МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Дидактичні матеріали для підготовки до практичних занять з дисципліни «Економетрика»

для студентів денної та заочної форм навчання

Тернопіль-2022

УДК 519.2

Рецензенти:

С.В. Мартинюк – к. ф.-м. н., доцент кафедри математики та методики її викладання Тернопільського національного педагогічного університету імені Володимира Гнатюка

О.С. Башуцька – к. е. н., доцент кафедри економічної кібернетики та інформатики Західноукраїнського національного університету;

затверджено на засіданні кафедри прикладної математики, протокол № 1 від 26.08.2022 р.

Пласконь С.А., Березька К.М., Мартинюк О.М., Руська Р.В., Єрьоменко В.О., Дзюбановська Н.В. Дидактичні матеріали для підготовки до практичних занять з дисципліни «Економетрика» для студентів денної та заочної форм навчання. Тернопіль: ЗУНУ, 2022. 107 с.

У запропонованих «Дидактичних матеріалах для підготовки до практичних занять з дисципліни «Економетрика» для студентів денної та заочної форм навчання наведений основний теоретичний матеріал та приклади розв’язування задач.

Відповідальний за випуск: О.М. Мартинюк, кандидат фіз.-мат. наук, завідувач кафедри прикладної математики ЗУНУ

1. Загальні принципи побудови економетричних моделей. Моделі парної регресії

1.1. Регресійний аналіз та його особливості

Під регресією розуміють односторонню стохастичну залежність однієї випадкової змінної від другої чи декількох інших випадкових змінних. Таким чином, регресія встановлює відповідність між випадковими змінними. Наприклад, при визначенні залежності обсягу податкових надходжень (y) від ставки податку (x) мова йде про визначення одностороннього зв'язку, тобто про регресію. Обидві змінні є випадковими. Кожному значенню x відповідає множина значень y і навпаки, кожному значенню y відповідає множина значень x. Таким чином, ми маємо справу із статистичним розподілом значень x та y. Виходячи з цих розподілів, ми повинні знайти стохастичну залежність між y та x. Одностороння стохастична залежність виражається за допомогою функції, яка на відміну від строгої математичної залежності називається функцією регресії чи просто регресією.

Принциповою різницею між строгою функціональною залежністю та функцією регресії є те, що у першому випадку аргумент (незалежна змінна) повністю визначає значення функції, і для неї існує обернена (наприклад, функція $y = x^2$, тоді $x = \sqrt{y}$). Функція регресії цією властивістю не володіє. Отже, якщо між явищами відсутній функціональний зв'язок, а має місце тільки стохастичний, то функція регресії буде незворотною.

За числом змінних, введених у регресійне рівняння, розрізняють просту (парну) та множинну (багатофакторну) регресії. Відносно форми залежності моделі діляться на лінійну та нелінійну регресії.

При побудові регресійної функції спочатку потрібно провести ідентифікацію змінних, тобто визначити, яка із них є ознакою (залежною чи поянюваною змінною), а які є незалежними (факторами чи пояснюючими змінними). Якщо в загальному випадку рівняння регресії для двох змінних записати $y=f(x)$, то y – пояснювана змінна, а x – пояснююча. Потім потрібно провести специфікацію моделі, тобто встановити форму зв'язку між змінними.

1.2. Діаграма розсіювання регресійної функції

Для того, щоб визначити форму залежності між двома змінними використовують діаграму розсіювання, яка є графічною формою представлення інформації у прямокутній системі координат. На осі абсцис відзначають значення незалежної змінної (x), на осі ординат – значення залежної змінної (y). Результат кожного спостереження (x_i, y_i) деякого економічного процесу відображається точкою на площині. Сукупність цих точок утворює хмарку, яка відображає зв'язок між двома змінними.

За шириною розкиду точок можна зробити висновок про тісноту зв'язку сукупності. Якщо точки розміщені близько одна до одної (у вигляді вузької смужки), то можна стверджувати про наявність відносно тісного зв'язку. Якщо точки на діаграмі розкидані широко, то має місце слабкий зв'язок між змінними, або взагалі не існує зв'язку (рис. 1.2).

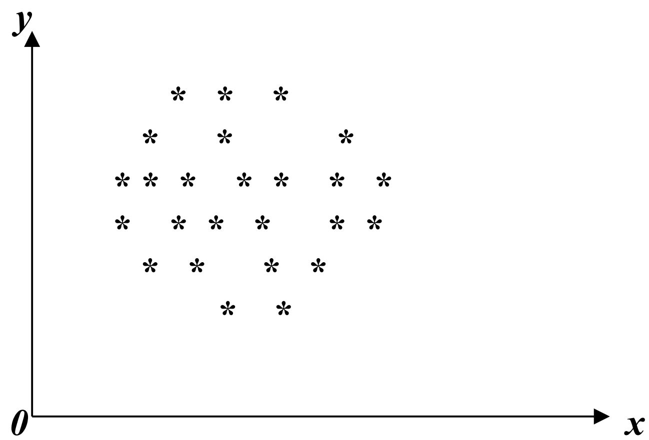


Рис. 1.2. Діаграма розсіювання у випадку відсутності зв'язку

На рис.1.3 представлені основні форми залежностей.

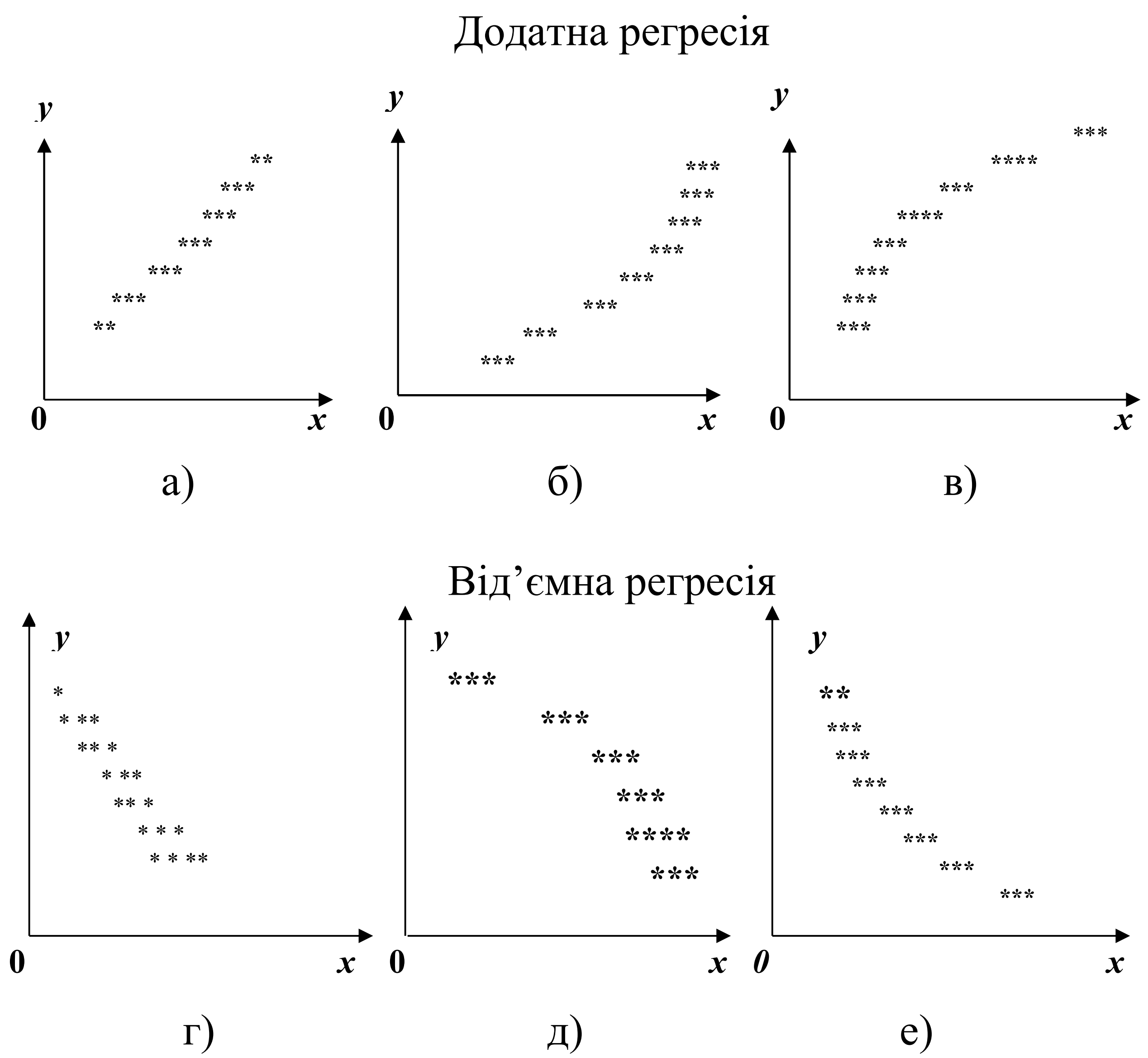


Рис. 1.3. Основні форми регресій

Отже, розрізняють додатну лінійну (а) та нелінійну (б, в) і від'ємну лінійну (г) та нелінійну (д, е) регресії.

За виглядом скупчення точок можна висунути гіпотезу про лінійність чи нелінійність взаємозв'язку між змінними. Так, на діаграмі 1.3 (а, г) маємо явно виражені лінійні тенденції скупчення точок. Спробуємо апроксимувати залежності, зображені на цих діаграмах, лінійною функцією регресії. Звичайно, ці тенденції існують лише в середньому. Вони порушені відхиленням окремих точок. Відхилення від прямої пояснюється впливом інших неврахованих факторів.

1.3. Модель парної лінійної регресії

Припустимо, що за виглядом діаграми розсіювання ми встановили, що між двома змінними існує лінійний зв'язок. Проста (парна) лінійна регресія встановлює лінійну залежність між двома змінними. При цьому одна із змінних (y) вважається залежною змінною (екзогенна, пояснювана, результативна змінна, регресант, відгук) і розглядається як функція від другої (x) незалежної змінної (ендогенна, пояснююча, регресор).

В загального випадку проста лінійна модель запишеться так:

$$y = \alpha + \beta x + u \tag{1.1}$$

Величина $y=\{y_1, y_2, \ldots, y_n\}$ (n – число спостережень) складається з двох складових:

1) невипадкової складової $\alpha+\beta x$, де $x=\{x_1,x_2,\ldots, x_n\}$, α і β – параметри рівняння;

2) випадкової складової u (збурення, помилки) $u=\{u_1,u_2,\ldots,u_n\}$.

Якщо y та x – це кількісні показники деяких економічних явищ чи процесів, то рівняння (1.1) називають економетричною моделлю.

Розглянемо основні причини існування збурення:

1. *Невключення в модель інших пояснюючих змінних*. Встановлення зв’язку тільки між двома факторами y та x є дуже великим спрощенням. Насправді існують інші фактори, що суттєво впливають на результативний показник, які не враховані чи не можуть бути врахованими у формулі (1.1). Вплив цих факторів приводить до того, що істинні точки лежать поза прямою. Об’єднавши всі такі можливі складові впливу на результативний показник, ми якраз отримаємо величину u. Наприклад, при вивченні залежності попиту на товар (y) від ціни на товар (x), збурена змінна u включала би в себе вплив на попит таких чинників: величина сімейного бюджету, якість товару та інші випадкові фактори. Якби ми точно знали, які змінні необхідно включати в модель, і мали можливість точно їх виміряти, то можна на їх основі будувати рівняння і тим самим виключати відповідний елемент збурення.

2. *Неправильна функціональна специфікація*. Функціональне співвідношення між y та x математично може бути визначене неправильно. Наприклад, фактична

залежність не є лінійною, а може бути більш складною. Проте, використання рівняння регресії, яке найкраще описує залежність між змінними є деяким наближенням.

3. *Помилки вимірювання*. Якщо у вимірюваннях однієї чи більше взаємозв'язаних змінних є помилки, то знайдені значення не будуть відповідати точному співвідношенню, а існуюча розбіжність буде вносити свій вклад у структуру збуреної змінної.

4. *Людський фактор*, який неможливо математично описати і нічим, крім випадкової складової відобразити не можна. Наприклад, при встановленні залежності попиту на товар від ціни на нього це можуть бути вподобання покупця.

Отже, збурення є сумарним проявом перелічених вище факторів.

Практично побудувати економетричну модель у вигляді рівняння (1.1) неможливо через випадкову складову, тому лінійну залежність між двома змінними будують у вигляді оціночного рівняння економетричної моделі, відкинувши випадкову складову:

$$\hat{y} = a + bx, \tag{1.2}$$

де a та b відповідно, представляють собою оцінки параметрів α та β рівняння (1.1). Знак "^" над y означає оцінку залежної змінної, отриману з рівняння (1.2) при деяких усереднених умовах.

Розглянемо геометричну інтерпретацію оцінок параметрів економетричної моделі.

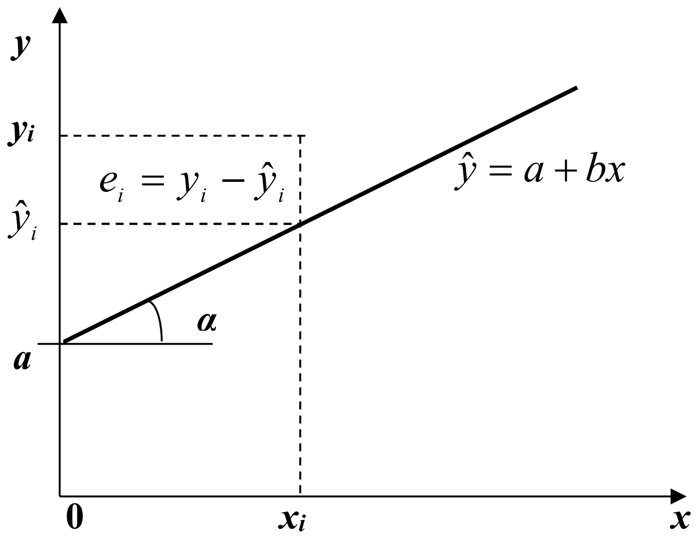


Рис. 1.4. Регресійна пряма та її параметри

Постійна величина *a* визначає точку перетину прямої регресії з віссю ординат (рис. 1.4) і є значенням *y* в точці x_0=0.

Коефіцієнт *b* характеризує нахил прямої до осі абсцис. Позначимо через α кут, який пряма утворює з віссю абсцис. Тоді b=tgα. Він є мірою залежності змінної *y* від *x* або мірою впливу, виявленою зміною *x* на *y*. Відповідно до рівняння (1.2) *b* визначає середню величину зміни результативного показника при зміні пояснювальної змінної *x* на одиницю. Знак *b* визначає напрямок цієї зміни.

Після визначення числових оцінок параметрів можна за рівнянням (1.2) обрахувати значення $\hat{y}_i$ для кожного значення пояснюючої змінної x_i. Це значення називають розрахунковим.

При лінійній функції сукупність розрахункових значень утворює пряму регресії. Як відзначалося раніше, через випадковий вплив сторонніх факторів для кожного значення x_i може спостерігатися декілька емпіричних значень y_i, тобто кожному значенню x відповідає розподіл значень змінної y. Значення функції регресії $\hat{y}_i$ таким чином є оцінками середніх значень змінної y для кожного фіксованого значення змінної x.

Звідси випливає економічна інтерпретація $\hat{y}_i$. Значення $\hat{y}_i$ показують середнє значення залежної змінної y при заданому x_i пояснюючої змінної у припущенні, що єдиною причиною зміни y є змінна x, а випадкова збурена змінна u прийняла значення, рівне нулеві. Розкид фактичних значень змінної y довкола $\hat{y}_i$ зумовлений впливом множини неврахованих факторів. Різницею між емпіричним значенням y_i і розрахунковим $\hat{y}_i$ назвемо залишком, який дає числову оцінку значенням збуреної змінної u (рис. 1.4). Отже, числове значення e визначається як $y_i - \hat{y}_i = e_i, i = \overline{1,n}$. Зрозуміло, що, чим менше значення e_i, тим вдаліше вибрана пряма.

Таким чином, ми підійшли до проблеми оцінювання невідомих параметрів α та β.

1.4. Методи знаходження оцінок параметрів економетричної моделі з двома змінними

а) Метод найменших квадратів (МНК) за системою нормальних рівнянь. Цей метод є найточнішим методом знаходження оцінок параметрів економетричної моделі. Провівши економічний аналіз певного процесу з врахуванням характеру хмарки точок на діаграмі розсіювання, переходимо до вирівнювання дослідних даних, яке полягає у побудові гіпотетичної лінії. Основною вимогою при цьому є зведення до мінімуму помилок специфікації форм зв'язку між змінними. Ці помилки визначаються через відхилення емпіричних даних y_i від значень регресії $\hat{y}_i$, тобто вони формують значення збуреної змінної e:

$$y_i - \hat{y}_i = e_i$$

З графіка (рис. 1.4) бачимо, що e_i – відхилення дослідної точки від оціночної прямої, виміряне по вертикалі. Це відхилення може бути додатнім чи від'ємним в залежності від того, по яку сторону від лінії розміщена конкретна точка.

При виборі прямої можна висунути вимогу, щоб сума відхилень всіх точок від лінії регресії була рівна нулеві, тобто

$$\sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n} e_i = 0.$$

Ця умова означає, що сума додатних відхилень повинна бути рівною сумі від'ємних. Дотримання даної умови не дає можливості однозначно визначити розміщення вирівнюючої прямої на площині. Дану умову задовольняє нескінчена множина прямих, тобто через задану точку з координатами ($x_i;y_i$) проходить пучок прямих.

Для знаходження однозначного розв'язку використовуємо один із показників розсіювання випадкової величини – дисперсію. Якщо всі відхилення піднести до квадрату і просумувати, то результат буде безпосередньо залежати від розкиду точок довкола шуканої лінії. Із множини можливих прямих має бути вибрана така, для якої міра розсіювання дослідних точок буде мінімальною. Це можна представити наступним чином:

$$F(a,b)=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-a-bx_i)^2\rightarrow \min, \quad (1.3)$$

тобто сума квадратів відхилень емпіричних значень змінної y від значень, обчислених за рівнянням прямої, повинна бути мінімальною.

Метод, в основу якого покладена вимога мінімізації суми квадратів відхилень, називається методом найменших квадратів (МНК). З його допомогою знаходять такі оцінки параметрів рівняння регресії, які зводять до мінімуму вибрану міру розсіювання. При цьому проходить вирівнювання емпіричних значень в одну лінію регресії. У випадку лінійного зв'язку між змінними ця лінія є прямою.

Таким чином, проблема визначення прямої регресії зводиться до мінімізації функції (1.3). Необхідною умовою цього є перетворення в нуль перших частинних похідних цієї функції по кожній змінній *a* та *b*. Візьмемо ці частинні похідні та прирівняємо їх до нуля:

$$\begin{cases} \dfrac{\partial F}{\partial a} = 2\sum_{i=1}^{n}(y_i - a - bx_i)(-1) = 0 \\ \dfrac{\partial F}{\partial b} = 2\sum_{i=1}^{n}(y_i - a - bx_i)(-x_i) = 0 \end{cases} \Rightarrow \begin{cases} -2\sum_{i=1}^{n}(y_i - a - bx_i) = 0 \\ -2\sum_{i=1}^{n}(y_i - a - bx_i)x_i = 0 \end{cases} \Rightarrow$$

$$\begin{cases} \sum_{i=1}^{n}(y_i - a - bx_i) = 0 \\ \sum_{i=1}^{n}(y_i - a - bx_i)x_i = 0 \end{cases} \Rightarrow \begin{cases} \sum_{i=1}^{n} y_i - na - b\sum_{i=1}^{n} x_i = 0 \\ \sum_{i=1}^{n} x_i y_1 - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_1^2 = 0 \end{cases}$$

В результаті виконання перетворень отримаємо наступну систему, яка називається системою нормальних рівнянь:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i, \end{cases} \qquad (1.4)$$

розв'язавши яку, знаходимо *a* та *b*:

Розв'язки системи (1.4) методом Крамера можна знайти за формулами:

$$a = \frac{\begin{vmatrix} \sum_{i=1}^{n} y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix}} = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \tag{1.5}$$

$$b = \frac{\begin{vmatrix} n & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i y_i \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix}} = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \tag{1.6}$$

Значення a та b, обчислені за формулами (1.5) і (1.6), є оцінками параметрів α та β регресії, отримані МНК. Маючи значення a та b можна обрахувати значення регресії для заданої області значень пояснюючої змінної x. Ці значення представляють собою найкращу з точки зору МНК лінійну апроксимацію для емпіричних значень y_i, оскільки вибрана міра розсіювання – стандартне відхилення зводиться при цьому до мінімуму.

б) МНК через відхилення від середніх**.** Розглянемо методику оцінювання параметрів з допомогою методу відхилень від середніх арифметичних. Основу даного методу складають властивості оцінок, знайдених МНК, які полягають в тому, що лінія регресії обов'язково проходить через точку середніх значень $(\bar{x}, \bar{y})$.

Поділимо перше рівняння системи (1.4) на n:

$$\frac{na}{n} + b\,\frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

У результаті отримаємо

$$a + b\,\frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} y_i}{n}, \quad \text{або}$$

$$\bar{y} = a + b\bar{x}. \tag{1.7}$$

Далі віднімемо від розрахункового значення змінної y (1.2) її середнє значення, знайдене за формулою (1.7):

$$\hat{y} - \bar{y} = a + bx - (a + b\bar{x}) = b(x - \bar{x}).$$

Звідси
$$\hat{y} = \bar{y} + b(x - \bar{x}). \tag{1.8}$$

Тоді відхилення фактичних значень змінної y від розрахункових, знайдених по формулі (1.8):

$$e_i = y_i - \hat{y}_i = (y_i - \bar{y}) - b(x_i - \bar{x}).$$

$y_i - \bar{y}$ та $x_i - \bar{x}$ – це відхилення змінних y та x від їх середніх значень. Для простоти позначимо їх Δy_i та Δx_i.

Дальше розглядаємо функцію, що представляє собою суму квадратів відхилень дійсних значень змінної y від розрахункових і досліджуємо її на *min*.

$$F(b)=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}\left(\Delta y_i-b\Delta x_i\right)^2 \quad \rightarrow \quad \min$$

Оскільки отримана функція залежить тільки від *b*, то знаходимо частинну похідну цієї функції по *b* і прирівнюємо її до нуля.

$$\frac{\partial F}{\partial b}=-2\sum_{i=1}^{n}\left(\Delta y_i-b\Delta x_i\right)\Delta x_i=-2\sum_{i=1}^{n}\left(\Delta x_i\Delta y_i-b(\Delta x_i)^2\right)=$$

$$=-2\sum_{i=1}^{n}\left(\Delta x_i\Delta y_i\right)+2b\sum_{i=1}^{n}(\Delta x_i)^2 .$$

$$\frac{\partial F}{\partial b}=0 \qquad \Rightarrow \qquad -2\sum_{i=1}^{n}\left(\Delta x_i\Delta y_i\right)+2b\sum_{i=1}^{n}(\Delta x_i)^2=0 \quad \Rightarrow$$

$$\sum_{i=1}^{n}\left(\Delta x_i\Delta y_i\right)-b\sum_{i=1}^{n}(\Delta x_i)^2=0 \quad \Rightarrow \qquad \sum_{i=1}^{n}\left(\Delta x_i\Delta y_i\right)=b\sum_{i=1}^{n}(\Delta x_i)^2 .$$

Звідси знаходимо параметр *b*:

$$b=\frac{\sum_{i=1}^{n}\left(\Delta x_i\Delta y_i\right)}{\sum_{i=1}^{n}\left(\Delta x_i\right)^2}. \tag{1.9}$$

Значення оцінки *a* знайдемо з формули (1.7):

$$a=\bar{y}-b\bar{x} \tag{1.10}$$

Формули (1.9) та (1.10) – це формули для знаходження оцінок параметрів економетричної моделі методом найменших квадратів через відхилення від середніх.

Формулу (1.9) можна представити в іншому вигляді. Якщо розділити чисельник і знаменник на $\frac{1}{n}$, то отримаємо:

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

Ми отримали в чисельнику коефіцієнт коваріації між змінними x та y, який обчислюється за формулою:

$$\operatorname{cov}(x,y)=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x})(y_i-\overline{y}\right)}{n},$$

а в знаменнику маємо дисперсію змінної x, яка знаходиться за формулою:

$$\sigma^2(x)=\operatorname{var}(x)=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}.$$

Тоді формула (1.9) матиме вигляд:

$$b=\frac{\operatorname{cov}(x,y)}{\operatorname{var}(x)}.$$

Завдання. Для десяти підприємств регіону за деякий умовний період відомі числові значення двох економічних показників: валова продукція y (млн. грн.) і вартість основних виробничих фондів x (млн. грн.), (табл.1). Для дослідження характеристики впливу вартості основних виробничих фондів (x) на випуск валової продукції (y) підприємства з допомогою економетричної моделі необхідно:

1. Зробити специфікацію моделі.
2. Знайти оцінки параметрів моделі з допомогою МНК (за системою нормальних рівнянь та через відхилення від середніх).

3. Побудувати оціночну пряму.

Таблиця 1.1

№ підприємства	Валовий випуск продукції, млн.грн., y_i	Основні виробничі фонди, млн.грн., x_i
1	2,2	1,4
2	4,2	2,2
3	5,7	3,3
4	6,8	2,6
5	5,9	3,2
6	7,6	4,5
7	9,5	5,1
8	8,4	6,7
9	10,1	7,3
10	12,3	8,9

♦***Розв'язування.***

1. Побудуємо діаграму розсіювання залежності валового випуску продукції (y) від вартості основних виробничих фондів підприємства (x):

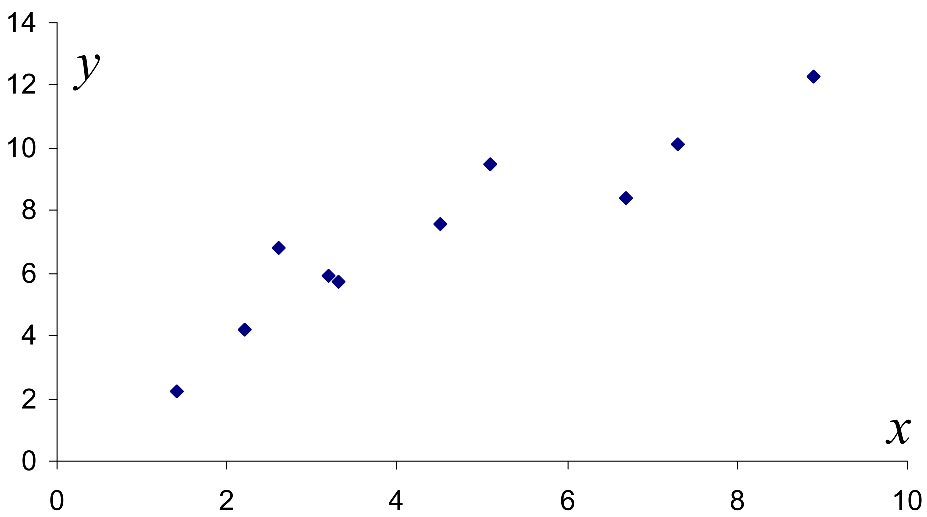


Рис.1.5. Діаграма розсіювання

Розміщення точок на діаграмі розсіювання дає можливість зробити припущення про існування лінійної форми зв'язку у вигляді функції (1.2):

$$\hat{y}=a+bx,$$

де $\hat{y}$ – розрахунковий обсяг випуску валової продукції, млн. грн.; x – вартість основних виробничих фондів, млн. грн.

2. Для спрощення розрахунків при знаходженні оцінок a та b параметрів економетричної моделі побудуємо таблицю:

№ п/п	y_i	x_i	x_i^2	$xi \cdot yi$	Δx_i	$(\Delta x_i)^2$	Δy_i	$\Delta x_i \cdot \Delta y_i$
1	2,2	1,4	1,96	3,08	-3,12	9,73	-5,07	15,82
2	4,2	2,2	4,84	9,24	-2,32	5,38	-3,07	7,12
3	5,7	3,3	10,89	18,81	-1,22	1,49	-1,57	1,92
4	6,8	2,6	6,76	17,68	-1,92	3,69	-0,47	0,9
5	5,9	3,2	10,24	18,88	-1,32	1,74	-1,37	1,81
6	7,6	4,5	20,25	34,2	-0,02	0	0,33	-0,01
7	9,5	5,1	26,01	48,45	0,58	0,34	2,23	1,29
8	8,4	6,7	44,89	56,28	2,18	4,75	1,13	2,46
9	10,1	7,3	53,29	73,73	2,78	7,73	2,83	7,87
10	12,3	8,9	79,21	109,47	4,38	19,18	5,03	22,03
Сума	**72,7**	**45,2**	**258,34**	**389,82**	**0**	**54,04**	**0**	**61,22**

$$\bar{x}=\frac{\sum_{i=1}^{n} x_i}{n}=\frac{45,2}{10}=4,52; \qquad \bar{y}=\frac{\sum_{i=1}^{n} y_i}{n}=\frac{72,7}{10}=7,27.$$

$$\Delta x_1 = x_1 - \overline{x} = 1{,}4 - 4{,}52 = -3{,}12;$$
$$\Delta x_2 = x_2 - \overline{x} = 2{,}2 - 4{,}52 = -2{,}32;$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\Delta x_{10} = x_{10} - \overline{x} = 8{,}9 - 4{,}52 = 4{,}38.$$

Запишемо систему нормальних рівнянь:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases} \qquad \begin{cases} 10a + 45{,}2b = 72{,}7 \\ 45{,}2a + 258{,}34b = 389{,}82 \end{cases}$$

Розв'яжемо її:

$$\begin{cases} a = -4{,}52b + 7{,}27 \\ 45{,}2(-4{,}52b + 7{,}27) + 258{,}34b = 389{,}82 \end{cases} \Rightarrow$$

$$\begin{cases} a = -4{,}52b + 7{,}27 \\ -204{,}3b + 328{,}6 + 258{,}34b = 389{,}82 \end{cases} \Rightarrow$$

$$\begin{cases} a = -4{,}52b + 7{,}27 \\ 54{,}04b = 61{,}22 \end{cases} \Rightarrow \begin{cases} a = -4{,}52b + 7{,}27 \\ b = 1{,}13 \end{cases} \Rightarrow$$

$$\begin{cases} a = -4{,}52 \cdot 1{,}13 + 7{,}27 \\ b = 1{,}13 \end{cases} \Rightarrow \begin{cases} a = 2{,}17 \\ b = 1{,}13 \end{cases}$$

Знайдемо ці ж оцінки за формулами відхилень від середніх (1.9) та (1.10):

$$b = \frac{\sum_{i=1}^{n}(\Delta x_i \Delta y_i)}{\sum_{i=1}^{n}(\Delta x_i)^2} = \frac{61{,}22}{54{,}04} = 1{,}13;$$

$$a = \overline{y} - b\overline{x} = 7{,}27 - 1{,}13 \cdot 4{,}52 = 7{,}27 - 5{,}1 = 2{,}17.$$

Отже, нами отримано оціночне рівняння економетричної моделі

$$\hat{y} = 2{,}17 + 1{,}13x.$$

3. Побудуємо оціночну пряму $\hat{y} = 2{,}17 + 1{,}13x$.

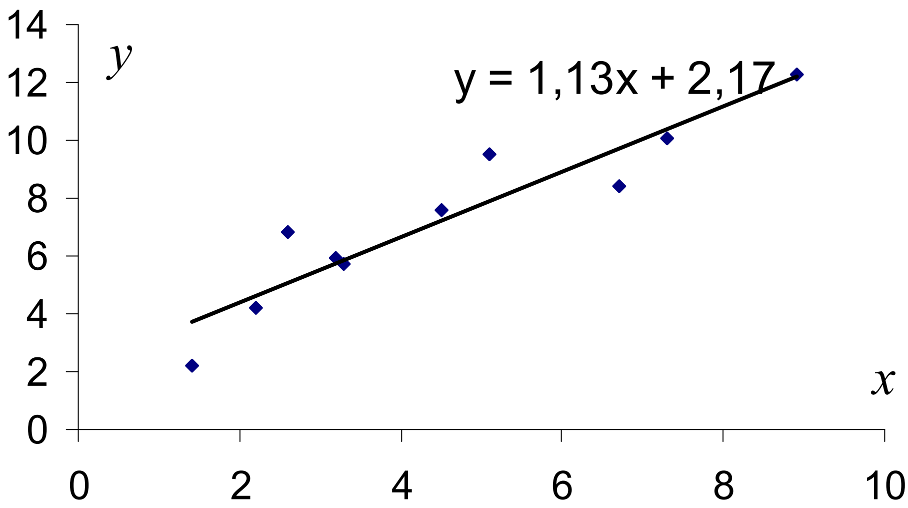


Рис.1.6. Діаграма розсіювання та оціночна пряма

1.5. Загальна, пояснююча та непояснююча дисперсії. Коефіцієнти кореляції та детермінації

Побудова рівняння регресії дає нам можливість розкласти значення y_i в кожному спостереженні на дві складові:

$$y_i = \hat{y}_i + e_i. \tag{1.11}$$

Величина $\hat{y}_i$ – розраховане за оціночним рівнянням економетричної моделі значення результативного показника в i-му випадку. Залишок e_i є розбіжністю між фактичним і прогнозним значенням змінної y, тобто, та частина y, яку ми не можемо пояснити з допомогою рівняння регресії. Знайдемо

$$\mathrm{var}(y) = \mathrm{var}(\hat{y} + e) = \mathrm{var}(\hat{y}) + \mathrm{var}(e) + 2\,\mathrm{cov}(\hat{y}, e). \quad (1.12)$$

Врахувавши, що $\mathrm{cov}(\hat{y}, e) = 0$, будемо мати:

$$\mathrm{var}(y) = \mathrm{var}(\hat{y}) + \mathrm{var}(e). \quad (1.13)$$

Дане співвідношення означає, що ми можемо розкласти загальну дисперсію var(y) на дві складові: $\mathrm{var}(\hat{y})$ – частина, яка пояснює рівняння регресії (пояснююча дисперсія) і var(e) – непояснююча частина (дисперсія помилок або не пояснююча дисперсія).

У лівій частині (1.13) маємо варіацію залежної змінної y навколо свого вибіркового середнього значення $\bar{y}$, а у правій – варіації розрахункових значень $\hat{y}$ навколо середнього значення $\bar{y}$ та фактичних значень y. За означенням дисперсії (1.13) прийме наступний вид:

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

$\sigma^2_{\text{заг}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$ – загальна дисперсія; (1.14)

$$\sigma^2_{\text{поясн.}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$ – пояснююча дисперсія; (1.15)

$$\sigma^2_{\text{непоясн.}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$ – дисперсія помилок, або (1.16)

непояснююча дисперсія.

Тобто, з (1.14) маємо: $\sigma^2_{\text{заг}} = \sigma^2_{\text{поясн.}} + \sigma^2_{\text{непоясн.}}$.

Якщо загальну дисперсію вважати незмінною, то чим більша буде пояснююча дисперсія, тим менша непояснююча, а значить менший розкид точок на діаграмі розсіювання відносно оціночної прямої.

Далі поділимо обидві частини (1.13) на $var(y)$ і отримаємо:

$$1 = \frac{\operatorname{var}(\hat{y})}{\operatorname{var}(y)} + \frac{\operatorname{var}(e)}{\operatorname{var}(y)} \tag{1.17}$$

Як можна побачити з виразу (1.17) перша частина $\frac{\operatorname{var}(\hat{y})}{\operatorname{var}(y)}$ є складовою дисперсії, яку можна пояснити через лінію регресії. Друга частина $\frac{\operatorname{var}(e)}{\operatorname{var}(y)}$ є пропорцією дисперсії помилок у загальній дисперсії, тобто являє собою частину дисперсії, яку не можна пояснити через регресійний зв’язок.

Частина дисперсії, що пояснюється регресією, називається коефіцієнтом детермінації і позначається d або r^2:

$$d = r^2 = \frac{\operatorname{var}(\hat{y})}{\operatorname{var}(y)} = \frac{\sigma^2_{поясн.}}{\sigma^2_{заг.}} = 1 - \frac{\sigma^2_{непоясн.}}{\sigma^2_{заг.}} \tag{1.18}$$

З цієї формули видно, що коефіцієнт детермінації завжди додатній і знаходиться в межах від нуля до одиниці. Максимальне значення d=1, за умови, що лінія регресії точно відповідає всім спостереженням ($\hat{y}_i = y_i$ і всі залишки рівні нулю). Тоді $\operatorname{var}(\hat{y}) = \operatorname{var}(y)$. Якщо ж для вибірки відсутній зв'язок між змінними y та x, то коефіцієнт d буде близький до нуля.

Після побудови регресійної моделі необхідно оцінити тісноту зв'язку між результативною та факторною змінними. Для цього необхідно розрахувати коефіцієнт кореляції, який саме характеризує ступінь щільності лінійної залежності між випадковими величинами x та y. Він позначається r і розраховується за формулою:

$$r = \frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{var}(x)\operatorname{var}(y)}} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

де cov(x,y) – коефіцієнт коваріації між змінними x та y, var(x), var(y) – дисперсії змінних x та y, а σ_x, σ_y – їх середні квадратичні відхилення.

Коефіцієнт кореляції, на відміну від коефіцієнта коваріації, є вже не абсолютною, а відносною мірою зв'язку між двома факторами і приймає значення з інтервалу [-1;1]. Додатне значення кореляції свідчить про наявність прямого зв'язку між змінними, а від'ємне – про зворотній зв'язок. Якщо коефіцієнт кореляції прямує до ±1, то мова йде про наявність тісного лінійного зв'язку між змінними. У той же час, коли він прямує до нуля, то лінійний зв'язок між змінними слабкий. Але, якщо нами отримано r=0, то не треба спішити з висновками про відсутність зв'язку між змінними. Можна лише робити висновок про відсутність лінійного зв'язку, але між вибраними змінними може існувати тісний нелінійний зв'язок. Коефіцієнт кореляції дає можливість робити висновок про тісноту саме лінійного зв'язку між змінними.

Подивимось, чи існує зв'язок між коефіцієнтом детермінації і коефіцієнтом кореляції. Для цього здійснимо наступні перетворення для коефіцієнта детермінації:

$$d=\frac{\sigma^2_{поясн.}}{\sigma^2_{заг.}}=\frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2}=\frac{\frac{1}{n}\sum_{i=1}^{n}(a+bx_i-(a-b\overline{x}))^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2}=$$

$$=\frac{\frac{1}{n}\sum_{i=1}^{n}(b(x_i-\overline{x}))^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2}=\frac{b^2\cdot\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2}=b^2\,\frac{\sigma^2_x}{\sigma^2_y}.$$

Виконаємо аналогічні перетворення для коефіцієнта кореляції, врахувавши, що $b = \dfrac{\operatorname{cov}(x, y)}{\operatorname{var}(x)}$:

$$r = \frac{\operatorname{cov}(x, y)}{\sqrt{\operatorname{var}(x)\operatorname{var}(y)}} = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} \cdot \frac{\sigma_x}{\sigma_x} = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} \cdot \frac{\sigma_x}{\sigma_y} = b \cdot \frac{\sigma_x}{\sigma_y}.$$

З останньої формули видно, що знак коефіцієнта кореляції визначається знаком оцінки b.

Ми бачимо, що коефіцієнт кореляції є коренем квадратним з коефіцієнта детермінації:

$$r = \pm\sqrt{d}\,. \tag{1.19}$$

Приклад 1.2. На основі даних прикладу 1.1 потрібно:

1. Обчислити загальну, пояснюючу та непояснюючу дисперсію.

2. Знайти значення коефіцієнтів детермінації та кореляції.

♦ ***Розв'язування.***

1. Для знаходження дисперсій використаємо наступні формули:

$$\sigma^2_{\text{заг.}} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n};\ \sigma^2_{\textit{поясн.}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n};\ \sigma^2_{\text{непояснн.}} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}.$$

Для спрощення підрахунків побудуємо таблицю, взявши середні значення змінних з прикладу 1.1:

$$\bar{y} = 7{,}27; \qquad \bar{x} = 4{,}52.$$

№ п/п	y_i	x_i	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$\hat{y}_i$	$\hat{y}_i-\bar{y}$	$(\hat{y}_i-\bar{y})^2$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$
1	2,2	1,4	-5,07	25,7	3,74	-3,53	12,49	-1,54	2,36
2	4,2	2,2	-3,07	9,42	4,64	-2,63	6,91	-0,44	0,2
3	5,7	3,3	-1,57	2,46	5,89	-1,38	1,91	-0,19	0,04
4	6,8	2,6	-0,47	0,22	5,09	-2,18	4,73	1,71	2,91
5	5,9	3,2	-1,37	1,88	5,77	-1,50	2,24	0,13	0,02
6	7,6	4,5	0,33	0,11	7,25	-0,02	0	0,35	0,12
7	9,5	5,1	2,23	4,97	7,93	0,66	0,43	1,57	2,47
8	8,4	6,7	1,13	1,28	9,74	2,47	6,1	-1,34	1,79
9	10,1	7,3	2,83	8,01	10,42	3,15	9,92	-0,32	0,1
10	12,3	8,9	5,03	25,3	12,23	4,96	24,62	0,07	0
Сума	**72,7**	**45,2**	**0**	**79,36**	**72,7**	**0**	**69,35**	**0**	**10,01**

Отже, маємо:

$$\sigma^2_{\text{заг}} = \frac{79{,}36}{10} = 7{,}936;\sigma^2_{\text{поясн.}} = \frac{69{,}35}{10} = 6{,}935;\sigma^2_{\text{непояснс.}} = \frac{10{,}01}{10} = 1{,}001.$$

2. Знайдемо значень коефіцієнтів детермінації та кореляції.

Для обчислення коефіцієнта детермінації використовуємо формулу:

$$d = \frac{\sigma^2_{\text{поясн.}}}{\sigma^2_{\text{заг.}}} = \frac{6{,}935}{7{,}936} = 0{,}87,$$

а це означає, що 87 % загальної дисперсії пояснюється оціночною прямою, на долю неврахованих факторів припадає 13 %.

Коефіцієнт кореляції знайдемо за формулою:

$$r = \pm\sqrt{d} = \sqrt{0{,}87} = 0{,}93.$$

Знак коефіцієнта кореляції визначається знаком кутового коефіцієнта оціночного рівняння b (в нашому випадку він додатний). Отримане значення коефіцієнта кореляції вказує на ступінь тісноти лінійного зв'язку між змінними. Значення коефіцієнта кореляції рівне 0,93 (близьке до одиниці), а це значить, що лінійна форма зв'язку між змінними y та x вибрана вірно і цей зв'язок тісний.

1.6. Умови Гаусса-Маркова для випадкової змінної

При використанні МНК для знаходження оцінок параметрів моделі випадкова змінна (збурення, випадковий член) повинна задовольняти чотири умови, що мають назву умови Гаусса-Маркова.

1. Математичне сподівання випадкової змінної для всіх спостережень рівне нулю:

$$M(u_i)=0,\ i=\overline{1,n},$$

де n – кількість спостережень.

Інколи випадковий член моделі може бути додатнім, а інколи – від'ємним. Однак він не має мати систематичного зміщення в жодному можливому напрямку.

2. Дисперсія випадкової змінної повинна бути постійною для всіх спостережень (властивість гомоскедастичності):

$$M(u_i^2)=\sigma_u^2, \quad i=\overline{1,n}.$$

3. Відсутність систематичного зв'язку між значеннями

випадкової змінної в будь-яких спостереженнях:

$$M(u_i, u_j) = 0, \quad i \neq j, \quad i, j = \overline{1, n}.$$

Це означає, що випадкові величини незалежні між собою, тобто значення випадкової змінної u в i-му спостереженні не залежить від того, яке значення вона прийме в j–му спостереженні.

4. Випадкова змінна повинна бути розподіленою незалежно від пояснюючих змінних:

$$M(x_i, u_j) = 0, \quad i, j = \overline{1, n}.$$

Значення будь-якої незалежної змінної в кожному спостереженні вважається екзогенним, повністю визначеним зовнішніми причинами, які не враховані в рівнянні регресії.

Крім приведених вище умов припускається дотримання нормального закону розподілу випадкового члена з нульовим математичним сподіванням і постійною дисперсією. Припущення про нормально розподілені величини ґрунтується на центральній граничній теоремі, яка стверджує, що, якщо випадкова величина є загальним результатом взаємодії значного числа інших випадкових величин, то вона буде мати приблизно нормальний розподіл, навіть якщо окремі складові не матимуть його.

1.7. Властивості оцінок параметрів моделі парної регресії

Розрізняють точкові та інтервальні оцінки параметрів економетричної моделі.

Точковою оцінкою параметра економетричної моделі називають знайдену оцінку цього параметру.

Інтервальною оцінкою параметра економетричної моделі називають інтервал, в межах якого з певною заданою ймовірністю знаходиться істинне значення цього параметру.

Оцінки параметрів економетричної моделі мають такі властивості:

1) Незміщеність.

Вибіркова оцінка b параметра β називається *незміщеною*, якщо вона задовольняє умову $M(b)=\beta$.

2) Обґрунтованість (спроможність).

Вибіркова оцінка b параметра β називається *обґрунтованою (спроможною)*, якщо для як завгодно малого числа ε справджується умова

$$\lim_{n\to\infty} p\left(|b-\beta|\le\varepsilon\right)=1,$$

тобто при збільшенні об'єму вибірки до безмежності оцінка як завгодно близько наближається до істинного параметру.

Ми бачимо, що оцінка є обґрунтованою, якщо вона задовольняє закон великих чисел. Обґрунтованість помилки означає, що чим більші будуть вибірки, тим більша ймовірність того, що помилка оцінки не перевищує як завгодно малого числа ε.

3) Ефективність.

Вибіркова оцінка b параметра β називається *ефективною*, якщо вона має найменшу дисперсію.

Теорема Гауса-Маркова. Якщо для залишкового члена рівняння (1.1) виконуються умови Гауса-Маркова, то оцінки a і b, розраховані за методом найменших квадратів, мають найменшу дисперсію в класі всіх лінійних незміщених оцінок.

Оцінки a і b, знайдені методом найменших квадратів є незміщеними, ефективними та спроможними.

1.8. Побудова довірчих інтервалів

Надійність оцінки визначається ймовірністю, з якою стверджується що побудований за результатами вибірки довірчий інтервал містить невідомий параметр генеральної сукупності. Ймовірність інтервальної оцінки параметра називають довірчою і позначають p. Тоді можна сподіватися, що при множині спостережень параметр генеральної сукупності буде правильно оцінений (тобто довірчий інтервал покриє дійсне значення цього параметра) приблизно у $p \cdot 100\%$ випадків і лише у $(100-p)\,\%$ випадків оцінка буде помилковою.

Ризик помилки визначається рівнем значущості α, який називається довірчим рівнем даного інтервалу.

Позначимо параметр генеральної сукупності через λ, а його оцінку – μ. Тоді, за означенням довірчого інтервалу, будемо мати наступну формулу:

$$P(\mu - k\sigma_{\mu} \le \lambda \le \mu + k\sigma_{\mu}) = 1 - \alpha ,$$

де k – довірчий множник, який відображає частку стандартного відхилення, яка повинна бути врахована, щоб із заданою ймовірністю p довірчий інтервал $\lfloor \mu - k\sigma_{\mu} ; \mu + k\sigma_{\mu} \rfloor$ покрив параметр λ генеральної сукупності.

Перейдемо до побудови довірчих інтервалів для параметрів парної лінійної регресії. Знайдемо довірчий інтервал для оціночного рівняння. Для цього нам необхідно мати похибку оцінки, яку знайдемо за формулою: $\Delta_{yx} = t_p \cdot \sigma_{\text{непояс.}}$, де t_p – ймовірнісний коефіцієнт, значення якого при заданому рівні ймовірності p знаходимо за таблицями нормального розподілу. Значення t_p є коренем рівняння $2\Phi(t_p) = p$, де $\Phi(t_p)$ – інтегральна функція Лапласа.

Тоді довірчий інтервал для оціночного рівняння матиме вигляд:

$$\hat{y}_i - \Delta_{yx} \le y_i \le \hat{y}_i + \Delta_{yx} . \qquad (1.20)$$

Графічно довірчий інтервал можна зобразити таким чином:

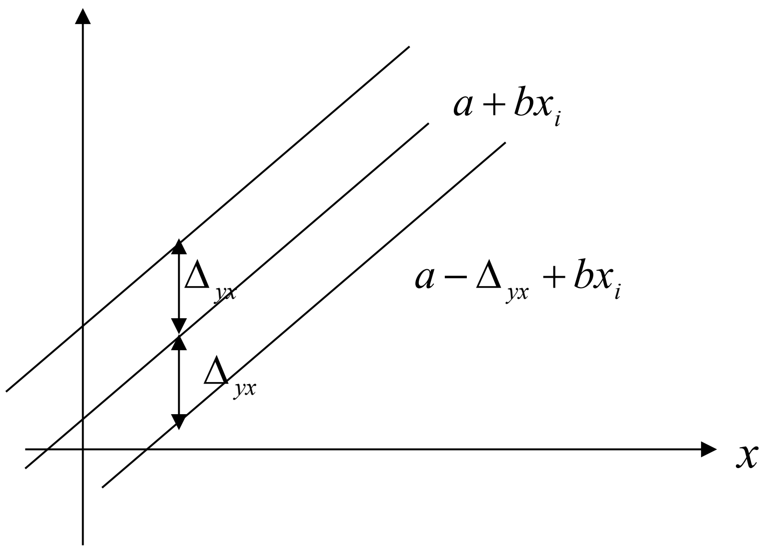


Рис. 1.7. Довірчий інтервал оціночного рівняння

Процедура побудови довірчих інтервалів для параметрів α та β аналогічна попередній процедурі. Спочатку знаходимо граничні похибки оцінок відповідних параметрів за формулами:

$$\Delta_a = \sigma_a \cdot t_p; \quad \Delta_b = \Delta_b \cdot t_p \qquad (1.21)$$

де σ_a^2, σ_b^2 – відповідно дисперсії оцінок *a* та *b*, значення яких обчислюються за формулами:

$$\sigma_a^2 = \frac{\sigma^2_{\text{непоясн.}} \cdot \sum_{i=1}^{n} x_i^2}{n \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2}, \qquad \sigma_b^2 = \frac{\sigma^2_{\text{непоясн.}}}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \qquad (1.22)$$

Отже, довірчі інтервал для параметрів α та β матимуть вигляд:

$$a-\Delta_a \le \alpha \le a+\Delta_a, \tag{1.23}$$

$$b-\Delta_b \le \beta \le b+\Delta_b. \tag{1.24}$$

Приклад 1.3. На основі даних прикладів 1.1 та 1.2 побудувати довірчі інтервали оціночного рівняння і параметрів α та β економетричної моделі для ймовірності p=0,9.

♦ ***Розв'язування.***

Обчислимо граничну похибку $\Delta_{yx} = t_p \cdot \sigma_{\text{непояcн.}}$. Значення t_p знайдемо, розв'язавши рівняння: $2\Phi(t_p) = p$, де $\Phi(t_p)$ - інтегральна функція Лапласа. Для p=0,9 маємо: t_p=1,65. Також знайдемо $\sigma_{\text{непояcн.}} = \sqrt{\sigma^2_{\text{непояcн.}}} = \sqrt{1{,}001} = 1{,}0005$

Тоді $\Delta_{yx} = 1{,}65 \cdot 1{,}0005 = 1{,}651$.

Довірчий інтервал матиме вигляд:

$$\hat{y}_i - 1{,}651 \le y_i \le \hat{y}_i + 1{,}651,$$

$$2{,}17 + 1{,}13x_i - 1{,}651 \le y_i \le 2{,}17 + 1{,}13x_i + 1{,}651,$$

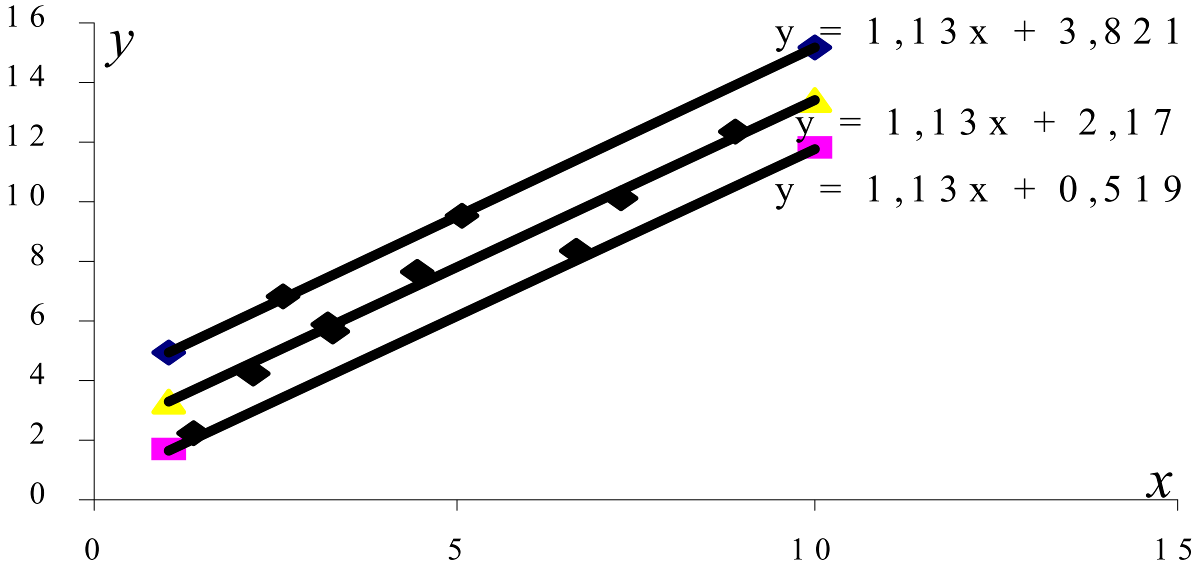
$$1{,}13x_i + 0{,}519 \le y_i \le 1{,}13x_i + 3{,}821.$$

Для побудови оціночного рівняння, верхньої та нижньої межі довірчого інтервалу достатньо по дві точки. Обчислимо їх:

Нижня межа $y=1{,}13x+0{,}519$	
x	y
1	1,65
10	11,82

Оціночна пряма $y=1{,}13x+2{,}17$	
x	y
1	3,30
10	13,47

Верхня межа $y=1{,}13x+3{,}821$	
x	y
1	4,95
10	15,12



Знайдемо довірчі інтервали для α та β. Почнемо із обчислення граничних похибок оцінок цих параметрів:

$$\Delta_a=\sigma_a t_p\ ,\quad \Delta_b=\sigma_b t_p\,.$$

Для цього спочатку знайдемо значення дисперсій і середніх квадратичних відхилень оцінок a та b за формулами:

$$\sigma_a^2=\frac{\sum_{i=1}^{n}x_i^2\sigma^2_{\text{непоясн.}}}{n\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=\frac{258{,}34\cdot 1{,}001}{10\cdot 54{,}04}=0{,}48,\qquad \sigma_a=\sqrt{0{,}48}=0{,}69.$$

$$\sigma_b^2=\frac{\sigma^2_{\text{непоясн.}}}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=\frac{1{,}001}{54{,}04}=0{,}02,\qquad \sigma_b=\sqrt{0{,}02}=0{,}14.$$

Далі знайдемо

$$\Delta_a=\sigma_a\cdot t_p=0{,}69\cdot 1{,}65\approx 1{,}14;$$

$$\Delta_b=\sigma_b\cdot t_p=0{,}14\cdot 1{,}65=0{,}23.$$

Отже, нами отримано наступні інтервали довіри для оцінки α:

$$a \in [a - \Delta_a; \ \ a + \Delta_a],$$
$$a \in [2{,}17 - 1{,}14; \ \ 2{,}17 + 1{,}14],$$
$$1{,}03 \le \alpha \le 3{,}31,$$

для оцінки β:

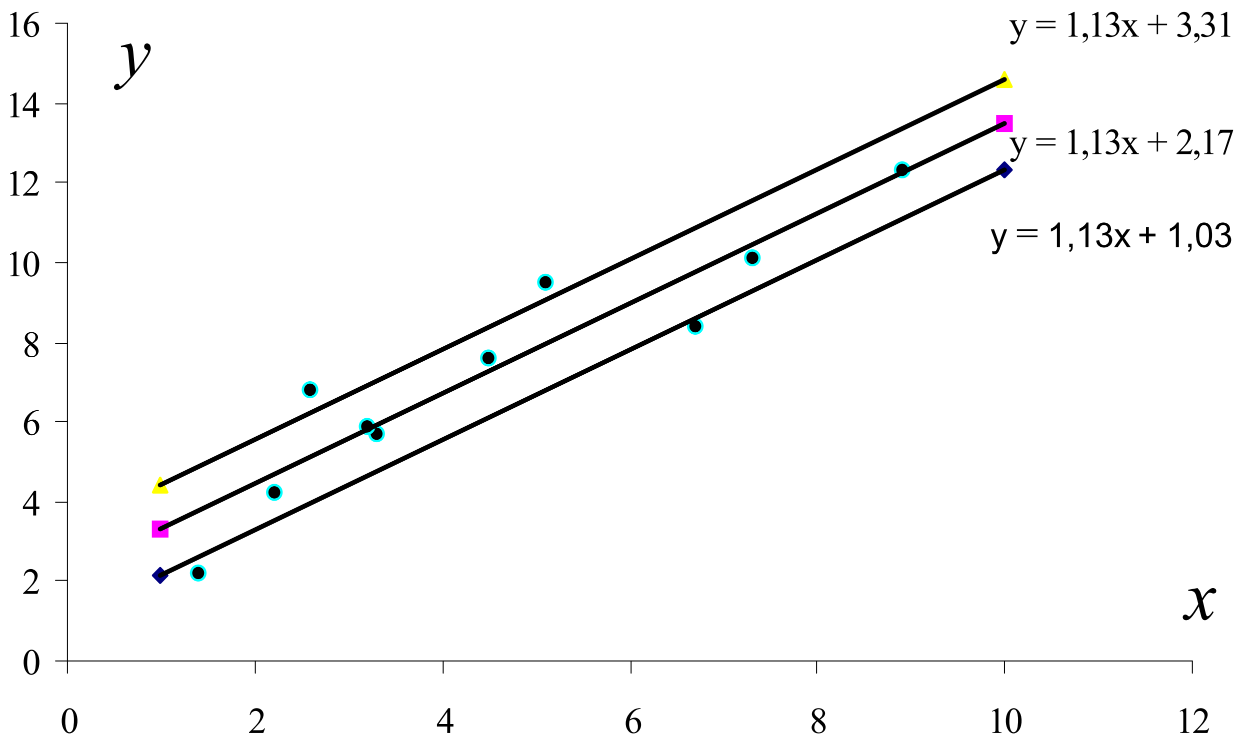
$$\beta \in [b - \Delta_b; \ \ b + \Delta_b],$$
$$\beta \in [1{,}13 - 0{,}23; \ \ 1{,}13 + 0{,}23]$$
$$0{,}9 \le \beta \le 1{,}36.$$

Відобразимо ці довірчі інтервали графічно. Обчислимо координати двох точок для побудови нижньої та верхньої меж довірчого інтервалу для α.

Нижня межа $y = 1{,}13x + 1{,}03$	
x	y
1	2,16
10	12,33

Оціночна пряма $y = 1{,}13x + 2{,}17$	
x	y
1	3,30
10	13,47

Верхня межа $y = 1{,}13x + 3{,}31$	
x	y
1	4,44
10	14,61

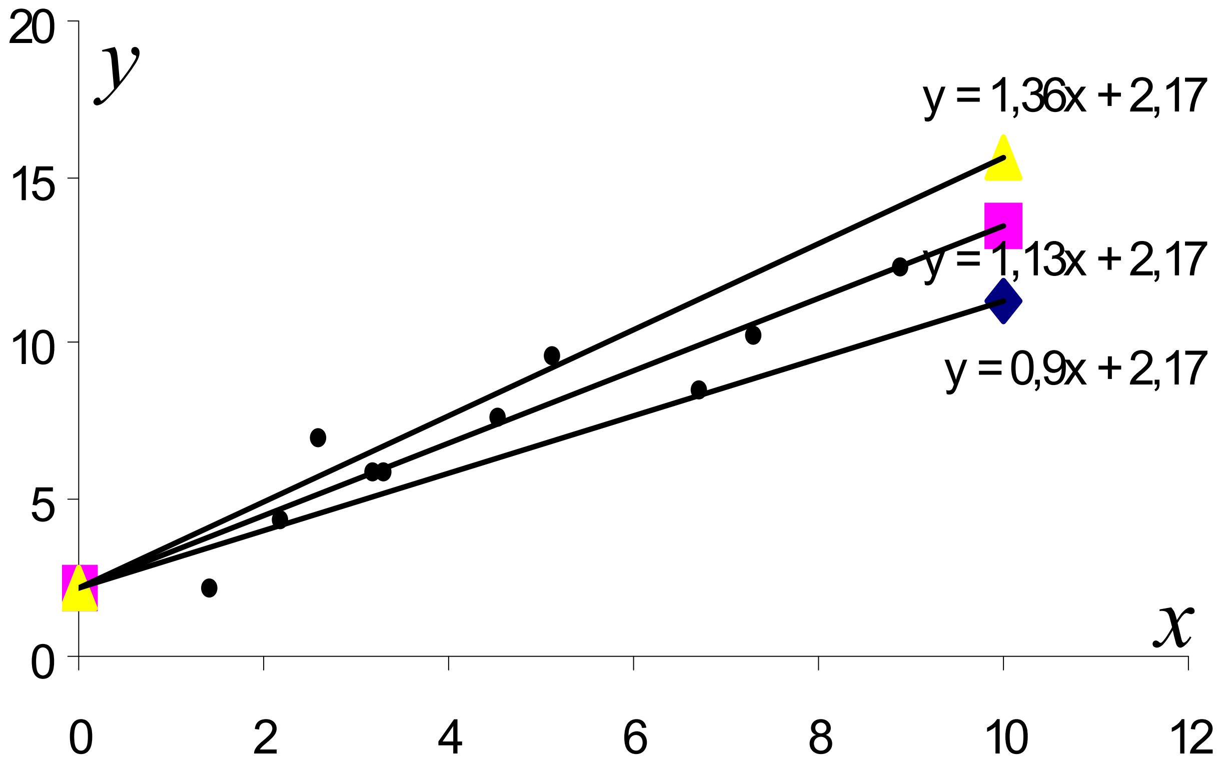


Для β:

Нижня межа $y = 0{,}9x + 2{,}17$	
x	y
0	2,17
10	11,17

Оціночна пряма $y = 1{,}13x + 2{,}17$	
x	y
0	2,17
10	13,47

Верхня межа $y = 1{,}36x + 2{,}17$	
x	y
0	2,17
10	15,77



1.9. Перевірка нульових гіпотез

Оскільки статистичні дані, які ми досліджуємо, створені різними випадковими факторами, то більшість статистичних досліджень супроводжується перевіркою деяких припущень або гіпотез про джерела цих даних.

Основне припущення, яке перевіряється, називається нульовою гіпотезою і позначається H_o переважно формулюється як відсутність різниць, відсутність впливу фактора, рівність нулю значень вибіркових характеристик і т.д.

Друге припущення, яке перевіряється (не завжди строго протилежне чи обернене першому), називається конкуруючою або альтернативною гіпотезою і позначається H_1.

Для статистичного висновку про наявність або відсутність кореляційного зв'язку між досліджуваними змінними необхідно провести перевірку рівня значущості вибіркового коефіцієнта кореляції. Використаний критерій для розв'язку задач даного типу ґрунтується на розподілі різних статистик і називається критерієм значущості.

Процедура перевірки значущості починається з формулювання нульової гіпотези H_o. У загальному випадку вона полягає в тому, що між параметром вибірки і параметром генеральної сукупності немає ніяких суттєвих різниць. Альтернативна гіпотеза H_1 полягає в тому, що між цими параметрами є суттєві різниці. Наприклад, при перевірці наявності кореляції в генеральній сукупності нульова гіпотеза полягає в тому, що істинний коефіцієнт кореляції рівний нулю (Ho: R=0). Якщо в результаті перевірки виявиться, що нульова гіпотеза неприйнятна, то нульова гіпотеза відкидається і приймається альтернативна H_1. Іншими словами, припущення відносно некорельованості випадкових змінних у генеральній сукупності треба визнати необґрунтованим. І навпаки, якщо на основі критерію значущості нульова гіпотеза приймається, тобто r міститься в допустимій зоні випадкового розсіяння, тоді немає підстави вважати сумнівним припущення відносно некорельованості змінних у генеральній сукупності.

При перевірці значущості встановлюють значення її рівня α, який дає певну впевненість в тому, що помилкові висновки будуть зроблені в дуже рідких випадках. Рівень значущості виражає ймовірність того, що нульова гіпотеза H_o відкидається тоді, коли вона в дійсності вірна. Зрозуміло, що має сенс вибрати дану ймовірність як можна меншою.

Перевіряючи значущість коефіцієнта парної кореляції, встановлюють наявність або відсутність кореляційного зв'язку між досліджуваними явищами. При відсутності зв'язку коефіцієнт кореляції генеральної сукупності рівний нулю (R=0). Процедура перевірки гіпотези починається з формулювання нульової та альтернативної гіпотез:

H_o: різниці між вибірковим коефіцієнтом кореляції r та R=0 немає;

H_1: різниця між r та R=0 значна, і як наслідок, між змінними y та x в генеральній сукупності є суттєвий лінійний зв'язок.

Для оцінки значущості коефіцієнта кореляції використовуємо t-критерій Стьюдента з $n-m-1$ ступенями вільності.

Під кількістю ступенів вільності розуміють різницю між кількістю спостережень та кількістю параметрів, які встановлені у результаті цих спостережень, незалежно один від одного.

За результатами вибірки обчислюємо t-статистику, або емпіричне значення параметру t:

$$t_{емп} = \frac{r\sqrt{n-m-1}}{\sqrt{1-r^2}}, \qquad (1.25)$$

яке порівнюється з критичним значенням $t_{кр}$, що знаходиться за таблицями розподілу Стьюдента при заданому рівні значущості α та $k=n-m-1$ ступенях вільності.

Правило використання критерію полягає у наступному:

- якщо $\left|t_{емп}\right| > t_{kp}$, то нульова гіпотеза H_o на рівні значущості α відкидається і приймається альтернативна гіпотеза H_1 про існування лінійної залежності між даними змінними в генеральній сукупності;
- якщо $\left|t_{емп}\right| \le \mathrm{t}_{\text{кр}}$, то нульова гіпотеза H_o на рівні

значущості α приймається.

Перевірку нульових гіпотез стосовно параметрів α та β економетричної моделі проводять аналогічно. Спочатку висуваємо нульові гіпотези:

$$H_0: \alpha=0, \quad H_0: \beta=0.$$

Альтернативними будуть гіпотези:

$$H_1: \alpha\neq 0, \quad H_1: \beta\neq 0.$$

Потім обчислюємо емпіричні значення параметра t за формулами:

$$t_\beta = \frac{|\beta|}{\sigma_\beta}; \qquad t_\alpha = \frac{|\alpha|}{\sigma_\alpha}.$$

Емпіричне значення параметру порівнюють з критичним, знайденим за таблицями Стьюдента для заданого рівня значущості α та $k=n-m-1$ ступенів вільності. Якщо $t_{емп} > t_{kp}$, то нульова гіпотеза H_o із рівнем значущості α відкидається і приймається альтернативна гіпотеза H_1. Тоді відповідна оцінка вважається статистично значимою. Якщо ж $t_{емп} \leq \mathrm{t}_{\text{кр}}$, то нульова гіпотеза H_o для рівня значущості α приймається, а відповідна оцінка не є статистично значимою.

<u>Приклад 1.4.</u> На основі даних прикладів 1.1-1.3 виконати перевірки нульових гіпотез стосовно коефіцієнта кореляції і параметрів α та β економетричної моделі.

♦***Розв'язування.***

Висуваємо нульову гіпотезу H_o: $R_{ген}=0$ (робимо припущення, що коефіцієнт кореляції генеральної сукупності рівний нулю). Альтернативною гіпотезою буде H_1: $R_{ген}\neq 0$.

Далі для заданої вибірки обчислимо емпіричне значення параметру t:

$$t_{\text{емп}} = \frac{r\sqrt{n-m-1}}{\sqrt{1-r^2}} = \frac{0,93\sqrt{10-1-1}}{\sqrt{1-0,87}} = 7,31$$

Для заданої ймовірності p=0,9 (α=1-p=1-0,9=0,1) і k=10-1-1=8 ступенів вільності знаходимо табличне значення $t_{кр.}$=2,306.

Оскільки $|t_{\text{емп}}| > t_{kp}$, то з надійністю p=0,9 гіпотезу H_o необхідно відкинути і прийняти альтернативну гіпотезу H_1 про існування залежності між змінними. Отже, у 90 % вибірок із генеральної сукупності коефіцієнт кореляції не дорівнює нулю.

Далі виконаємо перевірку нульової гіпотези відносно β (H_o: $\beta=0$) проти альтернативної H_1: $\beta\neq 0$. Для цього знаходимо емпіричне значення за формулою:

$$t_{\text{емп}} = \frac{|b|}{\sigma_b} = \frac{1,13}{0,14} = 8,07$$

Оскільки емпіричне значення t більше критичного ($t_{емп}>t_{кр}$), то нульова гіпотеза відхиляється і робиться висновок, що кутовий коефіцієнт b розрахований за даною вибіркою є статистично значущим з ймовірністю p=0,9.

Перевіримо нульову гіпотезу H_o: α=0. Обчислимо

$$t_{емп} = \frac{|a|}{\sigma_a} = \frac{2,17}{0,69} = 3,14.$$

$t_{емп}>t_{кр}$, значить нульова гіпотеза стосовно параметру α теж відхиляється, а значить α не може бути рівним нулю в генеральній сукупності.

1.10. Адекватність економетричної моделі

Відповідність побудованої економетричної моделі можна перевірити з допомогою коефіцієнта детермінації. Якщо його значення близьке до одиниці, то можна вважати, що отримана економетрична модель адекватна. В цьому випадку зміна значення результативної змінної y лінійно залежить саме від зміни пояснюючої змінної x, а не через вплив випадкових факторів. Якщо ж значення коефіцієнта детермінації близьке до нуля, то модель вважають неадекватною, тобто лінійний зв’язок між y та x відсутній. Якщо ж значення коефіцієнта детермінації нечітке, тобто близьке 0,5, то для перевірки адекватності економетричної моделі використовують критерій Фішера F. Він дозволяє оцінити, чи значно нахил b відрізняється від нуля, тобто перевірити гіпотезу H_o: β=0, оскільки, якщо значення b близьке до нуля, то з формули (1.8) маємо:

$$\hat{y} = \bar{y} + b(x - \overline{x}) \approx \bar{y}$$

Іншими словами, як краще апроксимувати дані за середнім значенням, чи регресійною прямою.

Альтернативна гіпотеза полягає в тому, що значення параметру β не дорівнює нулю, і має вигляд: H_1: $\beta \neq 0$.

Обчислюємо емпіричне значення параметру F за формулою:

$$F_{емп} = \frac{r^2 (n - m - 1)}{m(1 - r^2)}\ , \tag{1.26}$$

де m – число незалежних змінних (для простої регресії m=1).

Після цього знаходимо з таблиці величину $F_{кр}$ – критичне значення F-розподілу Фішера з (k_1=m=1, k_2=n-m-1) ступенями вільності і рівнем значущості α. Наприклад, якщо α=0,05, то ми можемо помилятися в 5 % випадків, а в 95 % випадків наші висновки будуть правильними.

Якщо розраховане нами значення $F_{емп}$>$F_{кр}$, то ми відкидаємо гіпотезу про те, що β=0 з ризиком помилитися не більше ніж у 5 % випадків. У такому випадку побудована нами економетрична модель адекватна реальній дійсності.

В протилежному випадку, тобто якщо розраховане

$F_{емп} \leq F_{кр}$, то гіпотезу про те, що $\beta=0$ приймаємо і в цьому випадку побудована нами економетрична модель неадекватна реальній дійсності.

Якщо ми отримали, що оціночне рівняння економетричної моделі відповідає реальній дійсності, то на його основі можна здійснювати прогноз, тобто передбачати шляхи розвитку досліджуваних явищ і процесів на найближче майбутнє. Прогноз може бути точковим або інтервальним.

Точковий прогноз на наступний $n+1$ період отримаємо, коли в оціночне рівняння економетричної моделі підставимо значення пояснюючої змінної x_{n+1}:

$$\hat{y}_{n+1} = a + bx_{n+1}.$$

Інтервальний прогноз – це інтервал, в який з заданою ймовірністю $p=1-\alpha$ попаде дійсне значення результативної змінної y. В загальному випадку він має вигляд:

$$\left(\hat{y}_{n+1} - t_\alpha \cdot \sigma_{yx} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}};\ \hat{y}_{n+1} + t_\alpha \cdot \sigma_{yx} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \right),$$

або

$$\left(a + bx_{n+1} - t_\alpha \cdot \sigma_{yx} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}};\ a + bx_{n+1} + t_\alpha \cdot \sigma_{yx} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \right)$$

Нижню межу цього інтервалу називають песимістичним прогнозом, а верхню – оптимістичним.

<u>Приклад 1.5.</u> На основі даних попередніх прикладів перевірити на адекватність побудовану економетричну модель.

♦***Розв’язування.***

Для оцінки рівня адекватності побудованої економетричної моделі експериментальним даним використовуємо критерій Фішера F. Обчислимо:

$$F_{емп} = \frac{r^2(n-m-1)}{m(1-r^2)} = \frac{0{,}87\cdot(10-1-1)}{1\cdot(1-0{,}87)} = 53{,}54$$

Знайдемо табличне значення даного критерію ($F_{кр.}$) для рівня надійності p=0,95 та числа ступенів вільності $k_1=m=1$, $k_2=n\text{-}m\text{-}1=10\text{-}1\text{-}1=8$:

$$F_{кр} = 5{,}32.$$

Оскільки $F_{емп.}>F_{кр.}$, то отримане нами оціночне рівняння економетричної моделі

$$\hat{y} = 2{,}17 + 1{,}13\,x$$

адекватне реальній дійсності і на його основі можна здійснювати прогнози.

2. Класична лінійна багатофакторна економетрична модель

2.1. Загальні відомості про лінійну багатофакторну економетричну модель

Відомо, що більшість економічних показників формується під впливом багатьох різноманітних факторів. Тому постає задача їх виявлення та оцінювання ступеня впливу різних факторів на результативний показник.

Припустимо, що деяка змінна y залежить від множини незалежних змінних $x_1, x_2, \dots x_m$. Тоді, у випадку лінійної форми зв’язку між ними, економетрична модель матиме вигляд:

$$y = b_0 + b_1 x_1 + \ldots + b_m x_m + u, \tag{2.1}$$

де y – залежна змінна; $x_1, x_2, \dots x_m$ – незалежні змінні; $b_0, b_1, \dots b_m$ – параметри моделі, для яких потрібно буде знайти оцінки; u – збурення або залишок (випадкова складова).

Тоді оціночне рівняння для даної моделі буде:

$$\hat{y} = a_0 + a_1 x_1 + \ldots + a_m x_m, \tag{2.2}$$

де $a_0, a_1, \dots a_m$ – відповідно оцінки невідомих параметрів $b_0, b_1, \dots b_m$.

Нехай нам задано сукупність n спостережень за залежною змінною $y=\{y_1, y_2, \dots y_n\}$ та m незалежних змінних $x_j=\{x_{j1}, x_{j2}, \dots x_{jn}\}$, $j=1, 2, \dots m$.

Для того, щоб знайти оцінки параметрів багатофакторної економетричної моделі, як і для моделі парної регресії, висуваються деякі припущення щодо випадкової складової:

1) математичне сподівання випадкової складової для всіх спостережень рівне нулю:

$$M(u_i) = 0, \quad i = \overline{1,n}.$$

2) Дисперсія випадкової складової однакова для всіх спостережень (властивість гомоскедастичності):

$$M(u_i^2) = \sigma^2, \quad i = \overline{1,n}.$$

3) Значення випадкової складової *u* в *i*–му спостереженні не залежить від того, яке значення вона прийняла в *j*–му спостереженні, тобто випадкові величини незалежні між собою:

$$M(u_i u_j) = 0, \quad i \neq j; \quad i, j = \overline{1,n}, \qquad \text{або}$$

$$\operatorname{cov}(u_i, u_j) = 0, \quad i \neq j; \quad i, j = \overline{1,n}.$$

4) Випадкова змінна *u* розподілена незалежно від змінної *x*:

$$M(x_i u_i) = 0, \qquad \text{або} \qquad \operatorname{cov}(x_i, u_i) = 0, \quad i, = \overline{1,n}.$$

5) Випадкова змінна *u* розподілена за нормальним законом з математичним сподіванням, рівним нулю, і постійною дисперсією.

6) Змінні x_i та x_j (фактори моделі) незалежні між собою, тобто між ними немає явища мультиколінеарності.

2.2. Метод найменших квадратів знаходження оцінок параметрів багатофакторної лінійної економетричної моделі

Як і випадку моделі парної регресії значення оцінок параметрів багатофакторної знайдемо таким чином, щоб сума квадратів відхилень фактичних даних від розрахункових була найменшою:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1i} - \ldots - a_m x_{mi})^2 = F(a_0, a_1, \ldots, a_m) \to \min, \quad (2.3)$$

де e_i являється оцінкою для u_i та різницею між фактичним значенням змінної y в i-му спостереженні та значенням $\hat{y}_i$, розрахованим за оціночним рівнянням (2.2).

Таким чином, наша задача зводиться до мінімізації функції (2.3). Необхідною умовою цього є рівність нулю перших частинних похідних цієї функції по кожній невідомій $a_0, a_1, \ldots a_m$. Візьмемо частинні похідні функції (2.3) по $a_0, a_1, \ldots a_m$ і прирівняємо їх до нуля. Отримаємо наступну систему рівнянь:

$$\begin{cases} \dfrac{\partial F}{\partial a_0} = -2\sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_{1i} - \ldots - a_m x_{mi} \right) = 0 \\ \dfrac{\partial F}{\partial a_1} = -2\sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_{1i} - \ldots - a_m x_{mi} \right) x_{1i} = 0 \\ \vdots \\ \dfrac{\partial F}{\partial a_m} = -2\sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_{1i} - \ldots - a_m x_{mi} \right) x_{mi} = 0 \end{cases} \quad (2.4)$$

Після виконання відповідних перетворень (2.4) прийме наступний вид:

$$\begin{cases} na_0 + a_1\sum\limits_{i=1}^{n} x_{1i} + a_2\sum\limits_{i=1}^{n} x_{2i} + \ldots + a_m\sum\limits_{i=1}^{n} x_{mi} = \sum\limits_{i=1}^{n} y_i \\ a_0\sum\limits_{i=1}^{n} x_{1i} + a_1\sum\limits_{i=1}^{n} x_{1i}^2 + a_2\sum\limits_{i=1}^{n} x_{2i}x_{1i} + \ldots + a_m\sum\limits_{i=1}^{n} x_{mi}x_{1i} = \sum\limits_{i=1}^{n} y_i x_{1i} \\ \vdots \\ a_0\sum\limits_{i=1}^{n} x_{mi} + a_1\sum\limits_{i=1}^{n} x_{1i}x_{mi} + a_2\sum\limits_{i=1}^{n} x_{2i}x_{mi} + \ldots + a_m\sum\limits_{i=1}^{n} x_{mi}^2 = \sum\limits_{i=1}^{n} y_i x_{mi} \end{cases} \quad (2.5)$$

Система (2.5) називається системою нормальних рівнянь для знаходження оцінок a_0, a_1, ... a_m.

У випадку, коли результативний показник залежить від двох факторів, тобто $m=2$, система нормальних рівнянь матиме вигляд:

$$\begin{cases} na_0 + a_1\sum\limits_{i=1}^{n} x_{1i} + a_2\sum\limits_{i=1}^{n} x_{2i} = \sum\limits_{i=1}^{n} y_i \\ a_0\sum\limits_{i=1}^{n} x_{1i} + a_1\sum\limits_{i=1}^{n} x_{2i}^2 + a_2\sum\limits_{i=1}^{n} x_{1i}x_{2i} = \sum\limits_{i=1}^{n} y_i x_{1i} \\ a_0\sum\limits_{i=1}^{n} x_{2i} + a_1\sum\limits_{i=1}^{n} x_{1i}x_{2i} + a_2\sum\limits_{i=1}^{n} x_{2i}^2 = \sum\limits_{i=1}^{n} y_i x_{2i} \end{cases} \quad (2.6)$$

Розв’язавши (2.6), отримаємо значення a_0, a_1 та a_2 і можемо записати рівняння оціночної площини лінійної залежності результативного показника y від двох факторів x_1 та x_2:

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 .$$

2.3. Лінійна багатофакторна економетрична модель в матричній формі

Знаходження значень оцінок $a_0, a_1, \ldots a_m$ з допомогою системи нормальних рівнянь (2.5) є досить громіздким, тому доцільно це зробити з допомогою апарату матричної алгебри.

Для кожного окремого спостереження економетрична модель (2.1) матиме вигляд:

$$y_i = b_0 + b_1 x_{1i} + \ldots + b_m x_{mi} + u_i, \quad i = \overline{1, n}, \tag{2.7}$$

або у вигляді системи:

$$\begin{cases} y_1 = b_0 + b_1 x_{11} + \ldots + b_m x_{m1} + u_1 \\ y_2 = b_0 + b_1 x_{12} + \ldots + b_m x_{m2} + u_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_0 + b_1 x_{1n} + \ldots + b_m x_{mn} + u_n. \end{cases} \tag{2.8}$$

Якщо ввести позначення:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \; X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{m1} \\ 1 & x_{12} & \cdots & x_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & \cdots & x_{mn} \end{bmatrix}; \; B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}; \; U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

то система (2.8) в матричній формі матиме вигляд:

$$Y = XB + U. \tag{2.9}$$

де Y – вектор значень залежної змінної; X – матриця значень незалежних змінних, перший стовпчик якої відповідає фіктивній змінній x_0={1, 1, …, 1}, яку вводимо для параметру b_0; B – вектор параметрів моделі; U – вектор залишків моделі.

Позначимо:

$$\hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}; \quad A = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix},$$

де $\hat{Y}$ – вектор розрахункових значень; A – вектор оцінок параметрів B моделі;

Тоді оціночне рівняння економетричної моделі (2.9) прийме вигляд:

$$\hat{Y} = XA. \tag{2.10}$$

Для знаходження оцінок параметрів A використаємо МНК.

Розглянемо відхилення фактичних значень від розрахункових, знайдених за формулою (2.10). Позначимо ці відхилення $e = Y - \hat{Y} = Y - XA$. В матричному вигляді це буде вектор

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$$

Далі запишемо суму квадратів залишків наступним чином:

$$e(A)=\sum_{i=1}^{n}e_i^2=e'e=(Y-\hat{Y})'(Y-\hat{Y})=(Y-XA)'(Y-XA)=$$
$$=Y'Y-2A'X'Y+A'X'XA\to\min, \tag{2.11}$$

де символ штрих (′) означає операцію транспонування.

Знайдемо частинну похідну даного виразу за компонентами вектора A і прирівняємо до нуля:

$$\frac{\partial e(A)}{\partial A}=-2X'Y+2X'XA=0.$$

Звідси отримаємо систему рівнянь, яка є системою нормальних рівнянь для знаходження оцінок a_0, a_1, ... a_m в матричній формі:

$$X'XA=X'Y. \tag{2.12}$$

Якщо визначник матриці $X'X$ не дорівнює нулю, то існує обернена їй матриця $(X'X)^{-1}$, і розв’язком системи (2.12) буде вектор-стовпець шуканих оцінок:

$$A=\left(X'X\right)^{-1}X'Y. \tag{2.13}$$

Якщо знайдемо значення оцінок A параметрів моделі, то можна знайти міру їх розсіювання, використавши для цього математичний апарат матричної алгебри. Знайдені оцінки є випадковими величинами, що мають певний розподіл імовірностей. Вони розсіюються навколо дійсного значення параметра B. Апарат матричної алгебри допомагає знайти не тільки дисперсії оцінок параметрів A, але й розрахувати значення коваріацій між двома параметрами a_k та a_j $(k \neq j, k = \overline{0,m}, j = \overline{0,m})$. Ці величини служать характеристиками випадкових величин a_j і утворюють дисперсійно-коваріаційну матрицю оцінок параметрів моделі.

За означенням коваріаційна матриця для оператора A буде:

$$\operatorname{cov}(A) = M\left[\left(A-B\right)\cdot\left(A-B\right)'\right] = \sigma_u^2\left(X'\cdot X\right)^{-1} = \begin{bmatrix} \sigma_{a_0}^2 & \dots & \sigma_{a_0 a_m} \\ \vdots & \vdots & \vdots \\ \sigma_{a_m a_0} & \dots & \sigma_{a_m}^2 \end{bmatrix} \quad (2.14)$$

Дана матриця є симетричною відносно головної діагоналі, елементами якої є дисперсії оцінок параметрів моделі a_j, $\left(j = \overline{0,m}\right)$, а поза діагоналлю – їх коваріація.

Оцінена дисперсія випадкової величини σ_u^2 знаходиться за формулою:

$$\sigma_u^2 = \frac{e'e}{n-m-1} = \frac{Y'\cdot Y - A'X'Y}{n-m-1}. \quad (2.15)$$

<u>Приклад 2.1.</u> Знайти оцінки параметрів економетричної моделі, яка описує залежність обсягу отриманого прибутку підприємствами регіону від розміру основних виробничих фондів та затрат праці (лінійна форма зв’язку). Вхідні дані приведено в таблиці 2.1.

Таблиця 2.1

Номер підприємства	Прибуток, млн.грн., y	Основні фонди, млн.грн., x_1	Затрати праці, млн.год., x_2
1	1,2	2,2	2,9
2	1,4	2,9	3,2
3	1,8	3,1	2,5
4	2,1	3,7	2,6
5	2,8	2,8	3,3
6	3,2	4,3	3,2
7	3,5	4,6	2,5
8	4,4	5,4	4,9
9	4,9	4,9	2,1
10	5,3	6,4	4,7

♦***Розв'язування.***

Оціночне рівняння економетричної моделі у випадку лінійної форми зв'язку між вибраними економічними показниками має вигляд:

$$y = a_0 + a_1 x_1 + a_2 x_2,$$

де y – прибуток, млн.грн; x_1 – вартість основних виробничих фондів, млн. грн; x_2 – затрати праці, млн. год. Для знаходження оцінок параметрів моделі використаємо математичний апарат матричної алгебри.

Введемо позначення:

$$A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix};$$

$$Y=\begin{pmatrix}1{,}2\\1{,}4\\1{,}8\\2{,}1\\2{,}8\\3{,}2\\3{,}5\\4{,}4\\4{,}9\\5{,}3\end{pmatrix};\qquad X=\begin{bmatrix}1&2{,}2&2{,}9\\1&2{,}9&3{,}2\\1&3{,}1&2{,}5\\1&3{,}7&2{,}6\\1&2{,}8&3{,}3\\1&4{,}3&3{,}2\\1&4{,}6&2{,}5\\1&5{,}4&4{,}9\\1&4{,}9&2{,}1\\1&6{,}4&4{,}7\end{bmatrix};$$

$$X'=\begin{bmatrix}1&1&1&1&1&1&1&1&1&1\\2{,}2&2{,}9&3{,}1&3{,}7&2{,}8&4{,}3&4{,}6&5{,}4&4{,}9&6{,}4\\2{,}9&3{,}2&2{,}5&2{,}6&3{,}3&3{,}2&2{,}5&4{,}9&2{,}1&4{,}7\end{bmatrix}.$$

Вектор оцінок A знайдемо користуючись формулою $A=\left(X'X\right)^{-1}X'Y$, де X' – матриця, транспонована до матриці X.

Знаходимо добуток двох матриць, використовуючи стандартну офісну програму Excel. Для цього заносимо елементи вектора *Y,* матриць X та X'. Транспоновану матрицю X' можна отримати, використавши майстер функцій f_x, вибравши в категоріях “Ссылки и массивы”, а в функціях “ТРАНСП”. Щоб перемножити матриці, відмічаємо поле, де буде знаходитись результат, входимо в майстер функцій (f_x), в категоріях вибираємо “Математические”, а в функціях “МУМНОЖ”. Дальше в область “Масив 1” вводимо адрес першої матриці X', а в “Масив 2” – область другої – *X*. Дальше натискаємо кнопку ”Ok”. Щоб отримати значення всіх елементів матриці добутку, одночасно натискаємо клавіші F2 та Ctrl+Shift+Enter. В результаті отримаємо:

$$\left(X' \cdot X\right) = \begin{bmatrix} 10 & 40,3 & 31,9 \\ 40,3 & 178,17 & 134,36 \\ 31,9 & 134,36 & 109,55 \end{bmatrix}.$$

Знаходимо обернену матрицю до матриці $\left(X'X\right)$. Знову відмічаємо поле, де буде знаходитись результат, входимо в f_x, вибираємо категорію “ Математические”, а в функціях “МОБР”, потім ”Ok”. Після одночасного натискання клавіш F2 та Ctrl+Shift+Enter отримаємо:

$$\left(X'X\right)^{-1} = \begin{pmatrix} 1,65 & -0,14 & -0,3 \\ -0,14 & 0,09 & -0,07 \\ -0,3 & -0,07 & 0,18 \end{pmatrix}.$$

Потім знаходимо добуток матриці X' та вектора *Y*:

$$X' \cdot Y = \begin{pmatrix} 30,6 \\ 139,44 \\ 102,91 \end{pmatrix}.$$

Значення елементів вектора *A* знайдемо, перемноживши матрицю $\left(X'X\right)^{-1}$ на вектор $X' \cdot Y$:

$$A=\left(X'X\right)^{-1}\left(X'Y\right)=\begin{pmatrix}-0,87\\1,06\\-0,11\end{pmatrix}.$$

Отже, ми отримали оціночне рівняння економетричної моделі:

$$\hat{y}=-0,87+1,06x_1-0,11x_2.$$

2.4. Коефіцієнти парної кореляції. Кореляційна матриця

При побудові економетричної моделі залежності результативного показника від одного фактора щільність кореляційного зв'язку ми визначали з допомогою коефіцієнта кореляції. В багатофакторних економетричних моделях є коефіцієнти парної, частинної та множинної кореляції.

Коефіцієнти парної кореляції використовуються для вимірювання сили лінійних зв'язків різних пар змінних (ознак) із заданої множини. Значення парних коефіцієнтів кореляції між результативною змінною y та незалежними змінними x_j обчислюються за формулою:

$$r_{yx_j}=\frac{\sum_{i=1}^{n}(y_i-\overline{y})(x_{ji}-\overline{x_j})}{\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}\cdot\sqrt{\sum_{i=1}^{n}(x_{ji}-\overline{x_j})^2}},\quad j=\overline{1,m}, \tag{2.16}$$

а між незалежними змінними x_l та x_j:

$$r_{x_l x_j} = \frac{\sum_{i=1}^{n}(x_{li} - \overline{x_l})(x_{ji} - \overline{x_j})}{\sqrt{\sum_{i=1}^{n}(x_{li} - \overline{x_l})^2} \cdot \sqrt{\sum_{i=1}^{n}(x_{ji} - \overline{x_j})^2}}, \qquad l, j = \overline{1, m}. \quad (2.17)$$

Парні коефіцієнти кореляції утворюють кореляційну матрицю (матрицю коефіцієнтів парної кореляції):

$$[R] = \begin{bmatrix} r_{yy} & r_{yx_1} & r_{yx_2} & \dots & r_{yx_m} \\ r_{x_1 y} & r_{x_1 x_1} & r_{x_1 x_2} & \dots & r_{x_1 x_m} \\ r_{x_2 y} & r_{x_2 x_1} & r_{x_2 x_2} & \dots & r_{x_2 x_m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{x_m y} & r_{x_m x_1} & r_{x_m x_2} & \dots & r_{x_m x_m} \end{bmatrix}.$$

Дана матриця є симетричною відносно головної діагоналі ($r_{x_l x_j} = r_{x_j x_l}$), елементи якої рівні одиниці, тобто:

$$[R] = \begin{bmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_m} \\ r_{x_1 y} & 1 & r_{x_1 x_2} & \dots & r_{x_1 x_m} \\ r_{x_2 y} & r_{x_2 x_1} & 1 & \dots & r_{x_2 x_m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{x_m y} & r_{x_m x_1} & r_{x_m x_2} & \dots & 1 \end{bmatrix}. \quad (2.18)$$

Припустимо, що змінна y залежить від двох факторів x_1 та x_2. Тоді ми будемо мати три коефіцієнти парної кореляції:

- коефіцієнт парної кореляції між y та x_1, або коефіцієнт кореляції економетричної моделі $y = \widetilde{a} + \widetilde{b} x_1 + u$:

$$r_{yx_1} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_{1i} - \overline{x_1})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2} \cdot \sqrt{\sum_{i=1}^{n}(x_{1i} - \overline{x_1})^2}}; \quad (2.19)$$

- коефіцієнт парної кореляції між *y* та *x₂*, або коефіцієнт кореляції економетричної моделі $y = a' + b'x_2 + u$:

$$r_{yx_2} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_{2i} - \overline{x_2})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2} \cdot \sqrt{\sum_{i=1}^{n}(x_{2i} - \overline{x_2})^2}}; \qquad (2.20)$$

- коефіцієнт парної кореляції між незалежними змінними x_1 та x_2, або коефіцієнт кореляції економетричної моделі $x_1 = a'' + b''x_2 + u$:

$$r_{x_1x_2} = \frac{\sum_{i=1}^{n}(x_{1i} - \overline{x_1})(x_{2i} - \overline{x_2})}{\sqrt{\sum_{i=1}^{n}(x_{1i} - \overline{x_1})^2} \cdot \sqrt{\sum_{i=1}^{n}(x_{2i} - \overline{x_2})^2}}. \qquad (2.21)$$

Кореляційна матриця матиме вигляд:

$$[R] = \begin{bmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{x_1y} & 1 & r_{x_1x_2} \\ r_{x_2y} & r_{x_2x_1} & 1 \end{bmatrix}.$$

<u>Приклад 2.2.</u> Використавши умову прикладу 2.1 знайти коефіцієнти парної кореляції.

♦***Розв'язування.***

Коефіцієнти парної кореляції знайдемо, використавши формули (2.19) – (2.21):

Для спрощення обчислень побудуємо розрахункову таблицю 2.2:

Таблиця 2.2

№	y	x_1	x_2	$y-\overline{y}$	$x_{1i}-\overline{x}_1$	$x_{2i}-\overline{x}_2$	$(y-\overline{y})^2$	$(x_{1i}-\overline{x}_1)^2$	$(x_{2i}-\overline{x}_2)^2$	$(y-\overline{y})(x_{1i}-\overline{x}_1)$	$(y-\overline{y})(x_{2i}-\overline{x}_2)$	$(x_{1i}-\overline{x}_1)(x_{2i}-\overline{x}_2)$
1	1,2	2,2	2,9	-1,86	-1,83	-0,29	3,46	3,35	0,08	3,40	0,54	0,53
2	1,4	2,9	3,2	-1,66	-1,13	0,01	2,76	1,28	0,00	1,88	-0,02	-0,01
3	1,8	3,1	2,5	-1,26	-0,93	-0,69	1,59	0,86	0,48	1,17	0,87	0,64
4	2,1	3,7	2,6	-0,96	-0,33	-0,59	0,92	0,11	0,35	0,32	0,57	0,19
5	2,8	2,8	3,3	-0,26	-1,23	0,11	0,07	1,51	0,01	0,32	-0,03	-0,14
6	3,2	4,3	3,2	0,14	0,27	0,01	0,02	0,07	0,00	0,04	0,00	0,00
7	3,5	4,6	2,5	0,44	0,57	-0,69	0,19	0,32	0,48	0,25	-0,30	-0,39
8	4,4	5,4	4,9	1,34	1,37	1,71	1,80	1,88	2,92	1,84	2,29	2,34
9	4,9	4,9	2,1	1,84	0,87	-1,09	3,39	0,76	1,19	1,60	-2,01	-0,95
10	5,3	6,4	4,7	2,24	2,37	1,51	5,02	5,62	2,28	5,31	3,38	3,58
$\sum$	**30,6**	**40,3**	**31,9**	**0**	**0**	**0**	**19,2**	**15,8**	**7,79**	**16,12**	**5,3**	**5,8**

Підставимо отримані значення сум в формули:

$$r_{yx_1} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_{1i} - \bar{x}_1)}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} \cdot \sqrt{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}} = \frac{16{,}12}{\sqrt{19{,}2} \cdot \sqrt{15{,}76}} = 0{,}93,$$

а це означає, що зв'язок між змінними y та x_1 тісний і змінна x_1 пояснює 93 % загальної дисперсії змінної y.

$$r_{yx_2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_{2i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} \cdot \sqrt{\sum_{i=1}^{n}(x_{2i} - \bar{x}_2)^2}} = \frac{5{,}3}{\sqrt{19{,}2} \cdot \sqrt{7{,}79}} = 0{,}43,$$

а це означає, що зв'язок між змінними y та x_2 середній і змінна x_2 пояснює 43 % загальної дисперсії змінної y.

$$r_{x_1x_2} = \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2} \cdot \sqrt{\sum_{i=1}^{n}(x_{2i} - \bar{x}_2)^2}} = \frac{5{,}8}{\sqrt{15{,}76} \cdot \sqrt{7{,}79}} = 0{,}52,$$

а значить зв'язок між незалежними змінними x_1 та x_2 середній і змінна x_2 пояснює 52 % дисперсії змінної x_1.

2.5. Коефіцієнти частинної кореляції

Коефіцієнти частинної кореляції так само представляють лінійні зв'язки ознак, але при цьому береться до уваги чистий зв'язок пари ознак за умови, що зв'язки всіх інших ознак з ознаками із даної пари не діють. Отже, частинною кореляцією між ознаками x_i та x_j називається кореляційна залежність між цими ознаками при фіксованих значеннях інших ознак.

Нехай число незалежних змінних $m=2$. Для знаходження коефіцієнтів частинної кореляції використовуємо наступні формули:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1 x_2}^2)}}, \quad r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1 x_2}^2)}} \qquad (2.22)$$

де $r_{yx_1 \cdot x_2}$ – коефіцієнт частинної кореляції між змінними y та x_1 при виключенні впливу x_2. Він показує яку долю загальної дисперсії змінної y, що не пояснила змінна x_2 пояснить введення в економетричну модель змінної x_1.

$r_{yx_2 \cdot x_1}$ – коефіцієнт частинної кореляції між змінними y та x_2 при виключенні впливу x_1. Він показує яку долю загальної дисперсії змінної y, що не пояснила змінна x_1 пояснить введення в економетричну модель нової змінної x_2.

$r_{yx_1}, r_{yx_2}, r_{x_1 x_2}$ – відповідно коефіцієнти парної кореляції.

Для $m=3$ маємо такі коефіцієнти частинної кореляції:

$$r_{yx_1 . x_2 x_3} = \frac{r_{yx_1 . x_3} - r_{yx_2 . x_3} r_{x_1 x_2 \cdot x_3}}{\sqrt{(1 - r_{yx_2 . x_3}^2)(1 - r_{x_1 x_2 . x_3}^2)}}; \quad r_{yx_2 . x_1 x_3} = \frac{r_{yx_2 . x_3} - r_{yx_1 . x_3} r_{x_2 x_1 \cdot x_3}}{\sqrt{(1 - r_{yx_1 . x_3}^2)(1 - r_{x_2 x_1 . x_3}^2)}};$$

$$r_{yx_3 . x_1 x_2} = \frac{r_{yx_3 . x_2} - r_{yx_1 . x_2} r_{x_3 x_1 \cdot x_2}}{\sqrt{(1 - r_{yx_1 . x_2}^2)(1 - r_{x_3 x_1 . x_2}^2)}}.$$

Узагальнюючи приведені вище формули для будь-якого числа пояснюючих змінних, отримаємо:

$$r_{yx_1.x_2...x_m} = \frac{r_{yx_1.x_3...x_m} - r_{yx_2.x_3...x_m} r_{x_1x_2 \cdot x_3...x_m}}{\sqrt{(1 - r^2_{yx_2.x_3...x_m})(1 - r^2_{x_1x_2.x_3...x_m})}}.$$

Останнє співвідношення показує, що обчислення коефіцієнта частинної кореляції порядку *m* зводиться до визначення таких же коефіцієнтів, але порядку (*m-1*). Тому спочатку необхідно знайти коефіцієнти парної кореляції, а потім перейти до обчислення коефіцієнтів вищих порядків.

<u>Приклад 2.4.</u> Розрахувати коефіцієнти частинної кореляції на основі даних прикладу 2.1.

♦***Розв'язування.***

Коефіцієнти частинної кореляції обчислюємо за формулами (2.22):

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_2})(1 - r^2_{x_1x_2})}} = \frac{0,93 - 0,43 \cdot 0,52}{\sqrt{(1 - 0,43^2)(1 - 0,52^2)}} = 0,92,$$

а це говорить про те, що введення змінної x_1 в економетричну модель пояснить 92 % з тих 57 % загальної дисперсії змінної *y*, що їх не пояснила змінна x_2. Значить змінну x_1 доцільно вводити в економетричну модель.

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_1})(1 - r^2_{x_1x_2})}} = \frac{0,43 - 0,93 \cdot 0,52}{\sqrt{(1 - 0,93^2)(1 - 0,52^2)}} = 0,17.$$

Цей коефіцієнт показує, що з 7 % загальної дисперсії змінної y, що не пояснила змінна x_1 17 % пояснить введення в економетричну модель нової змінної x_2, що свідчить про недоцільність введення в економетричну модель змінної x_2.

2.6. Коефіцієнти множинної кореляції та детермінації

Основним показником тісноти кореляційного зв'язку між результативним показником y та всіма незалежними змінними $x_j\left(j=\overline{1,m}\right)$, а також ступеня близькості вибраного виду математичної залежності до вибіркових даних є коефіцієнти множинної кореляції та детермінації.

Коефіцієнт множинної детермінації обчислюється за формулою:

$$R^2=\frac{\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2},$$

Враховуючи рівність:

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

маємо:

$$\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2=\sum_{i=1}^{n}(y_i-\overline{y})^2-\sum_{i=1}^{n}(y_i-\hat{y}_i)^2.$$

Тоді вираз для R^2 прийме такий вигляд:

$$R^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2-\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}.$$

З останньої формули випливає, що коли $\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=0$, то $R^2=1$. Отже, якщо всі вибіркові значення показника розміщені на лінії регресії, то коефіцієнт множинної детермінації дорівнює одиниці. А значить, чим ближче вибіркові значення наближаються до лінії регресії, тим ближче R^2 наближається до одиниці, а отже, тим більше варіація залежної змінної визначається варіацією незалежних факторів. Як бачимо, коефіцієнт множинної детермінації показує частку варіації результативної ознаки, яка знаходиться під впливом досліджуваних факторів, тобто визначає, яка частка варіації ознаки y врахована в моделі та викликана впливом вибраних факторів. Його числове значення змінюється від нуля до одиниці, тобто $R^2\in[0;1]$. Якщо R^2 прямує до нуля, то у вибірці відсутній лінійний зв'язок між залежною та незалежними змінними.

Характерною особливістю коефіцієнта детермінації R^2 є те, що він – неспадна функція від кількості факторів, які входять до моделі. Отже, якщо кількість незалежних факторів зростає, то значення R^2 так само зростає.

З останньої рівності маємо:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

причому значення знаменника у формулі не залежить від кількості факторів $\{x_j, j = \overline{1,m}\}$, тоді як чисельник, навпаки, знаходиться у зворотній залежності. Тобто, при зростанні числа незалежних факторів, величина $\sum_{i=1}^{n} e_i^2$ спадає, або на крайній випадок не зростає.

Тому при співставленні між собою двох регресійних моделей для однакових залежних змінних, але з різною кількістю незалежних факторів, перевагу треба віддати тій моделі, для якої значення R^2 є більшим.

Одним з основних показників тісноти кореляційного зв'язку результативного показника y з факторами x_j $(j = \overline{1,m})$, а також мірою ступеня відповідності даних $\hat{y}_i$ $(i = \overline{1,n})$ є коефіцієнт множинної кореляції. Він визначається як коефіцієнт кореляції між y та $\hat{y}$ має вигляд:

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}}.$$

Квадрат коефіцієнта множинної кореляції, як і випадку простої регресії є коефіцієнтом детермінації, тобто має місце $R = \sqrt{R^2}$.

На практиці для обчислення значень коефіцієнтів множинної детермінації та кореляції, коли розраховані значення коефіцієнтів парної кореляції, використовують наступні формули:

- коефіцієнт множинної детермінації:

$$R^2 = \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1x_2}}{1 - r_{x_1x_2}^2}. \qquad (2.23)$$

- коефіцієнт множинної кореляції: $R = \sqrt{R^2}$. (2.24)

Приклад 2.5. Використавши дані прикладу 2.1 знайти коефіцієнт множинної детермінації та кореляції.

♦ ***Розв'язування***.

Для знаходження коефіцієнта множинної детермінації використаємо формулу (2.23), використавши значення коефіцієнтів парної кореляції з прикладу 2.3:

$$R^2 = \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0{,}93^2 + 0{,}43^2 - 2 \cdot 0{,}93 \cdot 0{,}43 \cdot 0{,}52}{1 - 0{,}52^2} = 0{,}85$$

а значить 85 % загальної дисперсії пояснюється залежністю обсягу валової продукції від вартості основних виробничих фондів та затрат праці. І тільки 15 % загальної дисперсії не може бути пояснено отриманою нами залежністю. Таким чином, можна зробити висновок, що побудована модель статистично значима.

Тоді коефіцієнт множинної кореляції буде:

$$R = \sqrt{R^2} = \sqrt{0,85} = 0,92,$$

а це означає, що між валовою продукцією та основними виробничими фондами і затратами праці існує тісний лінійний зв'язок.

2.7. Основні характеристики виробничої функції

Оціночне рівняння (2.2) економетричної моделі інколи поєднують із поняттям виробничої функції. Формально загальне визначення виробничої функції можна дати як функції, яка кількісно описує існуючі взаємозв'язки та взаємозалежності між результативними показниками функціонування деякого економічного процесу або явища і чинниками під впливом яких вони формуються. До основних показників можна віднести дохід, прибуток, обсяг податкових надходжень, валова продукція, рентабельність, продуктивність праці, собівартість та ін.

Виробничі функції найперше почали використовувати для дослідження причинно-наслідкових відносин виробничих процесів у сфері мікроекономіки. Потім вони стали дуже популярним та ефективним, у певній мірі, інструментом, економетричного аналізу економічних процесів і явищ, як на мезо- так і на макрорівнях.

Вид виробничої функції визначається видом алгебраїчного рівняння, з допомогою якого описується її математична модель.

Розглянемо основні характеристики виробничої функції, заданої у вигляді:

$$\hat{y} = f(x_1, x_2, ..., x_m).$$

Першою характеристикою для довільного *j*-го показника (ресурсу) є його середня продуктивність при фіксованих обсягах інших ресурсів, яка визначається за формулою:

$$C_j = \frac{\bar{f}(x_1, x_2, ..., x_m)}{\bar{x}_j},$$

де C_j – середня продуктивність *j*-го виду ресурсу; $\bar{f}(x_1, x_2, ..., x_m)$ – середнє значення результативного показника; $\bar{x}_j$ – середнє значення *j*-го чинника.

Другою характеристикою є гранична продуктивність (віддача, ефективність) *j*-го виду ресурсу, яка характеризує приріст результату виробництва при одиничному прирості *j*-го ресурсу і визначається таким чином:

$$\Gamma_j = \frac{\partial y}{\partial x_j} = f'_{x_j}(x_1, x_2, ..., x_m).$$

Наприклад, якщо результативний показник відображає обсяг прибутку (млн. грн.) у залежності від вартості основних виробничих фондів (млн. грн.) та витрати праці (млн. люд.-год.), то гранична продуктивність першого фактора показує величину зміни результативного при одиничній зміні факторного (на 1млн. грн.).

Представляє інтерес з'ясування характеру зміни граничної продуктивності із зміною обсягу j-го при незмінному обсязі інших ресурсів. З цією метою знайдемо:

$$\frac{\partial^2 y}{\partial x_j^2} = f''_{x_j}(x_1, x_2, ..., x_m).$$

Якщо $f''_{x_j}(x_1, x_2, ..., x_m) > 0$, то гранична віддача j-го ресурсу зростає, а в протилежному випадку гранична продуктивність – спадає.

Третя характеристика – відносна зміна результату виробництва на одиницю відносної зміни витрат j-го ресурсу. Цей показник називається еластичністю випуску по витратах j-го ресурсу:

$$E_j = \frac{\partial y}{\partial x_j} \cdot \frac{\bar{x}_j}{\bar{\hat{y}}} = \frac{\bar{x}_j f'_{x_j}(x_1, x_2, ..., x_m)}{\bar{f}(x_1, x_2, ..., x_m)}.$$

Коефіцієнт еластичності j-го виду ресурсу показує на скільки процентів зміниться результативний показник (прибуток) при зміні факторного (вартість основних виробничих фондів) на один процент.

Додатні коефіцієнти при збільшенні факторного показника означають збільшення результативного показника, а від'ємні – зменшення.

Четверта характеристика – норма заміщення ресурсу. Для довільної пари ресурсів k та j можна визначити граничну норму h_{kj} заміщення j-го ресурсу k-им. Ця величина є співвідношення, взяте із знаком "-", граничних продуктивностей j-го та k-го ресурсів:

$$h_{kj} = -\frac{\partial y / \partial x_j}{\partial y / \partial x_k}.$$

У випадку лінійної форми зв'язку коефіцієнт граничної продуктивності співпадає з відповідною оцінкою параметру економетричної моделі, тобто має місце:

$$\Gamma_j = \frac{\partial y}{\partial x_j} = a_j,\ j = \overline{1, m}, \tag{2.25}$$

а формула для обчислення коефіцієнта еластичності по кожному фактору:

$$E_j = \frac{\partial y}{\partial x_j} \cdot \frac{\bar{x}_j}{\bar{y}} = a_j \cdot \frac{\bar{x}_j}{\bar{y}}, \quad j = \overline{1, m}. \tag{2.26}$$

Приклад 2.6. Використавши умову прикладу 2.1 знайти коефіцієнти граничної продуктивності та еластичності.

♦ ***Розв'язування***.

В прикладі 2.1 ми знайшли вигляд оціночного рівняння, що описує залежність валової продукції від основних виробничих фондів та затрат праці у вигляді:

$$\hat{y} = -0,87 + 1,06x_1 - 0,11x_2.$$

Обчислимо коефіцієнти граничної продуктивності за формулою (2.25):

$Г_1 = \dfrac{\partial y}{\partial x_1} = a_1 = 1,06$, а це означає, що при збільшенні основних виробничих фондів на 1 млн. грн. валова продукція зросте на 1,06 млн. грн.

$Г_2 = \dfrac{\partial y}{\partial x_2} = a_2 = -0,11$, а значить випуск валової продукції зменшиться на 0,11 млн.грн. при збільшенні затрат робочого часу на 1 млн. люд.-год.

Для знаходження значень коефіцієнтів еластичності використаємо формулу (2.24).

Спочатку знайдемо середні значення змінних y, x_1 та x_2, взявши значення їх сум з прикладу 2.2:

$$\bar{x}_1 = \frac{\sum_{i=1}^{n} x_{1i}}{n} = \frac{40,3}{10} = 4,03; \qquad \bar{x}_2 = \frac{\sum_{i=1}^{n} x_{2i}}{n} = \frac{31,9}{10} = 3,19;$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{30,6}{10} = 3,06.$$

Тоді:

$$E_1 = a_1 \frac{\bar{x}_1}{\bar{y}} = 1{,}06 \frac{4{,}03}{3{,}06} = 1{,}4,$$

а це значить, що при збільшенні основних виробничих фондів на 1 % валовий випуск продукції зросте на 1,4 %.

$$E_2 = a_2 \frac{\bar{x}_2}{\bar{y}} = -0{,}11 \frac{3{,}19}{3{,}06} = -0{,}12,$$

а значить при збільшенні затрат робочого часу на 1 % валовий випуск продукції зменшиться на 0,12 %.

Коефіцієнт сумарної еластичності

$$E = E_1 + E_2 = 1{,}4 - 0{,}12 = 1{,}28$$

вказує на те, що при одночасному збільшенні і основних виробничих фондів і затрат робочого часу на 1 % валовий випуск продукції зросте на 1,28 %.

2.8. Перевірка значущості оцінок параметрів економетричної моделі

Значущість коефіцієнта множинної кореляції та оцінок параметрів економетричної моделі перевіряється аналогічно моделі парної регресії за t–критерієм Стьюдента.

Для оцінки значущості коефіцієнта множинної кореляції обчислюємо емпіричне значення параметру t:

$$t_{емп} = \frac{R\sqrt{n-m-1}}{\sqrt{1-R^2}}, \tag{2.27}$$

яке порівнюється з критичним значенням $t_{кр}$, що знаходиться за таблицями розподілу Стьюдента при заданому рівні значущості α та $k= n$-m-1 ступенях вільності.

Правило використання критерію полягає у наступному:

- якщо $|t_{емп}| > t_{кр}$, то нульова гіпотеза H_o на рівні значущості α відкидається і приймається альтернативна гіпотеза H_1 про існування залежності між змінними;
- якщо $|t_{емп}| \le \mathrm{t}_{\text{кр}}$, то нульова гіпотеза H_o на рівні значущості α приймається.

Перевірку нульових гіпотез стосовно параметрів b_0, b_1 та b_2 економетричної моделі проводять аналогічно. Спочатку висуваємо нульові гіпотези:

$$H_0: b_0=0, \quad H_0: b_1=0, \quad H_0: b_2=0.$$

Альтернативними будуть гіпотези:

$$H_1: b_0 \neq 0, \quad H_1: b_1 \neq 0, \quad H_1: b_2 \neq 0.$$

Потім обчислюємо емпіричні значення параметра t за формулами:

$$t_{a_0} = \frac{|a_0|}{\sigma_{a_0}}; \qquad t_{a_1} = \frac{|a_1|}{\sigma_{a_1}}; \qquad t_{a_0} = \frac{|a_2|}{\sigma_{a_2}}. \qquad (2.26)$$

Емпіричне значення параметру порівнюють з критичним, знайденим за таблицями Стьюдента для заданого рівня значущості α та k=n-m-1 ступенів вільності. Якщо

$t_{емп} > t_{kp}$, то нульова гіпотеза H_o із рівнем значущості α відкидається і приймається альтернативна гіпотеза H_1. Тоді відповідна оцінка вважається статистично значимою. Якщо ж $t_{емп} \le \mathrm{t}_{\text{кр}}$, то нульова гіпотеза H_o для рівня значущості α приймається, а відповідна оцінка не є статистично значимою.

Приклад 2.7. На основі даних прикладу 2.1 виконати перевірки нульових гіпотез стосовно коефіцієнта кореляції та параметрів економетричної моделі.

♦***Розв'язування.***

Висуваємо нульову гіпотезу H_o: $R_{ген}$=0 (робимо припущення, що коефіцієнт кореляції генеральної сукупності рівний нулю). Альтернативною гіпотезою буде H_1: $R_{ген} \neq 0$.

Далі для заданої вибірки з $k=n-m-1$ ступенями вільності обчислимо емпіричне значення критерію Стьюдента:

$$t_{\text{емп}} = \frac{R\sqrt{n-m-1}}{\sqrt{1-R^2}} = \frac{0{,}92\sqrt{10-2-1}}{\sqrt{1-0{,}92^2}} = 6{,}23$$

Для заданої ймовірності p=0,9 (α=1-p=1-$0{,}9$=$0{,}1$) і $k=10$-2-$1=7$ ступенів вільності знаходимо табличне значення $t_{кр.}$=1,89.

Оскільки $|t_{\text{емп}}| > t_{kp}$, то з надійністю p=0,9 гіпотезу H_o необхідно відкинути і прийняти альтернативну гіпотезу H_1 про існування залежності між змінними. Отже, у 90 % вибірок із генеральної сукупності коефіцієнт множинної кореляції не дорівнює нулю.

Далі виконаємо перевірку нульових гіпотез відносно *b_0, b_1* та *b_2*. Для цього спочатку обчислимо елементи дисперсійно-коваріаційної матриці, по головній діагоналі якої знаходяться дисперсії оцінок *a_0, a_1* та *a_2*, використавши формулу (2.14):

$$\operatorname{cov}(A) = \sigma_u^2 \left(X' \cdot X\right)^{-1} = \begin{bmatrix} \sigma_{a_0}^2 & \dots & \sigma_{a_0 a_m} \\ \vdots & \vdots & \vdots \\ \sigma_{a_m a_0} & \dots & \sigma_{a_m}^2 \end{bmatrix},$$

а для обчислення дисперсії помилок формулу (2.15):

$$\sigma_u^2 = \frac{e'e}{n-m-1} = \frac{Y' \cdot Y - A'X'Y}{n-m-1}.$$

Елементи матриці $(X'X)^{-1}$, векторів Y, $X'Y$ та A візьмемо з прикладу 2.1:

$$\left(X'X\right)^{-1} = \begin{pmatrix} 1{,}65 & -0{,}14 & -0{,}3 \\ -0{,}14 & 0{,}09 & -0{,}07 \\ -0{,}3 & -0{,}07 & 0{,}18 \end{pmatrix}; X' \cdot Y = \begin{pmatrix} 30{,}6 \\ 139{,}44 \\ 102{,}91 \end{pmatrix}; A = \begin{pmatrix} -0{,}87 \\ 1{,}06 \\ -0{,}11 \end{pmatrix}.$$

$$Y' = (1{,}2 \quad 1{,}4 \quad 1{,}8 \quad 2{,}1 \quad 2{,}8 \quad 3{,}2 \quad 3{,}5 \quad 4{,}4 \quad 4{,}9 \quad 5{,}3);$$

$$A' = (-0{,}87 \quad 1{,}06 \quad -0{,}11);$$

$$Y = \begin{pmatrix} 1,2 \\ 1,4 \\ 1,8 \\ 2,1 \\ 2,8 \\ 3,2 \\ 3,5 \\ 4,4 \\ 4,9 \\ 5,3 \end{pmatrix}$$

Тоді обчислимо $Y'Y = 112,84$ і $A'X'Y = 110,2$, а потім:

$$\sigma_u^2 = \frac{e'e}{n-m-1} = \frac{Y' \cdot Y - A'X'Y}{n-m-1} = \frac{112,84 - 110,2}{10-2-1} = 0,377.$$

Перемноживши дисперсію помилок на діагональні елементи матриці $(X'X)^{-1}$, отримаємо дисперсії оцінок, коренем з яких є середні квадратичні відхилення:

$$\sigma_{a_0}^2 = 0,377 \cdot 1,65 = 0,62; \qquad \sigma_{a_0} = \sqrt{0,62} = 0,79;$$

$$\sigma_{a_1}^2 = 0,377 \cdot 0,09 = 0,03; \qquad \sigma_{a_1} = \sqrt{0,03} = 0,18;$$

$$\sigma_{a_2}^2 = 0,377 \cdot 0,18 = 0,07; \qquad \sigma_{a_2} = \sqrt{0,07} = 0,26.$$

Дальше висуваємо гіпотезу H_o: b_0=0 проти альтернативної H_1: $b_0 \neq 0$. Для цього знаходимо емпіричне значення за формулою:

$$t_{\text{емп}} = \frac{|a_0|}{\sigma_{a_0}} = \frac{|-0,87|}{0,79} = 1,1$$

Оскільки емпіричне значення t менше критичного ($t_{кр.}$=1,89), то нульова гіпотеза приймається і робиться висновок, що параметр b_0 може бути рівним нулю в генеральній сукупності.

Перевіримо нульову гіпотезу H_o: b_1=0. Обчислимо

$$t_{\text{емп}} = \frac{|a_1|}{\sigma_{a_1}} = \frac{1,06}{0,18} = 5,86 .$$

$t_{емп}>t_{кр}$, тому нульова гіпотеза відхиляється, значить b_1 не може бути рівним нулю в генеральній сукупності, а отже оцінка a_1, розрахована за даними вибірки, є статистично значимою.

Здійснимо перевірку нульової гіпотези стосовно параметру b_2. Порахуємо

$$t_{\text{емп}} = \frac{|a_2|}{\sigma_{a_2}} = \frac{|-0,11|}{0,14} = 0,44 .$$

Оскільки емпіричне значення t менше критичного, то нульова гіпотеза приймається, значить параметр b_2 може бути рівним нулю в генеральній сукупності, а отже оцінка a_2, розрахована за даними вибірки, не є статистично значимою.

2.9. Оцінка якості економетричних моделей

Якість лінійних економетричних моделей оцінюється стандартним для економіко-математичних задач способом: за адекватністю та точністю.

Для перевірки адекватності множинної регресійної моделі, як і у випадку парної регресії, використовується F-критерій Фішера. У даному випадку нульова гіпотеза узагальнюється:

$$H_0 : b_1 = b_2 = ... = b_m = 0.$$

Тоді альтернативною гіпотезою буде H_1 : хоча б одне значення b_j відмінне від нуля. У випадку невиконання гіпотези H_0 приймається гіпотеза H_1. Отже, не всі параметри незначною мірою відрізняються від нуля. Це свідчить про те, що включені до моделі фактори пояснюють змінну результативного показника.

Для перевірки гіпотези H_0 використовують F-критерій Фішера з m та (n-m-1) ступенями вільності:

$$F = \frac{\dfrac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{m}}{\dfrac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n-m-1}} = \frac{(n-m-1)\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{m\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2},$$

де m – кількість незалежних факторів, n – загальна кількість спостережень.

На практиці для обчислення емпіричного значення параметру F, у випадку, коли знайдене значення коефіцієнта множинної детермінації R^2, використовують формулу:

$$F = \frac{(n-m-1)R^2}{m(1-R^2)}. \tag{2.29}$$

Далі для заданого рівня значущості α і ступенів вільності $k_1=m$ та $k_2=n–m-1$ знаходимо табличне (критичне) значення критерію Фішера – $F_{кр.}(k_1, k_2, \alpha)$. Знайдене розрахункове значення критерію $F = F_{емп}$ порівнюємо з табличним: якщо $F_{емп}>F_{кр.}$, тоді гіпотеза H_0 відхиляється і приймається альтернативна, що свідчить про адекватність побудованої моделі, іншими словами підтверджується наявність істотного зв'язку між залежною та незалежними змінними побудованої економетричної моделі. У протилежному випадку нульова гіпотеза приймається, і модель вважається неадекватною.

Приклад 2.8. Здійснити перевірку адекватності економетричної моделі, отриманої в прикладі 2.1.

♦***Розв'язування.***

Обчислимо емпіричне значення параметру F:

$$F_{емп} = \frac{R^2(n-m-1)}{m(1-R^2)} = \frac{0{,}85\cdot(10-2-1)}{1\cdot(1-0{,}85)} = 23{,}8$$

Знайдемо табличне значення даного критерію ($F_{кр.}$) для рівня надійності p=0,95 та числа ступенів вільності k_1=m=2; k_2=n-m-1=10-2-1=7:

$$F_{кр} = 4,74.$$

Оскільки $F_{емп.} > F_{кр.}$, то отримане нами оціночне рівняння економетричної моделі

$$\hat{y} = -0,87 + 1,06x_1 - 0,11x_2$$

адекватне реальній дійсності і на його основі можна здійснювати прогнози.

3. Нелінійна економетрична модель

3.1. Загальні відомості про нелінійні функції регресій

У більшості випадків при побудові ЕМ залежності результативного показника від пояснюючих змінних зустрічаються випадки нелінійної форми зв'язку між ними. В загальному випадку модель парної регресії можна представити у вигляді:

$$y = f(x) + u,$$

де $f(x)$ - функція від x, а u – випадкова складова.

Нелінійна багатофакторна економетрична модель:

$$y = f(x_1, ..., x_m) + u.$$

При побудові економетричної моделі з двома змінними вид залежності, яка адекватно відобразить причинно-наслідкові зв'язки певного економічного процесу чи явища, визначаємо по діаграмі розсіювання.

Основні форми нелінійних регресійних моделей, які найчастіше використовуються при встановленні залежності між двома економічними показниками:

1) парабола $\hat{y} = a + b \cdot x + c \cdot x^2$

2) поліноміальна $\hat{y} = \sum_{i=1}^{n} a_i \cdot x^i$

3) степені 1/2 $\hat{y} = a + b \cdot \sqrt{x}$

4)логарифмічна $\hat{y} = a + b \cdot \ln(x)$

5) степенева $\hat{y} = a + bx^c$

6) експонента $\hat{y} = e^{a+bx}$

7) гіпербола $\hat{y} = a + b / x$

8) логістична $\hat{y} = a + \frac{b}{1 + e^{c+dx}}$

3.2. Класи нелінійних регресій

Розрізняють два класи нелінійних регресій:

1) I клас – це клас регресій, які нелінійні відносно пояснюючих змінних, але лінійні відносно оцінок параметрів економетричної моделі. Приклади таких функцій:

$$\hat{y} = a + b\ln x;$$

$$\hat{y} = a + b\frac{1}{x};$$

$$\hat{y} = a + b\cos x;$$

$$\hat{y} = a + b\sqrt{x};$$

$$\hat{y} = a_0 + a_1 x + a_2 x^2$$

$$\hat{y} = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m.$$

Цей клас нелінійних регресій ще називають квазілінійними регресіями, тому що їх можна привести до лінійного виду шляхом заміни, тобто лінеаризувати.

2) II клас нелінійних регресій – це клас економетричних моделей, які нелінійні і відносно пояснюючої змінної (чи пояснюючих змінних) і відносно оцінок параметрів ЕМ. Приклади таких функцій:

$$\hat{y} = ax^b; \qquad \hat{y} = ab^x; \qquad \hat{y} = a + \frac{1}{b + e^{c+dx}}.$$

степенева показникова логістична.

Такий клас регресій часто зустрічається при дослідженні економічних процесів чи явищ. Суттєвим його недоліком є те, що неможливо провести лінеаризацію та застосувати метод найменших квадратів.

Великий економічний інтерес серед функцій нелінійної регресії другого класу представляють виробничі функції.

Першою класичною виробничою функцією є степенева функція Кобба-Дугласа:

$$\hat{y} = aK^{\alpha} \cdot L^{1-\alpha},$$

де $\hat{y}$ – обсяг випуску продукції; K – затрати капіталу; L – затрати праці; a – коефіцієнт пропорційності; α - параметр функції або коефіцієнт еластичності по затратах праці.

Для загального випадку степенева функція Кобба-Дугласа має вигляд:

$$\hat{y} = a_0 x_1^{a_1} x_2^{a_2} \ldots x_m^{a_m}.$$

де $\hat{y}$ – обсяг випуску продукції (національний дохід);

$x_1, x_2, \ldots x_m$ – фактори впливу на результативний показник;

$a_1, a_2, \ldots a_m$ – коефіцієнти еластичності.

Виробничі функції використовуються в двох аспектах: як самостійні економіко-математичні моделі для аналізу зв'язків між економічними показниками, прийняття рішень, прогнозування; як складові частини складніших моделей для оптимального планування та управління виробництвом.

Нелінійні функції регресій I та II класу інакше ще називають відповідно суттєво лінійними та суттєво нелінійними.

3.3. Приведення нелінійних економетричних моделей до лінійного виду

Для того, щоб привести до лінійного виду нелінійні функції регресії першого класу використовують метод заміни змінних. Розглянемо це на прикладах функцій регресій цього класу, приведених в п.3.2:

1) $\hat{y} = a + b\ln x$

Здійснимо заміну $z = \ln x$ і в результаті матимемо функцію регресії лінійного виду $\hat{y} = a + bz$. Аналогічно для інших функцій:

2) $\hat{y} = a + b\frac{1}{x}$; заміна: $z = \frac{1}{x}$, отримаємо $\hat{y} = a + bz.$

3) $\hat{y} = a + b\cos x$; заміна: $z = \cos x$, маємо $\hat{y} = a + bz.$

4) $\hat{y} = a + b\sqrt{x}$; заміна: $z = \sqrt{x}$, отримаємо $\hat{y} = a + bz.$

5) $\hat{y} = a_0 + a_1 x + a_2 x^2$; заміна: $z_1 = x$; $z_2 = x^2$,
маємо $\hat{y} = a_0 + a_1 z_1 + a_2 z_2.$

6) $\hat{y} = a_0 + a_1 x + a_2 x^2 + ... + a_m x^m$; заміна:
$z_1 = x$; $z_2 = x^2$; $z_m = x^m$, отримаємо $\hat{y} = a_0 + a_1 z_1 + a_2 z_2 + ... + a_m z_m.$

Для знаходження оцінок параметрів нелінійних економетричних моделей другого класу використовують лінійне перетворення функцій. Наприклад прологарифмуємо степеневу функцію:

$$\ln \hat{y} = \ln(ax^b);$$

$$\ln \hat{y} = \ln a + \ln x^b;$$

$$\ln \hat{y} = \ln a + b \ln x.$$

Тоді здійснюємо заміни:

$z = \ln \hat{y};$ $\qquad$ $a^* = \ln a;$ $\qquad$ $x^* = \ln x.$

В результаті отримаємо лінійну функцію: $z = a^* + bx^*.$

Аналогічні лінійні перетворення можна зробити й для показникової функції:

$$\ln \hat{y} = \ln(a \cdot b^x);$$

$$\ln \hat{y} = \ln a + x \ln b.$$

В результаті замін:

$z = \ln \hat{y};$ $\qquad$ $a^* = \ln a;$ $\qquad$ $b^* = \ln b.$

отримаємо лінійну функцію: $z = a^* + b^* x.$

Ми бачимо, що з допомогою методу заміни змінних чи лінійних перетворень (наприклад, логарифмування) нелінійні економетричні моделі можна привести до лінійного виду, а тоді вже до отриманих лінійних функцій застосовуємо метод найменших квадратів для знаходження оцінок параметрів одно- чи багатофакторної економетричної моделі та проводимо весь економетричний аналіз: будуємо довірчі інтервали, здійснюємо перевірку нульових гіпотез та перевіряємо модель на адекватність. Коли будуть знайдені інтервальні оцінки для отриманих лінійних функцій регресій, то з допомогою обернених перетворень (зворотної заміни чи використання антилогарифмів) знаходимо довірчі інтервали для нелінійних економетричних моделей.

Слід зазначити, що для побудови нелінійної однофакторної економетричної моделі доцільно використовувати можливості офісної програми EXCEL.

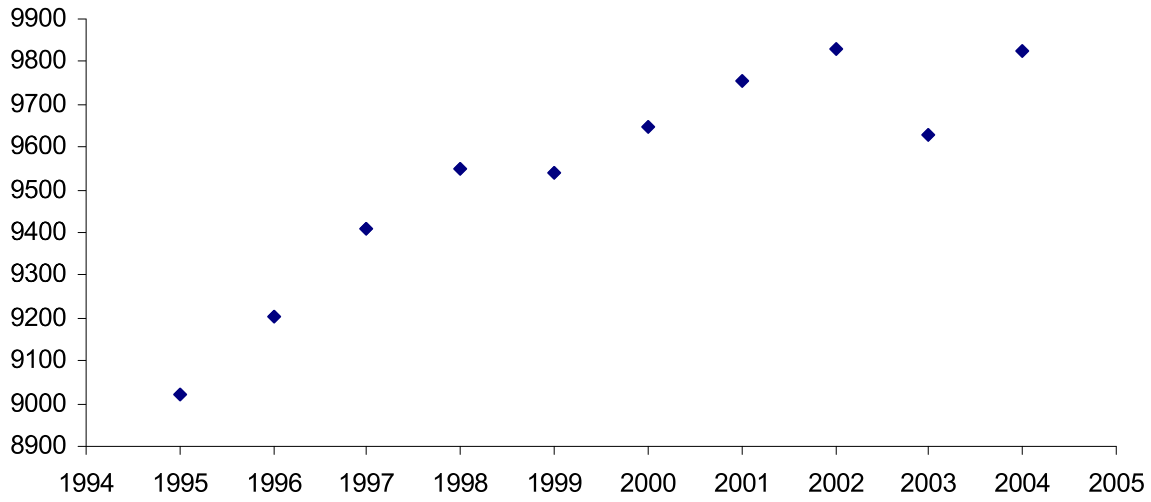
Приклад 3. На основі наведених в таблиці 3 по роках даних валового випуску продукції для деякого підприємства знайти рівняння лінії, яка найточніше опише тенденцію.

Таблиця 3

Роки	Валова продукція, млн. грн.
1995	9021
1996	9205
1997	9411
1998	9550
1998	9542
2000	9648
2001	9753
2002	9829
2003	9630
2004	9827

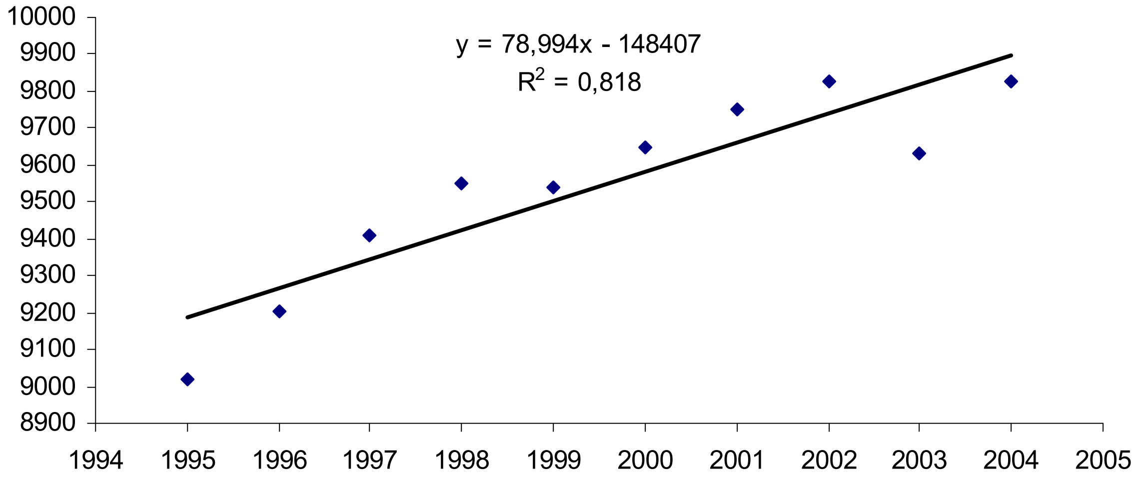
♦***Розв'язування.***

Занесемо дані валового випуску і роки в таблицю EXCEL і на їх основі будуємо діаграму розсіювання. Для цього лівою клавішею миші натискаємо на „Майстер діаграм”, отримуємо вікно „Мастер диаграмм: тип диаграммы”, вибираємо тип „Точечная”, натискаємо кнопку „Далее”, отримуємо вікно „Мастер диаграмм: источник данных диаграммы”, у віконечко „Диапазон” заносимо адрес області, де знаходяться вхідні дані, відмічаємо у віконечку яким чином записані дані – в рядках чи стовпцях, „Далее”, отримуємо вікно „Мастер диаграмм: параметры диаграммы”, забираємо лінії сітки, „Далее”, „Готово” і отримуємо діаграму розсіювання:

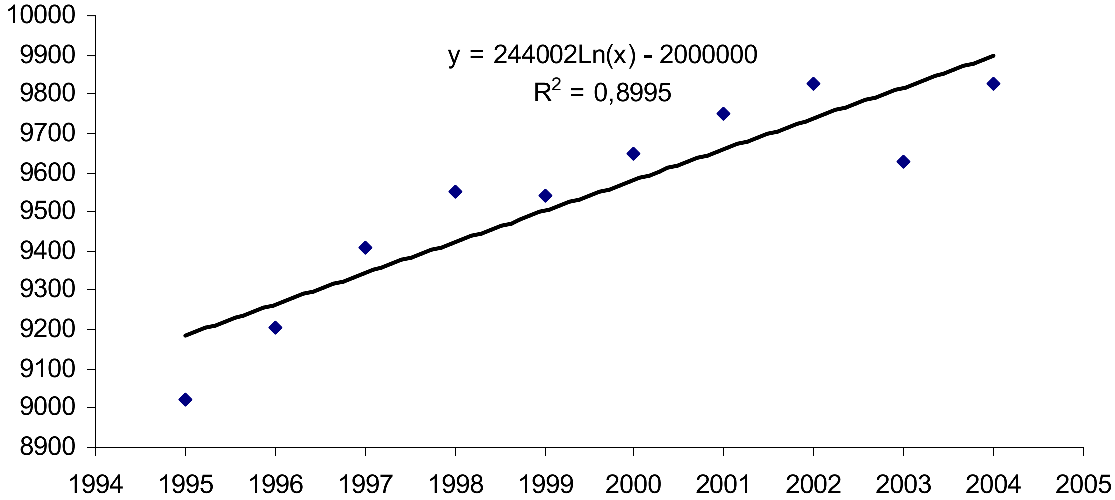


Дальше правою кнопкою миші натискаємо на одній із точок діаграми і вибираємо з отриманого списку „Добавить линию тренда”. Потім одержимо вікно „Линия тренда”, в якому вибираємо тип лінії тренду, а у вікні „Параметры” ставимо відмітки у ллітинках „Показывать уравнение на диаграмме” та „Поместить на диаграмму R^2”, „Готово”. Розглянемо для даних нашого прикладу можливі лінії тренду:

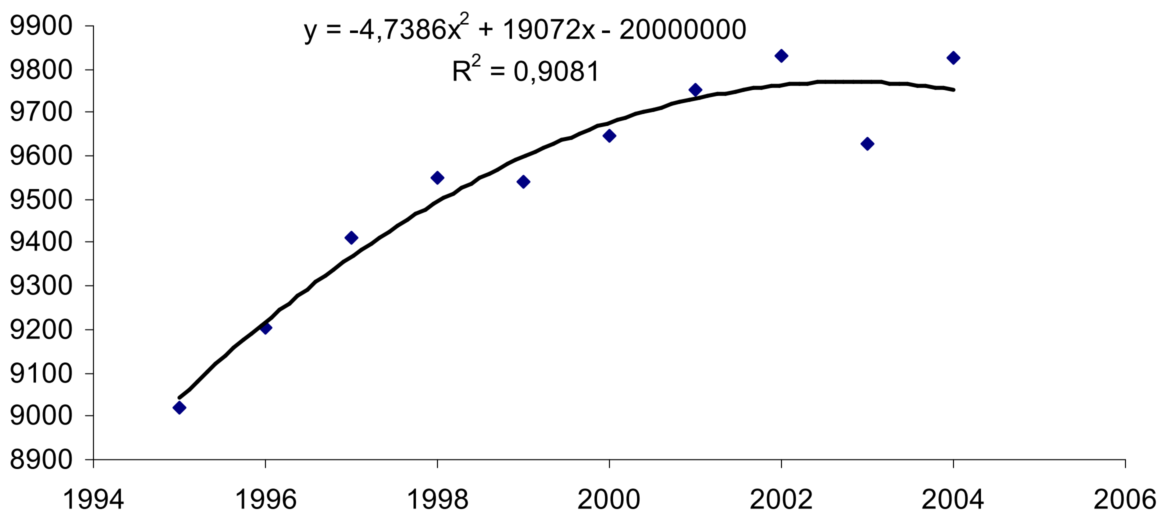
- лінійна



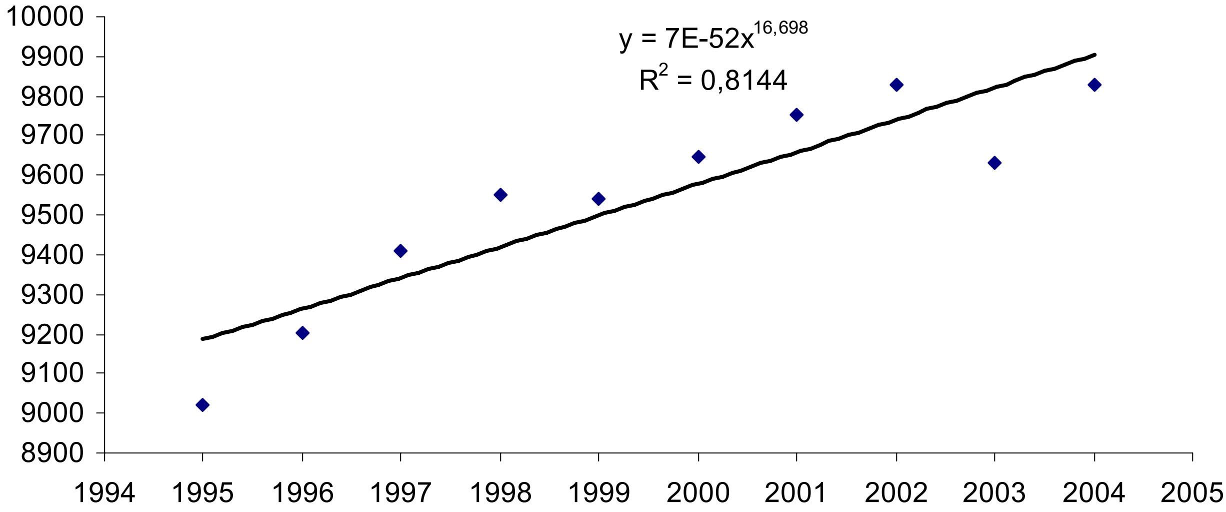
- логарифмічна



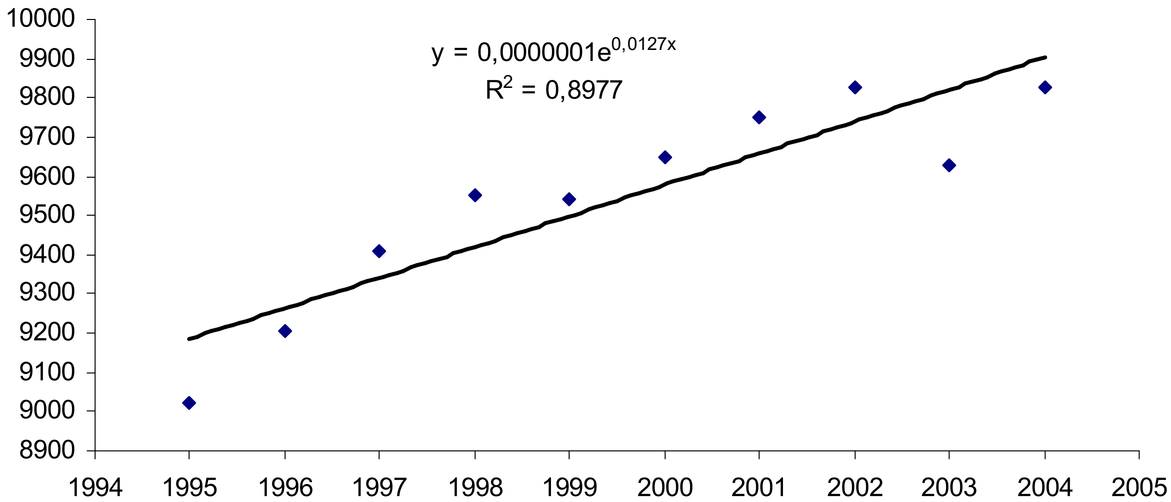
- поліноміальна



- степенева



- експоненціальна



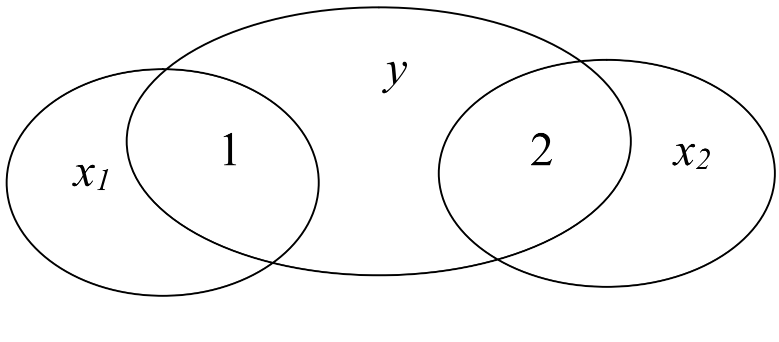
Бачимо, що найбільше значення коефіцієнта детермінації R^2 для поліноміальної залежності, тобто найкраще апроксимує наші дані нелінійна економетрична модель виду

$$\hat{y} = -4{,}7386x^2 + 19072x - 20000000.$$

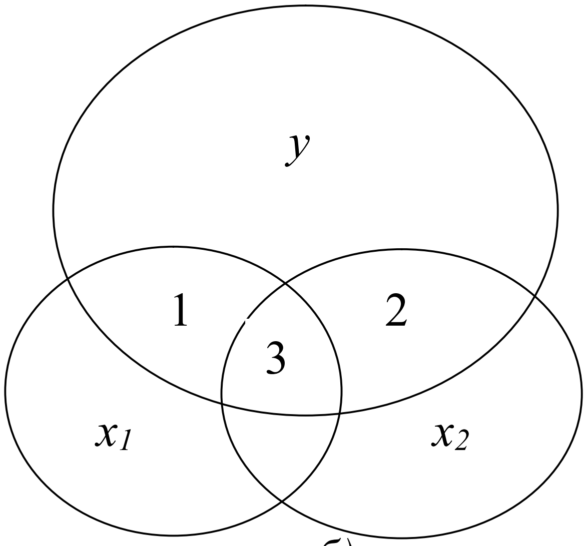
4. Мультиколінеарність в багатофакторних економетричних моделях

4.1. Визначення мультиколінеарності

Досить часто на практиці при побудові багатофакторних економетричних моделей зустрічаються випадки, коли при перевірці нульових гіпотез стосовно параметрів економетричної моделі розрахункові (емпіричні) значення параметру *t* досить малі, що свідчить про незначущість цих параметрів, а значить і незалежних змінних, що їм відповідають, а емпіричне значення *F*–критерію при перевірці економетричної моделі на адекватність велике (більше табличного), а значить отримана економетрична модель адекватна реальній дійсності, тобто економетрична модель є значущою. Ця ситуація є сама по собі суперечливою. Причиною цього може бути те, що між незалежними змінними існує тісний лінійний зв’язок, тобто присутнє явище мультиколінеарності. Це явище схематично можна подати у вигляді кругової діаграми:



а)



б)

Рис. 4. Зв’язок між результативним показником
і факторами моделі

На рис.4 *а)* немає зв’язку між факторами, а у випадку *б)* зв’язок є.

Області 1 та 2 – окремий вплив факторів x_1 та x_2 на результативний показник y.

Область 3 – спільний вплив обох факторів на результативний показник. У випадку спільного впливу двох змінних важко визначити вплив кожного окремого фактора на результативний показник.

Мультиколінеарність багатофакторної економетричної моделі – це залежність однієї пояснюючої змінної від іншої чи декількох інших незалежних змінних.

Мультиколінеарність буває строга і нестрога. Якщо фактори x_s та x_l залежні між собою і цю залежність можна подати у вигляді $x_s = cx_l$, то між ними існує строга мультиколінеарність, а якщо $x_s = cx_l + g$, де g – певне відхилення, то мультиколінеарність нестрога. При строгій мультиколінеарності неможливо точно знайти оцінки параметрів економетричної моделі, а при нестрогій вони ненадійні, тобто в любому випадку явище мультиколінеарності небажане при побудові багатофакторних економетричних моделей.

4.2. Способи виявлення мультиколінеарності

Виявити явище мультиколінеарності можна декількома способами:

1) Якщо значення коефіцієнтів парної кореляції між двома незалежними змінними більше 0,8, то це означає, що між ними існує мультиколінеарність;
2) Якщо значення визначника кореляційної матриці близьке до нуля, то це свідчить про те, що в масиві незалежних змінних існує мультиколінеарність;
3) Якщо невелика зміна (додавання чи відкидання невеликого числа спостережень) приводить до суттєвої зміни значень оцінок параметрів економетричної моделі, чи знаків цих оцінок, то в даному випадку між незалежними змінними може існувати мультиколінеарність;

4) При побудові багатофакторної економетричної моделі ми отримали значення оцінок параметрів, що мають з економічної точки зору невірні знаки або невиправдано великі чи малі значення, то причиною цього може бути явище мультиколінеарності.

На практиці найчастіше для виявлення мультиколінеарності використовують метод Фаррара-Глобера. В основі алгоритму цього методу є три статистичні критерії:

- критерій χ^2, з допомогою якого встановлюємо чи існує мультиколінеарність у всьому масиві незалежних змінних;
- *F*-критерій Фішера, за яким визначаємо мультиколінеарність кожної незалежної змінної з масивом інших;
- *t*-критерій Стьюдента, на основі якого перевіряємо наявність мультиколінеарності кожної пари незалежних змінних.

Сам алгоритм методу складається з таких етапів:

1) Обчислюємо емпіричне значення χ^2 за формулою:

$$\chi^2_{емп.} = -[n - 1 - \frac{1}{6}(2m + 5)]\ln\det[R],$$

де n – об’єм вибірки (кількість спостережень);

m – кількість незалежних змінних;

$\det[R]$ – визначник кореляційної матриці $[R]$;

ln – натуральний логарифм.

2) Для заданої ймовірності p (рівня значимості α=1–p) і числа ступенів вільності $k = \frac{1}{2}m(m-1)$ знаходимо табличне значення $\chi^2_{кр.}$ Дальше порівнюємо емпіричне і табличне значення. Якщо $\chi^2_{емп.} \le \chi^2_{кр.}$, то з заданою ймовірністю p, яку називають довірчою ймовірністю, або надійністю, можна стверджувати, що загальна мультиколінеарність відсутня і дослідження мультиколінеарності на цьому закінчується. Якщо ж $\chi^2_{емп.} > \chi^2_{кр.}$, то з заданою надійністю p вважаємо, що в масиві незалежних змінних існує мультиколінеарність.

3) Розраховуємо емпіричне значення параметру F за формулою:

$$F_{i\,емп.} = (z_{ii} - 1)\frac{n-m}{m-1}; \qquad i = \overline{1,m},$$

де z_{ii} – елементи головної діагоналі матриці $[Z]$, оберненої до кореляційної $[R]$.

4) Для заданої ймовірності p (рівня значимості α=1–p) і числа ступенів вільності $k = n - m$ знаходимо табличне значення $F_{кр.}$ Дальше порівнюємо емпіричне і табличне значення. Якщо $F_{i\,емп.} \le F_{кр.}$, то з заданою ймовірністю p,

можна стверджувати, що змінна x_i не мультиколінеарна з масивом інших незалежних змінних. Якщо ж $F_{i\,емп.} > F_{кр.}$, то існує мультиколінеарність змінної x_i з масивом інших незалежних змінних.

5) Знаходимо емпіричне значення параметру :

$$t_{ls} = r_{ls}\sqrt{\frac{n-m-1}{1-r_{ls}^2}}; \quad l,s=\overline{1,m}; \quad l \neq s,$$

де r_{ls} – коефіцієнти частинної кореляції між незалежними змінними x_l та x_s, значення яких обчислюємо за формулою:

$$r_{ls} = -\frac{z_{ls}}{\sqrt{z_{ll}\cdot z_{ss}}}; \quad l=\overline{1,m}; \quad s=\overline{1,m}; \quad l \neq s,$$

а z_{ls}, z_{ll} та z_{ss} – елементи матриці [Z].

6) За таблицями Стьюдента для рівня значущості α та числа ступенів вільності $k = n - m - 1$ знаходимо табличне значення $t_{кр.}$ і порівнюємо його з емпіричним. Якщо $|t_{ls}| \leq t_{кр.}$, то з заданою ймовірністю p можна стверджувати, що між змінними x_l та x_s немає мультиколінеарності. Якщо ж $|t_{ls}| > t_{кр.}$, то мультиколінеарність між цими змінними існує.

Таким чином, при побудові багатофакторних економетричних моделей ми можемо встановити, чи існує мультиколінеарність, але перед нами стоїть питання як її позбутися, або зменшити її вплив.

4.3. Методи усунення мультиколінеарності

Для зменшення впливу чи усунення мультиколінеарності існує декілька способів:

- якщо між двома факторами моделі існує мультиколінеарність, то один із них виключаємо з розгляду;
- якщо між двома факторами моделі x_l та x_s існує мультиколінеарність і з економічної точки зору можна здійснити заміну $x_l^* = x_l - x_s$ то досить часто змінні x^*_l та x_s вже не сильно корелюють;
- інколи причиною мультиколінеарності є замалий обсяг спостережень і часто при збільшенні об'єму вибірки мультиколінеарність зникає.
-

<u>Приклад 4.</u> Дослідити наявність мультиколінеарності між незалежними змінними з допомогою методу Фаррара-Глобера на основі умовних даних таблиці 4 та побудувати отриману економетричну модель.

Таблиця 4

Прибуток y	Вартість ОВФ, x_1	Затрати праці, x_2	Ціна одиниці продукції, x_3	Собівартість одиниці продукції, x_4
8	12	21	6	3
17	16	26	9	5
13	15	25	7	3,2

14	14	23	9	4,6
9	18	30	8	4,1
16	11	18	6	3,1
12	15	25	10	4,9
15	14	24	9	3,3
10	19	32	9	4,7
18	10	17	8	4,4

♦Розв’язування.

Знаходимо елементи кореляційної матриці, використавши в офісній програмі EXCEL ***Сервис – Анализ данных – Корреляция:***

$$[R]=\begin{bmatrix} 1 & 0{,}996 & 0{,}496 & 0{,}434 \\ 0{,}996 & 1 & 0{,}473 & 0{,}381 \\ 0{,}496 & 0{,}473 & 1 & 0{,}808 \\ 0{,}434 & 0{,}381 & 0{,}808 & 1 \end{bmatrix}$$

і обчислюємо визначник кореляційної матриці $\det[R]=0{,}00109$. Знаходимо емпіричне значення χ^2 за формулою:

$$\chi^2_{емп.}=-[n-1-\frac{1}{6}(2m+5)]\ln\det[R]=-[10-1-\frac{1}{6}(2\cdot 4+5)]\ln 0{,}00109=$$
$$=46{,}63.$$

За таблицями знаходимо:

$$\chi^2_{кр.}=12{,}592\,(\alpha=0{,}05;\quad k=\frac{1}{2}\cdot 4(4-1)=6).$$

Оскільки $\chi^2_{емп.} > \chi^2_{кр.}$, то з ймовірністю p=0,95 можемо стверджувати, що в масиві незалежних змінних існує мультиколінеарність.

Дальше знаходимо матрицю, обернену до кореляційної

$$[Z] = \begin{bmatrix} 248,02 & -243,25 & 11,59 & -24,26 \\ -243,25 & 239,85 & -11,98 & 23,8 \\ 11,59 & -11,98 & 3,71 & -3,46 \\ -24,26 & 23,8 & -3,46 & 5,25 \end{bmatrix}$$

і емпіричні значення параметру F для кожної незалежної змінної:

$$F_{1емп.} = (z_{11} - 1)\frac{n-m}{m-1} = (248,02 - 1)\frac{10-4}{4-1} = 494,05;$$

$$F_{2емп.} = (239,85 - 1)\frac{10-4}{4-1} = 477,71;$$

$$F_{3емп.} = (3,71 - 1)\frac{10-4}{4-1} = 5,42;$$

$$F_{4емп.} = (5,25 - 1)\frac{10-4}{4-1} = 8,5.$$

Знаходимо табличне значення $F_{кр.}$=4,76 (для p=0,95 і $k = n - m = 10 - 4 = 6$) і порівнюємо емпіричне і табличне значення. Оскільки $F_{1емп.} > F_{кр.}$ та $F_{2емп.} > F_{кр.}$, то робимо висновок, що змінні x_1 та x_2 мультиколінеарні з масивом інших незалежних змінних.

Перевіримо між якими саме парами незалежних змінних існує мультиколінеарність з допомогою критерію Стьюдента. Знайдемо коефіцієнти частинної кореляції:

$$r_{12} = -\frac{z_{12}}{\sqrt{z_{11} \cdot z_{22}}} = -\frac{-243,25}{\sqrt{248,02 \cdot 239,85}} = 0,997;$$

$$r_{13} = -\frac{11,59}{\sqrt{248,02 \cdot 3,71}} = -0,39;$$

$$r_{14} = -\frac{-24,26}{\sqrt{248,02 \cdot 5,25}} = 0,67;$$

$$r_{23} = -\frac{-11,98}{\sqrt{239,85 \cdot 3,71}} = 0,4;$$

$$r_{24} = -\frac{23,8}{\sqrt{239,85 \cdot 5,25}} = -0,67;$$

$$r_{34} = -\frac{-3,46}{\sqrt{3,71 \cdot 5,25}} = 0,78.$$

Дальше розраховуємо емпіричні значення параметру t для кожної пари незалежних змінних:

$$t_{12} = r_{12}\sqrt{\frac{n-m-1}{1-r_{12}^2}} = 0,997\sqrt{\frac{10-4-1}{1-0,997^2}} = 30,43;$$

$$t_{13} = -0,39\sqrt{\frac{10-4-1}{1-(-0,39)^2}} = -0,93;$$

$$t_{14} = 0,67\sqrt{\frac{10-4-1}{1-0,67^2}} = 2,03;$$

$$t_{23} = 0,4\sqrt{\frac{10-4-1}{1-0,4^2}} = 0,98;$$

$$t_{24} = -0,67\sqrt{\frac{10-4-1}{1-(-0,67)^2}} = -2,02;$$

$$t_{34} = 0,78\sqrt{\frac{10-4-1}{1-0,78^2}} = 2,82.$$

За таблицями Стьюдента для рівня значущості α=0,05 та числа ступенів вільності $k = n - m - 1 = 10 - 4 - 1 = 5$ знаходимо табличне значення $t_{кр}$=2,447. і порівнюємо його з емпіричним. Оскільки $|t_{12}| > t_{кр.}$, то існує мультиколінеарність між змінними x_1 та x_2, $|t_{34}| > t_{кр.}$, значить змінні x_3 та x_4 теж мультиколінеарні. Виключимо з масиву незалежних змінних x_2 і перевіримо, чи існує ще мультиколінеарність в масиві змінних, що залишилися.

Знаходимо елементи нової кореляційної матриці

$$[R] = \begin{bmatrix} 1 & 0,496 & 0,434 \\ 0,496 & 1 & 0,808 \\ 0,434 & 0,808 & 1 \end{bmatrix}$$

визначник якої $\det[R] = 0,26$. Емпіричне значення χ^2:

$$\chi^2_{емп.} = -[10 - 1 - \frac{1}{6}(2 \cdot 3 + 5)]\ln 0,26 = 9,63.$$

$$\chi^2_{кр.} = 7,815(\alpha = 0,05; \quad k = \frac{1}{2} \cdot 3(3-1) = 3).$$

Оскільки $\chi^2_{емп.} > \chi^2_{кр.}$, то в масиві незалежних змінних ще існує мультиколінеарність.

Матриця, обернена до кореляційної

$$[Z] = \begin{bmatrix} 1{,}33 & -0{,}56 & -0{,}13 \\ -0{,}56 & 3{,}11 & -2{,}27 \\ -0{,}13 & -2{,}27 & 2{,}89 \end{bmatrix}.$$

Тоді

$$F_{1емп.} = (z_{11} - 1)\frac{n-m}{m-1} = (1{,}33-1)\frac{10-3}{3-1} = 1{,}16;$$

$$F_{3емп.} = (3{,}11-1)\frac{7}{2} = 7{,}39; \qquad ,$$

$$F_{4емп.} = (2{,}89-1)\frac{7}{2} = 6{,}61.$$

$F_{кр.}$=4,74 (для p =0,95 і $k = n - m = 10 - 3 = 7$). Оскільки $F_{3емп.} > F_{кр.}$ та $F_{4емп.} > F_{кр.}$, то змінні x_3 та x_4 мультиколінеарні з масивом інших незалежних змінних.

Знаходимо коефіцієнти частинної кореляції:

$$r_{13} = -\frac{z_{13}}{\sqrt{z_{11} \cdot z_{33}}} = -\frac{-0{,}56}{\sqrt{1{,}33 \cdot 3{,}11}} = 0{,}28;$$

$$r_{14} = -\frac{-0{,}13}{\sqrt{1{,}33 \cdot 2{,}89}} = 0{,}06;$$

$$r_{34} = -\frac{-0{,}13}{\sqrt{3{,}11 \cdot 2{,}89}} = 0{,}76.$$

Дальше розраховуємо емпіричні значення параметру t для кожної пари незалежних змінних:

$$t_{13} = r_{13}\sqrt{\frac{n-m-1}{1-r_{13}^2}} = 0{,}28\sqrt{\frac{10-3-1}{1-0{,}28^2}} = 0{,}7;$$

$$t_{14} = 0{,}06\sqrt{\frac{6}{1-0{,}06^2}} = 0{,}16;$$

$$t_{34} = 0{,}76\sqrt{\frac{6}{1-0{,}76^2}} = 2{,}84.$$

$t_{кр}$=2,365. (для α=0,05 і $k = n - m - 1 = 10 - 3 - 1 = 6$). $|t_{34}| > t_{кр.}$, значить існує мультиколінеарність між змінними x_3 та x_4. Виключимо з масиву незалежних змінних x_4 і перевіримо, чи існує ще мультиколінеарність між змінними x_1 та x_3.

$$[R] = \begin{bmatrix} 1 & 0{,}496 \\ 0{,}496 & 1 \end{bmatrix}; \quad \det[R] = 0{,}75.$$

$$\chi^2_{емп.} = -[10 - 1 - \frac{1}{6}(2\cdot 2 + 5)]\ln 0{,}75 = 2{,}12.$$

$$\chi^2_{кр.} = 3{,}84\,(\alpha = 0{,}05; \quad k = \frac{1}{2}\cdot 2(2-1) = 1).$$

Оскільки $\chi^2_{емп.} < \chi^2_{кр.}$, то в масиві незалежних змінних немає мультиколінеарністі.

Слід зазначити, що при дослідженні мультиколінеарності використання критерію Фішера не обов’язкове, оскільки інколи з його допомогою не можна визначити, яку із незалежних змінних потрібно виключати з розгляду. Точну відповідь на це питання ми отримуємо за критерієм Стьюдента.

Тепер залишається побудувати економетричну модель (оціночне рівняння) залежності прибутку від вартості основних виробничих фондів та ціни одиниці продукції. Алгоритм побудови оціночного рівняння багатофакторної економетричної моделі подано в п.2.2. (приклад 2.1). Зауважимо, що оцінки параметрів економетричної моделі можна знайти також, використавши стандартну офісну програму EXCEL (***Сервис – Анализ данных – Регрессия)***.

Слід зазначити, що при дослідженні мультиколінеарності використання критерію Фішера не обов’язкове, оскільки інколи з його допомогою не можна визначити, яку із незалежних змінних потрібно виключати з розгляду. Точну відповідь на це питання ми отримуємо за критерієм Стьюдента.

Тепер залишається побудувати економетричну модель (оціночне рівняння) залежності прибутку від вартості основних виробничих фондів та ціни одиниці продукції. Алгоритм побудови оціночного рівняння багатофакторної економетричної моделі подано в п.2.2. (приклад 2.1).

Зауважимо, що оцінки параметрів економетричної моделі можна знайти також, використавши стандартну офісну програму EXCEL (***Сервис – Анализ данных – Регрессия)***.