Предмет економетрії.
Поняття моделі та моделювання

Економетрія виникла на початку 20-х років минулого століття як самостійна галузь наукових досліджень. Термін “економетрія” означає “вимірювання в економіці”.

Економетрія – це наука, що вивчає кількісні закономірності та взаємозв’язки економічних об’єктів і процесів з допомогою математико-статистичних методів та моделей. Це самостійна економіко-математична дисципліна, що об’єднує сукупність теоретичних результатів, способів, методів та моделей, призначених для того, щоб на базі економічної теорії, статистики математичної економіки та математико-статистичного інструментарію надавати конкретне кількісне вираження загальним (якісним) закономірностям та прогнозувати результати розвитку складних економічних процесів. Таким чином, сутність економетрії полягає в синтезі трьох складових: економіки, економічної статистики та математики.

Тому знання основ прикладної математики та методів економетричного аналізу даних є невід’ємною складовою професійних навичок майбутніх спеціалістів усіх освітніх рівнів, а відповідні програмні продукти – робочим інструментом кожного спеціаліста-аналітика.

Предметом економетрії є економіка в кількісному аспекті. Об’єкт вивчення – зв’язки між народно-господарським комплексом і його найпростішими складовими та поведінкою цих складових. Область економетрії визначається моделями математичної економіки. Економетрія шукає числові параметри моделей математичної економіки. Вона займається кількісним виміром економічних об’єктів як аналогічних еквівалентів математичної моделі.

Більш конкретне і загальне визначення економетрії та її предмету дослідження міститься в енциклопедії кібернетики: “Економетрія – напрямок в економіці, що ґрунтується на використанні математичних моделей для аналізу і прогнозування економічних явищ і пов’язаний із визначенням і оцінкою адекватності реальних явищ математичним уявленням про них”. Там же: “… важко розділити математичну економіку і економетрію, з одного боку, економетрію і економічну статистику, з іншого боку. Можна лише підкреслювати зв’язок математичної економіки і економетрії”. Побудова математичної моделі економіки завжди підтверджується оцінками адекватності такої моделі в реальній дійсності. Економічна статистика має справу з сталими і відносно нескладними економічними обчисленнями. Поява економетрії пов’язана з твердженням про недостатність таких економіко-статистичних обчислень для економічного аналізу і прогнозування. Найбільшого розвитку досягли в економетрії методи множинної кореляції.

Тобто, економетрія – прикладна економіко-математична дисципліна, яка синтезує економіку, математику та статистику і не може обмежуватись лише використанням регресійного аналізу в економіці, а повинна використовувати весь наявний математичний інструментарій прикладної математики для проведення економічних досліджень. Крім цього, економетрія виступає базовою основою для визначення альтернативних варіантів у системах прийняття управлінських рішень. В умовах конкуренції прогнозні рішення повинні прийматися на основі модельних альтернативних варіантів в тісному поєднанні з наявною інформаційною базою та сучасними програмними продуктами.

Метод “проб і помилок” у наші дні непридатний, дуже мало часу залишається для “проб” і дуже дорогими можуть бути помилки. В ринкових умовах не повинно бути місця свавільним, так званим “вольовим” рішенням. Стратегічні рішення повинні прийматися не інтуїтивно, а на підставі всебічного статистичного аналізу та математичних розрахунків. І не випадково, саме в наш час, відзначається активний ріст використання математичних методів в макро- та мікроекономічних дослідженнях. Замість того, щоб “пробувати і помилятися” на реальних об’єктах, аналітики роблять це з допомогою економіко-математичних моделей. Побудова таких моделей є однією з найважливіших задач прикладної математики.

Моделювання – процес побудови моделі, за допомогою якого вивчається функціонування об’єктів різної природи. Він складається з трьох основних елементів: суб’єкта, об’єкта дослідження та моделі, за допомогою якої суб’єкт пізнає об’єкт.

Модель – це такий матеріально або розумово зображуваний об’єкт, який у процесі дослідження замінює об’єкт-оригінал таким чином, що його безпосереднє вивчення дає нові знання про цей об’єкт.

Схематичне зображення процесу моделювання представлене на рис 1.1.

Математичне моделювання – універсальний та ефективний інструмент пізнання внутрішніх закономірностей, властивих явищам і процесам. Воно дає можливість вивчити кількісні взаємозв’язки і взаємозалежності системи, що вивчається, і вдосконалити її подальший розвиток та функціонування.

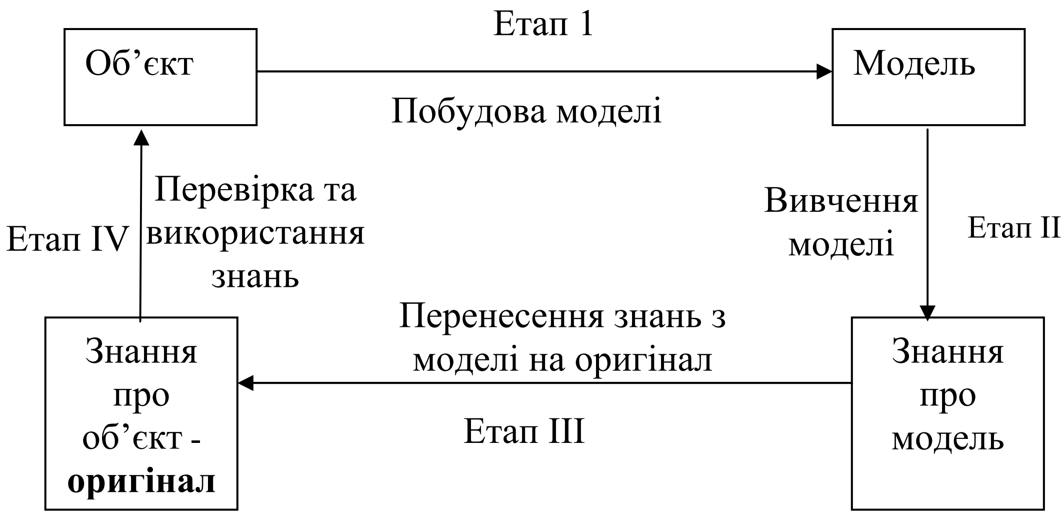


Рис. 1.1. Основні етапи моделювання

Серед існуючих систем економічні є найскладнішими, тому при побудові їх моделей слід відображати тільки найважливіші та найхарактерніші властивості процесів або явищ, що вивчаються. Внаслідок цього всі моделі є спрощеним відображенням реальної системи, але якщо цей процес виконано коректно, то отримане наближене відображення реальної ситуації дає можливість мати достатньо точні характеристики об'єкта дослідження.

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).

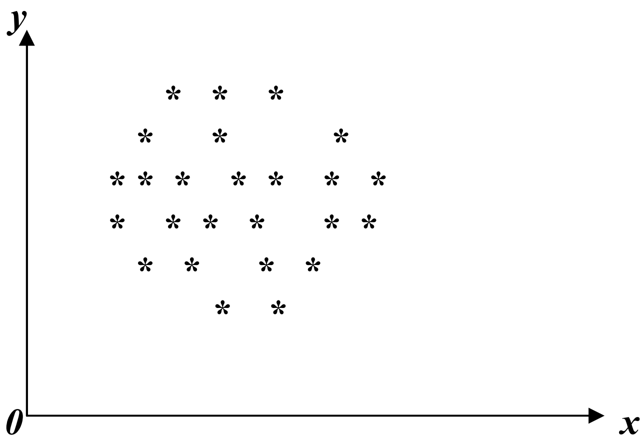
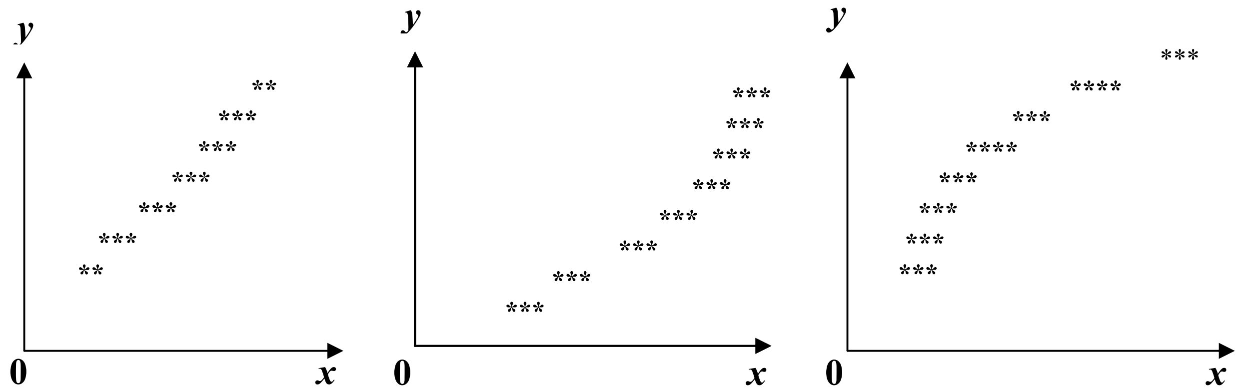


Рис. 1.2. Діаграма розсіювання у випадку відсутності зв'язку

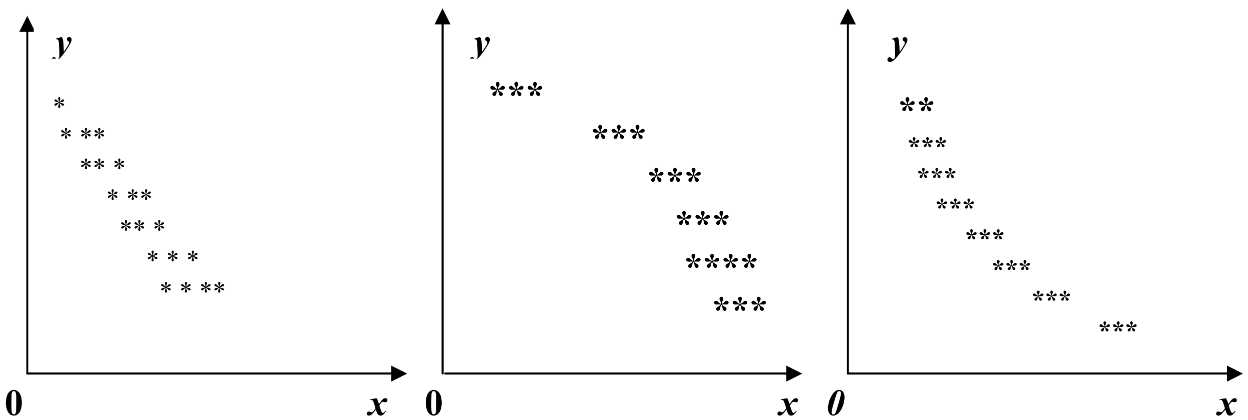
На рис.1.3 представлені основні форми залежностей.

Додатна регресія



а) б) в)

Від'ємна регресія



г) д) е)

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.

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:

$$y = \alpha + \beta x + u \qquad (1.1)$$

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

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

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

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.

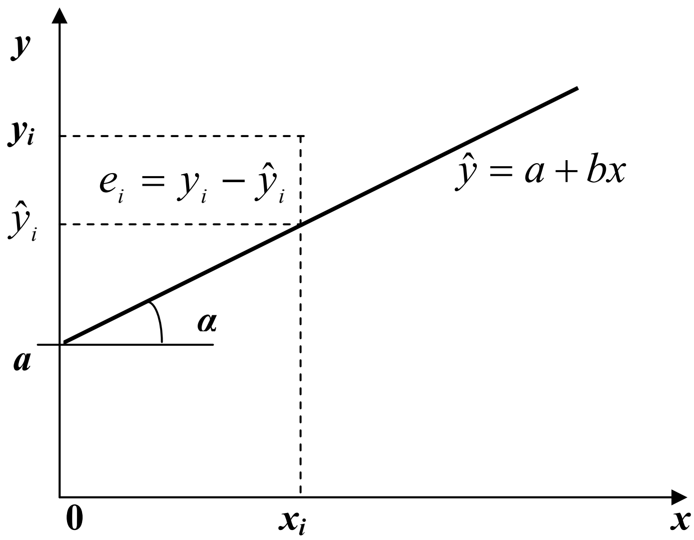


Рис. 1.4. Регресійна пряма та її параметри

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:

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

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:

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

$$\begin{cases}\sum\limits_{i=1}^{n}(y_i-a-bx_i)=0\\ \sum\limits_{i=1}^{n}(y_i-a-bx_i)x_i=0\end{cases}\Rightarrow\begin{cases}\sum\limits_{i=1}^{n}y_i-na-b\sum\limits_{i=1}^{n}x_i=0\\ \sum\limits_{i=1}^{n}x_iy_1-a\sum\limits_{i=1}^{n}x_i-b\sum\limits_{i=1}^{n}x_1^2=0\end{cases}$$

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

$$\begin{cases}na+b\sum\limits_{i=1}^{n}x_i=\sum\limits_{i=1}^{n}y_i\\ a\sum\limits_{i=1}^{n}x_i+b\sum\limits_{i=1}^{n}x_i^2=\sum\limits_{i=1}^{n}x_iy_i,\end{cases} \tag{1.4}$$

розв’язавши яку, знаходимо a та b:

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

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

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

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:

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

У результаті отримаємо

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

$$\bar{y} = a + b\bar{x}. \qquad (1.7)$$

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

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

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

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

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

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

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

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:

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

Приклад 1.1. Для десяти підприємств регіону за умовний деякий період відомі числові значення двох економічних показників: валова продукція y (млн. грн.) і вартість основних виробничих фондів x (млн. грн.), (табл.1.1). 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):

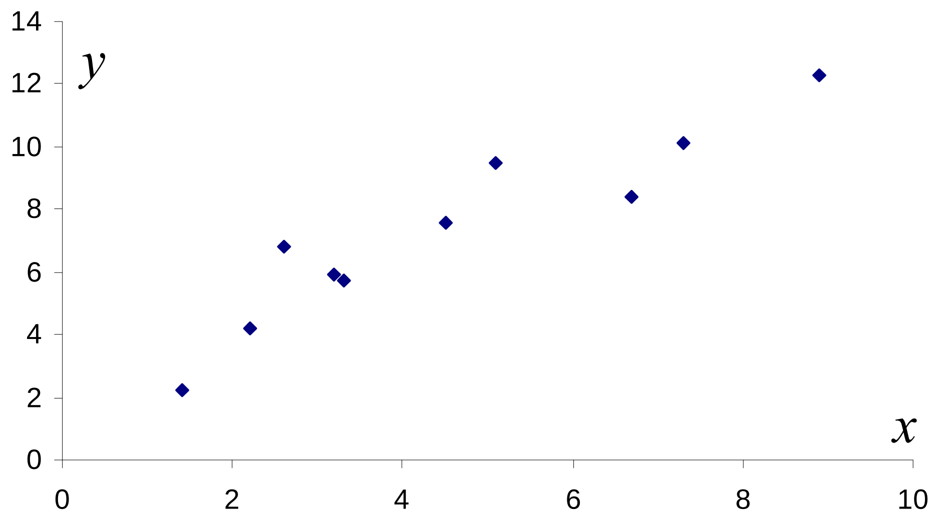


Рис.1.5. Діаграма розсіювання

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**

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

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

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):

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

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

Отже, нами отримано оціночне рівняння економетричної моделі

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

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

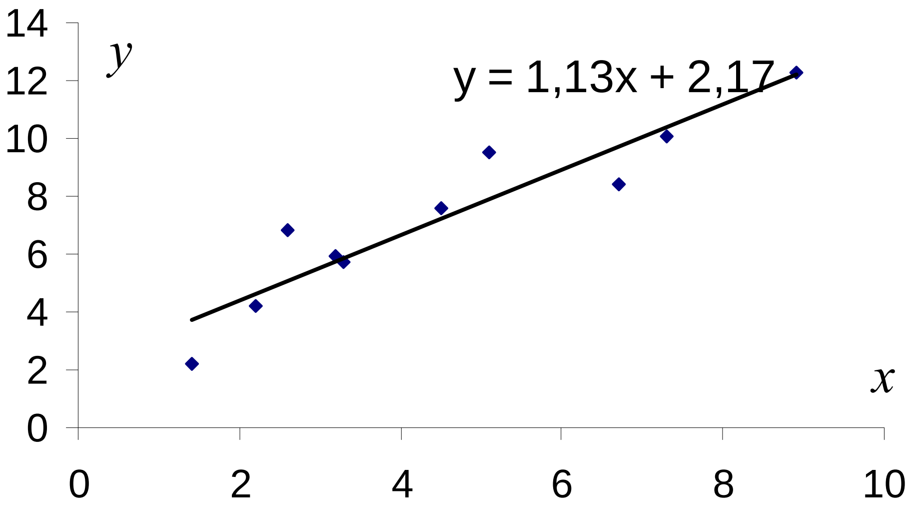


Рис.1.6. Діаграма розсіювання та оціночна пряма

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:

$$y_i = \hat{y}_i + e_i. \qquad (1.11)$$

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

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

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

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

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:

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

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

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

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

непояснююча дисперсія.

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

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:

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

Як можна побачити з виразу (1.17) перша частина $\frac{\text{var}(\hat{y})}{\text{var}(y)}$ є складовою дисперсії, яку можна пояснити через лінію регресії. Друга частина $\frac{\text{var}(e)}{\text{var}(y)}$ 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:

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

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 $\text{var}(\hat{y}) = \text{var}(y)$. 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:

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

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:

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

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

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: $r = \pm\sqrt{d}$. (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:

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

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:

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

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):

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

3. Відсутність систематичного зв’язку між значеннями випадкової змінної в будь-яких спостереженнях:

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

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:

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

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

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

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:

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

де k – довірчий множник, який відображає частку стандартного відхилення, яка повинна бути врахована, щоб із заданою ймовірністю p довірчий інтервал $\left[\mu-k\sigma_\mu;\mu+k\sigma_\mu\right]$ 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:

$\Delta_{yx} = t_p \cdot \sigma_{\text{непоясн.}}$, 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:

$$\hat{y}_i - \Delta_{yx} \le y_i \le \hat{y}_i + \Delta_{yx}. \tag{1.20}$$

Графічно довірчий інтервал можна зобразити таким чином:

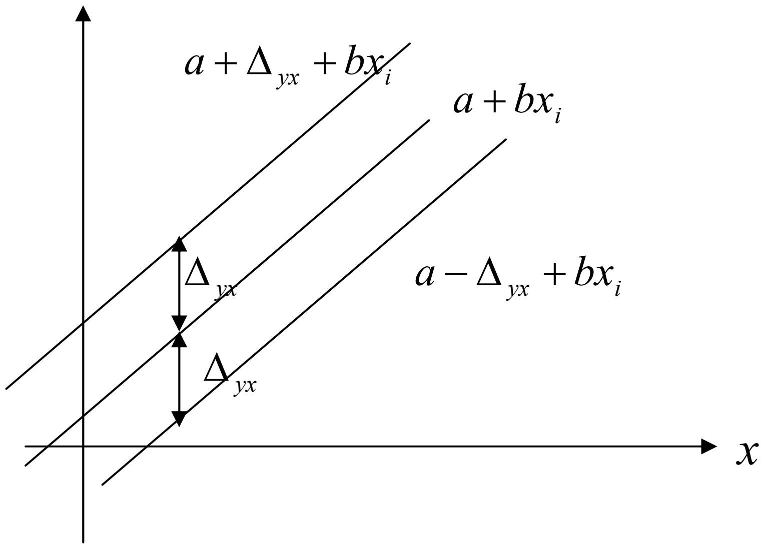


Рис. 1.7. Довірчий інтервал оціночного рівняння

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

$$\Delta_a = \sigma_a \cdot t_p; \quad \Delta_b = \Delta_b \cdot t_p \tag{1.21}$$

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

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

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

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

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

<u>Приклад 1.3.</u> 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 $\Delta_{yx} = t_p \cdot \sigma_{\text{непоясн.}}$. 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 $\sigma_{\text{непоясн.}} = \sqrt{\sigma^2_{\text{непоясн.}}} = \sqrt{1{,}001} = 1{,}0005$

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

Довірчий інтервал матиме вигляд:

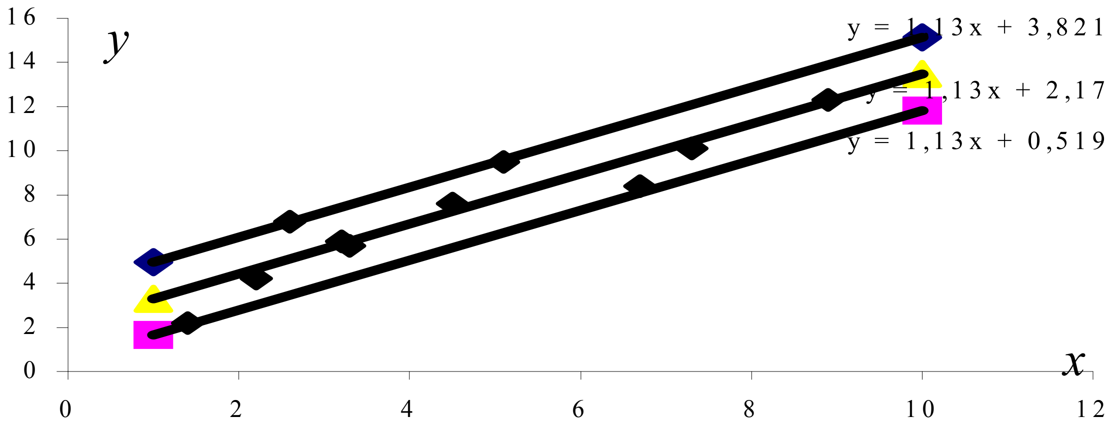
$$\hat{y}_i - 1{,}651 \le y_i \le \hat{y}_i + 1{,}651,$$
$$2{,}17 + 1{,}13x_i - 1{,}651 \le y_i \le 2{,}17 + 1{,}13x_i + 1{,}651,$$
$$1{,}13x_i + 0{,}519 \le y_i \le 1{,}13x_i + 3{,}821.$$

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:

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

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

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:

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

для оцінки β:

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

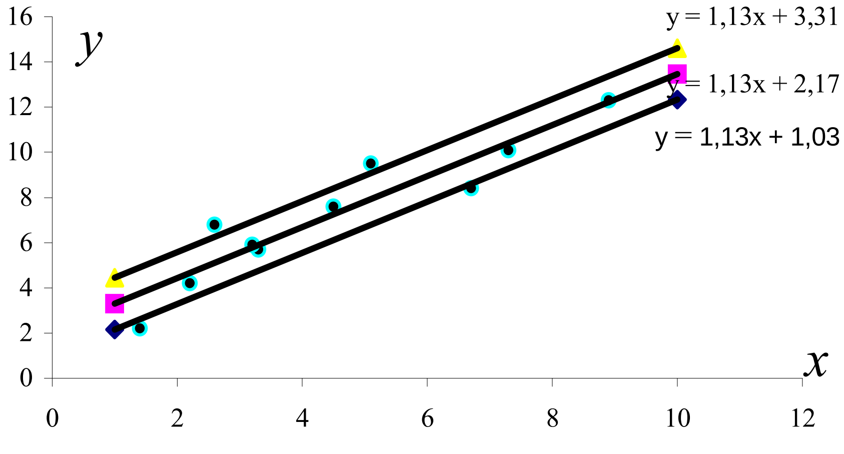
Відобразимо ці довірчі інтервали графічно. Обчислимо координати двох точок для побудови нижньої та верхньої меж довірчого інтервалу для α.

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:

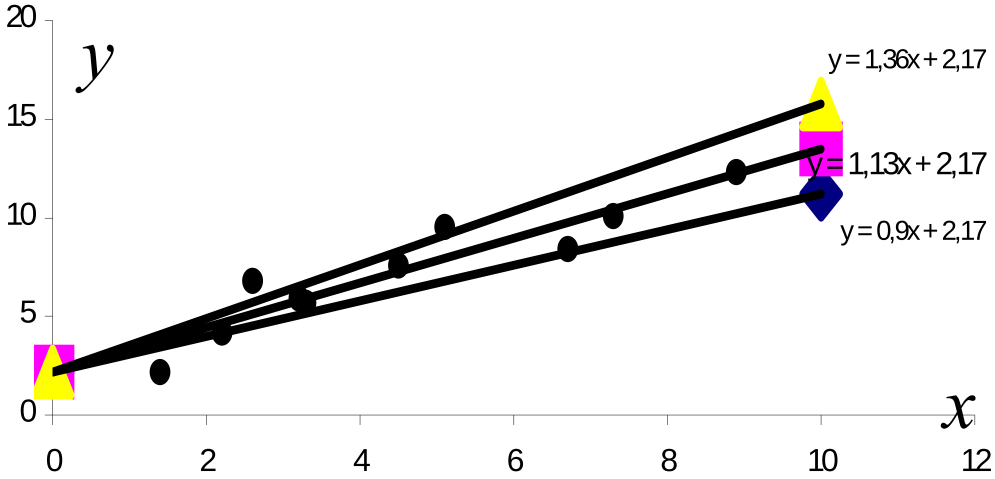


Нижня межа $y = 0{,}9x + 2{,}17$	
x	y
0	2,17
10	11,17

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_0: 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:

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

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 $|t_{емп}| > t_{кр}$, 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 $|t_{емп}| \le \mathrm{t}_{\text{кр}}$, 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 $t_{емп} > t_{кр}$, то нульова гіпотеза H_o із рівнем значущості α відкидається і приймається альтернативна гіпотеза H_1. Тоді відповідна оцінка вважається статистично значимою. Якщо ж $t_{емп} \le \text{t}_{\text{кр}}$, 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: $R_{ген}$≠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: β=0) проти альтернативної H_1: β≠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

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

$t_{емп}$>$t_{кр}$, 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:

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

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 β=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:

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

або

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

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

Приклад 1.5. 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 $k_1 = m = 1$, $k_2 = n - m - 1 = 10 - 1 - 1 = 8$:

$$F_{кр} = 5,32.$$

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

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

адекватне реальній дійсності і на його основі можна здійснювати прогнози.

Задача *Торгівельне підприємство має велику кількість філій і його керівництво вивчає питання про залежність* Y *(річний товарообіг однієї філії, млн. грн.) від* X *(торгівельної площі, тис. м*^{2}*). Для десяти філій за певний рік зафіксовані такі значення показників* Y *і* X*:*

i	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
y_i	1,5	2,9	3,1	3,2	4,3	5,7	5,8	7	7,2	7,5
x_i	0,2	0,3	0,5	0,6	0,8	1	1,1	1,2	1,3	1,4

На обсяг товарообігу впливають такі чинники: середньоденна інтенсивність потоку покупців, об'єм основних фондів, їх структура, середньоспискова чисельність працівників, площа підсобних приміщень тощо. Припускається, що в досліджуваній групі філій значення цих чинників приблизно однакові, тому вплив відмінностей їх значень на зміну обсягу товарообігу є незначним.

Вважаючи, що виконуються передумови 1-4*, потрібно:*

1) знайти статистичні оцінки параметрів лінійного рівняння регресії;

2) точкову оцінку та довірчий інтервал дисперсії збурень із надійністю $\gamma = 0{,}9$*;*

3) для рівня значущості $\alpha = 0{,}05$ *перевірити значущість коефіцієнтів регресії* α_0 *та* α_1*;*

4) знайти довірчі інтервали коефіцієнтів регресії з надійністю $\gamma = 0{,}95$*;*

5) знайти вибіркові коефіцієнт детермінації, коефіцієнт кореляції, а також інші показники якості лінійної регресії (MAPE, MPE);

6) знайти та побудувати довірчу зону функції регресії з надійністю $\gamma = 0{,}95$*;*

7) знайти прогнозне значення річного товарообігу для нової філії, торгівельна площа якої складає 1,8 тис. м^2*, а також із надійністю* $\gamma = 0{,}95$ *побудувати довірчий інтервал для цього прогнозного значення.*

○ **1)** Статистичні оцінки a_0, a_1 параметрів α_0 та α_1 лінійного рівняння регресії задовільняють системі нормальних рівнянь (2.12):

$$\begin{cases} a_0 + \overline{x} a_1 = \overline{y}, \\ \overline{x} a_0 + \overline{x^2} a_1 = \overline{xy}. \end{cases}$$

Для знаходження коефіцієнтів цієї системи складемо розрахункову табл. 2.1, останній стовпець якої потрібний для обчислення σ_y.

Таблиця 2.1

i	x_i	y_i	x_i^2	$x_i\ y_i$	y_i^2
1	0,2	1,5	0,04	0,3	2,25
2	0,3	2,9	0,09	0,87	8,41
3	0,5	3,1	0,25	1,55	9,61
4	0,6	3,2	0,36	1,92	10,24
5	0,8	4,3	0,64	3,44	18,49
6	1	5,7	1	5,7	32,49
7	1,1	5,8	1,21	6,38	33,64
8	1,2	7	1,44	8,4	49
9	1,3	7,2	1,69	9,36	51,84
10	1,4	7,5	1,96	10,5	56,25
Σ	8,4	48,2	8,68	48,42	272,22

Використовуючи нижній рядок табл. 2.1, отримаємо (обсяг вибірки $n=10$):

$$\bar{x}=\sum_{i=1}^{10}x_i/n=8{,}4/10=0{,}84; \qquad \bar{y}=\sum_{i=1}^{10}y_i/n=48{,}2/10=4{,}82;$$

$$\overline{x^2}=\sum_{i=1}^{10}x_i^2/n=8{,}68/10=0{,}868; \qquad \overline{xy}=\sum_{i=1}^{10}x_iy_i/n=48{,}42/10=4{,}842;$$

$$\overline{y^2}=\sum_{i=1}^{10}y_i^2/n=272{,}22/10=27{,}222;$$

$$\begin{cases} a_0+0{,}84a_1=4{,}82, \\ 0{,}84a_0+0{,}868a_1=4{,}842. \end{cases}$$

Єдиний розв'язок цієї системи рівнянь згідно із формулами (2.13):

$$a_1=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\overline{x^2}-\left(\bar{x}\right)^2}=\frac{4{,}842-0{,}84\cdot 4{,}82}{0{,}868-\left(0{,}84\right)^2}=\frac{0{,}7932}{0{,}1624}=4{,}884,$$

$$a_0=\bar{y}-a_1\bar{x}=4{,}82-0{,}84\cdot 4{,}884=0{,}717.$$

Отже, емпіричне рівняння регресії має такий вигляд:

$$\hat{y}=0{,}717+4{,}884x. \qquad (2.57)$$

2) Незміщену точкову оцінку S_u^2 невідомої дисперсії збурень σ_u^2 знайдемо за формулою (2.22):

$$S_u^2=\frac{1}{n-2}\sum_{i=1}^{n}u_i^2=\frac{1}{n-2}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2,$$

попередньо обчисливши $\hat{y}_i=0{,}717+4{,}884x_i$ та $u_i^2=\left(y_i-\hat{y}_i\right)^2$, $i=\overline{1{,}10}$, (табл. 2.2).

Таблиця 2.2

i	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Σ
y_i	1,5	2,9	3,1	3,2	4,3	5,7	5,8	7	7,2	7,5	–
$\hat{y}_i$	1,6938	2,1822	3,159	3,6474	4,6242	5,601	6,0894	6,5778	7,0662	7,5546	–
u_i	-0,1938	0,7178	-0,059	-0,4474	-0,3242	0,099	-0,2894	0,4222	0,1338	-0,0546	0,0044
u_i^2	0,0376	0,5152	0,0035	0,2002	0,1051	0,0098	0,0838	0,1783	0,0179	0,003	1,1544

Зауваження. *Згідно із (2.21):* $\overline{u} = 0$*, в той час як у нашому випадку* $\sum u_i / 10 = 0{,}00044$. *Цим значенням можна ігнорувати (вважати практично рівним нулю). Разом з тим з'ясуємо причину такого відхилення від нуля. Значення* a_1 *та* a_0 *з точністю до шести знаків після коми відповідно складають* $4{,}884236$ *та* $0{,}717242$*, тобто обидва ці значення (хай і несуттєво) більші тих, які взяті у моделі (2.57). Накопичення додатних похибок у різницях* $y_i - \hat{y}_i$ *і привело до того, що* $\sum(y_i - \hat{y}_i)$ *незначно перевищує нуль. Відмітимо також, що значення* $a_1 = 4{,}88$*,* $a_0 = 0{,}72$ *приводять до рівності* $\sum u_i = 0{,}008$.

Використавши підсумок останнього рядка, отримаємо:

$$S_u^2 = \frac{1}{10-2} \cdot 1{,}1544 = 0{,}1443.$$

Ліва і права межі довірчого інтервалу для σ_u^2 визначаються згідно (2.29) за формулами відповідно $\dfrac{(n-2)S_u^2}{\chi_2^2}$ і $\dfrac{(n-2)S_u^2}{\chi_1^2}$, де у відповідності із (2.26) та (2.27) χ_1^2 та χ_2^2 є коренями рівнянь

$$P\left(\chi^2(k) > \chi_1^2(p;k)\right) = p, \quad p = (1+\gamma)/2 = 0{,}95,$$
$$P\left(\chi^2(k) > \chi_2^2(p;k)\right) = p, \quad p = (1-\gamma)/2 = 0{,}05.$$

За табл. 4 додатків для $k = n - 2 = 8$ знайдемо: $\chi_1^2(0{,}95;8) = 2{,}73$ і $\chi_2^2(0{,}05;8) = 15{,}51$. Тоді ліва межа довірчого інтервалу дорівнює $\dfrac{8 \cdot 0{,}1443}{15{,}51} = 0{,}0744$, а права – $\dfrac{8 \cdot 0{,}1443}{2{,}73} = 0{,}4229$. Тобто остаточно з надійністю 0,9

$$0{,}0744 < \sigma_u^2 < 0{,}4229.$$

3) Згідно з п.8, якщо виконується нерівність (2.33): $\left|\dfrac{a_m}{S_{a_m}}\right| > t_{кр.}$ $(m=0, m=1)$, тоді на рівні значущості α приймається гіпотеза $H_1 : a_m \neq 0$. Значення S_{a_0} та S_{a_1} знайдемо із виразів (2.30):

Знайдемо дисперсію *x*:

$$\sigma_x{}^2 = \overline{x^2} - \left(\overline{x}\right)^2 = 0,868 - \left(0,84\right)^2 = 0,162$$

Звідки

$$\sigma_x = \sqrt{\sigma_x{}^2} = \sqrt{0,162} = 0,403\,.$$

Отже,

$$S_{a_0} = \sqrt{\frac{S_u^2\overline{x^2}}{n\sigma_x^2}} = \frac{S_u}{\sigma_x}\sqrt{\frac{\overline{x^2}}{n}} = \frac{0,1443}{0,403}\sqrt{\frac{\overline{x^2}}{10}} = 0,1443\sqrt{\frac{0,868}{10}} = 0,1055\,;$$

$$S_{a_1} = \sqrt{\frac{S_u^2}{n\sigma_x^2}} = \frac{S_u}{\sigma_x\sqrt{n}} = \frac{0,1443}{0,403\sqrt{10}} = 0,1132\,.$$

Тоді спостережені значення критерію:

$$\left|\frac{a_0}{S_{a_0}}\right| = \frac{0,717}{0,1055} = 6,796\,,\quad \left|\frac{a_1}{S_{a_1}}\right| = \frac{4,884}{0,1132} = 43,145\,.$$

Критична точка для двосторонньої критичної області $t_{кр.} = t_{двост.}(\alpha, k)$ при значеннях $\alpha = 0,05$, $k = n - 2 = 8$ знаходиться за верхньою частиною табл. 3 додатків: $t_{кр.} = 2,306$.

Оскільки $6,796 > t_{кр.} = 2,306$ і $43,145 > t_{кр.} = 2,306$, то на рівні значущості $\alpha = 0,05$ робимо висновки, що $\alpha_0 \neq 0$ і $\alpha_1 \neq 0$.

4) Згідно з (2.39) та (2.40) довірчі інтервали з надійністю γ для невідомих параметрів регресії a_0 та a_1 мають такий вигляд:

$$a_m - t_m(\gamma, k)S_{a_m} < \alpha_m < a_m + t_m(\gamma, k)S_{a_m}\,,$$

де $m = 0,1$, $t_m = t_m(\gamma, k)$ – корінь рівняння $P(|t_m| < t) = \gamma$, t_0 та t_1 – випадкові величини, розподілені за законом Ст'юдента.

У нашому випадку $\gamma = 0{,}95$, число ступенів вільності $k = n - 2 = 8$. За табл. 2 додатків знаходимо $t_0(0{,}95;8) = t_1(0{,}95;8) = 2{,}306$. Тоді з врахуванням знайдених значень $S_{a_0} = 0{,}1055$, $S_{a_1} = 0{,}1132$ отримаємо:

$$0{,}717 - 2{,}306 \cdot 0{,}1055 < \alpha_0 < 0{,}717 + 2{,}306 \cdot 0{,}1055,$$

$$4{,}884 - 2{,}306 \cdot 0{,}1132 < \alpha_1 < 4{,}884 + 2{,}306 \cdot 0{,}1132$$

або остаточно

$$0{,}4737 < \alpha_0 < 0{,}9603,$$

$$4{,}6236 < \alpha_1 < 5{,}145.$$

5) Коефіцієнт детермінації R^2 знайдемо за формулою (2.46*):

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

Із табл. 2.2 (останнє число нижнього рядка) $\sum_{i=1}^{10} u_i^2 = \sum_{i=1}^{10}\left(y_i - \hat{y}_i\right)^2 = 1{,}1544$.

Для знаходження $\sigma_y^2 = \overline{y^2} - \left(\overline{y}\right)^2$ використаємо табл. 2.1:

$$\sigma_y = \sqrt{\overline{y^2} - \left(\overline{y}\right)^2} = \sqrt{272{,}22/10 - \left(48{,}2/10\right)^2} = 1{,}9974.$$

Отже,

$$R^2 = 1 - \frac{1{,}1544}{10 \cdot 1{,}9974} = 0{,}9422.$$

Таким чином, варіація залежної змінної Y на 94,22% пояснюється варіацією пояснюючої змінної.

Вибірковий коефіцієнт кореляції згідно із (2.46):

$$r = \sqrt{R^2} = \sqrt{0{,}9422} = 0{,}9707.$$

При цьому додатний знак цього числа обрано в зв'язку з тим, що $a_1 > 0$.

Обчислимо абсолютну середню відсоткову помилку MAPE за формулою (2.49):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \cdot 100\%.$$

Для цього використаємо другий і четвертий рядки табл. 2.2:

$$\sum_{i=1}^{10}\left|\frac{\hat{y}_i - y_i}{y_i}\right| = \frac{0{,}1938}{1{,}5} + \frac{0{,}7178}{2{,}9} + \frac{0{,}059}{3{,}1} + \frac{0{,}4474}{3{,}2} + \frac{0{,}3242}{4{,}3} + \frac{0{,}099}{5{,}7} +$$

$$+\frac{0,2894}{5,8}+\frac{0,4222}{7}+\frac{0,1338}{7,2}+\frac{0,0546}{7,5}=0,1292+0,2475+0,019+0,1398+$$
$$+0,0754+0,0174+0,0499+0,0603+0,0186+0,0073=0,7644.$$

Отже, $MAPE=\frac{1}{10}\cdot 0,7644\cdot 100\%=7,644\%<10\%$, тобто відповідає високій точності прогнозу за моделлю.

Середню відсоткову помилку MPE знайдемо за формулою (2.50):

$$MPE=\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{y}_i-y_i}{y_i}\cdot 100\%,$$

використавши розрахунки при обчисленні MAPE:

$$\sum_{i=1}^{10}\frac{\hat{y}_i-y_i}{y_i}=0,1292-0,2475+0,019+0,1398++0,0754-0,0174+$$
$$+0,0499-0,0603-0,0186+0,0073=0,0768.$$

Остаточно

$$MPE=\frac{1}{10}\cdot 0,0768\cdot 100\%=0,768\%<5\%.$$

***<u>Висновок</u>**: всі знайдені показники вказують на високу якість моделі.*

6) Побудова довірчої зони для функції регресії передбачає побудову точок з координатами $\{x_i;\hat{y}_i-t(\gamma,n-2)S_{\hat{y}_i}\}$, $i=\overline{1,n}$, з наступним з'єднанням сусідніх (по індексу i) точок прямолінійними відрізками, а потім здійснення аналогічної процедури для послідовності точок $\{x_i;\hat{y}_i+t(\gamma,n-2)S_{\hat{y}_i}\}$.

Величину $S_{\hat{y}_i}$ знайдемо із формули (2.36) $S_{\hat{y}_i}=S_u\sqrt{\left[1+\frac{(x_i-\bar{x})^2}{\sigma_x^2}\right]\frac{1}{n}}$.

Використовуючи табл. 2.1 і знайдене значення $S_u=\sqrt{0,1443}=0,3799$, отримаємо:

$$\sigma_x^2=\overline{x^2}-(\bar{x})^2=0,1624;$$

$$S_{\hat{y}_1}=0,3799\sqrt{\left[1+\frac{(0,2-0,84)^2}{0,1624}\right]\frac{1}{10}}=0,2255;$$

$$S_{\hat{y}_2}=0,3799\sqrt{\left[1+\frac{(0,3-0,84)^2}{0,1624}\right]\frac{1}{10}}=0,201;$$

$S_{\hat{y}_3} = 0{,}1572$; $S_{\hat{y}_4} = 0{,}1398$; $S_{\hat{y}_5} = 0{,}1207$; $S_{\hat{y}_6} = 0{,}1293$;

$S_{\hat{y}_7} = 0{,}143$; $S_{\hat{y}_8} = 0{,}1611$; $S_{\hat{y}_9} = 0{,}1823$; $S_{\hat{y}_{10}} = 0{,}2057$.

За табл. 2 додатків знайдемо $t(0{,}95;8) = 2{,}306$. Використовуючи табл. 2.2 і знайдені $S_{\hat{y}_i}$, отримаємо ординати точок нижньої межі довірчої зони:

$y_1 - tS_{\hat{y}_1} = 1{,}6938 - 2{,}306 \cdot 0{,}2255 = 1{,}1738$;

$y_2 - tS_{\hat{y}_2} = 2{,}1822 - 2{,}306 \cdot 0{,}201 = 1{,}7187$;

$y_3 - tS_{\hat{y}_3} = 3{,}159 - 2{,}306 \cdot 0{,}1572 = 2{,}7965$;

$y_4 - tS_{\hat{y}_4} = 3{,}6474 - 2{,}306 \cdot 0{,}1394 = 3{,}3259$;

$y_5 - tS_{\hat{y}_5} = 4{,}6242 - 2{,}306 \cdot 0{,}1207 = 4{,}3459$

$y_6 - tS_{\hat{y}_6} = 5{,}601 - 2{,}306 \cdot 0{,}1293 = 5{,}3028$;

$y_7 - tS_{\hat{y}_7} = 6{,}0894 - 2{,}306 \cdot 0{,}143 = 5{,}7596$;

$y_8 - tS_{\hat{y}_8} = 6{,}5778 - 2{,}306 \cdot 0{,}1611 = 6{,}2063$;

$y_9 - tS_{\hat{y}_9} = 7{,}0662 - 2{,}306 \cdot 0{,}1823 = 6{,}6458$;

$y_{10} - tS_{\hat{y}_{10}} = 7{,}5546 - 2{,}306 \cdot 0{,}2057 = 7{,}0803$.

Тоді ординати точок верхньої межі довірчої зони набирають таких значень:

$\hat{y}_1 + tS_{\hat{y}_1} = 1{,}6938 + 2{,}306 \cdot 0{,}2255 = 2{,}2139$;

$\hat{y}_2 + tS_{\hat{y}_2} = 2{,}1822 + 2{,}306 \cdot 0{,}201 = 2{,}6457$;

$\hat{y}_3 + tS_{\hat{y}_3} = 3{,}159 + 2{,}306 \cdot 0{,}1572 = 3{,}5215$;

$\hat{y}_4 + tS_{\hat{y}_4} = 3{,}6474 + 2{,}306 \cdot 0{,}1394 = 3{,}9689$;

$\hat{y}_5 + tS_{\hat{y}_5} = 4{,}6242 + 2{,}306 \cdot 0{,}1207 = 4{,}9025$

$\hat{y}_6 + tS_{\hat{y}_6} = 5{,}601 + 2{,}306 \cdot 0{,}1293 = 5{,}8992$;

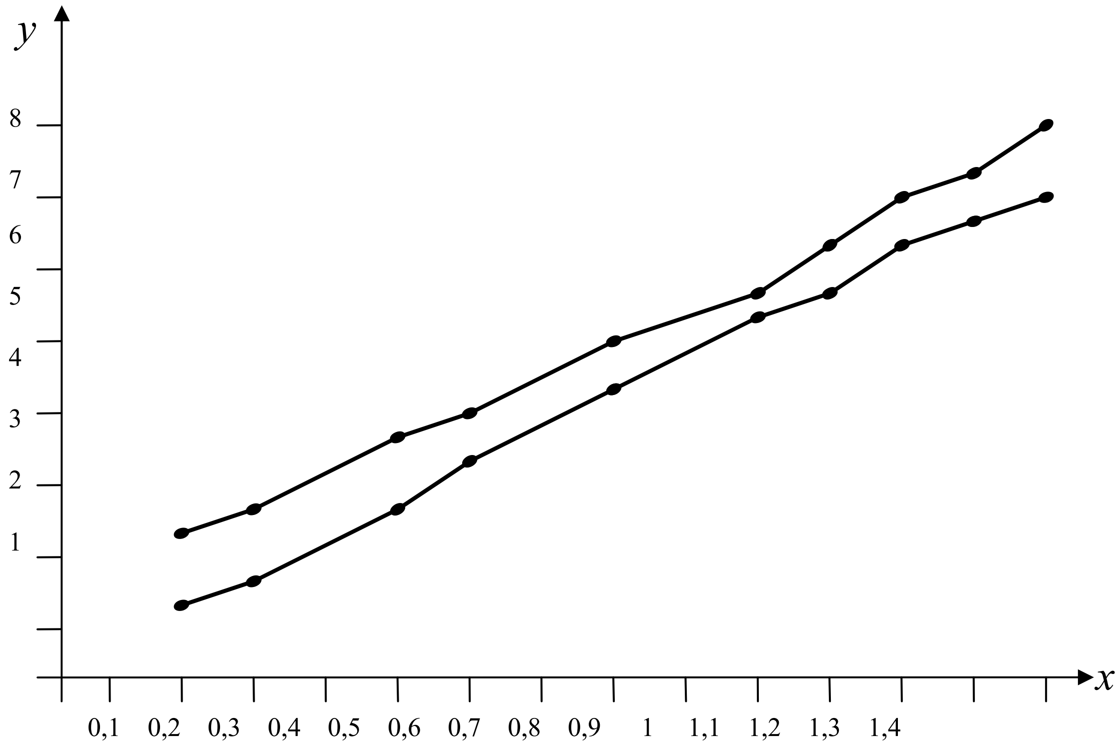
$\hat{y}_7 + tS_{\hat{y}_7} = 6{,}0894 + 2{,}306 \cdot 0{,}143 = 6{,}4192$;

$\hat{y}_8 + tS_{\hat{y}_8} = 6{,}5778 + 2{,}306 \cdot 0{,}1611 = 6{,}9493$;

$\hat{y}_9 + tS_{\hat{y}_9} = 7{,}0662 + 2{,}306 \cdot 0{,}1823 = 7{,}4866$;

$\hat{y}_{10} + tS_{\hat{y}_{10}} = 7{,}5546 + 2{,}306 \cdot 0{,}2057 = 8{,}1263$.

Довірча зона (з надійністю 0,95) для функції регресії зображена на мал. 2.2.



Малюнок 2.2.

7) Прогнозне значення річного товарообігу для нової філії із торгівельною площею 1,8 тис. м^2 знайдемо із рівняння (2.57):

$$\hat{y}_{n+1} = 0{,}717 + 4{,}884 \cdot 1{,}8 = 9{,}508.$$

Довірчий інтервал для прогнозного значення y_{n+1} із надійністю $\gamma = 0{,}95$ визначається (2.56). З допомогою виразу (2.55) знайдемо

$$S_{u_{n+1}} = S_u \sqrt{1 + \frac{1}{n} + \frac{\left(x_{n+1} - \bar{x}\right)^2}{n\sigma_x^2}} = 0{,}3799 \sqrt{1 + \frac{1}{10} + \frac{\left(1{,}8 - 0{,}84\right)^2}{10 \cdot 0{,}1624}} = 0{,}4906.$$

Тоді шуканий довірчий інтервал має вид

$$9{,}508 - 2{,}306 \cdot 0{,}4906 < y_{n+1} < 9{,}508 + 2{,}306 \cdot 0{,}4906$$

або остаточно

$$8{,}3767 < y_{n+1} < 10{,}6393.$$

ДОДАТКИ

Таблиця 1

Значення функції Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

x	Соті долі x									
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37285	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41308	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	44352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49835	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	49977	49978	49978	49979	49980	49981	48881	49982	49983	49983
3,6	49984	49985	40085	49986	49986	49987	49987	49988	49988	49989
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997

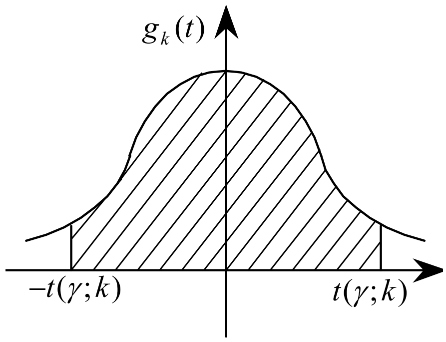
x	Десяті долі x				
	0	*2*	*4*	*6*	*8*
4,	0,4999683	4999867	4999946	4999979	4999992
5,	4999997				

Таблиця 2

Значення $t = t(\gamma; k)$,
що задовільняють рівнянню

$$P\left(\left| T \right| < t \right) = 2\int_0^t g_k(t)dt = \gamma,$$

де $g_k(t)$ – густина **розподілу Ст'юдента** (***t*-розподілу**), $k = n - 1$ – число ступенів вільності



$k = n - 1$	γ				
	0,9	*0,95*	*0,98*	*0,99*	*0,999*
1	6,314	12,706	31,821	63,657	636,619
2	2,920	4,303	6,965	9,925	31,599
3	2,353	3,182	4,541	5,841	12,924
4	2,132	2,776	3,747	4,604	8,610
5	2,015	2,571	3,365	4,032	6,869
6	1,943	2,447	3,143	3,707	5,969
7	1,895	2,365	2,998	3,499	5,408
8	1,860	2,306	2,896	3,355	5,041
9	1,833	2,262	2,821	3,250	4,781
10	1,812	2,228	2,764	3,169	4,587
11	1,796	2,201	2,718	3,106	4,437
12	1,782	2,179	2,681	3,055	4,318
13	1,771	2,160	2,650	3,012	4,221
14	1,761	2,145	2,624	2,977	4,140
15	1,753	2,131	2,602	2,947	4,073
16	1,746	2,120	2,583	2,921	4,015
17	1,740	2,110	2,567	2,898	3,965
18	1,734	2,101	2,552	2,878	3,922
19	1,729	2,093	2,539	2,861	3,883
20	1,725	2,086	2,528	2,845	3,850
25	1,708	2,060	2,485	2,785	3,725
30	1,697	2,042	2,457	2,750	3,646
40	1,684	2,021	2,423	2,704	3,551
50	1,676	2,009	2,403	2,678	3,496
60	1,671	2,000	2,390	2,660	3,460
70	1,667	1,994	2,381	2,648	3,435
80	1,664	1,990	2,374	2,639	3,416
90	1,662	1,987	2,368	2,632	3,402
100	1,660	1,984	2,364	2,626	3,390
120	1,658	1,980	2,358	2,617	3,373
∞	1,645	1,960	2,326	2,576	3,291

Таблиця 3

Критичні точки розподілу Ст'юдента (*t*-розподілу)

Для двосторонньої критичної області критична точка $t_{двост.кр}(\alpha;k)=t_\alpha$ є коренем рівняння $\int_0^{t_\alpha} g_k(t)dt=(1-\alpha)/2$; для односторонньої (правосторонньої) критичної області точка $t_{правост.кр}(\alpha;k)=t_{2\alpha}$ є коренем рівняння $\int_{t_{2\alpha}}^{\infty} g_k(t)dt=(1-2\alpha)/2$, де $g_k(t)$ – густина розподілу Ст'юдента, $k=n-1$ – число ступенів вільності. Для лівосторонньої критичної області $t_{лівост.кр}(\alpha;k)=-t_{2\alpha}$.

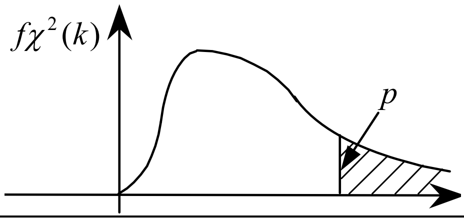
Число ступенів вільності $k=n-1$	Рівень значущості α (**двостороння** критична область)					
	0,10	*0,05*	*0,02*	*0,01*	*0,002*	*0,001*
1	6,314	12,71	31,82	63,66	318,3	637,0
2	2,920	4,303	6,965	9,925	22,33	31,60
3	2,353	3,182	4,541	5,841	10,22	12,94
4	2,132	2,776	3,747	4,604	7,173	8,610
5	2,015	2,571	3,365	4,032	5,893	6,859
6	1,943	2,447	3,143	3,707	5,208	5,959
7	1,895	2,365	2,998	3,499	4,785	5,405
8	1,860	2,306	2,896	3,355	4,501	5,041
9	1,833	2,262	2,821	3,250	4,297	4,781
10	1,812	2,228	2,764	3,169	4,144	4,587
11	1,796	2,201	2,718	3,106	4,025	4,437
12	1,782	2,179	2,681	3,055	3,930	4,318
13	1,771	2,160	2,650	3,012	3,852	4,221
14	1,761	2,145	2,624	2,977	3,787	4,140
15	1,753	2,131	2,602	2,947	3,733	4,073
16	1,746	2,120	2,583	2,921	3,686	4,015
17	1,740	2,110	2,567	2,898	3,646	3,965
18	1,734	2,101	2,552	2,878	3,611	3,922
19	1,729	2,093	2,539	2,861	3,579	3,883
20	1,725	2,086	2,528	2,845	3,562	3,850
21	1,721	2,080	2,518	2,831	3,527	3,819
22	1,717	2,074	2,508	2,819	3,505	3,792
23	1,714	2,069	2,500	2,807	3,485	3,767
24	1,711	2,064	2,492	2,797	3,467	3,745
25	1,708	2,060	2,485	2,787	3,450	3,725
26	1,706	2,056	2,479	2,779	3,435	3,707
27	1,703	2,052	2,473	2,771	3,421	3,690
28	1,701	2,048	2,467	2,763	3,408	3,674
29	1,699	2,045	2,462	2,756	3,396	3,659
30	1,697	2,042	2,457	2,750	3,385	3,646
40	1,684	2,021	2,423	2,704	3,307	3,551
60	1,671	2,000	2,390	2,660	3,232	3,460
120	1,658	1,981	2,362	2,624	3,172	3,374
∞	1,645	1,960	2,326	2,576	3,090	3,291
Число ступенів вільності $k=n-1$	*0,05*	*0,025*	*0,01*	*0,005*	*0,001*	*0,0005*
	Рівень значущості α (**одностороння** критична область)					

Таблиця 4

Значення

$$P\left(\chi^2(k) > \chi^2(p;k)\right) = p,$$

де k – число ступенів вільності



k	p							
	0,999	*0,99*	*0,95*	*0,90*	*0,10*	*0,05*	*0,01*	*0,001*
1	$0{,}157 \cdot 10^{-5}$	0,0002	0,004	0,02	2,71	3,84	6,63	10,83
2	0,002	0,02	0,10	0,21	4,61	5,99	9,21	13,82
3	0,02	0,12	0,35	0,58	6,25	7,82	11,34	16,27
4	0,09	0,30	0,71	1,06	7,78	9,49	13,28	18,47
5	0,21	0,55	1,15	1,61	9,24	11,07	15,08	20,51
6	0,38	0,87	1,64	2,20	10,65	12,59	16,81	22,46
7	0,60	1,24	2,17	2,83	12,02	14,06	18,48	24,32
8	0,86	1,65	2,73	3,49	13,36	15,51	20,09	26,12
9	1,15	2,09	3,33	4,17	14,68	16,92	21,67	27,88
10	1,48	2,56	3,94	4,87	15,99	18,31	23,21	29,59
11	1,83	3,05	4,58	5,58	17,28	19,68	24,72	31,26
12	2,21	3,57	5,23	6,30	18,55	21,03	26,22	32,91
13	2,62	4,11	5,89	7,04	19,81	22,36	27,69	34,53
14	3,04	4,66	6,57	7,79	21,06	23,69	29,14	36,12
15	3,48	5,23	7,26	8,55	22,31	25,00	30,58	37,70
16	3,94	5,81	7,96	9,31	23,54	26,30	32,00	39,25
17	4,42	6,41	8,67	10,09	24,77	27,59	33,41	40,79
18	4,90	7,02	9,39	10,86	25,99	28,87	34,81	42,31
19	5,41	7,63	10,12	11,65	27,20	30,14	36,19	43,82
20	5,92	8,26	10,85	12,44	28,41	31,41	37,57	45,31
21	6,45	8,90	11,59	13,24	29,62	32,67	38,93	46,80
22	6,98	9,54	12,34	14,04	30,81	33,92	40,29	48,27
23	7,53	10,20	13,20	14,85	32,01	35,17	41,64	49,73
24	8,08	10,86	13,85	15,66	33,19	36,42	43,98	51,18
25	8,65	11,52	14,61	16,47	34,38	37,65	44,31	52,62
26	9,22	12,20	15,37	17,29	35,56	38,89	45,64	54,05
27	9,80	12,88	16,15	18,11	36,74	40,11	46,96	55,48
28	10,39	13,56	16,93	18,94	37,92	41,34	48,28	56,89
29	10,99	14,26	17,71	19,77	39,09	42,56	49,59	58,30
30	11,59	14,95	18,49	20,60	40,26	43,77	50,89	59,70
40	17,92	22,16	26,51	29,05	51,81	55,76	63,69	73,40
50	24,67	29,71	34,76	37,69	63,17	67,51	76,15	86,66
100	61,92	70,07	77,93	82,36	118,50	124,34	135,81	149,45

Таблиця 5

Критичні точки $F_{кр}(\alpha; k_1, k_2)$ розподілу Фішера-Снедокора,

що задовільняють рівнянню $P\lfloor F > F_{кр}(\alpha; k_1, k_2)\rfloor = \alpha$ при $\alpha = 0{,}05$

k_2	k_1									
	1	*2*	*3*	*4*	*5*	*6*	*8*	*12*	*24*	∞
1	161,45	199,50	215,71	224,58	230,16	233,99	238,88	243,91	249,05	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3.52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
25	4,24	3,38	2,99	2,76	2,60	2.49	2,34	2.16	1,96	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00