

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКА НАЦІОНАЛЬНА АКАДЕМІЯ МІСЬКОГО ГОСПОДАРСТВА

В. Т. Доля

Економетрія

Навчальний посібник

**для студентів вищих навчальних закладів
за спеціальностями з економіки і менеджменту**

Рекомендовано

Міністерством освіти і науки України

ХАРКІВ

ХНАМГ

2010

УДК 330.43(075)
ББК 65.053я7
Д64

Рецензенти:

В. М. Тимофеев, доктор економічних наук, професор, зав. кафедрою економічного аналізу і обліку Національного технічного університету «Харківський політехнічний інститут»;

П. Т. Бубенко, доктор економічних наук, професор, зав. кафедрою регіональної і міської економіки Харківської національної академії міського господарства.

Рекомендовано Міністерством освіти і науки України як навчальний посібник для студентів вищих навчальних закладів (лист № 14/18-Г-271 від 09.01.2009 р.)

Доля В.Т.

Д64 Економетрія: навч. посібник / В.Т. Доля; Харк. нац. акад. міськ. госп-ва. – Х.: ХНАМГ, 2010. – 171 с.

ISBN 978-966-695-177-2

У навчальному посібнику викладений теоретичний матеріал за трьома змістовими модулями економетрії: постановка задачі економетричного моделювання; специфікація економетричної моделі; оцінювання параметрів економетричної моделі. Викладення матеріалу ілюструється наскрізним прикладом розробки економетричної моделі. Для контролю і самоконтролю навчальних досягнень студента за кожним змістовим модулем наведені тестові завдання, логічні та розрахункові вправи.

У посібнику викладені поради студентам із самостійної підготовки, а також завдання для проведення практичних занять та виконання індивідуальних розрахунково-графічних і контрольних робіт.

Розрахований на студентів бакалаврату з напрямів підготовки «Економіка підприємства», «Облік і аудит», «Менеджмент».

**УДК 330.43(075)
ББК 65.053я7**

ISBN 978-966-695-177-2

© Доля В.Т., 2010

© ХНАМГ, 2010

ВСТУП

Економетрія – порівняно молода галузь науки. Її виникнення можна датувати груднем 1930 року, коли в м. Клівленді (США) було засноване Міжнародне економетричне товариство. Воно визначило себе так: «Міжнародне товариство для розвитку економічної теорії і її зв'язку зі статистикою та математикою». В установчих документах товариства міститься і перше визначення сутності економетрії – «математична статистика для економістів». Пізніше відомий економетрист Цві Грілліхес визначив сутність економетрії більш емоційно: *«Економетрія є одночасно нашим телескопом і нашим мікроскопом для вивчення оточуючого нас економічного світу»*.

Термін «економетрія» вперше запропонував львівський учений Петро Чомпа у книзі «Нариси економетрії і природної теорії бухгалтерії, яка ґрунтується на політичній економії», що вийшла друком у Львові в 1910 році. Цим терміном користувалися у своїх працях з економетрії вчені ще у 10-20-х роках минулого століття в США, Європі, СРСР. Уже тоді вони намагалися скласти так звані «барометри розвитку» економіки, поведження товарного і грошового ринків, циклічного розвитку виробництва товарів і послуг. Потужного поштовху надала економетричним дослідженням велика депресія в економіці США на початку 30-х років. Саме тоді була розроблена класична економетрична модель – виробнича функція Кобба-Дугласа.

Отже, економетрія як галузь економічної науки пройшла досить тривалий шлях формування найголовніших ознак самостійної дисципліни – предмету і методології. Великий внесок у розвиток економетричних методів моделювання належить лауреатам Нобелівської премії з економіки Р. Фрішу і Я. Тінбергену (1969 р.), Л. Клейну (1980 р.), Т. Хаавельмо (1989 р.), Дж. Хекману (2000 р.).

Дисципліна «економетрія» вперше в Україні введена до навчального плану підготовки економістів в Харківській національній академії міського господарства за ініціативою автора в 1977/1978 навчальному році. Отже, ви розгорнули посібник, який є втіленням 30-річного авторського досвіду викладання

економетрії. Він містить стислий виклад теоретичних положень економетричного моделювання та методів оцінювання параметрів економетричних моделей.

Відповідно до Галузевих стандартів Вищої освіти МОН України *метою* вивчення економетрії є формування сучасного економічного мислення та спеціальних знань і вмінь щодо використання методів економетричного аналізу як складової прийняття рішень з розвитку економічних об'єктів різної складності, ієрархії та організації.

Предметом вивчення в економетрії є економіко-математичні методи дослідження економічних явищ і процесів на макро- та мікрорівнях економіки.

Для вивчення економетрії студент має володіти необхідними попередніми знаннями, а саме: знати макро- та мікроекономіку, економіку підприємства, статистику, вищу математику (матричну алгебру), теорію ймовірностей, математичну статистику; уміти користуватися документами обліку і звітності підприємства, вибірковим методом, сучасними комп'ютерними технологіями обробки статистичних даних.

Відповідно до вимог кредитно-модульної системи організації навчального процесу зміст дисципліни поділено на три цілісні частини – змістові модулі:

ЗМ 1. «Постановка задачі економетричного моделювання»,

ЗМ 2. «Специфікація економетричної моделі»,

ЗМ 3. «Оцінювання параметрів економетричної моделі»,

засвоєння яких підлягає поточному модульному контролю з обов'язковим оцінюванням у заліково-екзаменаційній відомості.

За кожним змістовим модулем посібник містить: виклад навчальних елементів, тобто мінімальних одиниць змісту, які підлягають обов'язковому контролю; тести для суцільного контролю знань; комплекси логічних і розрахункових вправ для вибіркового контролю здатності до економіко-математичного мислення і вміння з економетричного моделювання.

Внаслідок вивчення економетрії студент повинен:

знати:

- загальну методику економетричного моделювання;
- принципи й правила формування інформаційних масивів статистичних даних для економетричного моделювання;
- методи ідентифікації незалежних змінних економетричних моделей;
- методи специфікації рівнянь регресії; виявлення мультиколінеарності, авторегресії, гетероскедастичності й засоби їх усунення або врахування;
- методи оцінювання параметрів рівнянь регресії: метод найменших квадратів, узагальнений метод найменших квадратів (метод Ейткена);
- оцінку довірчих границь параметрів рівнянь регресії та прогнозу залежної змінної;

вміти:

- формувати інформаційні масиви статистичних даних для економетричного моделювання;
- розробляти прості і багатофакторні економетричні моделі;
- користуватися програмними комплексами економетричного моделювання, наприклад, STATGRAPHICS, або Мезозавр.

Ці кваліфікаційні вимоги є мінімальними. Для тих, хто хоче досконаліше оволодіти економетричними методами, слід звернутися до підручників і лабораторних практикумів, рекомендованих МОН України для вивчення економетрії.

Нехай вас не лякають щойно наведені нові терміни і поняття – сумлінною працею ви оволодієте ними і поповните свій професійний інструментарій могутнім засобом економетричного аналізу явищ і процесів.

1. ЗМІСТОВИЙ МОДУЛЬ 1: ПОСТАНОВКА ЗАДАЧІ ЕКОНОМЕТРИЧНОГО МОДЕЛЮВАННЯ

1.1. ПРЕДМЕТ, ЗАВДАННЯ І ЗМІСТ ЕКОНОМЕТРИЧНОГО МОДЕЛЮВАННЯ

1.1.1. Предмет економетрії

Термін «економетрія» означає вимірювання в економіці, але не всі прикладні дослідження економіки засобами математики відносяться до економетрії. Економетрія вивчає *моделі і методи кількісного оцінювання параметрів моделей, що характеризують взаємозв'язки між економічними показниками на макро- і мікрорівнях економіки*. Саме тому процес вивчення економічних явищ і процесів методами економетрії називається *економетричним моделюванням*.

В економетрії будь-який результат господарської діяльності на макро- і мікрорівнях економіки розглядається як дуже складна функція як за аналітичною формою, так і за невизначено великою кількістю умов-факторів. Загальний її вигляд такий

$$y = f(x_1, x_2, \dots), \quad (1.1.)$$

де y – залежна змінна (результат);

x_i – незалежні змінні (фактори), що впливають на рівень і варіацію y .

Такий складний об'єкт моделювання об'єктивно неможливо аналітично апроксимувати однозначним (детермінованим) рівнянням. В економетричну модель не можуть бути включені всі умови – фактори, тому що:

- багато умов-факторів аналітику невідомі;
- частина факторів не мають кількісного виміру;
- про деякі фактори відсутня інформація;
- вплив багатьох факторів дуже незначний і його можна залишити поза увагою.

В економетричну модель включається обмежена кількість найбільш важливих факторів і вона набуває виду

$$\hat{y} = f(x_1, x_2, \dots, x_m), \quad (1.2)$$

де \hat{y} – математичне сподівання, або найбільш ймовірне, середнє з можливих значень y .

Таке спрощення дійсності зумовлює виникнення помилки апроксимації, або прогнозу, бо в такому разі

$$y = f(x_1, x_2, \dots, x_m) + \mathcal{E}, \quad (1.3)$$

де \mathcal{E} – випадкова складова y , що пояснюється впливом інших (не врахованих і невідомих) факторів.

Це означає, що економетричні моделі є *стохастичними, ймовірністими*.

Стохастичні залежності в економетрії називаються також *кореляційними*. Термін «кореляція» (лат. correlatio) ввів до наукової термінології англійський вчений Ф. Гальтон у 1877 р. Цей термін буквально означає співвідношення, пропорцію, зв'язок, залежність.

Для визначення залежностей в економетрії використовуються такі означення: y – *залежна* (ендогенна) змінна; x_i – *незалежні* (екзогенні), пояснюючі змінні, або фактори. Рівняння (1.2) називається *рівнянням регресії*. Така назва введена англійським антропологом К. Пірсоном на початку ХХ сторіччя. Отже, предметом економетрії, по-перше, є методологія побудови – *специфікації* – рівнянь регресії в економіці.

Рівняння регресії можуть мати лінійну

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m + \mathcal{E},$$

і нелінійну форму, наприклад, степеневу

$$y = \alpha_0 x_1^{\alpha_1} x_2^{\alpha_2} \dots x_m^{\alpha_m} + \mathcal{E},$$

де α_i ($i = \overline{0, m}$) – *параметри рівняння регресії*, або коефіцієнти регресії. Вони визначають вплив змінення факторів x_i на рівень y .

Якщо модель лінійна, то a_i показує на скільки одиниць змінюється y при зміні x_i на одиницю. У степеневій моделі коефіцієнт регресії a_i є відносною мірою змінення, він визначає, наскільки відсотків зміниться y , якщо x_i зміниться на один відсоток. Отже, по-друге, предметом економетрії є **методи оцінювання параметрів рівнянь регресії**.

Оскільки кількість незалежних змінних у дійсності невизначено велика і істинна аналітична форма залежностей y від x_i невідома, коефіцієнти регресії a_i і величину \mathcal{E} визначити абсолютно точно об'єктивно неможливо. Оцінювання цих параметрів моделі зводиться до визначення їх емпіричних, ймовірнісних оцінок, відповідно a_i і e . Тоді, наприклад рівняння лінійної регресії набуває такого виду

$$y = a_0 + a_1x_1 + \dots + a_mx_m + e.$$

Отже, підсумовуючи сказане, предметом економетрії є **специфікація** (визначення змінних та аналітичної форми) і **оцінювання параметрів рівнянь регресії**.

1.1.2. Проблеми і завдання економетричного моделювання

У процесі економетричного моделювання вирішуються загалом дві проблеми:

- 1) **побудова рівняння регресії**, тобто залежності залежної (ендогенної) змінної y від незалежних (екзогенних) змінних x_i ($i = \overline{1, m}$);
- 2) **визначення довірчих границь** для математичного сподівання \hat{y} , в межах яких знаходиться випадкова складова \mathcal{E} , що означає помилку прогнозу або апроксимації y .

Вирішення **першої проблеми** здійснюється шляхом розв'язання низки задач, а саме:

- ідентифікація змінних;
- специфікація аналітичної форми рівняння регресії;
- оцінювання параметрів рівняння регресії та його якості.

Ідентифікація, тобто обґрунтування вибору незалежних змінних спочатку ґрунтується на попередньому теоретичному, якісному аналізі економічної сутності залежності, що моделюється. Потім попередній перелік незалежних змінних повинен піддаватися кількісній оцінці сили та автономності впливу їх на залежну змінну і тільки після цього можна визначити остаточний склад незалежних змінних. Для розуміння змісту і мети ідентифікації введемо поняття **коефіцієнта детермінації d_i** . Він визначає питому вагу варіації (коливання) залежної змінної y під впливом незалежної змінної x_i у загальній змінюваності залежної змінної під впливом абсолютно всіх незалежних змінних. Очевидно, що

$$d_1 + d_2 + d_3 + \dots + d_m \leq 1 \quad (1.4)$$

Тоді, звичайно, ідентифікація незалежних змінних за складом і кількістю мусить відповідати меті моделювання, а саме: рівняння регресії повинне апроксимувати (пояснювати) наперед визначену долю варіації залежної змінної

$$d_1 + d_2 + \dots + d_m \geq d$$

і заодно визначити відносну помилку апроксимації, яка дорівнює $1 - d$.

Специфікація аналітичної форми рівняння регресії полягає в обґрунтуванні лінійної або певного виду нелінійної форми залежностей y від x_i . Воно зводиться до вибору найбільш адекватного теоретичній гіпотезі аналітика типу із відомих у математиці рівнянь. Прості форми, наприклад, такі:

- лінійна $\hat{y} = a_0 + a_1 x$;
- парабола (квадратична) $\hat{y} = a_0 + a_1 x + a_2 x^2$;
- зворотна (гіпербола) $\hat{y} = a_0 + a_1 / x$;
- степенева $\hat{y} = a_0 x^{a_1}$;
- експоненційна $\hat{y} = a_0 a_1^x$.

Оцінювання параметрів рівняння регресії – дуже складна і трудомістка задача моделювання. Це зумовлено особливостями масивів економічної інформації про розвиток економічних об'єктів. Кількісний аналіз кореляції між

варіаційними або динамічними рядами масивів даних про y та x_i пов'язаний з властивими їм *мультиколінеарністю, гетероскедастичністю, авторегресією, наявністю часового лагу* та інших ознак. Ми до цих проблем повернемося пізніше, але зауважимо, що обґрунтування вибору методу розрахунку параметрів моделі a_i потребує копіткої роботи аналітика. Зауважимо також наперед, що оцінювання параметрів рівнянь регресії потребує визначення їх *незміщеності, обґрунтованості і ефективності*.

Вирішення другої проблеми економетричного моделювання полягає у визначенні довірчих границь для оцінки \hat{y} , в межах яких знаходиться істинне значення y , або помилка апроксимації e , тобто різниця $y - \hat{y}$. Позначимо довірчу границю Δy . Тоді очевидно, що

$$0 \leq |e| \leq \Delta y,$$

$$\hat{y} - \Delta y \leq y \leq \hat{y} + \Delta y.$$

Це означає, що за рівнянням регресії розраховується точкове значення залежної змінної y (його математичне сподівання), а істинне значення може бути іншим, але воно знаходиться в інтервалі $2\Delta y$ (на осі y). Зауважимо, що цей інтервал повинен гарантуватися з певною ймовірністю.

Отже, вирішення проблем економетричного моделювання шляхом послідовного розв'язання низки прикладних задач має привести до отримання економетричної моделі у вигляді

$$y = f(x_1, x_2, \dots, x_m) \pm \Delta y. \quad (1.5)$$

1.1.3. Зміст (послідовність) економетричного моделювання

Економетричне моделювання як робочий процес має алгоритмічний характер і включає послідовне виконання наступних комплексів теоретичних, організаційних, аналітичних і розрахункових завдань і операцій.

1. Постановка задачі моделювання. На цьому логікотеоретичному етапі проводиться формування попереднього переліку ендогенних та екзогенних

змінних (y, x_1, x_2, \dots, x_m) і вибір їх операційних характеристик. Цей перелік повинен бути достатньо повним тому, що в подальшому процесі деякі фактори можуть вилучатися.

2. Формування матриці статистичних даних про змінні y, x_1, x_2, \dots, x_m за певною кількістю об'єктів спостереження (n). Кількість об'єктів спостереження не може бути якою завгодно, а принаймні достатньою за розрахунком. Аномальні об'єкти спостереження, якщо вони є, можуть вилучатися з матриці.

3. Ідентифікація незалежних змінних, тобто вибір з попереднього переліку найбільш значущих факторів для введення у рівняння регресії. На цій стадії моделювання методами кореляційного і дисперсійного аналізу визначаються сила, суттєвість (невипадковість) і автономність впливу кожного з факторів на залежну змінну, оцінюється явище мультиколінеарності факторів. Несуттєві, випадкові й сильно корельовані між собою фактори можуть вилучатися з подальшого процесу економетричного моделювання.

4. Специфікація аналітичної форми рівняння регресії, тобто обґрунтування лінійної або певного типу криволінійної форми регресії. Перевагу слід віддавати теоретичному аналізу форми залежностей, використовуючи як допоміжні графічні й аналітичні засоби обґрунтування.

5. Оцінювання параметрів рівняння регресії a_i методом найменших квадратів (МНК). При цьому перевіряється дотримання умов використання цього методу. Якщо вони не виконуються, аналізуються такі явища, притаманні матрицям економічної інформації, як гетероскедастичність, авторегресія і наявність часового лагу. Внаслідок цього знаходиться можливість усунення даних явищ, наприклад, шляхом зміни специфікації рівняння регресії або врахування їх шляхом застосування іншого методу оцінювання параметрів a_i , наприклад, узагальненого методу найменших квадратів (методу Ейткена). Параметри a_i повинні бути *незміщеними, обґрунтованими і ефективними*.

П'ятий етап моделювання називають регресійним аналізом, його результат є одночасно результатом вирішення першої проблеми економетричного моделювання.

6. Встановлення довірчих границь помилок апроксимації або прогнозу величини залежної змінної. Це дає підстави для оцінки якості економетричних моделей і можливості їх застосування у прийнятті рішень з розвитку економічних об'єктів, зокрема в економічному аналізі та прогнозуванні.

1.2. ФОРМУВАННЯ МАТРИЦІ ДАНИХ ДЛЯ ЕКОНОМЕТРИЧНОГО МОДЕЛЮВАННЯ

1.2.1. Загальна характеристика матриці

Матриця статистичних даних для економетричного моделювання являє собою прямокутну таблицю кількісних значень певних змінних у певній кількості об'єктів спостереження. Оскільки серед змінних матриці щонайменше одна є залежною змінною (y), а решта незалежними змінними (x_i), матриця має такий вигляд

$$\begin{bmatrix} y_1 & x_{11} & x_{21} & \dots & x_{m1} \\ y_2 & x_{12} & x_{22} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots & \dots \\ y_n & x_{1n} & x_{2n} & \dots & x_{mn} \end{bmatrix} \quad (1.6)$$

де $i = 1, 2, \dots, m$ – перелік незалежних змінних;

$j = 1, 2, \dots, n$ – перелік об'єктів спостереження.

Отже, x_{ij} – значення i -тої змінної у j -го об'єкту спостереження.

Матриця статистичних даних характеризується:

- **мірністю** («шириною»), рівною $m + 1$, тобто кількістю змінних, якими фіксуються найбільш суттєві і значущі властивості об'єктів спостереження i , значить, кількісні відмінності цих об'єктів;
- **об'ємом вибірки** («довжиною»), рівним кількості об'єктів спостереження n , що складають їх вибірккову сукупність;
- **об'ємом матриці**, тобто кількістю елементів інформаційного масиву (дорівнює добутку $(m + 1)n$);
- **повнотою опису об'єктів спостереження** незалежними змінними, тобто відношенням m/M , де M – загальна кількість незалежних змінних. Відношення

m/M є гіпотетичною характеристикою матриці тому, що число M невизначено велике і досліднику невідоме;

- **репрезентативністю вибірки**, або долею вибіркової сукупності об'єктів спостереження в загальній, генеральній їх кількості n/N , де N – чисельність генеральної сукупності об'єктів спостереження.

Із наведених характеристик матриці статистичних даних випливає, що в економетричному моделюванні об'єктивно виникають помилки двох видів: 1) **помилки апроксимації** (відображення, відтворення, прогнозу) – внаслідок неповного врахування незалежних змінних і 2) **помилки вибірки** – внаслідок неповної репрезентативності вибіркової сукупності об'єктів спостереження.

Для зменшення помилок апроксимації і помилок вибірки є тільки один шлях – збільшення переліку змінних і об'єму вибірки, тобто об'єму інформаційного масиву даних щодо явища або процесу, що моделюються.

1.2.2. Змінні у матриці

Мірність матриці статистики, як ми уже знаємо, визначається кількістю змінних у рівнянні

$$y = f(x_1, x_2, \dots, x_m) + \mathcal{E}.$$

Залежна змінна (y) називається **ендогенною** тому, що вона визначається в залежності від змінних x_1, x_2, \dots, x_m , включених до матриці. Незалежні змінні x_1 називаються **екзогенними** в тому сенсі, що їх величини визначаються поза матрицею статистики.

Незалежні (екзогенні) змінні – це певні кількісні ознаки об'єктів спостереження, які системно (скрізь і завжди) впливають на рівень залежної (ендогенної) змінної. Незалежні змінні можуть бути здебільшого **прямими** вимірами певної ознаки і непрямими: **опосередкованими, штучними**. Останнє трапляється, коли якусь ознаку можна виразити лише атрибутивно, а не кількісно. Тому таким ознакам штучно присвоюється певна кількісна міра. Наприклад, рівень якості продукції можна виразити номером її сорту – 1-й, 2-й, 3-й; кваліфікацію робітників – номером тарифного розряду – 1-й, 2-й тощо. Ще приклад: в моделюванні

вартості квартир наявність в них телефону можна виразити числом «1», а відсутність числом «0», розташування квартир в місті – рангом відповідного місця розташування на карті кадастрової оцінки вартості землі. У класичній економетричній виробничій функції Кобба-Дугласа

$$Q = a_0 \times L^{a_1} \times K^{a_2} \times e^t,$$

де Q – національний доход;

L – зайнятість населення;

K – авансований у виробництво і сферу послуг капітал, змінна t є штучним показником науково-технічного і соціального прогресу і виражається номером року у відрізку часу, за який спостерігалися динамічні ряди показників Q , L , і K .

Формування первинного переліку екзогенних змінних (x_1, x_2, \dots, x_m) здійснюється насамперед при **логіко-професійному аналізі**, успіх якого залежить від рівня економіко-теоретичної підготовленості аналітика, його досвіду і ступеню вивченості явища або процесу, що моделюється. Корисним може стати ознайомлення з **аналогами** подібних досліджень. Важливо скористатися **евристичними методами експертних оцінок**, коли для вирішення проблеми вибору факторів мобілізуються знання, досвід та інтуїція групи експертів.

Відбір факторів – дуже відповідальний етап економетричного моделювання. По-перше, факторів у дійсності невизначено багато, тому аналітик вимушений із дуже великого переліку змінних вибрати декілька найбільш вагомих, системно діючих і невинуватих (суттєвих) факторів, які обумовлюють **левоу частку варіації залежної змінної**.

По-друге, перелік екзогенних змінних з самого початку формується на категоріальному рівні визначення: обсяг виробництва, капітал, продуктивність праці, фондівдача тощо. Це – **економічні категорії**, змістовні поняття, які мають різноманітні **операційні характеристики**, тобто конкретні одиниці виміру. Наприклад, обсяг виробництва може бути виражений валовою продукцією, товарним випуском, обсягом реалізації (продаж), оборотом, або виручкою. Тому

залежно від мети економетричного моделювання у кожному випадку вибору операційних характеристик змінних слід давати належне обґрунтування.

По-третє, великі економічні об'єкти спостереження (національні, регіональні економіки) є досить *інерційними системами*, тому в економетричному моделюванні на макрорівні кількість змінних має бути обмеженою – не більше трьох. Навпаки, показники діяльності невеликих економічних об'єктів (підприємств, тим паче малих) дуже *чутливі до змін умов їх діяльності*. В цьому разі економетричні моделі повинні включати більше змінних (до 5-7) для підвищення їх гнучкості і чутливості.

1.2.3. Об'єкти спостереження в матриці

Об'єктами спостереження в економетричному моделюванні передусім є *виробничі одиниці* (суб'єкти господарювання):

- на мікрорівні економіки – цехи, дільниці, підприємства, (юридичні особи), об'єднання підприємств (концерни, корпорації, асоціації тощо);
- на макрорівні – регіональні економіки (територій, населених пунктів), економіки галузевих комплексів (паливно-енергетичного, промислового, транспорту і зв'язку, агропромислового), національні економіки країн.

Об'єкти спостереження у межах однієї матриці мають бути *якісно однорідними*. Їх необхідно вибирати на одному і тому ж рівні ієрархії економіки, наприклад, всі об'єкти спостереження є тільки, підприємства, тільки об'єднання підприємств, або тільки виробничі дільниці. Крім того, об'єкти спостереження повинні бути одного і того ж виду економічної діяльності (наприклад, усі об'єкти в матриці – цегельні заводи, або цукрові заводи, або всі вони бригади шлюсарів-складальників машин і обладнання). Ці вимоги якісної однорідності щодо об'єктів спостереження є дуже важливими умовами одержання якісних економетричних моделей тому, що такий підхід суттєво звужує загальну кількість ознак різноманітності об'єктів спостереження і тим самим не тільки полегшує вибір змінних, але робить його більш повним і зменшує помилку апроксимації (прогнозу).

В економетричному моделюванні рекомендуються для відбору об'єктів спостереження з генеральної сукупності такі способи: випадковий, механічний, типічний. **Випадковий відбір** – це відбір наугад, за жеребкуванням, «як вийде». Цей найбільш простий спосіб малопридатний уже тому, що не забезпечує однакової щільності вибору із різних ділянок генеральної сукупності. **Механічний відбір** у цьому сенсі кращий: генеральна сукупність об'єктів вистроюється у певний ранжований ряд і потім із нього вибирається, наприклад, кожен п'ятий, десятий, сотий тощо. Щільність відбору при цьому буде однаковою на усіх ділянках генеральної сукупності об'єктів, але такий вибір організаційно набагато складніший. **Типічний відбір** – це комбінація перших двох способів, тому він має переваги механічного і простоту випадкового. Генеральна сукупність об'єктів спостереження ділиться на групи (наприклад, за територіальним принципом), а вже в межах кожної групи проводиться випадковий відбір у кількості, пропорційній чисельності відповідних груп.

Кожний j -й об'єкт спостереження включається в матрицю статистики рядком значень залежної і незалежних змінних $y_j, x_{1j}, x_{2j}, \dots, x_{mj}$. Оскільки об'єкти спостереження є, як правило, виробничими одиницями, значення змінних беруть із документів обліку (первинних документів обліку, облікових реєстрів, внутрішньогосподарської звітності), статистичних і фінансових звітів тощо.

Якщо об'єктами спостереження є галузеві, регіональні і національні економіки, значення змінних беруться із статистичних довідників, які щоквартально і щорічно друкують державні органи статистики адміністративно-територіальних одиниць і країн.

У спеціальних економетричних дослідженнях на мікрорівні об'єктами спостереження можуть бути, наприклад, окремі види продукції. У такому разі залежною змінною виступають показники собівартості, трудомістськості, або ціни одиниці продукції, а незалежними змінними є техніко-технологічні, організаційні і цінові фактори виробництва і збуту продукції.

Тут необхідно розглянути ще одне дуже важливе питання, яке стосується використання так званого методу «заводо-років». За цим методом з метою

збільшення об'єму вибірки один і той же об'єкт спостереження включається до матриці статистики не одним рядком, а декількома (наприклад, за декілька років спостереження), які умовно розглядаються як **незалежні об'єкти спостереження**. В економетричному моделюванні на макрорівні цей засіб часто поширюється на всю матрицю – спостерігається один єдиний економічний об'єкт, але за n років. Тому вектори значень змінних матриці можуть бути:

- **варіаційними рядами** змінних y, x_i , коли один об'єкт спостереження займає в матриці тільки один рядок;
- **часовими рядами** цих змінних, коли спостереження велись за одним економічним об'єктом n періодів часу поспіль (наприклад, місяців, років);
- **комбінованими рядами**, коли до матриці включаються різні об'єкти спостереження і деякі з них, або всі, спостерігаються по декілька разів.

В останніх двох випадках принцип незалежності об'єктів спостереження порушується, внаслідок чого у моделюванні виникає проблема **автокореляції**, тобто наявності залежності значень змінних за наступний період часу від їх значень за попередні періоди часу. Проблему автокореляції ми будемо розглядати далі.

1.2.4. Вимоги до розмірів матриці

Розміри матриці статистики визначаються, як уже відомо із підрозділу 1.2.1, її мірністю ($m + 1$) і обсягом вибірки об'єктів спостереження, тобто «довжиною» (n). Чим більше змінних, тим менша помилка апроксимації, прогнозу і, чим більше об'єктів спостереження, тим менша помилка вибірки. Отже, чим більший інформаційний масив статистичних даних, тим краще.

Вихідним моментом у визначенні розмірів матриці статистики є формування первісного переліку незалежних (екзогенних) змінних (див. підрозділ 1.2.2).

Накопичений досвід свідчить, що для отримання **незміщених, обґрунтованих і ефективних** оцінок коефіцієнтів регресії кількість об'єктів спостереження повинна перевищувати кількість змінних щонайменше у вісім разів. Отже мінімальна кількість об'єктів спостереження визначається так

$$n_{\min} \geq 8(m+1). \quad (1.7)$$

Без доказів ясно, що, чим більше n , тим краще – при збільшенні n зменшуються помилки вибірки. Однак об'єм вибірки не може бути нескінченим! Значить є межа для n , коли кількість об'єктів спостереження стає достатньою. Вона визначається чотирма умовами:

- 1) об'ємом генеральної сукупності об'єктів спостереження (N);
- 2) кількісною однорідністю об'єктів спостереження, що характеризується середньоквадратичним відхиленням залежної змінної (σ_y);
- 3) прийнятною (наперед визначеною) граничною помилкою вибірки для оцінок залежної змінної за допомогою рівняння регресії, що моделюється (Δy);
- 4) ймовірністю неперевищення граничної помилки і відповідним їй коефіцієнтом довіри (t).

Достатня кількість об'єктів спостереження визначається за формулою

$$n_{\text{дост}} = \frac{t^2 \sigma_y^2}{\Delta y^2 + t^2 \sigma_y^2 / N}. \quad (1.8)$$

У випадках невизначено великої або нескінченої генеральної сукупності об'єктів спостереження вираз (1.8) приймає вигляд

$$n_{\text{дост}} = t^2 \sigma_y^2 / \Delta y^2. \quad (1.9)$$

Середньоквадратичне відхилення σ_y , на даному етапі моделювання невідоме. Його необхідно визначити за даними невеликої пробної вибірки із тієї ж генеральної сукупності. Коефіцієнт довіри приймається рівним $t = 1,96$, чому відповідає ймовірність $P = 0,95$ неперевищення помилки вибірки для середнього арифметичного \bar{y} наперед заданого значення Δy . При бажанні прийняти припущення з вищою або нижчою ймовірністю, треба звернутися до таблиці нормального розподілу випадкової величини і прийняти інше значення t залежно від бажаної ймовірності (додаток 1).

Об'єм вибірки приймається не меншим більшої із двох визначених величин: n_{\min} і $n_{\text{дост}}$.

Розглянемо розрахунок об'єму вибірки на прикладі економетричного моделювання залежності рентабельності витрат від двох факторів: енергоозброєності праці і коефіцієнту постійного складу працюючих на підприємствах певної галузі*. Введемо позначення:

P – рентабельність витрат – відношення прибутку від реалізації продукції до суми витрат на її виробництво і збут, коп./грн.;

E – енергоозброєність праці – відношення сумарної потужності двигунів машин до чисельності працюючих, квт/чол.;

K – коефіцієнт постійного складу працюючих – відношення чисельності працюючих, які пропрацювали на підприємстві весь рік (з 1-го січня до 31-го грудня) до середньоспискової чисельності працюючих за рік, %.

Економетрична модель має вигляд

$$P = f(E, K) + \mathcal{E}.$$

Для визначення об'єму вибірки підприємств зроблена пробна вибірка 15 підприємств з фіксацією рівня рентабельності витрат (табл. 1.1). У цій же таблиці розраховані значення середньої рентабельності і середнього квадрату рентабельності (для 15 підприємств).

Таблиця 1.1 – Розрахунок \bar{P} і $\overline{P^2}$

j	P	P^2	j	P	P^2
1	6,7	44,89	8	11,4	129,96
2	7,3	53,29	9	7,9	62,41
3	11,5	132,25	10	8,9	79,21
4	10,9	118,81	11	9,5	90,25
5	5,4	29,16	12	12,7	161,29
6	9,7	94,09	13	18,9	357,21
7	10,1	102,01	14	9,8	96,04
			15	10,4	108,16
			Сума	151,1	1659,03
			Середнє	10,0733	110,6020

* Цей приклад буде використовуватися на всіх подальших етапах моделювання.

Таким чином, за формулою (1.13) середнє квадратичне відхилення рентабельності витрат складає

$$\sigma_p = \sqrt{110,6020 - 10,0733^2} = 3,0217 \text{ коп.},$$

а достатній об'єм вибірки за (1.9)

$$n_{\text{дост}} = \frac{1,96^2 \times 3,0217^2}{1,1^2} = 28,99 \text{ одиниць},$$

де 1,1 коп – прийнятна помилка вибірки для середнього рівня рентабельності.

Оскільки мінімальний об'єм вибірки за (1.7) складає $(2+1) \times 8 = 24$, приймається об'єм вибірки 30 підприємств.

В таблиці 1.2 наведені значення змінних P , E , і K за вибіркою 30 підприємств, тобто матриця статистичних даних для розробки економетричної моделі, до якої ми будемо неодноразово звертатися.

Таблиця 1.2 – Матриця статистичних даних

Підприємства (j)	Рентабельність, коп./грн. (P)	Енергоозброє- ність праці, квт/чол. (E)	Коефіцієнт постійності ПВП, % (K)
1	6,7	1,5	46
2	7,3	1,7	49
3	11,5	6,4	68
4	10,9	4,7	71
5	5,4	1,6	31
...
26	10,6	6,4	68
27	9,3	3,6	47
28	9,2	4,5	51
29	12,0	7,1	56
30	9,5	5,2	60

1.2.5. Показники варіації змінних

Для подальшого процесу економетричного моделювання необхідно заздалегідь визначити відомі студенту із курсу загальної теорії статистики, або математичної статистики показники варіації змінних. До них належать:

- варіаційний розмах абсолютний

$$R_x = x_{max} - x_{min}, \quad (1.10)$$

- варіаційний розмах відносний

$$i_x = x_{max} / x_{min}, \quad (1.11)$$

- середнє арифметичне

$$\bar{x} = \sum_{j=1}^n x_j / n, \quad (1.12)$$

- дисперсія (середній квадрат відхилення)

$$D_x = \sum_{j=1}^n (x_j - \bar{x})^2 / n,$$

що після нескладних перетворень дає більш прийнятну для розрахунків формулу

$$D_x = \overline{x^2} - \bar{x}^2, \quad (1.13)$$

- середнє квадратичне відхилення

$$\sigma_x = \sqrt{D_x}, \quad (1.14)$$

- коефіцієнт варіації

$$v_x = \sigma_x / \bar{x}. \quad (1.15)$$

Найбільш важливими і затребуваними показниками варіації в економетричному моделюванні є середнє арифметичне значення і середнє квадратичне відхилення змінних.

Розрахунки показників варіації за формулами (1.10)-(1.15) зручно виконувати в таблиці, форма якої з розрахунками для нашого прикладу наведена в таблиці 1.3.

Таблиця 1.3 – Розрахунок показників варіації змінних

j	P	E	K	P^2	E^2	K^2
1	6,7	1,5	46	44,89	2,25	2116
2	7,3	1,7	49	53,29	2,89	2401
3	11,5	6,4	68	132,25	40,96	4624
...
29	12,0	7,1	56			
30	9,5	5,2	60	90,25	27,04	3600
Σ	301,5	145,7	1807	3209,79	837,53	112653
\bar{x}	10,0500	4,8567	60,2333	106,9930	27,9177	3755,1000
\bar{x}^2	101,0025	23,5875	3628,0504			
D_x	5,9905	4,3302	127,0496			
σ_x	2,4475	1,8751	11,2716			
U_x	0,2435	0,3861	0,1871			
R_x	13,5	7,6	48			
i_x	3,5	6,1	2,6			

Як бачимо у нашому прикладі, для вибіркової сукупності 30 підприємств:

- середня рентабельність складає 10,05 коп./грн. з розмахом варіації від 5,4 до 18,9 (коефіцієнт варіації 0,24);
- середня енергоозброєність праці коливається в межах від 1,5 до 9,1 квт/чол. (коефіцієнт варіації 0,386) навколо середнього рівня 4,86 квт/чол.;
- коефіцієнт постійності промислово-виробничого персоналу знаходиться в межах 31÷79% при середньому значенні 60,2% (коефіцієнт варіації 0,19).

1.2.6. Поля кореляції та їх аналіз

Поле кореляції – це точкова діаграма двомірного розсіювання об'єктів спостереження в системі координат y, x_i (рис. 1.1).

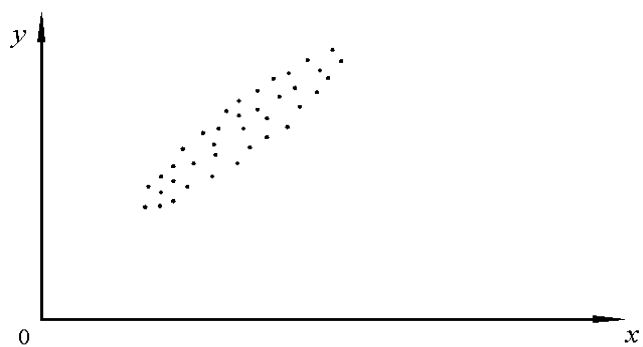


Рис. 1.1 – Поле кореляції y по x_i

Точка в полі – це «адреса», місце кожного j -го об'єкта спостереження з координатами y_i, x_{ij} . Число точок ідентичне об'єму вибірки n . Побудова поля кореляції ведеться за наступними правилами:

- градація осей y і x_i не обов'язково розпочинається з нуля, вона повинна узгоджуватися з абсолютним розмахом варіації: початок і кінець шкал на осях повинен включати відповідно y_{\min} і y_{\max} та $x_{i \min}$ і $x_{i \max}$, порожніх областей на діаграмі не повинно бути;
- співвідношення сторін поля кореляції (довжини шкал на осях) повинне приблизно відповідати співвідношенню коефіцієнтів варіації змінних. Ці вимоги необхідні для бачення розсіювання об'єктів спостереження на діаграмі таким, яким воно є у дійсності.

Поля кореляції відіграють у процесі моделювання велику роль. Їх візуальний аналіз дає підстави для:

- 1) визначення за графічними критеріями наявності, або відсутності кореляційних залежностей y від x_i (рис. 1.2);



Рис. 1.2 – Графічні критерії наявності кореляційної залежності

- 2) попередніх уявлень про математичну форму кореляційних залежностей y від x_i – лінійна, криволінійна, яка саме;
- 3) виявлення аномальних об'єктів спостереження, тобто таких, що не «вписуються» в закономірність розташування поля в системі координат (рис. 1.3);

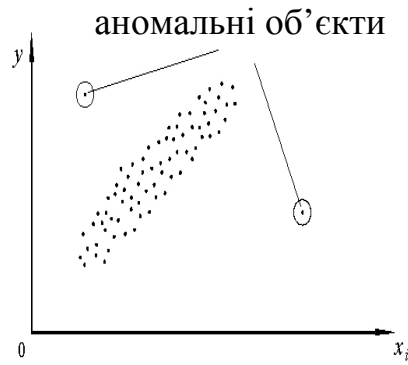


Рис. 1.3 – Аномальні об'єкти спостереження

4) виявлення розривів, розшарування або перехрещування полів кореляції що є ознакою неоднорідності об'єктів спостереження і вимагає розділення вибіркової сукупності принаймні на дві, або більше менших за об'ємом, але однорідних підсукупностей об'єктів (рис. 1.4);



Рис. 1.4 – Неоднорідність об'єктів спостереження

5) визначення типу розподілу змінних відносно середнього арифметичного \bar{x}_i : нормального (симетричного), асиметричного, або рівномірного, що дає додаткові аргументи для судження про однорідність вибіркової сукупності об'єктів спостереження за типом розподілу. Сукупність об'єктів однорідна, якщо виконується умова:

- при нормальному розподілі $v_{x_i} \leq 0,33$;
- при асиметричному розподілі $v_{x_i} \leq 0,40$;
- при рівномірному розподілі $v_{x_i} \leq 0,58$.

Аналіз полів кореляції приводить до наступних висновків:

- 1) на обох полях спостерігається додатна кореляційна залежність рентабельності витрат відповідно від енергоозброєності праці (рис. 1.5) і коефіцієнта постійності промислово-виробничого персоналу (рис. 1.6): чим більше значення E і K , тим вище рівень P , що відповідає теоретичним уявленням щодо сутності цих залежностей;
- 2) щільність поля кореляції P по E вища, ніж поля P по K , що дає підставу вважати, що енергоозброєність сильніше впливає на рентабельність ніж другий фактор;
- 3) аналітична форма залежностей на обох полях очевидно лінійна (принаймні в межах розсіювання змінних);
- 4) підприємство № 13 із-за дуже високої рентабельності ймовірно є аномальним;
- 5) вибіркова сукупність підприємств є кількісно однорідною як за візуальним аналізом полів кореляції, так і за критеріальними значеннями коефіцієнтів варіації.

1.2.7. Вилучення аномальних об'єктів спостереження

На полях кореляції об'єкти спостереження можуть мати ознаки аномальності двох родів:

- 1) занадто велике відхилення від центру двомірного розсіювання з координатами \bar{y} і \bar{x}_i ;
- 2) занадто велике відхилення від лінії регресії, що уявляється на полі кореляції y по x_i .

Для виявлення об'єктів спостереження з ознаками аномальності 1-го роду на поле кореляції y по x_i наноситься прямокутний шаблон двомірного розсіювання з центром в точці \bar{y} , \bar{x}_i і півсторонами, рівними відповідно $\omega_y = t\sigma_y$ і $\omega_{x_i} = t\sigma_{x_i}$. Коефіцієнти довіри t приймаються залежно від кількості об'єктів спостереження n і ймовірності надійності висновку P_ω за спеціальною таблицею. Фрагмент спеціальної таблиці наведений в додатку 2.

Об'єкти спостереження, що знаходяться за межами прямокутного шаблону розсіювання, вважаються аномальними 1-го роду.

У нашому прикладі (табл. 1.2 і рис. 1.5) на полі кореляції P по E півсторони прямокутного шаблону дорівнюють (зауважимо, що $n = 30$; $\sigma_P = 2,45$; $\sigma_E = 1,88$; $t = 2,08$ при $\rho = 0,95$):

$$\text{по } P \longrightarrow \omega_P = 2,08 \times 2,45 = 5,09;$$

$$\text{по } E \longrightarrow \omega_E = 2,08 \times 1,88 = 3,91.$$

Враховуючи, що $\bar{P} = 10,05$, маємо: верх шаблону $10,05 + 5,09 = 15,14$, а низ $10,05 - 5,09 = 4,96$. Враховуючи, що $\bar{E} = 4,86$ маємо: лівий край шаблону $4,86 - 3,91 = 0,95$, а правий $4,86 + 3,91 = 8,77$.

Отже, на полі кореляції P по E аномальним 1-го роду виявилось підприємство № 13. На полі кореляції P по K таким же чином виявлені аномальні підприємства № 5 і № 13 (рис. 1.6).

Для виявлення об'єктів спостереження з ознаками аномальності 2-го роду на поле кореляції наноситься коридор регресії з півстороною $Z_y = t \times \sigma_y \times q$. Віссю коридора слугує діагональ прямокутного шаблону розсіювання, вздовж якої розташовуються точки поля кореляції. В даному разі коефіцієнт довіри t приймається по таблиці нормального розподілу (додаток 1). Для ймовірності $0,95t = 1,96$.

Множник q призначений для врахування ступеню щільності поля кореляції за шкалою:

дуже щільне	– 0,53;
щільне	– 0,72;
середньої щільності	– 0,80;
«пухке»	– 0,92.

Об'єкти спостереження, що знаходяться за межами коридора регресії, вважаються аномальними 2-го роду.

У нашому прикладі півширина коридору щільної регресії P по E дорівнює $Z = 1,96 \times 2,45 \times 0,72 = 3,46$, а середньої щільності регресії P по K – $Z = 1,96 \times$

$\times 2,45 \times 0,80 = 3,84$. На обох полях кореляції аномальним 2-го роду є підприємство № 13.

Для з'ясування питання що до об'єктів спостереження, які мають ознаки аномальності, необхідно скласти зведення у формі таблиці. У нашому прикладі це зведення таке (табл. 1.4):

Таблиця 1.4 – Зведення аномальних об'єктів спостереження

№ аномальних об'єктів	1-го роду		2-го роду	
	<i>E</i>	<i>K</i>	<i>E</i>	<i>K</i>
5	–	+	–	–
13	+	+	+	+

Підприємство № 5 має тільки одну ознаку аномальності із чотирьох можливих, а № 13 – всі чотири.

Об'єкти спостереження, що переважно мають ознаки аномальності (у нас № 13) повинні вилучатися із матриці статистичних даних. При рівності кількості ознак аномальності (+) й нормальності (–) вилучають із матриці ті об'єкти, у яких переважають ознаки аномальності 2-го роду. Всі інші об'єкти залишаються в матриці.

У нашому прикладі вилучається об'єкт № 13, вибіркова сукупність відтепер нараховує 29 підприємств.

Без пояснень зрозуміло, що в разі вилучення аномальних об'єктів спостереження, необхідно скоригувати розраховані раніше показники варіації змінних (табл. 1.2). Наводимо для використання надалі скориговані показники варіації (табл. 1.5).

Таблиця 1.5 – Скориговані показники варіації змінних

<i>j</i>	<i>P</i>	<i>E</i>	<i>K</i>	<i>P</i> ²	<i>E</i> ²	<i>K</i> ²
1						
...						
30						
Σ	282,6	136,6	1740	2852,58	754,72	108164
\bar{x}	9,7448	4,7103	60,0000	98,3648	26,0248	3729,7931
\bar{x}^2	94,9611	22,1869	3600,0000			
D_x	3,4037	3,8379	129,7931			
σ_x	1,8449	1,9591	11,3927			

Зауважимо, що виявлення і видалення аномальних об'єктів спостереження має велике значення для одержання незміщених обґрунтованих і ефективних оцінок параметрів рівняння регресії.

1.3. КОМПЛЕКС КОНТРОЛЬНИХ ЗАВДАНЬ ЗА ЗМ 1 «ПОСТАНОВКА ЗАДАЧІ ЕКОНОМЕТРИЧНОГО МОДЕЛЮВАННЯ»

Навчальні елементи, що підлягають контролю і оцінюванню:

історіографія економетрії, поняття економетрії; предмет економетрії; економетрична модель, її параметри; завдання економетрії; зміст економетричного моделювання; матриця даних; вимоги до розмірів матриці даних; ендогенні і екзогенні змінні; Думмі-змінні; об'єкти спостереження; кількість об'єктів спостереження; показники варіації змінних; поля кореляції; аномальні об'єкти спостереження.

1.3.1. Тестові завдання

Виберіть правильну відповідь (відповіді):

T1.01. Термін «економетрія» вперше ввів у науковий обіг:

- а) К. Маркс; б) Ф. Гальтон; в) П. Чомпа; г) К. Пірсон.

T1.02. Міжнародне економетричне товариство було засноване в:

- а) 1877 р.; б) 1859 р.; в) 1909 р.; г) 1930 р.

T1.03. Ініціаторами заснування Міжнародного економетричного товариства були:

- а) Р. Фріш; б) О. Браве; в) І. Фішер; г) Ч. Русс.

T1.04. Економетрія є:

- а) дисципліною економічної теорії;
б) математичною дисципліною;
в) самостійною дисципліною;
г) поєднанням математики, теорії ймовірностей, математичної статистики і економічної теорії.

T1.05. В економетрії вивчаються такі залежності між економічними показниками:

- а) детерміновані; б) кореляційні; в) стохастичні; г) ймовірнісні.

T1.06. Які рівняння залежності в економетрії записано правильно?

- а) $y = f(x_1, x_2, x_3, \dots)$;
- б) $y = f(x_1, x_2, x_3, \dots, x_m)$;
- в) $y = f(x_1, x_2, x_3, \dots, x_m) + \mathcal{E}$;
- г) $\hat{y} = f(x_1, x_2, x_3, \dots, x_m)$.

T1.07. Предметом економетрії є:

- а) засоби спеціалізації рівнянь регресії;
- б) методи оцінювання параметрів рівнянь регресії;
- в) економічна постановка задачі;
- г) обґрунтування змінних рівнянь регресії.

T1.08. Завданнями економетричного моделювання є:

- а) економічна постановка задачі;
- б) специфікація змінних і аналітичної форми рівнянь регресії;
- в) оцінювання параметрів рівнянь регресії;
- г) точковий і інтервальний прогноз залежних змінних.

T1.09. Змінні в рівняннях регресії класифікуються так:

- а) залежні; б) незалежні; в) ендогенні; г) екзогенні.

T1.10. У рівнянні регресії $\hat{y} = f(x_1, x_2, x_3, \dots, x_m)$ величина \hat{y} є:

- а) приблизною оцінкою y ;
- б) математичним сподіванням y ;
- в) найбільш ймовірним значенням y ;
- г) точним (однозначним) значенням y .

T1.11. У рівнянні регресії $y = \alpha_0 + \alpha_1 x + \mathcal{E}$ параметрами, що підлягають оцінці, є:

- а) y, x ; б) y, \mathcal{E} ; в) α_0, α_1 ; г) $\alpha_0, \alpha_1, \mathcal{E}$.

T1.12. У рівнянні регресії $y = f(x_1, x_2, x_3, \dots, x_m) + \mathcal{E}$ величина \mathcal{E} :

- а) є помилкою оцінки y ;
- б) є випадковою невідомою величиною;
- в) дорівнює нулю;
- г) дорівнює $(y - \hat{y})$.

T1.13. Кому із вчених належить пріоритет введення в науковий обіг поняття «кореляція»?

- а) П. Чомпа; б) Ф. Гальтон; в) К. Пірсон; г) О. Курно.

T1.14. Кому із вчених належить пріоритет введення в науковий обіг поняття «регресія»?

- а) П. Чомпа; б) Ф. Гальтон; в) К. Пірсон; г) О. Курно.

T1.15. Помилки прогнозу за рівнянням регресії обумовлені:

- а) неповним переліком врахованих екзогенних змінних;
- б) недостатнім обсягом вибірки об'єктів спостереження;
- в) помилкою специфікації рівняння регресії;
- г) великою дисперсією ендогенної змінної.

T1.16. Dummy-змінні – це змінні:

- а) штучні; б) непрямі; в) атрибутивні; г) бінарні.

T1.17. Матриця даних для економетричного моделювання – це:

- а) таблиця числових даних про змінні;
- б) матриця розміром $(m + 1)n$;
- в) матриця значень $(m + 1)$ змінних у n об'єктів спостереження;
- г) перелік об'єктів спостереження і змінних.

T1.18. Об'єктами спостереження на макrorівні економіки можуть бути:

- а) національні економіки країн;
- б) регіональні економіки;
- в) економіки галузевих комплексів;
- г) підприємства.

T1.19. Об'єктами спостереження на мікрорівні економіки можуть бути:

- а) національні економіки країн;
- б) регіональні економіки;
- в) економіки галузевих комплексів;
- г) підприємства.

T1.20. Якісна однорідність об'єктів спостереження забезпечується єдністю:

- а) видів економічної діяльності;
- б) сегмента ринку збуту;
- в) форми власності;
- г) щабля в ієрархії економіки.

T1.21. Стовпці матриці даних можуть бути рядами:

- а) варіаційними;
- б) динамічними (часовими);
- в) змішаними;
- г) випадкових величин.

T1.22. Кількість об'єктів спостереження у матриці повинна бути:

- а) не менше $8 \times (m + 1)$;
- б) не менше $t^2 \sigma_y^2 / \Delta y^2$;
- в) рівним N ;
- г) не більше $t^2 \sigma_y^2 / \Delta y^2$.

T1.23. Варіація змінної x характеризується:

- а) \bar{x} ; б) D_x ; в) σ_x ; г) v_x .

T1.24. Поле кореляції y по x це:

- а) точкова діаграма залежності y від x ;
- б) діаграма двомірного розсіювання об'єктів спостереження;
- в) графік залежності y від x ;
- г) графічне зображення матриці даних.

T1.25. Найкращим способом вибірки об'єктів спостереження є:

- а) випадковий; б) механічний; в) типовий (серійний); г) суцільний.

T1.26. Кількісна неоднорідність об'єктів спостереження характеризується такими ознаками полів кореляції:

- а) розривами;
- б) розшаруванням;
- в) наявністю аномалій;
- г) криволінійною конфігурацією.

T1.27. Аномальні об'єкти спостереження:

- а) мають найбільші значення y або x ;
- б) мають найменші значення y або x ;
- в) знаходяться за межами прямокутного шаблону двомірного розсіювання об'єктів;
- г) знаходяться за межами «коридору» регресії.

T1.28. Змінна, що пояснюється, це змінна:

- а) ендогенна; б) екзогенна; в) залежна; г) незалежна.

T1.29. Пояснююча змінна – це змінна:

- а) ендогенна; б) екзогенна; в) залежна; г) незалежна.

T1.30. Побудова і аналіз полів кореляції призначенні для графічного тестування:

- а) наявності кореляційної залежності;
- б) теоретичної гіпотези щодо аналітичної форми залежності;
- в) кількісної однорідності об'єктів спостереження;
- г) наявності аномальних об'єктів спостереження.

1.3.2. Логічні вправи

Л1.01. Поясніть, чому бурхливий розвиток економетрії розпочався саме в 30-х роках минулого століття.

Л1.02. Чи є, на ваш погляд, економетрія математичною дисципліною, дисципліною економічної теорії або самостійною дисципліною? Аргументуйте свою точку зору.

Л1.03. Дайте змістовне, графічне та аналітичне тлумачення кореляції та кореляційної залежності.

Л1.04. Приведіть змістовну, графічну і аналітичну інтерпретацію регресії та рівняння регресії.

Л1.05. Поясніть принципову різницю між функціональною (детермінованою) та кореляційною (стохастичною) залежностями на прикладах.

Л1.06. Доведіть, що результати господарської діяльності є кореляційними функціями умов-факторів господарювання.

Л1.07. Доведіть, що собівартість одиниці продукції є кореляційною функцією кількості виробленої продукції.

Л1.08. Доведіть, що зростання фондоозброєності праці обумовлює зниження витрат виробництва.

Л1.09. Перелічіть внутрішньовиробничі та зовнішні умови-фактори, які впливають на прибуток підприємства. Чи можливо ці залежності апросиміювати детерміновано?

Л1.10. Визначить теоретично, чому дорівнює помилка апроксимації для рівняння площі круга $S = \pi \times r^2$.

Л1.11. Обґрунтуйте шляхи зменшення помилок рівняння регресії.

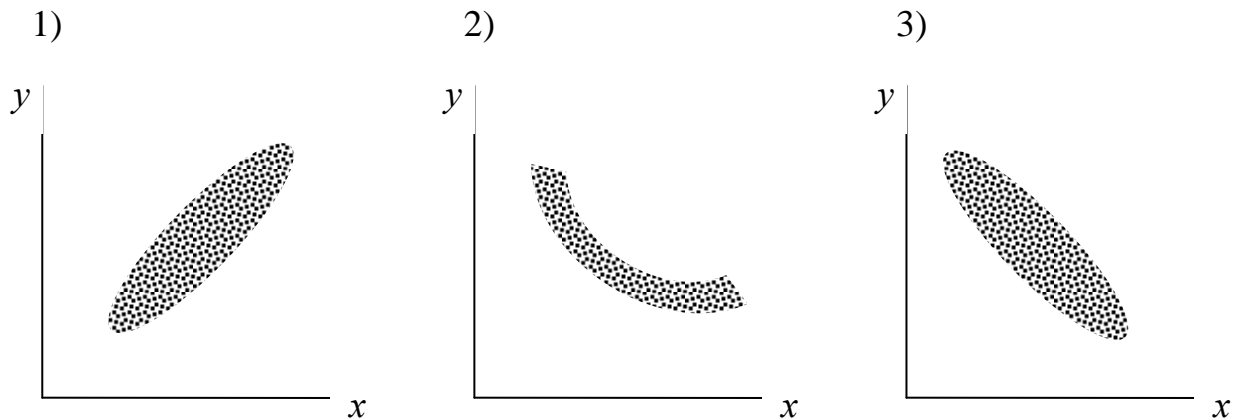
Л1.12. Поясніть, чому економісту необхідно кількісно вимірювати вплив мінливих умов господарювання на результати господарської діяльності.

Л1.13. Визначить різновиди економічної роботи на підприємстві, в яких корисне застосування економетричного моделювання.

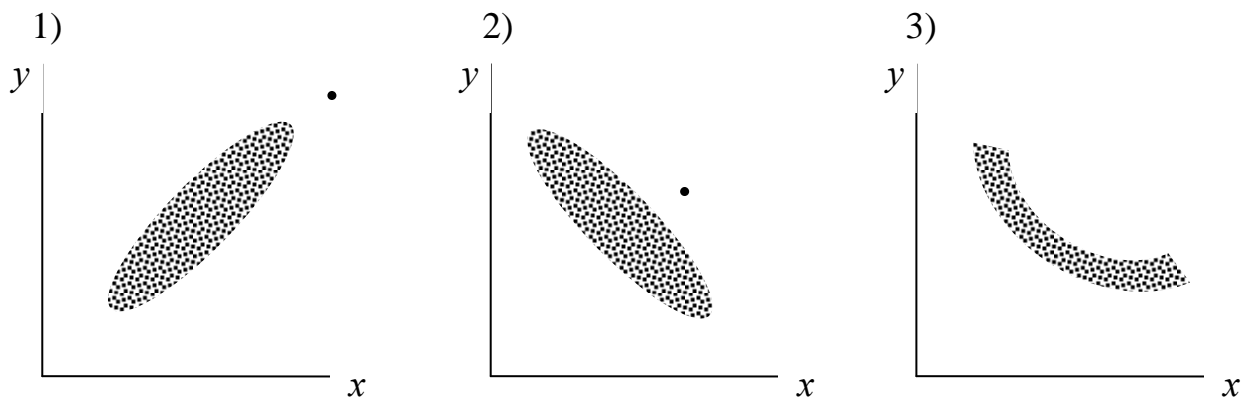
Л1.14. Наведіть перелік і охарактеризуйте показники, що визначають обсяг господарської діяльності підприємства.

Л1.15. Поясніть, чому взаємозв'язки показників господарської діяльності мають стохастичний характер.

Л1.16. На якому з приведених полів кореляції щільність (сила) зв'язку є найбільшою? Поясніть, чому.



Л1.17. На якому з полів кореляції є аномальні об'єкти спостереження 2-го роду? Поясніть, чому.



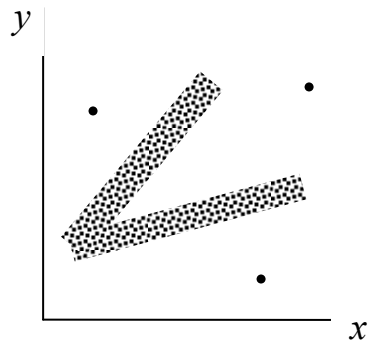
Л1.18. Наведіть приклади операційних характеристик обсягу виробництва продукції на підприємстві.

Л1.19. Наведіть приклади операційних характеристик продуктивності праці, рентабельності виробництва.

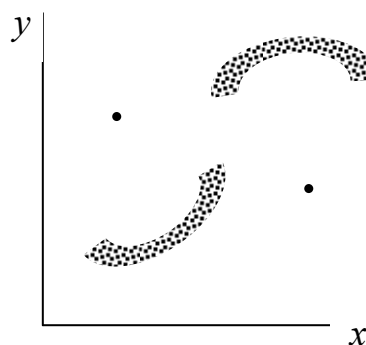
Л1.20. Наведіть операційні характеристики фондоозброєності праці, спеціалізації та кооперування виробництва, кваліфікації та плинності робітників на підприємстві.

Л1.21. Поясніть, чому механічний спосіб відбору об'єктів спостереження кращий за випадковий.

Л1.22. Розгляньте поле кореляції. Зробіть висновки щодо однорідності вибірки об'єктів спостереження та можливості складання рівняння регресії $\hat{y} = f(x)$.



Л1.23. Розгляньте поле кореляції. Зробіть висновки та пропозиції щодо можливості та альтернативи складання рівняння регресії $\hat{y} = f(x)$.



Л1.24. Доведіть, що мінімальний об'єм матриці статистики містить 16 елементів.

Л1.25. Доведіть, що $D_x = \overline{x^2} - \bar{x}^2$.

Л1.26. Поясніть, чому аномальні об'єкти спостереження необхідно вилучати із матриці статистики.

Л1.27. Викладіть джерела аргументації рішень щодо аномальності об'єктів спостереження.

Л1.28. Поясніть, чому в економетричному моделюванні на макрорівні використовують меншу кількість екзогенних змінних, а на мікрорівні – більшу?

Л1.29. Викладіть ваші міркування з приводу застосування Dummy-змінних: чи вони обов'язкові завжди? Чи можна без них обійтись? У чому полягає їх значення?

Л1.30. Побудуйте умовне поле кореляції, на якому були б наявні всі ознаки кількісної неоднорідності об'єктів спостереження.

Л1.31. Визначить, для яких цілей будуються і аналізуються поля кореляції. Наведіть приклади.

Л1.32. Запишіть будь-яке рівняння регресії. Укажіть його параметри, що підлягають оцінюванню в процесі моделювання.

1.3.3. Розрахункові вправи

Р1.01. Визначить мінімальний та достатній об'єм вибірки об'єктів спостереження за умови: $m = 6$; $t = 1,96$; $\sigma_y = 2$; $\Delta y = 0,5$.

Р1.02. Обчисліть, яким повинен бути об'єм вибірки об'єктів спостереження за умови: $m + 1 = 6$; $P = 0,95$; $\sigma_y = 2$; $\Delta y = 0,6$.

Р1.03. Обчисліть дисперсію змінної x за такими даними:

j	x_j
1	2
2	4
3	3
4	5
5	4

Р1.04. Перевірте, чи є кількісно однорідними об'єкти спостереження в наступній матриці статистики.

j	x_j	y_j	j	x_j	y_j
1	13	1,9	6	46	2,9
2	33	3,1	7	54	2,9
3	59	2,6	8	15	2,7
4	19	2,8	9	3	1,2
5	7	1,7	10	50	3,0

P1.05. Зробіть висновки щодо кількісної однорідності об'єктів спостереження в наступній матриці статистики.

j	x_j	y_j	j	x_j	y_j
1	19	1,3	6	48	1,8
2	39	3,7	7	30	1,5
3	60	2,0	8	60	5,0
4	49	4,6	9	8	1,1
5	41	1,7	10	29	2,7

P1.06. Побудуйте поле кореляції і визначить, чи є аномальні об'єкти спостереження у наступній матриці статистики.

j	x_j	y_j	j	x_j	y_j
1	37	6,1	6	13	3,9
2	53	6,5	7	62	6,0
3	22	5,2	8	30	5,9
4	29	5,3	9	44	6,1
5	48	3,8	10	17	4,7

P1.07. Побудуйте в системі координат (y, x) прямокутний шаблон дво-мірного розсіювання за умови: $\bar{y} = 5$; $\bar{x} = 3$; $\sigma_y = 1,5$; $\sigma_x = 1,2$; $t = 1,7$.

P1.08. Побудуйте в системі координат «коридор регресії» за умови: $\bar{y} = 5$; $\bar{x} = 3$; $\omega_y = 2,55$; $\omega_x = 2,05$; $z_y = 2,4$. Примітка: y є додатною залежністю від x .

P1.09. Перевірте, чи є аномальним 1-го роду об'єкт спостереження з координатами $y = 2,3$; $x = 1,5$ за умови: $\bar{y} = 5$; $\bar{x} = 3$; $\omega_y = 2,55$; $\omega_x = 2,05$.

P1.10. Перевірте, чи є аномальним 2-го роду об'єкт спостереження з координатами $y = 6,5$; $x = 4,5$ за умови: $\bar{y} = 5$; $\bar{x} = 3$; $\omega_y = 2,55$; $\omega_x = 2,05$; $z_y = 2,4$. Примітка: y є додатною залежністю від x .

P1.11. Визначить дисперсію і середньоквадратичне відхилення для варіаційного ряду змінної x : 4, 5, 3, 7, 6.

P1.12. Визначить дисперсію і середньоквадратичне відхилення для варіаційного ряду змінної y : 8, 9, 6, 12, 10.

P1.13. Задана матриця статистики. Розрахуйте середньоквадратичні відхилення σ_y та σ_x .

<i>j</i>	<i>y</i>	<i>x</i>
1	8	3
2	9	1
3	7	4
4	4	6
5	2	11

P1.14. За умови вправи P1.13 визначить сигмальним способом наявність залежності *y* від *x* або її відсутність, яка вона: додатна чи від'ємна?

P1.15. За умови вправи P1.13 визначить наявність або відсутність аномальних об'єктів спостереження.

P1.16. За умови вправи P1.13 і враховуючи, що $\sigma_y = 2,608$; $\sigma_x = 3,406$, визначить коефіцієнти варіації v_y та v_x .

P1.17. За умови вправи P1.13 визначить графічним способом наявність або відсутність залежності *y* від *x*, яка вона: додатна чи від'ємна?

P1.18. За умови вправи P1.13 обчисліть середньоквадратичне відхилення і коефіцієнт варіації змінної *x*.

P1.19. Визначить, яким повинно бути поле кореляції *y* по *x*: «стоячим» чи «лежачим»?

	<i>y</i>	<i>x</i>
<i>min</i>	6	2
<i>max</i>	18	14

P1.20. Визначить, яким повинно бути поле кореляції *y* по *x*: «стоячим» чи «лежачим»?

	<i>y</i>	<i>x</i>
<i>min</i>	5	28
<i>max</i>	19	46

P1.21. Перевірте, чи є аномальним 1-го роду об'єкт спостереження з координатами *y* = 6; *x* = 4 за умови: $\bar{y} = 7$; $\bar{x} = 4$; $\sigma_y = 2,5$; $\sigma_x = 3,0$.

P1.22. Задано: $m = 4$; $D_y = 4$; $\Delta y = 0,3$. Визначить, яким повинен бути достатній об'єм вибірки об'єктів спостереження.

P1.23. Задано: $m = 4$. Визначить, яким повинен бути мінімальний об'єм вибірки об'єктів спостереження.

P1.24. Дано: $\sum x = 100$; $\sum x^2 = 1120$; $n = 10$. Розрахуйте дисперсію і середньоквадратичне відхилення змінної x .

P1.25. Дано: $\sum x = 100$; $\sigma_x = 3$; $n = 10$. Розрахуйте коефіцієнт варіації змінної x .

P1.26. Дано: $m = 6$; $\sigma_y = 1,5$; $\Delta y = 0,25$. Визначить мінімальний і достатній об'єм вибірки об'єктів спостереження.

P1.27. Заданий ряд змінної x : 3, 7, 6, 5, 9, 2, 4, 8. Визначить дисперсію і середньоквадратичне відхилення змінної x .

1.3.4. Відповіді до розрахункових вправ

1.01. 56; 62.

1.02. 48.

1.03. 1,04.

1.04. Так, є.

1.05. Вибірка неоднорідна.

1.06. Так, є, № 5.

1.09. Так, є.

1.10. Ні.

1.11. 2,0; 1,414.

1.12. 4,0; 2,0.

1.13. 2,608; 3,406.

1.14. Наявна від'ємна залежність.

1.15. Аномальні об'єкти відсутні.

1.16. 0,43; 0,68.

1.17. Наявна від'ємна залежність.

1.18. 3,406; 0,68.

1.19. «Лежачим».

1.20. «Стоячим».

1.21. Ні.

1.22. 172.

1.23. 40.

1.24. 12; 3,46.

1.25. 0,3.

1.26. 56; 138.

1.27. 5,25; 2,29.

2. ЗМІСТОВИЙ МОДУЛЬ 2: СПЕЦИФІКАЦІЯ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ

2.1. ІДЕНТИФІКАЦІЯ НЕЗАЛЕЖНИХ ЗМІННИХ

2.1.1. Мета і послідовність ідентифікації

Перелік факторів (незалежних, екзогенних змінних $x_i (i = \overline{1, m})$), як нам уже відомо (див. підрозділ 1.2.2), попередньо формується виключно на теоретичному аналізі і професійних знаннях. Потім, коли сформовано матрицю статистичних даних про змінні за певними об'єктами спостереження і певний період часу, з'являється можливість і виникає об'єктивна потреба емпіричної ідентифікації факторів. Мета цієї стадії моделювання полягає в послідовній кількісній оцінці:

- 1) щільності (тісноти) зв'язків, тобто сили залежності y від факторів x_i ;
- 2) суттєвості (невипадковості) залежності y від факторів x_i ;
- 3) мультиколінеарності факторів, тобто щільності зв'язків між x_i і $x_j (i \neq j)$;
- 4) вкладу кожного з факторів x_i у множинний коефіцієнт детермінації d .

Таким чином, на емпіричній стадії проводиться, по суті, тестування факторів з метою визначення такого складу факторів, які: **сильно** впливають на залежну змінну і цей вплив є суттєвим (невипадковим); достатньо **автономно** (самостійно) впливають на залежну змінну; мають **вагомий внесок** у множинний коефіцієнт детермінації, тобто частку в поясненні варіації залежної змінної.

2.1.2. Коефіцієнти парної кореляції і детермінації

Для кількісної оцінки щільності (сили) лінійної залежності між двома змінними y і x_i використовується **коефіцієнт лінійної парної кореляції**, запропонований Ф. Гальтоном у 1877 р. Цей найпростіший критерій для кількісної оцінки зв'язку між двома змінними можна визначити за декількома по суті тотожними формулами. Найбільш зручною (особливо при розрахунках вручну на калькуляторі) є така

$$r_{yx_i} = \frac{\overline{yx_i} - \bar{y} \times \bar{x}_i}{\sigma_y \sigma_{x_i}}. \quad (2.1)$$

Різниця у чисельнику є ознакою наявності кореляційної залежності між y і x_i . Якщо $\overline{yx_i} > \bar{y} \times \bar{x}_i$, маємо пряму, додатну залежність, якщо, навпаки, $\overline{yx_i} < \bar{y} \times \bar{x}_i$ – зворотну, від'ємну. У випадку, якщо середній добуток змінних дорівнює добутку їх середніх арифметичних значень, $\overline{yx_i} = \bar{y} \times \bar{x}_i$, зв'язку між змінними немає.

Доведено, що r_{yx_i} набуває значення від -1 до $+1$, а за абсолютною величиною він завжди знаходиться в межах $0 < |r_{yx_i}| < 1$. Коли коефіцієнт кореляції за абсолютною величиною прямує до 1 , це свідчить про наявність сильного зв'язку. У випадках, коли він прямує до 0 , маємо слабкий зв'язок. Для якісної оцінки сили зв'язку можна користуватися шкалою:

Абсолютне значення r_{yx_i}	Якісна оцінка сили зв'язку
0	Немає
0-0,25	Слабка
0,26-0,5	Помірна
0,51-0,75	Сильна
0,76-1,00	Дуже сильна
1	Абсолютна

Коли $r_{yx_i} = 1$, це свідчить про наявність однозначної (детермінованої, функціональної) залежності y від x_i . Залежна змінна у такому випадку є функцією виключно даного фактора, ніякі інші фактори на неї не впливають.

Зауважимо, що коефіцієнт кореляції є *індикатором тільки на лінійну залежність*. Тому коли $r_{yx_i} = 0$, це не значить, що між змінними взагалі немає зв'язку. На рис. 3.1 наведено випадок, коли коефіцієнт кореляції дорівнює нулю, хоча залежність дуже сильна, але вона криволінійна (параболічна).

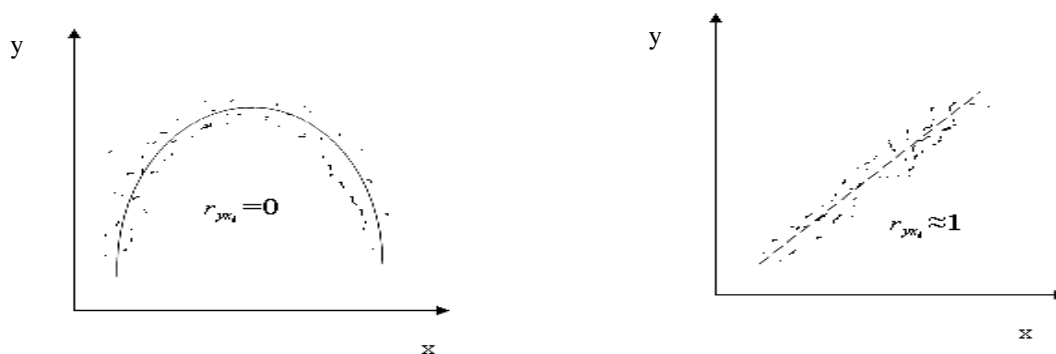


Рис. 2.1 – Залежність коефіцієнта кореляції від форми зв'язку

Коефіцієнт детермінації d_i (його ще позначають R^2) є квадратом коефіцієнта кореляції

$$d_i = R^2 = r_{yx}^2. \quad (2.2)$$

Цей коефіцієнт означає питому вагу дисперсії залежної змінної, яка пояснюється впливом даного фактора, в загальній дисперсії залежної змінної. Цю властивість коефіцієнта детермінації можна довести, але не станемо цього робити.

Повернемося до нашого прикладу і розрахуємо коефіцієнти кореляції між рентабельністю і енергоозброєністю та постійністю ПВП. Попередньо визна-

чимо середні добутки змінних (табл. 2.1) за даними табл. 1.2 (без аномального об'єкта № 13).

Таблиця 2.1 – Розрахунок середніх добутоків змінних

<i>j</i>	<i>PE</i>	<i>PK</i>	<i>EK</i>
1	10,05	308,2	69,0
2	12,41	357,7	83,3
3	73,60	782,0	435,2
...
28	41,40	469,2	229,5
29	85,20	672,0	397,6
30	49,40	570,0	312,0
Сума	1425,26	17413,9	8567,7
Середнє	49,1469	600,4793	295,4379

Тепер за даними табл. 1.5 і 2.1 за формулою (2.1) розрахуємо коефіцієнти кореляції:

$$r_{PE} = \frac{49,1469 - 9,7448 \times 4,7103}{1,8449 \times 1,9591} = 0,8981,$$

$$r_{PK} = \frac{600,4793 - 9,7448 \times 60,0000}{1,8449 \times 11,3927} = 0,7513.$$

Коефіцієнти детермінації за формулою (2.2) складають відповідно $d_E = 0,8981^2 = 0,724$ і $d_K = 0,7513^2 = 0,564$.

Отримані коефіцієнти кореляції і детермінації свідчать, що обидва фактори дуже сильно (*E*) і сильно (*K*) впливають на рентабельність.

Наперед підкреслимо, що для випадку множинної кореляції, коли оцінюють силу впливу на залежну змінну одночасно декількох факторів, формулу (2.2) для визначення коефіцієнта детермінації від одного окремого фактора застосувати не можна.

Зауважимо ще раз, що коефіцієнт лінійної парної кореляції при криво-лінійних формах зв'язку між y і x_i не використовується. До щойно наведених зауважень ми повернемося пізніше.

2.1.3. Тестування суттєвості (невипадковості) коефіцієнтів кореляції

Оскільки коефіцієнт кореляції визначається за вибірковою сукупністю об'єктів спостереження, необхідно оцінювати його значущість і будувати інтервали довіри для істинного коефіцієнта кореляції в генеральній сукупності об'єктів.

Значущість коефіцієнта кореляції перевіряють за допомогою t -статистики, яка має розподіл Стюдента з $(n - 2)$ ступенями свободи. Величину t визначають за формулою

$$t_i = r_{yx_i} \sqrt{\frac{n-2}{1-r_{yx_i}^2}} \quad (2.3)$$

і співставляють з критичним значенням за t -критерієм Стюдента (див. додаток 3). Якщо $t_i > t_{крит}$, то з ймовірністю 0,95 і більше можна стверджувати, що коефіцієнт кореляції значущий (не дорівнює 0) і зв'язок між y і x_i є суттєвим, не випадковим.

Визначимо t -статистики для нашого прикладу:

$$t_E = 0,8981 \sqrt{\frac{29-2}{1-0,8981^2}} = 10,61,$$

$$t_K = 0,7513 \sqrt{\frac{29-2}{1-0,7513^2}} = 5,92.$$

За додатком 3 при $n - m - 1 = 29 - 1 - 1 = 27$ і $P = 0,95$, критичне значення t становить 2,052. Оскільки $10,61 > 2,052$ і $5,92 > 2,052$, робимо висновок, що обидва фактори значущо і достатньо сильно впливають на рентабельність витрат і що цей вплив не випадковий.

2.1.4. Інтервали довіри для коефіцієнтів кореляції

Очевидно, що вибіркового коефіцієнта кореляції як випадкова величина має асиметричний розподіл (особливо при малих вибірках і великих за модулем значеннях). Це неважко зрозуміти, коли уявити, що, наприклад, при $r_{yx_i} = 0,8$ вибіркового коефіцієнта в інших вибірках може перебільшити цю величину тільки на 0,2, тоді як у бік зменшення вони можуть знаходитись від неї на 0,8. Доведено,

що при збільшенні об'єму вибірки ця асиметрія зменшується. Як же побудувати *інтервали довіри* (довірчі границі) для коефіцієнта кореляції?

Зручний підхід для розв'язання цієї задачі запропонував у 1921 р. Р. Фішер. Він сконструював перетворення випадкової величини r у випадкову величину z :

$$z = \frac{1}{2} \ln \frac{1+r}{1-r}, \quad (2.4)$$

яка розподілена майже за нормальним законом (симетрично) з середнім квадратичним відхиленням

$$\sigma_z = \frac{1}{\sqrt{n-3}}.$$

Отже, для випадкової величини z , як і для будь-якої нормально розподіленої величини, гранична помилка вибірки з ймовірністю 0,95 не перевищить

$$\mu_z = \frac{1,96}{\sqrt{n-3}}. \quad (2.5)$$

Знайдемо нижню і верхню межі інтервалу довіри для z , тобто

$$z_n - \frac{1,96}{\sqrt{n-3}} \quad \text{і} \quad z_o + \frac{1,96}{\sqrt{n-3}},$$

а потім і межі довіри для коефіцієнта кореляції. Для цього можна скористатися спеціальною таблицею z -перетворення Р. Фішера (додаток 4).

Побудуємо інтервали довіри для коефіцієнтів кореляції в нашому прикладі. Нагадаємо, що $r_{PE} = 0,8981$; $r_{PK} = 0,7513$; $n = 29$. За формулою (2.4) знаходимо

$$z_E = \frac{1}{2} \ln \frac{1+0,8981}{1-0,8981} = 1,462,$$

$$z_K = \frac{1}{2} \ln \frac{1+0,7513}{1-0,7513} = 0,973$$

і визначаємо за формулою (2.5) нижні й верхні межі інтервалів довіри для z .

нижня	верхня
для $z_E = 1,462 - \frac{1,96}{\sqrt{29-3}} = 1,078;$	$1,462 + \frac{1,96}{\sqrt{29-3}} = 1,847,$
для $z_K = 0,973 - \frac{1,96}{\sqrt{29-3}} = 0,589;$	$0,973 + \frac{1,96}{\sqrt{29-3}} = 1,357.$

Тепер за таблицею z -перетворення Р. Фішера (додаток 4) по z_i знаходимо довірчі межі для теоретичних коефіцієнтів кореляції (ρ_{pi}):

$$0,79 \leq \rho_E \leq 0,95,$$

$$0,53 \leq \rho_K \leq 0,88.$$

Звернемо увагу, що одержані межі (0,79-0,95) і (0,53-0,88) не знаходяться на однаковій відстані від вибірових значень коефіцієнтів кореляції 0,90 і 0,75, що свідчить про асиметричний розподіл вибірових коефіцієнтів кореляції.

На закінчення розгляду питання щодо оцінки значущості й інтервалів довіри для коефіцієнтів кореляції додатково зауважимо наступне:

- 1) якщо коефіцієнт кореляції від'ємний, то в розрахунках t_i (див. формулу (2.3)) ніяких особливостей або ускладнень не виникає, але в розрахунках z_i (формула 2.4) коефіцієнт кореляції треба приймати за модулем;
- 2) якщо коефіцієнт кореляції має невелике значення, наприклад, 0,25, тоді за формулою (2.4) $z = 0,255$, $\mu_z = 0,384$ за формулою (2.5) і нижня границя інтервалу довіри для z набуває від'ємного значення: $-0,129 \times (0,255 - 0,384)$. Це означає, що і інтервал довіри для r також включає значення 0. Така незалежна змінна, безумовно, є незначущою і її слід вилучити з подальшого процесу моделювання.

2.1.5. Мультиколінеарність

Термін «*мультиколінеарність*» вперше ввів Р. Фріш (1934 р.). За Фрішем мультиколінеарність означає, що в багатofакторній регресійній моделі дві або більше, навіть всі незалежні змінні (фактори) зв'язані між собою і мають певний ступінь кореляції ($r_{ij} \neq 0, i \neq j$). Це наслідок глобальної тенденції одночасної

зміни економічних показників. В економіці взагалі важко уявити відсутність кореляції, її наявність – об'єктивна.

У нашому прикладі кореляція між енергоозброєністю праці і коефіцієнтом постійності *ПВП* складає (табл. 1.5 і 2.1):

$$r_{EK} = \frac{295,4379 - 4,7103 \times 60,0000}{1,9591 \times 1139,3927} = 0,5744,$$

що свідчить про більш ніж помірну колінеарність цих двох факторів.

Чому наявність мультиколінеарності є проблемою, які негативні наслідки в економетричному моделюванні вона породжує? Відповідь проста: мультиколінеарність не дозволяє обґрунтовано оцінити параметри рівняння регресії, помилки цих оцінок великі. При наявності сильної мультиколінеарності ($r_{ij} \rightarrow 1$, $i \neq j$) може статися, що не тільки абсолютні величини коефіцієнтів регресії оцінені неправильно, а і їх знаки.

Тестування наявності і визначення сили мультиколінеарності не має єдиного підходу. Тому наведемо кілька простих методів тестування.

1. Мультиколінеарність наявна, якщо сума парних коефіцієнтів детермінації перевищує 1, тобто

$$d_1 + d_2 + \dots + d_m > 1.$$

У нашому прикладі саме це і спостерігається $0,8981^2 + 0,7513^2 = 0,81 + 0,56 = 1,37 > 1$. Це означає, що оцінки d_1 і d_2 завищені саме через колінеарність цих факторів.

Чим більше $\sum d_i$ перевищує одиницю, тим більша сила мультиколінеарності, тому що за відсутності мультиколінеарності

$$d_1 + d_2 + \dots + d_m < 1.$$

2. Поширеним тестом на наявність мультиколінеарності є великі величини коефіцієнтів кореляції між окремими парами змінних. Якщо значення хоча б одного коефіцієнта кореляції більше 0,8, то мультиколінеарність є серйозною проблемою. Але недоліком цього тесту є те, що високе значення парних коефі-

цієнтів кореляції – необхідна, але не достатня умова наявності мультиколінеарності.

3. Для визначення наявності й сили (рівня) колінеарності можна застосувати порівняння коефіцієнтів взаємної кореляції (r_{ij}) з коефіцієнтами кореляції залежної змінної з факторами (r_{yx_i}):

матриця коефіцієнтів взаємної кореляції r_{ij} ($i \neq j$)

вектор коефіцієнтів кореляції r_{yx_i}

$$\begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1m} \\ r_{21} & 1 & r_{23} & \dots & r_{2m} \\ r_{31} & r_{32} & 1 & \dots & r_{3m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & r_{m3} & \dots & 1 \end{bmatrix}$$

$$\begin{bmatrix} r_{yx_1} \\ r_{yx_2} \\ r_{yx_3} \\ \dots \\ r_{yx_m} \end{bmatrix}$$

Тести такі:

- якщо $r_{ij} = 0$, мультиколінеарність наявна;
- якщо всі $r_{ij} < r_{yx_i \min}$, мультиколінеарність є, але вона слабка;
- якщо всі r_{ij} знаходяться в межах від $r_{yx_i \min}$ до $r_{yx_i \max}$, мультиколінеарність помірною;
- якщо хоча б один $r_{ij} > r_{yx_i \max}$, мультиколінеарність сильна.

У нашому прикладі з двома незалежними змінними це порівняння полягає в співставленні двох матриць

$$\begin{bmatrix} 1 & 0,5744 \\ 0,5744 & 1 \end{bmatrix} \quad \begin{bmatrix} 0,8981 \\ 0,7513 \end{bmatrix}.$$

Мультиколінеарність, як бачимо, наявна, оскільки $r_{12} = r_{21} = 0,5744 \neq 0$. Але її сила або рівень невисокий, слабкий, тому що $0,5774 < 0,8981$ і $0,7513$.

Що робити, коли мультиколінеарність виявлено? Безпомилкових порад немає, оскільки вона є прикладною проблемою. Можна запропонувати декілька простих методів видалення явища мультиколінеарності, принаймні зниження її рівня:

- 1) перетворення незалежних змінних;
- 2) вилучення незалежних змінних;
- 3) тестування незалежних змінних на автономність.

Перетворення незалежних змінних полягає, наприклад, в заміні якоїсь із сильно корельованих змінних x_i на $\lg x_i$, заміні абсолютних значень x_i відносними величинами, заміні операційної характеристики змінної x_i іншою і т.п. Це може привести до послаблення мультиколеніарності факторів.

Вилучення незалежних змінних як засіб вилучення або послаблення мультиколеніарності полягає в тому, що з двох змінних з високою коленіарністю ($r_{ij} \rightarrow 1$) вилучають ту, яка має менший коефіцієнт кореляції із залежною змінною (r_{yx_i}). Але вилучення незалежної змінної з моделі може призводити до помилки ідентифікації. Так, за економічною теорією, для моделювання залежності споживання необхідно включати фактори доходу і багатства, але дохід і багатство є сильно корельованими факторами. Тоді вилучення будь-якої з них створюватиме помилку ідентифікації моделі. У таких випадках треба шукати інший спосіб зниження рівня мультиколінеарності.

Сутність методу **тестування незалежних змінних на автономність** розглянемо в наступному підрозділі.

2.1.6. Бета-коефіцієнти

В економетрії для багатьох цілей використовують β -коефіцієнти. Вони є індикаторами зміни залежної змінної за рахунок впливу кожної окремої незалежної змінної за умови закріплення всіх інших незалежних змінних на середньому рівні.

Ці коефіцієнти визначають, на скільки σ_y змінюється залежна змінна y при зміні незалежної змінної x_i на одне σ_{x_i} . β -коефіцієнти розраховують на основі **матриці коефіцієнтів взаємної кореляції** x_{ij} і коефіцієнтів кореляції r_{yx_i} за формулою

$$\beta_i = \frac{\Delta_i}{\Delta_0}, \quad (2.6)$$

де Δ_0 – визначник (детермінант) матриці коефіцієнтів взаємної кореляції r_{ij} ;

Δ_i – визначник тієї ж матриці із заміною в ній i -го стовпчика стовпчиком r_{yx_i} .

У нашому прикладі з двома незалежними змінними вихідна інформація для тестування – це дві матриці

<i>матриця r_{ij}</i>	<i>матриця r_{yx_i}</i>
$\begin{bmatrix} 1 & 0,5744 \\ 0,5744 & 1 \end{bmatrix}$	$\begin{bmatrix} 0,8981 \\ 0,7513 \end{bmatrix}$

За формулою (2.6) β_E дорівнює

$$\beta_E = \frac{\begin{vmatrix} 0,8981 & 0,5744 \\ 0,7513 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 0,5744 \\ 0,5744 & 1 \end{vmatrix}} = \frac{0,8981 \times 1 - 0,7513 \times 0,5744}{1 - 0,5744^2} = \frac{0,4666}{0,6701} = 0,6963,$$

β_K дорівнює

$$\beta_K = \frac{\begin{vmatrix} 1 & 0,8981 \\ 0,5744 & 0,7513 \end{vmatrix}}{\begin{vmatrix} 1 & 0,5744 \\ 0,5744 & 1 \end{vmatrix}} = \frac{0,7513 - 0,5744 \times 0,8981}{1 - 0,5744^2} = \frac{0,2354}{0,6701} = 0,3513.$$

Величина β_E означає, що внаслідок зміни енергоозброєності праці на $\sigma_E = 1,9591$ кВт/чол. (див. табл. 1.5) рентабельність витрат змінюється на $\beta_E \sigma_P = 0,6963 \times 1,8449 = 1,2846$ коп./грн., а зміна коефіцієнта постійності ПВП на $\sigma_K = 11,3927\%$ обумовлює зміну рентабельності витрат на $\beta_K \sigma_P = 0,3513 \times 1,8449 = 0,6481$ коп./грн.

β -коефіцієнти завдяки таким властивостям відіграють в економетричному моделюванні значну роль. Передусім їх використовують для тестування неза-

лежних змінних на наявність мультиколінеарності факторів. Як і коефіцієнти кореляції r_{yx_i} , β_i -коефіцієнти знаходяться в межах від -1 до $+1$. Тому їх можна визначати як кількісну міру сили зв'язку між y і x_i за умови виключення впливу явища мультиколінеарності.

2.1.7. Тестування автономії екзогенних змінних

За модулем у коректній моделі значення β_i завжди менше значення r_{yx_i} . Це означає, що за наявності мультиколінеарності r_{yx_i} є завищеною мірою сили зв'язку. При відсутності мультиколінеарності $\beta_i = r_{yx_i}$, що можна легко показати. Нехай $y = f(x_1, x_2)$ і встановлено, що $r_{12_i} = 0$, тобто x_1, x_2 не корелюють. Визначимо β_1, β_2 :

$$\beta_1 = \frac{\begin{vmatrix} r_{yx_1} & 0 \\ r_{yx_2} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}} = \frac{r_{yx_1} \times 1 - r_{yx_2} \times 0}{1 - 0^2} = r_{yx_1},$$

$$\beta_2 = \frac{\begin{vmatrix} 1 & r_{yx_1} \\ 0 & r_{yx_2} \end{vmatrix}}{\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}} = \frac{1 \times r_{yx_2} - 0 \times r_{yx_1}}{1 - 0^2} = r_{yx_2}.$$

Як бачимо, при відсутності мультиколінеарності $\beta_i = r_{yx_i}$. З прикладу випливає, що якщо $r_{ij} = 0$, чисельник у розрахунку зменшується в більшій мірі, ніж знаменник, що і обумовлює зменшення β_i порівняно з r_{yx_i} .

Саме на цій властивості β -коефіцієнтів побудований γ -тест на наявність мультиколінеарності:

$$\gamma_i = \frac{\beta_i}{r_{yx_i}}, \quad (0 < \gamma_i < 1). \quad (2.7)$$

Чим сильніше x_i корелює з іншими факторами, тим γ -відношення менше і навпаки. Отже, γ_i -відношення є своєрідним показником *автономності впливу*

факторів на залежну змінну в умовах мультиколінеарності. Це відношення за умови слабкої і помірної мультиколінеарності знаходиться в межах $0 \div 1$.

У нашому прикладі

$$\gamma_E = \frac{0,6963}{0,8981} = 0,775, \quad \gamma_K = \frac{0,3513}{0,7513} = 0,468,$$

що свідчить про певну автономність впливу цих факторів на рентабельність і правомірність їх включення до рівняння регресії.

При сильній мультиколінеарності умова (2.7) порушується. Припустимо, що $r_{yx1} = 0,5$, $r_{yx2} = 0,7$ і $r_{12i} = 0,8$. Тоді

$$\beta_1 = \frac{\begin{vmatrix} 0,5 & 0,8 \\ 0,7 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 0,8 \\ 0,8 & 1 \end{vmatrix}} = -0,17, \quad \beta_2 = \frac{\begin{vmatrix} 1 & 0,5 \\ 0,8 & 0,7 \end{vmatrix}}{\begin{vmatrix} 1 & 0,8 \\ 0,8 & 1 \end{vmatrix}} = 0,78,$$

відповідно,

$$\gamma_1 = \frac{-0,17}{0,5} = -0,34, \quad \gamma_2 = \frac{0,78}{0,7} = 1,11.$$

Від'ємний знак γ_i свідчить про відсутність будь-якої автономності впливу і самостійного значення фактора x_i . Такий фактор через його сильну корельованість з іншими повинен *безумовно* вилучатися із подальшого процесу моделювання. Отже в наведеному вище умовному прикладі x_1 не можна вводити в рівняння регресії. Якщо цього не зробити, то в рівнянні регресії коефіцієнт регресії α_1 матиме знак, протилежний уяві і економічній теорії.

Використання властивостей β -коефіцієнтів для інших цілей розглядаємо далі.

2.1.8. Коефіцієнти множинної кореляції і детермінації

Сила впливу на залежну змінну однієї окремо взятої незалежної змінної, як уже відомо, визначається коефіцієнтом парної кореляції. А як оцінити силу одночасного впливу на неї множини факторів x_1, x_2, \dots, x_m разом узятих та ще

і при наявності мультиколінеарності? Для цього використовують **коефіцієнт множинної кореляції**, який можна визначити за формулою Боярського:

$$R_{y x_1 \dots x_m} = \sqrt{\frac{|\Delta_*|}{\Delta_0}} = \sqrt{\frac{\begin{vmatrix} 1 & r_{12} & r_{13} & \dots & r_{1m} & r_{y x_1} \\ r_{21} & 1 & r_{23} & \dots & r_{2m} & r_{y x_2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & r_{m3} & \dots & 1 & r_{y x_m} \\ r_{y x_1} & r_{y x_2} & r_{y x_3} & \dots & r_{y x_m} & 0 \end{vmatrix}}{\begin{vmatrix} 1 & r_{12} & r_{13} & \dots & r_{1m} \\ r_{21} & 1 & r_{23} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & r_{m3} & \dots & 1 \end{vmatrix}}}. \quad (2.8)$$

У чисельнику (2.8) – визначник повної матриці коефіцієнтів кореляції із заміною одиниці у правому нижньому куті на нуль, а у знаменнику – визначник матриці коефіцієнтів взаємної кореляції. Визначник у чисельнику приймають в розрахунок за модулем.

Розрахуємо коефіцієнт множинної кореляції для нашого прикладу

$$R_{PEK} = \sqrt{\frac{\begin{vmatrix} 1 & 0,5744 & 0,8981 \\ 0,5744 & 1 & 0,7513 \\ 0,8981 & 0,7513 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & 0,5744 \\ 0,5744 & 1 \end{vmatrix}}} = \sqrt{\frac{0,5947}{0,6701}} = 0,9434,$$

а також множинний коефіцієнт детермінації

$$d_{EK} = R_{PEK}^2 = 0,9434^2 = 0,8892.$$

Останній показує, що 88,92% варіації рентабельності пояснюється впливом двох факторів разом узятих, а 11,08% варіації, яку називають залишковою, є середньою помилкою апроксимації, притаманною рівнянню регресії, що розробляється.

Але ж раніше (підрозділ 2.1.2) ми визначали, що фактор E пояснює 81%, а фактор K – 56%, або 137% разом! Це ще раз свідчить про наслідки колінеарності факторів.

Як же розділити 88,92% між факторами? Доведено, що у множинній лінійній кореляції при наявності мультиколінеарності **коефіцієнт частинної парної детермінації** становить не $r_{yx_i}^2$, а

$$d_i = \beta_i r_{yx_i}. \quad (2.9)$$

У нашому прикладі маємо

$$d_E = 0,6963 \times 0,8981 = 0,6253; 62,5\%,$$

$$d_K = 0,3513 \times 0,7513 = 0,2639; 26,4\%,$$

або 88,9% разом. Таким чином, різний рівень рентабельності на 29 підприємствах на 62,5% пояснюється варіацією енергоозброєності праці і на 26,4% – варіацією коефіцієнта постійності складу *ПВП*.

Розглянута властивість β -коефіцієнтів дозволяє набагато простіше розрахувати коефіцієнт множинної кореляції. Замість формули (2.8) можна скористатися

$$R_{yx_1 \dots m} = \sqrt{\sum_{i=1}^m \beta_i r_{yx_i}}. \quad (2.10)$$

У нашому прикладі, як і очікувалось, за формулою (2.10) отримаємо вже відому міру сили множинної кореляції

$$R_{PEK} = \sqrt{0,6963 \times 0,8981 + 0,3513 \times 0,7513} = \sqrt{0,8892} = 0,9434.$$

2.1.9. Тестування значущості вкладу факторів до множинної детермінації

Абсолютна величина **вкладу будь-якого фактора у множинну детермінацію** визначається коефіцієнтом частинної детермінації (2.9), тобто

$$d_i = \beta_i r_{yx_i},$$

бо $d = \beta_1 r_{yx_1} + \beta_2 r_{yx_2} + \dots + \beta_m r_{yx_m} < 1$.

Отже, мова йде про те, чи достатньо значущий доданок d_i у сумарній величині d . Для з'ясування цього використовуємо ρ -тест:

$$\rho_i = 0,742 \times \sqrt{n} \times \sqrt{R_{yx_1 \dots m}^2 - R_{yx_1 \dots m(i)}^2}, \quad (2.11)$$

де $R_{yx_1 \dots m(i)}$ – коефіцієнт множинної кореляції за умови виключення i -го фактора. Застосування ρ -тесту доволі трудомістке, адже треба розрахувати окрім $R_{yx_1 \dots m}$ ще m коефіцієнтів множинної кореляції $R_{yx_1 \dots m(i)}$. Наприклад, якщо факторів чотири, треба визначити п'ять коефіцієнтів множинної кореляції: $R_{y_{1234}}$, $R_{y_{234}}$, $R_{y_{134}}$, $R_{y_{124}}$, $R_{y_{123}}$. Зауважимо, що для розрахунку будь-якого з них необхідно в кожному разі заново розрахувати значення β_j .

Тестування факторів за ρ -тестом ґрунтується саме на порівнянні значень двох коефіцієнтів множинної детермінації: 1) за всіма m факторами і 2) за $m-1$ факторами, тобто без фактора, що тестується. Якщо розрахункове значення ρ_i значуще, то з певною ймовірністю маємо $\rho_i \geq \rho_{крит.}$. Для ймовірності 0,95 цей тест відповідає умові $\rho_i \geq 1,96$.

У нашому прикладі значення ρ_i такі:

$$\rho_E = 0,742 \sqrt{29} \sqrt{0,9434^2 - 0,7513^2} = 2,277,$$

$$\rho_K = 0,742 \sqrt{29} \sqrt{0,9434^2 - 0,8981^2} = 1,148.$$

Вклад фактора енергоозброєності праці в множинну детермінацію рентабельності значущий і з ймовірністю більше 0,95 повинен вводитися до рівняння регресії ($2,277 > 1,96$). Про вклад фактора постійності *ПВП* з такою впевненістю стверджувати не можна, адже $1,148 < 1,96$. Але це не значить, що він взагалі незначущий. Звернемося до таблиці нормального розподілу (додаток 1) і визначимо ймовірність, що відповідає коефіцієнту довіри $t = 1,148$. Вона становить 0,75. Отже, значущість вкладу цього фактора можна стверджувати з ймовірністю 0,75, що не так уже й сумнівно.

2.1.10. Вилучення екзогенних змінних

Рішення щодо включення незалежних змінних (факторів) у рівняння регресії приймається за наслідками тестування на значущість (невипадковість) їхнього впливу на залежну змінну за t-статистикою Стьюдента, автономність впливу за γ -тестом і достатність вкладу у множинний коефіцієнт детермінації за ρ -тестом. Для зручності визначення складають зведення результатів тестування. Для нашого прикладу таке зведення надане в табл. 2.2.

Таблиця 2.2 – Зведення результатів тестування факторів

Фактори	t-статистика Стьюдента ($t_i \geq 2,052$)	γ -тест ($0 \leq \gamma_i \leq 1$)	ρ -тест ($\rho_i \geq 1,96$)
<i>E</i>	10,61	0,775	2,277
<i>K</i>	5,92	0,468	1,148

Фактор енергоозброєності праці (*E*) за всіма тестами безумовно включається до рівняння регресії. Фактор постійності ПВП (*K*) помітно слабкіший, але з деяким сумнівом відносно його вагомості у множинному коефіцієнті детермінації також можна ввести в рівняння регресії.

Такі рішення, як правило, остаточні, хоча на стадії оцінювання параметрів рівняння регресії іноді трапляються випадки вилучення факторів.

2.2. СПЕЦИФІКАЦІЯ АНАЛІТИЧНОЇ ФОРМИ РІВНЯНЬ РЕГРЕСІЇ

2.2.1. Мета і способи специфікації

Парні й багатофакторні лінійні економетричні моделі, які часто використовуються в економіці, не завжди відображають дійсні закономірності розвитку і зв'язків між соціально-економічними явищами і процесами. Тому на практиці для побудови адекватних економетричних моделей застосовують нелінійну форму рівнянь регресії.

Нелінійні форми рівнянь регресії використовують з метою, по-перше, *максимально апроксимувати сутність явища за економічною теорією* і, по-

друге, *максимально зменшити помилку оцінювання параметрів і прогнозу залежної змінної.*

Наведемо приклади апріорі криволінійних форм регресії. З ростом доходу, наприклад, сім'ї обсяг її споживання зростає. За економічною теорією це зростання необмежене, але воно має монотонно уповільнений характер. Отже, зростання споживання описується кривою лінією зростання з від'ємним прискоренням. Другий приклад: собівартість одиниці продукції за рахунок збільшення обсягу її виробництва знижується, але за теорією це зниження не може бути лінійним. Воно має обмеження – величину постійних витрат на одиницю продукції. Отже, ця залежність криволінійна. Очевидно, в економетричних моделях теоретично вірні уяви щодо форми залежностей повинні лежати в основі обґрунтування форми регресії.

Застосування криволінійної форми регресії, має і суто прагматичну мету – одержання ефективних оцінок параметрів рівняння регресії, тобто **зменшення залишкової дисперсії**, що не пояснюється рівнянням регресії. На рис. 2.2 наведене поле кореляції і дві форми рівняння регресії: 1) лінійна ($\hat{y} = a_0 + a_1x$) і 2) параболічна ($\hat{y} = a_0 + a_1x + a_2x^2$). Залишки (відхилення точок поля від лінії регресії) при лінійній формі набагато більші, ніж при параболічній. Відповідно і оцінка параметрів a_0 і a_i у параболічній формі рівняння регресії набагато ефективніша, прогноз залежної змінної в даному разі набагато точніший.

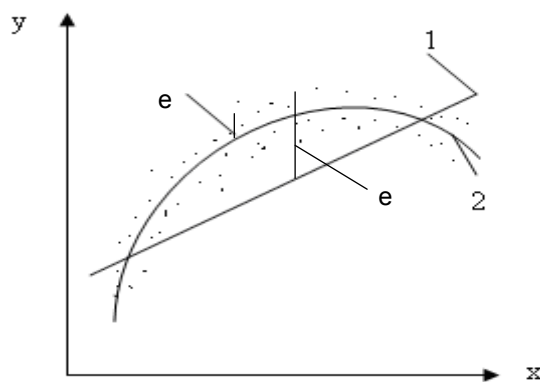


Рис. 2.2 – Лінійна і параболічна форми регресії

Способи специфікації (обґрунтування) форми рівняння регресії такі : теоретичний, графічний, аналітичний і сигмальний.

Теоретичний спосіб обґрунтування форми рівняння регресії базується на економіко-теоритичному аналізі сутності явища, що моделюється, із застосуванням професійних знань. Цей спосіб безпомилковий, найкращий, але він вимагає від аналітика високої професійної підготовки і розвинутого економічного мислення.

Графічний спосіб специфікації заснований на візуальному аналізі полів кореляції. Цей спосіб доступний для будь-кого: розглядай поля кореляції і підбирай найбільш адекватну форму рівняння регресії «на око». Дуже просто, але не завжди ефективно. Тому графічний спосіб можна рекомендувати як допоміжний до теоретичного.

Аналітичні способи багаточисельні. Один з них (спосіб перших різниць) ми розглянемо далі.

Сигмальний спосіб полягає в тому, що форма рівняння регресії взагалі спеціально не обґрунтовується, а вибирається з декількох варіантів. Це ефективний спосіб, але він трудомісткий. Приймаються кілька варіантів форми рівняння регресії, для кожного з них оцінюються параметри регресії і розраховується залишкова дисперсія. Той варіант специфікації, який забезпечує мінімальну дисперсію залишків ($\sum_1^n e^2$), вибирається як остаточний.

2.2.2. Аналітичні форми рівнянь регресії

В економетричному моделюванні часто і обґрунтовано застосовують *лінійні* форми рівнянь регресії як парні (прості),

$$\hat{y} = a_0 + a_1x, \quad (2.12)$$

так і багатофакторні

$$\hat{y} = a_0 + a_1x + a_2x_2 + \dots + a_mx_m. \quad (2.13)$$

У цих рівняннях, як ми вже знаємо, параметр a_0 означає величину y , яка не залежить від варіації незалежних змінних x_i і обумовлюється іншими факторами.

Його називають у парній регресії перетином. Параметри a_i визначають зміну y при зміні x_i на одну одиницю. У парній регресії його називають нахилом (рис. 2.3).

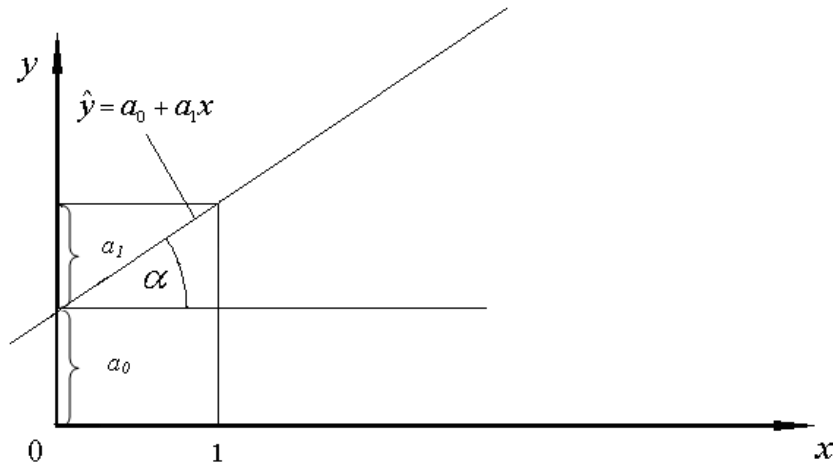


Рис. 2.3 – Параметри парної лінійної регресії

Дійсно, a_0 – це відрізок до точки перетину лінії регресії з віссю y , а a_1 – тангенс кута нахилу лінії регресії до осі x . У множинній регресії такі назви параметрів (перетин, нахил) не прийнятні, бо в даному разі ми маємо не площину, а багатовимірний простір.

Розрізняють два види (класи) нелінійних регресій. До першого належать регресії, що нелінійні за факторами x_i , але лінійні відносно параметрів a_i , які підлягають оцінюванню. Такі регресії називаються **квазілінійними**. До таких регресій відносяться, наприклад, параболічні другого порядку (квадратичні)

$$\hat{y} = a_0 + a_1x + a_2x^2 \quad (2.14)$$

і гіперболічні (зворотні)

$$\hat{y} = a_0 + \frac{a_1}{x} = a_0 + a_1 \frac{1}{x}. \quad (2.15)$$

За допомогою парної регресії квадратичної форми можна апроксимувати прискорене або уповільнене зростання, а також прискорене або уповільнене падіння y при збільшенні x . Парабола другого порядку має вершину, тобто **точку насичення** зростання або падіння, після якої тенденція зміни y змінюється

на симетрично протилежну. Парабола третього порядку має *S*-подібну форму з двома точками насичення.

Парна гіперболічна регресія придатна для апроксиміювання уповільненого падіння y при зростанні x і має **лінію насичення** $y = a_0$.

Другий клас нелінійних регресій складають **суттєво нелінійні** регресії, що характеризуються нелінійністю за своїми параметрами. До них належать часто використовувані степеневі

$$\hat{y} = a_0 x^{a_1}; \quad \hat{y} = a_0 x^{a_1} x_2^{a_2} \dots x_m^{a_m} \quad (2.16)$$

і показникова форми

$$\hat{y} = a_0 a_1^x; \quad \hat{y} = a_0 a_1^{x_1} a_2^{x_2} \dots a_m^{x_m}. \quad (2.17)$$

Парні степеневі й показникові форми регресії придатні для апроксиміювання прискореного або уповільненого зростання, а також прискореного або уповільненого падіння залежної змінної, але ці **тенденції не обмежені**, як у параболи другого порядку, насичення відсутнє.

В окремих випадках застосовують більш складні форми суттєво нелінійної регресії:

- модифікована експонента

$$\hat{y} = a_0 + a_1 a_2^x, \quad (2.18)$$

- крива Гомперця

$$\hat{y} = e^{a_0 + a_1 a_2^x}, \quad (2.19)$$

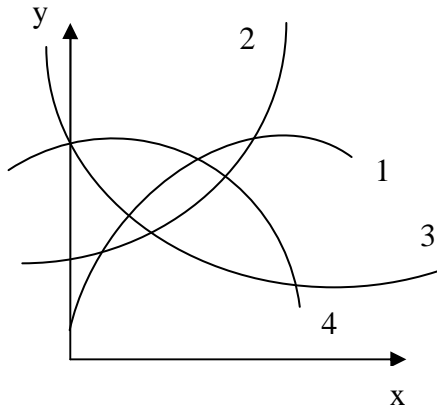
- логістична крива (крива Перла-Ріда)

$$\hat{y} = 1/(a_0 + a_1 a_2^x). \quad (2.20)$$

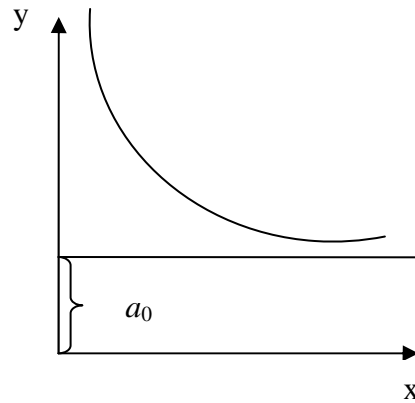
У цих формах a_0 , a_1 і a_2 – невідомі параметри рівняння регресії, що підлягають оцінюванню.

На рис. 2.4 наведені приклади графіків криволінійних форм парної регресії (для $y > 0$; $x > 0$).

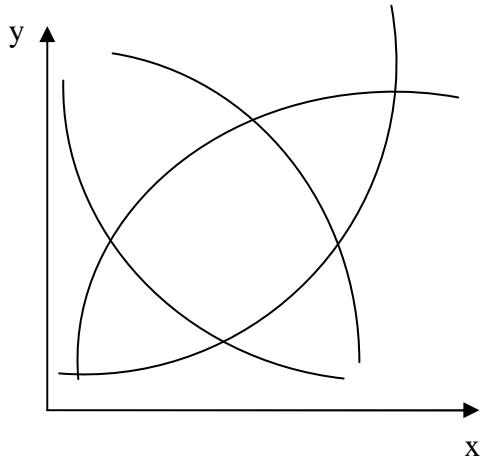
А) квадратична
 $\hat{y} = 1/(a_0 + a_1 a_2^x)$



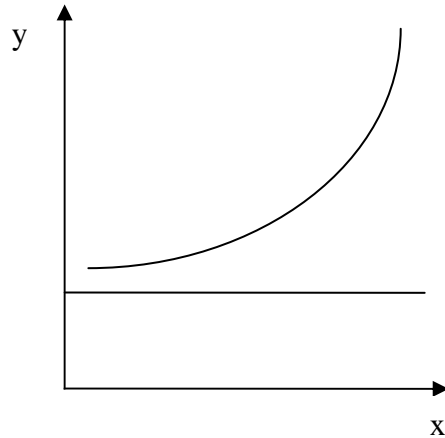
Б) гіперболічна
 $\hat{y} = a_0 + \frac{a_1}{x}$



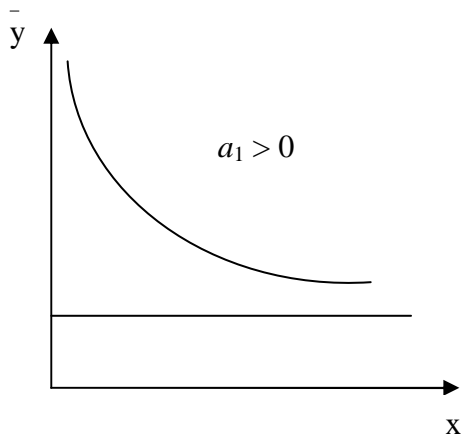
В) степенева
 $\hat{y} = a_0 x^{a_1}$



Г) модифікована експонента
 $\hat{y} = a_0 + a_1 a_2^x$



Д) крива гем перця
 $\hat{y} = e^{a_0 + a_1 a_2^x}$



Е) логістична
 $\hat{y} = 1/(a_0 + a_1 a_2^x)$

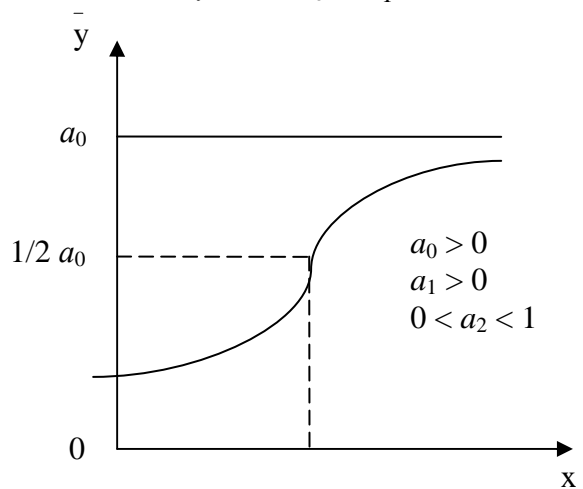


Рис. 2.4 – Графіки квазілінійних і суттєво нелінійних форм залежностей

2.2.3. Спосіб перших різниць

Спосіб перших різниць для обґрунтування аналітичної форми рівняння регресії передбачає послідовне виконання наступних дій:

1) групування об'єктів спостереження за фактором (факторами) x_i , для чого спочатку розраховують інтервал групування за формулою

$$i_{x_i} = \frac{x_{i \max} - x_{i \min}}{1 + 3,32 \times \lg n}. \quad (2.21)$$

У нашому прикладі інтервал, наприклад, за фактором E дорівнює

$$i_E = \frac{13,0 - 5,4}{1 + 3,32 \times \lg 29} \cong 1,0.$$

Отже, утворилось шість груп об'єктів спостереження (рис. 1.5).

2) у кожній групі визначають середні арифметичні значення \tilde{y}_f ($f = 1, 2, \dots, k$).

3) визначають перші різниці між наступним і попереднім значенням \tilde{y}_f за групами. Такі розрахунки для аналітичного обґрунтування форм залежності у від x_i , на прикладі залежності P від E подані в табл. 2.3.

Таблиця 2.3 – Перші різниці зростання P при збільшенні фактора E

Інтервал по E	n_t	P_j	$\sum_1^{n_t} P_j$	\tilde{P}_f	$\tilde{P}_{f+1} - \tilde{P}_f$
1-2	3	6,7; 7,3; 5,4	19,4	6,47	–
2-3	4	7,9; 8,9; 8,9; 7,5	32,3	8,08	+1,61
3-4	4	9,5; 9,1; 7,9; 9,3	35,8	8,95	+0,87
4-5	5	10,9; 9,7; 10,6; 8,5; 9,2	48,9	9,78	+0,83
5-6	4	9,8; 11,2; 11,5; 9,5;	42,0	10,5	+0,72
6-7	4	11,5; 11,4; 13,04; 10,6	46,5	11,63	+1,13
7-8	5	10,1; 12,7; 10,4; 12,5; 12,0	57,7	11,54	–0,09
Сума	29		282,6	9,74	

4) якщо перші різниці \tilde{y}_f однакові (приблизно однакові) або різні, але не виявляють закономірності зменшення, чи збільшення, то слід вибрати лінійну форму регресії. Саме такий випадок (табл. 2.3) і спостерігається в залежності P від E у нашому прикладі. Отже аналітична форма регресії P - E найбільш ймовірно є лінійною.

Якщо перші різниці виявляють явну тенденцію до збільшення або зменшення, треба вибирати криволінійну форму регресії з одним поворотом (квадратичну, гіперболічну, степеневу, експоненту).

Якщо ж перші різниці нагадують ординати нормального розподілу (спочатку збільшується, а потім зменшується і навпаки), то вибирають криволінійну форму регресії з двома поворотами (парабола 3-го порядку, логістична крива, або крива Гомперця).

2.2.4. Лінеаризація нелінійних рівнянь регресії

Оцінювання параметрів рівняння регресії, як буде показано далі (див. підрозділ 3.1.2), ґрунтується на ряді об'єктивних припущень.

Перше припущення вимагає, щоб рівняння регресії мало лінійну форму. Але в економіці часто зустрічаються залежності як квазілінійні, так і суттєво нелінійні.

Використання цього класу регресії для побудови економетричних моделей пов'язано з обчислювальними труднощами тому, що вказані регресії не допускають безпосереднього застосування класичного методу найменших квадратів для оцінювання параметрів рівнянь регресії. Щоб зробити це можливим, початкові дані спостережень піддають *перетворенням з метою лінеаризації* залежностей, тобто наданням їм лінійної форми.

Для квазілінійної регресії таке перетворення означає заміну значень незалежних змінних. Наприклад, в регресійному рівнянні гіперболічної форми

$$y = a_0 + a_1/x$$

$1/x$ можна замінити змінною $z = 1/x$ і отримати рівняння лінійної форми

$$y = a_0 + a_1z.$$

Аналогічно можна лінеаризувати квадратичну функцію (параболу)

$$y = a_0 + a_1x + a_2x^2,$$

замінивши $z = x^2$, тоді

$$y = a_0 + a_1x + a_2z.$$

Для суттєво нелінійних регресій лінеаризація досягається, як правило, шляхом логарифмування рівнянь регресії. Нижче наведені логарифмічні перетворення найуживаніших суттєво нелінійних регресій, що розглядалися в підрозділі 2.2.2:

степенева

$$y = a_0 x^{a_1} \rightarrow \lg y = \lg a_0 + a_1 \lg x,$$

показникова (експоненційна)

$$y = a_0 x^{a_1} \rightarrow \lg y = \lg a_0 + x \lg a_1,$$

Звідси степеневе рівняння регресії y від x стає лінійною регресією $\lg y$ від $\lg x$. В даному разі невідомими параметрами рівняння регресії є $\lg a_0$ і a_1 .

2.3. КОМПЛЕКС КОНТРОЛЬНИХ ЗАВДАНЬ

ЗА ЗМ 2 «СПЕЦИФІКАЦІЯ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ»

Навчальні елементи, що підлягають контролю і оцінюванню:

коефіцієнти парної кореляції і детермінації; суттєвість (невипадковість) і довірчі межі коефіцієнта кореляції; тестування мультиколінеарності; β -коефіцієнти; тестування екзогенних змінних на автономність; коефіцієнти множинної кореляції і детермінації; тестування екзогенних змінних на вклад до множинної детермінації; аналітичні форми рівнянь регресії: лінійні, квазілінійні, суттєво нелінійні; лінеаризація нелінійних рівнянь регресії; способи обґрунтування аналітичної форми моделі.

2.3.1. Тестові завдання

Виберіть правильну відповідь (відповіді):

T2.01. Коефіцієнт парної кореляції визначається за формулою:

а) $r_{yx} = \overline{y \times x} - \overline{yx}$;

б) $r_{yx} = \overline{yx} - \overline{y \times x}$;

в) $r_{yx} = \frac{\overline{yx}}{\overline{y \times x}}$;

г) $r_{yx} = (\overline{yx} - \overline{y \times x}) / \sigma_y \sigma_x$.

T2.02. Коефіцієнт парної кореляції може бути:

- а) $r_{yx} = 0$; б) $r_{yx} = 1$; в) $r_{yx} > 1$; г) $r_{yx} < -1$.

T2.03. Коефіцієнт парної кореляції між продуктивністю праці і кваліфікацією робітників може бути:

- а) $-1,83$; б) $+1,83$; в) $+0,83$; г) $-0,83$.

T2.04. Коефіцієнт парної кореляції між собівартістю одиниці продукції і обсягом її виробництва може бути:

- а) $-1,83$; б) $+1,83$; в) $+0,83$; г) $-0,83$.

T2.05. Коефіцієнт парної кореляції є додатним, якщо:

- а) $\overline{yx} < \bar{y} \times \bar{x}$; б) $\overline{yx} > \bar{y} \times \bar{x}$; в) $\overline{yx} = \bar{y} \times \bar{x}$; г) $\bar{y} > \bar{x}$.

T2.06. Коефіцієнт парної детермінації дорівнює:

- а) $d = r_{yx}$; б) $d = r_{yx}^2$; в) $d = \sqrt{r_{yx}}$; г) $d = |r_{yx}|$.

T2.07. Якщо t -статистика Стьюдента $> t_{\text{крит}}$, то коефіцієнт парної кореляції:

- а) значущий; б) суттєвий; в) випадковий; г) не випадковий.

T2.08. Коефіцієнт парної кореляції як випадкова величина має розподіл:

- а) нормальний;
б) асиметричний з лівостороннім ексцесом;
в) асиметричний з правостороннім ексцесом;
г) рівномірний.

T2.09. Z-перетворення Фішера для r_{yx} необхідне для:

- а) встановлення інтервалів довіри до r_{yx} ;
б) вилучення випадкових екзогенних змінних;
в) розрахунку коефіцієнта детермінації;
г) визначення значущості r_{yx} .

T2.10. Мультиколінеарність екзогенних змінних – це взаємозв’язок між:

- а) у і всіма x_i ;
- б) деякими парами екзогенних змінних;
- в) всіма парами екзогенних змінних;
- г) відсутність взаємозв’язку між екзогенними змінними.

T2.11. Якщо $r_{ij} (i \neq j) \leq r_{yx_i \min}$, то мультиколінеарність:

- а) слабка; б) помірна; в) сильна; г) відсутня.

T2.12. Якщо $r_{ij} (i \neq j) \geq r_{yx_i \max}$, то мультиколінеарність:

- а) слабка; б) помірна; в) сильна; г) відсутня.

T2.13. Якщо $r_{yx_i \min} \leq r_{ij} (i \neq j) \leq r_{yx_i \max}$, то мультиколінеарність:

- а) слабка; б) помірна; в) сильна; г) відсутня.

T2.14. Якщо $r_{ij} (i \neq j) = 0$, то мультиколінеарність:

- а) слабка; б) помірна; в) сильна; г) відсутня.

T2.15. β -коефіцієнт дорівнює:

- а) коефіцієнту кореляції;
- б) коефіцієнту детермінації;
- в) коефіцієнту кореляції за умови, що матриця $[r_{ij}]$ є квадратною одиничною матрицею;
- г) кількості σ_y змінення у при зміненні x на одне σ_x .

T2.16. Екзогенна змінна має певну автономність впливу на y за умови:

- а) $\beta_i / r_{yx_i} > 1$; б) $\beta_i / r_{yx_i} < 0$; в) $0 < \beta_i / r_{yx_i} < 1$; г) $\beta_i = r_{yx_i}$.

T2.17. Коефіцієнт множинної кореляції розраховується за формулою:

- а) $R_{yx_1 \dots x_m} = \sqrt{r_{yx_1}^2 + r_{yx_2}^2 + \dots + r_{yx_m}^2}$;
- б) $R_{yx_1 \dots x_m} = \sqrt{\beta_1^2 + \beta_2^2 + \dots + \beta_m^2}$;
- в) $R_{yx_1 \dots x_m} = \sqrt{r_{yx_1} \beta_1 + r_{yx_2} \beta_2 + \dots + r_{yx_m} \beta_m}$;
- г) $R_{yx_1 \dots x_m} = r_{yx_1} + r_{yx_2} + \dots + r_{yx_m}$.

T2.18. Коефіцієнт частинної детермінації у багатofакторній регресії дорівнює:

а) $d_i = r_{yx_i}^2$; б) $d_i = \beta_i^2$; в) $d_i = r_{yx_i} \beta_i$; г) $d_i = r_{yx_i} / \beta_i$.

T2.19. Рівняння регресії можуть бути:

- а) простими;
- б) множинними;
- в) лінійними;
- г) квазілінійними;
- д) суттєво нелінійними.

T2.20. Коефіцієнт множинної кореляції може бути:

а) > 1 ; б) < 0 ; в) > 0 ; г) < 1 .

T2.21. Рівняння регресії $y = a_0 + a_1x + e$ є:

- а) лінійним; б) квазілінійним; в) суттєво нелінійним.

T2.22. Рівняння регресії $y = a_0 + a_1x + a_2x^2 + e$ є:

- а) лінійним; б) квазілінійним; в) суттєво нелінійним.

T2.23. Рівняння регресії $y = a_0x^{a_1} + e$ є:

- а) лінійним; б) квазілінійним; в) суттєво нелінійним.

T2.24. Рівняння регресії $\hat{y} = a_0 + a_1x + a_2x^2$ є рівнянням:

- а) прямої; б) гіперболи; в) параболи; г) степеневі функції; д) експоненти.

T2.25. Рівняння регресії $\hat{y} = a_0 + a_1/x$ є рівнянням:

- а) прямої; б) гіперболи; в) параболи; г) степеневі функції; д) експоненти.

T2.26. Рівняння регресії $\hat{y} = a_0x^{a_1}$ є рівнянням:

- а) прямої; б) гіперболи; в) параболи; г) степеневі функції; д) експоненти.

T2.27. Рівняння регресії $\hat{y} = a_0 \times a_1^x$ є рівнянням:

- а) прямої; б) гіперболи; в) параболи; г) степеневі функції; д) експоненти.

T2.28. Лінії або точки насичення мають такі типи регресії:

а) $\hat{y} = a_0 + a_1x$; б) $\hat{y} = a_0 + a_1/x$; в) $\hat{y} = a_0 + a_1x + a_2x^2$; г) $\hat{y} = a_0x^{a_1}$.

T2.29. Найкращим способом обґрунтування аналітичної форми регресії є:

- а) теоретичний; б) графічний; в) аналітичний.

T2.30. У простій регресії $\hat{y} = a_0 + a_1x$ параметр a_0 визначає:

- а) точку на осі y , де ця вісь перетинається з лінією регресії (перетин);
б) тангенс кута нахилу лінії регресії до осі x (нахил);
в) результат впливу на y всіх інших факторів;
г) завжди дорівнює нулю.

T2.31. У простій регресії $\hat{y} = a_0 + a_1x$ параметр a_1 визначає:

- а) точку на осі y , де ця вісь перетинається з лінією регресії (перетин);
б) тангенс кута нахилу лінії регресії до осі x (нахил);
в) результат впливу на y всіх інших факторів;
г) завжди дорівнює одиниці.

T2.32. У простій регресії $\hat{y} = 0,34 + 1,2x$ нахил дорівнює:

- а) x ; б) y ; в) 0,34; г) 1,2.

T2.33. У простій регресії $\hat{y} = 0,34 + 1,2x$ перетин дорівнює:

- а) x ; б) y ; в) 0,34; г) 1,2.

T2.34. За рівнянням регресії заробітної плати в гривнях (y) від освіти в роках навчання (x) $\hat{y} = 12,2 + 525x$ особа, яка додатково навчалась один рік, може розраховувати на таке збільшення оплати:

- а) 12,2; б) 525; в) $12,2 + 525$; г) 24,4.

2.3.2. Логічні вправи

Л2.01. Обґрунтуйте теоретично наявність кореляційної залежності собівартості одиниці продукції від фондоозброєності праці робітників.

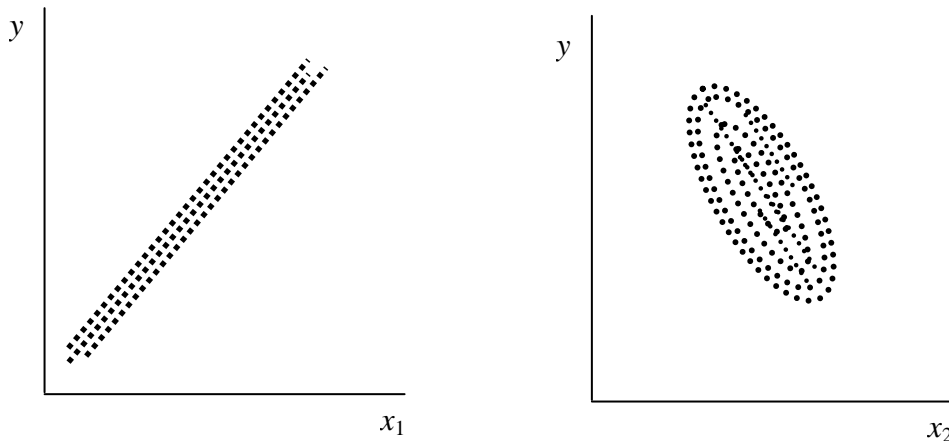
Л2.02. Теоретично доведіть, що рентабельність витрат підприємства кореляційно залежить від продуктивності праці робітників.

Л2.03. Поясніть, за якими ознаками витрати виробництва поділяються на постійні та змінні, як трансформується цей поділ у розрахунку витрат на

одиницю продукції. Покажіть у цьому контексті, що собівартість одиниці продукції є кореляційною функцією обсягу її виробництва.

Л2.04. Доведіть, що коефіцієнт лінійної парної кореляції неточно визначає силу (тісноту) кореляційної залежності y від x_i .

Л2.05. Розгляньте два поля кореляції. Визначте, чим відрізняється коефіцієнт кореляції r_{yx_1} від r_{yx_2} .



Л2.06. Поясніть, чи може бути відсутньою мультиколінеарність в матрицях економічної статистики. Покажіть це на прикладах.

Л2.07. Доведіть, що у разі відсутності мультиколінеарності факторів на прикладі двофакторної залежності $\hat{y} = f(x_1, x_2)$, $\beta_i = r_{yx_i}$.

Л2.08. Доведіть, що у разі відсутності мультиколінеарності факторів на прикладі двофакторної залежності $\hat{y} = f(x_1, x_2)$, $R_{yx_1x_2} = \sqrt{\sum_1^2 r_{yx_i}^2}$.

Л2.09. За критеріями оцінки значущості, автономності та вкладу факторів x_1, x_2, x_3 (табл.) обґрунтуйте рішення щодо включення цих факторів до рівняння регресії $\hat{y} = f(x_1, x_2, x_3)$.

	t	γ_i	ρ_i
x_1	8,12	0,72	3,14
x_2	6,03	0,37	2,08
x_3	3,14	-0,61	1,09

Л2.10. Визначить (з'єднайте стрілками) види наведених рівнянь регресії:

$\hat{y}_t = a_0 + a_1x_t + a_2x_{t-1} + a_3x_{t-2}$
$\hat{y}_t = a_0 + a_1x_t$
$\hat{y}_t = a_0 + a_1x_t + a_2y_{t-1}$
$\hat{y}_t = a_0 + a_1x_t + a_2y_{t-1} + a_3y_{t-2}$

дистрибутивно-лагова
авторегресивна
динамічна

Л2.11. Визначить (з'єднайте стрілками) тип математичної форми наведених рівнянь регресії:

лінійна
квазілінійна
суттєво нелінійна
нелінійна

$\hat{y} = a_0 + a_1x$
$\hat{y} = a_0x^{a_1}$
$\hat{y} = a_0a_1^x$
$\hat{y} = a_0 + \frac{a_1}{x}$
$\hat{y} = a_0 + a_1x + a_2x_2$

Л2.12. Визначить (з'єднайте стрілками) математичну форму наведених рівнянь регресії:

пряма
зворотна
гіперболічна
квадратична
параболічна
степенева
показникова

$\hat{y} = a_0 + \frac{a_1}{x}$
$\hat{y} = a_0 + a_1x$
$\hat{y} = a_0 + a_1x + a_2x^2$
$\hat{y} = a_0x^{a_1}$
$\hat{y} = a_0a_1^x$

Л2.13. Наведені рівняння регресії мають лінії або точки насичення (екстремуми) – з'єднайте стрілками:

$\hat{y} = a_0 + a_1x$
$\hat{y} = a_0 + \frac{a_1}{x}$
$\hat{y} = a_0x^{a_1}$
$\hat{y} = a_0 + a_1x + a_2x^2$

лінія насичення
точка насичення (екстремум)

Л2.14. Покажіть на графіках, що являють собою точки або лінії насичення залежної змінної для різних форм криволінійної регресії.

Л2.15. Доведіть, що найкращім способом обґрунтування форми регресії є теоретичний. Яке значення мають щодо цього графічний та аналітичні способи?

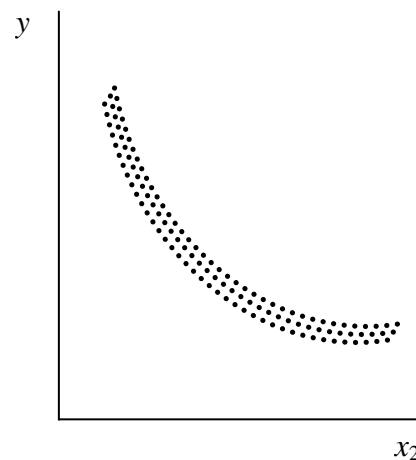
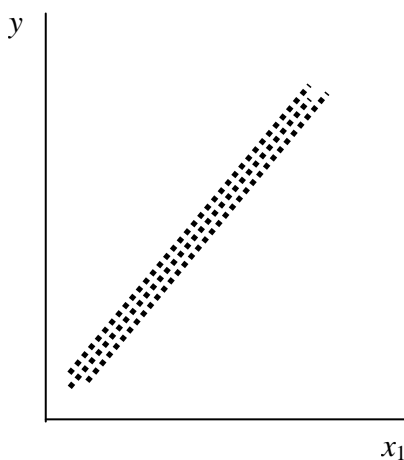
Л2.16. Обґрунтуйте теоретично форму регресії капітальних вкладень у будівництво цукрових заводів (K) від потужності цих заводів (M). Укладіть рівняння регресії $\hat{K} = f(M)$.

Л2.17. Обґрунтуйте теоретично форму регресії собівартості 1 тис. шт. цегли (C) від обсягу виробництва цегли на цегельному заводі (Q). Укладіть рівняння регресії $\hat{C} = f(Q)$.

Л2.18. Обґрунтуйте рівняння множинної регресії продуктивності праці робітників ($П$) від рівня кваліфікації (K), виробничого стажу (C) та їхнього віку (B).

Л2.19. Обґрунтуйте рівняння множинної регресії продуктивності праці робітників ($П$) від фондоозброєності праці (Φ), плинності їх на підприємстві (T) та температури зовнішнього повітря (t).

Л2.20. За графічними ознаками полів кореляції (рис. 2.5) обґрунтуйте окремі рівняння парної регресії $\hat{y} = f(x_i)$.



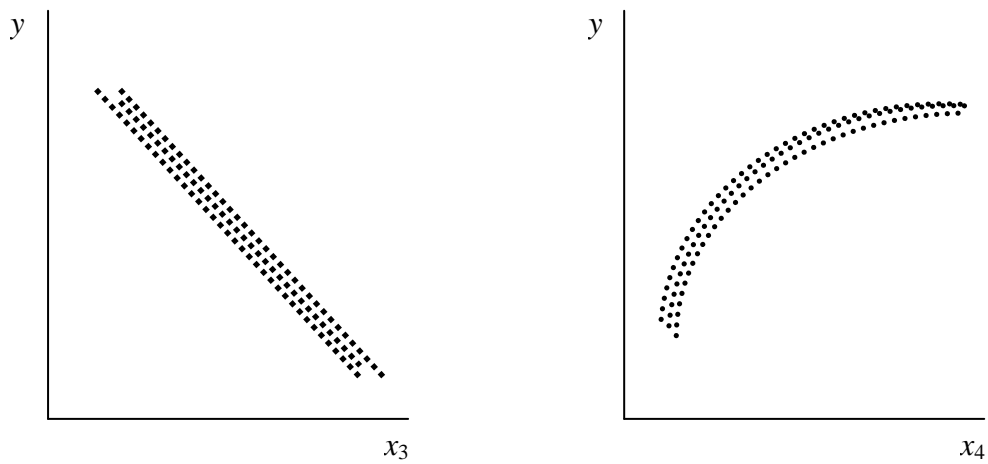


Рис. 2.5

Л2.21. За графічними ознаками полів кореляції (рис. 2.5) обґрунтуйте рівняння множинної регресії $\hat{y} = f(x_1, x_2, x_3, x_4)$.

Л2.22. За графічними ознаками полів кореляції (рис. 2.6) обґрунтуйте рівняння множинної регресії $\hat{y} = f(x_1, x_2, x_3)$.

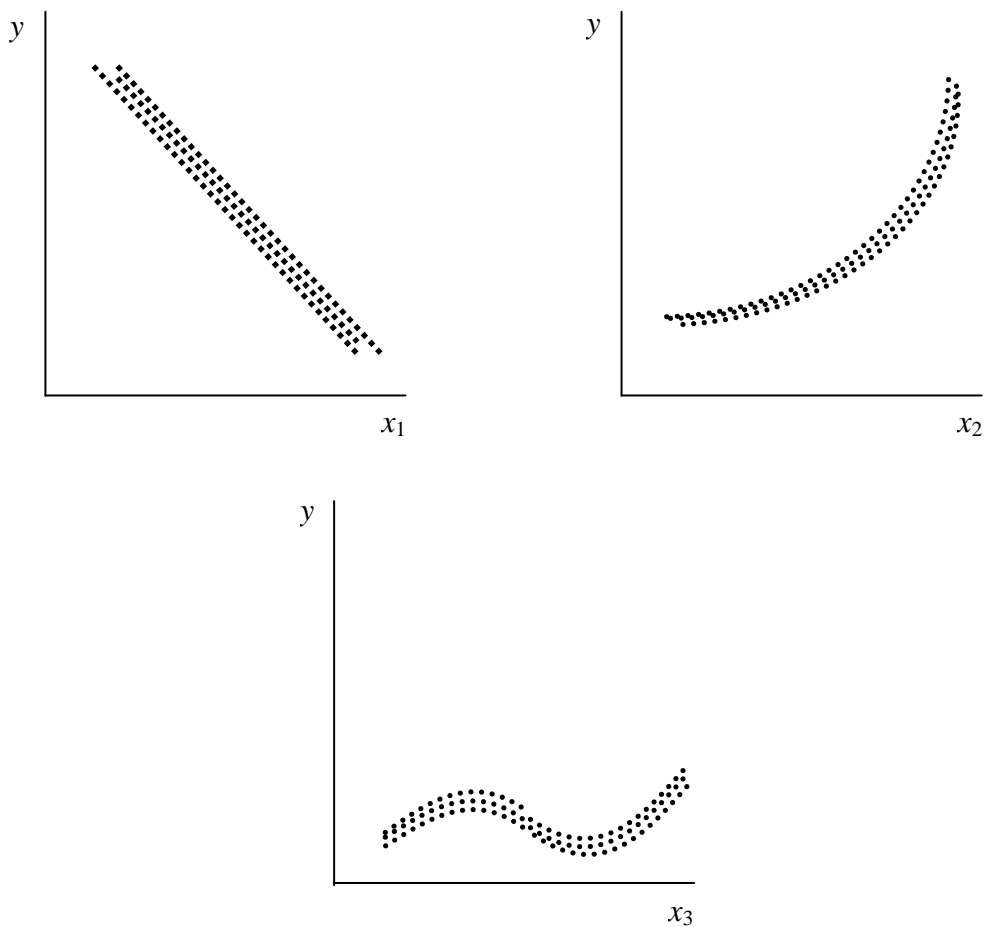


Рис. 2.6

Л2.23. Викладіть ваші уявлення щодо емпіричної та теоретичної лінії регресії. Яку роль вони відіграють в обґрунтуванні форми регресії?

Л2.24. В таблиці наведено три варіанти перших різниць між груповими середніми значеннями змінної y в залежності від фактора x . Визначте форму рівняння регресії $\hat{y} = f(x)$ за трьома варіантами окремо (поля кореляції будувати не треба).

x_j	$(\hat{y}_j - \hat{y}_{j-1})$ за варіантами			x_j	$(\hat{y}_j - \hat{y}_{j-1})$ за варіантами		
	1	2	3		1	2	3
1	4	2	1	6	4	4	4
2	4	2	2	7	4	5	3
3	4	3	3	8	4	5	2
4	4	3	4	9	4	6	1
5	4	4	5				

Л2.25. Поясніть переваги та недоліки критеріального способу обґрунтування форми регресії.

Л2.26. Поясніть, чому коефіцієнт парної кореляції, як випадкова величина, має асиметричний розподіл з правостороннім ексцесом.

Л2.27. Поясніть, чому коефіцієнт множинної кореляції завжди додатний.

Л2.28. Лінеаризуйте квазілінійні рівняння регресії:

$$\hat{y} = a_0 + \frac{a_1}{x}; \quad \hat{y} = a + a_1x + a_2x^2.$$

Л2.29. Лінеаризуйте суттєво нелінійні рівняння регресії:

$$\hat{y} = a_0x^{a_1}; \quad \hat{y} = a_0a_1^x.$$

2.3.3. Розрахункові вправи

P2.01. За матрицею статистики за допомогою графічних критеріїв обґрунтуйте форму кореляційної залежності y від x .

j	x_i	y_i
1	10	7
2	9	4
3	15	10
4	11	5
5	7	2

j	x_i	y_i
6	8	3
7	13	8
8	12	6
9	14	8
10	8	2

P2.02. За матрицею статистики за допомогою сигмального критерію визначити спрямування кореляційної залежності y від x .

j	x_i	y_i
1	7	2
2	9	4
3	9	3

j	x_i	y_i
4	13	12
5	10	6

P2.03. За допомогою графічного критерію обґрунтуйте за наведеною матрицею спрямованість кореляційної залежності y від x .

j	x_i	y_i
1	7	2
2	9	4
3	13	9

j	x_i	y_i
4	9	7
5	16	10
6	8	3

P2.04. За матрицею статистики за допомогою сигмального критерію визначити напрям кореляційної залежності y від x .

j	x_i	y_i
1	9	4
2	10	7
3	7	2

j	x_i	y_i
4	8	3
5	15	10
6	11	4

P2.05. Обчисліть коефіцієнт парної кореляції за умови: $\bar{y} = 2,5$; $\bar{x} = 3,5$; $\sum ux = 100$; $\sigma_y = 2$; $\sigma_x = 1$; $n = 10$. Оцініть силу (тісноту) кореляційної залежності y від x .

P2.06. Обчисліть коефіцієнт парної кореляції за умови: $\sum y = 80$; $\sum x = 40$; $\overline{yx} = 5$; $n = 40$; $\sigma_y = 2,5$; $D_x = 4$. Визначте силу (тісноту) кореляційної залежності y від x .

P2.07. Обчисліть коефіцієнт парної кореляції за умови: $\sum y = 50$; $\bar{x} = 2$; $\sum ux = 250$; $n = 50$; $D_y = 4$; $\sigma_x = 2,5$. Оцініть силу (тісноту) кореляційної залежності y від x .

P2.08. За матрицею статистики обчисліть коефіцієнт парної кореляції, визначте силу (тісноту) кореляційної залежності y від x .

j	x_i	y_i
1	6	3
2	5	1

j	x_i	y_i
3	10	8
4	7	4

P2.09. За повною матрицею коефіцієнтів парної кореляції обчисліть коефіцієнт β_1 .

	x_1	x_2	y
x_1	1	0,5	0,8
x_2	0,5	1	0,6
y	0,8	0,6	1

P2.10. За даними вправи P2.09. обчисліть коефіцієнт β_2 .

P2.11. За даними вправи P2.09. обчисліть коефіцієнт множинної кореляції $R_{yx_1x_2}$.

P2.12. Задані коефіцієнти парної кореляції $r_{yx_i} = \begin{bmatrix} 0,6 \\ 0,7 \end{bmatrix}$, $r_{ij} = \begin{bmatrix} 1 & 0,2 \\ 0,2 & 1 \end{bmatrix}$.

Обчисліть коефіцієнт β_1 .

P2.13. За даними вправи P2.12. обчисліть коефіцієнт β_2 .

P2.14. За даними вправи P2.12. обчисліть коефіцієнт множинної кореляції

$$R_{yx_1x_2}.$$

P2.15. Обчисліть коефіцієнт множинної кореляції $R_{yx_1x_2}$ за умови: $r_{yx_1} = 0,8$;

$$r_{yx_2} = 0,6; \beta_1 = 0,7; \beta_2 = 0,3.$$

P2.16. Оцініть за допомогою γ -критерію автономність впливу на y трьох факторів за такими даними:

x_i	x_1	x_2	x_3
r_{yx_i}	0,6	0,8	0,7
β_i	0,8	0,7	-0,3

P2.17. Оцініть за допомогою ρ -критерію значущість фактора x_1 за умови:

$$R_{yx_1x_2} = 0,94868; r_{yx_2} = 0,86023; n = 36.$$

P2.18. Дано $r_{yx} = 0,86$, $n = 42$. Перевірте суттєвість (невипадковість) зв'язку змінних y та x .

P2.19. Дано $r_{yx} = 0,86$, $n = 42$. Побудуйте інтервали довіри для коефіцієнта кореляції. Використовуйте z -перетворення Фішера.

P2.20. Дано $r_{yx} = 0,51$, $n = 37$. Перевірте суттєвість (невипадковість) зв'язку змінних y та x .

P2.21. Дано $r_{yx} = 0,51$, $n = 37$. Побудуйте інтервали довіри для коефіцієнта кореляції. Використовуйте z -перетворення Фішера.

P2.22. Дано $r_{yx} = 0,39$, $n = 44$. Перевірте суттєвість (невипадковість) зв'язку змінних y та x .

P2.23. Дано $r_{yx} = 0,39$, $n = 44$. Побудуйте інтервали довіри для коефіцієнта кореляції. Використовуйте z -перетворення Фішера.

P2.24. Дано $r_{yx} = 0,89$. Визначте питому вагу варіації y за рахунок впливу x .

P2.25. Охарактеризуйте силу колінеарності факторів x_1 і x_2 за таких умов:

$$[r_{ij}] = \begin{bmatrix} 1 & 0,48 \\ 0,48 & 1 \end{bmatrix}; [r_{yx_i}] = \begin{bmatrix} 0,67 \\ 0,89 \end{bmatrix}.$$

P2.26. Охарактеризуйте силу колінеарності факторів x_1 і x_2 за таких умов:

$$[r_{yx_i}] = \begin{bmatrix} 0,82 \\ 0,88 \end{bmatrix}; \quad [r_{ij}] = \begin{bmatrix} 1 & 0,91 \\ 0,91 & 1 \end{bmatrix}.$$

P2.27. Визначить та охарактеризуйте коефіцієнт множинної кореляції за такими вихідними даними: $[r_{ij}] = \begin{bmatrix} 1 & 0,48 \\ 0,48 & 1 \end{bmatrix}; \quad [r_{yx_i}] = \begin{bmatrix} 0,79 \\ 0,64 \end{bmatrix}.$

P2.28. Визначить та охарактеризуйте коефіцієнт множинної детермінації за такими вихідними даними: $[r_{ij}] = \begin{bmatrix} 1 & 0,62 \\ 0,62 & 1 \end{bmatrix}; \quad [r_{yx_i}] = \begin{bmatrix} 0,80 \\ 0,71 \end{bmatrix}.$

P2.29. Визначить та охарактеризуйте коефіцієнт множинної детермінації за такими вихідними даними:

$$\begin{aligned} \beta_1 &= 0,58; & r_{yx_1} &= 0,80; \\ \beta_2 &= 0,36; & r_{yx_2} &= 0,71. \end{aligned}$$

P2.30. Визначить та охарактеризуйте коефіцієнти частинної детермінації за такими вихідними даними:

$$\begin{aligned} \beta_1 &= 0,54; & r_{yx_1} &= 0,78; \\ \beta_2 &= 0,40; & r_{yx_2} &= 0,68. \end{aligned}$$

P2.31. Розрахуйте коефіцієнт множинної кореляції за такими вихідними даними:

$$\begin{aligned} \beta_1 &= 0,46; & r_{yx_1} &= 0,59; \\ \beta_2 &= 0,67; & r_{yx_2} &= 0,74. \end{aligned}$$

P2.32. Виконайте тестування факторів за ρ -критерієм значущості за такими даними: $R_{yx_1x_2} = 0,94; r_{yx_1} = 0,89; r_{yx_2} = 0,74; n = 40$. Зробіть висновки.

P2.33. Дано $[r_{ij}] = \begin{bmatrix} 1 & 0,56 \\ 0,56 & 1 \end{bmatrix}; \quad [r_{yx_i}] = \begin{bmatrix} 0,81 \\ 0,68 \end{bmatrix}$. Розрахуйте коефіцієнт множинної кореляції за формулою Боярського.

P2.34. За даними P2.33. розрахуйте β_1 .

P2.35. За даними P2.33. розрахуйте β_2 .

P2.36. За даними P2.33. розрахуйте коефіцієнти частинної детермінації.

P2.37. За даними 2.33. розрахуйте коефіцієнт множинної кореляції за допомогою β -коефіцієнтів.

P2.38. Дано:

x	r_{yx_i}	β_i
x_1	0,81	0,46
x_2	0,37	0,18
x_3	0,63	0,39

Оцініть автономність впливу факторів x_1 , x_2 , та x_3 на y .

P2.39. Дано:

j	y	x
1	8	4
2	9	5
3	6	3
4	12	7
5	10	6

Розрахуйте коефіцієнт кореляції змінних y та x .

2.3.4. Відповіді до розрахункових вправ

2.01. Наявна додатна лінійна залежність.

2.02. Наявна додатна залежність.

2.03. Наявна додатна лінійна залежність.

2.04. Наявна додатна залежність.

2.05. 0,625.

2.06. 0,6.

2.07. 0,6.

2.08. 0,9961.

2.09. 0,667.

2.10. 0,267.

2.11. 0,833.

- 2.12. 0,477.
- 2.13. 0,604.
- 2.14. 0,842.
- 2.15. 0,860.
- 2.16. x_3 не має самостійного значення.
- 2.17. $1,724 < 1,96$.
- 2.18. $10,66 > 2,021$.
- 2.19. $0,75 \leq \rho \leq 0,92$.
- 2.20. $3,507 > 2,035$.
- 2.21. $0,22 \leq \rho \leq 0,72$.
- 2.22. $2,745 > 2,021$.
- 2.23. $0,12 \leq \rho \leq 0,51$.
- 2.24. 79,2%.
- 2.25. Слабка.
- 2.26. Сильна.
- 2.27. 0,844.
- 2.28. 0,845.
- 2.29. 0,720.
- 2.30. 0,421; 0,272.
- 2.31. 0,861
- 2.32. 2,720; 1,424.
- 2.33. 0,857.
- 2.34. 0,625.
- 2.35. 0,330.
- 2.36. 0,506; 0,224.
- 2.37. 0,857.
- 2.38. 0,568; 0,486; 0,624.
- 2.39. 0,990.

3. ЗМІСТОВИЙ МОДУЛЬ 3: ОЦІНЮВАННЯ ПАРАМЕТРІВ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ

3.1. ОЦІНЮВАННЯ ПАРАМЕТРІВ РІВНЯНЬ РЕГРЕСІЇ

3.1.1. Мета і вимоги до оцінювання параметрів

У попередніх розділах, розглядаючи наскрізний приклад, ми зробили припущення про лінійну залежність рентабельності витрат (P) від енергоозброєності праці (E) і коефіцієнта постійності складу промислово – виробничного персоналу (K):

$$P = \alpha_0 + \alpha_1 E + \alpha_2 K + \varepsilon,$$

де $\alpha_0, \alpha_1, \alpha_2$ – невідомі параметри рівняння регресії, що оцінюються;

ε – випадкова величина, яка характеризує вплив на рентабельність усіх інших неврахованих факторів.

Оцінювання, тобто визначення кількісних значень цих параметрів рівняння регресії і випадкової складової $\alpha_0, \alpha_1, \alpha_2$ і ε , повинне забезпечити отримання незміщених, обґрунтованих і ефективних оцінок.

Оцінка параметра α_i , тобто a_i називається *незміщеною*, якщо її математичне сподівання дорівнює значенню параметра, який оцінюється, тобто $E(a_i) = \alpha_i$.

Оцінка параметра α_i називається *обґрунтованою*, якщо при збільшенні кількості спостережень (об'єму вибірки) її значення наближається до параметра, що оцінюється, тобто a_i (при $n \rightarrow N$) $\rightarrow \alpha_i$.

Оцінки параметрів α_i називаються *ефективними*, якщо дисперсія D_e має найменше значення (зауважимо, що $e_j = y_j - \hat{y}_j$). Ця вимога означає, що відхилення точок поля (простору) кореляції від лінії (поверхні) регресії повинні бути найменшими, тобто $\sum e_j^2 \rightarrow \min$.

Перелічені вимоги до оцінювання параметрів рівнянь регресії (у нашому прикладі це оцінки a_0, a_1, a_2 і e), мають дуже важливе значення в економетричному моделюванні. Тільки при виконанні вказаних вимог можна побудувати якісну економетричну модель.

3.1.2. Основні припущення щодо оцінювання параметрів

Економетричне моделювання методом класичного кореляційно-регресійного аналізу засноване на ряді наведених нижче припущень. Вони стосуються майже всіх етапів моделювання, а саме: ідентифікації змінних; формування матриці вихідних даних щодо змінних; специфікації форми рівнянь регресії; вибору методу оцінювання параметрів. Розглянемо ці припущення.

1. Зв'язок між змінними y і x_i описується *залежністю лінійної форми*

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m + \varepsilon.$$

Якщо за економічною теорією лінійна форма неприйнятна, обґрунтовують вибір адекватної квазілінійної або суттєво нелінійної форми зв'язку, але такої, яка шляхом перетворення приводиться до лінійної (див. підрозділ 2.2.4).

2. Для будь-яких значень x_i *математичне сподівання випадкової величини ε дорівнює нулю, тобто $E(\varepsilon) = 0$* . Це припущення стверджує, що вплив усіх факторів, які не враховані в моделі і тому їх віднесено до ε , має випадковий, а не системний характер на математичне сподівання y . Додатні значення ε компенсують від'ємні ε і тому їхній усереднений чи очікуваний вплив на y дорівнює нулю. З цього випливає, що математичне сподівання y дорівнює $\hat{y} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m$.

3. *Дисперсія випадкової величини ε постійною* для всіх значень x_i , тобто $D_\varepsilon = const$. Це припущення означає незалежність дисперсії D_ε від значень x_i , яка називається *гомоскедастичністю*. Випадок, коли умова гомоскедастичності не виконується, називається *гетероскедастичністю*.

4. *Значення випадкової величини ε є незалежними*, не корелюють між собою, тобто $r_{\varepsilon_i \varepsilon_j} = 0$ ($i \neq j$). Це припущення вказує на некорельованість залишків ε_j для різних спостережень. Воно часто порушується у випадках, коли дані спостережень є часовими рядами і тоді $r_{\varepsilon_i \varepsilon_j} \neq 0$ ($i \neq j$). Це явище називається *автокореляцією залишків*, у простих випадках автокореляції першого порядку її наявність підтверджується тим, що $r_{\varepsilon_i \varepsilon_{j-1}} \neq 0$.

5. *Випадкова величина ε розподілена за нормальним законом* з математичним сподіванням, що дорівнює нулю, тобто $E(\varepsilon) = 0$, і дисперсією D_ε або σ_ε^2 .

Це припущення вимагає, щоб залишки були розподілені за нормальним законом. Воно є необхідним в подальшому для побудови інтервалів довіри для параметрів лінійної регресійної моделі.

6. *Випадкова величина ε і змінні x_i не корелюють між собою*, тобто $r_{\varepsilon x_i} = 0$. Це припущення виконується автоматично, коли змінні x_i є детермінованими величинами, що ми і будемо припускати далі.

7. *Незалежні змінні не корелюють між собою*, тобто $r_{x_i x_j} = 0$ ($i \neq j$), принаймні сила їх попарних зв'язків невисока порівняно з r_{yx_i} . Це припущення вимагає, щоби вплив кожного окремого x_i на y був достатньо автономним, незалежним. Порушення цієї вимоги породжує *проблему мультиколінеарності* факторів, яку ми уже розглядали в розділі 2.1.5.

Відомо, що коли наведені припущення виконуються, то оцінки a_0, a_1, \dots, a_m параметрів $\alpha_0, \alpha_1, \dots, \alpha_m$ можна отримати класичним методом найменших квадратів і вони будуть незміщеними, обґрунтованими і ефективними, тобто найкращими BLUE-оцінками.

3.1.3. Метод найменших квадратів

Ідею методу найменших квадратів (МНК) вчені сформулювали ще на початку XIX ст., а саме англієць Гаусс і француз Лежандр. Але МНК як метод оцінювання параметрів рівнянь регресії, був опрацьований пізніше російським математиком М. Чебишевим в монографії «Об интерполировании по методу наименьших квадратов», що вийшла друком у 1859 р.

Метод найменших квадратів (МНК) заснований на вимозі, щоб

$$\sum (y - \hat{y})^2 \rightarrow \min, \quad (3.1)$$

тобто щоб відхилення точок поля кореляції від прямої регресії (залишки e) були найменшими. Відхилення $y - \hat{y}$ є помилкою оцінювання, бо залишки e не пояснюються рівнянням регресії (див. рис. 3.1).

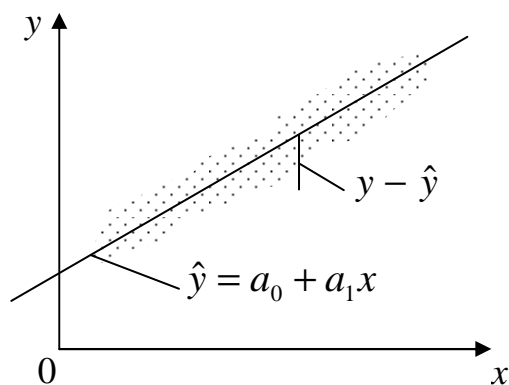


Рис. 3.1 – Залишки, що не пояснюються регресією

Ми можемо визначити величину цих помилок тільки для об'єктів спостережень (точок поля), але для інших можливих комбінацій y і x вони невідомі. Цілком природно за «найкращу» пряму лінію регресії вибрати таку, для якої б сума квадратів залишків приймала найменше значення, тобто

$$\sum_1^n e_j^2 = \sum_1^n (y - \alpha_0 - \alpha_1 x)^2 = \rightarrow \min.$$

Необхідною умовою мінімуму є рівність нулю частинних похідних цієї функції по α_0 і α_1 :

$$\frac{df}{da_0} = -2 \sum_1^n (y - a_0 - a_1 x) = 0;$$

$$\frac{df}{da_1} = -2 \sum_1^n (y - a_0 - a_1 x) x = 0.$$

Розкривши дужки, отримаємо систему нормальних рівнянь з невідомими α_0 і α_1 :

$$\sum_1^n y_j = a_0 \sum_1^n 1 + a_1 \sum_1^n x_j ; \tag{3.2}$$

$$\sum_1^n y_j x_j = a_0 \sum_1^n x_j + a_1 \sum_1^n x_j^2 .$$

Наведемо давно підмічені правила і приклади складання системи нормальних рівнянь для будь-яких форм рівняння регресії. Правило перше: перше рівняння системи отримуємо, підсумовуючи рівняння регресії за всіма спостереженнями змінних.

Наприклад, для трифакторної лінійної регресії

$$\hat{y} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3$$

перше рівняння системи за цим правилом матиме вигляд

$$\sum y = \alpha_0 \sum 1 + \alpha_1 \sum x_1 + \alpha_2 \sum x_2 + \alpha_3 \sum x_3.$$

Друге правило: друге, третє і всі інші рівняння системи отримують множенням першого рівняння на співмножники відповідно при α_1 , α_2 и т.д. у рівнянні регресії. У даному випадку множенням першого рівняння відповідно на x_1 , x_2 та x_3 отримаємо друге, третє і четверте рівняння системи

$$\sum yx_1 = \alpha_0 \sum x_1 + \alpha_1 \sum x_1^2 + \alpha_2 \sum x_2 x_1 + \alpha_3 \sum x_3 x_1,$$

$$\sum yx_2 = \alpha_0 \sum x_2 + \alpha_1 \sum x_1 x_2 + \alpha_2 \sum x_2^2 + \alpha_3 \sum x_3 x_2,$$

$$\sum yx_3 = \alpha_0 \sum x_3 + \alpha_1 \sum x_1 x_3 + \alpha_2 \sum x_2 x_3 + \alpha_3 \sum x_3^2.$$

Ще приклад. Нехай рівняння регресії має степеневу форму

$$y = \alpha_0 x^{\alpha_1}.$$

Лінеаризуємо його

$$\lg y = \lg \alpha_0 + \alpha_1 \lg x.$$

Система нормальних рівнянь для визначення $\lg \alpha_0$ і α_1 , складена за наведеними щойно правилами, приймає такий вигляд:

$$\sum \lg y = \lg \alpha_0 \sum 1 + \alpha_1 \sum \lg x,$$

$$\sum \lg y \lg x = \lg \alpha_0 \sum \lg x + \alpha_1 \sum (\lg x)^2.$$

У нашому наскрізному прикладі моделювання залежності рентабельності від двох факторів

$$\hat{P} = \alpha_0 + \alpha_1 E + \alpha_2 K$$

система нормальних рівнянь для оцінювання α_0 , α_1 , та α_2 складена за наведеними вище правилами, така:

$$\sum P = \alpha_0 \sum 1 + \alpha_1 \sum E + \alpha_2 \sum K,$$

$$\sum PE = \alpha_0 \sum E + \alpha_1 \sum E^2 + \alpha_2 \sum KE,$$

$$\sum PK = \alpha_0 \sum K + \alpha_1 \sum EK + \alpha_2 \sum K^2.$$

За даними табл. 1.5. і 2.1 перепишемо цю систему рівнянь в числах:

$$\begin{aligned}282,6 &= 29\alpha_0 + 136,6\alpha_E + 1740\alpha_K, \\1425,26 &= 136,6\alpha_0 + 754,72\alpha_E + 8567,7\alpha_K, \\17413,9 &= 1740\alpha_0 + 8567,7\alpha_E + 108164\alpha_K.\end{aligned}\tag{3.3}$$

Нижче розглянемо декілька можливих способів розв'язання системи нормальних рівнянь.

Спосіб Гаусса

Якщо вирішення системи нормальних рівнянь (3.3) виконується вручну, можна скористатися методом Гаусса послідовного вилучення невідомих. Поділимо рівняння системи на числові співмножники при α_0 , тобто відповідно на 29, 136,6 і 1740 та отримаємо

$$\begin{aligned}9,7448 &= \alpha_0 + 4,7103\alpha_E + 60,0000\alpha_K, \\10,4338 &= \alpha_0 + 5,5250\alpha_E + 62,7211\alpha_K, \\10,0080 &= \alpha_0 + 4,9240\alpha_E + 62,1632\alpha_K.\end{aligned}$$

Відніmemo перше рівняння від другого і третього та отримаємо систему двох рівнянь з невідомими α_E і α_K .

$$\begin{aligned}0,6890 &= 0,8147\alpha_E + 2,7211\alpha_K, \\0,2632 &= 0,2137\alpha_E + 2,1632\alpha_K.\end{aligned}$$

Поділимо систему рівнянь на числові співмножники при α_E і отримаємо

$$\begin{aligned}0,8457 &= \alpha_E + 3,3400\alpha_K, \\1,2316 &= \alpha_E + 10,1226\alpha_K.\end{aligned}$$

Відніmemo перше рівняння від другого, маємо

$$0,3859 = 6,7826\alpha_K,$$

звідки $\alpha_K = 0,0569$. Повертаючись послідовно до отриманих раніше рівнянь, знаходимо

$$\alpha_E = 0,6556 \quad \text{і} \quad \alpha_0 = 3,2427.$$

Отже, рівняння регресії рентабельності за енергоозброєністю праці і коефіцієнтом постійності складу *ПВП* має такий вигляд:

$$\hat{P} = 3,2427 + 0,6556E + 0,0569K. \quad (3.4)$$

У цьому рівнянні регресії:

3,2427 – частка рентабельності, яка не залежить від *E* і *K* і обумовлена всіма іншими факторами, коп./грн.;

0,6556 – підвищення рентабельності в коп./грн. при зростанні енергоозброєності праці на 1 кВт/чол.;

0,0569 – підвищення рентабельності в коп./грн. при зростанні коефіцієнта постійності складу *ПВП* на 1%.

Нехай вас не дивують десяті й соті частки копійки коефіцієнтів регресії! Допустимо, що витрати підприємства складають 10 млн. грн. Тоді вони становитимуть відповідно 65,6 та 5,69 тис. грн. зростання прибутку!

З першого рівняння системи (3.2) з урахуванням того, що

$$e_j = y_j - \hat{y}_j = y_j - a_0 - a_1x_j,$$

впливає, що сума помилок (залишків) e_j , а значить і їх середньоарифметичне значення дорівнюють нулю. Отже, $\sum e_j = 0$ і тоді $\sum \hat{y}_j = \sum y_j$. Це дуже важлива і необхідна ознака правильності оцінювання параметрів рівняння регресії за методом найменших квадратів.

Спосіб детермінантів

Спосіб детермінантів (визначників) заснований на теорії матричної алгебри і пов'язаний з розрахунком визначників, кількість яких на одиницю перевищує кількість параметрів, що оцінюються.

Побудуємо матрицю числових коефіцієнтів правої частини системи нормальних рівнянь за МНК (3.3). У нашому наскрізному прикладі вона має вигляд

$$A = \begin{bmatrix} 29 & 136,6 & 1740 \\ 136,6 & 756,72 & 8567,7 \\ 1740 & 8567,7 & 108164 \end{bmatrix}.$$

Знайдемо визначник матриці A

$$\Delta A = \begin{bmatrix} 29 & 136,6 & 1740 \\ 136,6 & 756,72 & 8567,7 \\ 1740 & 8567,7 & 108164 \end{bmatrix} = \left. \begin{array}{l} -2,284990272 \\ -2,128759015 \\ -2018292648 \\ +2367372488 \\ +2036405207 \\ +2036405207 \end{array} \right\} = 8140967.$$

Далі послідовно замінюватимемо стовпці матриці A стовпцем вільних членів системи нормальних рівнянь за МНК, тобто

$$\begin{bmatrix} 282,6 \\ 1425,26 \\ 17413,9 \end{bmatrix}.$$

і розрахуємо визначники ΔA_0 , ΔA_1 і ΔA_2 . Покажемо визначення ΔA_0 :

$$\Delta A_0 = \begin{bmatrix} 282,6 & 136,6 & 1740 \\ 1425,26 & 756,72 & 8567,7 \\ 17413,9 & 8567,7 & 108164 \end{bmatrix} = \left. \begin{array}{l} -22868156000 \\ -20744390000 \\ -21058505000 \\ +23069637000 \\ +21247488000 \\ +20380320000 \end{array} \right\} = 26394000.$$

Замінюючи далі в матриці A другий стовпчик стовпцем вільних членів, отримуємо

$$\Delta A_1 = \begin{bmatrix} 29 & 282,6 & 1740 \\ 136,6 & 1425,26 & 8567,7 \\ 1740 & 17413,9 & 108164 \end{bmatrix} = 5337218$$

і, нарешті, аналогічно

$$\Delta A_2 = \begin{bmatrix} 29 & 136,6 & 282,6 \\ 136,6 & 756,72 & 1425,26 \\ 1740 & 8567,7 & 17413,9 \end{bmatrix} = 463221.$$

Параметри рівняння регресії, включаючи a_0 розраховуємо за формулою

$$a_i = \frac{\Delta A_i}{\Delta A}, \quad (3.5)$$

тобто

$$a_0 = 26394000/8140967 = 3,2427,$$

$$a_1 = 5337218/8140967 = 0,6556,$$

$$a_2 = 463221/8140967 = 0,0569.$$

Отже, маємо $\hat{P} = 3,2427 + 0,6556E + 0,0569K$.

Таким чином ми отримали методом детермінантів, як і сподівались, те саме рівняння регресії (3.4), що і раніше, шляхом розв'язання системи нормальних рівнянь за схемою Гаусса.

Спосіб оберненої матриці

Побудуємо матрицю, обернену до матриці A , тобто матрицю A^{-1} . Для цього до кожного елемента a_{ij} матриці A знаходимо його алгебраїчне доповнення, тобто визначники відповідних матриць 2-го порядку з певним знаком. Отже, матриця A така:

$$\begin{bmatrix} 29 & 136,6 & 1740 \\ 136,6 & 756,72 & 8567,7 \\ 1740 & 8567,7 & 108164 \end{bmatrix}.$$

Алгебраїчне доповнення до елемента a_{11} – це визначник матриці без i -го стовпця і j -го рядка (i – номер стовпця, j – номер рядка), тобто мінор з індикатором знака -1^{i+j} . Отже, для елемента $a_{11} = 29$

$$\Delta A_{11} = \begin{bmatrix} 754,72 & 8567,7 \\ 8567,7 & 108164 \end{bmatrix} = 81633534 - 73405483 = 8228051.$$

Далі розрахуємо вісім інших алгебраїчних доповнень:

до елемента $a_{12} = 136,6$:

$$\Delta A_{12} = - \begin{bmatrix} 136,6 & 1740 \\ 8567,7 & 108164 \end{bmatrix} = + 132596,$$

до елемента $a_{13} = 1740$:

$$\Delta A_{13} = \begin{bmatrix} 136,6 & 1740 \\ 754,72 & 8567,7 \end{bmatrix} = -142865,$$

до елемента $a_{21} = 136,6$:

$$\Delta A_{21} = - \begin{bmatrix} 136,6 & 8567,7 \\ 1740 & 108164 \end{bmatrix} = +132596,$$

до елемента $a_{22} = 754,72$:

$$\Delta A_{22} = \begin{bmatrix} 29 & 1740 \\ 1740 & 108164 \end{bmatrix} = 109156,$$

до елемента $a_{23} = 8567,7$:

$$\Delta A_{23} = - \begin{bmatrix} 29 & 1740 \\ 136,6 & 8567,7 \end{bmatrix} = -10779,$$

до елемента $a_{31} = 1740$:

$$\Delta A_{31} = \begin{bmatrix} 136,6 & 754,72 \\ 1740 & 8567,7 \end{bmatrix} = -142865,$$

до елемента $a_{32} = 8567,7$:

$$\Delta A_{32} = - \begin{bmatrix} 29 & 136,6 \\ 1740 & 8567,7 \end{bmatrix} = -10779,$$

до елемента $a_{33} = 108164$:

$$\Delta A_{33} = \begin{bmatrix} 29 & 136,6 \\ 136,6 & 754,72 \end{bmatrix} = 3227,$$

Обернена матриця A^{-1} визначається діленням матриці алгебраїчних доповнень до елементів матриці A на її визначник ΔA :

$$A^{-1} = \frac{1}{\Delta A} [\Delta A_{ij}]. \quad (3.6)$$

У нашому прикладі обернена матриця A^{-1} набуває вигляду

$$A^{-1} = \frac{1}{8140967} \times \begin{bmatrix} 8228051 & 132596 & -142865 \\ 132596 & 109156 & -10779 \\ -142865 & -10779 & 3227 \end{bmatrix} = \begin{bmatrix} 1,010697 & 0,016288 & -0,017549 \\ 0,016288 & 0,013408 & -0,001324 \\ -0,017549 & -0,001324 & 0,000396 \end{bmatrix}.$$

Отже, обернена матриця A^{-1} така

$$A^{-1} = \begin{bmatrix} 1,010697 & 0,016288 & -0,017549 \\ 0,016288 & 0,013408 & -0,001324 \\ -0,017549 & -0,001324 & 0,000396 \end{bmatrix}.$$

За допомогою оберненої матриці A^{-1} можна дуже просто визначити коефіцієнти регресії. Для цього достатньо помножити обернену матрицю A^{-1} на стовпець вільних членів системи нормальних рівнянь за МНК.

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = A^{-1} \begin{bmatrix} \sum P \\ \sum P\phi \\ \sum P \end{bmatrix}. \quad (3.7)$$

У нашому прикладі такий розрахунок a_i виглядає так:

$$\begin{bmatrix} a_0 \\ a_E \\ a_K \end{bmatrix} = \begin{bmatrix} 1,010697 & 0,016288 & -0,017549 \\ 0,016288 & 0,013408 & -0,001324 \\ -0,017549 & -0,001324 & 0,000396 \end{bmatrix} \times \begin{bmatrix} 282,6 \\ 1425,26 \\ 17413,9 \end{bmatrix}.$$

Отже,

$$a_0 = 1,010697 \times 282,6 + 0,016288 \times 1425,26 - 0,017549 \times 17413,9 = 3,2427,$$

$$a_E = 0,016288 \times 282,6 + 0,013408 \times 1425,26 - 0,001324 \times 17413,9 = 0,6556,$$

$$a_K = -0,017549 \times 282,6 - 0,001324 \times 1425,26 + 0,000396 \times 17413,9 = 0,0569.$$

Ми отримали, як і сподівались, ті самі коефіцієнти регресії, як і раніше в цьому підрозділі, а також і рівняння регресії

$$\hat{P} = 3,2427 + 0,6556E + 0,0569K.$$

Спосіб β -коефіцієнтів

Для розрахунку коефіцієнтів регресії a_i можна скористатися β -коефіцієнтами, що зазвичай розраховуються раніше:

$$a_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}. \quad (3.8)$$

У нашому прикладі $\beta_E = 0,6963$, $\beta_K = 0,3513$, (див. підрозділ 2.1.6), $\sigma_p = 1,8449$, $\sigma_E = 1,9591$, $\sigma_K = 11,3927$ (див. табл 1.5). За цими даними за формулою (3.8) маємо

$$a_E = 0,6963 \frac{1,8449}{1,9591} = 0,6556,$$

$$a_K = 0,3513 \frac{1,8449}{11,3927} = 0,0569.$$

Діленням першого рівняння системи рівнянь для розрахунку параметрів a_i на n маємо

$$\bar{y} = \alpha_0 + \alpha_1 \bar{x}_1 + \alpha_2 \bar{x}_2 + \dots + \alpha_m \bar{x}_m,$$

звідки знаходимо

$$\alpha_0 = \bar{y} - \alpha_1 \bar{x}_1 - \alpha_2 \bar{x}_2 - \dots - \alpha_m \bar{x}_m. \quad (3.9)$$

У нашому прикладі (табл 1.4)

$$\alpha_0 = 9,7448 - 0,6556 \times 4,7103 - 0,0569 \times 60 = 3,2427.$$

Ми отримали за формулами (3.8) і (3.9) оцінки параметрів рівняння регресії, які ідентичні оцінкам, визначеним шляхом вирішення системи рівнянь за методом найменших квадратів будь-яким із наведених раніше способів.

Це і не дивно! Адже β -коефіцієнти є параметрами рівняння регресії, розрахованими методом найменших квадратів за стандартизованими значеннями y' та x'_i :

$$y' = \frac{y - \bar{y}}{\sigma_y}, \quad x'_i = \frac{x_i - \bar{x}}{\sigma_x}.$$

У стандартизованому масштабі одиницею виміру змінних, як бачимо, є середньоквадратичні відхилення цих змінних. Тому і зв'язок між a_i і β_i такий:

$$a_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}, \quad \beta_i = a_i \frac{\sigma_{x_i}}{\sigma_y}.$$

3.1.4. Виконання за МНК основних припущень щодо оцінювання параметрів

У табл. 3.1 для нашого наскрізного прикладу за економетричною моделлю рентабельності (3.3) представлені розрахунки оцінок рентабельності й визначені помилки оцінок (залишки). Отримані результати, по-перше, підтверджують формальну правильність оцінок параметрів рівняння регресії тому, що $\sum P = \sum \hat{P} = 282,6$ і $\sum e_j = 0$, і, по друге, вони знадобляться нам у подальших оцінках якості отриманої моделі.

Таблиця 3.1 – Розрахунок оцінок рентабельності за моделлю $\hat{P} = 3,2427 + 0,6556E + 0,0569K$

<i>j</i>	<i>P</i>	<i>E</i>	<i>K</i>	0,6556E	0,0569K	\hat{P}	<i>e_j</i>
1	6,7	1,5	46	0,98	2,62	6,84	-0,14
2	7,3	1,7	49	1,12	2,79	7,15	0,15
3	11,5	6,4	68	4,20	3,87	11,31	0,19
...
29	12,0	7,1	56	4,66	3,19	11,10	0,90
30	9,5	5,2	60	3,41	3,41	10,06	-0,56
\sum	282,6	-	-	-	-	282,60	0

Викладений і проілюстрований на прикладі метод найменших квадратів дозволяє отримати незміщені, обгрунтовані й ефективні оцінки, тобто найкращі BLUE-оцінки параметрів рівняння регресії (див. підрозділ 3.1), але тільки за умови виконання припущень, викладених в підрозділі 3.2.

Припущення 1 для лінійної і квазілінійної регресії виконується автоматично, а для суттєво нелінійної – лише шляхом лінеаризації рівняння регресії.

Припущення 2 ($E(\varepsilon) = 0$) завдяки властивостям методу найменших квадратів також не порушується. Такі ж висновки можна зробити відносно припущень 5, 6. Припущення 7 про некорельованість незалежних змінних в принципі порушується завжди. Але, якщо за γ -тестом вилучити сильно корельовані змінні x_i , які

можуть мати ненадійні і навіть алогічні оцінки параметрів a_i , явище мульти-колінеарності можна достатньо суттєво усунути (див. підрозділи 2.1.5 і 2.1.6).

Отже, лишаються припущення 3 щодо гомоскедастичності і припущення 4 щодо відсутності автокореляції залишків e_j .

3.1.5. Гетероскедастичність

Нагадаємо, що припущення 3 про гомоскедастичність залишків e_j полягає в тому, що $D_e = \sigma_e^2 = \text{const}$, тобто що $D_e \neq f(x_i)$.

Досить простим і наочним методом тестування припущення про наявність або відсутність гомоскедастичності є візуальний аналіз полів кореляції. На рис. 3.2а спостерігається гомоскедастичність, а на рис. 3.2б – найбільш поширений вид гетероскедастичності. Очевидно, що в останньому випадку $\sigma_e^2 = f(x)$.

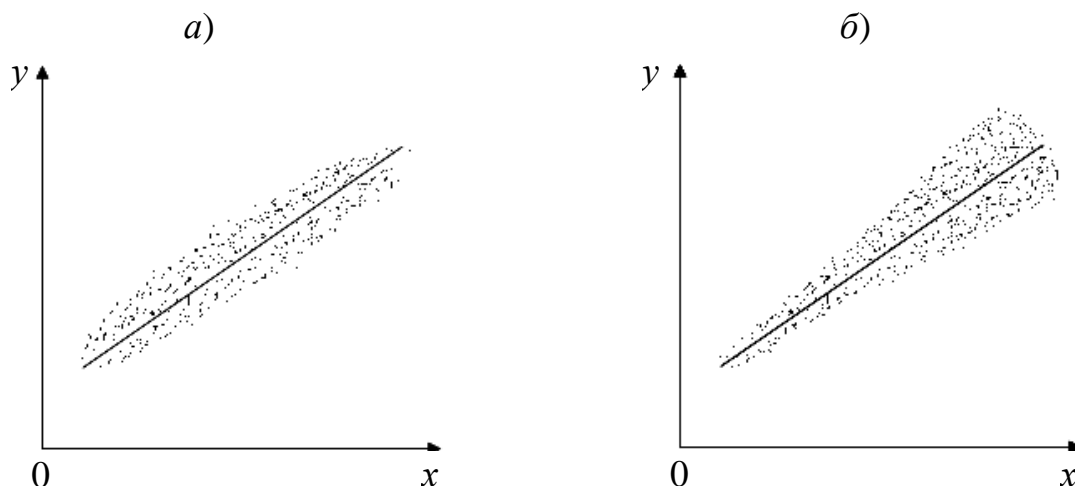


Рис. 3.2 – Графічні ознаки відсутності й наявності гетероскедастичності

Серед аналітичних тестів наявності (відсутності) гетероскедастичності найпростішим є тест рангової кореляції Спірмана

$$r_s = 1 - \frac{6 \times \sum_{j=1}^n d_j^2}{n \times (n^2 - 1)}, \quad (3.10)$$

де d_i – різниця в рангах незалежної змінної x_j і залишку e_j .

Для множинної кореляції, виходячи, що x_i це x_1, x_2, \dots, x_m і те, що $y = f(x_1, x_2, \dots, x_m)$, замість x_i можна без суттєвої помилки ранжувати y_j .

Спочатку ранжуємо y_j від найменшого до найбільшого значень і присвоюємо їм ранги від 1 до n . Потім визначаємо за тим же принципом ранги відповідних e_j і, нарешті, знаходимо різницю в рангах для кожної пари y_j і e_j , тобто d_j .

Для нашого наскрізного прикладу такі розрахунки наведені в табл. 3.2.

Таблиця 3.2 – Розрахунок d_j для визначення коефіцієнта рангової кореляції Спірмана

j	P_j	Ранг P_j	e_j	Ранг e_j	d_j	d_j^2
1	6,7	2	-0,14	5	-3	9
2	7,3	3	0,15	6	-3	9
3	11,5	24	0,19	10	14	196
...
29	12,0	26	0,90	24	2	4
30	9,5	14	-0,56	18	-4	16
Σ	282,6	-	0	-	-	3492

За даними табл. 3.2 коефіцієнт рангової кореляції у нашому прикладі становить за формулою (3.10)

$$r_s = 1 - \frac{6 \times 3492}{29 \times (29^2 - 1)} = 0,14.$$

Перевіримо значущість отриманого коефіцієнта рангової кореляції за t -статистикою Стюдента. Для цього визначимо t -статистику за формулою

$$t = \frac{r_s \times \sqrt{n-2}}{\sqrt{1-r_s^2}}, \quad (3.11)$$

де r_s – коефіцієнт рангової кореляції, і порівняємо його з критичним за таблицею t -розподілу Стюдента (додаток 3). Якщо $t \leq t_{кр}$, то коефіцієнт рангової кореляції є незначущим і гетероскедастичність залишків e_j вважається відсутньою і навпаки.

У нашому прикладі за формулою (3.11) маємо

$$t = \frac{0,14 \times \sqrt{29-2}}{\sqrt{1-0,14^2}} = \frac{0,727}{0,990} = 0,734.$$

За додатком 3 визначаємо при числі ступенів свободи $29 - 2 - 1 = 26$ і $P = 0,95$, що $t_{кр} = 2,05$. Отже, залишки e_j в моделі (3.3) гомоскедастичні, тому що $0,734 < 2,05$. Таким чином, за тестуванням на наявність гетероскедастичності параметри моделі (3.3) незміщені, обгрунтовані і ефективні.

3.1.6. Автокореляція

Як нам уже відомо, за припущенням 4 залишки e_j є випадковими незалежними між собою величинами, ще мають нульове очікування. Порушення цього припущення називається автокореляцією. Вона призводить до того, що оцінки a_i лінійної моделі, знайдені за методом найменших квадратів, будуть незміщеними, обгрунтованими, але неефективними. Автокореляція може виникати, якщо залежність описується лінійною моделлю, а насправді зв'язок у з пояснюючими змінними x_i має нелінійний характер. Таким випадкам можна протиставити більш адекватну форму специфікації моделі. Іншою причиною корельованості залишків може бути дія якогось суттєвого фактора, не включеного в модель і який може не випадковим чином впливати на величину відхилень e_j , наприклад, сезонні умови. Такі фактори можна знайти. Серйозною причиною автокореляції залишків є інерційність економічних процесів у часі, коли економетрична модель будується на основі часових рядів.

Автокореляція залишків – це кореляція залишків з тими ж залишками, що взяті з деяким запізненням, тобто це кореляція ряду e_1, e_2, \dots, e_n з рядом $e_{1-k}, e_{2-k}, \dots, e_{n-k}$. Кореляція між сусідніми членами ряду ($k = 1$) називається автокореляцією першого порядку. Найпоширенішим тестом на наявність автокореляції залишків першого порядку є тест Дарбіна-Уотсона, який ґрунтується на статистиці

$$DW = \frac{\sum_{j=2}^n (e_j - e_{j-1})^2}{\sum_{j=1}^n e_j^2}. \quad (3.12)$$

Доведено, що $DW \approx 2$ і може набувати значення в проміжку (0-4). При відсутності автокореляції ($r_{e_j e_{j-1}} = 0$) статистика DW набуває значень, близьких до 2, при додатній автокореляції ($0 < r_{e_j e_{j-1}} < 1$) – значення від 0 до 2, при від’ємній ($-1 < r_{e_j e_{j-1}} < 0$) – від 2 до 4.

Перевірка наявності або відсутності автокореляції залишків проводиться шляхом порівняння фактичного значення DW з критичними, які знаходимо за спеціальною таблицею залежно від рівня значущості P , кількості факторів m і кількості спостережень n (додаток 5). Таблиця містить два критичних значення: нижнє d_n для визначення додатної автокореляції і верхнє d_e – для визначення від’ємної автокореляції (рис. 3.3).

додатна автокореляція	?	автокореляція відсутня	?	від’ємна автокореляція
0	d_n	d_e	2	$4 - d_n$
			$4 - d_e$	4

Рис. 3.3 – Тест Дарбіна-Уотсона

Перевіримо за допомогою тесту Дарбіна-Уотсона існування автокореляції залишків першого порядку для економетричної моделі рентабельності витрат (3.3).

Таблиця 3.3 – Попередні розрахунки для визначення статистики DW

j	P_j	\hat{P}_j	e_j	e_j^2	$e_j - e_{j-1}$	$(e_j - e_{j-1})^2$
1	6,7	6,84	-0,14	0,020	–	–
2	7,3	7,15	0,15	0,021	0,001	0
3	11,5	11,31	0,19	0,036	0,015	0
...
29	12,0	11,10	0,90	0,81	0,798	0,637
30	9,5	10,06	-0,56	0,314	-0,496	0,246
Σ	282,6	282,6	0	10,673	–	11,936

У табл. 3.3 наведені фактичні значення рентабельності за спостереженнями P , значення оцінок рентабельності \hat{P} , розрахованих за економетричною моделлю (3.3), значення залишків $e_j = P_j - \hat{P}_j$, й відповідні суми квадратів, які необхідні для обчислення фактичного значення статистики DW . Підставляючи їх у формулу (3.12), отримуємо

$$DW = \frac{11,936}{10,673} = 1,118.$$

Порівняємо це значення з табличними (див. додаток б) для рівня значущості 0,99, кількості факторів $m = 2$ та кількості спостережень $n = 29$. Критичні значення DW такі: $d_n = 1,05$ – нижня межа, $d_g = 1,33$ – верхня межа. Оскільки фактичне значення DW задовольняє умову (див. рис. 3.3) $d_n < DW < d_g$ ($1,05 < 1,118 < 1,33$), то приймаємо гіпотезу щодо невизначеності наявності автокореляції залишків в економетричній моделі (3.3). Отже немає достатніх підстав брати під сумнів незміщеність, обгрунтованість та ефективність МНК – оцінок параметрів регресійного рівняння рентабельності (3.3), тобто

$$\hat{P} = 3,2427 + 0,6556E + 0,0569K.$$

Виникає запитання: що робити, коли виявлена гетероскедастичність або автокореляція залишків? Параметри регресії можуть бути незміщеними і обгрунтованими, але не ефективними, тобто дисперсія залишків буде великою.

Можна дати декілька порад. По-перше, можна віднайти якийсь новий фактор, який апроксиміював би явище гетероскедастичності або автокореляції. По-друге, їх негативний вплив можна зменшити шляхом підбору більш адекватної форми регресії, або перетворенням змінних. Ці дві поради пов'язані зі зміною специфікації моделі. По-третє, для оцінювання параметрів рівняння регресії треба застосовувати узагальнений метод найменших квадратів (метод Ейткена).

3.1.7. Значущість (адекватність) рівняння регресії

Для перевірки значущості (адекватності) рівняння регресії використовують F -статистику Фішера, або t -статистику Стьюдента, що еквівалентно (бо $F = t^2$).

У випадках парної (простой) регресії твердження про значущість (адекватність) моделі справедливе, якщо

$$F \geq F_{\text{крит}},$$

де

$$F = S_{\hat{y}} / S_e \quad (3.13)$$

і, відповідно,

$$S_{\hat{y}} = \frac{\sum_1^n (\hat{y}_j - \bar{y})^2}{1}, \quad S_e = \frac{\sum_1^n e_j^2}{n-2}. \quad (3.14)$$

Для оцінки F -відношення порівнюємо з критичним значенням за таблицею F -розподілу Фішера (додаток 5) при ступенях свободи $\kappa_1 = n$ і $\kappa_2 = n - 2$.

Наприклад, $S_{\hat{y}} = 86,35$, $S_e = 2,10$, $n = 26$. Тоді

$$F = \frac{86,35}{2,10} = 41,12,$$

а $F_{\text{крит}}$ при $\kappa_1 = 1$, $\kappa_2 = 26 - 2 = 24$ і $P = 0,95$ становить 4,26. Отже рівняння регресії адекватне, оскільки $41,12 > 4,26$.

У випадках багатofакторної регресії значущість (адекватність) моделі перевіряється аналогічно, але при ступенях свободи для чисельника $\kappa_1 = m$ і $\kappa_2 = n - m - 1$ – для знаменника, тоді формули (3.14) набувають вигляду

$$S_{\hat{y}} = \frac{\sum_1^m (\hat{y}_j - \bar{y})^2}{m}, \quad S_e = \frac{\sum e_j^2}{n-m-1}. \quad (3.15)$$

Перевіримо побудовану нами двофакторну економетричну модель рентабельності (3.3) на значущість (адекватність) за формулами (3.13) і (3.15).

Для розрахунку $S_{\hat{y}}$ попередньо знайдемо $\sum (\hat{y}_j - \bar{y})^2$. Використаємо співвідношення дисперсійного ANOVA-аналізу

$$\sum_1^n (y - \bar{y})^2 = \sum_1^n (\hat{y} - \bar{y})^2 + \sum_1^n (y - \hat{y})^2. \quad (3.16)$$

Із табл 1.5 знаходимо $D_p = 3,4037$, а з табл 3.3 $\sum_1^n e_j^2 = 10,673$. Враховуючи, що $n = 29$, маємо

$$\sum_1^n (\hat{y} - \bar{y})^2 = 3,4037 \times 29 - 10,683 = 98,707 - 10,683 = 88,024.$$

До речі, за даними ANOVA-аналізу можна визначити і коефіцієнт множинної кореляції за формулою

$$R_{1,2,\dots,m} = \sqrt{1 - \frac{\sum e_j^2}{\sum (y - \bar{y})^2}}. \quad (3.17)$$

У нашому прикладі він становить

$$R_{1,2} = \sqrt{1 - \frac{10,683}{98,707}} = \sqrt{1 - 0,1082} = 0,9434,$$

що співпадає з раніше визначеним за формулами (2.8) і (2.10).

Повернемося до перевірки адекватності рівняння регресії (3.3). За формулами (3.15) отримуємо

$$S_{\hat{y}} = \frac{88,024}{2} = 44,012, \quad S_e = \frac{10,673}{29 - 2 - 1} = 0,411,$$

отже $F = \frac{44,012}{0,411} = 107,08$.

За додатком 5 при $\kappa_1 = m = 2$ і $\kappa_2 = n - m - 1 = 29 - 2 - 1 = 26$ знаходимо $F_{крит} = 3,38$. Робимо висновок, що побудована нами модель адекватна, оскільки $107,88 > 3,38$ при $P = 0,95$.

Між F -критерієм Фішера і вже засвоєним нами коефіцієнтом детермінації $R^2(d)$ існує простий зв'язок. Доведено, що

$$F = \frac{R_{1,2,\dots,m}^2}{1 - R_{1,2,\dots,m}^2} \frac{n - m - 1}{m}. \quad (3.18)$$

У нашому наскрізному прикладі за формулою (3.18) маємо розрахункове значення F

$$F = \frac{0,9434^2}{1 - 0,9434^2} * \frac{29 - 2 - 1}{2} = 107,08,$$

як і раніше.

3.1.8. Перевірка значущості параметрів моделі

Для простої (парної) регресії значущість параметрів a_0 і a_1 перевіряється за допомогою t -статистики Стьюдента:

$$t_{a_0} = \frac{a_0}{\sigma_{a_0}}, \quad t_{a_1} = \frac{a_1}{\sigma_{a_1}}, \quad (3.19)$$

де

$$\sigma_{a_0} = \frac{\overline{x^2} \times \sum e^2}{\sigma_x^2 \times n \times (n - 2)}, \quad \sigma_{a_1} = \frac{\sigma_e^2}{\sigma_x^2 \times (n - 2)}. \quad (3.20)$$

Для множинної регресії значущість параметрів a_i також перевіряється за допомогою t -статистики Стьюдента (3.19) з визначенням середньо-квадратичних помилок коефіцієнтів регресії за формулою

$$\sigma_{a_i} = \frac{\sigma_y \times \sqrt{1 - R_{12\dots m}^2}}{\sigma_{x_i} \times \sqrt{1 - R_{12\dots m(\text{без } i)}^2} \times \sqrt{n}}. \quad (3.21)$$

Визначимо σ_{a_i} для побудованої нами економетричної моделі рентабельності. Попередньо за табл. 1.5 знайдемо $\sigma_P = 1,8449$, $\sigma_E = 1,9591$, $\sigma_K = 11,3927$, $r_{PE} = 0,8981$, $r_{PK} = 0,7513$; тоді

$$\sigma_{a_E} = \frac{1,8449 \times \sqrt{1 - 0,9434^2}}{1,9591 \times \sqrt{1 - 0,7513} \times \sqrt{29}} = 0,0876,$$

$$\sigma_{a_K} = \frac{1,8449 \times \sqrt{1 - 0,9434^2}}{11,3927 \times \sqrt{1 - 0,8981} \times \sqrt{29}} = 0,0227.$$

На жаль такого «алгебраїчного» способу для розрахунку помилки параметра a_0 не існує. Таку можливість дає застосування методу оберненої матриці. Нагадаємо, що для нашого прикладу вона така (3.7)

$$A^{-1} = \begin{bmatrix} 1,010697 & 0,016288 & -0,017549 \\ 0,016288 & 0,013408 & -0,001324 \\ -0,017549 & -0,001324 & 0,000396 \end{bmatrix}.$$

Елементи головної діагоналі оберненої матриці дозволяють дуже просто розрахувати середньоквадратичні відхилення (помилки) коефіцієнтів регресії $a_i (i = \overline{0, m})$ за формулою

$$\sigma_{a_i} = \sqrt{\frac{\sum_1^n e_j^2}{n - m - 1}} \times b_{ij}, \quad (i = j), \quad (3.22)$$

де b_{ij} – елементи головної діагоналі оберненої матриці A^{-1} .

У нашому прикладі, враховуючи, що

$$\frac{\sum_1^n e_j^2}{n - m - 1} = \frac{10,673}{29 - 2 - 1} = 0,411,$$

помилки коефіцієнтів регресії за формулою (3.22) відповідно складають:

$$\sigma_{a_0} = \sqrt{0,411 \times 1,010697} = 0,644,$$

$$\sigma_{a_E} = \sqrt{0,411 \times 0,013408} = 0,0876,$$

$$\sigma_{a_K} = \sqrt{0,411 \times 0,000396} = 0,0227,$$

що співпадає з раніше визначеними помилками параметрів σ_{a_E} і σ_{a_K} .

Далі за формулою (3.19) визначаємо t -статистики Стьюдента:

$$t_0 = \frac{3,2427}{0,644} = 5,035; \quad t_E = \frac{0,6556}{0,0876} = 7,527; \quad t_K = \frac{0,0569}{0,0227} = 2,529.$$

За додатком 3 при $29 - 2 - 1 = 26$ і $P = 0,95$ знаходимо, що $t_{крит} = 2,048$. Отже, є підстави вважати, що всі параметри рівняння регресії (3.3) статистично значущі: $5,035, 7,527$ і $2,529 > 2,048$.

3.1.9. Інтервали довіри до коефіцієнтів регресії

Коефіцієнти регресії a_i , як і коефіцієнти кореляції, мають t -розподіл Стьюдента. Тому довірчі інтервали для невідомих нам істинних коефіцієнтів регресії a_i визначаються так:

$$a_i - \sigma_{a_i} \leq \alpha_i \leq a_i + \sigma_{a_i}. \quad (3.23)$$

У нашому прикладі $a_E = 0,6556$, $\sigma_{a_E} = 0,0876$, $a_K = 0,0569$, $\sigma_{a_K} = 0,0227$, $a_0 = 3,2427$, $\sigma_{a_0} = 0,644$. Отже, можна стверджувати, що істинні коефіцієнти регресії знаходяться в межах

$$0,6556 - 0,0876 \leq \alpha_E \leq 0,6556 + 0,0876,$$

$$0,0569 - 0,0227 \leq \alpha_K \leq 0,0569 + 0,0227,$$

$$3,2427 - 0,644 \leq \alpha_0 \leq 3,2427 + 0,644$$

або

$$0,568 \leq \alpha_E \leq 0,743,$$

$$0,0333 \leq \alpha_K \leq 0,0796,$$

$$2,599 \leq \alpha_0 \leq 3,887.$$

Зауважимо, що коли інтервал довіри включає число «нуль», відповідний фактор безумовно незначущий, бо помилка коефіцієнта регресії перевищує його величину. Зауважимо також, що такі випадки не трапляються, якщо на стадії ідентифікації фактори тестувалися на автономність впливу (див. підрозділ 2.1.7).

3.2. ПРОГНОЗУВАННЯ ЗАЛЕЖНОЇ ЗМІННОЇ

3.2.1. Прогнозування на парних моделях

Парну (просту) лінійну регресію

$$\hat{y} = a_0 + a_1x$$

після перевірки її на адекватність і значущість параметрів a_0 і a_1 (див. підрозділ 3.1.8) ми можемо використовувати для прогнозування (розрахунків) залежної змінної y для будь-яких прогнозних значень змінної x . Слід мати на увазі, що

рівняння регресії надійно відображає залежність в межах абсолютного розмаху варіації x ($x_{min} \div x_{max}$). Тому при прогнозних значеннях x , що виходять за ці межі, надійність прогнозу знижується. Отже, інтерполяція за регресією незаперечна, а екстраполяція має бути обережною.

Розрізняють два типи прогнозів за рівнянь регресії: точкові й інтервальні.

Точковий прогноз залежної змінної \hat{y}_n визначають шляхом підстановки у рівняння регресії прогнозного значення фактора x_n :

$$\hat{y} = a_0 + a_1 x_n.$$

Прогнозне значення \hat{y}_n є точковою оцінкою дійсного значення y_n , яке, як відомо, дорівнює

$$\hat{y}_n = \alpha_0 + \alpha_1 x_n + \varepsilon,$$

де ε – значення неспостережуваної випадкової величини.

Оскільки дійсне значення y_n відрізняється від прогнозного, то необхідна побудова **інтервального прогнозу** – інтервалу довіри для \hat{y}_n , в якому із заданою ймовірністю буде знаходитися дійсне значення y_n .

Інтервал довіри Δy_n для математичного сподівання \hat{y}_n визначається за формулою

$$\sigma_{\hat{y}_g} = \pm \sigma_e \sqrt{\frac{1}{n} + \frac{(x_n - \bar{x})^2}{\sum_1^n (x_j - \bar{x})^2}}, \quad (3.24)$$

де, як і раніше, σ_e – середньоквадратичне відхилення випадкової величини e . Нагадаємо, що

$$\sigma_e = \sqrt{\frac{\sum_1^n e_j^2}{n-2}}.$$

Розглянемо умовний приклад. Рівняння регресії має такий вигляд:

$$\hat{y} = 5,67 + 2,72x;$$

а $n = 22$; $\sigma_e = 0,75$; $\bar{x} = 8,06$; $\sum (x_j - \bar{x})^2 = 22,3$. Тоді, для прогнозного значення $x_n = 10$ точковий прогноз \hat{y} становить

$$\hat{y} = 5,67 + 2,72 \times 10 = 32,87.$$

Інтервал довіри за формулою (3.24) дорівнює

$$\sigma \hat{y} = 0,75 \sqrt{\frac{1}{22} + \frac{(10 - 8,06)^2}{22,3}} = 0,35.$$

Отже, дійсне значення y з ймовірністю 0,95 має знаходитися в межах

$$32,87 - 0,35 \leq y \leq 32,87 + 0,35,$$

або

$$32,52 \leq y \leq 33,22.$$

Зауважимо, що помилка прогнозу $\sigma_{\varepsilon n}$ не постійна. Як видно з формули (3.24), в міру збільшення або зменшення прогнозних значень x_n відносно \bar{x} помилка прогнозу зростає. Наприклад, при $x_n = 15$ помилка прогнозу зростає за формулою (3.24) до 1,11. якщо $x_n = 5$, то вона також буде більшою, а саме 0,51. Найменша помилка прогнозу буде для $x_n = \bar{x} = 8,06$; в цьому разі вона дорівнює всього 0,16.

Отже, на віддалених ділянках варіації незалежної змінної помилки прогнозу залежної змінної зростають, а довіра до прогнозу зменшується (рис. 3.4).

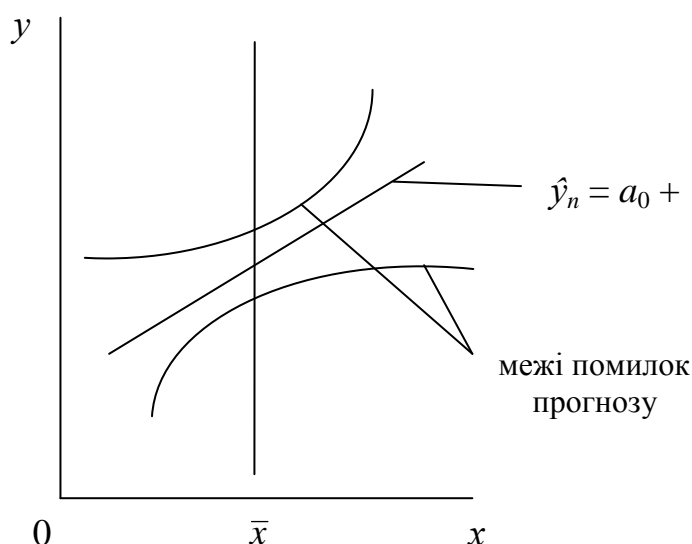


Рис. 3.4 – Межі помилок прогнозу для парної лінійної регресії

Інтервали довіри до прогнозу \hat{y}_n будуються заздалегідь шляхом варіювання x_n в розумних межах. Це усуває необхідність визначати інтервали довіри в кожному окремому випадку прогнозування залежної змінної.

3.2.2. Прогнозування на множинних моделях

Множинна лінійна регресія

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m$$

містить оцінки коефіцієнтів регресії, які, як ми вже знаємо, є випадковими величинами, тому мають дисперсію і середньоквадратичні відхилення (помилки). Отже оцінка y_n є також випадковою величиною з дисперсією $\sigma_{\hat{y}}^2$. Тому і у випадках множинної регресії є потреба визначати точковий прогноз залежної змінної та інтервали довіри для нього, тобто інтервальний прогноз.

Точковий прогноз \hat{y} визначається залежно від прогнозних значень x_{in} за економетричною моделлю.

Допустимо, що в нашому наскрізному прикладі ми маємо зробити прогноз рентабельності витрат, якщо відомі прогнози енергоозброєності праці ($E = 7$ кВт/чол.) й коефіцієнта постійності ПВП ($K = 65\%$). Тоді за моделлю (3.3) точковий прогноз рентабельності дорівнюватиме

$$\hat{P}_{II} = 3,2427 + 0,6556 \times 7 + 0,0569 \times 65 = 11,53 \text{ (коп./грн.)}.$$

Довірча межа для прогнозу y за рівнянням множинної регресії визначається за формулою (у матричному записі).

$$\sigma_{\hat{y}}^2 = \sigma_e^2 [X'] [A^{-1}] [X], \quad (3.25)$$

де $\sigma_{\hat{y}}^2$ – оцінка дисперсії випадкової величини ε ;

$[X']$ – матриця-рядок прогнозних значень незалежних змінних $[1 \quad X_{1II} \quad X_{2II} \quad \dots \quad X_{mII}]$;

$[A^{-1}]$ – обернена матриця числових коефіцієнтів правої частини системи нормальних рівнянь;

$[X]$ – матриця-стовпець прогнозних значень незалежних змінних

$$\begin{bmatrix} 1 \\ x_{1n} \\ x_{2n} \\ \dots \\ x_{mn} \end{bmatrix}.$$

Розглянемо застосування формули (3.25) для визначення прогнозу рентабельності за моделлю (3.3). Нагадаємо, що $\sigma_e^2 = 0,411$, $x_{EII} = 7$, $x_{KII} = 65$, обернена матриця $[A^{-1}]$ за (3.7) така:

$$\begin{bmatrix} 1,010697 & 0,016288 & -0,017549 \\ 0,016288 & 0,013408 & -0,001324 \\ -0,017549 & -0,001324 & 0,000396 \end{bmatrix}.$$

Отже, за формулою (3.25) довірна межа прогнозу \hat{y}_n дорівнює

$$\sigma_{\hat{p}}^2 = 0,411 \times [1 \quad 7 \quad 65] \times \begin{bmatrix} 1,010697 & 0,016288 & -0,017549 \\ 0,016288 & 0,013408 & -0,001324 \\ -0,017549 & -0,001324 & 0,000396 \end{bmatrix} \times \begin{bmatrix} 1 \\ 7 \\ 65 \end{bmatrix}.$$

Спочатку помножимо $[A^{-1}]$ на $[X]$, отримаємо

$$\begin{bmatrix} 1,010697 \times 1 + 0,01628 \times 7 + (-0,017549) \times 65 \\ 0,016288 \times 1 + 0,013408 \times 7 + (-0,001324) \times 65 \\ -0,017549 \times 1 + (-0,001324) \times 7 + 0,000396 \times 65 \end{bmatrix} = \begin{bmatrix} -0,015972 \\ 0,024084 \\ -0,001077 \end{bmatrix}.$$

Подальші наші дії такі

$$\begin{aligned} \sigma_{\hat{p}}^2 &= 0,411 \times [1 \quad 7 \quad 65] \times \begin{bmatrix} -0,015972 \\ 0,024084 \\ -0,001077 \end{bmatrix} = \\ &= 0,411 \times [1 \times (-0,015972) + 7 \times 0,024084 + 65 \times (-0,001077)] = \\ &= 0,411 \times (-0,015972 + 0,168588 - 0,070005) = 0,034. \end{aligned}$$

Отже, середньоквадратичне відхилення (помилка) прогнозу рентабельності складає

$$\sigma_{\hat{p}} = \sqrt{\sigma_{\hat{p}}} = \sqrt{0,034} = 0,184 \text{ коп./грн.}$$

Таким чином, ми визначили інтервальний прогноз рентабельності в межах

$$11,53 - 0,18 \leq P_n \leq 11,53 + 0,18,$$

або

$$11,35 \leq P_n \leq 11,71.$$

Як бачимо, розрахунок інтервального прогнозу досить трудомістке завдання. У практичній роботі для таких розрахунків є немала низка ліцензованих комп'ютерних програмних комплексів для економетричного моделювання, що включають використання матричної алгебри.

На практичних заняттях в лабораторіях інформаційних технологій є можливість опанувати ППП STATGRAPHICS, MEZOZABP, тощо.

3.3. КОМПЛЕКС КОНТРОЛЬНИХ ЗАВДАНЬ

ЗА ЗМ 3 «ОЦІНЮВАННЯ ПАРАМЕТРІВ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ»

Навчальні елементи, що підлягають контролю і оцінюванню:

вимоги до оцінювання параметрів; основні припущення щодо застосування МНК; МНК: способи Гауса, детермінантів, оберненої матриці; оцінювання параметрів за β -коефіцієнтами; гетероскедистичність: ранговий тест Спірмана; автокореляція: тест Дарбіна-Уотсона; узагальнений метод найменших квадратів (метод Ейткена); значущість та інтервали довіри для параметрів рівняння регресії; точкове і інтервальне прогнозування ендогенної змінної.

3.3.1. Тестові завдання

T3.01. Параметри рівняння регресії є незміщеними, коли:

$$\text{а) } E(a_i) = \alpha_i; \quad \text{б) } a_{i(n \rightarrow \infty)} \rightarrow \alpha_i; \quad \text{в) } \sum e_i^2 \rightarrow \min.$$

T3.02. Параметри рівняння регресії є обґрунтованими, коли:

$$\text{а) } E(a_i) = \alpha_i; \quad \text{б) } a_{i(n \rightarrow \infty)} \rightarrow \alpha_i; \quad \text{в) } \sum e_i^2 \rightarrow \min.$$

T3.03. Параметри рівняння регресії є ефективними, коли:

а) $E(a_i) = \alpha_i$; б) $a_{i(n \rightarrow \infty)} \rightarrow \alpha_i$; в) $\sum e_i^2 \rightarrow \min$.

T3.04. Умовами застосування МНК для оцінювання a_i є наступні припущення:

- а) можливість лінеаризації рівняння регресії;
- б) гомоскедастичність дисперсії залишків e_j ;
- в) наявність автокореляції залишків e_j ;
- г) відсутність мультиколінеарності екзогенних змінних.

T3.05. Умовами застосування МНК для оцінювання a_i є наступні припущення:

- а) $E(a_i) = 0$;
- б) $D_\varepsilon = \text{const}$;
- в) гетероскедастичність залишків e_j ;
- г) розподіл залишків e_j за нормальним законом.

T3.06. Основним принципом МНК є:

а) $\sum (y - \hat{y}) = 0$; б) $\sum (y - \hat{y})^2 \rightarrow \min$; в) $\sum e_i^2 \rightarrow \min$.

T3.07. Розв'язок системи нормальних рівнянь за МНК для оцінювання a_i може бути виконаний:

- а) способом послідовного вилучення невідомих (способом Гаусса);
- б) способом детермінантів (визначників) за матрицею A ;
- в) за допомогою оберненої матриці A^{-1} .

T3.08. Параметри a_i можуть бути оцінені за допомогою β -коефіцієнтів за формулою:

а) $a_i = \beta_i \frac{\sigma_y}{\sigma_{xi}}$; б) $a_i = \beta_i \frac{\sigma_{xi}}{\sigma_y}$; в) $a_i = \beta_i \sigma_y \sigma_{xi}$.

T3.09. Дисперсія фактичних значень y розраховується за відхиленнями:

а) $y - \bar{y}$; б) $y - \hat{y}$; в) $\hat{y} - \bar{y}$.

T3.10. Дисперсія оцінок \hat{y} розраховується за відхиленнями:

а) $y - \bar{y}$; б) $y - \hat{y}$; в) $\hat{y} - \bar{y}$.

T3.11. Дисперсія залишків e розраховується за відхиленнями:

а) $y - \bar{y}$; б) $y - \hat{y}$; в) $\hat{y} - \bar{y}$.

ТЗ.12. Дисперсійний ANOVA-аналіз полягає у розкладанні:

- а) $y - \bar{y} = (\hat{y} - \bar{y}) + (y - \hat{y})$;
- б) $\sum (y - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2$;
- в) $D_y = D_{\hat{y}} + D_e$.

ТЗ.13. Коефіцієнт множинної кореляції за ANOVA-аналізом може бути визначений за формулою:

- а) $R^2 = D_{\hat{y}}/D_y$;
- б) $R^2 = (D_y - D_e)/D_y$;
- в) $R^2 = 1 - \sum e^2 / \sum (y - \bar{y})^2$;
- г) $R^2 = 1 - \sum (\hat{y} - \bar{y})^2 / \sum (y - \bar{y})^2$.

ТЗ.14. Значущість (адекватність) рівняння регресії перевіряється за F -статистикою Фішера:

- а) залежно від співвідношення $\sum (\hat{y} - \bar{y})^2$ і $\sum e^2$;
- б) виходячи з R^2 ;
- в) за величиною помилок оцінок a_i .

ТЗ.15. Значущість (невипадковість) коефіцієнтів регресії a_i перевіряється за t -статистикою Стьюдента:

- а) за відношенням a_i / σ_{a_i} ;
- б) за відношенням σ_{a_i} / a_i ;
- в) виходячи із $\sum e^2$, a_i і елементів головної діагоналі оберненої матриці A^{-1} .

ТЗ.16. Інтервали довіри для коефіцієнтів регресії a_i визначаються залежно від:

- а) a_i ; б) σ_{a_i} ; в) a_i та σ_{a_i} .

ТЗ.17. Наявність або відсутність гетероскедастичності залишків e_j перевіряється за допомогою:

- а) графічного тесту;
- б) тестування за ранговою кореляцією Спірмана за t -статистикою Стьюдента;
- в) теоретичного аналізу.

T3.18. *DW*-статистика Дарбіна-Уотсона:

- а) приймає значення від 0 до 1;
- б) приймає значення від 0 до 4;
- в) при середньому значенні *DW* означає повну відсутність автокореляції;
- г) при $DW \rightarrow \min$ означає наявність додатної автокореляції, а при $DW \rightarrow \max$ – від'ємної.

T3.19. Прогноз залежної змінної за рівнянням регресії може бути:

- а) точковим; б) точним; в) інтервальним.

T3.20. У простій (парній) регресії помилка прогнозу залежної змінної:

- а) мінімальна для $x_n = \bar{x}_i$;
- б) зменшується в міру віддалення x_n від \bar{x}_i ;
- в) може встановлюватися заздалегідь для будь-яких значень x_n .

T3.21. У множинній регресії помилка прогнозу залежної змінної:

- а) мінімальна для $x_n = \bar{x}_i$;
- б) збільшується в міру віддалення x_{n_i} від \bar{x}_i у будь-яку сторону;
- в) може встановлюватися заздалегідь для будь-якої комбінації прогнозних значень екзогенних змінних;
- г) повинна визначатися в кожному окремому випадку прогнозування.

T3.22. Наявність або відсутність автокореляції залишків перевіряється за допомогою:

- а) графічного тесту;
- б) тесту Дарбіна-Уотсона за *DW*-статистикою;
- в) теоретичного аналізу.

3.3.2. Логічні вправи

Л3.01. Викладіть Ваші уявлення про сутність вимоги $E(a_i) = \alpha_i$ щодо незміщеності оцінок параметрів регресії.

Л3.02. Викладіть Ваші уявлення про сутність вимоги $a_{i(n \rightarrow \infty)} \rightarrow \alpha_i$ щодо обґрунтованості оцінок параметрів регресії.

ЛЗ.03. Викладіть Ваші уявлення про сутність вимоги. $\sum e_j \rightarrow \min$, щодо ефективності оцінок параметрів регресії.

ЛЗ.04. Назовіть розробників методу найменших квадратів, дайте змістовну і графічну інтерпретацію вимоги найменших квадратів.

ЛЗ.05. Поясніть, чому при застосуванні методу найменших квадратів припускається, що:

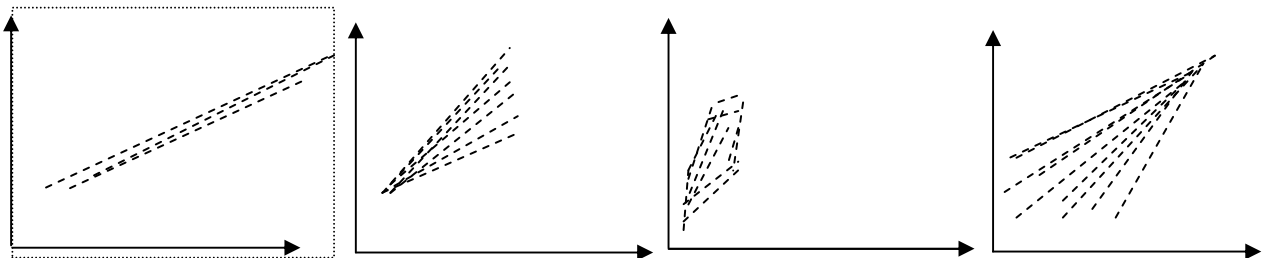
$$E(a_i) = a_i; \quad D_e = \text{const}; \quad \sum e = 0.$$

ЛЗ.06. Поясніть, чому мультиколінеарність екзогенних змінних призводить до зміщеності і необґрунтованості оцінок коефіцієнтів регресії, до їх великих помилок.

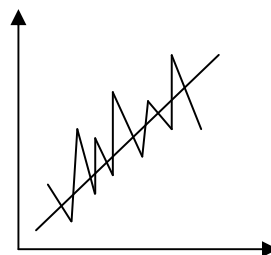
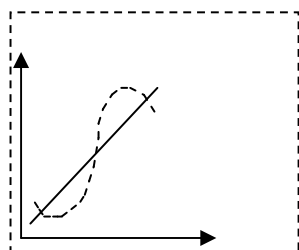
ЛЗ.07. Поясніть поняття гомоскедастичності і гетероскедастичності, наведіть приклади.

ЛЗ.08. Поясніть поняття автокореляції, наведіть приклади.

ЛЗ.09. Розгляньте поля кореляції і визначіть, на яких із них спостерігається гомоскедастичність і гетероскедастичність залишків



ЛЗ.10. Розгляньте поля кореляції, побудовані за часовими рядами. Чи вбачаєте Ви в цих випадках явище автокореляції залишків. Якщо так, поясніть у чому воно полягає.



ЛЗ.11. Поясніть, який із способів розв'язку системи за МНК є найефективнішим. Якщо це спосіб оберненої матриці, обґрунтуйте, чому.

ЛЗ.12. Викладіть Ваші уявлення про помилки оцінювання залежної змінної, тобто залишки. Якими шляхами можна їх зменшити?

ЛЗ.13. Покажіть на полі кореляції графічну інтерпретацію виникнення залишків.

ЛЗ.14. Поясніть ідею дисперсійного ANOVA-аналізу.

ЛЗ.15. Покажіть на полі кореляції відхилення, за якими обчислюються загальна, систематична та залишкова дисперсії залежної змінної. Якими факторами обумовлені ці відхилення?

ЛЗ.16. Покажіть, як за результатами ANOVA-аналізу можна обчислити коефіцієнт кореляції.

ЛЗ.17. Доведіть, що коефіцієнт кореляції за модулем не може бути менше нуля і більше одиниці. Дайте тлумачення цим граничним значенням коефіцієнта кореляції.

ЛЗ.18. Логічно доведіть, що коефіцієнт кореляції для залежності $s = pr^2$ дорівнює одиниці.

ЛЗ.19. Покажіть (на графіках), як вибір аналітичної форми регресії впливає на величину помилки апроксимації і прогнозу.

3.3.3. Розрахункові вправи

РЗ.01. Складіть систему нормальних рівнянь за МНК для розрахунку коефіцієнтів регресії a_0, a_1, a_2 :

$$\hat{y} = a_0 + a_1x_1 + \frac{a_2}{x_2}.$$

РЗ.02. Складіть систему нормальних рівнянь за МНК для розрахунку коефіцієнтів регресії a_0, a_1, a_2, a_3 :

$$\hat{y} = a_0 + \frac{a_1}{x_1} + a_2x_2 + a_3x_3^2.$$

Р3.03. За матрицею статистики зробіть необхідні розрахунки і складіть систему нормальних рівнянь за МНК для обчислення коефіцієнтів регресії.

j	y_i	x_i	j	y_i	x_i
1	15	8	4	12	6
2	8	5	5	4	3
3	4	2	–	–	–

Р3.04. За матрицею статистики зробіть необхідні розрахунки і складіть систему нормальних рівнянь за МНК для обчислення коефіцієнтів регресії.

j	y_i	x_i	j	y_i	x_i
1	6	5	4	7	6
2	8	8	5	10	3
3	14	2	–	–	–

Р3.05. Обчисліть коефіцієнти регресії та складіть рівняння парної лінійної регресії

$$45 = 5a_0 + 24a_1$$

$$252 = 24a_0 + 138a_1.$$

Р3.06. Обчисліть коефіцієнти регресії та складіть рівняння парної гіперболічної регресії

$$45 = 5a_0 + 1,325a_1$$

$$13,85 = 1,325a_0 + 0,461a_1.$$

Р3.07. Обчисліть коефіцієнти регресії та складіть рівняння множинної лінійної регресії за такими даними: $\bar{y} = 12$; $\bar{x}_1 = 7$; $\sigma_y = 5$; $\sigma_{x_1} = 3$; $\sigma_{x_2} = 2,5$, $\beta_1 = 0,6$; $\beta_2 = -0,5$.

Р3.08. Обчисліть \hat{y} при $x = 7,5$ та $x = 9$ за рівнянням регресії

$$\hat{y} = 2,137x^{0,513}.$$

Р3.09. Складіть систему рівнянь та обчисліть коефіцієнти парної лінійної регресії за умови: $\sum y = 40$; $\sum x = 26$; $\sum x^2 = 136$; $\sum yx = 245$; $n = 9$.

Р3.10. Складіть систему рівнянь та обчисліть коефіцієнти парної гіперболічної регресії за умови: $\sum y = 40$; $\sum \frac{1}{x} = 1,3$; $\sum \left(\frac{1}{x}\right)^2 = 0,45$; $\sum \frac{y}{x} = 12,5$; $n = 9$.

Р3.11. Складіть систему рівнянь для розрахунку коефіцієнтів регресії a_0, a_1, a_2, a_3

$$\hat{y} = a_0 + a_1x + a_2x^2 + a_3x^3.$$

Р3.12. Складіть систему рівнянь для розрахунку коефіцієнтів регресії a_0, a_1, a_2

$$\hat{y} = a_0x_1^{a_1}x_2^{a_2}.$$

Р3.13. Обчисліть коефіцієнти регресії і складіть рівняння парної параболічної регресії за системою рівнянь

$$16 = 8a_0 + 20a_1 + 90a_2$$

$$35 = 20a_0 + 90a_1 + 151a_2$$

$$78 = 90a_0 + 151a_1 + 236a_2.$$

Р3.14. Задана система нормальних рівнянь за МНК

$$45 = 5a_0 + 25a_1$$

$$239 = 25a_0 + 135a_1.$$

Розрахуйте коефіцієнти лінійної регресії a_0 та a_1 методом Гаусса.

Р3.15. Задана система нормальних рівнянь за МНК

$$45 = 5a_0 + 25a_1$$

$$239 = 25a_0 + 135a_1.$$

Розрахуйте коефіцієнти лінійної регресії a_0 та a_1 методом детермінантів (визначників).

Р3.16. Задана матриця числових коефіцієнтів системи нормальних рівнянь за МНК

$$[A] = \begin{bmatrix} 5 & 25 \\ 25 & 135 \end{bmatrix}.$$

Розрахуйте визначник матриці $[A]$.

Р3.17. Задана матриця числових коефіцієнтів системи нормальних рівнянь за МНК

$$[A] = \begin{bmatrix} 5 & 25 \\ 25 & 135 \end{bmatrix}$$

і її визначник $\Delta A = 50$. Отримайте обернену матрицю $[A]$.

Р3.18. Розрахуйте коефіцієнти регресії a_0 і a_1 , якщо

$$[A] = \begin{bmatrix} 2,7 & -0,5 \\ -0,5 & 0,1 \end{bmatrix}, \text{ а } \sum y = 45 \text{ і } \sum yx = 239.$$

Р3.19. Дано:

j	y	x
1	8	4
2	9	5
3	6	3
4	12	7
5	10	6

Складіть систему нормальних рівнянь для оцінювання параметрів a_0 і a_1 за методом найменших квадратів (МНК)

Р3.20. Дано:

j	y	\hat{y}
1	8	7,6
2	9	9
3	6	6,2
4	12	11,8
5	10	10,4

Розрахуйте D_y та $D_{\hat{y}}$. Виконайте ANOVA-аналіз.

Р3.21. За даними Р3.20 визначить коефіцієнт детермінації.

Р3.22. За даними Р3.20 визначить коефіцієнт кореляції.

Р3.23. За даними задачі Р3.20 розрахуйте суму квадратів залишків $\sum e_j^2$.

Р3.24. Дано рівняння регресії $\hat{y} = 2 + 1,4x$, обернена матриця $[A^{-1}] = \begin{bmatrix} 2,7 & -0,5 \\ -0,5 & 0,1 \end{bmatrix}$ і сума квадратів залишків $\sum e_j^2 = 0,40$. Обчисліть стандартні помилки і інтервали довіри для коефіцієнтів регресії a_0 і a_1 .

Р3.25. Дана матриця статистики:

<i>j</i>	<i>y</i>	<i>x</i>
1	8	3
2	9	1
3	7	4
4	4	6
5	2	11

Зробіть необхідні розрахунки і складіть систему нормальних рівнянь за МНК для оцінювання параметрів лінійної регресії.

Р3.26. Дана система нормальних рівнянь за МНК:

$$30 = 5a_0 + 25a_1$$

$$107 = 25a_0 + 183a_1.$$

Обчисліть параметри лінійної регресії способом Гаусса.

Р3.27. Обчисліть визначник (детермінант) матриці A за системою рівнянь задачі Р3.26.

Р3.28. Дана матриця A і визначник, що дорівнює 290. Отримайте обернену матрицю A^{-1} . $[A] = \begin{bmatrix} 5 & 25 \\ 25 & 135 \end{bmatrix}$.

Р3.29. Дана обернена матриця, а також стовпчик вільних членів системи нормальних рівнянь за МНК:

$$\begin{bmatrix} -0,63103 & -0,08601 \\ -0,08601 & 0,01724 \end{bmatrix} \begin{bmatrix} \sum y = 30 \\ \sum yx = 107 \end{bmatrix}.$$

Розрахуйте параметри лінійної парної регресії.

Р3.30. Виконайте за даними задачі Р3.26 розв'язок системи нормальних рівнянь за МНК і обчисліть параметри методом детермінантів.

Р3.31. За даними задачі Р3.25 за рівнянням регресії $\hat{y} = 9,707 - 0,7414x$ розрахуйте оцінки \hat{y} , відхилення $(y - \hat{y})$ і суму квадратів відхилень (залишків) $\sum e_y^2$.

Р3.32. Дано рівняння регресії $\hat{y} = 9,707 - 0,7414x$, обернена матриця $[A^{-1}] = \begin{bmatrix} -0,63103 & -0,08621 \\ -0,08621 & 0,01724 \end{bmatrix}$ і сума квадратів залишків $\sum e^2 = 1,92$.

Обчисліть стандартні помилки і інтервали довіри для коефіцієнтів регресії a_0 і a_1 .

Р3.33. Дано:

j	y	\hat{y}
1	8	7,48
2	9	8,97
3	7	6,74
4	4	5,26
5	2	1,55

Розрахуйте D_y , $D_{\hat{y}}$ або $\sum e_j^2$. Виконайте ANOVA-аналіз.

Р3.34. За даними задачі Р3.33 визначіть коефіцієнти детермінації і кореляції.

Р3.35. Дано: $\hat{y} = 2 - 1,4x$; прогноз: $x_n = 11$; $\sigma_e = 0,365$; $\bar{x} = 5$; девіація $\sum (x - \bar{x})^2 = 58$. Установіть точковий і інтервальний прогноз y .

Р3.36. Дано: $\hat{y} = 9,707 - 0,7414x$; прогноз: $x_n = 8$; $\bar{x} = 5$; $\sigma_e = 0,843$; девіація $\sum (x - \bar{x})^2 = 58$. Установіть точковий і інтервальний прогноз y .

Р3.37. Обчисліть коефіцієнт кореляції за умови, що $D_{\hat{y}} = 14,78$; $\sigma_y = 4,73$.

Р3.38. За даними про загальну і залишкову дисперсію (відповідно 15,62 та 4,07) обчисліть коефіцієнт кореляції.

Р3.39. Обчисліть коефіцієнт детермінації за такими даними: $\sigma_{\hat{y}} = 4,51$; $D_e = 4,07$.

Р3.40. Задані $D_{\hat{y}} = 16,22$; $\sigma_{\varepsilon} = 3,17$. Обчисліть коефіцієнти кореляції і детермінації.

Р3.41. Задані: $D_{\hat{y}} = 11,36$; $\sigma_{\varepsilon} = 1,93$ обчисліть коефіцієнт детермінації і кореляції.

Р3.42. Обчисліть коефіцієнт кореляції за даними девіаціями: $\sum (y - \bar{y})^2 = 116,4$; $\sum (y - \bar{y})^2 = 102,3$.

Р3.43. Задані: $\sum (y - \bar{y})^2 = 386,2$; $\sigma_{\varepsilon} = 1,93$; $n = 20$. Обчисліть коефіцієнт детермінації і кореляції.

Р3.44. Задані: $\hat{y} = 2 + 1,4x$; $x_{\pi} = 11$; $\sigma_{\varepsilon} = 0,365$; $[A^{-1}] = \begin{bmatrix} 2,7 & -0,5 \\ -0,5 & 0,1 \end{bmatrix}$.

Установіть точковий і інтервальний прогноз змінної y .

Р3.45. Задані: $\hat{y} = 9,707 - 0,7414x$; $x_{\pi} = 8$; $[A^{-1}] = \begin{bmatrix} -0,63103 & -0,08621 \\ -0,08621 & 0,01724 \end{bmatrix}$;
 $\sigma_{\varepsilon} = 0,843$. Установіть точковий і інтервальний прогноз змінної y .

3.3.4. Відповіді до розрахункових вправ

3.03. $43 = 5a_0 + 24a_1$

$$252 = 24a_0 + 138a_1.$$

3.04. $43 = 5a_0 + 1,325a_1$

$$13,85 = 1,325a_0 + 0,444a_1.$$

3.05. $\hat{y} = -1 + 2x$.

3.06. $\hat{y} = 3,449 + \frac{20,757}{x}$.

3.07. $\hat{y} = x_1 - x_2$.

3.08. 6,008; 6,597.

3.09. $\hat{y} = -1,6971 + 2,1259x$.

3.10. $\hat{y} = 0,7442 + \frac{25,6244}{x}$.

3.13. $\hat{y} = 0,379 + 0,1003x - 0,1218x^2$.

3.14. $a_0 = 2, a_1 = 1,4$.

3.15. $a_0 = 2, a_1 = 1,4$.

3.16. 50.

3.17. $\begin{bmatrix} 2,7 & -0,5 \\ -0,5 & 0,1 \end{bmatrix}$.

3.18. $a_0 = 2, a_1 = 1,4$.

3.19. $43 = 5a_0 + 24a_1$

$$239 = 25a_0 + 135a_1.$$

3.20. $D_y = 4,00; D_{\hat{y}} = 3,92; D_e = 0,08$.

3.21. 0,9800.

3.22. 0,9899.

3.23. 0,40.

3.24. $\sigma_{a_0} = 0,6; \sigma_{a_1} = 0,115; 1,4 \leq \alpha_0 \leq 2,6; 1,285 \leq \alpha_0 \leq 1,515$.

3.25. $30 = 5a_0 + 25a_1$

$$107 = 25a_0 + 183a_1.$$

3.26. $a_0 = 9,707, a_1 = -0,7414$.

3.27. 290.

3.28. $\begin{bmatrix} -0,63103 & -0,08621 \\ -0,08621 & 0,01724 \end{bmatrix}$.

3.29. $a_0 = 9,707, a_1 = -0,7414$.

3.30. $a_0 = 9,707, a_1 = -0,7414$.

3.31. 2,13.

3.32. $9,038 \leq \alpha_0 \leq 10,376$.

3.33. $D_y = 6,8; D_{\hat{y}} = 6,38; D_e = 0,42$.

3.34. $d = 0,938; r_{yx} = 0,969$.

3.35. $\hat{y}_n = 17,4; 16,69 \leq y_n \leq 18,11$.

3.36. $y_n = 3,78; 3,28 \leq y_n \leq 4,28$.

3.37. 0,8128.

3.38. 0,8599.

3.39. 0,8332.

- 3.40. 0,6175; 0,7858.
3.41. 0,6721; 0,8198.
3.42. 0,9375.
3.43. 0,8071; 0,8984.
3.44. $y_n = 17,4; 16,69 \leq y_n \leq 18,11$.
3.45. $y_n = 3,78; 3,28 \leq y_n \leq 4,28$.

4. САМОСТІЙНЕ ВИВЧЕННЯ ЕКОНОМЕТРІЇ

4.1. ЗМІСТ І САМОКОНТРОЛЬ САМОСТІЙНОЇ ТЕОРЕТИЧНОЇ ПІДГОТОВКИ

Позааудиторна теоретична підготовка – це опанування знаннями та оволодіння вміннями шляхом самоініційованого вивчення рекомендованих підручників, навчальних посібників, інших навчальних видань – див. рекомендовану навчальну літературу. Цей шлях є по суті єдиним для студентів заочної, дистанційної, екстернатної форм навчання.

Навчальний посібник [1], конспект лекцій [4] і методичне видання [7] можна отримати в тимчасове користування у бібліотеці. З підручником [3] та навчальними посібниками [2], [5] і [6] Вам запропонують працювати у читальному залі бібліотеки.

Розполіт теоретичного матеріалу рекомендованих навчальних видань за окремими змістовими модулями такий:

ЗМ 1. Постановка задачі економетричного моделювання

- [1], розділи 1, 2;
[2], розділи 1, 2, 4;
[3], розділи 1, 2;
[4], розділи 1, 2;
[5], глави 2, 3.

Розділ 1 цього посібника.

ЗМ 2. Специфікація економетричних моделей

[1], розділи 3, 4;

[2], розділи 2, 3, 4, 5;

[3], розділи 4, 5, 6;

[4], розділи 3, 4;

[5], глави 2, 3, 6;

[6], розділи 1, 2, 3, 4.

Розділ 2 цього посібника.

ЗМ 3. Оцінювання параметрів економетричних моделей

[1], розділи 5, 6;

[2], розділи 2, 4;

[3], розділи 7, 8;

[4], розділ 5;

[5], глави 2, 3;

[6], розділи 1, 2, 3.

Розділ 3 цього посібника.

Самостійне опрацювання теоретичного матеріалу вимагає від студента самоконтролю власних навчальних досягнень. Для самоконтролю необхідно звернутися до підрозділів 1.3; 2.3 і 3.3 посібника і за відповідними змістовими модулями виконати тестові завдання і логічні (теоретичні) вправи (принаймні вибірково).

Закінчивши опрацювання теорії дисципліни, визначте, чи твердо Ви за-
своїли насамперед наступні основні поняття економетрії:

- стохастична (ймовірнісна) залежність;
- кореляція, кореляційна залежність;
- економетричне моделювання;
- економетрична модель;
- параметри економетричної моделі;
- ендогенні та екзогенні змінні;

- регресія, рівняння регресії;
- помилка апроксимації (прогнозу);
- ідентифікація факторів;
- специфікація аналітичної форми регресії;
- мультиколінеарність;
- авторегресія;
- гетероскедастичність;
- незміщеність, обґрунтованість, ефективність оцінок.

Запишіть напам'ять у загальному вигляді основні аналітичні форми моделей рівнянь регресії:

- лінійна модель;
- квазілінійна модель;
- суттєво нелінійна модель;
- множинна лінійна модель;
- множинна нелінійна модель;
- авторегресивна модель;
- дистрибутивно-лагова модель;
- симультативна модель.

Напишіть напам'ять формули основних вимірювальників кореляції і регресії:

- коефіцієнт парної кореляції;
- коефіцієнт парної детермінації;
- β -коефіцієнт;
- коефіцієнт частинної парної детермінації;
- ANOVA-аналіз;
- коефіцієнт множинної кореляції;
- коефіцієнт множинної детермінації;
- коефіцієнт регресії;
- коефіцієнт еластичності.

Перевірте, чи твердо Ви запам'ятали призначення різних статистичних тестів (критеріїв), зокрема:

- t -статистика нормального розподілу;
- t -статистика Стьюдента;
- F -критерій Фішера;
- DW -статистика Дарбіна-Уотсона.

4.2. ЗМІСТ І САМОКОНТРОЛЬ САМОСТІЙНОЇ ПРАКТИЧНОЇ ПІДГОТОВКИ

Самостійна практична підготовка призначена для набуття вмінь, навичок і первісного власного практичного досвіду з економетричного моделювання.

Практична самостійна робота з дисципліни здійснюється у двох формах:

- виконання розрахункових вправ;
- виконання комплексного індивідуального домашнього завдання.

Розділи 1, 2 і 3 за змістовими модулями містять понад 100 розрахункових вправ (задач), у тому числі:

ЗМ 1.	27,
ЗМ 2.	39,
ЗМ 3.	45.

Самостійне розв'язання цих задач слугує закріпленню засвоєних теоретичних знань і формуванню навичок виконання окремих розрахунків у процесі економетричного моделювання. Для самоконтролю правильності рішень у посібнику наводяться відповіді до розрахункових вправ (див. 1.3.4, 2.3.4 і 3.3.4).

Формуванню теоретичних уявлень і необхідних практичних навичок економетричного моделювання як цілісного творчого процесу слугує обов'язкове виконання кожним студентом індивідуального домашнього завдання. Вихідні дані, методика і приклади виконання завдання викладені у 5-му розділі Посібника. Розподіл покрокового процесу розробки рівняння регресії за змістовими модулями такий:

ЗМ 1. кроки 1÷5;

ЗМ 2. кроки 6÷14;

ЗМ 3. кроки 15÷26.

Виконання індивідуального завдання має регулярно контролюватися і оцінюватися викладачем. Студенти заочної форми навчання виконують цю роботу як контрольну.

5. ЗАВДАННЯ ДЛЯ ПРАКТИЧНИХ ЗАНЯТЬ, ВИКОНАННЯ ІНДИВІДУАЛЬНОГО ДОМАШНЬОГО ЗАВДАННЯ АБО КОНТРОЛЬНОЇ РОБОТИ

5.1. ЗМІСТ ЗАВДАНЬ ТА ВИХІДНІ ДАНІ

Зміст практичних занять та індивідуального домашнього завдання або контрольної роботи (для студентів заочної форми навчання) полягає у побудові економетричної моделі прибутковості (рентабельності), тобто рівняння регресії цієї функції на змінення двох внутрішньогосподарських факторів виробництва. Робота виконується на статистичних даних вибірки однорідних підприємств.

Перелік, позначення, зміст операційних характеристик та одиниці виміру змінних такі:

ендогенні (залежні) змінні:

- доходність витрат (Y_1, P_D) – відношення виручки підприємства від реалізації продукції (робіт, послуг) до витрат підприємства на їх виробництво, грн./грн.;
- рівень витрат ($Y_2; V_3$) – відношення витрат на вироблену підприємством продукцію (роботи, послуги) до виручки від їх реалізації, коп./грн.;
- прибутковість (рентабельність) реалізації (Y_3, P_P) – відношення прибутку підприємства від реалізації продукції (робіт, послуг) до виручки від їх реалізації, коп./грн.;
- прибутковість (рентабельність) витрат (Y_4, P) – відношення прибутку підприємства від реалізації продукції (робіт, послуг) до витрат на виробництво, коп./грн.;

екзогенні (незалежні) змінні:

фондоозброєність праці ($X_0; \Phi$)	– відношення первісної вартості основних виробничих засобів до чисельності працівників підприємства, тис.грн./чол.;
структура основних виробничих засобів ($X_1; A$)	– питома вага активних основних виробничих засобів (машин і обладнання) у загальній первісній вартості основних виробничих засобів, %;
коефіцієнт придатності ($X_2; \Gamma$)	– відношення залишкової вартості основних виробничих засобів до їх первісної вартості, %;
енергоозброєність праці ($X_3; \Xi$)	– відношення сумарної потужності двигунів машин, обладнання тощо до чисельності працівників підприємства, кВт/чол.;
спеціалізація підприємства ($X_4; C$)	– питома вага обсягу реалізації одного (головного) виду продукції (робіт, послуг) в загальному обсязі реалізації, %;
кооперування виробництва ($X_5; K$)	– питома вага вартості покупних комплектуючих деталей і вузлів у загальних матеріальних витратах на виробництво продукції (робіт, послуг), %;
кількість видів продукції (робіт, послуг) ($X_6; \Pi$)	– кількість найменувань продукції (робіт, послуг), що виробляється підприємством, одиниць;
питома вага зарплати ($X_7; \mathcal{Z}$)	– питома вага витрат на оплату праці в загальній сумі витрат на виробництво продукції (робіт, послуг), %;
бригадна організація праці ($X_8; B$)	– питома вага робітників, залучених до бригадної форми організації та оплати праці, у загальній чисельності робітників підприємства, %;
плинність кадрів ($X_9; T$)	– відношення чисельності звільнених за рік працівників за власним бажанням та за порушення трудової дисципліни до середньорічної чисельності працівників підприємства, %.

Статистичні дані 15 підприємств про числові значення чотирьох ендогенних (залежних) та десяти екзогенних (незалежних) змінних наведені у табл. 5.1.

5.2. ВАРІАНТИ ЗАВДАНЬ

Варіанти завдань на практичні заняття, виконання індивідуального домашнього завдання студентами денної та контрольної роботи студентами заочної форми навчання визначаються так.

а) Практичні заняття

Таблиця 4.1 – Статистичні дані 15 підприємств

j	Змінні														j
	ендогенні (залежні)				екзогенні (незалежні)										
	Y ₁	Y ₂	Y ₃	Y ₄	X ₀	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	
P _Д	Y ₃	P _Р	P	Φ	A	Г	Э	С	К	П	З	Б	Т		
1	1,067	93,7	6,3	6,7	1,5	36	36	4,5	58	46	18	21	24	33	1
2	1,073	93,2	6,8	7,3	1,7	48	44	4,8	68	49	17	23	35	27	2
3	1,115	89,7	10,3	11,5	6,4	62	65	16,3	73	68	14	34	62	12	3
4	1,109	90,2	9,8	10,9	4,7	61	67	19,4	84	71	11	35	68	11	4
5	1,054	94,9	5,1	5,4	1,6	41	28	3,5	60	31	21	18	15	31	5
6	1,097	91,2	8,8	9,7	4,2	51	48	11,4	74	54	16	29	71	25	6
7	1,101	90,8	9,2	10,1	7,6	54	50	14,1	78	52	18	33	67	11	7
8	1,114	89,8	10,2	11,4	6,9	76	64	14,9	88	64	12	37	84	7	8
9	1,079	92,7	7,3	7,9	2,4	53	49	6,7	72	51	22	21	44	24	9
10	1,089	91,8	8,2	8,9	2,9	68	61	11,5	81	68	13	28	52	21	10
11	1,095	91,3	8,7	9,5	3,8	64	65	11,9	77	74	15	32	61	23	11
12	1,127	88,7	11,3	12,7	7,3	73	72	21,5	86	79	9	39	77	7	12
13	1,189	84,1	15,9	18,9	9,1	82	77	27,6	91	67	14	38	51	4	13
14	1,098	91,1	8,9	9,8	5,6	49	54	12,1	75	72	9	31	64	19	14
15	1,104	90,6	9,4	10,4	7,1	54	58	13,8	82	58	15	36	69	17	15

На практичних заняттях протягом семестру в усіх групах виконується **єдиний** варіант завдання: моделювання рівняння регресії прибутковості (рентабельності) витрат (Y_4, P) на змінення двохзмінних – факторів: фондоозброєності праці ($X_0; \Phi$) та кооперування виробництва ($X_5; K$), тобто

$$P = f(\Phi; K) \pm \sigma_{\hat{p}} \quad (P = 0,95),$$

де $\sigma_{\hat{p}}$ – довірчі межі помилки апроксимації з ймовірністю 0,95*.

б) Індивідуальне домашнє завдання

Варіант індивідуального домашнього завдання студентам денної форми навчання **визначає викладач**, який веде практичні заняття, за такими ознаками:

- варіант ендогенної (залежної) змінної Y_φ ($\varphi = 1, 2, 3, 4$) визначається, як правило, єдиним для всіх студентів однієї академічної групи, тобто, $Y_1(P_D)$ або $Y_1(V_3)$, або $Y_3(P_P)$, або $Y_4(P)$;

* Далі у цьому розділі розрахункові формули, таблиці і приклади приводяться саме для цього варіанту завдання.

- варіанти двох екзогенних (незалежних) змінних або факторів можливі такі (№ змінних):

01
 02 12
 03 13 23
 04 14 24 34
 05 15 25 35 45
 06 16 26 36 46 56
 07 17 27 37 47 57 67
 08 18 28 38 48 58 68 78
 09 19 29 39 49 59 69 79 89.

У кожній академічній групі викладач визначає і веде поіменний облік виданих варіантів завдань. Отже, кожний студент розробляє рівняння регресії такого загального вигляду

$$Y_{\varphi} = f(X_i, X_j) \pm \Delta y \quad (P = 0,95),$$

де $i, j = 0, 1, 2, \dots, 9$, але $i \neq j$;

$\varphi = 1, 2, 3, 4$.

в) Завдання на контрольну роботу для студентів заочної форми навчання

Варіант завдання на контрольну роботу студент заочної форми навчання визначає самостійно. Зокрема, варіант ендогенної (залежної) змінної Y_{φ} ($\varphi = 1, 2, 3, 4$) беруть довільно, тобто будь-який з чотирьох можливих. Варіанти двох екзогенних (незалежних) змінних визначають за останніми двома цифрами номера залікової книжки. Наприклад:

№ залікової книжки	Склад екзогенних (незалежних) змінних-факторів
... 01	$X_0 (\Phi); X_1 (A)$
... 27	$X_2 (\Gamma); X_7 (З)$
... 30	$X_3 (\Xi); X_0 (\Phi)$

Повний перелік варіантів складу факторів наведений вище (див. п. б). Якщо останні дві цифри номера залікової книжки однакові (наприклад, 44, 00 тощо), варіант визначається за умови зміни однієї цифри на найближчу (наприклад, 44 → 43 або 45, або 34, або 54).

5.3. ПОСЛІДОВНІСТЬ РОЗРОБКИ РІВНЯННЯ РЕГРЕСІЇ

Загальна блок-схема алгоритму розробки рівняння регресії зображена на рис. 5.1 у двох варіантах: а) повний – для денної форми навчання; б) скорочений – для заочної форми навчання.

Блок-схеми включають наступні розрахункові та логічні процедури (кроки), об'єднані у три змістові модулі:

Змістовий модуль 1. Постановка задачі

1. Постановка задачі моделювання.
2. Формування матриці даних про змінні.
3. Розрахунок показників розмаху варіації змінних.
4. Побудова і аналіз полів кореляції.
5. Розрахунок показників варіації змінних.

Змістовий модуль 2. Специфікація рівняння регресії

6. Розрахунок коефіцієнтів кореляції.
7. Тестування факторів на значущість (невипадковість).
8. Оцінка мультиколінеарності факторів.
9. Розрахунок β -коефіцієнтів.
10. Тестування факторів на автономність.
11. Розрахунок коефіцієнта множинної кореляції.
12. Тестування факторів на вагомість у множинній детермінації.
13. Прийняття рішень щодо включення факторів у рівняння регресії.
14. Обґрунтування аналітичної форми рівняння регресії.

а) повний варіант

б) скорочений варіант

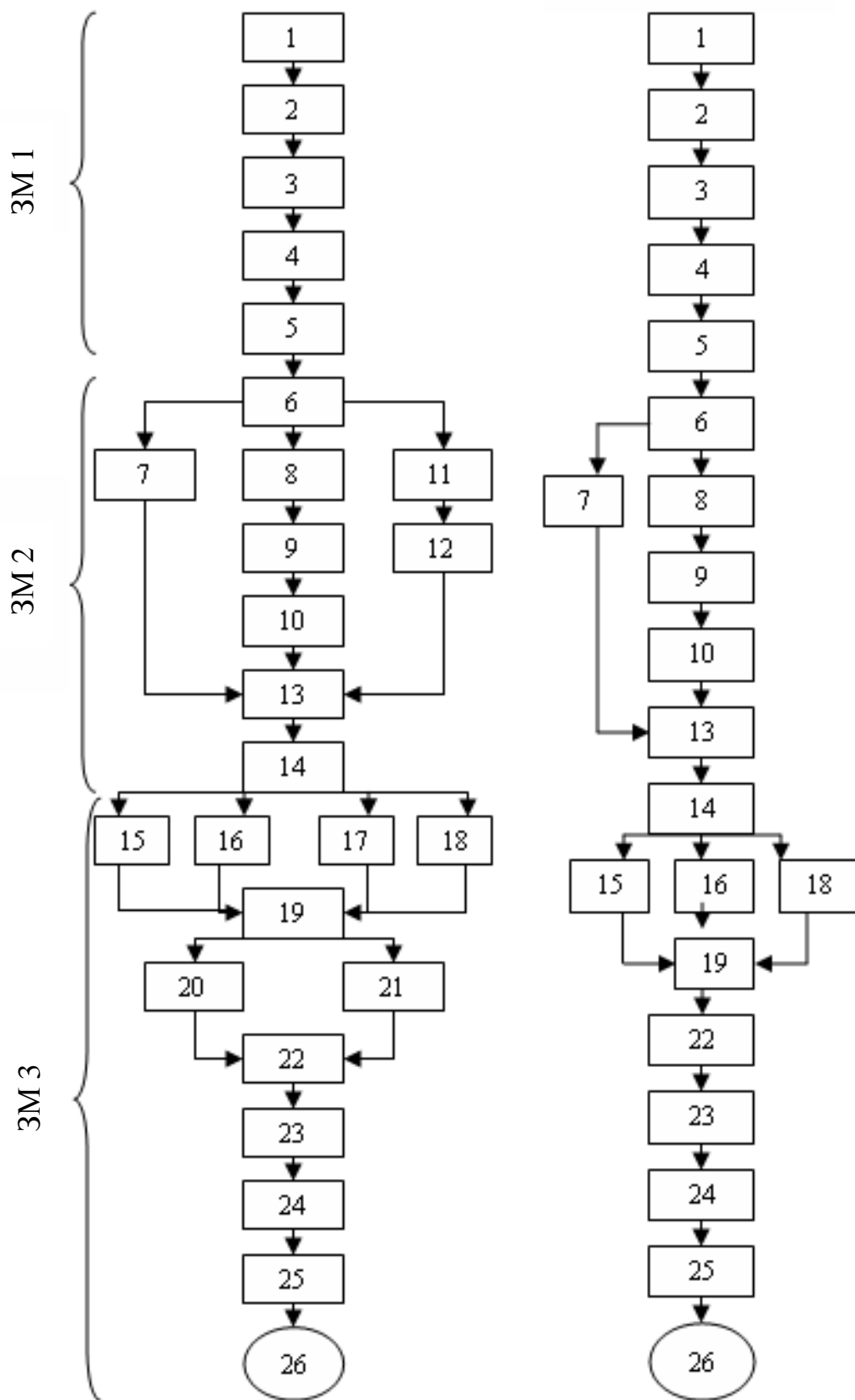


Рис.5.1 – Блок-схема розробки рівняння регресії

Змістовий модуль 3. Оцінювання параметрів економетричної моделі

15. Розрахунок коефіцієнтів регресії МНК способом Гаусса.
16. Те ж, способом детермінантів.
17. Те ж, способом оберненої матриці.
18. Те ж, за β -коефіцієнтами.
19. Розрахунок оцінок ендогенної змінної за рівнянням регресії.
20. Тестування гомоскедастичності залишків.
21. Тестування автокореляції залишків.
22. ANOVA-дисперсійний аналіз.
23. Тестування значущості (адекватності) рівняння регресії.
24. Тестування коефіцієнтів регресії на значущість (невипадковість).
25. Розрахунок інтервалів довіри для прогнозу рентабельності.
26. Економічна інтерпретація економетричної моделі.

5.4. МЕТОДИКА РОЗРОБКИ РІВНЯННЯ РЕГРЕСІЇ

Крок 1. Постановка задачі включає: а) вибір змінних та їх операційних характеристик (у прикладі – P , Φ , K) для складання рівняння регресії за варіантом завдання; б) теоретичне обґрунтування наявності та аналітичної форми залежності прибутковості від обох факторів; в) складання розроблюваного рівняння регресії у загальному вигляді. Початкове розуміння сутності розроблюваного рівняння регресії дуже важливе для кількісних і якісних оцінок багатьох проміжних і заключних результатів моделювання.

Крок 2. Матриця статистики складається за варіантом завдання за даними 15 підприємств. Наприклад, для варіанта на практичні заняття вона має вигляд (табл. 5.2).

Таблиця 5.2 – Матриця статистики (для варіанту $P = f(\Phi, K)$)

№ підприємства	P, коп./грн.	Φ, тис. грн./чол.	K, %
1	6,7	1,5	46
2	7,3	1,7	49
...			
15	10,4	7,1	58

Матриця статистики характеризується:

- мірністю, тобто кількістю змінних $(m + 1)$;
- об'ємом вибірки, тобто кількістю об'єктів спостереження (n) ;
- об'ємом матриці $(m + 1)n$;
- співвідношенням розмірів матриці $n/(m + 1)$, яке для отримання незміщених оцінок кореляції і регресії повинно бути не менше восьми. Ця умова порушується, що виправдовується навчальним характером роботи, у якій процес моделювання набагато вагоміший за кінцевий результат. Скорочення матриці дає виграш у трудомісткості процесу без жодних втрат щодо його змісту та методики виконання.

Крок 3. **Показники розмаху варіації змінних** розраховуються за формулами 1.10 і 1.11 (наприклад, для змінної P):

- абсолютний розмах варіації

$$R_p = P_{max} - P_{min};$$

- відносний розмах варіації

$$i_p = P_{max}/P_{min}.$$

Розрахунок показників слід оформити в табл. 5.3.

Таблиця 5.3 – Розрахунок показників розмаху варіації змінних (для варіанту $P = f(\Phi, K)$)

	P	Φ	K
1	6,7	1,5	46
2	7,3	1,7	49
...			
15	10,4	7,1	58
Мінімальне значення	5,4	1,4	31
Максимальне значення	18,9	9,1	79
Абсолютний розмах варіації	13,5	7,6	48
Відносний розмах варіації	3,5	6,1	2,5

Крок 4. **Поля кореляції** (графічні зображення залежності) будуються за матрицею статистики (табл. 5.2) на міліметровому папері формату А4. Масштаб зображення за осями координат вибирається таким, щоб поле кореляції виглядало

«стоячим», якщо $i_p > i_{x_i}$ (рис. 5.2, а), «лежачим», якщо $i_p < i_{x_i}$ (рис. 5.2, б) або квадратним, якщо $i_p \approx i_{x_i}$ (рис. 5.2, в).

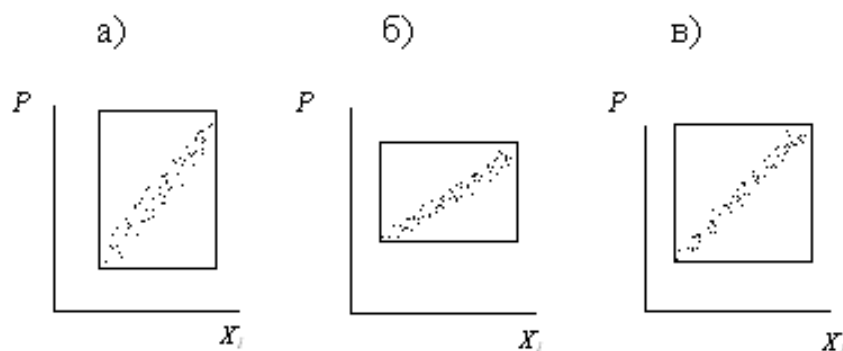


Рис. 5.2 – Типи полів кореляції

Розмітка координаційної сітки диктується мінімальними і максимальними значеннями змінних, тому площа полів повинна використовуватися повністю.

Аналіз полів кореляції проводиться з метою визначення за графічними критеріями:

- наявності кореляційних залежностей;
- напрямку та аналітичної форми залежностей;
- кількісної однорідності об'єктів спостереження, зокрема, наявності аномальних об'єктів.

Аномальні об'єкти спостереження – це об'єкти, що «випадають» із вибіркової сукупності на полях кореляції. Вони визначаються візуально. Аномальні об'єкти спостереження вилучаються із матриці статистики. У прикладі це об'єкт № 13.

Зробіть відповідні висновки за всіма аспектами аналізу полів кореляції.

Крок 5. Показники варіації змінних – дисперсія і середнє квадратичне відхилення – визначаються за формулами 1.12, 1.13, 1.14 (наприклад, для змінної P):

- середнє арифметичне

$$\bar{P} = \frac{1}{15} \sum_{i=1}^{15} P_i;$$

- дисперсія (середній квадрат відхилення)

$$D_p = \overline{P^2} - \overline{P}^2;$$

- середнє квадратичне відхилення

$$y_p = \sqrt{D_p}.$$

Розрахунок показників варіації змінних рекомендується викласти у табл. 5.4.

Таблиця 5.4 – Розрахунок показників варіації змінних (для варіанту $P = f(\Phi, K)$)

	<i>P</i>	<i>Φ</i>	<i>K</i>	<i>P</i>²	<i>Φ</i>²	<i>K</i>²
1	6,7	1,5	76	44,89	2,25	2116
2	7,3	1,7	49	53,29	2,89	2401
...						
15	10,4	7,1	58	108,16	50,41	3364
Сума	132,2	63,7	837	1301,82	357,43	52349
Середнє арифметичне	9,4429	4,5500	59,7857	92,9871	25,5307	3739,21
Квадрат середнього	89,1684	20,7025	3574,32			
Дисперсія	3,8187	4,8282	164,884			
Середнє квадратичне відхилення	1,9541	2,1973	12,8407			

Останні три стовпці табл. 5.4 призначені для вирахування середніх квадратів змінних, наприклад:

$$\overline{P^2} = 1301,82/14 = 92,9871.$$

Вони в свою чергу використовуються для розрахунку дисперсій змінних за формулою (1.13), наприклад:

$$D_p = 92,9871 - 89,1684 = 3,8187.$$

Крок 6. Для **розрахунку коефіцієнтів кореляції** необхідно попередньо виконати розрахунок середніх добутків змінних (табл. 5.5),

Таблиця 5.5 – Розрахунок середніх добутків змінних

<i>j</i>	<i>PΦ</i>	<i>PK</i>	<i>ΦK</i>
1	10,05	308,2	69,0
2			
...			
15			
Сума	654,39	8191,0	4026,9
Середній добуток	46,7421	585,0714	287,6357

після чого виконати розрахунок коефіцієнтів кореляції за формулою (2.1) для варіанта $P = f(\Phi, K)$:

$$r_{px_i} = (\overline{PX_i} - \bar{P} \times \bar{X}_i) / \sigma_p \sigma_{x_i}.$$

Наприклад, за даними табл. 5.4 і 5.5 знаходимо

$$r_{p\phi} = (46,7421 - 9,4429 \times 4,5500) / 1,9541 \times 2,1973 = 0,8796.$$

Аналогічно $r_{pk} = 0,8178$ і $r_{p\phi} = 0,5533$.

Коефіцієнти кореляції показують напрямок і силу впливу факторів на рентабельність («+» – доданий, «-» – від’ємний). Слід мати на увазі, що $-1 \leq r_{px_i} \leq +1$. За наслідками розрахунків зробіть висновки щодо сили (тісноти) кореляційних зв’язків.

Крок 7. Тестування факторів на значущість (невипадковість) їх впливу на рентабельність проводиться за t -статистикою Стьюдента. Розраховують спочатку вибіркові значення t_i за формулою (2.3) (для варіанта $P = f(\Phi, K)$)

$$t_{\phi} = |r_{p\phi_i}| \sqrt{\frac{n-2}{1-r_{p\phi_i}^2}}$$

і потім порівнюють їх з критеріальним ($t_{крит}$). $t_{крит}$ визначають за додатком 3 залежно від n .

З ймовірністю 0,95 зробіть Ваші висновки щодо значущості факторів. При відсутності посібника можна прийняти $t_{крит} = 2,179$. У нашому прикладі за формулою (2.3)

$$t_{\phi} = 0,8796 \sqrt{\frac{14-2}{1-0,8796^2}} = 6,405$$

і відповідно $t_k = 4,923$. Отже, вплив факторів на рентабельність є не випадковим, оскільки $6,405 > 2,179$ і $4,923 > 2,179$.

Крок 8. Наявність і сила **мультиколінеарності факторів**, тобто взаємозв’язку між ними, оцінюється за повною матрицею коефіцієнтів кореляції, тобто

$$\begin{array}{c}
 (\Phi) \quad (K) \quad (P) \\
 (\Phi) \begin{bmatrix} 1 & r_{\phi k} & r_{p\phi} \\ r_{k\phi} & 1 & r_{pk} \\ r_{p\phi} & r_{pk} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0,5533 & 0,8796 \\ 0,5533 & 1 & 0,8178 \\ 0,8796 & 0,8178 & 1 \end{bmatrix}
 \end{array}$$

за наступною шкалою оцінок (за модулями коефіцієнтів кореляції для загального випадку $y = f(X_1, X_2)$):

$$r_{12} = 0 \quad \text{– відсутня,}$$

$$0 < r_{12} < r_{yx_i \min} \quad \text{– слабка,}$$

$$r_{yx_i \min} < r_{12} < r_{yx_i \max} \quad \text{– помірна,}$$

$$r_{yx_i \max} < r_{12} < 1 \quad \text{– сильна,}$$

$r_{12} = 1$ – абсолютна, коли з двох факторів один є однозначно зайвим.

У прикладі колінеарність факторів є слабкою, оскільки $0,5533 < 0,8798$ і $0,8178$.

Крок 9. β -коефіцієнти, які визначають напрямок і силу впливу факторів на рентабельність за умови вилучення мультиколінеарності, розраховуються за формулою (2.6)

$$\beta_i = \Delta_i / \Delta_0,$$

де Δ_0 – визначник (детермінант) матриці взаємної кореляції (мультиколінеарності);

Δ_i – те ж із заміною в ній i -го стовпця стовпцем коефіцієнтів кореляції

r_{px_i} . Наприклад (див. крок 8):

$$\beta_{\phi} = \frac{\begin{vmatrix} r_{p\phi} & r_{\phi k} \\ r_{pk} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{\phi k} \\ r_{k\phi} & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 0,8796 & 0,5533 \\ 0,8178 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 0,5533 \\ 0,5533 & 1 \end{vmatrix}} = 0,6155.$$

Аналогічно $\beta_k = 0,4772$.

Крок 10. **Оцінка незалежності (автономності) впливу факторів** на рентабельність визначається за допомогою γ -критерія

$$0 < \gamma_i < 1,$$

де γ_i – рівень автономності, який розраховується за формулою (2.7)

$$\gamma_i = \beta_i / r_{px_i}.$$

Якщо $\gamma_i > 0$, фактор x_i має певну автономність впливу на рентабельність; $\gamma_i = 0$ – ніякої автономності він не має і його вплив на рентабельність завдяки мультиколінеарності проявляється через інші фактори, якщо $\gamma_i < 0$ – i -й фактор вилучається з подальшого процесу розробки рівняння регресії, бо знаки β_i і r_{px_i} протилежні, чого у належно специфікованих моделях не може бути.

Крок 11. Розрахунок **коефіцієнта множинної кореляції** необхідний для визначення сили впливу на рентабельність обох факторів разом, він розраховується за формулу Боярського (2.8)

$$R_{p.fk} = \sqrt{\frac{|\Delta_*|}{\Delta_0}},$$

де Δ_* – визначник повної матриці коефіцієнтів кореляції із заміною нижнього правого елемента нулем.

У нашому прикладі формулу Боярського застосовуєть так:

$$R_{p.fk} = \sqrt{\frac{\begin{vmatrix} 1 & 0,5533 & 0,8796 \\ 0,5533 & 1 & 0,8178 \\ 0,8796 & 0,8178 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & 0,5533 \\ 0,5533 & 1 \end{vmatrix}}} = \sqrt{\frac{0,6465}{0,6939}} = \sqrt{0,9317} = 0,9652.$$

З метою контролю правильності розрахунків рекомендується цей коефіцієнт визначати також за такою формулою (2.10)

$$R_{p.fk} = \sqrt{r_{p\phi} \beta_\phi + r_{pk} \beta_k}.$$

У прикладі (див. крок 9)

$$R_{\rho_{\phi\kappa}} = \sqrt{0,8796 \times 0,6155 + 0,8178 \times 0,4772} = \sqrt{0,5414 + 0,3903} = \sqrt{0,9317} = 0,9652.$$

Отже, коефіцієнт множинної детермінації 0,9317 показує, що варіація рентабельності у різних підприємств пояснюється впливом різних рівнів фондоозброєності праці на 54,14%, різного рівня кооперування виробництва на 39,03% і іншими не врахованими факторами – 6,83% (100 – 83,17).

Крок 12. Вагомість вкладу факторів у варіацію рентабельності оцінюється на основі коефіцієнтів множинної кореляції за ρ -критерієм

$$\rho_i \geq 1,96,$$

де (див. формулу 2.11)

$$\rho_i = 0,742\sqrt{n}\sqrt{R_{\rho_{x_1x_2}}^2 - r_{\rho_{x_j}}^2} \quad (i, j = 1, 2; i \neq j).$$

У нашому прикладі формула (2.11) для визначення ρ_ϕ і ρ_κ набуває такого вигляду:

$$\rho_\phi = 0,742\sqrt{n}\sqrt{R_{\rho_{\phi\kappa}}^2 - r_{\rho_\kappa}^2};$$

$$\rho_\kappa = 0,742\sqrt{n}\sqrt{R_{\rho_{\phi\kappa}}^2 - r_{\rho_\phi}^2}.$$

Якщо $\rho_i \geq 1,96$, то з ймовірністю 0,95 і більше можна стверджувати, що вплив фактора x_i достатньо значущий, вагомий для включення його у рівняння регресії (див. додаток 1).

Крок 13. Для прийняття рішення щодо включення факторів у рівняння регресії складається зведення результатів, що одержані на кроках 7, 10, 12 (табл. 5.6).

Таблиця 5.6 – Зведення оцінок не випадковості, автономності та значущості вкладу факторів

Фактори	$t_i \geq 2,179$	$0 < \gamma < 1$	$\rho_i \geq 1,96$	Рішення
ϕ	6,405	0,6997	1,423	Включається
κ	4,923	0,5835	1,104	Включається

Поради щодо прийняття рішень такі:

- якщо $\gamma_i < 0$, x_i вилючається з матриці (крок 10);
- значення ρ_i дуже залежать від об'єму вибірки (див. формулу 2.11), тому рішення приймається з огляду на порушення вимоги $n/(m + 1) \geq 8$.

Отже, якщо критерії ρ_i виконуються не в повній мірі, обидва фактори треба включати до рівняння регресії, принаймні з ймовірністю меншою 0,95.

Крок 14. Для *обґрунтування аналітичної форми рівняння регресії* необхідно скористатися раніше сформульованими щодо цього висновками із теоретичного обґрунтування (крок 1), а також із візуального аналізу полів кореляції (див. крок 4). Ці аргументи вибору аналітичної форми рівняння регресії (пряма, гіпербола, парабола тощо) достатньо надійні і ними можна обмежитися.

Не буде великої помилки, якщо форма регресії буде прийнята лінійною і модель рівняння регресії матиме такий вигляд (для нашого прикладу):

$$\hat{P} = a_0 + a_1\Phi + a_2K, \quad (5.1)$$

де a_0, a_1, a_2 – коефіцієнти регресії. Коефіцієнт a_0 показує частину P , що не залежить від факторів Φ і K ; a_1 визначає, на скільки копійок змінюється P за рахунок зміни Φ на одну тисячу гривень; a_2 визначає зміну P в копійках при зміні K на один відсоток.

Крок 15. Для *розрахунку коефіцієнтів регресії* a_0, a_1 та a_2 методом найменших квадратів (МНК) слід скласти систему нормальних рівнянь і вирішити її. У нашому прикладі в разі вибору лінійної форми регресії ця система така:

$$\left. \begin{aligned} \sum P &= n \times a_0 + a_1 \sum \Phi + a_2 \sum K \\ \sum P\Phi &= a_0 \sum \Phi + a_1 \sum \Phi^2 + a_2 \sum K\Phi \\ \sum PK &= a_0 \sum K + a_1 \sum \Phi K + a_2 \sum K^2 \end{aligned} \right\}. \quad (5.2)$$

Числові значення коефіцієнтів цієї системи беруть із табл. 5.4 (крок 5), та 5.5 (крок 6). У нашому прикладі система набуває вигляду

$$\begin{aligned} 132,2 &= 14a_0 + 63,7a_1 + 837a_2 \\ 654,39 &= 63,7a_0 + 357,43a_1 + 4026,9a_2 \\ 8191,0 &= 837a_0 + 4026,9a_1 + 52349a_2. \end{aligned}$$

Спосіб Гауса полягає у послідовному вилученні невідомих a_0 , a_1 та a_2 . Діленням системи на числові коефіцієнти при a_0 отримаємо

$$9,4429 = a_0 + 4,5500a_1 + 59,7857a_2$$

$$10,2730 = a_0 + 5,6115a_1 + 63,2166a_2$$

$$9,7861 = a_0 + 4,8111a_1 + 62,5436a_2.$$

Віднімемо 1-е рівняння із 2-го і 3-го, після чого маємо

$$0,8301 = 1,0615a_1 + 3,4304a_2$$

$$0,3432 = 0,2611a_1 + 2,7579a_2,$$

тобто систему рівнянь без a_0 . Повторивши цю саму процедуру ще раз, отримаємо одне рівняння з одним невідомим a_2

$$0,5324 = 7,3309a_2,$$

звідкіля $a_2 = 0,0726$. Далі зворотним шляхом знайдемо $a_1 = 0,5476$ і $a_0 = 2,6108$.

Отже,

$$\hat{P} = 2,6108 + 0,5476\Phi + 0,0726K. \quad (5.3)$$

Прокоментуйте сутність знайдених параметрів економетричної моделі рентабельності, тобто 2,6108; 0,5476 та 0,0726.

Крок 16. Для розрахунку коефіцієнтів регресії методом детермінантів попередньо необхідно за системою рівнянь МНК визначити чотири детермінанта (визначника) – ΔA ; ΔA_0 ; ΔA_1 і ΔA_2 . Зокрема у нашому прикладі це виглядає так:

$$\Delta A = \begin{bmatrix} 14 & 63,7 & 837 \\ 63,7 & 357,43 & 4026 \\ 837 & 357,43 & 52349 \end{bmatrix} = 1513628.$$

Послідовно замінюючи стовпці матриці A стовпцями вільних членів системи рівнянь за МНК знайдемо

$$\Delta A_0 = \begin{vmatrix} 132,2 & 63,7 & 837 \\ 654,39 & 357,43 & 4026,9 \\ 8191,0 & 4026,9 & 52349 \end{vmatrix} = 3951780,$$

$$\Delta A_1 = \begin{vmatrix} 14 & 132,2 & 837 \\ 63,7 & 654,39 & 4026,9 \\ 837 & 8191,0 & 52349 \end{vmatrix} = 828831,$$

$$\Delta A_2 = \begin{vmatrix} 14 & 63,7 & 132,2 \\ 63,7 & 357,43 & 654,39 \\ 837 & 4026,9 & 8191,0 \end{vmatrix} = 109889.$$

Оскільки розрахунки детермінантів виконуються вручну на калькуляторі, нагадаємо, як їх зручніше і простіше виконувати.

Елементи матриці A поділимо на 100, тобто на 10^2 , що потім обумовить множення результату на 1000000, тобто на 10^6 . Для розрахунку визначника матриці 3-го порядку допишемо до неї ще раз перші два рядки. Знайдемо добутки трьох елементів по кожній нісхідній діагоналі матриці (додатні) і такі ж добутки по висхідним діагоналям (від'ємні). Сумуючи шість добутків знаходимо детермінант. Наприклад

$$\Delta A' = \begin{vmatrix} 0,14 & 0,637 & 8,37 \\ 0,637 & 3,5743 & 40,269 \\ 8,37 & 40,269 & 523,49 \\ 0,14 & 0,637 & 8,37 \\ 0,637 & 3,5743 & 40,269 \end{vmatrix} = \begin{vmatrix} -250,40437 \\ -227,02393 \\ -212,41600 \\ +261,95544 \\ +214,70182 \\ +214,70182 \end{vmatrix} = 1,513628.$$

З урахуванням наведених раніше зауважень маємо

$$\Delta A = 1,513628 \times 10^6 = 1513628.$$

Параметри рівняння регресії способом детермінантів отримуємо за формулою (3.5)

$$a_i = \frac{\Delta A_i}{\Delta A},$$

тобто

$$a_0 = 3951780 : 1513628 = 2,6108;$$

$$a_1 = 828831 : 1513628 = 0,5476;$$

$$a_2 = 109889 : 1513628 = 0,0726.$$

Значення параметрів ідентичні, як і сподівалось, отриманим способом Гауса (крок 15).

Крок 17. Найбільш ефективним є **розрахунок коефіцієнтів регресії способом оберненої матриці**. Цей спосіб вигідний тим, що одночасно вирішує завдання визначення помилок оцінки коефіцієнтів регресії (крок 24) і апроксимації або прогнозу залежної змінної (крок 25), чого неможливо зробити інакше.

Спочатку отримуємо матрицю A^{-1} , обернену до матриці A , тобто

$$A = \begin{bmatrix} 14 & 63,7 & 837 \\ 63,7 & 357,43 & 4026,9 \\ 837 & 4026,9 & 52349 \end{bmatrix}.$$

Попередньо знайдемо мінори і алгебраїчні доповнення до кожного елемента матриці A і складемо матрицю $[\Delta A_{ij}]$. Наприклад алгебраїчне доповнення до елемента 14 складе

$$\Delta A_{11} = 1^{1+1} \begin{vmatrix} 3,5743 & 40,269 \\ 40,269 & 523,49 \end{vmatrix} \times 10^4 = 2,495173$$

до елемента 63,7 –

$$\Delta A_{12} = -1^{1+2} \begin{vmatrix} 0,637 & 40,269 \\ 8,37 & 523,49 \end{vmatrix} \times 10^4 = 35890.$$

До речі, індикатори знаків мінорів для матриці 3-го порядку такі

	j		
	+	-	+
i	-	+	-
	+	-	+

На основі матриці алгебраїчних доповнень

$$[\Delta A_{ij}] = \begin{bmatrix} 2495170 & 35890 & -42659 \\ 35890 & 32320 & -3060 \\ -42659 & -3060 & 946 \end{bmatrix}$$

за формулою (3.6) розрахуємо обернену матрицю

$$A^{-1} = \frac{1}{\Delta A} [\Delta A_{ij}],$$

Враховуючи, що $\Delta A = 1513628$ (крок 16), отримуємо

$$A^{-1} = \begin{bmatrix} 1,648475 & 0,023711 & -0,028183 \\ 0,023711 & 0,021353 & -0,002022 \\ -0,028183 & -0,002022 & 0,000625 \end{bmatrix}.$$

Параметри рівняння регресії знаходимо за формулою (3.7)

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = A^{-1} \times \begin{bmatrix} \sum P \\ \sum P\Phi \\ \sum PK \end{bmatrix}.$$

Отже, маємо:

$$a_0 = 1,648475 \times 132,2 + 0,023711 \times 654,39 - 0,028183 \times 8191 = 2,6108;$$

$$a_1 = 0,023711 \times 132,2 + 0,021353 \times 654,39 - 0,002022 \times 8191 = 0,5476;$$

$$a_2 = -0,028183 \times 132,2 - 0,002022 \times 654,39 - 0,000625 \times 8191 = 0,0726,$$

що ідентично отриманим раніше результатам.

Крок 18. Для контролю правильності розрахунків коефіцієнтів регресії рекомендується ще один варіант їх *розрахунку через β -коефіцієнти* за формулами (3.8) і (3.9)

$$a_i = \beta_i \frac{\sigma_p}{\sigma_{x_i}}, \quad a_0 = \bar{P} - \sum a_i \bar{X}_i.$$

Цей спосіб можливий лише за умови, що обидва фактори включені до рівняння регресії (див. підрозділ 5.5).

Отже, за формулами (3.8) і (3.9), отримуємо уже відомі нам значення коефіцієнтів регресії

$$a_\phi = 0,6155 \frac{1,9541}{2,1973} = 0,5476; \quad a_\kappa = 0,4772 \frac{1,9541}{12,8407} = 0,0726;$$

$$a_0 = 9,4429 - 0,5476 \times 4,5500 - 0,0726 \times 59,7857 = 2,6108.$$

Крок 19. *Розрахунок оцінок рентабельності* за рівнянням регресії (5.3) необхідний для *визначення помилок апроксимації*. Розрахунки доцільно ввести до табл. 5.7.

Таблиця 5.7 – Розрахунок оцінок рентабельності і помилок апроксимації

№ підприємства	P_j	Розрахунок оцінок				Помилка апроксимації $e_j = P_j - \hat{P}_j$	e_j^2
		2,61	0,548Ф	0,0726К	\hat{P}_j		
1	6,7	2,61	0,82	3,34	6,77	-0,07	0,0049
2	7,3	2,61	0,93	3,56	7,10	+0,20	0,0400
Всього	132,2	×	×	×	132,2	0	3,6164

За розрахунком $\sum e_j^2 = 3,6164$. Отже, дисперсія залишків D_e дорівнює $3,6164/14 = 0,2603$, що складає 6,83% загальної дисперсії 3,8187 (крок 5). Порівняйте цей результат з результатами кроку 11.

Крок 20. *Тестування гомоскедастичності залишків e_j* , полягає в перевірці того, що $D_e = \text{const}$. Це одне із основних припущень правомірності застосування МНК.

Використаємо тест рангової кореляції Спірмана (3.10)

$$r_s = 1 - \frac{6 \times \sum d_j^2}{n \times (n^2 - 1)}$$

Розрахунок $\sum d^2$ для нашого прикладу фрагментарно наведений у табл. 5.8.

Таблиця 5.8 – Розрахунок $\sum d^2$

J	P_j	Ранг P_j	$ e_j $	Ранг $ e_j $	d_j	d_j^2
1	6,7	2	0,07	1	7	7
...						
\sum	132,2	-	-	-	-	246

Таким чином, за формулою (3.10)

$$r_s = 1 - \frac{6 \times 246}{14 \times (14^2 - 1)} = 0,4593.$$

Оскільки t -статистика Стьюдента за формулою (2.3) дорівнює $t = r_s \sqrt{\frac{n-2}{1-r_s^2}} =$
 $= t = r_s \sqrt{\frac{n-2}{1-r_s^2}} = 0,4593 \sqrt{\frac{14-2}{1-0,4593^2}} = 1,791$ і не перевищує $t_{крит} = 2,179$ (див додаток 3), то з ймовірністю більшою 0,95 можна стверджувати, що залишки e_j гомоскедастичні і явище гетероскедастичності відсутнє.

Крок 21. Тестування автокореляції залишків e_j виконується за тестом DW Дарбіна-Уотсона (3.12).

$$DW = \frac{\sum (e_j - e_{j-1})^2}{\sum e_j^2}.$$

В табл. 5.9 виконані попередні розрахунки до формули (3.12).

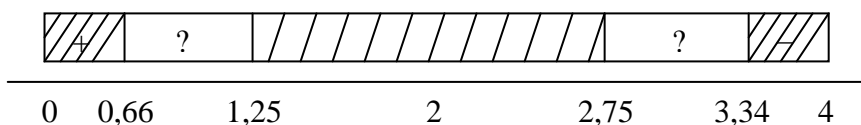
Таблиця 5.9 – Розрахунок $\sum e_j^2$ та $\sum (e_j - e_{j-1})^2$

j	e_j	e_j^2	$e_j - e_{j-1}$	$(e_j - e_{j-1})^2$
1	-0,07	0,0049	–	–
2	+0,20	0,0400	0,27	0,0729
3	+0,45	0,2025	0,25	0,0625
...
Σ	0	3,6164	–	8,7048

Таким чином, статистика Дарбіна-Уотсона складає

$$DW = \frac{8,7048}{3,6164} = 2,407.$$

За додатком 5 при $m = 2$ і $n = 14$ нижня межа $DW d_n = 0,66$, а верхня $d_g = 1,25$, тобто шкала DW така:



Отже, при $DW = 2,407$ маємо ситуацію повної відсутності автокореляції.

Висновок: параметри економетричної моделі рентабельності (5.3)

$$\hat{P} = 2,61 + 0,548\Phi + 0,0726K$$

знайдені за методом найменших квадратів, є незміщеними, обґрунтованими і ефективними.

Крок 22. ANOVA-дисперсійний аналіз виконується за результатами розрахунків на кроках 3 і 19 за формулою (3.16)

$$\sum (P - \bar{P})^2 = \sum (\hat{P} - \bar{P})^2 + \sum (P - \hat{P})^2.$$

За табл. 5.4 $D_p = 3,8187$ і відповідно $\sum (P - \bar{P})^2 = D_p \times n = 3,8187 \times 14 = 53,4618$,

а за табл. 5.7 $\sum (P - \hat{P})^2 = \sum e^2 = 3,6164$.

Отже, $\sum (P - \hat{P})^2 = 49,8454$, тобто за формулою (3.16).

$$53,4618 = 49,8454 + 3,6164.$$

$$(100\%) \quad (93,17\%) \quad (6,83\%)$$

З ANOVA-аналізу випливає, що:

1) коефіцієнт множинної кореляції за формулою (3.17)

$$R_{pфк} = \sqrt{1 - \frac{\sum e_j^2}{\sum (P - \bar{P})^2}}$$

складає

$$R_{pфк} = \sqrt{1 - \frac{3,6164}{53,4618}} = \sqrt{1 - 0,0683} = \sqrt{0,9317} = 0,9652,$$

що уже було двічі отримано раніше (крок 11);

2) коефіцієнт множинної детермінації дорівнює 0,9317, апроксимація складає 93,17%, а помилка апроксимації (і, доречі, прогнозу) – 6,83%.

Порівняйте ці результати з результатами кроків 11 і 19.

Крок 23. Тестування значущості (адекватності) рівняння регресії проводиться за F -статистикою Фішера (3.13).

$$\bar{F} \geq \bar{F}_{крит}; \quad F = S_y / S_e,$$

де S_y – систематична і S_e – залишкова дисперсії залежної змінної на одну ступінь свободи за формулами (3.15)

$$S_{\hat{y}} = \frac{\sum (\hat{y}_i - \bar{y})^2}{m}, \quad S_e = \frac{\sum e_j^2}{n - m - 1}.$$

Враховуючи, що $m = 2$; $n = 14$; $\sum (\hat{P} - \bar{P})^2 = 49,8454$; $\sum e_j^2 = 3,6164$ (див. крок 21), маємо

$$S_{\hat{P}} = \frac{49,8454}{2} = 24,923, \quad S_e = \frac{3,6164}{14 - 2 - 1} = 0,3298, \quad F = \frac{24,8454}{0,329} = 75,52.$$

$F_{крит}$ знаходимо за додатком 5, воно складає $4,02$. Оскільки $75,52 > 4,02$, робимо висновок, що економетрична модель рентабельності $\hat{P} = 2,61 + 0,548\Phi + 0,0726K$ є значущою, адекватною дійсності.

Крок 24. Тестування коефіцієнтів регресії на значущість (невипадковість) проводиться за t -статисткою Стьюдента (3.19).

$$t \geq t_{крит}; \quad t = \frac{a_i}{\sigma_{a_i}},$$

де σ_{a_i} – середньоквадратичні помилки коефіцієнта регресії. Ці помилки визначають за формулою (3.22)

$$\sigma_{a_i} = \sqrt{\frac{\sum e_j^2}{n - m - 1}} S_{ij} \quad (i = j),$$

де S_{ij} – елементи головної діагоналі оберненої матриці A^{-1} (див. крок 17).

У нашому прикладі маємо

$$\sigma_{a_0} = \sqrt{\frac{3,6164}{14 - 2 - 1}} \times 1,648475 = \sqrt{0,3288 \times 1,6485} = 0,736 \text{ коп.};$$

$$\sigma_{a_\phi} = \sqrt{0,3288 \times 0,02135} = 0,0838 \text{ коп.};$$

$$\sigma_{a_k} = \sqrt{0,3288 \times 0,000625} = 0,0143 \text{ коп.}$$

За формулою (3.19) знаходимо

$$t_0 = \frac{2,6108}{0,736} = 3,55; \quad t_\phi = \frac{0,5476}{0,0838} = 6,53; \quad t_k = \frac{0,0726}{0,0143} = 5,08.$$

За додатком 3 $t_{крит} = 2,201$. Отже, за t -критерієм Стьюдента всі параметри економетричної моделі рентабельності оцінені як значущі і не випадкові.

Крок 25. Побудова інтервалів прогнозу рентабельності за будь-яких прогнозних значень Φ і K може здійснюватися тільки за допомогою оберненої матриці. Довірча межа помилки прогнозу визначається за формулою (3.25)

$$\sigma_{\hat{y}}^2 = \frac{\sum e_j^2}{n - m - 1} \times [X'] \times [A^{-1}] \times [X]$$

або для нашого випадку

$$\sigma_{\hat{P}}^2 = 0,3288 \times [X'] \times \begin{bmatrix} 1,648475 & 0,023711 & -0,028183 \\ 0,023711 & 0,021353 & -0,002022 \\ -0,028183 & -0,002022 & 0,00625 \end{bmatrix} \times [X].$$

Нехай за прогнозом $\Phi = 7$ тис.грн. і $K = 65\%$. Точковий прогноз рентабельності складе $\hat{P} = 2,61 + 0,548 \times 7 + 0,0726 \times 65 = 11,17$ коп./грн.

Дисперсія помилки прогнозу рентабельності складе

$$\sigma_{\hat{P}}^2 = 0,3288 \begin{vmatrix} 1 & 7 & 65 \end{vmatrix} \begin{bmatrix} 1,648475 & 0,023711 & -0,028183 \\ 0,023711 & 0,021353 & -0,002022 \\ -0,028183 & -0,002022 & 0,00625 \end{bmatrix} \begin{bmatrix} 1 \\ 7 \\ 65 \end{bmatrix}.$$

Спочатку множимо $[A^{-1}]$ на $[X]$, отримуємо

$$\begin{bmatrix} 1,648475 \times 1 & 0,023711 \times 7 & -0,028183 \times 65 \\ 0,023711 \times 1 & 0,021353 \times 7 & -0,002022 \times 65 \\ -0,028183 \times 1 & -0,002022 \times 7 & 0,00625 \times 65 \end{bmatrix} = \begin{bmatrix} -0,017443 \\ 0,041752 \\ -0,001712 \end{bmatrix}.$$

Далі

$$\sigma_{\hat{P}}^2 = 0,3288 \begin{vmatrix} 1 & 7 & 65 \end{vmatrix} \begin{bmatrix} -0,017443 \\ 0,041752 \\ -0,001712 \end{bmatrix} = 0,3288 \times [1 \times (-0,017443 + 7 \times 0,041752 + 65 \times (-0,001712))] = 0,3288 \times (-0,017443 + 0,292264 - 0,111280) = 0,3288 \times 0,163541 = 0,05377.$$

Оскільки $\sigma_{\hat{P}}^2 = 0,05377$, $\sigma_{\hat{P}} = \sqrt{0,05377} = 0,23$ коп.

Таким чином, якщо прогнози фондоозброєності праці очікується 7 тис. грн./чол., а кооперування виробництва – 65%, точковий прогноз рентабельності

дорівнює 11,17 коп./грн., а інтервальний – у межах $11,17 - 0,23 \leq P \leq 11,17 + 0,23$, або $10,94 \div 11,40$.

Крок 26. Економічна інтерпретація результатів моделювання полягає:

- у чіткому викладі змісту і виміру змінних P , Φ і K , а також параметрів рівняння регресії 2,61, 0,548 і 0,0726;
- у конкретному тлумаченні довірчих меж помилки прогнозу $\pm 0,23$ коп.;
- у визначенні сфер застосування розробленої економетричної моделі.

5.5. ОСОБЛИВОСТІ ОЦІНЮВАННЯ ПАРАМЕТРІВ РІВНЯННЯ ПАРНОЇ РЕГРЕСІЇ

У деяких варіантах завдання на **13-му** кроці доведеться вилучати якийсь фактор із подальшого процесу моделювання. Нехай у нашому прикладі це фактор K . Тоді економетрична модель рентабельності стає рівнянням парної регресії (**крок 14**)

$$\hat{P} = a_0 + a_\phi \Phi.$$

Крок 15. Система рівнянь за МНК (див. табл. 5.4 і 5.5) має таке рішення способом Гаусса:

$132,2 = 14a_0 + 63,7a_\phi$ $654,39 = 63,7a_0 + 357,43a_\phi$: : 14 : 63,7
$9,4429 = a_0 + 4,5500a_\phi$ $10,2730 = a_0 + 5,6111a_\phi$	віднімемо 1-е рівняння із 2-го
$0,8301 = 1,0611a_\phi$	маємо: $a_\phi = 0,8301 : 1,0611 = 0,7823$
$a_0 = 10,2730 - 5,6111 \times 0,7823 = 5,8834$	

Таким чином, рівняння регресії набуває вигляду:

$$\hat{P} = 5,8834 + 0,7823\Phi.$$

Крок 16. Для розрахунку системи рівнянь за МНК способом детермінантів попередньо визначають детермінанти

$$\Delta A = \begin{vmatrix} 14 & 63,7 \\ 63,7 & 357,43 \end{vmatrix} = 5004,02 - 4057,69 = 946,33;$$

$$\Delta A_0 = \begin{vmatrix} 132,2 & 63,7 \\ 654,39 & 357,43 \end{vmatrix} = 47252,25 - 41684,64 = 5567,61;$$

$$\Delta A_\phi = \begin{vmatrix} 14 & 132,2 \\ 63,7 & 654,39 \end{vmatrix} = 9161,46 - 8421,14 = 740,32.$$

Отже, за формулою (3.5) маємо ті ж самі оцінки параметрів рівняння регресії

$$a_0 = \frac{5567,61}{946,33} = 5,8834; \quad a_\phi = \frac{740,32}{946,33} = 0,7823.$$

Крок 17. Вихідними даними для побудови **оберненої матриці** A^{-1} є (крок 16):

$$A = \begin{vmatrix} 14 & 63,7 \\ 63,7 & 357,43 \end{vmatrix}; \quad \Delta A = 946,33.$$

Спочатку побудуємо матрицю алгебраїчних доповнень до елементів матриці A

$$A_{11} = -1^{1+1} \times 357,43 = 357,43;$$

$$A_{12} = -1^{1+2} \times 63,7 = -63,7;$$

$$A_{21} = -1^{2+1} \times 63,7 = -63,7;$$

$$A_{22} = -1^{2+2} \times 14 = 14.$$

Отже, маємо:

$$A_{ij} = \begin{bmatrix} 357,43 & -63,7 \\ -63,7 & 14 \end{bmatrix}$$

і за формулою (3.6) обернену матрицю

$$A^{-1} = \frac{1}{946,33} \begin{bmatrix} 357,43 & -63,7 \\ -63,7 & 14 \end{bmatrix} = \begin{bmatrix} 0,377701 & -0,067313 \\ -0,067313 & 0,014794 \end{bmatrix}.$$

Насамкінець, за формулою (3.7) отримуємо коефіцієнти регресії:

$$a_0 = 0,377701 \times 132,2 - 0,067313 \times 654,39 = 5,8834;$$

$$a_\phi = -0,067313 \times 132,2 + 0,014794 \times 654,39 = 0,7823,$$

котрі ідентичні результатам оцінювання іншими способами.

Крок 18 у такому разі не виконується, бо користуватися β -коефіцієнтами, визначеними для двофакторної моделі, не можна.

Крок 19. Помилки апроксимації e_j визначимо у таблиці:

j	P_j	$\hat{P} = 5,8834 + 0,7823\Phi_j$	$e_j = P_j - \hat{P}$	e_j^2
1	6,7	7,06	-0,36	0,1296
2	7,3	7,21	0,09	0,0081
3	11,5	10,89	0,61	0,3721
...				
14	9,8	10,26	-0,46	0,2116
15	10,4	11,44	-1,04	1,0816
Σ	132,2	132,2	0,0	12,1382

Крок 20. Для тестування гомоскедастичності залишків e_j , тобто відсутності явища гетероскедастичності, розрахуємо коефіцієнт рангової кореляції Спірмана, для чого визначимо ранги P_j і e_j

j	P_j	Ранг P	e_j	Ранг $ e_j $	d	d^2
1	6,7	2	-0,36	4	2	4
2	7,3	3	0,09	1	2	4
3	11,5	13	0,61	7	6	36
...						
14	9,8	8	-0,46	5	3	9
15	10,4	10	-1,04	10	0	0
Σ	132,2	-	0,0	-	-	430

Таким чином, коефіцієнт рангової кореляції за формулою (3.10) становить

$$r_3 = 1 - \frac{6 \times 430}{14(14^2 - 1)} = 0,0549,$$

а t -статистика Стьюдента за формулою (2.3)

$$t = 0,0549 \sqrt{\frac{14-1-1}{1-0,0549^2}} = 0,19$$

при критичному значенні 2,179 (додаток 3). Оскільки $0,19 < 2,179$ гетероскедастичність відсутня. Параметри моделі є незміщеними і обґрунтованими.

Крок 21. Для вирахування DW -статистики Дарбіна-Уотсона виконаємо попередні розрахунки

j	e_j	$e_j - e_{j-1}$	$(e_j - e_{j-1})^2$
1	-0,36	-	-
2	0,09	0,45	0,2025
3	0,61	0,52	0,2704
...			
14	-0,46	-1,57	2,4649
15	-1,04	-0,58	0,3304
Σ	0,0	-	27,4620

Отже, (див. крок 21) за формулою (3.12)

$$DW = \frac{27,4620}{12,1382} = 2,26.$$

Критичні значення DW при $n = 14$ і $m = 1$ за додатком 6 знайдемо такі: $d_n = 0,78$, $d_g = 105$. Отже, значення 2,26 знаходиться у зоні від $1,05(0 + 1,05)$ до $2,95(4 - 1,05)$, що дає підстави вважати автокореляцію залишків відсутньою, а параметри моделі, знайдені за МНК, обґрунтованими і незміщеними.

Крок 22. ANOVA-аналіз полягає у розкладанні загальної девіації рентабельності на систематичну і залишкову (див. кроки 5 і 19, формула 3.16):

$$\sum (P - \bar{P})^2 = 3,8187 \times 14 = 53,4618;$$

$$\sum (P - \hat{P})^2 = \sum e_j^2 = 12,1382;$$

$$\sum (\hat{P} - \bar{P})^2 = 53,4618 - 12,1382 = 41,3236.$$

Отже,

$$53,4618 = 41,3236 + 12,1382.$$

$$(100\%) (77,2956\%) (22,7044)$$

Помилка апроксимації складає 22,704%. Коефіцієнт кореляції дорівнює (див. крок 22)

$$R_{p\phi} = \sqrt{0,772956} = 0,8796.$$

Крок 23. Адекватність (достовірність) моделі перевіряється за F -статистикою Фішера (формули 3.13, 3.15):

$$F \geq F_{крит}; \quad F = S_{\hat{P}}/S_e,$$

$$S_{\hat{P}}^2 = \frac{\sum (\hat{P} - \bar{P})^2}{m}; \quad S_e^2 = \frac{\sum e_j^2}{m - n - 1}.$$

Отже,

$$S_{\hat{P}}^2 = \frac{41,3236}{1} = 41,3236; \quad S_e^2 = \frac{12,1382}{14 - 1 - 1} = 1,0115.$$

Оскільки $F = 41,3236/1,0115 = 40,85$ і $40,85 > 4,62$ (див. додаток 5), робимо висновок про адекватність моделі дійсності.

Крок 24. Значущість параметрів моделі перевіряється за t -статистикою Стьюдента (формули 3.19)

$$t_i \geq t_{крит}; \quad t_i = \sigma_{a_i}/a_i.$$

Середньоквадратичні помилки параметрів a_0 і a_1 визначають за формулами (5.4) і (5.5)

$$\sigma_{a_0} = \sqrt{\frac{\Phi^2 \sum e_j^2}{n \times (n - 2) \sigma_{\phi}^2}}, \quad (5.4)$$

$$\sigma_{a_i} = \sqrt{\frac{\sum e_j^2}{n \times (n - 2) \sigma_{\phi}^2}}. \quad (5.5)$$

У нашому прикладі ці помилки дорівнюють

$$\sigma_{a_0} = \sqrt{\frac{25,5307 \times 12,1382}{14(14 - 2) \times 4,8282}} = \sqrt{\frac{309,8967}{811,1376}} = 0,6181;$$

$$\sigma_{a_1} = \sqrt{\frac{12,1382}{14(14 - 2) \times 4,8282}} = \sqrt{\frac{12,1382}{811,1376}} = 0,1223.$$

Помилки коефіцієнтів регресії можна визначити також за допомогою елементів головної діагоналі оберненої матриці за формулою (3.22):

$$\sigma_{a_0} = \sqrt{\frac{12,1382}{14-2}} \times 0,377701 = 0,6181;$$

$$\sigma_{a_1} = \sqrt{\frac{12,1382}{14-2}} \times 0,014794 = 0,1223.$$

Таким чином, t -статистики

$$t_0 = \frac{5,8834}{0,6181} = 9,52; \quad t_1 = \frac{0,7823}{0,1223} = 6,40.$$

значно перевищують $t_{крит} = 2,179$. Це значить, що параметри моделі значущі, не випадкові.

Крок 25. Побудова **інтервалів прогнозу** рентабельності за моделлю можлива у двох варіантах:

а) за формулою (3.24)

$$\sigma_{\hat{p}} = \sqrt{\frac{\sum e_j^2}{n-2}} \times \sqrt{\frac{1}{n} + \frac{(\Phi_n - \bar{\Phi})^2}{\sum (\Phi_j - \bar{\Phi})^2}};$$

б) за допомогою оберненої матриці за формулою – див. крок 25 для багатофакторної регресії.

Визначимо $\sigma_{\hat{p}}$ за формулою (3.24), що за даними нашого прикладу має вигляд

$$\sigma_{\hat{p}} = 1,0057 \sqrt{0,0714 + \frac{(\Phi_n - 4,55)^2}{67,5948}},$$

для ряду значень Φ , наведених у наступній таблиці:

Φ	$\sigma_{\hat{p}}$	Φ	$\sigma_{\hat{p}}$
1	0,511	5	0,274
2	0,412	6	0,322
3	0,329	7	0,405
4	0,279	8	0,500
4,55	0,268	9	0,607

Побудуємо за даними таблиці лінію точкових прогнозів \hat{P} і довірчі інтервали для прогнозу (рис. 5.3).

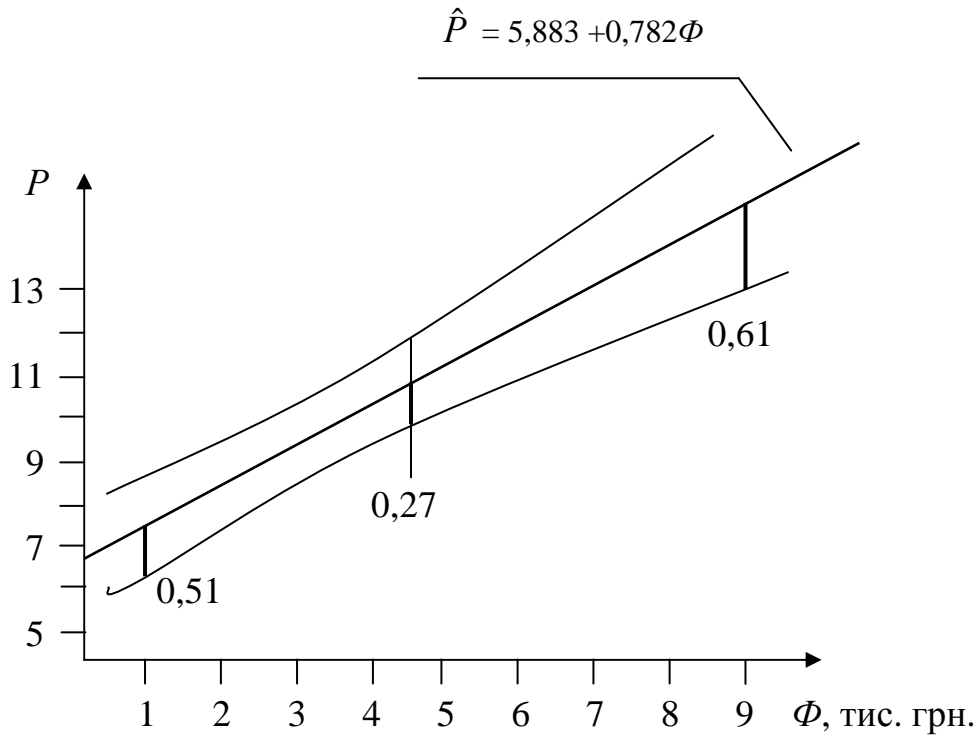


Рис. 5.3

Наприклад, для $\Phi = 9$ тис. грн./чол. – точковий прогноз:

$$\hat{P}_n = 5,883 + 0,782 \times 9 = 12,92 \text{ коп.},$$

а інтервальний

$$12,92 - 0,61 \leq \hat{P}_n \leq 12,92 + 0,61 \quad \text{або} \quad 12,31 \div 13,53.$$

Крок 26. Дивитися вище.

РЕКОМЕНДОВАНА НАВЧАЛЬНА ЛІТЕРАТУРА

1. Доля В.Т. Статистическое моделирование производственных систем и процессов: Уч. пособие. – К.: УМКВО, 1988. – 140 с.
2. Лук'яненко І.Г., Краснікова Л.І. – Економетрика: Підручник. – К.: Знання, КОО, 1998. – 494 с.
3. Наконечний С.І., Терещенко Т.О., Романюк Т.П. Економетрія: Підручник. – К.: КНЕУ, 2002. – 296 с.
4. Доля В.Т. Економетрія: Конспект лекцій. – Харків: ХНАМГ, 2006. – 93 с.
5. Лук'яненко І.Г., Краснікова Л.І. – Економетрика: Практикум з використанням комп'ютера. – К.: Знання, КОО, 1998. – 220 с.
6. Медведєв М.Г. Економетричні методи моделювання: Навч. посібник. – К.: Вид-во Європ. ун-ту, 2003. – 140 с.
7. Доля В.Т. Економетрія: Методичні вказівки з самостійного вивчення дисципліни (для студентів за напрямами підготовки «Економіка підприємства», «Облік і аудит», «Менеджмент»). Вид. 3-є, – Харків: ХНАМГ, 2008. – 48 с.

КОРОТКИЙ УКРАЇНСЬКО-РОСІЙСЬКИЙ СЛОВНИК МАТЕМАТИКО-СТАТИСТИЧНИХ ПОНЯТЬ І ТЕРМІНІВ

автокореляція	– автокорреляция
автономність впливу	– автономность влияния
авторегресія	– авторегрессия
авторегресивна модель	– авторегрессионная модель
агрегування	– агрегирование
адекватність	– адекватность
аномальний об'єкт спостереження	– аномальный объект наблюдения
апроксимація	– аппроксимация
багатофакторна модель	– многофакторная модель
варіаційний розмах	– вариационный размах
варіація	– вариация
вектор	– вектор
взаємозв'язок	– взаимосвязь
відношення	– отношение
відсоток	– процент
відсутність	– отсутствие
відхилення	– отклонение
від'ємний	– отрицательный
вісь	– ось
вибірка	– выборка
вибіркова сукупність	– выборочная совокупность
визначення	– определение
визначник	– определитель (детерминант)
викривлення	– искажение
вилучення	– исключение
вимірювач (вимірювальник)	– измеритель
вимога	– требование
випадкова змінна	– случайная переменная
вираз	– выражение
вплив	– влияние
гетероскедастичність	– гетероскедастичность
гомоскедастичність	– гомоскедастичность
граничний	– граничный, предельный
границя(межа)	– граница, предел
групування	– группирование

двомірне розсіювання
девіація
детермінант, визначник
детермінація
добуток
доведення
довільний
довірчий інтервал
додатний
достатній
дисперсійний аналіз
дистрибутивно-лагова модель

екзогенна змінна
економетрія
економетрична модель
експонента
еластичність
емпірична формула
ендогенна змінна

загальна дисперсія
залежність
зайвий фактор
залишок
залишкова дисперсія
запізнююча змінна
засіб
зведення
зв'язок
зіставлення
зміщені оцінки
знаменник
значущість

ідентифікація
імовірність, ймовірність
інтерпретація

квадратичне відхилення
кількісний вимір
кластерний аналіз
колінеарність
конфлюентний аналіз

Д

– двумерное рассеяние
– девиация
– детерминант, определитель
– детерминация
– произведение
– доказательство
– произвольный
– доверительный интервал
– положительный
– достаточный
– дисперсионный анализ
– дистрибутивно-лаговая модель

Е

– экзогенная переменная
– эконометрия
– эконометрическая модель
– экспонента
– эластичность
– эмпирическая формула
– эндогенная переменная

З

– общая дисперсия
– зависимость
– избыточный, излишний фактор
– остаток
– остаточная дисперсия
– запаздывающая переменная
– способ
– сводка
– связь
– сопоставление
– смещенные оценки
– знаменатель
– значимость

І

– идентификация
– вероятность
– интерпретация

К

– квадратическое отклонение
– количественное измерение
– кластерный анализ
– коллинеарность
– конфлюэнтный анализ

кореляція
кореляційна залежність
кореляційний аналіз
коридор регресії
коригування
крок
кут нахилу

лінеаризація
лінійна залежність

мала вибірка
математичне сподівання
матриця
мірність
множинна модель
множинна регресія
моделювання
мультиколінеарність

надійність оцінки
нахил прямої
наявність залежності
насичення регресії
нескінченність
нехтувати

обґрунтування
обернена матриця
обмеження
обсяг
об'єм
операційний
опуклий
оцінка
оцінювання

парціальна кореляція
перетворення
перетин
площина
поверхня
показник
поле кореляції
помилка апроксимації

– корреляция
– корреляционная зависимость
– корреляционный анализ
– коридор регрессии
– корректировка
– шаг
– угол наклона

Л

– линеаризация
– линейная зависимость

М

– малая выборка
– математическое ожидание
– матрица
– мерность
– множественная модель
– множественная регрессия
– моделирование
– мультиполинеарность

Н

– надежность оценки
– наклон прямой
– наличие зависимости
– насыщение регрессии
– бесконечность
– пренебрегать

О

– обоснование
– обратная матрица
– ограничение
– объем
– объем
– операционный
– выпуклый
– оценка
– оценивание

П

– парциальная корреляция
– преобразование
– пересечение
– плоскость
– поверхность
– показатель
– корреляционное поле
– ошибка аппроксимации

помилка вибірки
похідна
прийнятне рішення
припущення
прямокутник
псевдокореляція
питома вага

розкид значень
регресійний аналіз
регресія
рівняння регресії
різниця
розбіжність результатів
розмах
розподіл
розрахунок
розсіювання
розшарування
рядок матриці

симультаивна модель
скінченний
складання системи рівнянь
специфікація
співвідношення
ступені вільності, свободи
стовпець матриці
стохастичність
сукупність об'єктів
суттєвість (невипадковість) впливу

тіснота зв'язку
тотожність

умова

фактор (чинник)
форма зв'язку
формування

чисельник

– ошибка выборки
– производная
– приемлемое решение
– допущение
– прямоугольник
– псевдкорреляция
– удельный вес

Р

– разброс значений
– регрессионный анализ
– регрессия
– уравнение регрессии
– разность
– расхождение результатов
– размах
– распределение
– расчет
– рассеяние
– расслоение
– строка матрицы

С

– симультаивная модель
– конечный
– составление системы уравнений
– спецификация
– соотношение
– степени свободы
– столбец матрицы
– стохастичность
– совокупность объектов
– существенность (неслучайность) влияния

Т

– теснота связи
– тождество

У

– условие

Ф

– фактор
– форма связи
– формирование

Ч

– числитель

шар
штучний фактор

щільність поля кореляції
щільність розподілу

якісний показник

Ш

– слой
– искусственный фактор

Щ

– плотность поля корреляции
– плотность распределения

Я

– качественный показатель

**Значення P при розрахункових значеннях t
(закон нормального розподілу)***

$t (Z)$	$P (A)$	$t (Z)$	$P (A)$
0,67	0,4971	1,60	0,8904
0,68	0,5035	1,70	0,9109
0,69	0,5098	1,80	0,9281
0,70	0,5161	1,90	0,9426
0,75	0,5468	1,96	0,9500
0,80	0,5763	2,00	0,9545
0,85	0,6047	2,10	0,9643
0,90	0,6319	2,20	0,9722
0,95	0,6579	2,30	0,9786
1,00	0,6827	2,40	0,9836
1,05	0,7063	2,50	0,9876
1,10	0,7287	2,60	0,9907
1,15	0,7499	2,70	0,9931
1,20	0,7699	2,80	0,9949
1,25	0,7887	2,90	0,9963
1,30	0,8064	3,00	0,9973
1,35	0,8230	3,25	0,9986
1,40	0,8385	3,50	0,9995
1,45	0,8529	3,75	0,9998
1,50	0,8664	4,00	0,9999

* Наконечний С.І., Терещенко Т.О., Романюк Т.П. Економетрія: Підручник. – К.: КНЕУ, 2000. – С. 280-281.

**Критичні значення t для побудови прямокутного шаблону
двомірного розсіювання***

Об'єм вибірки (n)	Ймовірність	
	0,95	0,99
5	3,04	5,04
10	2,37	3,41
15	2,22	3,08
20	2,15	2,93
25	2,10	2,85
30	2,08	2,8
40	2,05	2,74
50	2,03	2,71
70	2,01	2,67
100	1,99	2,64
∞	1,96	2,58

* Румшиский Л.З. Математическая обработка результатов эксперимента: Справочное руководство. – М.: Наука, 1971. – С. 18 и 173.

**Значення t -критерію Стьюдента
(для $P = 0,95$)***

Число ступенів свободи ($n - m - 1$)	t	Число ступенів свободи ($n - m - 1$)	t
1	12,706	20	2,086
2	4,303	22	2,074
3	3,182	24	2,064
4	2,776	27	2,052
5	2,571	30	2,042
6	2,447	35	2,035
7	2,365	40	2,021
8	2,306	50	2,012
9	2,262	60	2,000
10	2,228	75	1,993
11	2,201	90	1,989
12	2,179	120	1,980
13	2,160	250	1,970
14	2,145	∞	1,960
16	2,120		
18	2,101		

* Медведєв М.Г. Економетричні методи моделювання: Навч. посібник. – К.: Вид-во Європ. ун-ту, 2003. – С. 131.

Z-перетворення Фішера

$$z = \frac{1}{2} \ln \frac{1+r}{1-r}$$

	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
0,1	0,100	0,110	0,121	0,131	0,141	0,151	0,161	0,172	0,182	0,192
0,2	0,203	0,213	0,224	0,234	0,245	0,255	0,266	0,277	0,288	0,299
0,3	0,310	0,321	0,332	0,343	0,354	0,365	0,377	0,388	0,400	0,412
0,4	0,424	0,436	0,448	0,460	0,472	0,485	0,497	0,510	0,523	0,536
0,5	0,549	0,563	0,576	0,590	0,604	0,618	0,633	0,648	0,663	0,678
0,6	0,693	0,709	0,725	0,741	0,758	0,775	0,793	0,811	0,829	0,848
0,7	0,867	0,887	0,908	0,929	0,951	0,973	0,996	1,020	1,045	1,071
0,8	1,099	1,127	1,157	1,188	1,221	1,256	1,293	1,333	1,376	1,422
0,9	1,472	1,528	1,589	1,658	1,738	1,832	1,946	2,092	2,298	2,647
0,99	2,647	2,700	2,759	2,826	2,903	2,995	3,106	3,250	3,453	3,800

Приклади.

1. Нехай $r = 0,69$, тоді значення z знаходиться на перетині рядка 0,6 і стовпчика 9, воно дорівнює 0,848. Можна перевірити за формулою z -перетворення Фішера

$$z = \frac{1}{2} \ln \frac{1+0,69}{1-0,69} = 0,848.$$

2. Нехай $z = 1,291$. тоді знаходимо в таблиці найближче до нього – 1,293 (рядок 0,8 і стовпчик 6). Коефіцієнт кореляції дорівнює 0,86.

Значення F -критерію Фішера
(для $P = 0,95$)

Чисельник Знаменник	1	2	3	4	5	7	10	15	20	25	30	40	60	120	∞
1	161	200	216	225	230	237	242	246	248	249	250	251	252	253	254
2	18,51	19,00	19,20	19,20	19,30	19,40	19,40	19,40	19,40	19,40	19,50	19,50	19,50	19,50	19,50
3	10,13	9,55	9,28	9,12	9,01	8,89	8,79	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	6,09	5,96	5,86	5,80	5,78	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,88	4,74	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37
7	5,59	4,74	4,35	4,12	3,97	3,79	3,64	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
10	4,96	4,10	3,71	3,48	3,33	3,14	2,98	2,85	2,77	3,74	2,70	2,66	2,62	2,58	2,54
15	4,54	3,68	3,29	3,06	2,9	2,71	2,54	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
20	4,35	3,49	3,10	2,87	2,71	2,51	2,35	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
25	4,24	3,39	2,99	2,76	2,60	2,40	2,24	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
30	4,17	3,32	2,92	2,69	2,53	2,33	2,16	2,01	1,93	1,88	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,25	2,08	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,17	1,99	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,09	1,91	1,75	1,66	1,60	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,6	2,37	2,21	2,01	1,83	1,67	1,57	1,51	1,46	1,39	1,32	1,22	1,00

Критичні значення критерію Дарбіна-Уотсона

(для $P = 0,99$)

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$	
	d_n	d_g	d_n	d_g	d_n	d_g	d_n	d_g	d_n	d_g
10	0,60	1,00	0,47	1,33	0,34	1,73	0,23	2,19	0,15	2,64
12	0,70	1,02	0,57	1,27	0,45	1,58	0,34	1,91	0,24	2,28
14	0,78	1,05	0,66	1,25	0,55	1,49	0,46	1,76	0,34	2,05
16	0,84	1,09	0,74	1,25	0,63	1,45	0,53	1,66	0,44	1,90
18	0,90	1,12	0,81	1,26	0,71	1,42	0,61	1,60	0,52	1,80
20	0,95	1,15	0,87	1,27	0,77	1,41	0,69	1,57	0,60	1,74
22	1,00	1,17	0,92	1,28	0,83	1,41	0,75	1,54	0,67	1,69
24	1,04	1,20	0,96	1,30	0,88	1,41	0,81	1,53	0,73	1,66
26	1,07	1,22	1,00	1,31	0,93	1,41	0,86	1,52	0,78	1,64
28	1,10	1,24	1,04	1,33	0,97	1,42	0,90	1,51	0,83	1,62
30	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
32	1,16	1,28	1,10	1,35	1,04	1,43	0,98	1,51	0,92	1,60
35	1,20	1,31	1,14	1,37	1,09	1,44	1,03	1,51	0,97	1,59
40	1,21	1,34	1,20	1,40	1,15	1,46	1,10	1,52	1,04	1,59
45	1,24	1,38	1,25	1,42	1,20	1,47	1,16	1,53	1,11	1,59
50	1,32	1,40	1,29	1,45	1,25	1,49	1,21	1,54	1,16	1,59
70	1,43	1,49	1,40	1,52	1,37	1,55	1,34	1,58	1,31	1,61
100	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63	1,44	1,65
200	1,66	1,68	1,65	1,69	1,64	1,70	1,63	1,72	1,62	1,73

ЗМІСТ

	Стор.
Вступ	3
1. Змістовий модуль 1: Постановка задачі економетричного моделювання	6
1.1. Предмет, завдання і зміст економетричного моделювання	6
1.1.1. Предмет економетрії	6
1.1.2. Проблеми і завдання економетричного моделювання	8
1.1.3. Зміст (послідовність) економетричного моделювання	10
1.2. Формування матриці даних для економетричного моделювання	12
1.2.1. Загальна характеристика матриці	12
1.2.2. Змінні у матриці	13
1.2.3. Об'єкти спостереження в матриці	15
1.2.4. Вимоги до розмірів матриці	17
1.2.5. Показники варіації змінних	21
1.2.6. Поля кореляції та їх аналіз	22
1.2.7. Вилучення аномальних об'єктів спостереження	26
1.3. Комплекс контрольних завдань за ЗМ 1 «Постановка задачі економетричного моделювання»	29
1.3.1. Тестові завдання	29
1.3.2. Логічні вправи	34
1.3.3. Розрахункові вправи	37
1.3.4. Відповіді до розрахункових вправ	40
2. Змістовий модуль 2: Специфікація економетричних моделей	41
2.1. Ідентифікація незалежних змінних	41
2.1.1. Мета і послідовність ідентифікації	41
2.1.2. Коефіцієнт парної кореляції і детермінації	42
2.1.3. Тестування суттєвості (невипадковості) коефіцієнтів кореляції	45
2.1.4. Інтервали довіри для коефіцієнтів кореляції	45
2.1.5. Мультиколінеарність	47
2.1.6. Бета-коефіцієнти	50
2.1.7. Тестування автономії екзогенних змінних	52
2.1.8. Коефіцієнти множинної кореляції і детермінації	53
2.1.9. Тестування значущості вкладу факторів до множинної детермінації	55
2.1.10. Вилучення екзогенних змінних	57

2.2. Специфікація аналітичної форми рівнянь регресії	57
2.2.1. Мета і способи специфікації	57
2.2.2. Аналітичні форми рівнянь регресії	59
2.2.3. Спосіб перших різниць	63
2.2.4. Лінеаризація нелінійних рівнянь регресії	64
2.3. Комплекс контрольних завдань за ЗМ 2 «Специфікація економетричних моделей»	65
2.3.1. Тестові завдання	65
2.3.2. Логічні вправи	69
2.3.3. Розрахункові вправи	75
2.3.4. Відповіді до розрахункових вправ	79
3. Змістовий модуль 3: Оцінювання параметрів економетричних моделей	81
3.1. Оцінювання параметрів рівнянь регресії	81
3.1.1. Мета і вимоги до оцінювання параметрів	81
3.1.2. Основні припущення щодо оцінювання параметрів	82
3.1.3. Метод найменших квадратів	83
Спосіб Гауса	86
Спосіб детермінантів	87
Спосіб оберненої матриці	89
Спосіб β -коефіцієнтів	91
3.1.4. Виконання за МНК основних припущень щодо оцінювання параметрів	93
3.1.5. Гетероскедастичність	94
3.1.6. Автокореляція	96
3.1.7. Значущість (адекватність) рівняння регресії	99
3.1.8. Перевірка значущості параметрів моделі	101
3.1.9. Інтервали довіри до коефіцієнтів регресії	103
3.2. Прогнозування залежної змінної	103
3.2.1. Прогнозування на парних моделях	103
3.2.2. Прогнозування на множинних моделях	106
3.3. Комплекс контрольних завдань за ЗМ 3 «Оцінювання параметрів економетричних моделей»	108
3.3.1. Тестові завдання	108
3.3.2. Логічні вправи	111
3.3.3. Розрахункові вправи	113
3.3.4. Відповіді до розрахункових вправ	119

4. Самостійне вивчення економетрії	121
4.1. Зміст і самоконтроль самостійної теоретичної підготовки	121
4.2. Зміст і самоконтроль самостійної практичної підготовки	124
5. Завдання для практичних занять, виконання індивідуального домашнього завдання або контрольної роботи	125
5.1. Зміст завдань та вихідні дані	125
5.2. Варіанти завдань	126
5.3. Послідовність розробки рівняння регресії	129
5.4. Методика розробки рівняння регресії	131
5.5. Особливості оцінювання параметрів рівняння парної регресії	149
Рекомендована навчальна література	156
Короткий українсько-російський словник математико-статистичних понять і термінів	157
Додаток 1. Значення P для розрахункових значень t (закон нормального розподілу)	162
Додаток 2. Критичні значення t для побудови прямокутного шаблону двовірного розсіювання	163
Додаток 3. Значення t -критерію Стьюдента	164
Додаток 4. Z -перетворення Фішера	165
Додаток 5. Значення F -критерію Фішера	166
Додаток 6. Критичні значення критерію Дарбіна-Уотсона	167

Навчальне видання

Доля Володимир Тимофійович

ЕКОНОМЕТРІЯ

НАВЧАЛЬНИЙ ПОСІБНИК

Для студентів вищих навчальних закладів
за спеціальностями з економіки і менеджменту

*Відповідальний за випуск Т.В. Момот
Редактор Д.Ф. Курильченко
Комп'ютерне верстання Г.О. Павлова*

Підп. до друку 02.08.2009 р.
Друк на ризографі.
Тираж 500 пр.

Формат 60×84 1/16
Ум. друк. арк. 7,4
Зам. №

Видавець і виготовлювач:
Харківська національна академія міського господарства,
вул. Революції, 12, Харків, 61002
Електронна адреса: rectorat@ksame.kharkov.ua
Свідоцтво суб'єкта видавничої справи: ДК №731 від 19.12.2001