного рівняння регресії:

$$M_x(Y) = \varphi(x) \text{ або } M_y(X) = \psi(y), \tag{1.2}$$

де $\varphi(x) \neq const$, $\psi(y) \neq const$.

Для точного опису рівняння регресії необхідно знати умовний закон розподілу залежної змінної *Y* за умови, що змінна *X* набуде значення x, тобто $X = x$. У статистичній практиці таку інформацію отримати досить складно, оскільки, зазвичай, дослідник використовує лише вибірку пар значень (x_i, y_i) обмеженого обсягу n. У цьому випадку мова може йти лише про оцінку (апроксимацію) за вибіркою функції регресії. Такою оцінкою є вибіркова лінія (крива) регресії:

$$\overline{y_x} = \overline{f}(x, b_0, \ldots, b_p), \tag{1.3}$$

де $\overline{y_x}$ – умовна середня змінної *Y* при фіксованому значенні змінної $X = x$, $b_0, \ldots, b_p$ – параметри кривої. Рівняння (1.3) називають вибірковим рівнянням регресії. За правильно визначеної апроксимаційної функції $\overline{f}(x, b_0, \ldots, b_p)$ зі збільшенням обсягу вибірки вона буде збігатися за ймовірністю до функції регресії $f(x)$.

1.2 Однофакторна лінійна регресія: побудова та оцінювання параметрів

Прості лінійні регресійні моделі встановлюють залежність між двома змінними: факторною ознакою і ознакою, що результує. Наприклад, між витратами на відпустку і складом сім'ї, між витратами на рекламу і обсягом реалізованої продукції.

У загальному вигляді **проста вибіркова лінійна** регресія записується так:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \tag{1.4}$$

де y_i – множина спостережень за змінною $y_i = \{y_1,..., y_n\}$;
x_i – множина спостережень за факторною ознакою $x_i = \{x_1,..., x_n\}$;
ε_i – множина відхилень (помилок) $\varepsilon_i = \{\varepsilon_1, ..., \varepsilon_n\}$;
β_0 та β_1 – істинні параметри зв’язку, значення яких нам невідомі.

Регресійна модель називається **лінійною**, якщо вона лінійна за своїми параметрами. Модель (1.4) називають узагальненою парною лінійною кореляційно-регресійною моделлю.

Параметр β_1 називають коефіцієнтом регресії, параметр β_0 називають вільним членом кореляційно-регресійної моделі. Випадкова величина ε містить різноманітні стохастичні збурення, помилки спостереження та вимірювання, елементи випадковості людської реакції тощо.

Наше завдання – на основі заданих статистичних значень змінних x та y отримати оцінки b_0 та b_1 параметрів β_0 та β_1 (в сенсі певного критерію), тобто побудувати рівняння регресії

$$\hat{y}_i = b_0 + b_1 x_i, \tag{1.5}$$

де $\hat{y}_i$ – оцінене (теоретичне) значення результуючої змінної y_i.

Приклад 1.1. Бюро економічного аналізу фабрики “ROSHEN” оцінює ефективність відділу маркетингу з продажу цукерок і тістечок. Для такого оцінювання є досвід роботи у 5 географічних зонах з майже однаковими умовами (потенційні клієнти, ставлення до товарного знака і т.п.). У цих зонах зафіксовано протягом однакового періоду обсяги продажу (млн. коробок), витрати (млн. грн.) фірми на просування товару на ринку. Вихідні дані наведено в табл. 1.1.

Реальні спостереження y_i зобразимо точками у системі координат (X,Y) на рис. 1.1. Візуально можна припустити, що між даними є лінійна залежність, тобто їх можна апроксимувати прямою лінією.

Табл.1.1 – Вихідні дані для прикладу 1.1

№	y_i	x_i
1	25	5
2	30	6
3	35	9
4	45	12
5	65	18

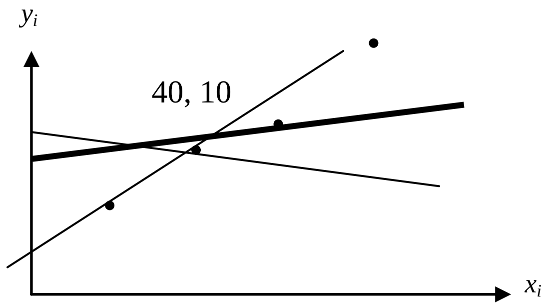


Рисунок 1.1 – Відхилення реальних даних від оцінюваних

Взагалі існує безліч прямих, які можна провести через множину спостережуваних точок. Для цього треба скористатися критерієм, який дає змогу обрати "найкращу" з них, з точки зору цього критерію. Найпоширенішим є *критерій мінімізації суми квадратів відхилень.*

На рис. 1.1 будь-яка з прямих, яку можна провести через дані точки, має точки, які знаходяться над прямою і під нею. Встановимо відхилення від безпомилкової прямої так:

$$e_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_i , \qquad (1.6)$$

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

Відхилення або помилки називають ще залишками. Зрозуміло, що логічно обирати таку пряму, щоб сума квадратів помилок була мінімальною.

У цьому полягає критерій найменших квадратів: невідомі параметри b_0 та b_1 визначаються таким чином, щоб мінімізувати суму квадратів помилок:

$$\sum_{i=1}^{n} e^2_i = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 = Q(b_0, b_1) \rightarrow min \qquad (1.7)$$

Дослідимо функцію $Q(b_0, b_1)$ на мінімум. Необхідною умовою екстремуму є

$$\frac{\partial Q(b_0, b_1)}{\partial b_0} = 0; \; \frac{\partial Q(b_0, b_1)}{\partial b_1} = 0. \tag{1.8}$$

У нашому випадку маємо:

$$\begin{aligned} &\frac{\partial Q(b_0, b_1)}{\partial b_0} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0; \\ &\frac{\partial Q(b_0, b_1)}{\partial b_1} = -2\sum_{i=1}^{n} x_i \cdot (y_i - b_0 - b_1 x_i) = 0, \end{aligned} \tag{1.9}$$

звідки отримаємо систему лінійних рівнянь:

$$\begin{cases} \sum_{i=1}^{n} y_i = n \cdot b_0 + b_1 \cdot \sum_{i=1}^{n} x_i \, ; \\ \sum_{i=1}^{n} y_i x_i = b_0 \cdot \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^{\,2}. \end{cases} \tag{1.10}$$

Систему (1.10) називають *системою нормальних рівнянь* для оцінювання параметрів рівняння регресії.

Виразимо b_0 з першого рівняння і підставимо у друге, отримаємо b_1:

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - \dfrac{\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \dfrac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}. \tag{1.11}$$

З метою спрощення помножимо чисельник і знаменник на $1/n$:

$$b_1 = \frac{\dfrac{1}{n}\left(\sum_{i=1}^{n} x_i y_i\right) - \overline{x}\,\overline{y}}{\dfrac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2} \quad , \quad \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Вираз (1.11) можна записати ще так:

$$b_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \frac{cov(x,y)}{var(x)}. \qquad (1.12)$$

Доведемо це:

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right) = \sum_{i=1}^{n} x_i y_i - \bar{x}\cdot\sum_{i=1}^{n} y_i - \bar{y}\cdot\sum_{i=1}^{n} x_i + n\bar{x}\bar{y} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y};$$

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \sum_{i=1}^{n}\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right) = \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2.$$

Чисельник (1.12) є коефіцієнтом коваріації між *x* та *y*. За визначенням, **коефіцієнт коваріації** оцінюють за формулою:

$$\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\cdot\bar{y}. \qquad (1.13)$$

Коваріація (cov) – це абсолютна міра зв'язку між двома величинами.

Знаменник (1.12) є **дисперсією** величини *x*, тобто:

$$D(X) = \operatorname{var}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2. \qquad (1.14)$$

Отже, b_1 – кут нахилу простої лінійної регресії можна встановити за формулами (1.11) чи (1.12).

З першого рівняння системи (1.10) знайдемо параметр b_0:

$$\frac{1}{n}\sum_{i=1}^{n} y_i - b_0 - b_1\frac{1}{n}\sum_{i=1}^{n} x_i = 0 \quad \Rightarrow \quad b_0 = \bar{y} - b_1\bar{x}.$$

Перевіримо, чи є точка (b_0, b_1) точкою мінімуму. Достатньою умовою мінімуму функції $Q(b_0, b_1)$ є:

$$\Delta = \begin{vmatrix} \frac{\partial^2 Q(b_0,b_1)}{\partial b_1^2} & \frac{\partial^2 Q(b_0,b_1)}{\partial b_1\,\partial b_0} \\ \frac{\partial^2 Q(b_0,b_1)}{\partial b_0\partial b_1} & \frac{\partial^2 Q(b_0,b_1)}{\partial b_0^2} \end{vmatrix} > 0 \quad \text{та} \quad \frac{\partial^2 Q(b_0,b_1)}{\partial b_1^2} > 0.$$

У нашому випадку маємо:

$$\frac{\partial^2 Q(b_0, b_1)}{\partial b_1^2} = 2\sum_{i=1}^{n} x_i^2 > 0;$$

для неперервних функцій $\frac{\partial^2 Q(b_0, b_1)}{\partial b_1 \partial b_0} = \frac{\partial^2 Q(b_0, b_1)}{\partial b_0 \partial b_1}$, тому

$$\frac{\partial^2 Q(b_0, b_1)}{\partial b_0 \partial b_1} = 2\sum_{i=1}^{n} x_i;$$

$$\frac{\partial^2 Q(b_0, b_1)}{\partial b_0^2} = 2n.$$

$$\Delta = \begin{vmatrix} 2\sum_{i=1}^{n} x_i^2 & 2\sum_{i=1}^{n} x_i \\ 2\sum_{i=1}^{n} x_i & 2n \end{vmatrix} = 4\left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \right) = \left\{ \begin{matrix} \textit{домножимо} \quad \textit{і} \quad \textit{поділимо} \\ \textit{даний} \quad \textit{вираз} \quad \textit{на} \quad n^2 \end{matrix} \right\} =$$

$$= 4n^2 \left(\frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n} \right)^2 \right) = 4n^2 \left(\frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2 \right) = 4n^2 \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = 4n\sum_{i=1}^{n} (x_i - \bar{x})^2 > 0,$$

якщо існує хоча б одне значення $x_i \neq \bar{x}$.

Приклад 1.2. Припустимо, що ви збираєте дані про витрати родини на відпустку та її доходи і оцінюєте це таким рівнянням:

$$\hat{y} = 5 + 0{,}86 \cdot x,$$

де $\hat{y}$ – споживання;

x – дохід.

Визначити, що є залежною змінною, а що − незалежною? Поясніть взаємозв'язок між витратами на відпустку родини і її доходом. Наскільки зростуть витрати, якщо дохід зросте на одиницю?

Розв'язання. Оскільки витрати на відпустку залежать від доходу родини, то витрати є залежною, а дохід – незалежною змінною.

Дане рівняння свідчить про те, що витрати складаються з двох частин. Перша частина подана перетином b_0=5, що означає, розмір витрат $\hat{y}$ = 5 при відсутності доходу. Друга частина складається з 0,86·x, тобто зростання доходу на одну одиницю зумовить зростання витрат на відпустку на 0,86.

Приклад 1.3. На підставі даних про валовий регіональний продукт та величину основних засобів по регіонах центру України за 2011 рік (табл. 1.2) побудувати лінійну кореляційно-регресійну модель, яка описує залежність величини валового регіонального продукту від величини основних засобів.

Таблиця 1.2 – Дані для прикладу 1.3

№	Область	Валовий регіональний продукт у фактичних цінах, млн. грн.	Основні засоби у фактичних цінах, млн. грн.
1	Вінницька	8123	25993
2	Дніпропетровська	30040	128686
3	Кіровоградська	5594	19729
4	Полтавська	13983	52493
5	Черкаська	6623	23238

Розв'язання. Для побудови системи нормальних рівнянь (1.10) доцільно скласти допоміжну таблицю (табл. 1.3).

Таблиця 1.3 – Допоміжна таблиця для системи нормальних рівнянь

№	x_i	y_i	x_i^2	$x_i y_i$
1	25993	8123	675636049	211141139
2	128686	30040	16560086596	3865727440
3	19729	5594	389233441	110364026
4	52493	13983	2755515049	734009619
5	23238	6623	540004644	153905274
Разом	250139	64363	2092045779	5075147498

Таким чином маємо:

$$\begin{cases} 64363 = 5 \cdot b_0 + b_1 \cdot 250139; \\ 5075147498 = b_0 \cdot 250139 + b_1 \cdot 2092045779. \end{cases}$$

Розв'язавши одержану систему методом Крамера, знаходимо:

$$b_0 = 21778{,}112276; \qquad b_1 = -0{,}178011.$$

Отже, лінійна кореляційно-регресійна модель, яка описує залежність величини валового регіонального продукту від величини основних засобів, є такою:

$$\hat{y}_i = 21778{,}112276 - 0{,}178011 x_i .$$

1.3 Основні припущення класичного кореляційного аналізу

Розглянемо припущення, які є основою класичного кореляційно-регресійного аналізу (згідно з методом найменших квадратів).

1. Вектор помилок ε_i будемо називати збуренням.
2. Математичне сподівання збурення дорівнює нулю:

$$M(\varepsilon_i) = 0 . \qquad (1.15)$$

(або математичне сподівання залежної змінної y_i дорівнює лінійній функції регресії: $M(y_i) = \beta_0 + \beta_1 x_i$)

3. Дисперсія збурення ε_i (або залежної змінної y_i) є сталою для довільного i:

$$D(\varepsilon_i) = \sigma^2 , \qquad (1.16)$$

(або $D(y_i) = \sigma^2$) – умова гомоскедастичності чи рівнозмінності збурень (залежної змінної)).

4. Збурення ε_i та ε_j є некорельованими (відсутність автокореляції):

$$M(\varepsilon_i \varepsilon_j) = 0, (i \neq j), \qquad (1.17)$$

або

$$cov(\varepsilon_i , \varepsilon_j) = 0 . \qquad (1.18)$$

5. Незалежність значень випадкової величини ε і незалежної змінної x:

$$cov(\varepsilon_i , x_i) = 0 ,\ i = \overline{1,n} . \qquad (1.19)$$

6. Випадкова величина ε_i розподілена за нормальним законом з математичним сподіванням, що дорівнює нулю, та дисперсією σ^2.

Якщо параметри вибіркової лінійної кореляційно-регресійної моделі розраховані за методом найменших квадратів, враховуючи вищевикладені припущення, то

$$M(b_0)=\beta_0,$$

$$M(b_1)=\beta_1,$$

і регресійну модель специфіковано правильно (обрано правильну функціональну форму моделі).

Економетричне дослідження містить етап специфікації моделі, що має бути адекватною економічному об’єкту, процесу, явищу, що вивчається. При специфікації моделі потрібно з’ясувати такі запитання:

1) які змінні потрібно вносити до моделі;
2) якою повинна бути функціональна форма моделі: лінійною чи нелінійною, якщо нелінійною, то якою: степеневою, показниковою тощо;
3) які можливі припущення щодо змінних x, y, ε можна зробити в моделі?

На етапі специфікації економетричної моделі потрібно залучати експертів або проводити послідовні економетричні дослідження для вдосконалення моделі.

Лема 1.1. Якщо під час кореляційно-регресійного аналізу виконуються перераховані вище припущення, то залежна змінна y має нормальний розподіл з математичним сподіванням

$$M(y_i)=\beta_0+\beta_1 x_i$$

та дисперсією

$$D(y_i)=\sigma^2.$$

Доведення

$$M(y_i)=M(\beta_0+\beta_1 x_i+\varepsilon_i)=M(\beta_0)+M(\beta_1 x_i)+M(\varepsilon_i)=\beta_0+\beta_1 x_i,$$

оскільки β_0, β_1 – константи, а за припущенням $M(\varepsilon_i)=0$.

$$D(y_i)=M\left(y_i-M(y_i)\right)^2=M(\beta_0+\beta_1 x_i+\varepsilon_i-\beta_0-\beta_1 x_i)^2=M(\varepsilon_i)^2=\sigma^2$$

Лема 1.2. Якщо параметри лінійної кореляційно-регресійної моделі (1.5) розраховано з використанням методу найменших квадратів за припущень 1–8, то

$$D(b_0)=\frac{\sigma^2\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2};$$

$$D(b_1)=\frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}.$$

Доведення

Дисперсію параметра b_1 визначимо за формулою

$$D(b_1)=M\left(b_1-M(b_1)\right)^2=M(b_1-\beta_1)^2.$$

Позначивши $\alpha_i=\frac{x_i-\overline{x}}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$, можна записати $b_1=\sum_{i=1}^{n}\alpha_i y_i$. Тоді

$$D(b_1)=D\left(\sum_{i=1}^{n}\alpha_i y_i\right)=\sum_{i=1}^{n}\alpha_i^2 D(y_i)=\sum_{i=1}^{n}\alpha_i^2\sigma^2.$$

Знайдемо $\sum_{i=1}^{n}\alpha_i^2$:

$$\sum_{i=1}^{n}\alpha_i^2=\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}\right)^2=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}{\left(\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2\right)^2}=\frac{1}{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}.$$

Остаточно формулу для обчислення дисперсії параметра b_1 подають так:

$$D(b_1)=\frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}.$$

Дисперсію параметра b_0 визначимо за формулою

$$D(b_0)=M\left(b_0-M(b_0)\right)^2=M(b_0-\beta_0)^2.$$

Оскільки $b_0=\overline{y_x}-b_1\overline{x}=\overline{y_x}-\overline{x}\sum_{i=1}^{n}\alpha_i y_i=\frac{\sum_{i=1}^{n}y_i}{n}-\overline{x}\sum_{i=1}^{n}\alpha_i y_i=\sum_{i=1}^{n}\left(\frac{1}{n}-\overline{x}\alpha_i\right)y_i$, то

$$D(b_0)=D\left(\sum_{i=1}^{n}\left(\frac{1}{n}-\bar{x}\alpha_i\right)y_i\right)=\sum_{i=1}^{n}\left(\frac{1}{n}-\bar{x}\alpha_i\right)^2 D(y_i)=\sum_{i=1}^{n}\left(\frac{1}{n}-\bar{x}\alpha_i\right)^2\sigma^2 .$$

Обчислимо $\sum_{i=1}^{n}\left(\frac{1}{n}-\bar{x}\alpha_i\right)^2$:

$$\sum_{i=1}^{n}\left(\frac{1}{n}-\bar{x}\alpha_i\right)^2=\sum_{i=1}^{n}\left(\frac{1}{n^2}-\frac{2\bar{x}\alpha_i}{n}+\bar{x}^2\alpha_i^2\right)=\frac{1}{n}-\frac{2\bar{x}}{n}\sum_{i=1}^{n}\alpha_i+\bar{x}^2\sum_{i=1}^{n}\alpha_i^2=$$

$$=\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2+n\bar{x}^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}\left(x_i^2-2\bar{x}x_i+\bar{x}^2\right)+n\bar{x}^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}=$$

$$=\frac{\sum_{i=1}^{n}x_i^2-2n\bar{x}^2+n\bar{x}^2+n\bar{x}^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

Остаточна формула для обчислення дисперсії параметра b_0 така:

$$D(b_0)=\frac{\sigma^2\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

Виникає природне запитання, чи є оцінки b_0 та b_1 параметрів β_0 та β_1 найкращими? Відповідь на це запитання дає теорема Гаусса-Маркова: *якщо регресійна модель задовольняє припущення 1-5, то оцінки b_0 та b_1 мають найменшу дисперсію в класі усіх лінійних несуміщених оцінок.*

Таким чином, в певному сенсі, оцінки b_0 та b_1 є найбільш ефективними лінійними оцінками параметрів β_0 та β_1.

1.4 Елементи дисперсійного аналізу. Поняття про ступені вільності

У математичній статистиці дисперсійний аналіз розглядають як самостійний метод статистичного аналізу, а в економетрії він застосовується як допоміжний засіб для вивчення якості регресійної моделі. Згідно з основною ідеєю дисперсійного аналізу

$$\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2=\sum_{i=1}^{n}\left[\left(\hat{y}_i-\overline{y}\right)+\left(y_i-\hat{y}_i\right)\right]^2=\sum_{i=1}^{n}\left(\hat{y}_i-\overline{y}\right)^2+\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2+$$

$$+2\sum_{i=1}^{n}\left(\hat{y}_i-\overline{y}\right)\left(y_i-\hat{y}_i\right), \tag{1.20}$$

або

$$SST = SSR + SSE, \tag{1.21}$$

де $SST=\sum_{i=1}^{n}(y_i-\overline{y})^2$ – загальна сума квадратів відхилень залежної змінної від середнього значення;

$SSR=\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2$ – сума квадратів відхилень, зумовлена регресією;

$SSE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ – залишкова сума квадратів, що характеризує невраховані фактори.

Переконаймося, що пропущений у (1.21) третій доданок $2\sum_{i=1}^{n}\left(\hat{y}_i-\overline{y}\right)\left(y_i-\hat{y}_i\right)=0$. Враховуючи, що $b_0=\overline{y}-b_1\overline{x}$ та $\hat{y}_i=\overline{y}+b_1(x_i-\overline{x})$, маємо:

$$\hat{y}_i-\overline{y}=b_1(x_i-\overline{x});$$

$$y_i-\hat{y}_i=y_i-b_0-b_1x_i=y_i-\left(\hat{y}_i-b_1\overline{x}\right)-b_1x_i=\left(y_i-\overline{y}\right)-b_1(x_i-\overline{x});$$

$$2\sum_{i=1}^{n}\left(\hat{y}_i-\overline{y}\right)\left(y_i-\hat{y}_i\right)=2b_1\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})-2b_1^2\sum_{i=1}^{n}(x_i-\overline{x})^2=0$$

(враховуючи рівність (1.12)).

Кожна сума квадратів пов'язана з числом, яке називається її **ступенем вільності**, це число показує, скільки незалежних елементів інформації, які утворюються з елементів ($y_1,...,y_n$) потрібно для розрахунку даної суми квадратів.

Розглянемо, скільки ступенів вільності має кожна, вивчена нами сума квадратів.

Для утворення **SST** потрібно (n - 1) незалежних чисел, тому що з чисел
{($y_1 - \overline{y}$), ($y_2 - \overline{y}$), ... , ($y_n - \overline{y}$)} незалежні тільки (n - 1) завдяки властивості:

$$\sum_{i=1}^{n}(y_i - \bar{y}) = 0$$

Сума квадратів, що пояснює регресію - **SSR** має тільки єдину незалежну одиницю інформації, яка утворюється з $y_1,...,y_n$, а саме b_1. Доведемо це.

Запишемо відхилення, що пояснює регресію, у вигляді:

$$\hat{y}_i - \bar{y} = b_1(x_i - \bar{x})$$

Візьмемо суми з обох боків рівняння і піднесемо їх до квадрату:

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = b_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2$$

Таким чином, дійсно SSR можна утворити, використовуючи лише єдину незалежну одиницю інформації b_1.

Сума квадратів помилок **SSE** має (n - 2) ступеня вільності:

SST = SSE + SSR

У разі простої лінійної регресії: n-1 = n-2 + 1

Ця сума базується на кількості ступенів вільності, яка дорівнює різниці між кількістю спостережень і кількістю параметрів, що оцінюються. У разі простої лінійної регресії оцінюються два параметри b_0 та b_1. Якщо позначити кількість спостережень через n, то для SSE маємо (n - 2) ступеня вільності. Ступені вільності позначаються так: df або Df.

Простий ANOVA-аналіз у лінійній регресії. Використовуючи суми квадратів та відповідні їм ступені вільності, введемо поняття про середні квадрати. Середнім квадратом називається сума квадратів, поділена на відповідний їй ступінь вільності і позначається через:

Середній квадрат помилки – називається сума квадратів помилок, поділена на відповідний ступінь вільності, який позначається через **MSE** і розраховується:

$$MSE = \frac{SSE}{n-m} = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n-m} \qquad (1.22)$$

Середній квадрат, що пояснює регресію, позначається через **MSR** та відповідно дорівнює сумі квадратів, що пояснює регресію, поділену на її ступінь вільності. У разі простої лінійної регресії сума квадратів, що пояснює регресію, має лише один ступінь вільності, тобто середній квадрат збігається з сумою квадратів, а саме:

$$MSR = \frac{SSR}{m-1} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{m-1}. \qquad (1.23)$$

Слід зазначити, що для загальної суми квадратів середній квадрат не розраховується.

Суми квадратів, пов'язані з певними джерелами варіації, а також їх ступенями вільності і середніми квадратами наведено у таблиці 1.4.

Таблиця 1.4 – **ANOVA-таблиця** (дані для дисперсійного аналізу)

Сума квадратів	Число ступенів вільності	Середні квадрати
$SSR = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$	m-1	$MSR = \frac{SSR}{m-1} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{m-1}$
$SSE = \sum_{i=1}^{n}(\hat{y}_i - y_i)^2$	n-m	$MSE = \frac{SSE}{n-m} = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n-m}$
$SST = \sum_{i=1}^{n}(y_i - \overline{y})^2$	n-1	

де n – число дослідів;

m – число параметрів в рівнянні регресії.

1.5 Коефіцієнти кореляції та детермінації

Найпростішим критерієм, який дає кількісну оцінку зв'язку між двома показниками є коефіцієнт *кореляції*. Він розраховується за допомогою такої формули:

$$r = \frac{\mathrm{cov}(x,y)}{\sqrt{\mathrm{var}(x)} \times \sqrt{\mathrm{var}(y)}} = \frac{\mathrm{cov}(x,y)}{\sigma_x \times \sigma_y} \qquad (1.24)$$

$$\sigma_x = \sqrt{\mathrm{var}(x)} \qquad \sigma_y = \sqrt{\mathrm{var}(y)}$$

Коефіцієнт кореляції, на відміну від коефіцієнта коваріації, є не абсолютною, а **відносною мірою зв'язку між двома факторами**, тому значення коефіцієнта кореляції (*corr*) змінюються у межах: -1 до + l, тобто -1 ≤ *corr* ≤ +1

Позитивне значення коефіцієнта кореляції свідчить про прямий зв'язок між показниками, а негативне – про зворотній зв'язок. Коли коефіцієнт кореляції прямує до $r_{xy} \to \pm 1$, це свідчить про наявність щільного зв'язку між двома факторами. У протилежному випадку – коефіцієнт кореляції прямує до нуля ($r_{xy} \to 0$), тобто зв'язку між змінними немає.

Отже, ***властивості коефіцієнта кореляції*** є такими:

1. Якщо $r>0$, то кореляційний зв'язок прямий, при якому збільшення однієї змінної приводить до збільшення іншої змінної.
2. Якщо $r<0$, то кореляційний зв'язок зворотний, при якому збільшення однієї величини приводить до зменшення іншої.
3. Якщо $r = \pm 1$, то кореляційна залежність лінійна, тобто всі емпіричні точки лежать на прямій.
4. $r \in [-1,1]$.
5. Якщо $r = 0$, то лінійний зв'язок відсутній і лінія регресії паралельна вісі Ox.

Зв'язок між ознаками традиційно оцінюють за шкалою Чеддока таким чином:

$0.1 < r < 0.3$ – зв'язок слабкий;
$0.3 < r < 0.5$ – помірний;
$0.5 < r < 0.7$ – помітний;
$0.7 < r_{xy} < 0.9$ – щільний;
$0.9 < r_{xy} < 1$ – дуже щільний зв'язок.

Коефіцієнт кореляції є безрозмірною величиною, яка у даному вигляді характеризує ступінь залежності цих величин, котра виявляється в тім, що при зростанні однієї випадкової величини друга також виявляє тенденцію до зростання (чи убування). У першому випадку говорять, що випадкові величини пов'язані позитивною кореляцією, а в другому – кореляція негативна.

Перевірка адекватності побудованої регресійної моделі реальному економічному явищу (об'єктові), тобто оцінка правильності вибору форми зв'язку між двома змінними здійснюється за допомогою **коефіцієнта детермінації**. Наприклад, якщо перевіряється адекватність оціненої лінійної моделі, то коефіцієнт детермінації визначає, чи справді змінна y лінійно залежить від змінної x. Таким критерієм є :

$$D = \frac{\sigma^2_{регр}}{\sigma^2_{заг}} = \frac{SSR}{SST}, \tag{1.25}$$

де $\sigma^2_{регр}$ – це дисперсія, що пояснює регресію;

$\sigma^2_{заг}$ – це загальна дисперсія;

SST – загальна сума квадратів;

SSR – сума квадратів, що пояснює регресію.

Перепишемо SSR, враховуючи вирази для b_0 та b_1, таким чином:

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n}\left(b_0 + b_1 x_i - \left[b_0 + b_1\bar{x}\right]\right)^2 = b_1^2\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (1.26)$$

Внесемо зміни до (1.25), враховуючи (1.26), помножуючи чисельник і знаменник на $1/n$. Отримаємо:

$$D = SSR/SST = \frac{b_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2 \times \frac{1}{n}}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \times \frac{1}{n}} = \frac{b_1^2 \times \sigma_x^2}{\sigma_y^2} \qquad (1.27)$$

Відомо, що коефіцієнт кореляції:

$$r = \frac{b_1 \times \sigma_x}{\sigma_y} \qquad (1.28)$$

Отже, порівнюючи вирази (1.27) та (1.28), отримаємо:

$$r^2 = D \qquad (1.29)$$

Тоді щільність будь-якої форми зв’язку між ознаками (як лінійної, так і нелінійної) оцінюють за допомогою універсального виразу для коефіцієнта кореляції, який обчислюється за формулою:

$$r = \pm\sqrt{D} = \pm\sqrt{\frac{SSR}{SST}} = \pm\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \qquad (1.30)$$

де $\hat{y}_i$ – оцінене значення змінної y;

$\bar{y}$ – середнє арифметичне змінної y;

y_i – емпіричні (дослідні) значення ознаки y (знак «–» в формулі ставиться при зворотному зв’язку).

Якщо значення коефіцієнта детермінації у межах [0.55 – 1], тобто близьке до одиниці, то модель є адекватною. Якщо ж значення коефіцієнта детермінації знаходиться у межах (0 – 0,44), то модель є неадекватною. У випадку, коли $D \in [0{,}45–0{,}54]$, для перевірки правильності припущення про форму зв’язку між ознаками (адекватності моделі) застосовують F-критерій (критерій Фішера).

Крім коефіцієнта детермінації придатність лінії регресії можна перевірити, визначивши **стандартну помилку оцінювання**:

$$\hat{\sigma}_e = \sqrt{\frac{SSE}{(n-k)}} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-k}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k}} \tag{1.31}$$

де k – кількість оцінюваних параметрів у регресійній моделі. У разі простої лінійної регресії $k = 2$.

Приклад 1.4. За даними прикладу 1.3 оцінити силу кореляційного зв'язку між величиною валового регіонального продукту та величиною основних засобів.

Розв'язання. Для оцінювання сили кореляційного зв'язку між величиною валового регіонального продукту та величиною основних засобів використаємо формулу (1.23), де $\overline{y_x} = 21778{,}112276 - 0{,}178011 x_i$, як було одержано в прикладі 1.3, $\overline{y} = \frac{64363}{5} = 12872{,}6$.

Для зручності розрахунку складемо допоміжну таблицю (табл. 1.5)

Таблиця 1.5 – Допоміжні дані для розрахунку коефіцієнта кореляції

№	x_i	y_i	$\hat{y}_i$	$(\hat{y}_i - \overline{y})^2$	$(y_i - \overline{y})^2$
1	25993	8123	17151,07	18305326	22558700
2	128686	30040	-1129,411	196056320	294719623
3	19729	5594	18266,13	29090201	52978018
4	52493	13983	12433,78	192562	1232988
5	23238	6623	17641,49	22742337	39057500
Разом	250139	64363	64363,07	266386746	410546829

Оскільки пряма $\hat{y}_i = 21778{,}112276 - 0{,}178011 x_i$ є спадною, то

$$r = -\sqrt{\frac{266386746}{410546829}} = -\sqrt{0{,}65} = -0{,}8.$$

Оскільки значення коефіцієнта кореляції близьке за абсолютним значенням до одиниці, то це свідчить про тісний зворотний зв'язок між величиною валового регіонального продукту та величиною основних засобів.

Приклад 1.5. Знайти рівняння регресії залежності між видобутком вугілля на одного шахтаря за зміну (y) та потужністю пласту (x) за такими даними, що характеризують процес видобутку вугілля на 10 шахтах (таблиця 1.6).

Таблиця 1.6 – Дані для прикладу 1.5

i	1	2	3	4	5	6	7	8	9	10
x_i	8	11	12	9	8	8	9	9	8	12
y_i	5	10	10	7	5	6	6	5	6	8

Обчислити коефіцієнт кореляції.

Розв'язання. Для знаходження рівняння регресії складемо систему нормальних рівнянь за допомогою допоміжної таблиці (табл. 1.7).

Таблиця 1.7 – Допоміжна таблиця для прикладу 1.5

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	8	5	64	40	25
2	11	10	121	110	100
3	12	10	144	120	100
4	9	7	81	63	49
5	8	5	64	40	25
6	8	6	64	48	36
7	9	6	81	54	36
8	9	5	81	45	25
9	8	6	64	48	36
10	12	8	144	96	64
$\sum_{i=1}^{10}$	94	68	908	664	496

$$\begin{cases} 68 = 10 \cdot b_0 + 94 b_1; \\ 664 = 94 b_0 + 908 b_1. \end{cases}$$

Розв'язавши цю системи, отримаємо

$$b_0 = 2{,}75, \; b_1 = 1{,}016.$$

Таким чином, рівняння регресії набуває вигляду:

$$\hat{y}_i = 2{,}75 + 1{,}016 x_i .$$

Для знаходження коефіцієнта кореляції використаємо формулу (1.25):

$$r = \frac{10 \cdot 664 - 94 \cdot 68}{\sqrt{10 \cdot 908 - 94^2} \cdot \sqrt{10 \cdot 496 - 68^2}} = 0{,}866,$$

тобто зв'язок між змінними досить тісний.

Одним із важливих питань кореляційного аналізу є перевірка значимості рівняння регресії, тобто встановлення адекватності математичної моделі експериментальним даним (реальному об’єкту). Однією з найбільш ефективних оцінок адекватності регресійної моделі є коефіцієнт детермінації, який обчислюється за формулою:

$$D = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{1.26}$$

або у відсотковому вигляді:

$$D = \frac{SSR}{SST} \cdot 100\%, \tag{1.27}$$

і показує, яка частина варіації змінної Y залежить від змінної X, а яка − від неврахованих факторів. Рівняння регресії буде адекватне експериментальним даним, якщо $D \in [0{,}55;\ 1]$.

1.6 Критерій Фішера

У попередніх підрозділах було показано, що адекватність простої лінійної регресійної моделі можна перевірити за допомогою коефіцієнта детермінації. Якщо його значення близьке до одиниці, то можна вважати, що модель адекватна. Якщо його значення наближується до 0, то модель неадекватна, тобто лінійного зв'язку між двома змінними немає. Але, який висновок можна зробити, якщо значення коефіцієнта детермінації має нечітко виражене граничне значення, наприклад 0.44, 0.45, 0.5, 0.54 і т.п. Зрозуміло, що в таких випадках важко зробити однозначний висновок про наявність зв'язку відповідної форми, тобто про адекватність моделі. Потрібен інший критерій, який би однозначно відповідав на питання про адекватність моделі.

Найпоширенішим серед таких критеріїв є **критерій Фішера**. Він дозволяє перевірити базову гіпотезу (в статистиці вона називається нульовою гіпотезою і позначається H_o), що краще апроксимувати дані середнім значенням ($\hat{y} = \overline{y}$), ніж регресійною прямою ($\hat{y} = \beta_0 + \beta_1 x$). Це в свою чергу дає змогу перевірити наявність або відсутність лінійного зв'язку між змінними, тобто перевірити адекватність побудованої регресійної моделі реальності.

Перевірка моделей на адекватність за F-критерієм Фішера передбачає такі **етапи**:

- На першому етапі розрахунок величини F-відношення:

$$F_{(1,\, n\text{-}2)} = \frac{MSR}{MSE},$$

де *MSR* - середній квадрат, який пояснює регресію;

MSE - середній квадрат помилок.

$$F_{1,n\text{-}2} = \frac{SSR/1}{SSE/n-2} = \frac{MSR}{MSE},$$

де 1, *n*-2 - ступені вільності, що відповідно пов'язані з MSR і MSE.

- На другому етапі задаємо рівень значимості α або $\alpha 100\%$. Наприклад, якщо ми вважаємо, що можлива помилка для нас становить 0,05 (5%), то це означає, що ми можемо помилитися не більше, ніж у 5% випадків, а у 95% = $(100(1\text{-}\alpha))\%$ наші висновки будуть вірними.
- На третьому етапі за статистичними таблицями *F*- розподілу Фішера з (1, *n*-2) ступенями вільності і рівнем значимості $(100(1\text{-}\alpha))\%$. Обчислимо критичне значення $F_{кр}$.
- Якщо розраховане нами значення $F > F_{кр}$, то побудована нами регресійна модель є адекватною (ризик помилитися є не більшим, ніж у 5% випадків) і нульова гіпотеза H_o не справджується.

Приклад 1.6. За даними прикладу 1.5 знайти коефіцієнт детермінації та пояснити його зміст.

Розв'язання. Використовуючи таблицю 1.6, отримаємо

$$SST = \sum_{i=1}^{10}\left(y_i - \overline{y}\right)^2 = \sum_{i=1}^{10} y_i^2 - \frac{\left(\sum_{i=1}^{10} y_i\right)^2}{10} = 496 - \frac{68^2}{10} = 33{,}6;$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 0{,}14 + 2{,}48 + 0{,}31 + 0{,}37 + 0{,}14 + 0{,}39 + 0{,}15 + 1{,}94 + 0{,}39 + 2{,}09 = 8{,}39$$

$$SSR = SST - SSE = 33{,}6 - 8{,}39 = 25{,}21.$$

За формулою (1.26) маємо $D = \frac{25{,}21}{33{,}6} = 0{,}75$. Одержаний результат означає, що варіація залежної змінної Y – добовий видобуток вугілля на одного шахтаря – на 75% пояснюється зміною X – потужністю пласту.

Якщо $D \in [0{,}45; 0{,}55]$, то для встановлення адекватності моделі реальному об'єкту використовують *F*-критерій Фішера, який застосовується так.

1) Обчислюється *F*-відношення за формулою (1.28).

2) Задається рівень значимості $\alpha \cdot 100\%$ (зазвичай, $\alpha \cdot 100\%$ становить 5% або 1%). Наприклад, якщо можлива помилка $\alpha \cdot 100\%$ складає 5%, це

означає, що ми можемо помилитись не більш як в 5% випадків, а в 95% випадків наші висновки правильні.

3) За статистичними таблицями F-розподілу Фішера обчислюється критичне значення $F_{кр} = F_{(\alpha,k_1,k_2)} = F_{(\alpha,m-1,n-m)}$ (додаток А).

4) Якщо $F > F_{кр}$, то з ризиком помилитися не більше ніж в $\alpha \cdot 100\%$ випадках можна стверджувати, що побудована модель адекватна реальному об'єкту.

Приклад 1.7. За даними прикладів 1.5 та 1.6 оцінити на рівні значущість рівняння регресії.

Розв'язання. За формулою (1.28) отримуємо значення F-відношення

$$F = \frac{25{,}21 \cdot (10-2)}{8{,}39} = 24{,}04.$$

За таблицею F-розподілу (додаток Д) $F_{0{,}05;1;8} = 4{,}20$. Оскільки розраховане значення більше за табличне, то з ймовірністю помилитися не більше ніж у 5% випадків можна стверджувати, що рівняння регресії є значущим.

1.6 Стандартна помилка оцінювання. Оцінювання коефіцієнта кореляції

Зобразимо у декартовій системі координат на площині пряму регресії $\overline{y_x} = b_0 + b_1 x$ і пряму середнього арифметичного значення змінної $y = \overline{y}$, яка результує (рис. 1.2).

Із сукупності заданих точок спостережень візьмемо довільну точку $D(x_i, y_i)$ й опустимо з неї перпендикуляр DA на вісь абсцис. Очевидно, що

$$AD = AB + BC + CD,$$

або

$$y_i = \overline{y} + \left(\overline{y_{x_i}} - \overline{y}\right) + \left(y_i - \overline{y_{x_i}}\right),$$

звідки

$$y_i - \overline{y} = \left(\overline{y_{x_i}} - \overline{y}\right) + \left(y_i - \overline{y_{x_i}}\right). \qquad (1.29)$$

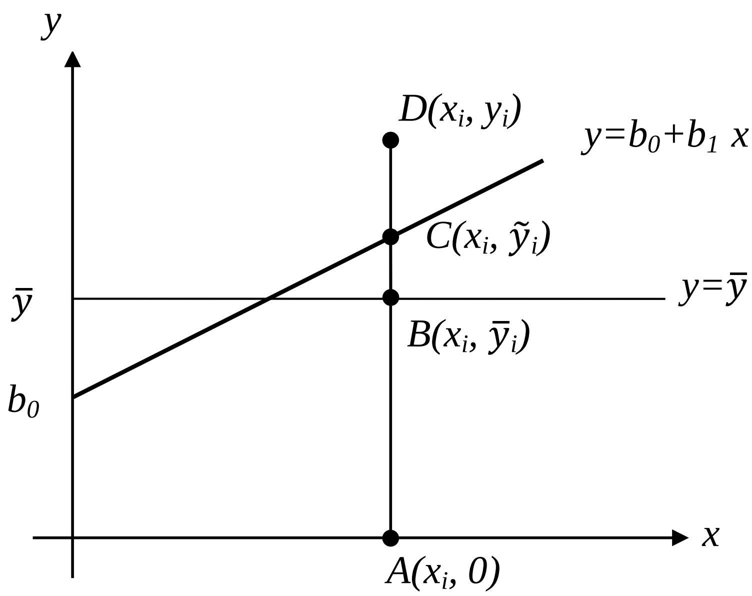


Рисунок 1.2 – Декомпозиція загального відхилення змінної, яка результує

У рівності (1.29) різницю $y_i - \overline{y}$ називають *загальним відхиленням* змінної, яка результує, різницю $\overline{y_{x_i}} - \overline{y}$ називають відхиленням, котре можна пояснити з точки зору кореляційно-регресійної моделі (*пояснене відхилення*). Справді, при будь-якому значенні факторної ознаки x завжди можна знайти величину цього відхилення, маючи тільки кореляційно-регресійну модель, бо $\overline{y}$ залишається незмінною величиною. Різницю $y_i - \overline{y_{x_i}}$ (*випадкове відхилення*) називають ще *непоясненим відхиленням*, тому що це відхилення не можна пояснити з точки зору кореляційно-регресійної моделі. Якщо факторна ознака x змінюється, то змінюються обидві величини y_i та $\overline{y_{x_i}}$, тому, маючи тільки кореляційно-регресійну модель, пояснити це відхилення неможливо.

Рівність (1.29) називають *формулою декомпозиції загального відхилення*: загальне відхилення змінної, яка результує, можна розкласти на пояснене відхилення та непояснене відхилення.

Стандартною похибкою кореляційно-регресійної моделі (*стандартною похибкою оцінки за рівнянням регресії*) називають величину

$$s_{yx} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \overline{y_{x_i}}\right)^2}{n}}\,. \qquad (1.30)$$

Стандартна похибка моделі характеризує розсіювання фактичних значень змінної, яка результує, в околі теоретичних, знайдених за рівнянням регресії. З геометричної точки зору коефіцієнт кореляції характеризує ступінь розсіяності емпіричних точок навколо теоретичної лінії регресії. Чим ближче значення коефіцієнта кореляції до одиниці, тим щільніше розташовані емпіричні точки до лінії регресії.

Щоб обчислити стандартну похибку парної лінійної кореляційно-регресійної моделі, можна використати формули

$$s_{yx} = \sqrt{\frac{\sum_{i=1}^{n} e_i^{\,2}}{n}}, \tag{1.31}$$

або

$$s_{yx} = \sqrt{\frac{\sum_{i=1}^{n} y_i^{\,2} - b_0 \sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i y_i}{n}}. \tag{1.32}$$

Зауважимо, що s_{yx}^2 є зміщеною оцінкою дисперсії випадкових відхилень e_i. Для вибірок малих обсягів використовують незміщену оцінку $S_{yx}^2 = \frac{n}{n-2} s_{yx}^2$.

Якщо взаємозв'язок між змінними x та y функціональний, то всі випадкові відхилення $y_i - \overline{y_{x_i}}$ дорівнюють нулю і, отже, $s_{yx} = 0$.

Якщо кореляційний зв'язок між змінними, що результують, та факторними змінними відсутній, то $\overline{y_{x_i}} = \overline{y}$ і, отже, $s_{yx} = s_y$, тобто стандартна похибка моделі збігається з середнім квадратичним значенням змінної, яка результує.

Отже,

$$0 \le s_{yx} \le s_y.$$

Усі випадкові величини, які ми оцінювали (випадкові відхилення, параметри b_0 та b_1, теоретичні значення змінної $\overline{y_x}$, яка результує), мали нормальний закон розподілу (або близький до нього), тому для їх оцінювання можна було будувати симетричний довірчий інтервал, використовуючи таблиці нормального розподілу або розподілу Стьюдента.

Коефіцієнт кореляції загалом не є нормально розподіленою випадковою величиною, його областю допустимих значень є інтервал [-1, 1].

Найчастіше відмінність розподілу коефіцієнта кореляції від нормального виражена при тісному зв'язку між змінними, тобто коли коефіцієнт кореляції за абсолютним значенням близький до одиниці (рис. 1.3).

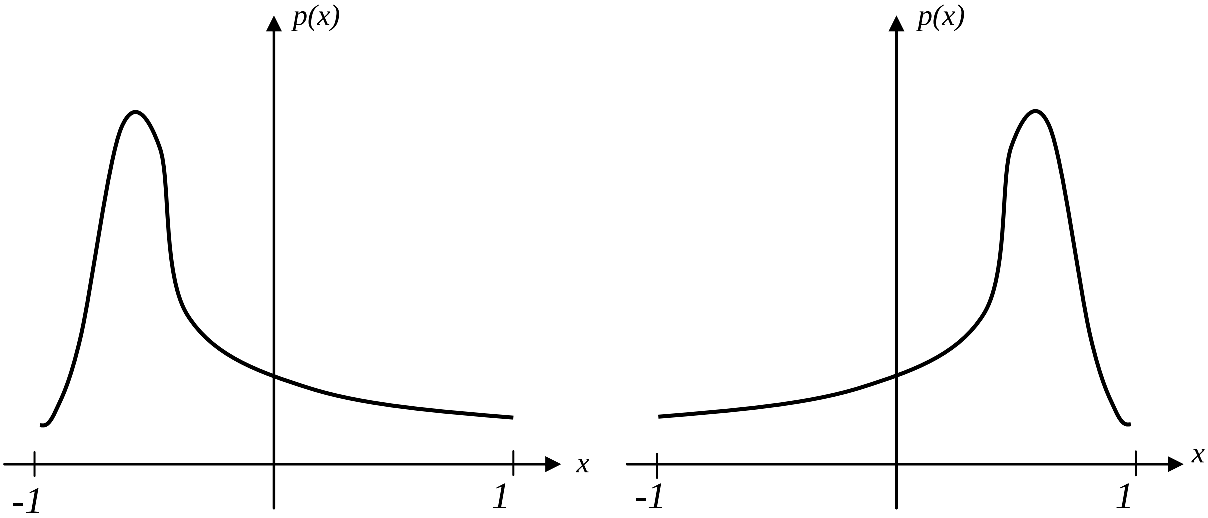


Рисунок 1.3 – Розподіл коефіцієнта кореляції при тісному зв'язку між змінними

Щоб мати змогу оцінити коефіцієнт кореляції у разі, коли його значення не наближене до нуля, Фішер у 1921 році запропонував такий метод.

Спочатку потрібно перейти від випадкової змінної r до випадкової змінної z:

$$z = \frac{1}{2} ln \frac{1+r}{1-r}. \qquad (1.33)$$

Фішер довів, що випадкова змінна z розподілена за нормальним законом з математичним сподіванням:

$$E(z) = \frac{1}{2} ln \frac{1+\rho}{1-\rho}, \qquad (1.34)$$

де ρ – істинне значення коефіцієнта кореляції для генеральної сукупності, та дисперсією

$$D_z = \frac{1}{n-3}. \qquad (1.35)$$

Стандартна вибіркова похибка випадкової змінної z, згідно з (1.35), складає

$$s_z = \frac{1}{\sqrt{n-3}}, \qquad (1.36)$$

а гранична вибіркова похибка z при заданому значенні довірчої ймовірності p –

$$\Delta_z = t_p \cdot s_z .$$

Припустимо, що ς – невідоме значення випадкової змінної z, яке відповідає істинному значенню коефіцієнта кореляції ρ. Тоді довірчий інтервал для змінної ξ має вигляд

$$z_r - t_p \cdot s_z \le \xi \le z_r + t_p \cdot s_z , \qquad (1.37)$$

де z_r – значення випадкової величини z, яке відповідає, згідно з перетворенням (1.33), вибірковому значенню коефіцієнта кореляції r.

Введемо позначення

$$z_1 = z_r - t_p \cdot s_z , \qquad z_2 = z_r + t_p \cdot s_z .$$

Тоді

$$z_1 \le \xi \le z_2 . \qquad (1.38)$$

Здійснивши обернене перетворення від змінної z до змінної r за формулою

$$r = \frac{e^{2z} - 1}{e^{2z} + 1}, \qquad (1.39)$$

отримаємо довірчий інтервал для істинного значення коефіцієнта кореляції ρ генеральної сукупності:

$$r_1 \le \rho \le r_2 , \qquad (1.40)$$

де $r_1 = \frac{e^{2z_1} - 1}{e^{2z_1} + 1}$, $r_2 = \frac{e^{2z_2} - 1}{e^{2z_2} + 1}$.

Якщо значення коефіцієнта кореляції r наближене до нуля, тобто зв’язок між змінними слабкий, то його розподіл наближається до нормального (зазвичай, при великих обсягах вибірки) (рис. 1.4)

У такому разі стандартну похибку коефіцієнта кореляції визначають за формулою

$$s_r = \frac{1-r^2}{\sqrt{n-1}}. \qquad (1.41)$$

Далі оцінювання істинного значення коефіцієнта кореляції ρ генеральної сукупності проводять за звичайною схемою.

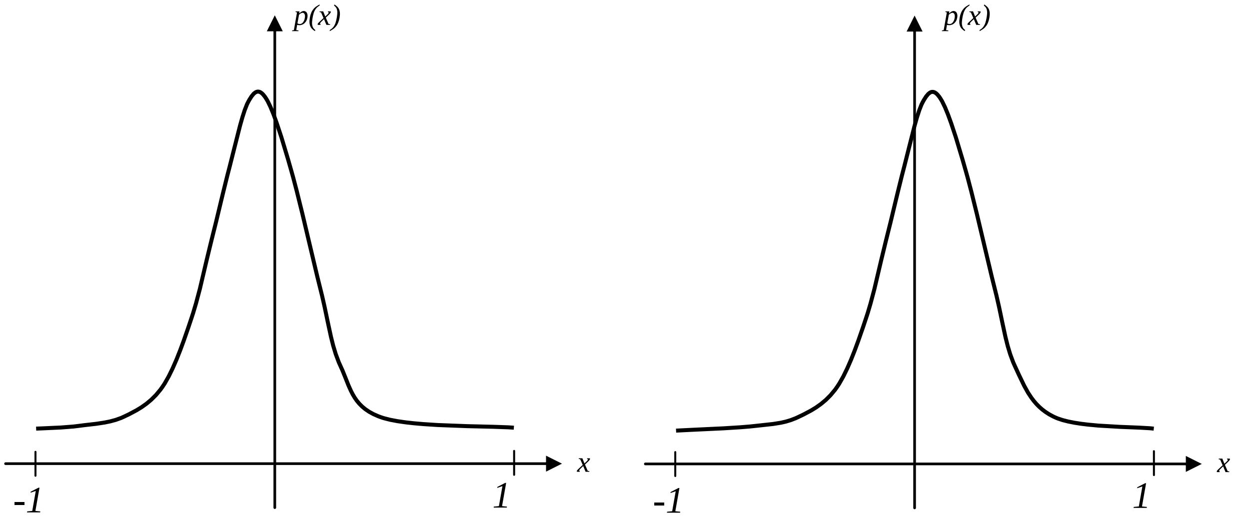


Рисунок 1.4 – Розподіл коефіцієнта кореляції при слабкому зв’язку між змінними

Приклад 1.8. За даними прикладу 1.4 оцінити істинне значення коефіцієнта кореляції між величиною валового регіонального продукту та величиною основних засобів для центральних регіонів України.

Розв’язання. Значення коефіцієнта кореляції, обчислене в прикладі 1.4, дорівнює $r = -0{,}8$, $n = 5$. Отож, для оцінювання коефіцієнта кореляції насамперед перейдемо від випадкової змінної r до випадкової змінної z (1.33):

$$z = \frac{1}{2} ln \frac{1-0{,}8}{1+0{,}8} = \frac{1}{2} ln 0{,}11 \approx -1{,}0985.$$

За формулою (1.36) знайдемо стандартну вибіркову похибку $s_z = \frac{1}{\sqrt{5-3}} = \frac{1}{\sqrt{2}} = 0{,}709;$ задаємо ймовірність $p = 0{,}95$ ($\alpha = 1 - p = 0{,}05$) та знаходимо за таблицею (додаток Б) значення функції Стьюдента $t_p = 3{,}182;$ враховуючи, що $\nu = 5 - 2 = 3$.

Гранична вибіркова похибка z при заданому значенні довірчої ймовірності $p = 0{,}95$ становить

$$\Delta_z = t_p \cdot s_z = 3{,}182 \cdot 0{,}709 = 2{,}256038.$$

За формулами (1.38) маємо:

$$z_1 = -1{,}0985 - 2{,}256038 = -3{,}354538;$$

$$z_2 = -1{,}0985 + 2{,}256038 = 1{,}157538.$$

Тоді за формулами (1.40) будуємо довірчий інтервал:

$$r_1 = \frac{e^{2z_1} - 1}{e^{2z_1} + 1} = \frac{e^{-2\cdot 3{,}354} - 1}{e^{-2\cdot 3{,}354} + 1} \approx \frac{e^{-6{,}708} - 1}{e^{-6{,}708} + 1} \approx \frac{0{,}00122 - 1}{0{,}00122 + 1} \approx -\frac{0{,}99878}{1{,}00122} \approx -0{,}99;$$

$$r_2 = \frac{e^{2z_2} - 1}{e^{2z_2} + 1} = \frac{e^{2\cdot 1{,}157} - 1}{e^{-2\cdot 1{,}157} + 1} \approx \frac{e^{2{,}315} - 1}{e^{2{,}315} + 1} \approx \frac{10{,}1397 - 1}{10{,}1397 + 1} \approx \frac{9{,}1397}{11{,}1397} \approx 0{,}82;$$

$$-0{,}99 \le \rho \le 0{,}82.$$

Отже, з довірчою ймовірністю 0,95 можна стверджувати, що істинне значення коефіцієнта кореляції ρ генеральної сукупності має лежати в межах від -0,99 до 0,82.

Приклад 1.9. За даними прикладу 1.5 оцінити істинне значення коефіцієнта кореляції між видобутком вугілля на одного шахтаря за зміну та потужністю пласту.

Розв'язання. Значення коефіцієнта кореляції, обчислене в прикладі 1.5, дорівнює $r = 0{,}866$; $n = 10$. Отож, для оцінювання коефіцієнта кореляції насамперед перейдемо від випадкової змінної r до випадкової змінної z (1.33):

$$z = \frac{1}{2} ln \frac{1 + 0{,}866}{1 - 0{,}866} = \frac{1}{2} ln 13{,}925 \approx 1{,}317.$$

За формулою (1.36) знайдемо стандартну вибіркову похибку $s_z = \frac{1}{\sqrt{10-3}} = \frac{1}{\sqrt{7}} = 0{,}378$; задаємо ймовірність $p = 0{,}95$ $(\alpha = 1 - p = 0{,}05)$ та знаходимо за таблицею (додаток Б) значення функції Стьюдента $t_p = 2{,}306$; враховуючи, що $\nu = 10 - 2 = 8$.

Гранична вибіркова похибка z при заданому значенні довірчої ймовірності $p = 0{,}95$ становить

$$\Delta_z = t_p \cdot s_z = 2{,}306 \cdot 0{,}378 = 0{,}871668.$$

За формулами (3.38) маємо:

$$z_1 = 1{,}317 - 0{,}871668 = 0{,}890664;$$

$$z_2 = 1{,}317 + 0{,}871668 = 2{,}188668.$$

Тоді за формулами (1.40) будуємо довірчий інтервал:

$$r_1 = \frac{e^{2z_1} - 1}{e^{2z_1} + 1} \approx \frac{e^{0{,}890664} - 1}{e^{0{,}890664} + 1} \approx \frac{2{,}4381 - 1}{2{,}4381 + 1} \approx \frac{1{,}4381}{3{,}4381} \approx 0{,}418;$$

$$r_2 = \frac{e^{2z_2} - 1}{e^{2z_2} + 1} \approx \frac{e^{4{,}377396} - 1}{e^{4{,}377396} + 1} \approx \frac{79{,}8509 - 1}{79{,}8509 + 1} \approx \frac{78{,}8509}{80{,}8509} \approx 0{,}975;$$

$$0{,}418 \le \rho \le 0{,}975.$$

Отже, з довірчою ймовірністю 0,95 можна стверджувати, що істинне значення коефіцієнта кореляції ρ генеральної сукупності має лежати в межах від 0,418 до 0,975.

У ряді прикладних задачах потрібно оцінити значущість коефіцієнта кореляції. При цьому виходять з того, що при відсутності кореляційного зв'язку статистика $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ має t-розподіл Стьюдента із $n-2$ ступенями вільності.

Коефіцієнт кореляції є значущим на рівні α, якщо

$$|t| = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} > t_{1-\alpha;n-2}, \tag{1.42}$$

де $t_{1-\alpha;n-2}$ – табличне значення t-критерію Стбюдента (додаток Б), обчислене на рівня значущості α при числі ступенів вільності $n-2$.

1.7 Приклади нелінійної кореляційної залежності

Пошук рівняння регресії – одна з найважливіших проблем кореляційного аналізу. На практиці в економічних дослідженнях іноді потрібно розглядати нелінійні рівняння регресії.

Як і у випадку лінійної регресії для знаходження оцінок параметрів рівнянь регресії потрібно розв'язати відповідні системи нормальних рівнянь. Наведемо деякі типи нелінійних кореляційних рівнянь та відповідних їм систем нормальних рівнянь, де для спрощення запису замість $\sum_{i=1}^{n}$ записано $\sum$.

Параболічна залежність: $\overline{y_x} = a_0 + a_1 x + a_2 x^2$,

$$\begin{cases} na_0 + a_1 \sum x_i + a_2 \sum x_i^2 = \sum y_i; \\ a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i^3 = \sum x_i y_i; \\ a_0 \sum x_i^2 + a_1 \sum x_i^3 + a_2 \sum x_i^4 = \sum x_i^2 y_i. \end{cases} \quad (1.43)$$

Гіперболічна залежність: $\overline{y_x} = a_0 + \frac{a_1}{x}$,

$$\begin{cases} na_0 + a_1 \sum \frac{1}{x_i} = \sum y_i; \\ a_0 \sum \frac{1}{x_i} + a_1 \sum \frac{1}{x_i^2} = \sum \frac{y_i}{x_i}. \end{cases} \quad (1.44)$$

Показникова залежність: $\overline{y_x} = a_0 \cdot a_1^{\,x}$,

$$\begin{cases} n \ln a_0 + \ln a_1 \sum x_i = \sum \ln y_i; \\ \ln a_0 \sum x_i + \ln a_1 \sum x_i^2 = \sum x_i \ln y_i. \end{cases} \quad (1.45)$$

Степенева залежність: $\overline{y_x} = a_0 \cdot x^{a_1}$,

$$\begin{cases} n \ln a_0 + a_1 \sum \ln x_i = \sum \ln y_i; \\ \ln a_0 \sum \ln x_i + a_1 \sum (\ln x_i)^2 = \sum \ln x_i \ln y_i. \end{cases} \quad (1.46)$$

Зауваження! 1) Якщо позначити $a = ln a_0$, $b = ln a_1$, то показникову залежність можна звести до лінійної $\bar{z} = a + bx$.

2) Щільність показників також оцінюють за допомогою коефіцієнта кореляції, який можна обчислити за однією з формул, наведених у підрозділі 1.5.

3) Адекватність нелінійної моделі встановлюють за допомогою коефіцієнта детермінації, який обчислюють за формулою (1.27).

До нелінійних моделей приводять дослідження динаміки поведінки економічних систем. П. Самуельсон та В. Фріск визначають динамічну систему так: "…систему називають динамічною, якщо її поведінка в часі визначена функціональними рівняннями, в яких змінні в різні моменти часу присутні в явному вигляді". Такі дослідження дають можливість визначити перспективи розвитку цих систем, виявити можливі резерви, розробити комплекс адаптивних управлінських рішень, які забезпечать ефективне функціонування економічних об'єктів. Дослідженням детермінованої поведінки економічних систем в часі та визначенням оптимальної траєкторії їх розвитку займається економічна динаміка. Трендові, лагові та факторні моделі економічної динаміки є економетричними моделями. В основі динамічного аналізу економічних систем, явищ, процесів лежить поняття траєкторії – функції часу, яка описує стан об'єкта дослідження (значення деякого показника):

$$Q = Q(t), \quad t \in [0, T], \tag{1.45}$$

де $[0, T]$ – скінченний відрізок часу, на якому визначена траєкторія.

У теоретичних моделях досліджують також нескінченні траєкторії та траєкторії, для яких початковий момент часу може бути від'ємним. Під час дослідження траєкторії час t можна враховувати дискретно (моментами, інтервалами) або неперервно. Якщо час враховувати дискретно, то моделі економічної динаміки будуть описані скінченно-різницевими рівняннями, якщо ж неперервно – диференціальними рівняннями. *Динамічний (часовий) ряд* – це таблиця значень траєкторії, для якої час змінюється дискретно. За часовою ознакою економічні показники поділяють на моментні та інтервальні. Моментні показники дають кількісну характеристику об'єкта дослідження на деякий момент часу: кількість населення Вінницької області на кінець 2010 року, обсяг основних фондів деякого підприємства на початок 2011 року. Інтервальні показники характеризують об'єкт досліджень за деякий період часу: обсяг продукції, виготовленої деяким підприємством протягом кварталу, прибуток за місяць, виторг торговельного підприємства за один робочий день. Інтервальні показники мають властивість динамічної адитивності, тобто їх можна підсумовувати під час переходу від невеликих проміжків часу до триваліших. Моментні показники не є адитивними в часі. Залежно від показника, значення якого утворює траєкторію, динамічні ряди поділяють на моментні та інтервальні.

У математичній статистиці динамічний ряд розглядають як реалізацію випадкового процесу. У стаціонарних випадкових процесах, для яких характерна рівновага щодо деякого середнього рівня, основні характеристики обчислюють за однією реалізацією процесу (за результатами одного досліду). Проте динамічні ряди економічних показників здебільшого нестаціонарні, їм властиві тенденції, які відображають динамічність економіки. До цих тенденцій належать: нарощування виробничих ресурсів, підвищення науково-технічного рівня, вдосконалення управління системою тощо. Разом з динамічністю, економічним процесам властива інерційність, яка, насамперед, виявляється у вигляді розвитку (напрямок, темпи, коливання).

У складі динамічного ряду можна виділити такі компоненти.

1. Головна (вікова) тенденція, або тренд.
2. Регулярні коливання відносно тренду (цикли).
3. Сезонні коливання відносно тренду.
4. Випадкова компонента, яка відображає вплив факторів стохастичного характеру.

Одним із найважливіших завдань дослідження економічної динаміки є встановлення загальної закономірності (тенденції) розвитку. Для виконання цього завдання використовують різноманітні методи зменшення коливальності ряду, серед яких можна виділити дві групи:

- згладжування ряду за допомогою середніх (методи згладжування);
- аналітичне вирівнювання ряду (моделі тренду).

Детальніше розглянемо аналітичне вирівнювання динамічного ряду – метод вираження головної тенденції розвитку у вигляді функції показника від часу. Таку функцію називають *моделлю тренду*. Динамічний ряд в межах періоду зі стабільними умовами розвитку має деяку закономірність динаміки – головну тенденцію. Різні економічні процеси, або один і той самий процес, але у різні періоди свого розвитку, можуть суттєво відрізнятись за характеристиками розвитку. За основу типізації економічного розвитку використовують динаміку абсолютних приростів.

Базовий абсолютний приріст розраховують як різницю поточного та базового (початкового) рівнів динамічного ряду:

$$\delta_{t/0} = Q_t - Q_0 .$$

Ланцюговий абсолютний приріст розраховують як різницю поточного та базового (попереднього) рівнів динамічного ряду:

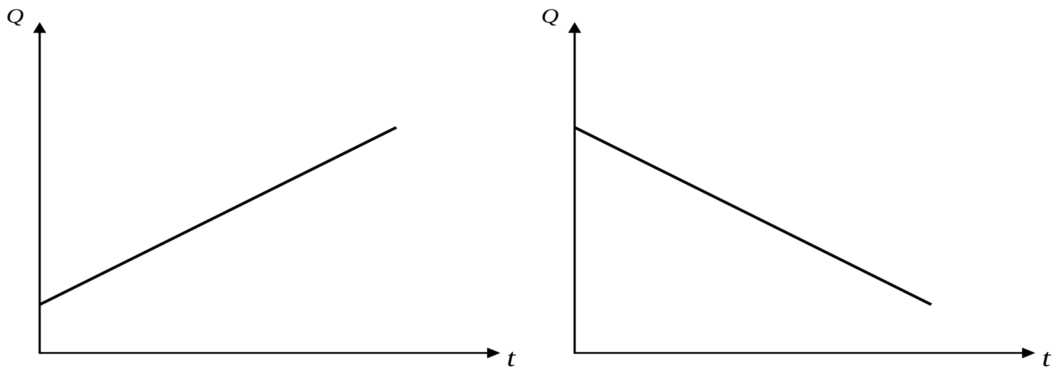
$$\delta_{t/t-1} = Q_t - Q_{t-1} .$$

Абсолютний приріст за одиницю часу характеризує швидкість динаміки. Знак абсолютного приросту засвідчує напрямок динаміки: якщо абсолютний приріст додатний, то значення рівнів ряду зростають, а якщо від'ємний, то спадають.

Базовий темп зростання розраховують як відношення поточного до базового (початкового) рівня динамічного ряду:

$$\eta_{t/0} = \frac{Q_t}{Q_0} .$$

Ланцюговий темп зростання розраховують як відношення поточного до базового (попереднього) рівня динамічного ряду:



$$\eta_{t/t-1} = \frac{Q_t}{Q_{t-1}}.$$

Темп зростання характеризує інтенсивність динаміки. Темп зростання може бути виражений у відсотках або числом, тоді його називають ще коефіцієнтом зростання.

Базовий темп приросту розраховують як відношення базового абсолютного приросту для поточного часового періоду до базового (початкового) рівня динамічного ряду:

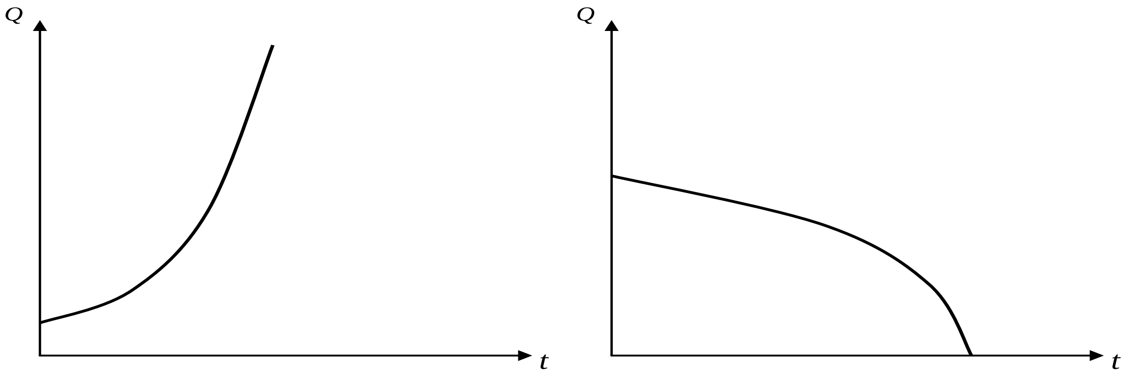
$$\rho_{t/0} = \frac{Q_t - Q_0}{Q_0} = \frac{\delta_{t/0}}{Q_0}.$$

Ланцюговий темп приросту розраховують як відношення ланцюгового абсолютного приросту для поточного часового періоду до базового (попереднього) рівня динамічного ряду:

$$\rho_{t/t-1} = \frac{Q_t - Q_{t-1}}{Q_{t-1}} = \frac{\delta_{t/t-1}}{Q_{t-1}}.$$

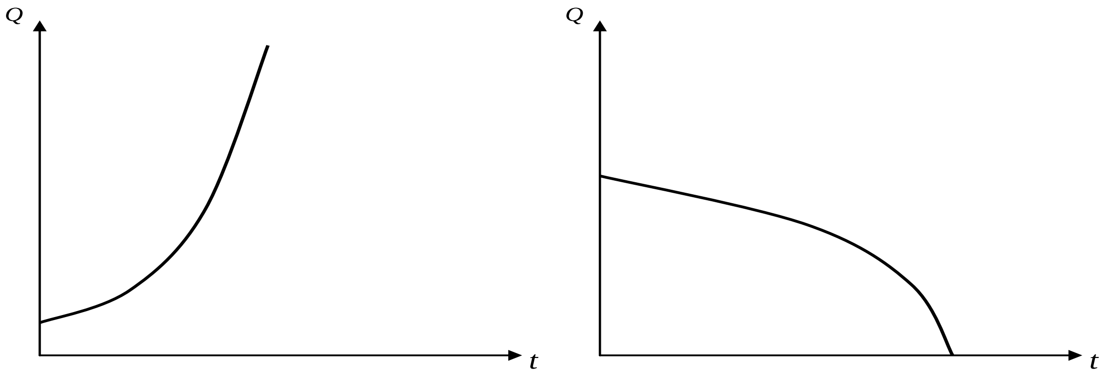
Темп приросту характеризує відносну швидкість, тобто прискорення динаміки. Темп приросту в прикладних застосуваннях виражають у відсотках.

Можна виділити щонайменше чотири типи економічного розвитку (рис. 1.5-1.8).



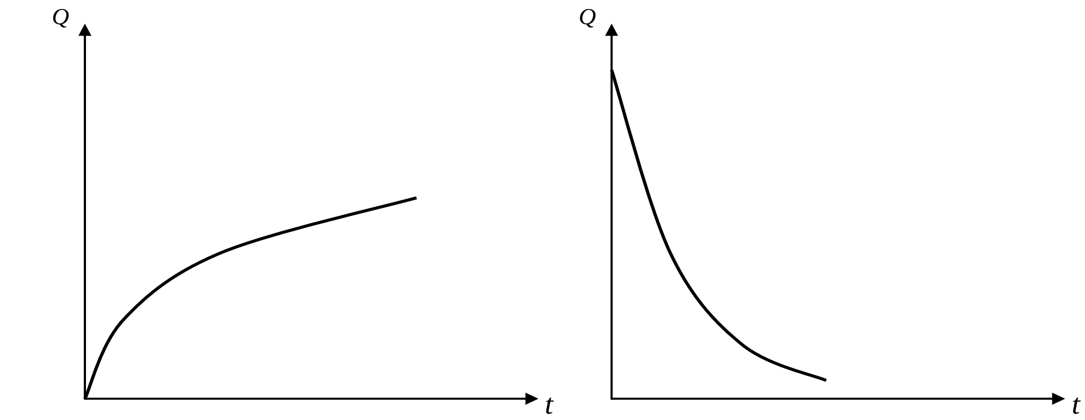
а) Рівномірне зростання б) Рівномірне спадання

Рисунок 1.5



а) Прискорене зростання б) Прискорене спадання

Рисунок 1.6



а) Уповільнене зростання б) Уповільнене спадання

Рисунок 1.7

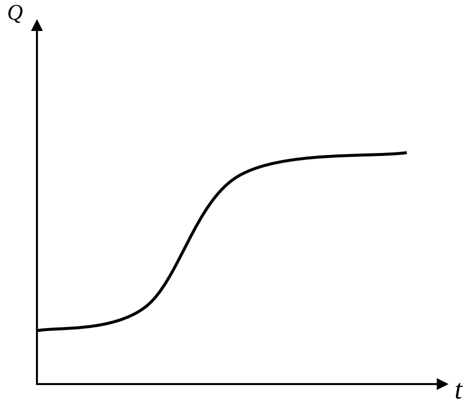


Рисунок 1.8 Розвиток зі зміною характеристик динаміки

1. *Рівномірний (постійний, стабільний) розвиток* – характеризується постійним або близьким до нього абсолютним приростом.
2. *Прискорений розвиток* – характеризується абсолютним приростом, який з плином часу збільшується (для спадання – за абсолютною величиною).
3. *Уповільнений розвиток* – характеризується абсолютним приростом, який з плином часу зменшується (для спадання – за абсолютною величиною).
4. *Розвиток із якісною зміною характеристик динаміки протягом розглядуваного періоду часу.*

Зауваження! 1) Для спрощення викладення матеріалу будемо вважати, що часовим інтервалом у динамічному ряді є один рік.
2) Для всіх трендових моделей вважатимемо, що $t \geq 0$.
3) Для теоретичних досліджень важливою є гладкість моделі тренду. З огляду на це, обираючи трендові моделі, перевагу надають диференційованим функціям.
4) На відміну від фактичної траєкторії динаміки $Q(t)$ модель тренду будемо позначати $x(t)$, а замість фактичного значення показника динаміки ряду Q_t будемо використовувати позначення x_t.

Розглянемо деякі приклади нелінійних моделей.

Найчастіше прискорений розвиток описують показниковою та експоненціальною моделями тренду. Зокрема, показникову модель тренду

$$x(t) = a(1+b)^t, \quad a > 0, \tag{1.47}$$

де a – теоретичний початковий рівень,
$b = \rho$ – дискретний темп росту, використовують для опису дискретних процесів.

Експоненціальну модель тренду

$$x(t) = ae^{bt}, \quad a > 0, \tag{1.48}$$

де a – теоретичний початковий рівень,

$b = \tilde{\rho}$ – неперервний темп приросту:

$$\tilde{\rho}(t) = \frac{d\ln x(t)}{dt} = \frac{d}{dt}(a + bt) = b,$$

використовують для опису неперервних процесів.

Якщо $b > 0$, то трендові моделі (1.47) та (1.48) описують прискорене зростання, а якщо $-1 < b < 0$ – то прискорене спадання.

У межах прискореного розвитку можна виділити динаміку з постійним абсолютним прискоренням $\tilde{\varphi}$. Такий розвиток описують трендовою моделлю параболічного типу:

$$x(t) = a + bt + ct^2, a > 0, b > 0, c > 0. \tag{1.49}$$

Абсолютний приріст динаміки, яку описують трендовою моделлю (1.49),

$$\tilde{\delta}(t) = \frac{dx(t)}{dt} = b + 2ct.$$

Абсолютне прискорення динаміки

$$\tilde{\varphi}(t) = \frac{d\tilde{\delta}(t)}{dt} = \frac{d^2x(t)}{dt^2} = 2c = const,$$

темп приросту динаміки

$$\tilde{\rho}(t) = \frac{d\ln x(t)}{dt} = \frac{b + 2ct}{a + bt + ct^2}.$$

Темп приросту динаміки, яку описують трендовою моделлю (1.49), може змінюватись двома способами: або монотонно спадати; або на початковому інтервалі часу зростати, а потім спадати.

Дослідимо зміну темпу приросту динаміки, яку описують параболічною моделлю, тобто знайдемо проміжки монотонності функції $\tilde{\rho}(t) = \dfrac{b + 2ct}{a + bt + ct^2}$. Для цього розв’яжемо сукупність нерівностей:

$$\left[\begin{array}{l}\dfrac{d\tilde{\rho}(t)}{dt} > 0;\\[2mm] \dfrac{d\tilde{\rho}(t)}{dt} < 0.\end{array}\right.$$

$$\frac{d\tilde{\rho}(t)}{dt} = \frac{2c(a+bt+ct^2)-(b+2ct)^2}{(a+bt+ct^2)^2}.$$

Зрозуміло, що знак похідної $\dfrac{d\tilde{\rho}(t)}{dt}$ збігається зі знаком чисельника

$$-2c^2t^2-2bct+2ac-b^2.$$

Знайдемо проміжки знакосталості параболи $y(t)=-2c^2t^2-2bct+2ac-b^2$.

Для цього дослідимо рівняння:

$$2c^2t^2-2bct+(2ac-b^2)=0;$$

$$2c^2t^2+2bct-(2ac-b^2)=0;$$

$$D=(2bc)^2+4\cdot 2c^2\cdot(2ac-b^2)=4b^2c^2+16ac^3-8b^2c^2=16ac^3-4b^2c^2;$$

$$D=4c^2(4ac-b^2).$$

Якщо $4ac-b^2<0$, то парабола $y(t)=-2c^2t^2-2bct+2ac-b^2$ не перетинається з віссю Ox і має завжди від'ємні значення. Якщо $t>0$, $y(t)=-2c^2t^2-2bct+2ac-b^2<0$, якщо $2ac-b^2<0$, тобто $2ac<b^2$.

Якщо $4ac-b^2\geq 0$, то парабола $y(t)$ має два корені:

$$t_{1,2}=\frac{-2bc\pm 2c\sqrt{4ac-b^2}}{4c^2}=\frac{-b\pm\sqrt{4ac-b^2}}{2c},$$

серед яких (при додатних параметрах b,c і $2ac>b^2$) лише один додатний:

$$t_0=\frac{-b+\sqrt{4ac-b^2}}{2c}.$$

У цій точці $\tilde{\rho}(t)$ досягає свого максимуму і відбувається перехід зростання на спадання.

Отже:

1. Якщо $2ac < b^2$, то $\tilde{\rho}(t)$ монотонно спадає на всьому інтервалі $[0, +\infty)$.

2. Якщо $2ac > b^2$, то на проміжку $\left[0, \dfrac{-b+\sqrt{4ac-b^2}}{2c}\right]$ $\tilde{\rho}(t)$ зростає, а на проміжку $\left[\dfrac{-b+\sqrt{4ac-b^2}}{2c}, \infty\right)$ спадає.

Трендова модель (1.48) відображає основні тенденції розвитку багатьох економічних процесів на сучасному етапі, коли абсолютні прирости продовжують збільшуватись, а темпи приросту спадають.

Моделлю тренду, яка описує динаміку уповільненого розвитку з насиченням може бути гіперболічна трендова модель першого порядку (узагальнена обернена модель) –

$$x(t) = a + \frac{b}{t}. \qquad (1.50)$$

Очевидно, що $\lim\limits_{t \to \infty} x(t) = a$ – межа насичення. Абсолютний приріст $\tilde{\delta}(t) = -\dfrac{b}{t^2}$ поступово спадає за абсолютною величиною.

Якщо $b < 0$, то узагальнена обернена модель описує динаміку уповільненого зростання з верхньою межею (рис. 1.9), а при $b > 0$ – динаміку уповільненого спадання з нижньою межею.

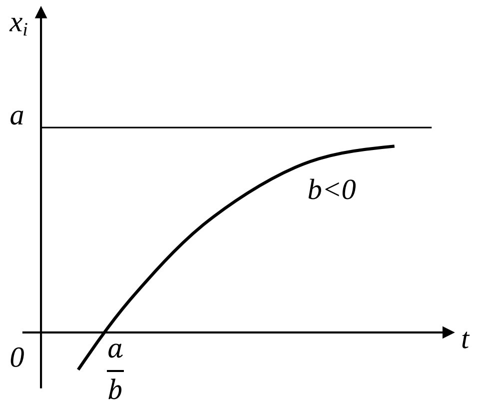


Рисунок 1.9 – Узагальнена обернена модель ($b < 0$)

Прикладами використання зворотної моделі в економічних дослідженнях є криві Енгеля та Філліпса.

Крива Енгеля:

$$x(t) = a + \frac{b}{t}, \quad a > 0, \quad b < 0.$$

Наприкінці XIX ст. німецький статистик Е. Енгель сформулював емпіричні закони споживання і побудував криві, відповідно до яких із зростанням доходу частка витрат на харчування зменшується, а частка витрат на одяг і житло залишається стабільною. Криві, які пов'язують споживчі витрати на товари із загальними витратами або доходом, називають кривими Енгеля. Крива Енгеля (рис. 1.10) для деякого товару вказує на такі особливості:

- критичний рівень доходу $-\frac{b}{a}$, нижче якого товар не буде куплено;
- межу насичення (стелю) a, яку не можна збільшити, скільки б не зростав дохід.

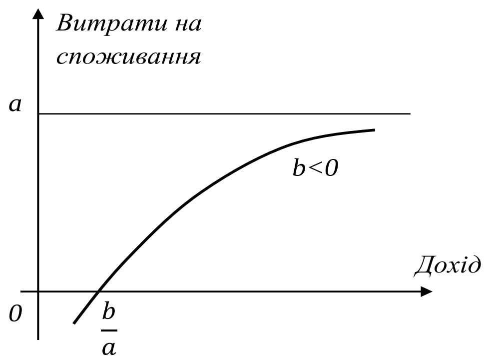


Рисунок 1.10 – Крива Енгеля

Крива Філліпса:

$$x(t) = a + \frac{b}{t}, \quad a < 0, \quad b > 0.$$

Економіст А. Філліпс, аналізуючи дані про норми зміни відсотка заробітної плати і відсотка безробіття для Англії за період від 1861 по 1957 рік, побудував криву, яка описує залежність норми зміни заробітної плати від норми безробіття і яку називають кривою Філліпса (рис. 1.11).

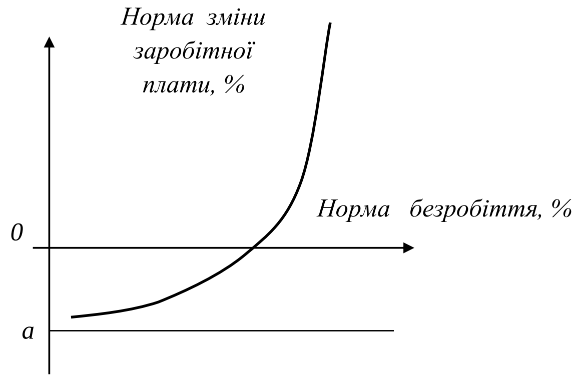


Рисунок 1.11 – Крива Філліпса

Крива Філліпса має такі особливості:
- межа насичення a є межею зміни заробітної плати;
- точка U^N є значенням природної норми безробіття: коли норма безробіття менша за U^N, то норма зміни заробітної плати додатна, коли норма безробіття більша, ніж U^N, норма зміни заробітної плати від'ємна.

Криву Філліпса в макроекономіці використовують, щоб розрахувати мінімальну заробітну плату, компенсацію за безробіття тощо.

Приклад 1.10. На підставі даних про валовий внутрішній продукт України за 2003-2007 роки (табл. 1.8) Побудувати показникову трендову модель. Оцінити адекватність отриманої моделі.

Таблиця 1.8 – Дані для прикладу 1.10

Рік	*Квартал*	*Номер часового періоду*	*Валовий внутрішній продукт, млн. грн*
2003	I	1	52583
	II	2	60798
	III	3	75812
	IV	4	78151
2004	I	5	66981
	II	6	78607
	III	7	99405
	IV	8	100120
2005	I	9	88104
	II	10	101707
	III	11	122861
	IV	12	128780
2006	I	13	105423
	II	14	124116
	III	15	150434
	IV	16	157694
2007	I	17	133108

Розв'язання. Для знаходження параметрів показникової трендової моделі $x(t) = a_0 \cdot a_1^t$ складемо систему нормальних рівнянь (1.44), використавши допоміжну таблицю (табл. 1.9).

Маємо

$$\begin{cases} 17 ln a_0 + 153 ln a_1 = 195{,}1775 \\ 153 ln a_0 + 1785 ln a_1 = 1780{,}895 \end{cases}.$$

Розв’язавши побудовану систему нормальних рівнянь, знаходимо: $ln a_1 \approx 0{,}059$, $a_1 \approx 1{,}0608$, $ln a_0 \approx 10{,}9459$, $a_0 \approx 56718{,}6105$.

Таблиця 1.9 – Таблиця для розрахунку системи нормальних рівнянь

i	t_i	x_i	t_i^2	$ln x_i$	$t_i \, ln x_i$
1	1	52583	1	10,8701	10,8701
2	2	60798	4	11,0153	22,0306
3	3	75812	9	11,236	33,708
4	4	78151	16	11,2664	45,0656
5	5	66981	25	11,1122	55,561
6	6	78607	36	11,2722	67,6332
7	7	99405	49	11,507	80,549
8	8	100120	64	11,5141	92,1128
9	9	88104	81	11,3863	102,4767
10	10	101707	100	11,5299	115,299
11	11	122861	121	11,7188	128,9068
12	12	128780	144	11,7659	141,1908
13	13	105423	169	11,5657	150,3541
14	14	124116	196	11,729	164,206
15	15	150434	225	11,9213	178,8195
16	16	157694	256	11,9684	191,4944
17	17	133108	289	11,7989	200,5813
$\sum$	**153**		**1785**	**195,1775**	**1780,859**

Таким чином, показникова трендова модель набуває вигляду

$$x(t) = 56718{,}6105 \cdot 1{,}0608^t.$$

Для оцінювання отриманої трендової моделі на адекватність використаємо коефіцієнт детермінації $D = \frac{Q_R}{Q} \cdot 100\%$. Для знаходження відповідних сум квадратів відхилень використаємо допоміжну таблицю (табл. 1.10).

Маємо

$$D = \frac{14146836905}{15434707332} \cdot 100\% \approx 92\%,$$

оскільки значення коефіцієнта детермінації близьке до 100%, то модель є адекватною.

Таблиця 1.10 – Допоміжна таблиця для розрахунку сум квадратів відхилень

i	t_i	x_i	$x(t_i)$	$Q_R = (x(t_i) - \bar{x})^2$	$Q = (x_i - \bar{x})^2$
1	1	52583	60167,1	1704442801	2388179161
2	2	60798	63825,26	1415771426	1652747716
3	3	75812	67705,84	1138803467	657409600
4	4	78151	71822,35	877916000,6	542936601
5	5	66981	76189,15	638211503,4	1188249841
6	6	78607	80821,45	425619504,9	521894025
7	7	99405	85735,4	247011627,7	4190209
8	8	100120	90948,11	110331735,9	1774224
9	9	88104	96477,75	24743127,91	178169104
10	10	101707	102343,6	794952,2512	65025
11	11	122861	108566,1	50610303,36	458345281
12	12	128780	115166,9	188098763,8	746819584
13	13	105423	122169,1	429196509,5	15768841
14	14	124116	129596,9	792137488,3	513656896
15	15	150434	137476,4	1297759625	2399236324
16	16	157694	145835	1969850507	3163162564
17	17	133108	154701,8	2835537561	1002102336
$\sum$	**153**		**1709548**	**14146836905**	**15434707332**

1.8 Моделювання сезонних коливань економічних явищ

При порівнянні квартальних та місячних даних багатьох соціально-економічних явищ можна виявити періодичні коливання, що виникають під впливом зміни пір року. Вони є результатом впливу природно-кліматичних умов, загальних економічних факторів Прикладом таких явищ може бути попит на деякі види продукції, пасажирські перевезення, виробництво у сезонних галузях економіки (цукровій, консервній).

У широкому розумінні до сезонних відносять явища, які мають у своєму розвитку чітко виражену закономірність змін протягом року. Періодичні коливання, що мають визначений та сталий період, називають *“сезонними коливаннями”* або *“сезонними хвилями”*, а динамічний ряд в цьому випадку називають *сезонним рядом динаміки*.

Повсякденна життєдіяльність людей за умов періодичної зміни сезонів супроводжується специфічними змінами інтенсивності динаміки соціально-економічних процесів. У більшості галузей народного господарства це проявляється у вигляді річних чергувань підйомів і спадів випуску продукції, різного використання сировини та енергії, коливання рівнів продуктивності праці, собівартості, прибутку й інших показників.

Для деяких сфер людської діяльності річна динаміка характеризується призупиненням процесів у міжсезонні періоди (цукроваріння, рибальство, мисливство, навігація, туризм і т.д.). Яскраво виражений сезонний характер має сільськогосподарське виробництво, особливо рослинництво за умов відкритого ґрунту. Це викликає нерівномірність використання трудових ресурсів, напруженість в роботі транспорту, сховищ, баз. Сезонне коливання характерне також для грошового та товарного оборотів.

У практичній діяльності люди, впливаючи на природу, створюють більш сприятливі умови праці та побуту. Однак, на даному етапі свого розвитку, людство не може керувати всіма силами природи. Наприклад, неможливо на власний розсуд змінювати час настання та тривалість несприятливих сезонів, наявність стихійних лих.

Знання сезонних особливостей попиту на окремі товари має важливе значення для торгівлі: розробка заходів підвищення ефективності торгівлі, покращення організації торгівлі, підвищення культури обслуговування покупців. З'ясування особливостей попиту населення на товари за сезонами важливе для розробки науково обґрунтованих нормативів, воно дозволяє уникнути нераціональних витрат та втрат.

Статистичні ряди внутрішньорічної динаміки зазвичай складають за матеріалами поточної звітності. Однією з необхідних умов вивчення сезонних коливань є зведення рядів динаміки до зіставного виду. При цьому необхідно мати на увазі, що різновеликі за тривалістю місяці та квартали річних періодів є однією з причин, які впливають на зміни рівнів рядів внутрішньорічної динаміки. Для усунення цієї причини об'ємні величини перераховують в середні, які характеризують інтенсивність розвитку досліджуваного явища за одиницю часу.

Для аналітичного вирівнювання, як методу виміру сезонних хвиль, використовують вирівнювання за рядом Фур'є, який подає сезонне явище у вигляді гармоніки.

Ряд Фур'є у загальному випадку можна записати у вигляді такої залежності:

$$\overline{y_t} = a_0 + \sum_{k=1}^{m}(a_k \cos kt + b_k \sin kt), \qquad (1.51)$$

де a_0, a_k, b_k – параметри досліджуваної моделі;

t – фактор часу;

k – порядковий номер гармоніки;

m – номер гармоніки, яка використовується з різним ступенем точності (зазвичай, від однієї до чотирьох).

Звичайно кількість гармонік при побудові ряду Фур’є необхідно приймати не більше чотирьох. Після побудови усіх гармонік проводять оцінювання їх адекватності, тобто визначають, яка з гармонік найкраще описує сезонні коливання економічних явищ.

Як і в будь-якій моделі, необхідно оцінити її параметри. Оцінювання параметрів ряду Фур’є проводиться методом найменших квадратів за такими залежностями:

$$a_0 = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i = \bar{y}, \tag{1.52}$$

$$a_k = \frac{2}{n} \sum_{i=1}^{n} y_i \cos kt_i , \tag{1.53}$$

$$b_k = \frac{2}{n} \sum_{i=1}^{n} y_i \sin kt_i , \tag{1.54}$$

де n − це кількість часових періодів, за які розглядається досліджуване явище (дні, місяці, квартали, роки), t_i – періоди, y_i – рівні динаміки.

Необхідно зауважити, що для побудови економетричної моделі потрібно для такого фактора, як час, зробити перерахунок від натурального масштабу до радіанного або ж градусного. Такий перехід пропонується здійснити за формулою:

$$t = \frac{2\pi}{n} \cdot t_i , \tag{1.55}$$

де t_i – фактор часу у радіанному або ж градусному вимірі;

n – кількість спостережень (кількість інтервалів часу).

Можна подати місяці року у радіанній формі (таблиця 1.11).

Таблиця 1.11 – Радіанна форма подання місяців року

№ місяця	Місяць	Радіани	№ місяця	Місяць	Радіани
1	Січень	0	7	Липень	π
2	Лютий	π/6	8	Серпень	7π/6
3	Березень	π/3	9	Вересень	4π/3
4	Квітень	π/2	10	Жовтень	3π/2
5	Травень	2π/3	11	Листопад	5π/3
6	Червень	5π/6	12	Грудень	11π/6

Перед тим, як визначати параметри економетричної моделі, необхідно скласти таблицю значень тригонометричних функцій (табл. 1.12).

Таблиця 1.12 – Таблиця значень тригонометричних функцій

№	t_i, рад.	$\cos t_i$	$\cos 2\,t_i$	$\sin t_i$	$\sin 2\,t_i$
1	0	1	1	0	0
2	π/6	0,866	0,5	0,5	0,866
3	π/3	0,5	-0,5	0,866	0,866
4	π/2	0	-1	1	0
5	2π/3	-0,5	-0,5	0,866	-0,866
6	5π/6	-0,866	0,5	0,5	-0,866
7	π	-1	1	0	0
8	7π/6	-0,866	0,5	-0,5	0,866
9	4π/3	-0,5	-0,5	-0,866	0,866
10	3π/2	0	-1	-1	0
11	5π/3	0,5	-0,5	-0,866	-0,866
12	11π/6	0,866	0,5	-0,5	-0,866

Розглянемо застосування ряду Фур'є на конкретному прикладі.

Приклад 1.11. Дано статистичні щомісячні дані щодо реалізації зимового одягу (табл. 1.13). Необхідно описати сезонні коливання реалізації зимового одягу на основі рядів Фур'є (з використанням двох гармонік) і обрати гармоніку, яка найбільш адекватно описує ці сезонні коливання.

Таблиця 1.13 – Дані прикладу 1.11

№	Місяці	t_i, рад.	y_i
1	Січень	0	37
2	Лютий	π/6	40
3	Березень	π/3	44
4	Квітень	π/2	52
5	Травень	2π/3	46
6	Червень	5π/6	70
7	Липень	π	60
8	Серпень	7π/6	48
9	Вересень	4π/3	46
10	Жовтень	3π/2	38
11	Листопад	5π/3	36
12	Грудень	11π/6	35
Сума			552

Розв'язання. Для подальших розрахунків необхідно скласти допоміжну таблицю (табл. 1.14) для першої гармоніки, що значно полегшить оцінювання параметрів моделі.

Таблиця 1.14 – Допоміжні розрахунки для першої гармоніки

№	$y_i \cos t_i$	$y_i \sin t_i$	$\overline{y_{1,t_i}}$	$\left(\overline{y_{1,t_i}} - \overline{y}\right)^2$	$\left(y_i - \overline{y}\right)^2$
1	37,000	0,000	34,96	121,8816	81
2	34,641	20,000	39,31	44,7561	36
3	22,000	38,105	45,45	0,3025	4
4	0,000	52,000	51,74	32,9476	36
5	-23,000	39,837	56,49	110,0401	0
6	-60,622	35,000	58,43	154,5049	576
7	-60,000	0,000	57,04	121,8816	196
8	-41,569	-24,000	52,69	44,7561	4
9	-23,000	-39,837	46,55	0,3025	0
10	0,000	-38,000	40,26	32,9476	64
11	18,000	-31,177	35,51	110,0401	100
12	30,311	-17,500	33,57	154,5049	121
Σ	**-66,239**	**34,428**	**552,0**	**928,8656**	**1218**

За формулами (1.51)–(1.53) розраховуємо параметри моделі для першої гармоніки:

$$a_{01} = \overline{y} = \frac{552}{12} = 46;\; a_1 \approx \frac{2 \cdot (-66{,}24)}{12} \approx -11{,}04;\; b_1 \approx \frac{2 \cdot 34{,}43}{12} \approx 5{,}74.$$

Отримавши значення параметрів можна записати ряд Фур'є для першої гармоніки:

$$\overline{y_{1t}} = 46 - 11{,}04 \cdot \cos t + 5{,}74 \cdot \sin t. \tag{1.56}$$

Підставивши значення t_i в рівняння (1.56) одержуємо теоретичні значення змінної $\overline{y_{1,t_i}}$ (див. табл. 1.14). Замість значень косинусів та синусів у рівняння підставляємо відповідні їм значення з таблиці 1.14.

Розраховуємо абсолютне значення коефіцієнта кореляції для першої гармоніки за формулою (1.23), використовуючи попередньо підраховані проміжні дані з таблиці 1.14,

$$r_{y_1/t} = \sqrt{\frac{928{,}8656}{1218}} \approx \sqrt{0{,}76} \approx 0{,}87.$$

Оскільки значення коефіцієнта кореляції близьке до одиниці, то це свідчить про щільний зв'язок між коливаннями реалізації зимового одягу та часом.

Надалі необхідно за тим же принципом, що і для першої гармоніки, оцінити параметри для другої гармоніки (табл. 1.15).

Таблиця 1.15 – Допоміжні розрахунки для другої гармоніки

$y_i \cos 2t_i$	$y_i \sin 2t_i$	$\overline{y_{2,t_i}}$	$\left(\overline{y_{2,t_i}} - \overline{y}\right)^2$
37	0,000	37,88	65,9344
20	34,641	39,64	40,4496
-22	38,105	42,87	9,7969
-52	0,000	48,82	7,9524
-23	-39,837	56,16	103,2256
35	-60,622	61,01	225,3001
60	0,000	59,96	194,8816
24	41,569	53,03	49,4209
-23	39,837	43,97	4,1209
-38	0,000	37,35	74,8225
-18	-31,177	35,18	117,0724
17,5	-30,311	36,15	97,0225
17,5	**-7,794**	**552,0**	**989,9998**

Таким чином, маємо значення параметрів для другої гармоніки:

$$a_{02} = \overline{y} = \frac{552}{12} = 46;\; a_2 \approx \frac{2 \cdot 17{,}5}{12} \approx 2{,}92;\; b_1 \approx \frac{2 \cdot (-7{,}79)}{12} \approx -1{,}3.$$

Запишемо ряд Фур'є для другої гармоніки:

$$\overline{y_{2t}} = 46 - 11{,}04 \cdot \cos t + 5{,}74 \cdot \sin t + 2{,}92 \cdot \cos 2t - 1{,}3 \cdot \sin 2t. \qquad (1.57)$$

Розраховуємо абсолютне значення коефіцієнта кореляції для першої гармоніки за формулою (1.23), використовуючи попередньо підраховані проміжні дані з таблиці 1.14,

$$r_{y_2/t} = \sqrt{\frac{989{,}9998}{1218}} \approx \sqrt{0{,}813} \approx 0{,}9.$$

Оскільки значення коефіцієнта кореляції близьке до одиниці, то це свідчить про щільний зв'язок між коливаннями реалізації зимового одягу та часом.

Порівнюючи отримані значення щільності зв'язку для двох гармонік можна зробити висновок, що ряд Фур'є для другої гармоніки є більш адекватним та краще описує сезонні коливання. На основі обраного рівняння будуємо графік залежності продажів від місяців, використовуючи прикладний пакет *Mathcad* (рис. 1.12).

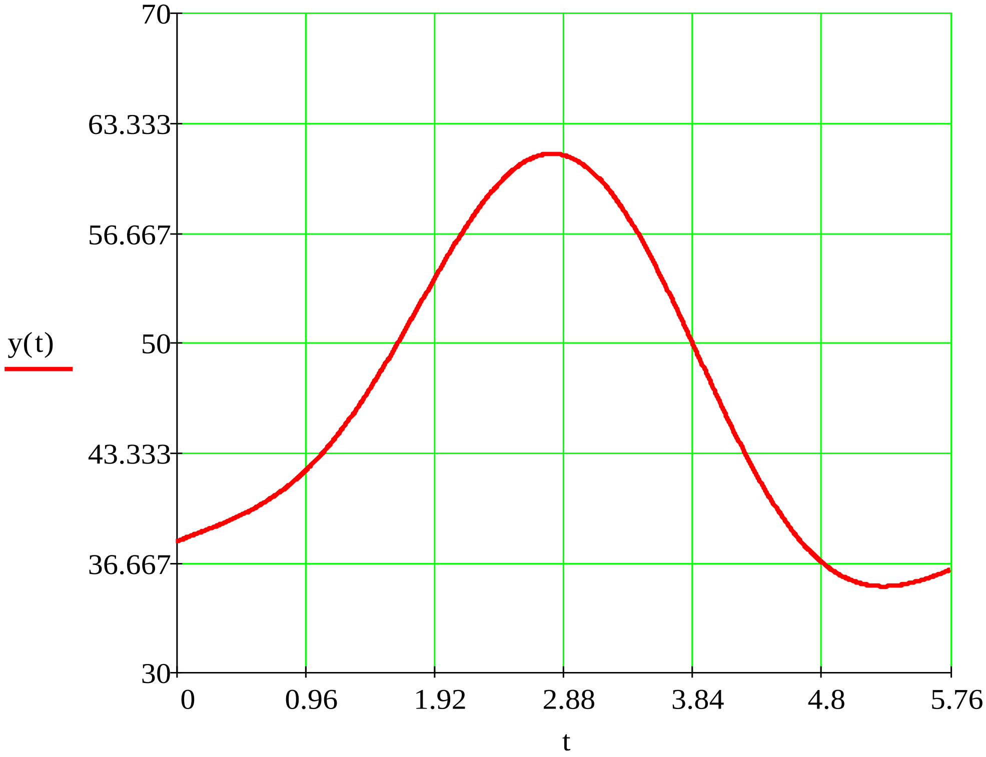


Рисунок 1.12 – Щомісячний графік продажів

Приклад 1.12. Дано статистичні дані про середньоденний товарообіг торговельного підприємства по місяцях 2010 року (табл. 1.16). Опишіть сезонні коливання товарообігу з використанням однієї гармоніки та визначіть розрахункові рівні товарообігу підприємства щомісячно.

Таблиця 1.16 – Дані прикладу 1.12

Місяць	Обсяг товарообігу, тис. грн.
Січень	65,1
Лютий	66,5
Березень	74,4
Квітень	73,6
Травень	67,2
Червень	100,0
Липень	90,0
Серпень	72,6
Вересень	68,9
Жовтень	70,4
Листопад	66,3
Грудень	77,2

Розв'язання. Для подальших розрахунків необхідно скласти допоміжну таблицю (табл. 1.17), використавши таблицю 1.13, що значно полегшить оцінювання параметрів моделі.

Таблиця 1.17 – Допоміжна таблиця для прикладу 1.12

t_i, рад.	Обсяг товарообігу, тис. грн. (y_i)	$cos\,t_i$	$sin\,t_i$	$y_i\,cos\,t_i$	$y_i\,sin\,t_i$
0	65,1	1	0	65,1	0
π/6	66,5	0,866	0,5	57,6	33,3
π/3	74,4	0,5	0,866	37,2	64,4
π/2	73,6	0	1	0	73,6
2π/3	67,2	-0,5	0,866	-33,6	58,2
5π/6	100,0	-0,866	0,5	-86,6	50,0
π	90,0	-1	0	-90,0	0
7π/6	72,6	-0,866	-0,5	-62,9	-36,3
4π/3	68,9	-0,5	-0,866	-34,5	-59,7
3π/2	70,4	0	-1	0	-70,4
5π/3	66,3	0,5	-0,866	33,2	-57,4
11π/6	77,2	0,866	-0,5	66,9	-38,6
Σ	**893,0**	*	*	**-47,6**	**17,1**

Застосовуючи першу гармоніку ряду Фур'є, визначимо параметри рівняння (1.50):

- за формулою (3.51): $a_0 = \frac{1}{12} \cdot 893 \approx 74{,}4$;

- за формулою (3.52): $a_1 = \frac{2}{12}(-47{,}6) \approx -7{,}9$;

- за формулою (3.53): $b_1 = \frac{2}{12} \cdot 17{,}1 \approx 2{,}9$.

За одержаними параметрами математична модель така:

$$\overline{y_t} = 74{,}4 - 7{,}9 \cdot cos\,t + 2{,}9 \cdot sin\,t\,. \qquad (1.58)$$

За даними моделі (1.58) визначаємо для кожного місяця розрахункові теоретичні рівні $\overline{y_{t_i}}$, одержані значення занесемо до таблиці (табл. 1.18):

$$\overline{y_{t\text{січень}}} = 74{,}4 - 7{,}9 \cdot 1 + 2{,}9 \cdot 0 = 66{,}5 \text{ тис. грн;}$$

$$\overline{y_{t\text{лютий}}} = 74{,}4 - 7{,}9 \cdot 0{,}866 + 2{,}9 \cdot 0{,}5 = 69{,}0 \text{ тис. грн.;}$$

…………………………………………..

$$\overline{y_{t\text{грудень}}} = 74{,}4 - 7{,}9 \cdot 0{,}866 + 2{,}9 \cdot (-0{,}5) = 66{,}1 \text{ тис. грн.;}$$

Таблиця 1.18 – Щомісячні розрахункові теоретичні рівні $\overline{y_{t_i}}$

Місяць	$\overline{y_{t_i}}$	Місяць	$\overline{y_{t_i}}$
Січень	66,5	Липень	82,3
Лютий	69,0	Серпень	79,8
Березень	73,0	Вересень	75,8
Квітень	77,3	Жовтень	71,5
Травень	80,9	Листопад	67,9
Червень	82,7	Грудень	66,1

Оскільки $\sum_{i=1}^{12} \overline{y_{t_i}} = 892{,}8$, то це свідчить про достатньо точний розподіл вирівняних даних. Відхилення на 0,2 пояснюється заокругленням у розрахунках.

1.9 Алгоритм розв'язання практичної задачі однофакторного кореляційного аналізу

Нехай для виробничого підприємства відомі такі показники: кількість виробленого та реалізованого продукту ***К*** *(тис. одиниць),* ціна ***Ц*** *(тис. грошових од. за одиницю продукції),* витрати ***В*** виробництва за повною собівартістю *(млн. грошових од.).* Необхідно провести аналіз на оптимальність обсягу реалізації продукції за критерієм максимального прибутку та розробити стратегію підприємства на майбутні періоди.

Відомо, що прибуток підприємства обчислюється за формулою:

$$П = К \cdot Ц - В.$$

Зрозуміло, що в даному випадку ціна та витрати виробництва за повною собівартістю, а отже і прибуток, кореляційно залежать від кількості виробленого та реалізованого продукту. Тому в наших розрахунках ми будемо одержувати функцію прибутку, аналітичний вираз якої в спрощеному вигляді можна записати так:

$$\overline{П}_к = К \cdot \overline{Ц}_к - \overline{В}_к. \tag{1.59}$$

При розв’язуванні даної практичної задачі потрібно дотримуватися певного ***алгоритму***.

1. *Знаходимо аналітичні рівняння регресії для ціни та витрат:* $\overline{Ц_к}$ *=Ц(К),* $\overline{В_к}$ *=В(К).*

Для виконання цього пункту потрібно спочатку нанести на координатні площини *КОЦ* та *КОВ* емпіричні дискретні точки та зробити припущення щодо форми відповідної лінії регресії. За допомогою нормальних систем (1.10) або (1.43)-(1.45) оцінюємо значення параметрів обраних рівнянь регресії.

2. Перевіряємо адекватність знайдених рівнянь регресії. Оцінюємо щільність зв’язку між кількістю виробленого та реалізованого продукту і ціною та кількістю і витратами.

Адекватність знайдених рівнянь регресії перевіряємо за допомогою коефіцієнта детермінації, який обчислюємо за формулою (1.26) або за (1.27). Варто нагадати, що рівняння будуть значущими, якщо коефіцієнт детермінації D належить відрізку $[0{,}55;\ 1]$. Якщо $\mathrm{D} \in [0{,}45;\ 0{,}55]$, то для встановлення адекватності моделі реальному об’єкту використовують *F*-критерій Фішера.

За умови, що $0 < D < 0{,}45$, знайдене рівняння регресії неадекватне, наше припущення щодо форми лінії регресії недостовірне і потрібно робити інше припущення.

Оцінити щільність зв’язку між відповідними показниками та зробити необхідні висновки можна за допомогою коефіцієнта кореляції ($r_{Ц/К}$ та $r_{В/К}$), який обчислюється за однією з формул (1.23)−(1.27).

Зауваження! Неадекватність одержаних рівнянь регресії може бути пов’язана з використанням нераціональних методів розв’язування систем нормальних рівнянь або неточних заокруглень. Для усунення останнього недоліку пропонується в розрахунках використовувати числа не менше як із п’ятьма знаками після коми.

3. Будуємо функцію прибутку та знаходимо кількість виробленого і реалізованого продукту, при якій прибуток буде максимальним.

Для побудови функції прибутку підставляємо знайдені рівняння регресії для ціни та витрат у рівність (1.59). Одержану функцію прибутку досліджуємо на екстремум (знаходимо точку максимуму $К_0$).

4. Робимо висновки, за даним критерієм, про стан підприємства та розробляємо стратегію на майбутні періоди.

Для одержання висновків щодо стану підприємства складається допоміжна порівняльна таблиця (табл. 1.19).

Значення $К_n$, $Ц_n$, та $В_n$ – статистичні дані динамічного ряду за останній період.

Оптимальне значення кількості виробленого та реалізованого продукту – це точка максимуму функції прибутку $К_0$. Підставивши це значення в рівняння регресії $\overline{Ц}_к = Ц(К)$, $\overline{В}_к = В(К)$, отримаємо оптимальні ціну та витрати.

Таблиця 1.19 – Допоміжна порівняльна таблиця

Показник	К	Ц	В	ЦК	П
Фактичний за останній період	$К_n$	$Ц_n$	$В_n$	$Ц_n К_n$	$П_n$
Оптимальний	$К_0$	$\overline{Ц}_к = Ц(К_0)$	$\overline{В}_к = В(К_0)$	$Ц(К_0)\, К_0$	$Ц(К_0)\, К_0 - В(К_0)$
Відхилення фактичного від оптимального	$К_n - К_0$	$Ц_n - Ц(К_0)$	$В_n - В(К_0)$	$Ц_n К_n - Ц(К_0)\, К_0$	$П_n - (Ц(К_0)\, К_0 - В(К_0))$

Аналізуємо значення відхилення фактичного значення показника від оптимального, робимо відповідні висновки та розробляємо стратегію підприємства на майбутні періоди.

Зауваження! Якщо в п.1 ми встановили, що крива прибутку має точку мінімуму, а не максимуму, то ми не зможемо розв’язати задачу за критерієм максимального прибутку. Ми можемо лише визначити мінімальний обсяг виробництва та мінімальні ціну й витрати.

Приклад 1.13. Для виробничого підприємства відомі такі показники за 6 періодів: кількість виробленого та реалізованого продукту *К* (тис. одиниць), ціна *Ц* (тис. грошових од. за одиницю продукції), витрати *В* виробництва за повною собівартістю (млн. грошових од.) (табл. 1.20).

Необхідно:

1) знайти кореляційну залежність ціни *Ц* та витрат *В* від кількості виробленого і реалізованого продукту *К*: $\overline{Ц} = Ц(К)$, $\overline{В} = В(К)$. Оцінити щільність зв’язку між відповідними ознаками за кореляційним відношенням, обчислити коефіцієнт детермінації;

2) провести аналіз на оптимальність обсягу реалізації продукції за критерієм максимального прибутку;

3) зробити висновки та розробити стратегію підприємства на майбутні періоди.

Таблиця 1.20 – Дані для прикладу 1.13

Показник \ Період	1	2	3	4	5	6
К	48	37	25	23	22	21
Ц	1,653	4,794	7,288	8,165	9,218	10,000
В	55	120	137	153	173	182

Розв'язання. Якщо побудувати емпіричні точки, то можна переконатися в тому, що залежність $\overline{Ц} = Ц(К)$ доцільно шукати у вигляді лінійної функції

$$\overline{Ц} = a_1 К + a_0 . \qquad (1.60)$$

Для пошуку параметрів функції (1.60) використаємо систему нормальних рівнянь (1.10):

$$\begin{cases} a_1 \sum К_i + n a_0 = \sum Ц_i ; \\ a_1 \sum К_i^2 + a_0 \sum К_i = \sum К_i Ц_i . \end{cases}$$

Складаємо таку допоміжну розрахункову таблицю (табл. 1.21):

Таблиця 1.21 – Допоміжна розрахункова таблиця для розрахунку $\hat{Ц}$

Період	$К_i$	$Ц_i$	$К_i^2$	$К_i Ц_i$	$\overline{Ц}_i$	$(\overline{Ц}_i - \overline{Ц})^2$	$(Ц_i - \overline{Ц})^2$
1	48	1,653	2304	79,344	1,575786	27,851099	27,040000
2	37	4,794	1369	177,378	4,685484	4,698126	4,239481
3	25	7,288	625	182,200	8,078100	1,500870	0,189225
4	23	8,165	529	187,795	8,643536	3,206019	1,721344
5	22	9,218	484	202,796	8,926254	4,298382	5,593225
6	21	10,000	441	210,000	9,208972	5,550604	9,903609
Σ	**176**	**41,118**	**5752**	**1039,513**	–	**47,105100**	**48,686884**

Тобто, в нашому випадку система нормальних рівнянь має вид:

$$\begin{cases} 176 a_1 + 6 a_0 = 41{,}118; \\ 5752 a_1 + 176 a_0 = 1039{,}513. \end{cases}$$

Розв'яжемо отриману систему методом Гаусса. Маємо:

$$\left(\begin{array}{cc|c} 176 & 6 & 41{,}118 \\ 5752 & 176 & 1039{,}513 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 3{,}492857 \cdot 10^{-2} & 2{,}33625 \cdot 10^{-1} \\ 1 & 3{,}0598052 \cdot 10^{-2} & 1{,}80722 \cdot 10^{-1} \end{array}\right) \sim$$

$$\sim \left(\begin{array}{cc|c} 0 & 3{,}492857 \cdot 10^{-3} & 5{,}2903 \cdot 10^{-2} \\ 1 & 3{,}0598052 \cdot 10^{-2} & 1{,}8072 \cdot 10^{-1} \end{array}\right).$$

Із першого рядка останньої матриці випливає, що $a_0 = \dfrac{5{,}2903 \cdot 10^{-2}}{3{,}492857 \cdot 10^{-3}} \approx 15{,}14605$. Використовуючи другий рядок останньої матриці знаходимо наступний параметр рівняння регресії:

$$a_1 = 1{,}8072 \cdot 10^{-1} - 3{,}0598052 \cdot 10^{-2} a_0 \approx -0{,}282718.$$

Тому шукане рівняння регресії має вигляд:

$$\underline{Ц}_i = -0{,}282718K + 15{,}14605. \qquad (1.61)$$

Підставляючи по черзі значення K_i у рівняння (1.61), отримаємо відповідні значення $\overline{Ц}_i$, якими заповнюємо шосту графу таблиці 1.21, а за цими значеннями та значеннями $Ц_i$ і $\overline{Ц}_i$, обчислюємо елементи сьомого та восьмого стовпців цієї таблиці.

Щільність зв’язку між ознаками Ц та К обчислюємо за допомогою коефіцієнта кореляції:

$$r_{Ц/К} = -\sqrt{\frac{\sum(\overline{Ц}_i - \overline{Ц})^2}{\sum(Ц_i - \overline{Ц})^2}} = -\sqrt{\frac{47{,}1051}{48{,}686884}} = -\sqrt{0{,}967511} \approx -0{,}984.$$

(Знак “–” для попередніх розрахунків обираємо з тієї причини, що пряма (1.61) є спадною).

Оскільки значення коефіцієнта кореляції близьке за абсолютним значенням до одиниці, то це свідчить про щільний зворотний зв’язок між ціною товару та його реалізованою кількістю. Коефіцієнт детермінації D є підкореневим виразом останньої формули. Цей коефіцієнт $D \approx 0{,}97$ показує, що варіація ознаки, яка результує (ціни товару), на 97% відбувається під впливом фактора кількості реалізованого продукту, а на 3% − під впливом неврахованих в моделі факторів.

Аналогічно, якщо побудувати емпіричні точки, то можна переконатися в тому, що залежність $\overline{B}_i = B(K)$ доцільно шукати у вигляді лінійної функції

$$\overline{B}_i = b_1 K + b_0. \qquad (1.62)$$

Для знаходження параметрів функції (1.62) використаємо систему нормальних рівнянь (1.10):

$$\begin{cases} b_1 \sum K_i + n b_0 = \sum B_i; \\ b_1 \sum K_i^2 + b_0 \sum K_i = \sum K_i B_i. \end{cases}$$

Складаємо таку допоміжну розрахункову таблицю (табл. 1.22):

Таблиця 1.22 – Допоміжна розрахункова таблиця для розрахунку $\overline{B}_i$

Період	K_i	B_i	K_i^2	$K_i\,B_i$	$\overline{B}_i$	$(\overline{B}_i - \overline{B})^2$	$(B_i - \overline{B})^2$
1	48	55	2304	2640	60,60632	5784,16224	6668,35560
2	37	120	1369	4440	105,42758	975,46405	277,55560
3	25	137	625	3425	154,3235	311,99923	0,1156
4	23	153	529	3519	162,47282	666,30167	266,99560
5	22	173	484	3806	166,54748	893,26146	1320,5056
6	21	182	441	3822	170,62214	1153,42695	2055,7156
Σ	**176**	**820**	**5752**	**21652**		**9784,6156**	**10589,3336**

Тобто, в нашому випадку система нормальних рівнянь має вигляд:

$$\begin{cases} 176b_1 + 6b_0 = 820; \\ 5752b_1 + 176a_0 = 21652. \end{cases}$$

Розв’яжемо отриману систему методом Гаусса. Маємо:

$$\left(\begin{array}{cc|c} 176 & 6 & 820 \\ 5752 & 176 & 21652 \end{array}\right) \sim \left(\begin{array}{cc|c} 176 & 6 & 820 \\ 0 & -3536 & -905888 \end{array}\right).$$

Із другого рядка останньої матриці випливає, що $b_0 = \dfrac{905888}{3536} \approx 256{,}1900$. Використовуючи перший рядок останньої матриці знаходимо наступний параметр рівняння регресії:

$$b_1 = \frac{1}{176}(820 - 6b_0) \approx -4{,}07466.$$

Тому шукане рівняння регресії має вигляд:

$$\overline{B}_i = -4{,}07466K + 256{,}1900. \tag{1.63}$$

Підставляючи по черзі значення K_i в рівняння (1.63), отримаємо відповідні значення $\overline{B}_i$, якими заповнюємо шосту графу таблиці 1.22, а за цими значеннями та значеннями B_i і $\overline{B}$, обчислюємо елементи сьомої та восьмої граф цієї таблиці.

Оскільки пряма (1.63) є спадною (коефіцієнт біля K від’ємний), то щільність зв’язку між ознаками B та K обчислюємо за допомогою коефіцієнта кореляції так:

$$r_{B/K} = -\sqrt{\frac{\sum(\overline{B}_i - \overline{B})^2}{\sum(B_i - \overline{B})^2}} = -\sqrt{\frac{9784{,}6156}{10589{,}3336}} = -\sqrt{0{,}92400621} \approx -0{,}961.$$

Оскільки значення коефіцієнта кореляції близьке за абсолютним значенням до одиниці, то це свідчить про щільний зворотний зв'язок між витратами на виробництво та кількістю реалізованої продукції. Коефіцієнт детермінації D є підкореневим виразом останньої формули. Цей коефіцієнт $D \approx 0{,}92$ показує, що варіація ознаки, яка результує (витрат), на 92% відбувається під впливом фактора кількості реалізованого продукту, а на 8% – під впливом неврахованих в моделі факторів.

Використовуючи рівняння (1.61) та (1.63), знаходимо формулу прибутку від реалізації продукту:

$$\overline{П} = \overline{Ц} \cdot K - \overline{B} = (-0{,}282718K + 15{,}14605)K + 4{,}07466K - 256{,}1900 =$$
$$= -0{,}282718K^2 + 19{,}22071K - 256{,}1900.$$

Прирівнюючи до нуля першу похідну отриманої функції, знаходимо оптимальне значення кількості реалізованого продукту за критерієм максимального прибутку:

$$(\overline{П})' = -0{,}565436 \cdot K + 19{,}22071 = 0.$$

Звідки $K \approx 34$ (оскільки $(\overline{П})'(30) > 0$, $(\overline{П})'(40) < 0$, то $K \approx 34$ є точкою максимуму).

Таким чином, оптимальний обсяг випуску та реалізації продукту складає 34 одиниці. За цією величиною можна обчислити оптимальну ціну продукту, за рівнянням (1.61), оптимальні витрати, за рівнянням (1.63), виручку від реалізації та оптимальний прибуток. Результати розрахунків наведено в таблиці 1.23. Для порівняння в даній таблиці наведено також фактичні дані за останній період.

Таблиця 1.23 – Порівняльна таблиця

Показник	K	$Ц$	B	$ЦK$	$П$
Фактичний за останній період	21	10,000	182	210	28
Оптимальний	34	5,534	117,652	188,156	70,504
Відхилення фактичного від оптимального	-13	4,466	64,348	21,844	-42,504

Висновки

1. Прибуток підприємства на 42,504 млн. грош. од. менший оптимального за рахунок відповідного перевищення витрат на 64,348 млн. грош. од.

2. Виручка від реалізації продукту перевищує оптимальну на 21,844 млн. грош. од. за рахунок збільшення ціни на 4,466 тис. грош. од.

Стратегія підприємства: зростання обсягу випуску та реалізації продукту, зменшення ціни та витрат.

Приклад 1.14. Для виробничого підприємства відомі такі показники за 6 періодів: кількість виробленого та реалізованого продукту *К* (тис. одиниць), ціна *Ц* (тис. грошових од. за одиницю продукції), витрати *В* виробництва за повною собівартістю (млн. грошових од.) (табл. 1.24).

Необхідно:

1) знайти кореляційну залежність ціни *Ц* та витрат *В* від кількості виробленого та реалізованого продукту К: $\overline{Ц} = Ц(К),\ \overline{В} = В(К)$.

Оцінити щільність зв'язку між відповідними ознаками за кореляційним відношенням, обчислити коефіцієнт детермінації;

2) провести аналіз на оптимальність обсягу реалізації продукції за критерієм максимального прибутку;

3) зробити висновки та розробити стратегію підприємства на майбутні періоди.

Таблиця 1.24 – Дані для прикладу 1.14

Період	1	2	3	4	5	6
К	68	31	26	25	14	12
Ц	6	17	19	19	24	25
В	401	420	520	484	280	255

Розв'язання. Якщо побудувати емпіричні точки, то можна переконатися в тому, що залежність $\overline{Ц_к} = Ц(К)$ доцільно шукати у вигляді лінійної функції

$$\overline{Ц_К} = a_1 К + a_0 . \qquad (1.64)$$

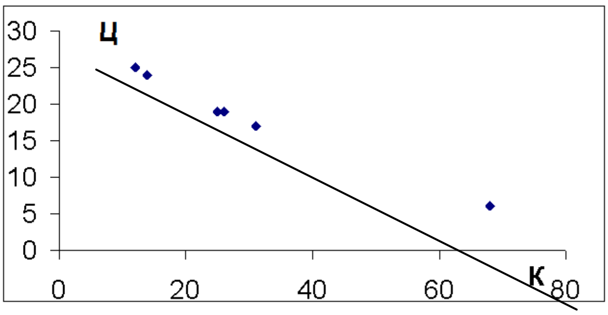


Рисунок 1.13 Апроксимація даних *Ц*(*К*) лінійною залежністю

Для знаходження параметрів функції (1.64) використаємо систему нормальних рівнянь (1.10):

$$\begin{cases} a_1 \sum K_i + na_0 = \sum Ц_i \\ a_1 \sum K_i^2 + a_0 \sum K_i = \sum K_i Ц_i \end{cases}.$$

Складаємо таку допоміжну розрахункову таблицю (табл. 1.25):

Таблиця 1.25

Період	$К_i$	$Ц_i$	$К_i^2$	$К_i Ц_i$	$\overline{Ц_К}$	$(\overline{Ц_К} - \overline{Ц})^2$	$(Ц_i - \overline{Ц})^2$
1	68	6	4626	408	5,465	165,587	152,111
2	31	17	961	527	17,779	0,308	1,778
3	26	19	676	494	19,443	1,231	0,444
4	25	19	625	475	19,775	2,08	0,444
5	14	24	196	336	23,436	26,039	32,111
6	12	25	144	300	24,102	33,275	44,444
$\sum$	**176**	**110**	**7226**	**2540**	**110**	**228,519**	**231,333**

Тобто, в нашому випадку система нормальних рівнянь має вигляд:

$$\begin{cases} 176a_1 + 6a_0 = 110; \\ 7226a_1 + 176a_0 = 2540. \end{cases}$$

Розв’язавши отриману систему методом Гаусса, знаходимо:

$$a_0 \approx 28{,}095315,\ a_1 \approx -0{,}333.$$

Тому шукане рівняння регресії має вигляд:

$$\overline{Ц_К} = -0{,}333К + 28{,}095315. \qquad (1.65)$$

Підставляючи по черзі значення K_i в рівняння (1.65), отримаємо відповідні значення $\overline{Ц_К}$, якими заповнюємо шосту графу таблиці 1.25, а за цими значеннями та значеннями $Ц_i$ і $\overline{Ц}$ обчислюємо елементи сьомої та восьмої граф цієї таблиці.

Щільність зв’язку між ознаками *Ц* та *К* обчислюємо за допомогою коефіцієнта кореляції:

$$r_{Ц/К} = \sqrt{\frac{\sum(\overline{Ц_К} - \overline{Ц})^2}{\sum(Ц_i - \overline{Ц})^2}} = \sqrt{\frac{228{,}519}{231{,}333}} = -\sqrt{0{,}987835} \approx -0{,}994.$$

(Знак “–” для попередніх розрахунків обираємо з тієї причини, що пряма (1.65) є спадною).

Оскільки значення коефіцієнта кореляції близьке за абсолютним значенням до одиниці, то це свідчить про щільний зворотний зв’язок між ціною товару та його реалізованою кількістю. Коефіцієнт детермінації D є підкореневим виразом останньої формули. Цей коефіцієнт $D \approx 0{,}98$ показує, що варіація ознаки, яка результує (ціни товару), на 98% відбувається під впливом фактора кількості реалізованого продукту, а на 2% - під впливом неврахованих в моделі факторів.

Аналогічно, якщо побудувати емпіричні точки, то

Визначимо тип залежності $B(K)$ за допомогою графіка, що апроксимує дані кореляційної таблиці до функції, зображеної на рис. 1.13:

З рисунку 1.13 можна припустити, що залежність між витратами та кількістю виробленої та реалізованої продукції має квадратичний характер:

$$\overline{B_{к}} = b_0 + b_1 \cdot K + b_2 \cdot K^2. \tag{1.66}$$

Для пошуку параметрів параболічної залежності скористаємося системою нормальних рівнянь (1.42):

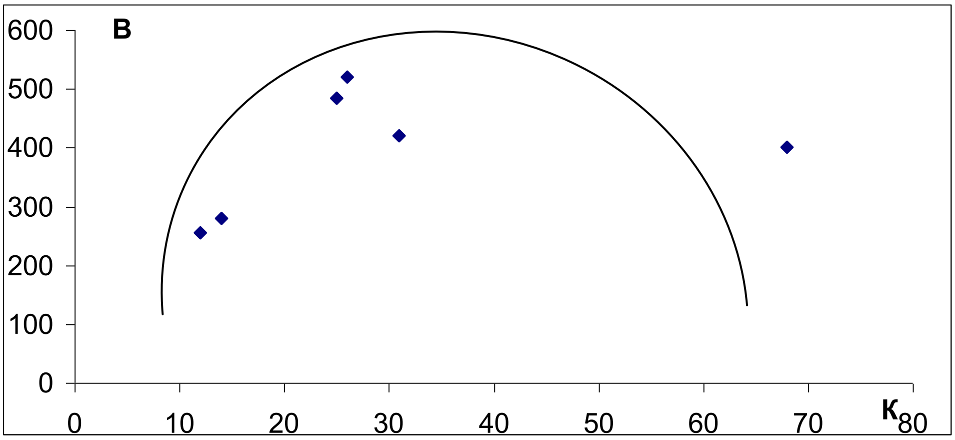


Рисунок 1.14 − Апроксимація даних $B(K)$ параболічною залежністю

$$\begin{cases} nb_0 + b_1 \sum K_i + b_2 \sum K_i^2 = \sum B_i; \\ b_0 \sum K_i + b_1 \sum K_i^2 + b_2 \sum K_i^3 = \sum K_i B_i; \\ b_0 \sum K_i^2 + b_1 \sum K_i^3 + b_2 \sum K_i^4 = \sum {K_i}^2 B_i. \end{cases}$$

Для розрахунку значень параметрів параболічної однофакторної моделі необхідно скласти відповідну допоміжну таблицю 1.26.

Таблиця 1.26

№	K_i	B_i	K_i^2	K_i^3	K_i^4	K_iB_i	$K_i^2B_i$
1	68	401	4626	314568	21390624	27268	1855026
2	31	420	961	29791	923521	13020	403620
3	26	520	676	17576	456976	13520	351520
4	25	484	625	15625	390625	12100	302500
5	14	280	196	2744	38416	3920	54880
6	12	255	144	1728	20736	3060	36720
$\sum$	**176**	**2360**	**7226**	**382032**	**23220898**	**72888**	**3004266**

Таким чином, система нормальних рівнянь набуває вигляду:

$$\begin{cases} 6b_0 + 176b_1 + 7226b_2 = 2360; \\ 176b_0 + 7226b_1 + 382032b_2 = 72888; \\ 7226b_0 + 382032b_1 + 23220898b_2 = 3004266. \end{cases}$$

Розв’язавши систему засобами *Mathcad* знаходимо:

$$b_0 = 2{,}471; \qquad b_1 = 24{,}788; \qquad b_2 = -0{,}279.$$

Тому шукане рівняння регресії таке:

$$\overline{B_к} = 2{,}471 + 24{,}788 \cdot K - 0{,}279 \cdot K^2. \qquad (1.67)$$

Підставляючи по черзі значення K_i в рівняння (1.67) знайдемо умовні середні значення витрат ($\overline{B_k}$) та відповідні суми квадратів відхилень (табл. 1.27), врахувавши, що $\overline{B} = \frac{2360}{6} \approx 393{,}333$.

Таблиця 1.27

№	$\overline{B_k}$	$(\overline{B_K} - \overline{B})^2$	$(B_i - \overline{B})^2$
1	397,959	21,39988	58,78289
2	502,78	11978,65000	711,12890
3	458,355	4227,86000	16044,53000
4	447,796	2966,21800	7220,50500
5	294,819	9705,00800	12844,37000
6	259,751	17844,15000	19136,02000
$\sum$	**2361,46**	**46843,28**	**57015,33**

Щільність зв'язку між ознаками В та К обчислюємо за допомогою коефіцієнта кореляції так:

$$r_{B/K} = \sqrt{\frac{\sum(\overline{B_K} - \overline{B})^2}{\sum(B_i - \overline{B})^2}} = \sqrt{\frac{46843{,}28}{57015{,}33}} = \sqrt{0{,}819837} \approx 0{,}91.$$

Оскільки значення коефіцієнта кореляції близьке до одиниці, то це свідчить про щільний прямий зв'язок між витратами на виробництво та кількістю реалізованої продукції. Коефіцієнт $D \approx 0{,}91$ показує, що варіація ознаки, яка результує (витрат), на 91% відбувається під впливом фактора кількості реалізованого продукту, а на 9% − під впливом неврахованих в моделі факторів. Отже, побудоване рівняння регресії є адекватиним реальному об'єкту.

Використовуючи рівняння (1.65) та (1.67), знаходимо формулу прибутку від реалізації продукту:

$$\overline{П_K} = \overline{Ц_K}K - \overline{B_K} = (-0{,}333K + 28{,}095315)K - 2{,}471 - 24{,}788K + 0{,}279K^2 =$$
$$= -0{,}054K^2 + 3{,}307315K - 2{,}471.$$

Прирівнюючи до нуля першу похідну отриманої функції, знаходимо оптимальне значення кількості реалізованого продукту за критерієм максимального прибутку: $(\overline{П_K})' = -0{,}108K + 3{,}307315 = 0$. Звідки $K \approx 31$ – точка максимуму, оскільки крива функції прибутку є параболою з від'ємним коефіцієнтом біля K^2.

Таким чином, оптимальний обсяг випуску та реалізації продукту складає 31 одиницю. За цією величиною можна обчислити оптимальну ціну продукту, за рівнянням (1.65), оптимальні витрати, за рівнянням (1.67), виручку від реалізації та оптимальний прибуток. Результати розрахунків подано в таблиці 1.28.

Таблиця 1.28

Показник	*К*	*Ц*	*В*	*ЦК*	*П*
Фактичний за останній період	12	25	255	300	45
Оптимальний	31	17,77	502	550,87	48,87
Відхилення фактичного від оптимального	-19	7,23	-247	-250,87	-3,87

Висновки

1) Лінійна залежність $\widehat{Ц} = Ц(K)$ та квадратична $\widehat{B} = B(K)$ є гарними наближеннями для вихідних даних, що засвідчується отриманими значеннями коефіцієнтів кореляції. Такий результат дозволяє

прогнозувати значення кількості виробленого та реалізованого продукту при наявності значень щодо ціни одиниці продукції або витрат на цей обсяг продукції за повною собівартістю. І навпаки, маючи заплановану оптимальну кількість виготовленого продукту можна визначити, виходячи з оцінених рівнянь, оптимальну ціну, яку необхідно встановити, та витрати, що може собі дозволити виробник.

2) Прибуток підприємства на 3,87 млн. грош. од. менший оптимального за рахунок недовиробництва.

3) Кількість виробленої та реалізованої продукції менша за оптимальну на 19 тис. одиниць за рахунок перевищення ціни на 7,23 тис. горш. од.

Стратегія підприємства: зростання обсягу випуску та реалізації продукту, зменшення ціни.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Який зв'язок між економічними показниками називають балансовим?
2. Що називають умовною середньою величиною?
3. Яку залежність називають кореляційною? Наведіть приклади.
4. Що таке рівняння та лінія регресії?
5. Запишіть загальний вигляд простої лінійної вибіркової регресії.
6. За яким критерієм оцінюють параметри лінійного рівняння регресії?
7. Охарактеризуйте метод найменших квадратів для оцінювання параметрів лінійної регресії та виведіть відповідну систему нормальних рівнянь.
8. Запишіть формулу для обчислення коефіцієнта коваріації.
9. Сформулюйте основні припущення кореляційного аналізу.
10. Доведіть, що за умови виконання основних припущень кореляційного аналізу залежна змінна y має нормальний розподіл.
11. Доведіть, що розраховані за методом найменших квадратів та з використанням основних припущень кореляційного аналізу параметри лінійної регресійної моделі мають такі властивості:

$$D(b_0)=\frac{\sigma^2\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2};\ D(b_1)=\frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}.$$

12. Охарактеризуйте елементи дисперсійного аналізу.
13. Сформулюйте поняття ступеня вільності.
14. Опишіть F-критерій Фішера-Снедекора.

15. Що характеризує коефіцієнт кореляції $r_{y/x}$ і як його можна обчислити?
16. Охарактеризуйте геометричний зміст коефіцієнта кореляції та його властивості.
17. За допомогою якого показника встановлюють адекватність рівняння регресії реальному об’єкту і як його обчислюють?
18. Опишіть стандартну помилку оцінювання.
19. Опишіть метод Фішера оцінювання коефіцієнта кореляції.
20. Які дослідження приводять до нелінійних моделей? Наведіть приклади таких моделей. Спробуйте вивести одну з систем нормальних рівнянь для оцінювання параметрів нелінійної кореляційної залежності.
21. Що називають траекторією та трендом? Що таке динамічний ряд? Які компоненти можна виділити у складі данамічного ряду?
22. Яку функцію називають моделлю тренду?
23. Охарактеризуйте динаміку абсолютних приростів.
24. Вкажіть типи економічного розвитку.
25. Якими моделями тренду описують прискорений розвиток? Опишіть ці моделі.
26. Якою трендовою моделлю описують динаміку з абсолютним прискоренням $\tilde{\varphi}$?
27. Дослідіть темпи приросту динаміки, яку описують параболічною моделлю.
28. Охарактеризуйте криві Енгеля і Філліпса та вкажіть сфери їх застосування.
29. Що називають сезонними хвилями? Для яких сфер життєдіяльності людини вони характерні?
30. Охарактеризуйте вирівнювання сезонних рядів динаміки рядом Фур’є.
31. Опишіть алгоритм розв’язання практичної задачі кореляційного аналізу.

ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

Завдання 1.1 Побудувати економетричну модель залежності y від фактора x. Початкові дані наведено у таблиці (відповідно до варіанта). Оцінити щільність зв’язку між змінними та адекватність побудованої моделі.

1.

i	y_i	x_i
1	6,25	1,25
2	7,5	1,5
3	8,75	2,25
4	11,25	3
5	16,25	4,5

2.

i	y_i	x_i
1	97,5	54
2	67,5	45
3	52,5	30
4	45	22,5
5	67,5	15

3.

i	y_i	x_i
1	8,75	1,75
2	10,5	2,1
3	12,25	3,15
4	15,75	4,2
5	22,75	6,3

4.

i	y_i	x_i
1	110,5	61,2
2	76,5	51
3	59,5	34
4	51	25,5
5	76,5	17

5.

i	y_i	x_i
1	11,25	2,25
2	13,5	2,7
3	15,75	4,05
4	20,25	5,4
5	29,25	8,1

6.

i	y_i	x_i
1	78	43,2
2	54	36
3	42	24
4	36	18
5	54	12

7.

i	y_i	x_i
1	13,75	2,75
2	16,5	3,3
3	19,25	4,95
4	24,75	6,6
5	35,75	9,9

8.

i	y_i	x_i
1	117	64,8
2	81	54
3	63	36
4	54	27
5	81	18

9.

i	y_i	x_i
1	16,25	3,25
2	19,5	3,9
3	22,75	5,85
4	29,25	7,8
5	42,25	11,7

10.

i	y_i	x_i
1	136,5	75,6
2	94,5	63
3	73,5	42
4	63	31,5
5	94,5	21

11.

i	y_i	x_i
1	18,75	3,75
2	22,5	4,5
3	26,25	6,75
4	33,75	9
5	48,75	13,5

12.

i	y_i	x_i
1	162,5	90
2	112,5	75
3	87,5	50
4	75	37,5
5	112,5	25

13.

i	y_i	x_i
1	21,25	4,25
2	25,5	5,1
3	29,75	7,65
4	38,25	10,2
5	55,25	15,3

14.

i	y_i	x_i
1	169	93,6
2	117	78
3	91	52
4	78	39
5	117	26

15.

i	y_i	x_i
1	23,75	4,75
2	28,5	5,7
3	33,25	8,55
4	42,75	11,4
5	61,75	17,1

16.

i	y_i	x_i
1	175,5	97,2
2	121,5	81
3	94,5	54
4	81	40,5
5	121,5	27

17.

i	y_i	x_i
1	50	10
2	60	12
3	70	18
4	90	24
5	130	36

18.

i	y_i	x_i
1	182	100,8
2	126	84
3	98	56
4	84	42
5	126	28

19.

i	y_i	x_i
1	52,5	10,5
2	63	12,6
3	73,5	18,9
4	94,5	25,2
5	136,5	37,8

20.

i	y_i	x_i
1	188,5	104,4
2	130,5	87
3	101,5	58
4	87	43,5
5	130,5	29

21.

i	y_i	x_i
1	57,5	11,5
2	69	13,8
3	80,5	20,7
4	103,5	27,6
5	149,5	41,4

22.

i	y_i	x_i
1	195	108
2	135	90
3	105	60
4	90	45
5	135	30

23.

i	y_i	x_i
1	67,5	13,5
2	81	16,2
3	94,5	24,3
4	121,5	32,4
5	175,5	48,6
Σ	540	135

24.

i	y_i	x_i
1	201,5	111,6
2	139,5	93
3	108,5	62
4	93	46,5
5	139,5	31

25.

i	y_i	x_i
1	145	57,2
2	91	41
3	64	14
4	51	20,5
5	91	27

Завдання 1.2. Для виробничого підприємства відомі такі показники за 6 періодів: кількість виробленого та реалізованого продукту ***К*** (тис. одиниць), ціна ***Ц*** (тис. грошових од. за одиницю продукції), витрати ***В*** виробництва за повною собівартістю (млн. грошових од.)

Необхідно:

1) знайти кореляційну залежність ціни *Ц* та витрат *В* від кількості виробленого та реалізованого продукту *К*: $\overline{Ц_к}$ =Ц(К), $\overline{В_к}$ =В(К).

Оцінити щільність зв'язку між відповідними ознаками за кореляційним відношенням, обчислити коефіцієнт детермінації;

2) провести аналіз на оптимальність обсягу реалізації продукції за критерієм максимального прибутку;

3) зробити висновки та розробити стратегію підприємства на майбутні періоди.

1.

Період / Показник	1	2	3	4	5	6
К	55	44	39	34	33	27
Ц	75	266	266	280	280	300
В	3240	7011	8989	8989	8500	7776

2.

Період / Показник	1	2	3	4	5	6
К	8	6	5,6	4,3	4	3
Ц	47	135	140	140	190	200
В	340	709	730	709	690	570

3.

Період / Показник	1	2	3	4	5	6
К	25	19	5	11	8	6
Ц	26	62	62,5	62,5	80	82
В	404	640	778	729	458	377

4.

Період / Показник	1	2	3	4	5	6
К	68	31	26	25	14	12
Ц	6	17	19	19	24	25
В	401	420	520	484	280	255

5.

Період / Показник	1	2	3	4	5	6
К	19	9,3	7,6	7,4	6	5
Ц	23	78	85	85	107	108
В	400	453	485	620	620	529

6.

Період / Показник	1	2	3	4	5	6
К	1560	778	555	545	348	300
Ц	5	14	15	15	22	23
В	5538	7859	7950	8100	7413	6555

7.

Період / Показник	1	2	3	4	5	6
К	45	38,5	37,5	37	33	20
Ц	8	20	21	21	29	36
В	248	736	785	790	930	713

8.

Період / Показник	1	2	3	4	5	6
К	130	117	92	60	31	28
Ц	16	53	61	110	165	170
В	1301	4350	5432	5199	2802	2666

9.

Період / Показник	1	2	3	4	5	6
К	2700	2413	2200	2150	1066	900
Ц	0,57	1,66	1,75	1,75	2,14	2,22
В	1486	4005	3850	3763	2047	1793

10.

Період / Показник	1	2	3	4	5	6
К	48	37	25	23	22	231
Ц	1,653	4,794	7,288	8,165	9,218	10
В	55	120	137	153	173	182

11.

Період / Показник	1	2	3	4	5	6
К	589	528	367	342	327	134
Ц	2,484	7,203	10,95	12,286	13,85	15,025
В	1027	3214	3428	4380	4017	1877

12.

Період / Показник	1	2	3	4	5	6
К	90	69	48	45	43	18
Ц	3,968	11,506	17,491	19,597	22,123	24,000
В	228	715	762	980	898	403

13.

Період / Показник	1	2	3	4	5	6
К	97	73	50	47	45	19
Ц	3,306	9,588	14,476	16,33	18,436	20
В	196	599	630	812	746	353

14.

Період / Показник	1	2	3	4	5	6
К	229	175	122	114	109	44
Ц	1,356	3,931	5,976	6,695	7,559	8,2
В	171	535	572	733	672	308

15.

Період / Показник	1	2	3	4	5	6
К	275	211	147	137	125	51
Ц	1,306	3,787	5,757	6,45	7,282	7,9
В	198	623	666	852	781	345

16.

Період / Показник	1	2	3	4	5	6
К	25	19	5	11	8	6
Ц	26	62	62,5	62,5	80	82
В	404	640	778	729	458	377

17.

Період / Показник	1	2	3	4	5	6
К	19	9,3	7,6	7,4	6	5
Ц	23	78	85	85	107	108
В	400	453	485	620	620	529

18.

Період / Показник	1	2	3	4	5	6
К	45	38,5	37,5	37	33	20
Ц	8	20	21	21	29	36
В	248	736	785	790	930	713

19.

Період / Показник	1	2	3	4	5	6
К	2700	2413	2200	2150	1066	900
Ц	0,57	1,66	1,75	1,75	2,14	2,22
В	1486	4005	3850	3763	2047	1793

20.

Період / Показник	1	2	3	4	5	6
К	589	528	367	342	327	134
Ц	2,484	7,203	10,95	12,286	13,85	15,025
В	1027	3214	3428	4380	4017	1877

21.

Період / Показник	1	2	3	4	5	6
К	97	73	50	47	45	19
Ц	3,306	9,588	14,476	16,33	18,436	20
В	196	599	630	812	746	353

22.

Період / Показник	1	2	3	4	5	6
К	275	211	147	137	125	51
Ц	1,306	3,787	5,757	6,45	7,282	7,9
В	198	623	666	852	781	345

23.

Період / Показник	1	2	3	4	5	6
К	19	9,3	7,6	7,4	6	5
Ц	23	78	85	85	107	108
В	400	453	485	620	620	529

24.

Період / Показник	1	2	3	4	5	6
К	90	69	48	45	43	18
Ц	3,968	11,506	17,491	19,597	22,123	24,000
В	228	715	762	980	898	403

25.

Період / Показник	1	2	3	4	5	6
К	48	37	25	23	22	231
Ц	1,653	4,794	7,288	8,165	9,218	10
В	55	120	137	153	173	182

ТЕМА 2 МНОЖИННИЙ РЕГРЕСІЙНИЙ АНАЛІЗ

Економічні явища, як правило, визначаються великою кількістю факторів. Кожна економічна змінна залежить не від одного, а від багатьох факторів. Наприклад, валовий регіональний продукт залежить не лише від величини основних засобів, а й від величини оборотних фондів, величини інвестицій в основний капітал, кількості людей, зайнятих на підприємствах регіону, технологій, які використовують на підприємствах, ефективності управлінських рішень тощо. Урожайність деяких культур залежить від родючості ґрунту, кількості внесених органічних і мінеральних добрив, сорту насіння, агротехнічного оброблення, природних умов тощо. Спільний вплив кількох факторів на одну змінну, яка результує, досліджують за допомогою багатофакторних економетричних моделей.

2.1 Класична нормальна лінійна модель множинної регресії

Припустимо, що потрібно дослідити залежність змінної Y від змінних X_1, $X_2, \ldots, X_p$. Позначимо i-те спостереження залежної змінної y_i, пояснювальних змінних $-$ x_{i1}, $x_{i2}, \ldots, x_{ip}$. Тоді модель множинної регресії така:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i \,, \qquad (2.1)$$

де ε_i – випадкова величина;

β_i – параметри моделі, $i=\overline{1,n}$, які відповідають припущенням (1.15)–(1.19).

Модель (2.1) називають класичною нормальною лінійною моделлю множинної регресії. Залежна змінна Y називається також *пояснюваною, ендогенною змінною,* незалежні змінні X_i – *пояснювальними, екзогенними змінними*.

Введення більшої кількості пояснювальних змінних ускладнює математичне оброблення даної моделі, тому виникла доцільність використання матричних позначень:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \; X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix}, \; \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \qquad (2.2)$$

Звертаємо увагу на те, що до матриці X введено додатковий стовпець, усі елементи якого дорівнюють 1, тобто умовно припускаємо, що в моделі (2.1) вільний член β_0 множиться на фіктивну змінну x_{i0}, яка дорівнює 1 для всіх значень i.

Враховуючи позначення (2.2) модель (2.1) набуває вигляду:

$$Y = X\beta + \varepsilon . \tag{2.3}$$

Оцінкою моделі (2.1) за вибіркою є рівняння виду:

$$Y = Xb + e , \tag{2.4}$$

де $b = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{pmatrix}$, $e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$.

2.2 Оцінювання параметрів класичної регресійної моделі методом найменших квадратів

За методом найменших квадратів параметр b обирається таким, щоб сума квадратів відхилень була мінімальною, тобто:

$$e^T e = \sum_{i=1}^{n} e_i^2 \to \min ,$$

або

$$Q_e = \sum_{i=1}^{n} (\overline{y_x} - y_i)^2 = \sum_{i=1}^{n} e_i^2 = (Y - Xb)^T (Y - Xb) \to min . \tag{2.5}$$

Врахувавши властивість транспонування добутку матриць $\left((AB)^T = B^T A^T\right)$, після розкриття дужок одержимо:

$$Q_e = Y^T Y - b^T X^T Y - Y^T Xb + b^T X^T Xb .$$

Добуток $Y^T Xb$ – матриця розмірності $(1\times n)[n\times(p+1)] \times [(p+1)\times 1]=(1\times 1)$, тобто скалярна величина. Це означає, що даний добуток не змінюється при транспонуванні, тобто $Y^T Xb$ =($Y^T Xb)^T$=$b^T X^T Y$. Тому умова мінімізації набуває вигляду:

$$Q_e = Y^T Y - 2b^T X^T Y + b^T X^T Xb \to min. \qquad (2.6)$$

Необхідною умовою екстремуму функції $Q(b_0, b_1, \ldots, b_p)$ є

$$\frac{\partial Q}{\partial b} = \begin{pmatrix} \frac{\partial Q}{\partial b_0} & \frac{\partial Q}{\partial b_1} & \cdots & \frac{\partial Q}{\partial b_p} \end{pmatrix}.$$

Для вектора частинних похідних можна довести такі твердження:

$$\frac{\partial}{\partial b}\left(b^T c\right) = c, \qquad \frac{\partial}{\partial b}\left(b^T Ab\right) = 2Ab,$$

де b та c – вектор-стопці, A – симетрична матриця, в якій елементи, розташовані симетрично відносно головної діагоналі, рівні.

Справедливість наведених формул проілюструємо на прикладі. Нехай $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$, $c = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$, $A = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$. Оскільки $b^T c = \begin{pmatrix} b_1 & b_2 \end{pmatrix}\begin{pmatrix} 3 \\ 4 \end{pmatrix} = 3b_1 + 4b_2$ і $b^T Ab = \begin{pmatrix} b_1 & b_2 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = 2b_1^2 + 6b_1 b_2 + 5b_2^2$, то $\frac{\partial}{\partial b}\left(b^T c\right) = \frac{\partial}{\partial b}(3b_1 + 4b_2) = \begin{pmatrix} 3 \\ 4 \end{pmatrix} = c$ та $\frac{\partial}{\partial b}\left(b^T Ab\right) = \frac{\partial}{\partial b}\left(2b_1^2 + 6b_1 b_2 + 5b_2^2\right) = \begin{pmatrix} 4b_1 + 6b_2 \\ 6b_1 + 10b_2 \end{pmatrix} = 2\begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = 2Ab$.

Тому, позначивши c=$X^T Y$, а матрицю A=$X^T X$ (яка є симетричною), знайдемо

$$\frac{\partial Q}{\partial b} = -2X^T Y^T + 2X^T Xb = 0,$$

звідки одержуємо систему нормальних рівнянь в матричній формі для визначення вектора b:

$$X^T Xb = X^T Y. \qquad (2.7)$$

Згідно з цим методом параметр b знаходиться за формулою:

$$b = (X^T X)^{-1} \cdot X^T Y. \qquad (2.8)$$

Тоді лінійна модель має вигляд: $\overline{Y} = X_0^T b$ (2.9), де $X_0^T = \begin{pmatrix} 1 & x_{10} & x_{20} & \ldots & x_{p0} \end{pmatrix}$.

Зауваження! 1) Для зручності практичної роботи із системою (2.7) знайдемо матриці, які до неї входять. Під знаком $\sum$ будемо розуміти знак $\sum\limits_{i=1}^{n}$.

$$X^T X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} =$$

$$= \begin{pmatrix} n & \sum x_{i1} & \dots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \dots & \sum x_{i1}x_{ip} \\ \dots & \dots & \dots & \dots \\ \sum x_{ip} & \sum x_{i1}x_{ip} & \dots & \sum x_{ip}^2 \end{pmatrix}. \qquad (2.9)$$

$$X^T Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \dots \\ \sum y_i x_{ip} \end{pmatrix}. \qquad (2.10)$$

В окремому випадку, використовуючи матричне рівняння (2.7) з врахуванням (2.9) та (2.10) для однієї пояснювальної змінної, нескладно одержати відому систему нормальних рівнянь (1.10):

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_i \end{pmatrix} \leftrightarrow \begin{cases} nb_0 + b_1 \sum x_i = \sum y_i \\ b_0 \sum x_i + b_1 \sum x_i^2 = \sum y_i x_i \end{cases}.$$

2) Система (2.7) має розв'язок, якщо матриця $X^T X$ є невласною, тобто її визначник не дорівнює нулю.

Із другого зауваження можна одержати ще одну передумову множинного регресійного аналізу: *вектори значень пояснювальних змінних, або стовпці матриці X повинні бути лінійно незалежними.*

Теорема Гаусса-Маркова, розглянута для парної регресійної моделі, є справедливою і для моделі множинної регресії: за умови виконання передумов множинного регресійного аналізу оцінювання методом найменших квадратів $b=(X^T X)^{-1} X^T Y$ є найбільш ефективним, тобто має дисперсію в класі лінійних несуміщених оцінок.

Приклад 2.1. Є такі дані (умовні) про видобуток вугілля за зміну на одного шахтаря Y (т), потужність пласту X_1 (м) та рівні механізації робіт X_2 (%), що характеризують видобуток вугілля на 10 шахтах (табл. 2.1)

Таблиця 2.1

i	x_{i1}	x_{i2}	y_i	i	x_{i1}	x_{i2}	y_i
1	8	5	5	6	8	8	6
2	11	8	10	7	9	6	6
3	12	8	10	8	9	4	5
4	9	5	7	9	8	5	6
5	8	7	5	10	12	7	8

Припускаючи, що між змінними Y, X_1, X_2 існує лінійна кореляційна залежність, знайти її аналітичний вираз.

Розв'язання

Позначимо

$$Y = \begin{pmatrix} 5 \\ 10 \\ 10 \\ 7 \\ 5 \\ 6 \\ 6 \\ 5 \\ 6 \\ 8 \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ 1 & 12 & 8 \\ 1 & 9 & 5 \\ 1 & 8 & 7 \\ 1 & 8 & 8 \\ 1 & 9 & 6 \\ 1 & 9 & 4 \\ 1 & 8 & 5 \\ 1 & 12 & 7 \end{pmatrix}.$$

Тоді

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 8 & 11 & 12 & 9 & 8 & 8 & 9 & 9 & 8 & 12 \\ 5 & 8 & 8 & 5 & 7 & 8 & 6 & 4 & 5 & 7 \end{pmatrix} \begin{pmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ 1 & 12 & 8 \\ 1 & 9 & 5 \\ 1 & 8 & 7 \\ 1 & 8 & 8 \\ 1 & 9 & 6 \\ 1 & 9 & 4 \\ 1 & 8 & 5 \\ 1 & 12 & 7 \end{pmatrix} =$$

$$=\begin{pmatrix} 10 & 94 & 63 \\ 94 & 908 & 603 \\ 63 & 603 & 417 \end{pmatrix};$$

$$X^TY=\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 8 & 11 & 12 & 9 & 8 & 8 & 9 & 9 & 8 & 12 \\ 5 & 8 & 8 & 5 & 7 & 8 & 6 & 4 & 5 & 7 \end{pmatrix}\begin{pmatrix} 5 \\ 10 \\ 10 \\ 7 \\ 5 \\ 6 \\ 6 \\ 5 \\ 6 \\ 8 \end{pmatrix}=\begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix}.$$

Знайдемо матрицю, обернену до X^TX:

$$det\, X^TX=\begin{vmatrix} 10 & 94 & 63 \\ 94 & 908 & 603 \\ 63 & 603 & 417 \end{vmatrix}=10\begin{vmatrix} 908 & 603 \\ 603 & 417 \end{vmatrix}-94\begin{vmatrix} 94 & 603 \\ 63 & 417 \end{vmatrix}+63\begin{vmatrix} 94 & 908 \\ 63 & 603 \end{vmatrix}=$$

$$=10\cdot 15027-94\cdot 1209+63\cdot(-522)=3738.$$

$$A_{11}=\begin{vmatrix} 908 & 603 \\ 603 & 417 \end{vmatrix}=15027; \quad A_{21}=-\begin{vmatrix} 94 & 63 \\ 603 & 417 \end{vmatrix}=-1209; \quad A_{31}=\begin{vmatrix} 94 & 63 \\ 908 & 603 \end{vmatrix}=-522;$$

$$A_{12}=-\begin{vmatrix} 94 & 603 \\ 63 & 417 \end{vmatrix}=-1209; \quad A_{22}=\begin{vmatrix} 10 & 63 \\ 63 & 417 \end{vmatrix}=201; \quad A_{32}=-\begin{vmatrix} 10 & 63 \\ 94 & 603 \end{vmatrix}=-108;$$

$$A_{13}=\begin{vmatrix} 94 & 908 \\ 63 & 603 \end{vmatrix}=-522; \quad A_{23}=-\begin{vmatrix} 10 & 94 \\ 63 & 603 \end{vmatrix}=-108; \quad A_{33}=\begin{vmatrix} 10 & 94 \\ 94 & 908 \end{vmatrix}=244.$$

Таким чином, маємо:

$$\left(X^TX\right)^{-1}=\frac{1}{3738}\begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}.$$

$$b = \left(X^T X\right)^{-1} \cdot X^T Y = \frac{1}{3738}\begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}\begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix} = \begin{pmatrix} -3{,}5393 \\ 0{,}8539 \\ 0{,}3670 \end{pmatrix}.$$

Із врахуванням рівності (2.9)

$$\overline{Y} = -3{,}5393 + 0{,}854 x_1 + 0{,}367 x_2 .$$

Одержане рівняння регресії показує, що при збільшенні потужності пласту X_1 (за незмінного значення X_2) лише на 1 м видобуток вугілля на одного шахтаря Y збільшується в середньому на 0,854 т, а при збільшенні рівня механізації робіт X_2 (за незмінного значення X_1) – в середньому на 0, 367 т.

На практиці досить часто необхідне порівняння відокремленого впливу на залежну змінну різних пояснювальних змінних, коли останні виражають різними одиницями виміру. В цьому випадку використовують коефіцієнти еластичності $E_j \ \left(j = \overline{1, p}\right)$:

$$E_j = b_j \frac{\overline{x_j}}{\overline{y}}, \tag{2.11}$$

де b_j – відповідний коефіцієнт з рівняння регресії;

$\overline{x_j}$ – середнє арифметичне змінної x_j,;

$\overline{y}$ – середнє арифметичне ендогенної змінної Y.

Коефіцієнт еластичності E_j показує, на скільки відсотків зміниться в середньому Y, якщо x_j збільшити на 1%.

Приклад 2.2. За даними прикладу 2.1 порівняти відокремлений вплив на видобуток вугілля за зміну двох факторів – потужності пласту та рівня механізації робіт.

Розв'язання

Оскільки

$$\overline{x_1} = \frac{\sum_{i=1}^{10} x_{i1}}{10} = \frac{94}{10} = 9{,}4; \quad \overline{x_2} = \frac{\sum_{i=1}^{10} x_{i2}}{10} = \frac{63}{10} = 6{,}3; \quad \overline{y} = \frac{\sum_{i=1}^{10} y_i}{10} = \frac{68}{10} = 6{,}8;$$

тому за формулами (2.11) коефіцієнти еластичності:

$$E_1 = 0{,}8539 \cdot \frac{9{,}4}{6{,}8} = 1{,}18; \qquad E_2 = 0{,}367 \cdot \frac{6{,}3}{6{,}8} = 0{,}34 .$$

Таким чином, збільшення потужності пласту на 1% приводить до збільшення видобутку вугілля на 1,18%, а при збільшенні рівня механізації робіт на 1% видобуток вугілля збільшується на 0,34%. Останній результат показує, що на видобуток вугілля більший вплив має фактор "потужність пласту" у порівнянні з фактором "рівень механізації робіт".

Приклад 2.3. Побудувати лінійну економетричну модель, що характеризує залежність між витратами обігу (Y), обсягом вантажообігу (X_1) та фондомісткістю бази (X_2). Визначити коефіцієнти еластичності. Вихідні дані наведені в таблиці 2.2.

Таблиця 2.2

№	Витрати обігу	Обсяг вантажообігу	Фондомісткість бази
1	2,48	16,8	117,7
2	2,62	16,9	97,5
3	2,88	16,1	113,7
4	2,68	15,0	122,3
5	2,52	18,0	102,0
6	2,74	17,2	106,7
7	2,56	17,1	108,5
8	2,68	16,4	114,3
9	2,55	16,7	94,3

Розв'язання

Позначимо

$$Y = \begin{pmatrix} 2{,}48 \\ 2{,}62 \\ 2{,}88 \\ 2{,}68 \\ 2{,}52 \\ 2{,}74 \\ 2{,}56 \\ 2{,}68 \\ 2{,}55 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 16{,}8 & 117{,}7 \\ 1 & 16{,}9 & 97{,}5 \\ 1 & 16{,}1 & 113{,}7 \\ 1 & 15{,}0 & 122{,}3 \\ 1 & 18{,}0 & 102{,}0 \\ 1 & 17{,}2 & 106{,}7 \\ 1 & 17{,}1 & 108{,}5 \\ 1 & 16{,}4 & 114{,}3 \\ 1 & 16{,}7 & 94{,}3 \end{pmatrix}.$$

Тоді

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 16,8 & 16,9 & 16,1 & 15 & 18 & 17,2 & 17,1 & 16,4 & 16,7 \\ 117,7 & 97,5 & 113,7 & 122,3 & 102 & 106,7 & 108,5 & 114,3 & 94,3 \end{pmatrix} \begin{pmatrix} 1 & 16,8 & 117,7 \\ 1 & 16,9 & 97,5 \\ 1 & 16,1 & 113,7 \\ 1 & 15,0 & 122,3 \\ 1 & 18,0 & 102,0 \\ 1 & 17,2 & 106,7 \\ 1 & 17,1 & 108,5 \\ 1 & 16,4 & 114,3 \\ 1 & 16,7 & 94,3 \end{pmatrix} =$$

$$= \begin{pmatrix} 9 & 150,2 & 977 \\ 150,2 & 2512,16 & 16266,1 \\ 977 & 16266,1 & 06762,6 \end{pmatrix}.$$

Застосувавши засоби *Mathcad* знаходимо:

$$\left(X^T X\right)^{-1} = \begin{pmatrix} 171,3396 & -6,80699 & -0,53086 \\ -6,80699 & 0,29993 & 0,016595 \\ -0,53086 & 0,016595 & 0,002339 \end{pmatrix}.$$

$$X^T Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 16,8 & 16,9 & 16,1 & 15 & 18 & 17,2 & 17,1 & 16,4 & 16,7 \\ 117,7 & 97,5 & 113,7 & 122,3 & 102 & 106,7 & 108,5 & 114,3 & 94,3 \end{pmatrix} \begin{pmatrix} 2,48 \\ 2,62 \\ 2,88 \\ 2,68 \\ 2,52 \\ 2,74 \\ 2,56 \\ 2,68 \\ 2,55 \end{pmatrix} =.$$

$$= \begin{pmatrix} 23,71 \\ 395,311 \\ 2576,513 \end{pmatrix}.$$

$$b=\left(X^T X\right)^{-1}\cdot X^T Y=\begin{pmatrix}171{,}3396 & -6{,}80699 & -0{,}53086\\ -6{,}80699 & 0{,}29993 & 0{,}016595\\ -0{,}53086 & 0{,}016595 & 0{,}002339\end{pmatrix}\begin{pmatrix}23{,}71\\ 395{,}311\\ 2576{,}513\end{pmatrix}=\begin{pmatrix}3{,}826004\\ -0{,}07058\\ -0{,}00013\end{pmatrix}.$$

Із врахуванням рівності (2.9)

$$\overline{Y}=3{,}826004-0{,}07058x_1-0{,}00013x_2.$$

Оскільки

$$\overline{x_1}=\frac{\sum_{i=1}^{9}x_{i1}}{9}=\frac{150{,}12}{9}=16{,}68; \qquad \overline{x_2}=\frac{\sum_{i=1}^{9}x_{i2}}{9}=\frac{976{,}95}{9}=108{,}55;$$

$$\overline{y}=\frac{\sum_{i=1}^{9}y_i}{9}=\frac{23{,}67}{9}=2{,}63;$$

то за формулами (2.11) коефіцієнти еластичності:

$$E_1=-0{,}07058\cdot\frac{16{,}68}{2{,}63}\approx-0{,}45.$$

Збільшення обсягу вантажообігу на 1% веде до зменшення витрат обігу на 0,45%;

$$E_2=-0{,}00013\cdot\frac{108{,}55}{2{,}63}\approx-0{,}005.$$

При збільшенні фондомісткості бази на 1% витрати обігу зменшаться на 0,005%.

Коефіцієнт $b_0=3{,}826004$ характеризує граничні витрати обігу.

2.3 Коваріаційна матриця та її вибіркова оцінка

Перетворимо вектор оцінок параметрів (2.8) з врахуванням (2.3):

$$b=\left(X^T X\right)^{-1}X^T(X\beta+\varepsilon)=\left(X^T X\right)^{-1}\left(X^T X\right)\beta+\left(X^T X\right)^{-1}X^T\varepsilon=E\beta+\left(X^T X\right)^{-1}X^T\varepsilon,$$

або

$$b=\beta+\left(X^T X\right)^{-1}X^T\varepsilon, \tag{2.12}$$

тобто оцінки параметрів (2.8), знайдені за вибіркою, будуть містити випадкові помилки.

Оскільки математичне сподівання оцінки *b* дорівнює параметру β

$$M(b)=M\left[\beta+\left(X^TX\right)^{-1}X^T\varepsilon\right]=M(\beta)+\left(X^TX\right)^{-1}X^TM(\varepsilon)=M(\beta)=\beta$$

($M(\varepsilon)=0$ за припущенням регресійного аналізу (1.15)), то очевидно, що вектор b є несуміщеною оцінкою параметра β. Зрозумілим є той факт, що варіації оцінок параметрів визначають точність рівняння множинної регресії. Для їх вимірювання в множинному регресійному аналізі розглядають так звану *коваріаційну матрицю* вектора оцінок параметрів $\sum_b$, яка є матричним аналогом дисперсії однієї змінної:

$$\sum\nolimits_b=\begin{pmatrix}\sigma_{00} & \sigma_{01} & \cdots & \sigma_{0p}\\ \sigma_{10} & \sigma_{11} & \cdots & \sigma_{1p}\\ \vdots & \vdots & \vdots & \vdots\\ \sigma_{p0} & \sigma_{p1} & \cdots & \sigma_{pp}\end{pmatrix},$$

де елементи σ_{ij} – коваріації (або кореляційні моменти) оцінок параметрів β_i та β_j. Коваріація двох змінних визначається як математичне сподівання добутку відхилення цих змінних від їх математичних сподівань. Тому

$$\sigma_{ij}=M\left[\left(b_i-M(b_i)\right)\left(b_j-M(b_j)\right)\right]. \qquad (2.13)$$

Каваріація характеризує як ступінь розсіювання значень двох змінних відносно їх математичних сподівань, так і взаємозв’язок цих змінних.

Оскільки оцінки параметрів, отримані методом найменших квадратів, є несуміщеними, то вираз (2.13) набуває вигляду:

$$\sigma_{ij}=M\left[\left(b_i-\beta_i\right)\left(b_j-\beta_j\right)\right].$$

Розглянуваши коваріаційну матрицю $\sum_b$, легко помітити, що на її головній діагоналі знаходяться дисперсії оцінок параметрів регресії, оскільки

$$\sigma_{jj}=M\left[\left(b_j-\beta_j\right)\left(b_j-\beta_j\right)\right]=M(b_j-\beta_j)^2=\sigma_{b_j}^2. \qquad (2.14)$$

В скороченому вигляді коваріаційна матриця вектора оцінки параметрів $\sum_b$ така:

$$\sum\nolimits_b = M\left[(b-\beta)(b-\beta)^T\right]$$

(в останньому результаті легко переконатись, перемноживши вектори ($b - \beta$) та $(b-\beta)^T$).

Враховуючи рівність (2.12) перетворюємо даний вираз:

$$\sum\nolimits_b = M\left\{\left[\left(X^TX\right)^{-1}X^T\varepsilon\right]\left[\left(X^TX\right)^{-1}X^T\varepsilon\right]^T\right\} = M\left[\left(X^TX\right)^{-1}X^T\varepsilon\varepsilon^TX\left(X^TX\right)^{-1}\right] =$$

$$= \left(X^TX\right)^{-1}X^TM(\varepsilon\varepsilon^T)X\left(X^TX\right)^{-1}, \qquad (2.15)$$

оскільки елементи матриці X – невипадкові величини.

Матриця $M(\varepsilon\varepsilon^T)$ – коваріаційна матриця вектора збурень

$$\sum\nolimits_\varepsilon = M(\varepsilon\varepsilon^T) = \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & \dots & M(\varepsilon_1\varepsilon_n) \\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & \dots & M(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & \vdots & \vdots \\ M(\varepsilon_n\varepsilon_1) & M(\varepsilon_n\varepsilon_2) & \dots & M(\varepsilon_n^2) \end{pmatrix},$$

в якій усі елементи, що не лежать на головній діагоналі, дорівнюють нулю в силу некорельованості збурень ε_i та ε_j між собою. А всі елементи, які лежать на головній діагоналі, виходячи із припущень кореляційного аналізу (1.15) та (1.16), дорівнюють одній і тій самій дисперсії σ^2:

$$M(\varepsilon_i^2) = M(\varepsilon_i - 0)^2 = D(\varepsilon_i^2) = \sigma^2.$$

Тому матриця

$$\sum\nolimits_\varepsilon = M(\varepsilon\varepsilon^T) = \sigma^2E_n,$$

де E_n – одинична матриця n-го порядку.

Врахувавши рівність (2.15), одержуємо, що коваріаційна матриця вектора оцінки параметрів:

$$\sum\nolimits_b = \left(X^TX\right)^{-1}X^T\sigma^2E_nX\left(X^TX\right)^{-1} = \sigma^2\left(X^TX\right)^{-1}\left(X^TE_nX\right)\left(X^TX\right)^{-1},$$

або

$$\sum\nolimits_b = \sigma^2\left(X^TX\right)^{-1}. \qquad (2.16)$$

Таким чином, *за допомогою оберненої матриці* $(X^T X)^{-1}$ *визначається не лише сам вектор* b*, але й дисперсії та коваріації його компонент.*

2.4 Доведення теореми Гаусса-Маркова. Оцінювання дисперсії збурень

Наразі у нас з'явилась можливість навести *доведення теореми Гаусса-Маркова*, сформульованої вище.

В попередньому підрозділі ми показали, що оцінка за методом найменших квадратів $b=\left(X^T X\right)^{-1} X^T Y$ є несуміщеною оцінкою для вектора параметрів β. Будь-яку іншу оцінку b_1 вектора β можна подати у вигляді

$$b_1=\left[\left(X^T X\right)^{-1} X^T+C\right]Y,$$

де C – деяка матриця розміру $(p+1)\times n$.

Оскільки оцінки, що їх розглядають у теоремі, належать до класу несумісних, то $M(b_1)=\beta$ або $M(b_1)=M\left(\left[\left(X^T X\right)^{-1} X^T+C\right]Y\right)=\beta$.

Врахувавши, що матриця в квадратних дужках невипадкова, а за припущенням кореляційно-регресійного аналізу $M(\varepsilon)=0$, одержуємо

$$M(b_1)=\left[\left(X^T X\right)^{-1} X^T+C\right]M(Y)=\left[\left(X^T X\right)^{-1} X^T+C\right]X\beta=$$

$$=\left[\left(X^T X\right)^{-1} X^T X+CX\right]\beta=(E+CX)\beta=\beta,$$

звідки випливає, що CX=0.

Тоді

$$b_1-\beta=\left[\left(X^T X\right)^{-1} X^T+C\right]Y-\beta=\left[\left(X^T X\right)^{-1} X^T+C\right](X\beta+\varepsilon)-\beta=$$

$$=\left(X^T X\right)^{-1} X^T X\beta+CX\beta+\left[\left(X^T X\right)^{-1} X^T+C\right]\varepsilon-\beta=E\beta+\left[\left(X^T X\right)^{-1} X^T+C\right]\varepsilon-\beta=$$

$$=\beta+\left[\left(X^T X\right)^{-1} X^T+C\right]\varepsilon-\beta=\left[\left(X^T X\right)^{-1} X^T+C\right]\varepsilon.$$

За допомогою перетворень, аналогічних наведеним при одержанні формул (2.15) та (2.16), знайдемо, що коваріаційна матриця вектора оцінок b_1 дорівнює

$$\sum_{b_1} = M\left[(b_1-\beta)(b_1-\beta)^T\right] = \sigma^2\left(X^TX\right)^{-1} + \sigma^2CC^T,$$

або, враховуючи (2.16),

$$\sum_{b_1} = \sum_{b} + \sigma^2CC^T.$$

Діагональні елементи матриці CC^T невід'ємні, оскільки це є суми квадратів елементів відповідних рядків цієї матриці. Оскільки діагональні елементи матриць $\sum_{b_1}$ та $\sum_{b}$ – дисперсії компонент векторів оцінок b_1 та b, то $\sigma^2_{b_{1i}} \ge \sigma^2_{b_i}$ $(i = 1,2,\ldots,p+1)$. Це означає, що оцінки коефіцієнтів регресії, знайдених за методом найменших квадратів, мають найменшу дисперсію, *що й треба було довести*.

Розглянемо вектор залишків e, який дорівнює, з огляду на (2.4), $e=Y–Xb$. З рівностей (2.3) та (2.8) маємо

$$e = (X\beta+\varepsilon) - X\left|\left(X^TX\right)^{-1}X^T(X\beta+\varepsilon)\right| =$$

$$= X\beta + \varepsilon - X\left(X^TX\right)^{-1}\left(X^TX\right)\beta - X\left(X^TX\right)^{-1}X^T\varepsilon =$$

$$= X\beta + \varepsilon - XE\beta - X\left(X^TX\right)^{-1}X^T\varepsilon,$$

або

$$e = \varepsilon - X\left(X^TX\right)^{-1}X^T\varepsilon$$

(врахували, що добуток $\left(X^TX\right)^{-1}X^TX = E$, тобто дорівнює одиничній матриці E_{p+1} $(p+1)$-го порядку).

Знайдемо транспонований вектор залишків e^T. Оскільки при транспонуванні матриця $(X^TX)^{-1}$ не змінюється, тобто

$$\left[(X^TX)^{-1}\right]^T = \left[\left(X^TX^T\right)^T\right]^{-1} = (X^TX)^{-1},$$

то

$$e^T = \left[\varepsilon - X(X^TX)^{-1}X^T\varepsilon\right] = \varepsilon^T - \varepsilon^TX(X^TX)^{-1}X^T.$$

Тепер

$$M\left(e^{T}e\right)=M\left[\left(\varepsilon^{T}-\varepsilon^{T}X(X^{T}X)^{-1}X^{T}\right)\left(\varepsilon-\varepsilon X(X^{T}X)^{-1}X^{T}\right)\right]=$$

$$=M\left(\varepsilon^{T}\varepsilon\right)-M\left|\varepsilon^{T}X\left(X^{T}X\right)^{-1}X^{T}\varepsilon\right|-M\left|\varepsilon^{T}X\left(X^{T}X\right)^{-1}X^{T}\varepsilon\right|+$$

$$+M\left|\varepsilon^{T}X\left(X^{T}X\right)^{-1}\left(X^{T}X\right)\left(X^{T}X\right)^{-1}X^{T}\varepsilon\right|.$$

Оскільки останні два доданки взаємно знищуються, то

$$M\left(e^{T}e\right)=M\left(\varepsilon^{T}\varepsilon\right)-M\left|\varepsilon^{T}X\left(X^{T}X\right)^{-1}X^{T}\varepsilon\right|. \tag{2.17}$$

Перший доданок виразу (2.17)

$$M\left(\varepsilon^{T}\varepsilon\right)=M\left(\sum_{i=1}^{n}\varepsilon_i^2\right)=\sum_{i=1}^{n}M\left(\varepsilon_i^2\right)=\sum_{i=1}^{n}\sigma^2=n\sigma^2, \tag{2.18}$$

оскільки за припущенням кореляційно-регресійного аналізу

$$M\left(\varepsilon_i^2\right)=M\left(\varepsilon_i-0\right)^2=D\left(\varepsilon_i\right)=\sigma^2.$$

Матриця $B=X\left(X^{T}X\right)^{-1}X^{T}$ симетрична, оскільки

$$B^{T}=\left[X\left(X^{T}X\right)^{-1}X^{T}\right]^{T}=X\left(X^{T}X\right)^{-1}X^{T}, \text{ тобто } B^{T}=B.$$

Тому $\varepsilon^{T}B\varepsilon$ є квадратичною формою $\sum_{i,j=1}^{n}b_{ij}\varepsilon_i\varepsilon_j$; її математичне сподівання

$$M\left(\varepsilon^{T}B\varepsilon\right)=M\sum_{i,j=1}^{n}b_{ij}\varepsilon_i\varepsilon_j=\sum_{i,j=1}^{n}b_{ij}M(\varepsilon_i\varepsilon_j).$$

Останню суму можна розбити на дві складові суми елементів на головній діагоналі матриці *B* та поза нею:

$$M\left(\varepsilon^{T}B\varepsilon\right)=\sum_{i=1}^{n}b_{ii}M(\varepsilon_i^{\,2})+\sum_{i,j=1\,(i\neq j)}^{n}b_{ij}M(\varepsilon_i\varepsilon_j).$$

З припущення (1.17) кореляційно-регресійного аналізу випливає, що другий доданок останньої суми дорівнює нулю. Сума діагональних елементів матриці *B* утворює слід матриці *tr*(*B*). Тому

$$M\left(\varepsilon^T B\varepsilon\right)=\sum_{i=1}^{n} b_{ii} M(\varepsilon_i^{\,2})=\sigma^2\sum_{i=1}^{n} b_{ii}=\sigma^2 tr(B). \qquad (2.19)$$

Врахувавши, що $B = X\left(X^T X\right)^{-1} X^T$, отримуємо

$$M\left(\varepsilon^T B\varepsilon\right)=\sigma^2 tr(B)=\sigma^2 tr\left(X\left(X^T X\right)^{-1} X^T\right)=\sigma^2 tr\left(\left(X^T X\right)^{-1} X^T X\right)=$$

$$=\sigma^2 tr\left(E_{p+1}\right)=\sigma^2(p+1),$$

оскільки слід матриці не змінюється при її транспонуванні, а слід одиничної матриці дорівнює порядку цієї матриці.

З формул (2.17), (2.18) та (2.19) випливає:

$$M\left(e^T e\right)=M\left(\sum_{i=1}^{n} e_i^2\right)=n\sigma^2-\sigma^2(p+1)=(n-p-1)\sigma^2. \qquad (2.20)$$

Рівність (2.20) означає, що несуміщена оцінка s^2 параметра σ^2 або вибіркова залишкова дисперсія s^2 визначається за формулою:

$$s^2=\frac{e^T e}{n-p-1}=\frac{\sum_{i=1}^{n} e_i^2}{n-p-1}. \qquad (2.21)$$

В знаменнику (2.21) стоїть n–(p+1). Це означає, що (p+1) ступінь вільності втрачається при визначенні невідомих параметрів, число яких разом із вільним членом дорівнює (p+1).

2.5 Перевірка двофакторної регресії на адекватність за допомогою коефіцієнта детермінації. Критерій Фішера

Як і у випадку парної регресійної моделі сума квадратів відхилень залежної змінної від середнього може бути розвинена на дві складові:

$$Q = Q_R + Q_e,$$

де Q_R та Q_e – відповідно, сума квадратів відхилень, зумовлених регресією, та залишкова сума квадратів, що характеризує вплив неврахованих в моделі факторів.

Знайдемо формули для знаходження відповідних сум квадратів.

$$Q=\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}y_i^2-\frac{\left(\sum\limits_{i=1}^{n}y_i\right)^2}{n}=Y^TY-n\overline{y}^2 \qquad (2.22)$$

(оскільки $\sum\limits_{i=1}^{n}y_i^2=y_1^2+y_2^2+...+y_n^2=\begin{pmatrix}y_1 & y_2 & \dots & y_n\end{pmatrix}\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}=Y^TY$).

Із врахуванням (2.6) маємо

$$Q_e=\sum_{i=1}^{n}(y_i-\overline{y_x})^2=Y^TY-2b^TX^TY+b^TX^TXb=Y^TY-b^TX^TY \qquad (2.23)$$

(оскільки згідно із рівністю (2.7) $X^TXb=X^TY$).

Тоді

$$Q_R=Q-Q_e=Y^TY-n\overline{y}^2-\left(Y^TY-b^TX^TY\right)=b^TX^TY-n\overline{y}^2. \qquad (2.24)$$

Рівняння множинної регресії є значущим (іншими словами – гіпотеза H_0 про рівність нулю параметрів регресійної моделі відкидається), якщо (враховуючи (1.22) при $m=p+1$)

$$F=\frac{Q_R(n-p-1)}{Q_e p}>F_{\alpha;p;n-p-1}, \qquad (2.25)$$

де $F_{\alpha;p;n-p-1}$ — табличне значення F-критерію Фішера-Снедекора, а Q_R та Q_e визначаються за формулами (2.23) та (2.24).

В підрозділі 1.5 було введено коефіцієнт детермінації D як одну із найбільш ефективних оцінок адекватності регресійної моделі, міри якості рівняння регресії, характеристики його прогностичної сили.

$$D=\frac{Q_R}{Q}=\frac{b^TX^TY-n\overline{y}^2}{Y^TY-n\overline{y}^2}, \qquad (2.26)$$

або

$$D=1-\frac{Q_e}{Q}=1-\frac{(Y-Xb)^T(Y-Xb)}{\left(Y-\overline{Y}\right)^T\left(Y-\overline{Y}\right)}=1-\frac{e^Te}{y^Ty}, \qquad (2.27)$$

де $e = Y - Xb$, $\overline{Y} = \left(\overline{y}, \quad \overline{y}, \quad \ldots, \quad \overline{y}\right)$, $y = \left(Y - \overline{Y}\right)$ – n-вимірні вектори;

$$e^T e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \overline{y_x})^2 ,$$

$$y^T y = \sum_{i=1}^{n} (y_i - \overline{y})^2 .$$

Нагадаємо, що *коефіцієнт детермінації D характеризує частку варіації залежної змінної, зумовленої регресією. Чим ближчий цей коефіцієнт до одиниці, тим краще регресія описує залежність між пояснювальними та залежною змінними*.

Разом з тим використання лише одного коефіцієнта детермінації для вибору найкращого рівняння регресії може виявитись недостатнім. На практиці зустрічаються випадки, коли неправильно визначена модель регресії може дати досить високий коефіцієнт детермінації.

Недоліком коефіцієнта детермінації є той факт, що він збільшується при додаванні нових пояснювальних змінних, хоча це не обов'язково означає покращення якості регресійної моделі. В цьому сенсі доцільніше використовувати *скоригований (адаптований, виправлений) коефіцієнт детермінації* $\overline{D}$, що визначається за формулою

$$\overline{D} = 1 - \frac{n-1}{n-p-1}\left(1 - D\right), \tag{2.28}$$

або, врахувавши рівність (2.27),

$$\overline{D} = 1 - \frac{(n-1)e^T e}{(n-p-1)y^T y} . \tag{2.29}$$

З рівності (2.28) випливає, що чим більше число пояснювальних змінних p, тим менший коефіцієнт $\overline{D}$ у порівнянні із D. Скоригований коефіцієнт детермінації $\overline{D}$, на відміну від D, може зменшуватись при введенні в модель нових пояснювальних змінних, що не мають істотного впливу на залежну змінну. Однак навіть збільшення скоригованого коефіцієнта детермінації при введенні в модель нової пояснювальної змінної не завжди означає, що її коефіцієнт регресії є значущим (це відбувається лише в тому випадку, коли відповідне значення t-статистики більше за одиницю (за абсолютною величиною), тобто $|t| > 1$). Іншими словами, збільшення $\overline{D}$ ще не означає покращення якості регресійної моделі.

Якщо є відомим коефіцієнт детермінації D, то критерій значущості (2.25) рівняння регресії може бути записаний у вигляді (критерій Фішера):

$$F = \frac{D(n-p-1)}{(1-D)p} > F_{\alpha;k_1;k_2}, \qquad (2.30)$$

де $k_1=p$, $k_2=n-p-1$, оскільки в рівнянні множинної регресії разом із вільним членом оцінюється $m=p+1$ параметрів, $F_{\alpha;k_1;k_2}$ – табличне значення функції Фішера-Снедекора.

Приклад 2.4. За даними прикладу 2.1 визначити множинний коефіцієнт детермінації та перевірити значущість одержаного рівняння регресії Y за X_1 та X_2 на рівні $\alpha = 0{,}05$. Перевірити значущість рівняння регресії за критерієм Фішера.

Розв'язання

Обчислимо добутки векторів (див. приклад 2.1):

$$b^T X^T Y = \begin{pmatrix} -3{,}54 & 0{,}854 & 0{,}367 \end{pmatrix} \begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix} = -3{,}54 \cdot 68 + 0{,}854 \cdot 664 + 0{,}367 \cdot 445 = 489{,}65$$

та $Y^T Y = \sum_{i=1}^{10} y_i^2 = 496$. З таблиці 2.1 знаходимо також $\sum_{i=1}^{10} y_i = 68$, звідки $\overline{y} = \frac{\sum_{i=1}^{10} y_i}{n} = \frac{68}{10} = 6{,}8$ (т).

Тепер за (2.26) множинний коефіцієнт детермінації

$$D = \frac{489{,}65 - 10 \cdot 6{,}8^2}{496 - 10 \cdot 6{,}8^2} = 0{,}811.$$

Такий коефіцієнт детермінації свідчить про те, що варіація досліджуваної змінної Y – видобуток вугілля за зміну на одного робітника – на 81,1% залежить від потужності пласту X_1 та рівня механізації робіт X_2, 18,9% припадає на невраховані в моделі фактори. Отже, одержане в прикладі 2.1 рівняння регресії є значущим.

Перевіримо значущість рівняння регресії за критерієм Фішера. Фактичне значення критерію за (2.30):

$$F=\frac{0{,}811(10-2-1)}{(1-0{,}811)\cdot 2}=15{,}0$$

більше табличного $F_{0,05;2;7}=4{,}74$; визначеного на рівні значимості α=0,05 при k_1=2 та k_2=10–2–1=7 ступенях вільності (додаток А), тобто рівняння регресії є значущим і залежна змінна Y достатньо якісно описується внесеними в модель змінними X_1 та X_2.

Проробивши аналогічні розрахунки за даними прикладу 1.5 для однієї пояснювальної змінної X_1, в прикладі 1.6 було обчислено коефіцієнт детермінації $D'=0{,}75$. Порівнюючи значення D та D' можна сказати, що внесення в модель другої пояснювальної змінної X_2 неістотно збільшило величину коефіцієнта детермінації, який визначає якість моделі.

За формулою (2.28) обчислимо скоригований коефіцієнт детермінації:

при $p=1$ $$\overline{D}=1-\frac{9}{8}(1-0{,}75)\approx 0{,}71875;$$

при $p=2$ $$\overline{D}=1-\frac{9}{8}(1-0{,}811)\approx 0{,}757.$$

Легко бачити, що хоча скоригований коефіцієнт детермінації і збільшився при внесенні в модель пояснювальної змінної X_2 (рівня механізації робіт), та це ще не свідчить про значущість коефіцієнта b_2.

Значущість коефіцієнта регресії b_j можна перевірити, якщо врахувати, що статистика $\frac{(b_j-\beta_j)}{s_{b_j}}$ має t-розподіл Стьюдента із $k=n-p-1$ ступенями вільності. Тому b_j істотно відрізняється від нуля (гіпотеза H_0 про рівність параметра β_j нулю, тобто H_0: β_j =0, відкидається) на рівні значимості α, якщо $|t|=\frac{|b_j|}{s_{b_j}}>t_{1-\alpha;n-p-1}$, де $t_{1-\alpha;n-p-1}$ – табличне значення t-критерію Стьюдента, визначене на рівні значимості α при числі ступенів вільності $k=n-p-1$.

В загальній постановці гіпотеза H_0 про рівність параметра β_j заданому числу β_{j0}, тобто H_0: β_{j0}=0, відкидається, якщо

$$|t|=\frac{|b_j-\beta_{j0}|}{s_{b_j}}>t_{1-\alpha;n-p-1}. \qquad (2.31)$$

Тому довірчий інтервал для параметра β_j:

$$b_j - t_{1-\alpha;n-p-1} \cdot s_{b_j} \le \beta_j \le b_j + t_{1-\alpha;n-p-1} \cdot s_{b_j} . \qquad (2.32)$$

Приклад 2.5. За даними прикладу 2.3 визначити множинний коефіцієнт детермінації та перевірити значущість одержаного рівняння регресії Y за X_1 та X_2 на рівні $\alpha = 0{,}05$.

Розв'язання

Обчислимо добутки векторів (див. приклад 2.3):

$$b^T X^T Y = \begin{pmatrix} 3{,}826004 & -0{,}07058 & -0{,}00013 \end{pmatrix} \begin{pmatrix} 23{,}71 \\ 395{,}311 \\ 2576{,}513 \end{pmatrix} = 62{,}48935$$

та $Y^T Y = \sum_{i=1}^{9} y_i^2 = 62{,}5881$. З таблиці 2.2 знаходимо також

$$\bar{y} = \frac{\sum_{i=1}^{9} y_i}{9} = \frac{23{,}67}{9} = 2{,}63$$

Тепер за (2.26) множинний коефіцієнт детермінації

$$D = \frac{62{,}48935 - 9 \cdot 2{,}63^2}{62{,}5881 - 9 \cdot 2{,}63^2} = \frac{0{,}23725}{0{,}336} \approx 0{,}7 .$$

Такий коефіцієнт детермінації свідчить про те, що варіація досліджуваної змінної Y – витрати обігу − на 70% залежить від обсягу вантажообігу X_1 та фондомісткості бази X_2, 30% припадає на невраховані в моделі фактори. Отже, одержане в прикладі 2.3 рівняння регресії, є значущим.

2.6 Моделювання нелінійної множинної регресії. Виробнича функція Кобба-Дугласа. Коефіцієнти часткової еластичності

До цих пір ми розглядали лінійні регресійні моделі. Однак співвідношення між соціально-економічними явищами та процесами далеко не завжди можна виразити лінійними функціями, оскільки при цьому можуть виникати невиправдано великі помилки.

Так, наприклад, нелінійними є *виробничі функції* (залежності між обсягом виробничої продукції та основними факторами виробництва –

працею, капіталом і т.п.), *функції попиту* (залежність між попитом на товари чи послуги та їх цінами або доходом) і інші.

Для оцінювання параметрів нелінійних моделей використовують два підходи.

Перший підхід заснований на *лінеаризації* моделі і полягає в тому, що за допомогою відповідних перетворень досліджувану залежність подають у вигляді *лінійного* співвідношення між *перетвореними* змінними.

Другий підхід зазвичай застосовують у випадку, коли підібрати відповідне лінеаризувальне перетворення не вдалось. В цьому випадку застосовують методи *нелінійної оптимізації* на базі вихідних змінних.

Для лінеаризації моделі в рамках першого підходу можуть використовуватись як моделі нелінійні за змінними, так і нелінійні за параметрами.

Якщо модель *нелінійна за змінними*, то введенням нових змінних її можна звести до лінійної моделі, для оцінювання параметрів якої можна використати звичайний метод найменших квадратів.

Так, наприклад, якщо нам потрібно оцінити параметри регресійної моделі

$$y_i = \beta_0 + \beta_1 \sqrt[3]{x_{i1}} + \beta_2 \ln x_{i2} + \varepsilon_i, \qquad i = \overline{1,n},$$

то, зробивши заміну $Z_1 = \sqrt[3]{X_1}$, $Z_2 = \ln X_2$, одержуємо лінійну модель

$$y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \varepsilon_i, \qquad i = \overline{1,n},$$

параметри якої знаходять методом найменших квадратів за формулою (2.8).

Слід, однак, відмітити і недолік такої заміни, пов'язаний з тим, що ми одержуємо вектор оцінок b не з умови мінімізації суми квадратів відхилень для вихідних змінних, а з умови мінімізації квадратів відхилень для перетворених змінних. У зв'язку з цим необхідне певне уточнення одержаних оцінок.

Більш складною проблемою є *нелінійність* моделі *за параметрами*, оскільки безпосереднє застосування методу найменших квадратів для їх оцінювання неможливе. До таких моделей можна віднести, наприклад, *мультиплікативну (степеневу)* модель

$$y_i = \beta_0 x_{i1}^{\beta_1} x_{i2}^{\beta_2} \varepsilon_i, \; i = \overline{1,n}, \tag{2.33}$$

експоненціальну модель

$$y_i = e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}} \varepsilon_i,\ i = \overline{1,n} \qquad (2.34)$$

і інші.

В окремих випадках, шляхом необхідної заміни, ці моделі можна звести до лінійної форми. Так, моделі (2.33) та (2.34) можуть бути зведені до лінійних шляхом логарифмування обох частин рівнянь. Тоді, наприклад, модель (2.33) стане така:

$$ln\, y_i = ln\, \beta_0 + \beta_1\, ln\, x_{i1} + \beta_2\, ln\, x_{i2} + ln\, \varepsilon_i,\ i = \overline{1,n}. \qquad (2.35)$$

Відмітимо, що до моделі (2.35) вже можна застосовувати звичайні методи дослідження лінійної регресії, розглянуті вище. Однак слід зауважити, що критерії значущості параметрів, які застосовують для нормальної лінійної регресії, вимагають, щоб нормальний закон розподілу в моделях (2.33), (2.34) мав логарифм вектора збурень ε (тобто $ln\varepsilon \approx N_n\left(0, \sigma^2 E_n\right)$), а зовсім не ε. Іншими словами, вектор збурень ε повинен мати логарифмічно нормальний розподіл.

Зауважимо, що до моделі

$$y_i = \beta_0 {x_{i1}}^{\beta_1} {x_{i2}}^{\beta_2} + \varepsilon_i,\ i = \overline{1,n}, \qquad (2.36)$$

які розглядають як альтернативні відносно до моделі (2.33), викладені вище методи дослідження лінійної регресії вже стають неможливими, оскільки модель (2.36) не можна звести до лінійного виду. В цьому випадку використовують спеціальні (ітеративні) процедури оцінювання параметрів.

Як приклад використання лінеаризувального перетворення регресії розглянемо ***виробничу функцію Кобба-Дугласа***

$$Y = AK^{\alpha} L^{\beta}, \qquad (2.36)$$

де Y – обсяг виробництва;
  K – витрати капіталу;
  L – витрати праці.

Показники α та β є коефіцієнтами часткової еластичності обсягу виробництва Y відповідно до витрат капіталу та праці. Це означає, що при збільшенні одних лише витрат капіталу (праці) на 1% обсяг виробництва збільшиться на α% (β%).

Зауважимо, що *коефіцієнт часткової еластичності* $E_{x_i}(y)$ функції $y = f\left(x_1, x_2, \ldots, x_n\right)$ відносно змінної $x_i \left(i = \overline{1,n}\right)$ називають границею відношення відносного часткового приросту функції до відносного приросту цієї змінної, коли останній прямує до нуля, тобто

$$E_{x_i}(y) = \lim_{\Delta x_i \to 0}\left(\frac{\Delta_{x_i} y}{y} \div \frac{\Delta x_i}{x_i}\right) = \lim_{\Delta x_i \to 0}\left(\frac{\Delta_{x_i} y}{\Delta x_i} \cdot \frac{x_i}{y}\right) = \frac{x_i}{y} \cdot \frac{\partial y}{\partial x_i}.$$

Для функції Кобба-Дугласа маємо:

$$E_K(y) = \frac{K}{AK^{\alpha}L^{\beta}} \cdot AL^{\beta}\alpha K^{\alpha-1} = \alpha; \qquad E_L(y) = \frac{L}{AK^{\alpha}L^{\beta}} \cdot A\beta L^{\beta-1}K^{\alpha} = \beta.$$

Враховуючи вплив випадкових збурень, функцію Кобба-Дугласа (2.36) можна записати у вигляді

$$Y = AK^{\alpha}L^{\beta}\varepsilon. \qquad (2.37)$$

Одержану мультиплікативну (степеневу) модель легко звести до лінійної шляхом логарифмування обох частин рівняння (2.37). Тоді для i-го спостереження одержимо

$$\ln y_i = \ln A + \alpha \ln K_i + \beta \ln L_i + \ln \varepsilon_i,\ i = \overline{1,n}. \qquad (2.38)$$

Якщо в моделі (2.37) $\alpha+\beta=1$ (тобто модель така, що при розширенні масштабу виробництва – збільшення витрат капіталу та праці в певну кількість разів – обсяг виробництва збільшується в те саме число разів) функцію Кобба-Дугласа подають у вигляді

$$Y = AK^{\alpha}L^{1-\alpha}\varepsilon,$$

або

$$\frac{Y}{L} = A\left(\frac{K}{L}\right)^{\alpha}\varepsilon. \qquad (2.39)$$

Таким чином, отримуємо залежність продуктивності праці $\left(\frac{Y}{L}\right)$ від її капіталоозброєності $\left(\frac{K}{L}\right)$. Для оцінювання параметрів моделі (2.39) шляхом логарифмування зводимо її до виду (для i-го спостереження)

$$\ln\left(\frac{Y}{L}\right)_i = \ln A + \alpha \ln\left(\frac{K}{L}\right)_i + \ln \varepsilon_i, \qquad i = \overline{1,n}. \qquad (2.40)$$

Функція Кобба-Дугласа з врахуванням технічного прогресу така:

$$Y = A_0 K^{\alpha} L^{\beta} e^{\theta t}, \qquad (2.41)$$

де A_0 – параметр масштабування і початкової ефективності виробництва (загальна базова ефективність);

t – час;

θ – параметр, який відображає темп приросту обсягу виробництва завдяки технічному прогресу (автономний темп).

Модель (2.41) зводиться до лінійного виду аналогічно моделі (2.37).

Ступінь однорідності функції Кобба-Дугласа визначають сумою $\alpha+\beta$. Для статичної функції приймають $\alpha+\beta=1$ (постійна віддача ресурсів), для динамічної функції – $\alpha+\beta>1$. Остання умова відображає зростання загальної ефективності виробничих факторів в економічній динаміці.

Здійснимо нормування коефіцієнтів еластичності ресурсів:

$$\frac{\beta}{\alpha+\beta} = \gamma, \qquad \frac{\alpha}{\alpha+\beta} = 1-\gamma,$$

тоді динамічна функція набуває вигляду:

$$Y = A_0 e^{\theta t} \left(L^{\gamma} K^{1-\gamma} \right)^{\alpha+\beta}. \qquad (2.42)$$

Прологарифмуємо обидві частини рівняння (2.41):

$$lnY = lnA_0 + \theta t + \alpha lnK + \beta lnL.$$

Одержану рівність продиференціюємо за t. Враховуючи, що для траєкторії $Q(t)\frac{d\,lnQ(t)}{dt} = \tilde{\rho}(t)$ – неперервний темп приросту, отримаємо рівність:

$$\rho_Y = \theta + \alpha\rho_K + \beta\rho_L, \qquad (2.43)$$

де ρ_Y – темп приросту національного доходу;

ρ_L, ρ_K – темпи приросту затрат праці та основних фондів відповідно.

Рівняння (2.43) показує, що темпи приросту національного доходу є сумою автономного темпу і зваженої суми темпів приросту виробничих факторів.

Якщо $\rho_Y > \rho_L$ і $\rho_Y > \rho_K$, то це означає збільшення ефективності обох виробничих факторів (зростання ефективності праці та зростання фондовіддачі).

У разі $\rho_K > \rho_Y > \rho_L > 0$ $\theta>0$ може бути лише при $\alpha<1$, тобто за умови зменшення фондовіддачі.

Приклад 2.6. Є дані (умовні) про продуктивність праці $\left(\frac{Y}{L}\right)$ та капіталоозброєність $\left(\frac{K}{L}\right)$ для n=10 підприємств Вінницької області (табл. 2.3). Оцінити виробничу функцію Кобба-Дугласа у вигляді (2.39) на рівні β=0,2.

Таблиця 2.3

№	$\frac{Y}{L}$	$\frac{K}{L}$
1	0,147619	7,005952
2	0,15503	5,769231
3	0,178882	7,062112
4	0,178667	8,153333
5	0,14	5,666667
6	0,159302	6,203488
7	0,149708	6,345029
8	0,163415	6,969512
9	0,152695	5,646707
10	0,135948	7,24183

Розв'язання

Від вихідних значень змінних перейдемо до їх натуральних логірифмів та, використовуючи метод найменших квадратів, розрахуємо оцінки параметрів моделі (2.40). Для цього складемо систему нормальних рівнянь, скориставшись допоміжною розрахунковою таблицею (табл. 2.4).

Таблиця 2.4

№	$ln\frac{Y}{L}$	$ln\frac{K}{L}$	$\left(ln\frac{K}{L}\right)^2$	$\left(ln\frac{Y}{L}\right)\left(ln\frac{K}{L}\right)$
1	-1,910543	1,946767	3,789901	-3,719382
2	-1,86433	1,752499	3,071252	-3,267236
3	-1,720369	1,954728	3,820963	-3,362855
4	-1,720369	2,098386	4,403224	-3,609999
5	-1,966113	1,73466	3,009045	-3,410537
6	-1,838851	1,825033	3,330746	-3,355964
7	-1,89712	1,847667	3,413874	-3,505246
8	-1,814005	1,941615	3,76987	-3,5221
9	-1,877317	1,731124	2,996792	-3,24987
10	-1,9951	1,979897	3,919994	-3,950094
$\sum$	**-18,60412**	**18,81238**	**35,52566**	**-34,95328**

В нашому випадку

$$\begin{cases} 10\,ln\,A + 18{,}81238\alpha = -18{,}60412 \\ 18{,}81238\,ln\,A + 35{,}52566\alpha = -34{,}95328 \end{cases}.$$

Розв'язавши систему методом Гаусса, знаходимо:

$$\alpha \approx 0{,}34; \quad ln\,A = -2{,}5; \; A \approx 0{,}08195.$$

Таким чином, виробнича функція Кобба-Дугласа набуває вигляду:

$$\overline{\left(\frac{Y}{L}\right)} = 0{,}08195\left(\frac{K}{L}\right)^{0{,}34}.$$

Коефіцієнт еластичності α=0,34 говорить про те, що при зміні капіталоозброєності праці на 1% продуктивність праці на підприємствах збільшується в середньому на 0,34%.

Легко показати, що рівняння $\overline{\frac{Y}{L}} = -2{,}5 + 0{,}34\,ln\left(\frac{K}{L}\right)$ є значущим, оскільки за допомогою t-критерію Стьюдента значення t-відношення t=1,5 більше за табличне значення $t_{1-0{,}3;8} = 1{,}4$ (додаток Б).

2.7 Часткова кореляція

Якщо змінні корелюють одна з одною, то на значення коефіцієнта кореляції частково впливають інші змінні. У зв'язку з цим часто виникає необхідність дослідити часткову кореляцію між змінними при вилучені (елімінуванні) впливу однієї або декількох змінних.

Вибірковим частковим коефіцієнтом кореляції (або просто частковим коефіцієнтом кореляції) між змінними X_i та X_j при фіксованих значеннях решти (p-2) змінних називають вираз

$$r_{ij.1,2,\dots,p} = \frac{-q_{ij}}{\sqrt{q_{ii}q_{jj}}}, \qquad (2.44)$$

де q_{ii} та q_{jj} – алгебраїчні доповнення елементів r_{ii} та r_{jj} матриці вибіркових коефіцієнтів кореляції

$$q_p = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix},$$

а r_{ij} визначають за формулами (1.23)-(1.25). Зокрема, у випадку трьох змінних (n=3) із (2.44) випливає, що

$$r_{ij.k} = \frac{r_{ij}^{\ 2} - r_{ik} r_{jk}}{\sqrt{\left(1 - r_{ik}^{\ 2}\right)\left(1 - r_{jk}^{\ 2}\right)}}. \tag{2.45}$$

Пояснимо одержану формулу (2.45). Припустимо, що є звичайна регресійна модель $x_i = \beta_0 + \beta_1 x_j + \beta_2 x_k + \varepsilon_i$ та необхідно оцінити кореляцію між залежною змінною X_i та пояснювальною змінною X_j при виключенні (елімінуванні) впливу іншої пояснювальної змінної X_k. З цією метою знайдемо рівняння парної регресії X_i за X_k ($\overline{x_i} = b_0 + b_1 x_k$) та X_j за X_k ($\overline{x_i} = b_0' + b_1' x_k$), а потім вилучимо вплив змінної X_k, обравши залишки $e_{x_i} = x_i - \overline{x_i}$ та $e_{x_j} = x_j - \overline{x_j}$. Очевидно, що коефіцієнти кореляції між залишками e_{x_i} та e_{xj} відображають щільність часткової кореляції між змінними X_i та X_j при вилученні змінної X_k. Можна показати, що знайдений за формулою (1.23) звичайний коефіцієнт кореляції між залишками e_{x_i} та e_{x_j} дорівнює частковому коефіцієнту кореляції $r_{ij.k}$, визначеного за формулою (2.45).

Частковий коефіцієнт кореляції $r_{ij.1,2,\dots,p}$, як і парний коефіцієнт r_{ij}, може набувати значень від -1 до +1. Окрім того, частковий коефіцієнт кореляції, обчислений за вибіркою обсягу n, має такий самий розподіл, як і звичайний коефіцієнт кореляції r_{ij}, обчислений за $n' = n - p + 2$ спостереженнями. Тому значущість часткового коефіцієнта кореляції $r_{ij.1,2,\dots,p}$ оцінюють так само, як і звичайного коефіцієнта кореляції r, але при цьому обирають $n' = n - p + 2$.

Приклад 2.7. Для дослідження залежності між продуктивністю праці (X_1), віком (X_2) та виробничим стажем (X_3) було проведено вибірку із 100 робітників однієї і тієї ж спеціальності. Обчислені парні коефіцієнти кореляції виявились значущими та склали: r_{12}=0,2; r_{13}=0,41; r_{23}=0,82. Обчислити часткові коефіцієнти кореляції та оцінити їх значущість на рівні α=0,05.

Розв'язання

За формулою (2.45) визначимо часткові коефіцієнти кореляції

$$r_{12.3}=\frac{0{,}2^2-0{,}41\cdot 0{,}82}{\sqrt{(1-0{,}41)^2(1-0{,}82)^2}}=-0{,}26;$$

$$r_{13.2}=\frac{0{,}41^2-0{,}2\cdot 0{,}82}{\sqrt{(1-0{,}2)^2(1-0{,}82)^2}}=0{,}44;$$

$$r_{23.1}=\frac{0{,}82^2-0{,}2\cdot 0{,}41}{\sqrt{(1-0{,}41)^2(1-0{,}2)^2}}=0{,}83.$$

Оцінимо значущість коефіцієнта $r_{12.3}$. Значення статистики t-критерію за при $n'=n-p+2=100-3+2=99$ (за абсолютною величиною)

$$|t|=\frac{|-0{,}26|\sqrt{99-2}}{\sqrt{1-(-0{,}26)^2}}=2{,}65$$

більше за табличне $t_{0,95;97}$=1,99 (додаток Б) отже, частковий коефіцієнт кореляції є значущим. Аналогічно встановлюємо значущість інших часткових коефіцієнтів кореляції.

Порівнюючи часткові коефіцієнти кореляції з відповідними парними коефіцієнтами помічаємо, що найбільші зміни відбулися з коефіцієнтом кореляції між продуктивністю праці (X_1) та віком робітників (X_2) (змінилось не лише його значення, але й знак: r_{12}=0,2; $r_{12.3}$=-0,26, причому обидва коефіцієнти є значущими).

Таким чином, між продуктивністю праці (X_1) та віком робітників (X_2) існує прямий кореляційний зв'язок (r_{12}=0,2). Якщо ж усунути (елімінувати) вплив змінної "виробничий стаж" (X_3), то в чистому вигляді продуктивність праці (X_1) знаходиться в зворотному за напрямом зв'язку з віком робітників (X_2) ($r_{12.3}$=-0,26). Цей факт легко пояснити, якщо розглядати вік лише як працеспроможність організму на певному етапі його життєдіяльності. Аналогічно можна пояснити й інші часткові коефіцієнти кореляції.

2.8 Багатофакторні моделі економічного зростання

Багатофакторна модель економічного зростання виражає залежність взаємодії динаміки обсягу виробництва від низки виробничих факторів із урахуванням зміни їх якості, ефективності використання, загальних наслідків науково-технічного прогресу та вдосконалення організації виробництва:

$$y(t) = f(X_t, A_t, t), \tag{2.46}$$

де X_t – вектор фізичних обсягів виробничих ресурсів; A_t – вектор параметрів, які характеризують якість та ефективність ресурсів.

Функція (2.46) є динамічним варіантом виробничої функції з взаємозамінними ресурсами, але, на відміну від статичної функції, її параметри змінюються у часі й вона може містити змінну часу. Модель (2.46) будують, використовуючи емпіричні динамічні ряди обсягів виробництва і виробничих ресурсів, тобто як багатофакторне рівняння регресії.

За функцією (2.46) визначають середні μ_{it} і граничні ν_{it} показники ефективності використання ресурсів:

$$\mu_{it} = \frac{f(X_t, A_t, t)}{x_{it}}, \tag{2.47}$$

$$\nu_{it} = \frac{\partial f(X_t, A_t, t)}{\partial x_{it}}. \tag{2.48}$$

Перевагами показників ефективності (2.47), (2.48) є врахування поєднаності всіх основних виробничих факторів. За їх допомогою можна дослідити поєднання виробничих факторів, які приводять, наприклад, до найбільшого зростання продуктивності праці чи до найбільшої економії деяких природних ресурсів тощо.

Ознакою підвищення ефективності використання виробничих факторів у динаміці виробництва є нерівності:

$$\frac{d\mu_{it}}{dt} > 0, \qquad \frac{d\nu_{it}}{dt} > 0.$$

Якщо середня ефективність *i*-го ресурсу зростає, то темп приросту виробництва буде вищим за темп приросту відповідного ресурсу.

В основі багатофакторного аналізу значення екстенсивних та інтенсивних факторів економічного зростання лежить розкладання повного приросту функції (2.46) на складові та економічна інтерпретація цього розкладання. Проте формально повний приріст функції *n* змінних як суму *n* доданків, які відповідають приростам аргументів (якщо вони не прямують до нуля), подати не можна. З огляду на це неможливо отримати точних однозначних оцінок внеску кожного фактора у приріст функції. Економічний зміст цього висновку очевидний: ефект взаємодії факторів не є сумою ефектів дії кожного фактора окремо, його досягають лише у сукупній взаємодії усіх факторів (властивість емерджентності будь-якої кібернетичної системи).

У разі використання багатофакторних моделей приріст обсягу виробництва умовно розкладають на складові таким чином:

$$\Delta f(X,A,t) = \Delta f(X) + \Delta f(A) + \Delta f(t), \tag{2.49}$$

де $\Delta f(X) = \sum_{i \in M} \Delta f(x_i),\ \Delta f(x_i)$ – приріст, отриманий за рахунок збільшення фізичного обсягу i-го ресурсу; $\Delta f(A) = \sum_{j \in N} \Delta f(a_j),\ \Delta f(a_j)$ – приріст, який отриманий за рахунок підвищення ефективності використання *j*-го ресурсу; $\Delta f(t)$ – приріст, отриманий за рахунок загального економічного прогресу виробництва, який відображає сукупну взаємодію всіх факторів.

Відповідно:

$\lambda_{\textbf{екс}} = \dfrac{\Delta f(X)}{\Delta f(X,A,t)}$ – частка екстенсивних факторів зростання;

$\lambda_{\textbf{інт}} = \dfrac{\Delta f(A)}{\Delta f(X,A,t)}$ – частка інтенсивних факторів зростання;

$\lambda_{\mathbf{t}} = \dfrac{\Delta f(t)}{\Delta f(X,A,t)}$ – частка факторів загального економічного прогресу (часто її приєднують до $\lambda_{\textbf{інт}}$).

При теоретичних дослідженнях багатофакторних моделей часто використовують неперервний (граничний) варіант формули (2.49):

$$dy(t) = \sum_{i \in M} \frac{\partial f(X,A,t)}{\partial x_i} dx_i + \sum_{j \in M} \frac{\partial f(X,A,t)}{\partial a_j} da_j + \frac{\partial f(X,A,t)}{\partial t} dt, \tag{2.50}$$

тобто повний диференціал функції $y(t) = f(X,A,t)$.

Приклад 2.8. Дослідити вплив екстенсивних та інтенсивних факторів зростання валового випуску підприємства залежно від взаємодії двох виробничих ресурсів: витрат праці та величини виробничих фондів. Вихідні дані, необхідні для проведення аналізу використання ресурсів, наведено в таблиці 2.5.

Таблиця 2.5

Показник	Базовий період	Поточний період
Валовий випуск підприємства	26879	31259
Кількість працівників, осіб	3680	4160
Продуктивність праці, шт./осіб	7,3041	7,5142
Величина основних виробничих фондів, грн.	51067	53014
Фондовіддача основних виробничих фондів, грн./грн.	0,5263	0,5896

Розв'язок

Припустимо, що обсяг валового випуску підприємства описують степеневою виробничою функцією, яка має такий вигляд:

$$y = x_1^{5/8} a_1^{5/8} x_2^{3/8} a_2^{3/8},$$

де y – обсяг валового випуску продукції; x_1 – витрати праці (кількість працівників); a_1 – продуктивність праці; x_2 – величина основних виробничих фондів; a_2 – фондовіддача основних виробничих фондів.

За даний період валовий випуск продукції підприємства зріс на 4380 (31259-26879) одиниць.

Приріст витрат ресурсу x_1 (фонду робочого часу) дорівнює

$$\Delta x_{1t} = 4160 - 3680 = 480.$$

Приріст ефективності використання ресурсу a_1 (продуктивності праці) становить

$$\Delta \mu_{1t} = 7{,}5142 - 7{,}3041 = 0{,}2101.$$

Приріст витрат ресурсу x_2 (величина основних виробничих фондів) становить

$$\Delta x_{2t} = 53014 - 41067 = 1947.$$

Приріст ефективності використання ресурсу a_2 (фондовіддачі основних виробничих фондів) дорівнює

$$\Delta\mu_{2t} = 0{,}5896 - 0{,}5263 = 0{,}0633.$$

За методом ланцюгових підстановок приріст валового випуску підприємства становить:

– за рахунок приросту фонду робочого часу (екстенсивна складова x_1)

$$\Delta f(x_{1t}) = x_{1зв}^{\frac{5}{8}} a_{1баз}^{\frac{5}{8}} x_{2баз}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} - x_{1баз}^{\frac{5}{8}} a_{1баз}^{\frac{5}{8}} x_{2баз}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} =$$

$$= 4160^{\frac{5}{8}} \cdot 7{,}3041^{\frac{5}{8}} \cdot 51067^{\frac{3}{8}} \cdot 0{,}5263^{\frac{3}{8}} - 3680^{\frac{5}{8}} \cdot 7{,}3041^{\frac{5}{8}} \cdot 51067^{\frac{3}{8}} \cdot 0{,}5263^{\frac{3}{8}} = 2140{,}6681;$$

– за рахунок приросту продуктивності праці (інтенсивна складова a_1)

$$\Delta f(a_{1t}) = x_{1зв}^{\frac{5}{8}} a_{1зв}^{\frac{5}{8}} x_{2баз}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} - x_{1зв}^{\frac{5}{8}} a_{1баз}^{\frac{5}{8}} x_{2баз}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} =$$

$$= 4160^{\frac{5}{8}} \cdot 7{,}5142^{\frac{5}{8}} \cdot 51067^{\frac{3}{8}} \cdot 0{,}5263^{\frac{3}{8}} - 4160^{\frac{5}{8}} \cdot 7{,}3041^{\frac{5}{8}} \cdot 51067^{\frac{3}{8}} \cdot 0{,}5263^{\frac{3}{8}} = 518{,}9359;$$

– за рахунок приросту величини основних виробничих фондів (екстенсивна складова x_2)

$$\Delta f(x_{2t}) = x_{1зв}^{\frac{5}{8}} a_{1зв}^{\frac{5}{8}} x_{2зв}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} - x_{1зв}^{\frac{5}{8}} a_{1зв}^{\frac{5}{8}} x_{2баз}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} =$$

$$= 4160^{\frac{5}{8}} \cdot 7{,}5142^{\frac{5}{8}} \cdot 53014^{\frac{3}{8}} \cdot 0{,}5263^{\frac{3}{8}} - 4160^{\frac{5}{8}} \cdot 7{,}5142^{\frac{5}{8}} \cdot 51067^{\frac{3}{8}} \cdot 0{,}5263^{\frac{3}{8}} = 417{,}3947;$$

– за рахунок приросту фондовіддачі основних виробничих фондів (інтенсивна складова a_2)

$$\Delta f(a_{2t}) = x_{1зв}^{\frac{5}{8}} a_{1зв}^{\frac{5}{8}} x_{2зв}^{\frac{3}{8}} a_{2зв}^{\frac{3}{8}} - x_{1зв}^{\frac{5}{8}} a_{1зв}^{\frac{5}{8}} x_{2зв}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} =$$

$$= 4160^{\frac{5}{8}} \cdot 7{,}5142^{\frac{5}{8}} \cdot 53014^{\frac{3}{8}} \cdot 0{,}5896^{\frac{3}{8}} - 4160^{\frac{5}{8}} \cdot 7{,}5142^{\frac{5}{8}} \cdot 53014^{\frac{3}{8}} \cdot 0{,}5263^{\frac{3}{8}} = 1303{,}0014.$$

Перевіримо формулу розкладання загального приросту на складові:

2140,6681+518,9359+417,3947+1303,0014=4380.

Частки екстенсивних та інтенсивних факторів зростання валового випуску продукції наведено в таблиці 2.6.

Результати дослідження економічного зростання валового випуску продукції показують, що стосовно виробничого фактора витрат праці на підприємстві переважає екстенсивний тип розвитку, а стосовно виробничого фактора витрат основних виробничих фондів переважає інтенсивний тип розвитку.

Таблиця 2.6

Приріст валового випуску продукції за рахунок	Частка загального приросту
фонду робочого часу	0,4887 (48,87%)
продуктивності праці	0,1185 (11,85%)
величини основних виробничих фондів	0,0953 (9,53%)
фондовіддачі основних виробничих фондів	0,2975 (29,75%)

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Запишіть класичну нормальну лінійну модель множинної регресії у звичайній та матричній формах.
2. Охарактеризуйте процес оцінювання параметрів класичної регресійної моделі методом найменших квадратів.
3. Що характеризує коефіцієнт еластичності та як його можна обчислити?
4. Що називають коваріаційною матрицею? Запишіть її. Що характеризує коефіцієнт коваріації?
5. Охарактеризуйте коваріаційну матрицю вектора збурень.
6. Оцініть коваріаційну матрицю.
7. Сформулюйте та доведіть теорему Гаусса-Маркова.
8. Оцініть дисперсію збурень.
9. Охарактеризуйте перевірку двофакторної регресії на адекватність за допомогою коефіцієнта детермінації.
10. Охарактеризуйте критерій Фішера.
11. Що називають скоригованим коефіцієнтом детермінації і як його обчислюють?
12. Вкажіть недолік коефіцієнта детермінації.
13. Як оцінити значущість коефіцієнта регресії b_j?
14. В чому полягає метод лінеаризації нелінійних множинних моделей? Вкажіть недоліки цього методу.
15. Знайдіть довірчий інтервал для параметра b_j лінійної множинної регресії.
16. Охарактеризуйте виробничу функцію Кобба-Дугласа.
17. Що називають коефіцієнтом часткової еластичності?
18. Запишіть функцію Кобба-Дугласа з врахуванням технічного прогресу.
19. За допомогою якого показника визначають ступінь однорідності функції Кобба-Дугласа.

20. Запишіть динамічну функцію Кобба-Дугласа, нормувавши коефіцієнт еластичності.
21. За якою формулою можна оцінити темп приросту національного доходу?
22. Що називають частковим коефіцієнтом кореляції? Поясніть формулу для обчислення цього коефіцієнта.
23. Що виражає багатофакторна модель економічного зростання?
24. Запишіть середні та граничні показники ефективності використання ресурсів та вкажіть їх переваги. Що можна дослідити за їх допомогою?
25. Вкажіть ознаку ефективності використання виробничих факторів у динаміці виробництва.
26. Як можна розкласти на складові приріст обсягу виробництва?
27. Запишіть формулу для знаходження частки екстенсивних факторів зростання.
28. Запишіть формулу для знаходження частки інтенсивних факторів зростання.
29. Запишіть формулу для знаходження частки факторів загального економічного прогресу.
30. Запишіть неперервний варіант формули розкладання на складові приросту обсягу виробництва.

Вправи

1. Є такі дані про виробіток литва на одного працівника X_1 (т), браковане литво X_2 (%) та про собівартість 1 т литва Y (грн.) для 25 цехів заводів (табл. 2.7).

Таблиця 2.7

i	x_{1i}	x_{2i}	y_i	i	x_{1i}	x_{2i}	y_i	i	x_{1i}	x_{2i}	y_i
1	14,6	4,2	239	10	25,3	0,9	198	19	17,0	9,3	282
2	135,	6,7	254	11	56,0	1,3	170	20	33,1	3,3	196
3	21,5	5,5	262	12	40,2	1,8	173	21	30,1	3,5	186
4	17,4	7,7	251	13	40,6	3,3	197	22	65,2	1,0	176
5	44,8	1,2	158	14	75,8	3,4	172	23	22,6	5,2	238
6	111,9	2,2	101	15	27,6	1,1	201	24	33,4	2,3	204
7	20,1	8,4	259	16	88,4	0,1	130	25	19,7	2,7	205
8	28,1	1,4	186	17	16,6	4,1	251				
9	22,3	4,2	204	18	33,4	2,3	195				

Необхідно: 1) знайти рівняння множинної регресії Y за X_1 та X_2, оцінити значущість цього рівняння та його коефіцієнтів на рівні α=0,05; 2) знайти множинний коефіцієнт детермінації та пояснити його зміст; 3) порівняти відокремлений вплив на залежну змінну кожної з пояснювальних змінних, використовуючи коефіцієнт еластичності; 4) знайти 95%-ні довірчі інтервали для коефіцієнтів регресії.

2. Є такі дані про річні ставки місячних доходів за трьома акціями за шість місяців:

Акція	Доходи по місяцях, %					
A	5,4	5,3	4,9	4,9	5,4	6,0
B	6,3	6,2	6,1	5,8	5,7	5,7
C	9,2	9,2	9,1	9,0	8,7	8,6

Є підстави припускати, що доходи Y акції C залежать від доходів X_1 та X_2 акцій A та B. Необхідно: 1) знайти множинний коефіцієнт детермінації та пояснити його зміст; 2) скласти рівняння регресії Y за X_1 та X_2; 3) перевірити значущість рівняння регресії на рівні α=0,2; 4) оцінити середній дохід акції C, якщо доходи акцій A та B склали відповідно 5.5% та 6%.

3. З метою дослідження впливу факторів X_1 − середньомісячної кількості профілактичного налаштування автоматичної лінії та X_2 – середньомісячного числа обривів нитки на показник Y – середньомісячну характеристику якості тканини (в балах) за даними 37 підприємств легкої промисловості було обчислено парні коефіцієнти кореляції: r_{y1}=0,105; r_{y2}=0,024 та r_{12}=0,996. Визначити часткові коефіцієнти кореляції $r_{y1.2}$ та $r_{y2.1}$ на оцінити їх значущість на 5%-му рівні.

ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

Завдання 2.1. Побудувати лінійну економетричну модель, що характеризує залежність між витратами обігу, обсягом вантажообігу та фондомісткістю бази. Визначити коефіцієнти еластичності та детермінації. Вихідні дані наведені в таблицях відповідно до варіанта.

Варіант 1

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,72	15,6	106,3
2	3,04	13,5	128,5
3	2,84	15,3	118,0
4	2,89	14,9	121,2
5	2,58	15,1	120.0
6	2,64	16,1	118,4
7	2,52	16,7	108,4
8	2,75	15,4	110,0
9	2,63	17,1	105,9

Варіант 2

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,58	15,1	120,0
2	2,64	16,1	118,4
3	2,52	16,7	108,4
4	2,75	15,4	110,0
5	2,63	17,1	105,9
6	2,48	16,8	117,7
7	2,62	16,9	97,5
8	2,88	16.1	113,7
9	2,68	15,0	122,3

Варіант 3

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,48	16.8	117,7
2	2,62	16,9	97,5
3	2,88	16,1	113,7
4	2,68	15,0	122,3
5	2,52	18,0	102,0
6	2,74	17,2	106,7
7	2,56	17.1	108,5
8	2,68	16,4	114,3
9	2,55	16.7	94,3

Варіант 4

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,75	16,8	110,0
2	2,63	15,5	105,9
3	2,48	17,0	117,7
4	2,62	16,8	97,5
5	2,88	16,9	113,7
6	2,68	16,1	122,3
7	2,56	15,0	102,0
8	2,74	18,0	106,7
9	2,60	17,2	108,5

Варіант 5

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,92	14,1	87,8
2	2,64	17,2	72,0
3	2,79	17,1	72,4
4	2,67	17,8	69,5
5	2,68	16,2	75,0
6	2,85	17,2	70-6
7	2,40	16,8	73,4
8	2.91	14,8	80,7
9	2,29	19,6	62.2

Варіант 6

№	***Витрати***	Вантажо-обіг	Фондо-міскість
1	2,45	17,1	71,3
2	2,38	19,5	61,7
3	3,04	12,5	96,2
4	2,67	16,5	72,9
5	2,70	16,0	75,0
6	2,65	16,1	74,6
7	2,79	16,2	74,1
8	2,49	18,0	66,9
9	3,27	11,4	98,6

Варіант 7

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,85	17,2	70,6
2	2,40	16,8	73,4
3|	2,91	14,8	80,7
4	2,29	19,6	62,2
5	3,27	11,4	98,6
6	2,45	17,1	71,3
7	2,38	19,5	61,7
8	3,04	12,5	96,2
9	2,67	16.5	72,9

Варіант 8

№	***Витрати***	Вантжо-обіг	Фондо-міскість
1	2,93	18,1	71,2
2	2,50	17,2	73,4
3	2,95	14,9	81,2
4	2,39	20,1	63.7
5	3,25	11,4	96,6
6	2,65	17,1	72,2
7	2,42	19,5	61,7
8	3,14	17,5	96,2
9	2,75	16,7	72,9

Варіант 9

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,14	17,4	80,3
2	2,94	13,8	102,5
3	2,67	15,0	94.3
4	2,44	18,6	76,0
5	2,83	16,2	87.3
6	2,92	15,7	90,1
7	2,61	17.9	82.8
8	2,72	15,3	96,9
9	2,68	16,3	83,7

Варіант 10

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,67	15,0	94,0
2	2,45	18,6	78,0
3	2,86	16.2	87.5
4	2.90	15,7	90,2
5	2,60	17.9	84,8
6	2,72	16,3	95,9
7	2,68	17,7	91,0
8	2,50	16,8	84,7
9	2,74	17,5	88,2

Варіант 11

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,83	13,8	68.0
2	2,75	14,8	64,3
3	2,40	16,9	55,1
4	2,30	16,8	55,5
5	2,47	14,8	63.3
6	2,45	17,9	52,7
7	2,48	17,6	53,7
8	2,41	15,7	60,2
9	2,34	15,2	62.2

Варіант 12

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,69	14,9	69,4
2	2,48	16,1	58,7
3	2,11	19,7	62,3
4	2,82	14,0	83,8
5	2,43	17,1	68,5
6	2,34	18,2	64,5
7	2,48	17,4	67,6
8	2,69	16,1	72,9
9	2,36	18,8	62,4

Варіант 13

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,35	16,8	56.5
2	2,48	15,0	64,3
3	2,45	17,5	53,7
4	2,47	17,6	54,7
5	2,42	15,7	60,2
6	2,34	15.2	62,4
7	2,70	14.9	69.5
8	2,48	16,1	58.7
9	2,15	19,7	62.3

Варіант 14

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,46	19,0	65,4
2	2,70	16,3	73,9
3	2,58	17,5	68,5
4	2,34	18,4	64,5
5	2,43	17.2	69,3
6	2,84	15,0	83,8
7	2.15	19,8	62,3
8	2,49	16.1	58,7
9	2,60	14,9	69,4

Варіант 15

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,27	32,1	112,5
2	1,94	31,0	116,4
3	2,32	32,4	111,6
4	2,49	33,2	108,9
5	2,57	31,2	116,5
6	2,01	34,8	104,5
7	1,87	35,4	102,7
8	2,39	33,0	110,2
9	2,18	34,8	104,7

Варіант 16

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,17	33,3	109,4
2	1,80	36,1	101,1
3	2,36	38,3	102,6
4	2,50	30,6	128,5
5	2.27	32,1	122,5
6	2,33	37,6	105,2
7	2,51	34,8	114,8
8	2,40	34,2	116,0
9	2,50	34,2	116,0

Варіант 17

№	***Витрати***	Вантажо-обіг	Фондо-місткіст
1	2,95	29,4	152,0
2	2,55	35,4	126,2
3	2,26	39,7	112,6
4	2,49	37,1	120,5
5	2,17	35,7	125,3
6	2,38	40,2	111,3
7	2,22	39,4	112,2
8	2,64	43,7	121,2
9	2,63	38,4	126.4

Варіант 18

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,32	38,8	114,0
2	2,19	39,9	103,1
3	2,83	30,1	153,8
4	2,75	31,7	146,0
5	2,59	37,2	124,8
6	2,27	39,7	103,6
7	2,05	36,9	119,0
8	1,95	38,2	108,7
9	2,08	40,1	106,5

Варіант 19

№	***Витрат и***	Вантажо-обіг	Фондо-місткість
1	3,95	29,4	152,0
2	3,55	35,4	126,2
3	3,26	39,7	112,6
4	3,49	37,1	120,5
5	3,17	35,7	125,3
6	3,38	40,2	111,3
7	3,22	39,4	112,2
8	3,64	43,7	121,2
9	3,63	38,4	126.4

Варіант 20

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,32	39,8	114,0
2	2,19	40,9	103,1
3	2,83	30,1	153,8
4	2,75	32,7	146,0
5	2,59	33,2	124,8
6	2,27	39,7	103,6
7	2,05	36,3	119,0
8	1,95	28,2	108,7
9	2,08	40,1	106,5

Варіант 21

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,95	29,41	153,0
2	2,55	35,42	127,2
3	2,26	39,73	113,6
4	2,49	37,14	121,5
5	2,17	35,75	126,3
6	2,38	40,24	112,3
7	2,22	39,43	113,2
8	2,64	43,72	122,2
9	2,63	38,41	127.4

Варіант 22

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,32	38,81	115,0
2	2,19	39,92	104,1
3	2,83	30,13	154,8
4	2,75	31,74	147,0
5	2,59	37,25	125,8
6	2,27	39,74	104,6
7	2,05	36,93	119,0
8	1,95	38,22	109,7
9	2,08	40,11	107,5

Варіант 23

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,08	32,5	81,6
2	1,99	33,4	79.4
3	1,96	37,8	69,5
4	2,18	35,8	85.4
5	1,91	34,2	84,3
6	2,37	37,2	71,4
7	1,92	38,2	78.1
8	2,15	29,4	90,8
9	2,41	37.2	72,1

Варіант 24

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,20	34,5	83,6
2	1,96	35,0	76,5
3	2,25	43,7	76,9
4	2,69	31.9	104,6
5	2,24	37.3	72,3
6	2,43	40,9	66,3
7	2,32	38,8	69,6
8	2,60	35,7	75,6
9	2,70	43,2	62,4

Варіант 25

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	4,95	32,1	69,4
2	4,55	31,0	58,7
3	4,26	32,4	62,3
4	4,49	33,2	83,8
5	4,17	31,2	68,5
6	4,38	34,8	64,5
7	4,22	35,4	67,6
8	4,64	33,0	72,9
9	4,63	34,8	62,4

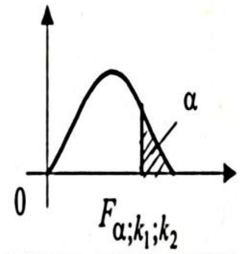
СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Грубер И. Эконометрия : учебное пособие для студентов економических специальностей / Грубер И. − К., 1996. − 397 с.
2. Данко П. Е. Высшая математика в упражнениях и задачах / П. Е. Данко, А. Г. Попов. − М. : Высшая школа, 1974. – 280 с.
3. Замков О. О. Математические методы в экономике / Замков О. О., Толстопятенко А. В., Черемных Ю. Н. − М. : ДИС, 1998. – 385 с.
4. Лук'яненко І. Економетрика : підручник / І. Лук'яненко, Л. Краснікова. – К. : Товариство "Знання", КОО, 1998. – 494 с.
5. Лотов А. В. Введение в экономико-математическое моделирование / Лотов А. В. − М. : Наука, 1984. − 392 с.
6. Сытник В. Ф. Математические модели в планировании и управлении предприятиями / В. Ф. Сытник, Е. А. Карагодова − Киев : Выща школа, 1985. − 214 с.
7. Терехов Л. Л. Экономико-математические методы / Терехов Л. Л. − М. : Статистика, 1972. – 250 с.
8. Толбатов Ю. А. Економетрика : підручник / Толбатов Ю. А. – К. : Четверта хвиля, 1997. − 362 с.
9. Наконечний С. І. Економетрія / Наконечний С. І., Терещенко Т. О., Романюк Т. П. – К. : КНЕУ, 2006. – 528 с.
10. Кремер Н. Ш. Эконометрика / Кремер Н. Ш., Путко Б. А. – М. : Юнити, 2004. – 311 с.
11. Здрок В. В. Економетрія : підручник / В. В. Здрок, Т. Я. Лагоцький. – К. : Знання, 2010. – 541 с.

Д О Д А Т К И

ДОДАТОК А

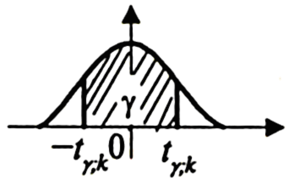
Значення $F_{\alpha,\,k_1,\,k_2}$ − критерію Фішера – Снедекора



α=0,05																			
k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	240	242	244	246	248	249	250	251	252	253	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6.59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84

ДОДАТОК Б

Значення t-критерію Стьюдента



Число ступенів вільності, k	Ймовірність, γ											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99
1	0,16	0,32	0,51	0,73	1,00	1,38	1,96	3,08	6,31	12,71	31,82	63,66
2	14	29	44	62	0,82	06	34	1,89	2,92	4,30	6,96	9,92
3	14	28	42	58	76	0,98	25	64	35	3,18	4,54	5,84
4	13	27	41	57	74	94	19	53	13	2,78	3,75	4,60
5	13	27	41	56	73	92	16	48	01	57	36	03
6	0,13	0,26	0,40	0,55	1,72	1,91	1,13	1,44	1,94	2,45	3,14	3,71
7	13	26	40	55	71	90	12	41	89	36	00	50
8	13	26	40	55	70	89	11	40	86	31	2,90	35
9	13	26	40	54	70	88	10	38	83	26	82	25
10	13	26	40	54	70	88	09	37	81	23	76	17
11	0,13	0,26	0,40	0,54	0,70	0,88	1,09	1,36	1,80	2,20	2,72	3,11
12	13	26	39	54	69	87	08	36	78	18	68	05
13	13	26	39	54	69	87	08	35	77	16	65	01
14	13	26	39	54	69	87	08	34	76	14	62	2,98
15	13	26	39	54	69	87	07	34	75	13	60	95
16	0,13	0,26	0,39	0,53	0,69	0,86	1,07	1,34	1,75	2,12	2,58	2,92
17	13	26	39	53	69	86	07	33	74	11	57	90
18	13	26	39	53	69	86	07	33	73	10	55	88
19	13	26	39	53	69	86	07	33	73	09	54	86
20	13	26	39	53	69	86	06	32	72	09	53	84
21	0,13	0,26	0,39	0,53	0,69	0,86	1,06	1,32	1,72	2,08	2,52	2,83
22	13	26	39	53	69	86	06	32	72	07	51	82
23	13	26	39	53	68	86	06	32	71	07	50	81
24	13	26	39	53	68	86	06	32	71	06	49	80
25	13	26	39	53	68	86	06	32	71	06	48	79
26	0,13	0,26	0,39	0,53	0,68	0,86	1,06	1,31	1,71	2,06	2,48	2,78
27	13	26	39	53	68	85	06	31	70	05	47	77
28	13	26	39	53	68	85	06	31	70	05	47	76
29	13	26	39	53	68	85	05	31	70	04	46	76
30	13	26	39	53	68	85	05	31	70	04	46	75
40	0,13	0,25	0,39	0,53	0,68	0,85	1,05	1,30	1,68	2,02	2,42	2,70
60	13	25	39	53	68	85	05	30	67	00	39	66
120	0,13	0,25	0,39	0,53	0,68	0,84	1,04	1,29	1,66	1,98	2,36	2,62
∞	13	25	38	52	67	84	04	28	64	96	33	58

Навчальне видання

Азарова Анжеліка Олексіївна
Сачанюк-Кавецька Наталія Василівна
Роїк Олександр Митрофанович
Міронова Юлія Володимирівна

ЕКОНОМЕТРІЯ
Частина 2

Навчальний посібник

Редактор В. О. Дружиніна
Оригінал-макет підготовлено А. Азаровою

Підписано до друку ……. р
Формат 29.7 × $42^1/_4$. Папір офісний.
Гарнітура Times New Roman.
Друк різографічний. Ум. друк. арк. …
Наклад … прим. Зам. №

Вінницький національний технічний університет
навчально-методичний відділ ВНТУ
21021, м. Вінниця, Хмельницьке шосе, 95
ВНТУ, к. 2201
Тел. (0432) 59-87-36.
Свідоцтво суб’єкта видавничої справи
серія ДК №3516 від 01.07.2009 р.

Віддруковано у Вінницькому національному технічному університеті
в комп’ютерному інформаційно-видавничому центрі
21021, м. Вінниця, Хмельницьке шосе, 95.
ВНТУ, ГНК, к. 114.
Тел. (0432) 59-87-38.
Свідоцтво суб’єкта видавничої справи
Серія ДК №3516 від 01.07.2009 р.