МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВНУТРІШНІХ СПРАВ
КРЕМЕНЧУЦЬКИЙ ЛЬОТНИЙ КОЛЕДЖ

Циклова комісія природничих дисциплін

ТЕКСТ ЛЕКЦІЇ

з навчальної дисципліни «Економіко-математичне моделювання»
обов’язкових компонент освітньо-професійної програми
першого (бакалаврського) рівня вищої освіти
Логістика

за темою − Принципи побудови економетричних моделей.
Парна лінійна регресія

Харків 2021

ЗАТВЕРДЖЕНО
Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від _23.09.2021__ № _8_

СХВАЛЕНО
Методичною радою Кременчуцького
льотного коледжу Харківського
національного університету
внутрішніх справ
Протокол від _22.09.2021__ № _2_

ПОГОДЖЕНО
Секцією Науково-методичної ради
ХНУВС з технічних дисциплін
Протокол від _22.09.2021__ № _8_

Розглянуто на засіданні циклової комісії природничих дисциплін, протокол від _10.09.2021__ № _2_

Розробник: викладач циклової комісії природничих дисциплін, спеціаліст першої категорії Подгорних Н.В.

Рецензенти:
1. Завідувач відділення фахової підготовки навчального відділу КЛК ХНУВС, старший викладач циклової комісії технічного обслуговування авіаційної техніки КЛК ХНУВС, к.т.н., спеціаліст вищої категорії, викладач-методист Владов С. І.
2. Завідувач кафедри інформатика та вищої математики Кременчуцького національного університету імені Михайла Остроградського, д. т. н., професор Ляшенко В. П.

План лекції

1. Економетрична модель. Причинні взаємозв'язки між змінними величинами. Класифікація змінних величин в економетричних моделях.
2. Особливості побудови економетричних моделей.
3. Лінійна парна регресія.
4. Основні положення регресійного аналізу. Оцінка параметрів регресійної моделі. Теорема Гаусса – Маркова.
5. Інтервальна оцінка функції регресії та її параметрів. Оцінка значущості рівняння регресії. Коефіцієнт детермінації.
6. Алгоритм побудови економетричної моделі та оцінка її достовірності.

Рекомендована література:

Основна

1. Економіко-математичне моделювання : навчальний посібник / за ред. О. Т. Іващука. – Тернопіль : ТНЕУ «Економічна думка», 2008. – 704 с.
2. Наконечний С. І., Савіна С. С. Математичне програмування: Навч. посіб. — К.: КНЕУ, 2003. - 452 с.
3. Вітлинський В. В. Математичне програмування : навчально-методичний посібник для самостійного вивчення дисципліни / В. В. Вітлинський, С. І. Наконечний, Т. О. Терещенко. – К. : КНЕУ, 2001. – 248 с.

Додаткова

4. Ржевський С. В. Дослідження операцій : підручник для ВНЗ / С. В. Ржевський, В. М. Александрова. – К. : Академвидав, 2006. – 560 с.
5. Кулян В. Р. Математичне програмування : навчальний посібник / В. Р. Кулян, О. О. Юнькова, О. Е. Жильцов. – К. :МАУП, 2006. – 184 с.
6. Лещинський О.Л. Економетрія / О.Л. Лещинський, В.В. Рязанцева, О.О. Юнькова. – К. : «Вид. дім «Персонал»», 2008. – 208 с.
7. Дослідження операцій у середовищі електронних таблиць Excel : навчальний посібник / [Леснікова І. Ю., Халіпова Н. В. та ін.]. – К. : Центр учбової літератури, 2007. – 186 с.
8. Розв'язування оптимізаційних задач за допомогою лінійного програмування: Навчальний посібник/ За ред. М.І.Бєлікова; А.М.Гуржія, В.Р.Кігеля, В.В.Самсонова. - К.: ІСДО, 1994. - 132 с.
9. Ульянченко О.В. Дослідження операцій в економіці: Підручник/ О.В. Ульянченко. - Х.: Гриф, 2002. - 580 с.

Текст лекції

1. Економетрична модель. Причинні взаємозв’язки між змінними величинами. Класифікація змінних величин в економетричних моделях.

Економетрія – це наука, що досліджує кількісні закономірності й взаємозалежності в економіці за допомогою методів математичної статистики. Основа цих методів – кореляційно-регресійний аналіз.

Методологія економетричного моделювання тісно пов'язана з використанням системного підходу. Соціально-економічна система може бути подана нескінченним числом структурних і функціональних інваріантів, що відображають взаємозв'язки між різними процесами, що відбуваються у цій системі (економічними, соціальними, екологічними, демографічними й т.д.). Опис системи здійснюється за допомогою її якісних і кількісних характеристик, що називаються *параметрами*. Параметри становлять основу мов опису систем, а при формалізації ототожнюються з незалежними змінними математичного опису процесу функціонування систем. При побудові економетричної моделі реалізується метод моделювання за принципом «чорного ящика», коли дослідникові невідомий механізм процесів, що відбуваються в системі, вивчити який можна за вхідними і вихідними характеристиками системи. Вхідні та вихідні характеристики системи часто ототожнюють із екзогенними й ендогенними змінними або в кореляційно - регресійному аналізі вживають терміни «незалежні (факторні) змінні» або ознаки й «залежні (результативні) змінні» або ознаки. Графічно принцип «чорного ящика» зображений на рис. 1.

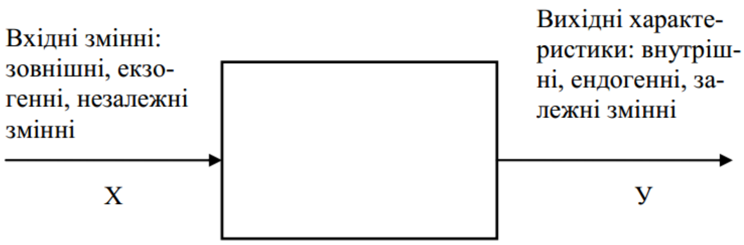


Рис. 1.

Досліднику необхідно виділити вхідні й вихідні характеристики та на підставі економетричних методів установити характер причинннонаслідкових зв'язків, що лежать в основі механізму функціонування соціально-економічної системи. Заключним етапом у процедурі послідовної формалізації опису процесів функціонування соціально-економічних систем є розробка математичних моделей.

2. Особливості побудови економетричних моделей.

Економетрична модель – це особливий клас економікоматематичних моделей, у яких дослідник вирішує цілий ряд задач:

- вибір форми математичної залежності, що описує поводження економічного об'єкта на основі системи спостережень;
- оцінка параметрів даної моделі різними методами (метод найменших квадратів, метод максимальної правдоподібності та ін.);
- перевірка статистичної значущості моделі.

Часто **економетрична модель** у загальному вигляді подається як система лінійних рівнянь:

$$\boldsymbol{BY=AX+E},$$

де B – матриця коефіцієнтів при ендогенних (залежних) змінних;
Y – вектор ендогенних змінних (спостережуване значення залежної змінної);
A – матриця коефіцієнтів при екзогенних (пояснюючих) змінних;
X – вектор екзогенних змінних;
AX – пояснювальна частина, що залежить від значень пояснюючих змінних;
E – вектор випадкових збурювань (помилок, відхилень).

Економетричні моделі включають досить широкий клас різних економіко-математичних моделей. Наведемо одну із **класифікацій економетричних моделей**.

1. За способом математичного подання економетричні моделі можна умовно розділити на прості й складні. *Прості* економетричні моделі подані одним рівнянням, однією залежністю, *складні* – декількома рівняннями, декількома залежностями.

2. За кількістю факторних ознак, що включає модель, прості економетричні моделі можна розділити на однофакторні й багатофакторні. *Однофакторні моделі* містять одну незалежну ознаку, *багатофакторні моделі* – ряд незалежних ознак. Однофакторні й багатофакторні моделі можуть бутиподані *лінійними* та *нелінійними* функціями.

3. Складні економетричні моделі можуть бути подані трьома видами систем одночасних рівнянь залежно від форми включення до правої частини ендогенних змінних. Звичайно виділяють три типи систем:

1) *системи, що розв'язані відносно ендогенних змінних;*

2) *рекурсивні системи;*

3) *системи, що не розв'язані відносно ендогенних змінних.*

4. ***Залежно від наявності (відсутності) у моделі фактора часу*** розрізняють *динамічні* й *статичні моделі.* Прикладами динамічних моделей є: *трендові моделі; моделі згладжування часових рядів; моделі декомпозиції часового ряду; авторегресійні моделі* й *моделі ковзної середньої; лагові моделі* та *динамічні регресійні моделі.*

Побудова економетричної моделі проводиться у кілька **основних етапів**:

1) якісний аналіз (постановка мети аналізу, визначення сукупності, визначення результативних і факторних ознак, вибір періоду, за який проводиться аналіз, вибір методу аналізу);

2) попередній аналіз сукупності, що моделюється (перевірка однорідності сукупності, виключення аномальних спостережень, уточнення необхідного обсягу ознак, установлення законів розподілу ознак);

3) побудова економетричної моделі (установлення переліку факторів, розрахунок оцінок параметрів рівнянь регресії, перебір конкуруючих варіантів моделі);

4) оцінка адекватності моделі (перевірка статистичної істотності рівняння залежності в цілому та його окремих параметрів; перевірка відповідності формальних властивостей оцінок завданням дослідження);

5) економічна інтерпретація та практичне використання моделі.

3. Лінійна парна регресія.

Нехай маємо ряди значень змінних і точки, які їм відповідають (xk, yk) нанесені на графік та з'єднані лінією. Якщо це реальні статистичні дані, то ми ніколи не одержимо просту лінію — лінійну, квадратичну, експоненціальну і т.д. Завжди будуть присутні відхилення залежної змінної, викликані помилками вимірювання, впливом неврахованих величин та випадкових факторів. Зв'язок змінних, на який накладається вплив випадкових факторів називається статистичним зв'язком. Викликає інтерес ситуація, коли змінні x та y, які задаються таблицею значень{xk}, {yk},k= не рівноправні між собою, але одна з них розглядається, як незалежна (пояснююча) змінна, а інша, як залежна від першої — пояснювана змінна. Це є той випадок, коли можна спробувати скласти аналітичний вираз y=f(x). Останнє рівняння носить назву рівняння регресії, і є формулою статистичного зв'язку між змінними. Якщо змінних дві — то така формула називається парною регресією, у випадку залежності кількох змінних — множинною регресією.

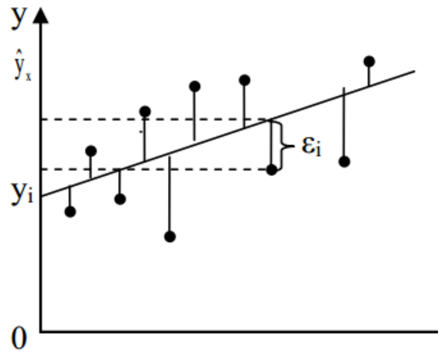
Якщо формула y = f(x) — лінійна, то мова йде про лінійну регресію. Аналогічно можна говорити про гіперболічну, експоненціальну та інші регресії.

Рівняння лінійної парної регресії має вигляд:

$$y^{\wedge} = b_0 + b_1x\ . \quad (1)$$

Відповідно до методу найменших квадратів (МНК) невідомі параметри b_0 й b_1 вибираються таким чином, щоб сума квадратів відхилень емпіричних значень y_i від значень $y^{\wedge}_i$, що знайдені за рівнянням регресії (1), була мінімальною:

$$S = \sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 = \sum_{i=1}^{n}\left(b_0 + b_1x_i - y_i\right)^2 \rightarrow \min\ . \quad (2)$$



На підставі необхідної умови екстремуму функції двох змінних S = S(b_0,b_1)

(2) прирівнюємо до нуля її частинні похідні, тобто

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = 2\sum_{i=1}^{n}(b_0 + b_1 x_i - y_i) = 0; \\ \dfrac{\partial S}{\partial b_1} = 2\sum_{i=1}^{n}(b_0 + b_1 x_i - y_i)x_i = 0, \end{cases}$$

звідки після перетворень одержимо систему нормальних рівнянь для визначення параметрів лінійної системи:

$$\begin{cases} b_0 n + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i; \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases} \qquad (3)$$

Тепер, розділивши обидві частини рівняння (3) на n, одержимо систему нормальних рівнянь у вигляді:

$$\begin{cases} b_0 + b_1 \bar{x} = \bar{y}; \\ b_0 \bar{x} + b_1 \overline{x^2} = \overline{xy}, \end{cases} \qquad (4)$$

де відповідні середні визначаються за формулами:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}; \qquad \overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n};$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}; \qquad \overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n}.$$

Підставляючи значення $b_0 = \bar{y} - b_1\bar{x}$ з першого рівняння системи (4) у рівняння регресії (1) одержимо

$$b_1 = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{K_{XY}}{s_x^2}, \qquad (5)$$

де b_1 – вибірковий коефіцієнт регресії; K_{XY} – вибірковий кореляційний момент або вибіркова кореляція; S_x^2 – вибіркова дисперсія змінної X:

$$K_{XY} = \overline{xy} - \bar{x}\cdot\bar{y}, \; s_x^2 = \overline{x^2} - \bar{x}^2.$$

Величина $r = \frac{b_1 \cdot S_x}{S_y}$ називається вибірковим коефіцієнтом кореляції. Він показує, наскільки величин s_y зміниться Y, якщо X зміниться на одне s_x.

$$r = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{s_x \cdot s_y}. \qquad (6)$$

Підставивши у вираз (6) вихідні дані, одержимо

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}. \qquad (7)$$

Коефіцієнт кореляції має такі *властивості*:

1) Коефіцієнт кореляції набуває значення на відрізку [–1;1], тобто $-1 \le \rho \le 1$. Чим ближче $|r|$ до 1, тим тісніше кореляційний зв'язок.

2) При $r = 1$ кореляційний зв'язок стає функціональним. При цьому всі спостережувані значення лежать на одній лінії.

3) При $r = 0$ кореляційний зв'язок відсутній та лінія регресії паралельна осі x. При $r > 0$ ($b1 > 0$) кореляційний зв'язок називається *прямим*. При $r < 0$ ($b1 < 0$) кореляційний зв'язок називається *оберненим*.

Приклад 1. Побудувати економетричну модель впливу вартості основних виробничих фондів на обсяг отриманого прибутку деяким умовним підприємством регіону. Статистичні дані для розрахунку і необхідні величини для побудови системи нормальних рівнянь наведені в таблиці 1.

Таблиця 1

Вплив вартості основних виробничих фондів на прибуток підприємства

№ підприємства	Прибуток, млн. грн. y_i	Основні фонди, млн. грн. x_i	x_i^2	$x_i y_i$
1	1,2	2,5	6,25	3,0
2	1,5	2,8	7,84	4,2
3	1,9	3,0	9,0	5,7
4	2,2	3,6	12,96	7,92
5	2,8	3,9	15,21	10,22
6	3,1	4,2	17,64	13,02
7	3,4	4,5	20,25	15,3
8	4,5	5,0	25,0	22,5
9	4,8	5,6	31,36	26,88
10	5,4	6,0	36,0	32,4
Всього	30,8	41,1	181,51	141,84

Розв'язування.

Побудуємо діаграму розсіювання залежності обсягу прибутку (y) від вартості основних виробничих фондів підприємства (x). Розміщення точок на діаграмі розсіювання (рис. 2) дає можливість зробити припущення про існування лінійної форми зв'язку у вигляді функції : $\hat{y}=a+bx$, де $\hat{y}$ – розрахунковий обсяг прибутку, млн. грн.; x – вартість основних виробничих фондів, млн. грн. Для знаходження параметрів a та b будуємо систему нормальних рівнянь:

$$\begin{cases} 10a + 41{,}1b = 30{,}8, \\ 41{,}1a + 181{,}5b = 141{,}84. \end{cases}$$

Розв'язавши окреслену систему рівнянь, отримуємо: a =-1,8989; b = =1,2114.

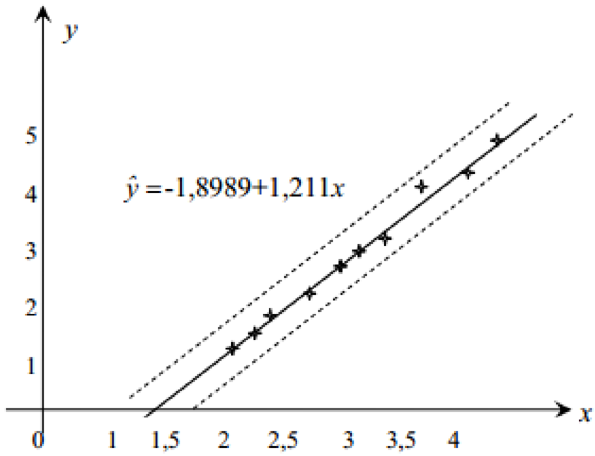


Рис. 2. Діаграма розсіювання та регресійна пряма, що відображає залежність прибутку від основних фондів

Отже, отримано регресійне рівняння $y^{\wedge}$ = -1,8989 + 1,2114x.

4. Основні положення регресійного аналізу. Оцінка параметрів регресійної моделі. Теорема Гаусса – Маркова.

Основні положення регресійного аналізу.

Як відзначено раніше, розглянута в регресійному аналізі залежність Y від X може бути подана у вигляді модельного рівняння регресії: $Mx(У)=\varphi(X)$. Окремі виміри величини Y будуть відрізнятися від обчислених значень за рахунок неврахованих факторів і помилок спостереження $Y=\varphi(X)+\varepsilon$, де ε – випадкова величина, що називається *збурюванням або помилкою*.

Для лінійної моделі ці рівняння мають вигляд: $Mx(У)=\beta_0+\beta_1 x$.

Спостережувані значення величини y визначаються за формулою

$$yi=\beta_0+\beta_1\cdot x+\varepsilon \qquad (8)$$

Основні передумови регресійного аналізу:

1. У моделі (8) збурювання εi (або yi) є випадковою величиною, а xi – невипадковою величиною.

2. Математичне очікування збурювання εi дорівнює нулю:

$$M(\varepsilon_i)=0,\ M(y_i)=\beta_0+\beta_i x_i\ .$$

3. Дисперсія збурювання εi (або yi) є постійною величиною для будь-якого номера i :

$$D(\varepsilon_i)=\sigma^2 \text{ або } D(yi)=\sigma^2.$$

Ця умова називається умовою *гомоскедастичності* (рівномінливості).

4. Збурювання εi та εj некорельовані: $M(\varepsilon i\cdot\varepsilon j)=0$, якщо $i\neq j$.

5. Збурювання εi (або yi) мають нормальні розподіли.

Оцінка параметрів регресійної моделі.

Вибірковою оцінкою рівняння (8) є рівняння регресії: $y^{\wedge}=b0+b1x$. Вплив неврахованих факторів і помилок спостережень у моделі (8) визначається за допомогою дисперсії збурювань або залишкової дисперсії. Незміщеною оцінкою залишкової дисперсії є:

$$s^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n-m} = \frac{\sum_{i=1}^{n} e_i^2}{n-2}, \qquad (9)$$

де yi – спостережуване значення; $yˆi$ – групова середня, що знайдена за рівнянням регресії; $ei = yˆi - yi$ – вибіркова оцінка збурювання εi або залишок регресії; $m = 2$ – число зв'язків (два рівняння для визначення b_0, b_1).

Теорема Гаусса – Маркова. Якщо регресійна модель (8) задовольняє умови 1–4, то оцінки b_0, b_1 мають найменшу дисперсію в класі всіх незміщених лінійних оцінок.

5. Інтервальна оцінка функції регресії та її параметрів. Оцінка значущості рівняння регресії. Коефіцієнт детермінації. Рангова кореляція.

Інтервальна оцінка функції регресії та її параметрів.

1. Знайдемо довірчий інтервал для умовного математичного очікування $M\ x(Y)$ із заданою надійністю. Припустимо, що γ – це задана надійність. Тоді мають місце наступні рівняння:

$$\hat{y} - \bar{y} = b_1(x - \bar{x}) \text{ – рівняння регресії;}$$

$$\hat{y} = \bar{y} + b_1(x - \bar{x});$$

$$\sigma_{\hat{y}}^2 = \sigma_{\bar{y}}^2 + (x - \bar{x})^2 \cdot \sigma_{b_1}^2 \text{ – рівняння дисперсії.}$$

Можна показати, що

$$\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}, \quad \sigma_{b_1}^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Тоді оцінкою дисперсії $\sigma_{\hat{y}}^2$ є:

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

де s^2 – вибіркова оцінка σ^2.

Можна показати, що статистика $t = \frac{\hat{y} - M_x(Y)}{s_{\hat{y}}}$ розподілена за t-розподілом Стьюдента з k ступенями волі. За таблицею цього розподілу для рівня значущості α та числа ступенів волі k знаходимо $t_{1-\alpha\,;\,k}$. Тоді довірчий інтервал для $M_x(Y)$ має вигляд:

$$\hat{y}_0 - t_{1-\alpha;k} \cdot s_{\hat{y}_0} \le M_x(Y) \le \hat{y}_0 + t_{1-\alpha;k} \cdot s_{\hat{y}_0},$$

де $s_{\hat{y}_0}^2 = s^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right).$

Прогнозовані значення залежної змінної Y залежать від відхилення незалежного фактора осі X . Чим більше це відхилення (останнє відхилення), тим більше помилка прогнозу.

2. Знайдемо довірчий інтервал для індивідуального значення y_0^*:

$$s_{y_0^*}^2 = s_{\hat{y}}^2 + s^2;$$

$$s_{y_0^*}^2 = s^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right).$$

Довірчий інтервал для y_0^* на рівні значущості α та при числі ступенів волі k має вигляд:

$$\hat{y}_0 - t_{1-\alpha;\,k} \cdot s_{y_0^*} \le y_0^* \le \hat{y}_0 + t_{1-\alpha;\,k} \cdot s_{y_0^*}.$$

3. Знайдемо довірчий інтервал для параметра β1. Можна показати, що статистика

$$t = \frac{b_1 - \beta_1}{s} \cdot \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

має t-розподіл Стьюдента. Тоді довірчий інтервал для параметра β_1 на рівні значущості α та при числі ступенів волі k має вигляд:

$$b_1 - t_{1-\alpha;\,k} \cdot \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \le \beta_1 \le b_1 + t_{1-\alpha;\,k} \cdot \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}.$$

4. Знайдемо довірчий інтервал для дисперсії збурювань σ^2.

Можна показати, що статистика $\frac{ns^2}{\sigma}$ має χ^2-розподіл зі k = n – 2 ступенями волі. Тоді довірчий інтервал для σ^2 на рівні значущості α має вигляд:

$$\frac{ns^2}{\chi^2_{\alpha/2;\,n-2}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{1-\alpha/2;\,n-2}}.$$

Оцінка значущості рівняння регресії. Коефіцієнт детермінації.

Перевірити значущість рівняння регресії – значить установити, чи відповідає математична модель експериментальним даним і чи досить включених у рівняння пояснюючих змінних (однієї або декількох) для опису результуючих факторів. Перевірка значущості рівняння регресії може бути проведена способом оцінки значущості коефіцієнта регресії b_1, що має t-розподіл з k ступенями волі. Рівняння парної регресії (або коефіцієнта регресії b_1) значуще на рівні α, якщо спостережуване значення статистики

$$t = \frac{b_1}{s}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

за абсолютною величиною більше табличного значення t1-α; k.

Обчислимо *суму квадратів відхилень*.

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

Можна показати, що третій доданок дорівнює нулю, тоді це рівняння запишемо так: $Q = Q_R + Qe$, де Q – загальна сума квадратів відхилень залежної змінної від середньої;Q_R – сума квадратів, що обумовлена регресією; Qe – сума квадратів, що обумовлена неврахованими факторами.

Ефективною оцінкою адекватності регресійної моделі є *коефіцієнт детермінації,* який обчислюється за допомогою наступної формули:

$$R^2 = \frac{Q_R}{Q} = 1 - \frac{Q_e}{Q}.$$

Коефіцієнт детермінації показує, яка частина зміни залежної перемінної обумовлена зміною незалежного фактора.

$$0 \le R^2 \le 1.$$

Чим ближче R^2 до 1, тим краще рівняння регресії наближає (апроксимує) експериментальні дані. У випадку парної регресії $R^2 = r^2$, де r – коефіцієнт кореляції.

6. Алгоритм побудови економетричної моделі та оцінка її достовірності.

Розглянемо алгоритм побудови економетричної моделі на прикладі.

Приклад 2. Досліджується залежність обсягів реалізації продукції підприємства (Y) від його кредиторської заборгованості (X) протягом року на основі даних його фінансової звітності, наведених в табл.2.

Необхідно:

1) визначити вид зв’язку між заданими показниками;
2) розрахувати оцінки параметрів методом найменших квадратів;
3) виконати аналіз якості побудованої моделі та оцінок її параметрів;

Таблиця 2

Період часу	Обсяги реалізації продукції, млн.грн.	Кредиторська заборгованість, млн.грн
I	*Y*	*X*
Січень	0,54	0,42
Лютий	0,47	0,37
Березень	0,65	0,66
Квітень	0,83	0,71
Травень	0,76	0,83
Червень	0,50	0,57
Липень	0,72	0,69
Серпень	0,84	0,75
Вересень	0,93	0,88
Жовтень	1,11	0,99
Листопад	0,94	1,01
Грудень	1,14	1,17

Розв’язання

1) *Специфікація моделі*

В даному прикладі розглядається зв’язок між двома показниками – обсягом реалізації продукції та кредиторською заборгованістю підприємства, отже, маємо парну регресію.

Позначимо залежну величину – обсяг реалізації продукції – через *y*, а незалежну величину – кредиторську заборгованість – через *x*.

У випадку парної регресії зв'язок між величинами може бути представлений у графічній формі, що дає можливість наочно впевнитись у виборі кращої форми регресійного рівняння (специфікації моделі).

Представимо графічно залежність між наведеними показниками на рис.3.

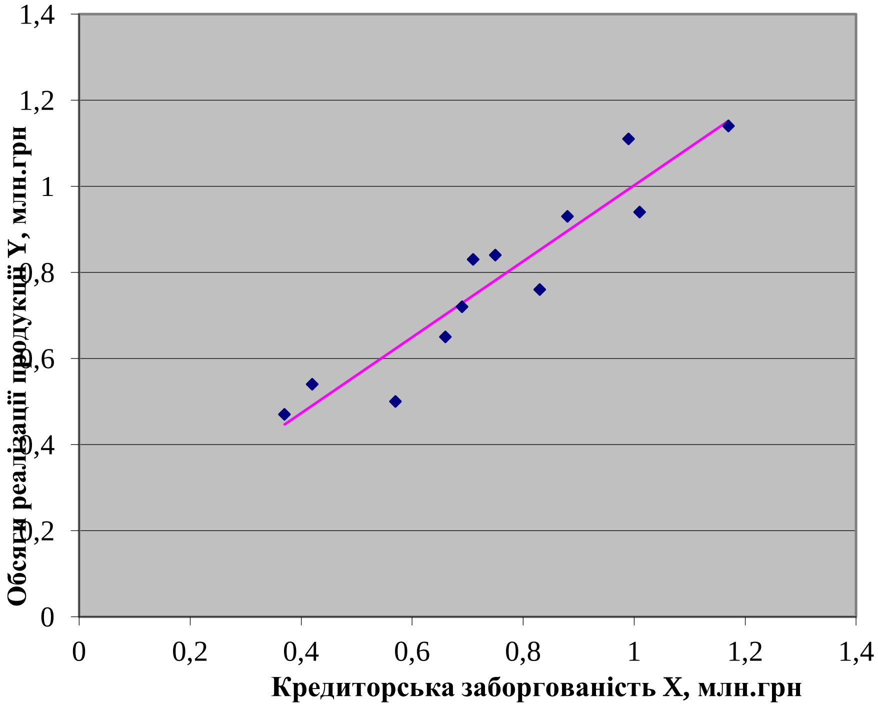


Рис. 3.

Очевидно, що в даному випадку зв'язок між наведеними показниками близький до лінійного, тому емпіричне рівняння матиме вигляд:

$$y_i = \beta_0^* + \beta_1^* x_i + e_i$$

2) *Визначення параметрів вибраного рівняння*

Розрахуємо значення коефіцієнтів:

$$\beta_1^* = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}; \; \beta_0^* = \bar{y} - \bar{x} \cdot \beta_1^*.$$

Для потрібних обчислень необхідно поетапно розрахувати наступні величини: $\bar{x}, \bar{y}$, для кожного спостереження $i = \overline{1,12}$ $(x_i - \bar{x})$, $(y_i - \bar{y})$, $(x_i - \bar{x})^2$, а також $\sum_{i=1}^{12}(x_i - \bar{x}) \cdot (y_i - \bar{y}), \sum_{i=1}^{12}(x_i - \bar{x})^2$.

Обчислимо $\bar{x} = \frac{0{,}54 + 0{,}47 + 0{,}65 + \ldots + 1{,}14}{12}$ =0,7858;

$$\bar{y} = \frac{0,42 + 0,37 + 0,66 + \ldots + 1,17}{12} = 0,7542.$$

Розрахунки значень інших величин подано в табл.3.

Таблиця 3

№ спостереження	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	x_i^2
1	0,54	0,42	-0,3342	-0,2458	0,1117	0,0821	0,1764
2	0,47	0,37	-0,3842	-0,3158	0,1476	0,1213	0,1369
3	0,65	0,66	-0,0942	-0,1358	0,0089	0,0128	0,4356
4	0,83	0,71	-0,0442	0,0442	0,0020	-0,0020	0,5041
5	0,76	0,83	0,0758	-0,0258	0,0058	-0,0020	0,6889
6	0,50	0,57	-0,1842	-0,2858	0,0339	0,0526	0,3249
7	0,72	0,69	-0,0642	-0,0658	0,0041	0,0042	0,4761
8	0,84	0,75	-0,0042	0,0542	0,0000	-0,0002	0,5625
9	0,93	0,88	0,1258	0,1442	0,0158	0,0181	0,7744
10	1,11	0,99	0,2358	0,3242	0,0556	0,0764	0,9801
11	0,94	1,01	0,2558	0,1542	0,0655	0,0394	1,0201
12	1,14	1,17	0,4158	0,3542	0,1729	0,1473	1,3689
$\sum$	**9,43**	**9,05**	**0,00**	**0,00**	**0,6237**	**0,5503**	7,4489

$$\beta_1^* = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{0,5503}{0,6237} \approx 0,8823,$$

$$\beta_0^* = \bar{y} - \bar{x} \cdot \beta_1^* \approx 0,7858 - 0,7542 \cdot 0,8823 = 0,1204.$$

Отже, емпіричне лінійне рівняння парної регресії (економетрична модель залежності між обсягом реалізації продукції та кредиторською заборгованістю) по визначеним β_0^*, β_1^* має вигляд:

$$y_i^* = 0,1204 + 0,8823 x_i.$$

Розрахунки значень y_i^* запишемо в табл.4. В цю ж таблицю включимо й відповідні значення e_i, обчислені як

$$e_i = y_i - y_i^*.$$

3) *Аналіз якості моделі*

3.1) *Перевірка загальної якості рівняння регресії*

Нагадаємо, що загальну якість рівняння регресії здійснюють через розрахунок коефіцієнта детермінації, що обчислюється за формулою

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \text{або} \quad R^2 = \frac{\sum_{i=1}^{n}\left(y_i^* - \bar{y}\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

Обчислимо суми $\sum_{i=1}^{12}(y_i - \bar{y})^2$, $\sum_{i=1}^{12}(y_i^* - \bar{y})^2$, $\sum_{i=1}^{12} e_i^2$. Для цього в стовпчиках таблиці 4 запишемо розрахунок величин $(y_i - \bar{y})$, $(y_i^* - \bar{y})$ $(i = \overline{1,12})$, та знайдемо відповідні значення $(y_i - \bar{y})^2$, $(y_i^* - \bar{y})^2$ $(i = \overline{1,12})$, а потім визначимо $\sum_{i=1}^{12}(y_i - \bar{y})^2$ та $\sum_{i=1}^{12}(y_ø^* - \bar{y})^2$:

Таблиця 4

№ спостереження	y_i^*	$(y_i^* - \bar{y})$	$(y_i^* - \bar{y})^2$	e_i	$e_ш^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
1	0,4910	-0,2948	0,0869	0,0490	0,0024	-0,2458	0,0604
2	0,4469	-0,3390	0,1149	0,0231	0,0005	-0,3158	0,0998
3	0,7027	-0,0831	0,0069	-0,0527	0,0028	-0,1358	0,0185
4	0,7469	-0,0390	0,0015	0,0831	0,0069	0,0442	0,0020
5	0,8527	0,0669	0,0045	-0,0927	0,0086	-0,0258	0,0007
6	0,6233	-0,1625	0,0264	-0,1233	0,0152	-0,2858	0,0817
7	0,7292	-0,0566	0,0032	-0,0092	0,0001	-0,0658	0,0043
8	0,7822	-0,0037	0,0000	0,0578	0,0033	0,0542	0,0029
9	0,8969	0,1110	0,0123	0,0331	0,0011	0,1442	0,0208
10	0,9939	0,2081	0,0433	0,1161	0,0135	0,3242	0,1051
11	1,0116	0,2257	0,0510	-0,0716	0,0051	0,1542	0,0238
12	1,1527	0,3669	0,1346	-0,0127	0,0002	0,3542	0,1254
$\sum$	9,4300	0,0000	0,4856	0,0000	0,0597	0,00	0,5453

Таким чином, маємо:

$$R^2 = 1 - \frac{\sum_{i=1}^{15} e_i^2}{\sum_{i=1}^{15}(y_i - \bar{y})^2} = 1 - \frac{0,0597}{0,5453} \approx 1 - 0,1095 = 0,8905.$$

Відомо, що R^2 є мірою, яка дозволяє визначити наскільки вдало емпіричне рівняння регресії узгоджується із статистичними даними, тобто наскільки реальні значення відхиляються від побудованої лінії регресії. Для нашого прикладу з рис.3 бачимо, що точки відповідні реальним спостереженням розташовуються дуже близько від лінії регресії, отже, отриманий результат $R^2 \approx 0,89$ є цілком закономірним. Зробимо висновок: на 89% зміна обсягів реалізації продукції пояснюється зміною кредиторської заборгованості підприємства, і лише на 11% такі зміни пов'язані з іншими (не врахованими в даній економетричній моделі) факторами.

Оскільки $R^2 = r_{xy}^2$, обчислимо значення парного коефіцієнта кореляції $r_{xy} = \sqrt{R^2}$ та здійснимо перевірку його статистичної значущості.

Маємо $r_{xy} = \sqrt{R^2} = \sqrt{0,89} \approx 0,9436$.

Для перевірки на статистичну значущість парного коефіцієнта кореляції r_{xy} обирається статистичний критерій

$$t^* = \frac{r_{xy}}{\sqrt{1-r_{xy}^2}}\sqrt{n-2}$$

по заданому рівню значущості α і ступенях свободи $k = n-2$. Обчислене значення t^* порівнюємо з табличним значенням, на основі чого робимо висновок стосовно прийняття гіпотези про значущість (незначущість) коефіцієнта кореляції.

$$t^* = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} = \frac{0{,}9436}{\sqrt{1-0{,}8905}}\sqrt{12-2} \approx 9{,}016$$

якщо α=0,05; $k = 12-2 = 10$, то табличне значення $t''_{кр}\left(\frac{\alpha}{2};k\right) = t''_{кр}(0{,}025;10)$ =2,228; $t'_{кр}(0{,}025;10) = -t''_{кр}(0{,}025;10) = -2{,}228$.

Оскільки $t^* \notin \left[t'_{кр}\left(\frac{\alpha}{2};k\right); t''_{кр}\left(\frac{\alpha}{2};k\right)\right]$, то приймається гіпотеза про статистичну значущість розрахованого для даного прикладу коефіцієнта кореляції.

Загалом проведений аналіз якості побудованої лінії регресії дає підстави вважати дану модель якісною, а тому можливо використовувати її для подальших досліджень.

3.2) *Перевірка статистичної значущості оцінок параметрів* β_0^*, β_1^* *економетричної моделі*

Для перевірки нульової гіпотези $H_0 : \beta_i = 0$ при альтернативній гіпотезі $H_\alpha : \beta_i \neq 0$ вибирають за статистичний критерій випадкову величину:

$$t_{\beta_j^*} = \frac{\beta_j^*}{S_{\beta_j^*}}, \ (j = 0, 1)$$

що має розподіл Стьюдента (*t*-розподіл) із $k = n-2 = 12-2 = 10$ ступенями свободи. По обраному рівню значущості $\alpha = 0{,}05$ та числу ступенів свободи *k=10* маємо $t''_{кр}\left(\frac{\alpha}{2};k\right) = t''_{кр}(0{,}025;10)$=2,228 та другу точку

$$t'_{кр}\left(\frac{\alpha}{2};k\right) = t'_{кр}(0{,}025;\,10) = -t''_{кр}(0{,}025;\,10) = -2{,}228.$$

Область прийняття гіпотези, що $\beta_i = 0$ визначається інтервалом:

$$t'_{кр}(0{,}025;\,10) < t_{\beta_i^*} < t''_{кр}(0{,}025;\,10) \qquad (j = 0, 1)$$

$$-2{,}228 < t_{\beta_i^*} < 2{,}228 \ (j = 0, 1).$$

Обчислимо спостережене значення обраного статистичного критерію, як

$$t^*_{\beta_j^*} = \frac{\beta_j^*}{S_{\beta_j^*}}, \ (j = 0, 1).$$

Нагадаємо, що коли $t^*_{\beta^*_i} \in \left[t'_{\text{кр}}\left(\frac{\alpha}{2}; k\right); t''_{\text{кр}}\left(\frac{\alpha}{2}; k\right)\right]$, то приймається гіпотеза про те, що $\beta_j = 0$, і, навпаки, $\beta_j \neq 0$ ($j = 0, 1$), якщо $t^*_{\beta^*_i} \notin \left[t'_{\text{кр}}\left(\frac{\alpha}{2}; k\right); t''_{\text{кр}}\left(\frac{\alpha}{2}; k\right)\right]$.

Для розрахунку $S_{\beta^*_0}$ та $S_{\beta^*_1}$ використаємо формули:

$$S_{\beta^*_0} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} \cdot \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}},$$

$$S_{\beta^*_1} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} \cdot \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}.$$

Для проведення необхідних обчислень виконаємо розрахунок $\sum_{i=1}^{12} x_i^2$.

Тепер

$$S_{\beta^*_0} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} \cdot \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{0{,}0597}{10}} \cdot \sqrt{\frac{7{,}4489}{12 \cdot 0{,}6237}} \approx 0{,}0773 \cdot 0{,}99763 \Rightarrow$$

$$S_{\beta^*_0} \approx 0{,}0771.$$

Також обчислимо $S_{\beta^*_1}$:

$$S_{\beta^*_1} = \frac{\sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{0{,}0597}{10}} \cdot \frac{1}{\sqrt{0{,}6237}} \approx 0{,}0773 \cdot 1{,}2662 \Rightarrow$$

$$S_{\beta^*_1} \approx 0{,}09786.$$

Отже маємо:

$t^*_{\beta^*_0} = \frac{\beta^*_0}{S_{\beta^*_0}} = \frac{0{,}1204}{0{,}0771} \approx 1{,}56 \in [-2{,}228; 2{,}228]$, тобто для оцінки параметру β^*_0 приймається гіпотеза $\beta^*_0 = 0$; аналогічно здійснюємо перевірку статистичної значущості параметру β^*_1:

$t^*_{\beta_1^*} = \dfrac{\beta_1^*}{S_{\beta_1^*}} = \dfrac{0{,}8823}{0{,}09786} \approx 9{,}016 \notin [-2{,}228; 2{,}228]$, а тому оцінка параметру β_1^* є статистично значущою, що свідчить про суттєвий вплив на залежну змінну обраної незалежної змінної, тоді як статистична незначущість коефіцієнту β_0^*, вказує на те, що всі інші фактори, які не були враховані в даній регресійній моделі не дають значного впливу на залежну змінну.

Отже, на даному етапі аналізу побудованої економетричної моделі (перевірка статистичної значущості оцінок параметрів β_0^*, β_1^*) можна визначити модель як якісну, оскільки статистично значуща оцінка параметру β_1^* вказує на суттєвий вплив обраного фактору на залежну змінну, тоді як всі інші невраховані фактори (статистично незначуща оцінка параметру β_0^*) не дають суттєвого впливу на зміну значень *y*.

3.3) *Довірчі інтервали для оцінок параметрів* β_0^*, β_1^* *економетричної моделі*

Довірчі інтервали обчислюється за формулами:

для β_0^*:

$$\beta_0^* - t(\gamma, k)S_{\beta_0^*} < \beta_0 < \beta_0^* + t(\gamma, k)S_{\beta_0^*}$$

для β_1^*:

$$\beta_1^* - t(\gamma, k)S_{\beta_1^*} < \beta_1 < \beta_1^* + t(\gamma, k)S_{\beta_1^*}$$

де $t(\gamma, k)$ визначається за таблицею розподілу Стьюдента по заданій надійності γ і числу ступенів свободи $k = n - 2$.

Оберемо рівень надійності $\gamma = 0{,}95$ та, враховуючи, що число ступенів свободи дорівнює $k = n - 2$=12–2=10 маємо $t(\gamma, k)$=2,228.

Таким чином, довірчі інтервали для оцінок параметрів визначаються межами:

$$\beta_0^* - t(\gamma, k)S_{\beta_0^*} < \beta_0 < \beta_0^* + t(\gamma, k)S_{\beta_0^*}$$

$$0{,}1204 - 2{,}228 \cdot 0{,}0771 < \beta_0 < 0{,}1204 + 2{,}228 \cdot 0{,}0771$$

$$-0{,}0514 < \beta_0 < 0{,}2922,$$

аналогічно для β_1 маємо:

$$\beta_1^* - t(\gamma, k)S_{\beta_1^*} < \beta_1 < \beta_1^* + t(\gamma, k)S_{\beta_1^*}$$

$$0{,}8823 - 2{,}228 \cdot 0{,}09786 < \beta_1 < 0{,}8823 + 2{,}228 \cdot 0{,}09786$$

$$0{,}6643 < \beta_1 < 1{,}1004.$$

3.4) Перевірка адекватності побудованої економетричної моделі.

Для оцінки рівня адекватності побудованої економетричної моделі експериментальним даним використовуємо критерій Фішера *F*, основу якого складає формула:

$$F = \frac{r^2}{1-r^2} \cdot \frac{n-m-1}{m}$$

$F_{розр.} = \frac{0,8905}{1-0,8905} \cdot \frac{12-1-1}{1} = 87,73.$

Знайдемо табличне значення даного кретерію (*F табл.)*, для рівня надійності Р=0,95 та числа ступенів вільності:

$$k_1 = m = 1,\ k_2 = n - m - 1 == 12 - 1 - 1 = 10,$$

$$F_{табл.} = 4,96,$$

$$F_{розр.} > F_{табл.}$$

Оскільки $F_{табл.} > F_{розр.}$, то отриману економетричну модель з надійністю Р=0,95 можна вважати адекватною експериментальним даним і на її основі доцільно проводити всебічний економетричний аналіз.

Контрольні питання:

1. Предмет економетрії.
2. Параметри системи, екзогенні та ендогенні змінні.
3. Економетрична модель, її загальний вигляд.
4. Класифікація та етапи побудови економетричних моделей.
5. Алгоритм методу найменших квадратів для парної регресії.
6. Коефіцієнт кореляції, його властивості.
7. Основні передумови регресійного аналізу.
8. Залишкова дисперсія, теорема Гаусса – Маркова.
9. Інтервальна оцінка функції регресії та її параметрів.
10. Оцінка значущості рівняння регресії. Коефіцієнт детермінації.
11. Алгоритм побудови економетричної моделі.