Міністерство освіти і науки України
Донбаська державна машинобудівна академія

Л. В. Васильєва, О. А. Кльованик

РЕГРЕСІЙНІ МОДЕЛІ ТА АНАЛІЗ ЧАСОВИХ РЯДІВ

Навчальний посібник
для студентів вищих навчальних закладів

Рекомендовано
Міністерством освіти і науки

Краматорськ 2010

УДК 330.43(075.8)
ББК 65.053
В 19

Рецензенти:

Калоєров С. О., д-р фіз.-мат. наук, професор (Донецький національний університет)

Зайцев Д. А., д-р техн. наук, професор (Одеська національна академія зв'язку)

Новіков О. О., канд. фіз.-мат. наук, доцент (Слов'янський державний педагогічний університет)

Рекомендовано Міністерством освіти і науки України
Лист № 14/18-Г-1340 від 01.08.2007

Васильєва Л. В.

В 19 Регресійні моделі та аналіз часових рядів : навчальний посібник для студентів вищих навчальних закладів / Л. В. Васильєва, О. А. Кльованик. – Краматорськ : ДДМА, 2010. – 176 с.
ISBN 978-966-379-453-2.

Містить теоретичні відомості та практичну частину за такими розділами економетрики: лінійна і нелінійна однофакторна регресія, перевірка адекватності моделі, довірчий інтервал і довірча область, прогноз за обраною моделлю; модель багатофакторної регресії, колінеарність і мультиколінеарність факторів; еластичність моделі; системи одночасних рівнянь, ендогенні та екзогенні змінні; часові ряди, метод ковзних середніх і експоненційного згладжування.

Посібник розрахований на студентів і аспірантів економічних спеціальностей, а також буде корисним для тих, хто бажає самостійно освоїти економетричні розрахунки.

УДК 330.43(075.8)
ББК 65.053

ISBN 978-966-379-453-2

ЗМІСТ

ЧАСТИНА 1

ВСТУП

Потреба в способах статистичного аналізу даних в економічній практиці дуже велика. Для успішного функціонування в умовах жорсткої конкуренції підприємства, банки, страхові компанії мають потребу в аналізі наявної інформації і отриманні обґрунтованих висновків. Аналіз такої інформації здійснюється за допомогою методів, об'єднаних у дисципліну «Економетрика».

Буквальний переклад слова «економетрика» означає «вимірювання економіки».

Економетрика – це наука, яка вивчає кількісні закономірності та взаємозв'язки економічних об'єктів і процесів за допомогою математико-статистичних методів і моделей. Тобто економетрика відновлює невідомі економіко-математичні залежності за статистичними даними і розглядає можливість використовування цих моделей в економічних дослідженнях.

Модель – це штучне відтворення деякого економічного процесу для досліджень. У економетриці під моделлю мають на увазі математичну модель, тобто опис економічного процесу за допомогою математичних формул.

Економетричні моделі кількісно описують зв'язок між вхідними факторами економічної системи x і результуючим показником (відгуком) y плюс вплив випадкової компоненти ε.

За моделлю одержують прогноз.

Прогноз – це розрахунок невідомого показника для заданих факторів на основі моделі.

1 ПРЕДМЕТ ЕКОНОМЕТРИКИ

1.1 Основні задачі економетрики

Економетрика повинна вирішувати п'ять *основних задач*:

1 *Вибір конкретного виду функції для деякого економічного процесу.* Наприклад, залежність між доходом і витратою можна описати так: $y = \beta_1 x + \beta_0$.

2 *Збір і підготовка економічної інформації*. Важливо вибрати правильні позначення для змінних і правильні одиниці вимірювання. Наприклад, якщо йдеться про зміну доходу з часом, то функція набуде вигляду $y = f(t)$, де y – дохід, t – час. Якщо йдеться про національний дохід, то одиниці вимірювання візьмемо такі: для y – млн грн., для t – рік. Якщо йдеться про підприємство, то для y – грн., для t – місяць.

3 *Оцінка на підставі наявних статистичних даних значень параметрів моделі.*

4 *Перевірка моделі на адекватність, оцінка якості вибраної моделі, її простоти, точності опису даних.*

5 *Економічний аналіз моделі.*

1.2 Етапи економетричного аналізу

Щоб провести економетричний аналіз, потрібно:

1 Висунути гіпотезу про вид залежності за статистичними даними відповідно до набору факторів.

2 Провести оцінку невідомих параметрів моделі.

3 Перевірити модель на адекватність.

4 Використовувати модель в економічних прогнозах і дослідженнях.

1.3 Інформаційна база економетрики

Рішення задач економетрики проводиться на базі статистичних даних. *Статистичні дані* – це дані, зібрані на реальних економічних об'єктах.

У економетриці статистичні дані можна розподілити на 2 типи: динамічні (часові) і варіаційні ряди.

Часові ряди – це послідовність спостережень за одним і тим же процесом або явищем у різні проміжки часу. Наприклад, дані про динаміку рівня інфляції за певний період.

Варіаційні ряди – послідовність спостережень за яким-небудь економічним показником для різних однотипних об'єктів. Усі виміри роблять в один і той же час. Значення варіаційного ряду розташовують у порядку зростання.

Варіаційні та часові (динамічні) ряди досліджуються різними методами, і для них будуються різні за своєю суттю моделі.

Варіаційні дані можна обробляти методами регресійного аналізу. Найбільш уживаний з цих методів – метод найменших квадратів, який буде розглянутий далі. Моделі, отримані таким чином, називають регресійними.

1.4 Класифікація регресійних економетричних моделей

1 Однофакторні $y = f(x)$.

- a) Лінійні виду $y = b_0 + b_1 x$.
- b) Нелінійні:
  1) що зводяться до лінійних;
  2) істотно нелінійні.

2 Багатофакторні $y = f(x_1, x_2, ..., x_p)$.

- a) Лінійні виду $y = b_0 + b_1 x_1 + ... b_p x_p$.
- b) Нелінійні:
  1) що зводяться до лінійних;
  2) істотно нелінійні.

1.5 Обробка інформаційних даних

Сукупність даних динамічних і варіаційних рядів обробляється за правилами, розробленими у математичній статистиці.

Генеральна сукупність – усі можливі реалізації показника, що нас цікавить. На практиці ми спостерігаємо випадково вибрані значення цього показника, це – *вибірка.* За генеральною сукупністю можна отримати точні значення параметрів, а за вибіркою – наближені, або оцінки.

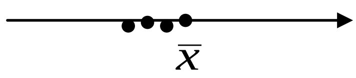
Об'єм вибірки n – сумарна кількість спостережень. Об'єми вибірок можуть бути невеликими (n≈10), великими (n≈100) і дуже великими ($n \approx 10^4$). На практиці найчастіше доводиться мати справу з великими і дуже великими вибірками, тому розрахунки проводяться за допомогою комп'ютера.

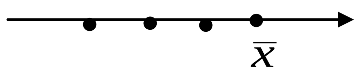
У всіх випадках усю сукупність вибіркових даних x_i (i=1…n) прагнуть охарактеризувати деякими усередненими параметрами, які враховують особливості вибірки. За вибірками розраховують основні *статистичні характеристики:*

1 Середнє значення $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

2 Варіація (дисперсія) $Var(x) = D(x) = \sigma_x^2 = \frac{1}{n}\sum (x_i - \bar{x})^2$.

Дисперсії характеризують, як сильно розсіяні значення вибірки щодо середнього значення:

D(X) мала 

D(X) велика 

3 Середньоквадратичне відхилення $\sigma_x = \sqrt{Var(x)}$ або стандартне відхилення. Ця величина характеризує відхилення вибіркових значень у середньому від $\bar{x}$.

2 ОДНОФАКТОРНА ЛІНІЙНА РЕГРЕСІЯ

Вивчення залежностей економічних показників починають з випадку двох змінних – x і y: $y = f(x)$. Цей метод найбільш простий і може бути поданий графічно.

Спершу потрібно встановити, чи існує функціональна залежність між фактором x і відкликом y, і якщо існує, то визначити формулу зв'язку.

Для аналізу дані подають у вигляді таблиці 1.

Таблиця 1

x	***y***
x_1	y_1
x_2	y_2
...	...
x_n	y_n

За такою таблицею будується кореляційне поле (діаграма розсіювання). Кореляційним полем називають систему точок (x_i, y_i), (i =1, …, n), зображену на координатній площині *xoy* (рис.1).

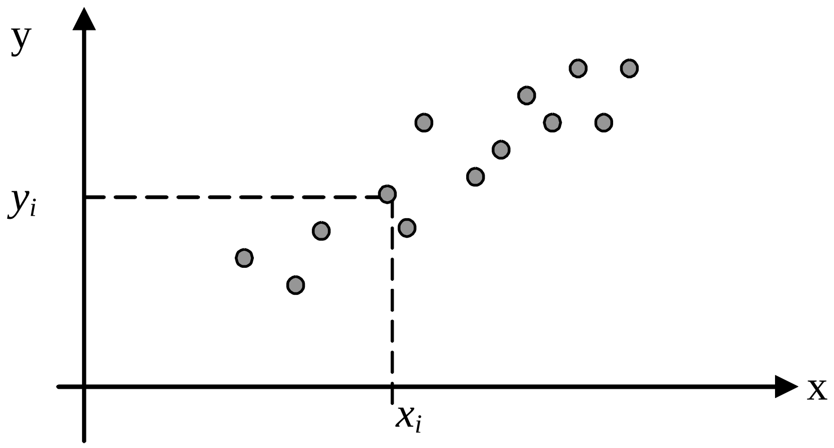


Рисунок 1

Точка з координатами $(\bar{x}, \bar{y})$ називається центром розсіяння.

З вигляду кореляційного поля висувається припущення, чи є залежність між y і x лінійною або нелінійною.

Значення σ_x (великі або малі) ще не дають характеристику того, чи є зв'язок між *x* і *y.* На рисунках 2, 3, 4 показані ситуації, коли σ_x, σ_y малі, але на рисунку 2 залежності вигляду *y = f(x)* немає, на рисунку 3 залежність є, і вона лінійна, на рисунку 4 є явно нелінійна залежність.

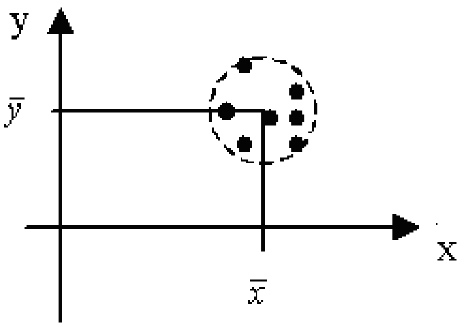


Рисунок 2

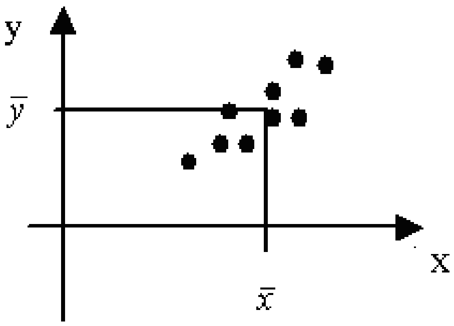


Рисунок 3

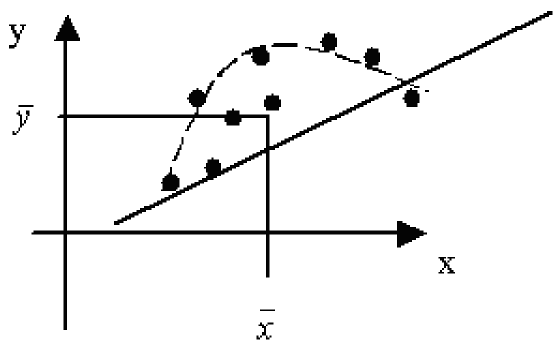


Рисунок 4

Тому вводиться ще одна статистика – *коваріація* x, y – $\text{cov}(x,y)$ (сумісна варіація):

$$\text{cov}(x,y) = \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y}).$$

Коваріація має ту властивість, що для випадків на рисунках 2 і 4 дорівнює 0, а для випадку на рисунку 3 не дорівнює 0, і тим більше за модулем, чим ближче кореляційне поле до прямої.

Якщо кореляційне поле починає розмиватися (рис.5), коваріація зменшується.

Для зручності роботи коваріацію поділяють на добуток σ_x, σ_y і називають *коефіцієнтом кореляції*. Позначають $\boldsymbol{r_{xy}}$.

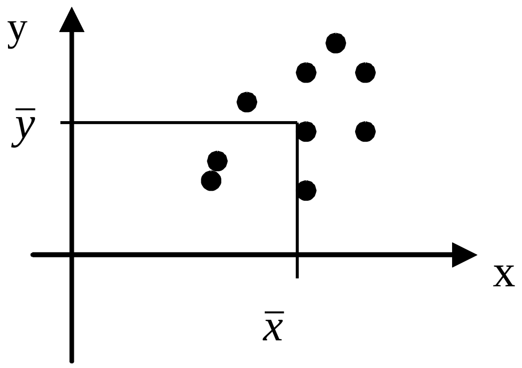


Рисунок 5

Коефіцієнт кореляції між змінними x і y обчислюється за формулою

$$r_{xy} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}.$$

Коефіцієнт кореляції є показником щільності *лінійного* взаємозв'язку.

Властивості коефіцієнта кореляції:

1 $-1 \le r_{xy} \le 1$.

2 Якщо $r_{xy} > 0$, то залежність між фактором x і y пряма, тобто із зростанням x показник y також зростає.

3 Якщо $r_{xy} < 0$, то залежність між фактором x і y зворотна.

4 Якщо $|r_{xy}| \approx 1$, зв'язок між x і y майже лінійний (див. рис. 3).

5 Якщо $|r_{xy}| \approx 0$, зв'язку немає (див. рис.2) або зв'язок суттєво нелінійний (див. рис.4).

Щільність лінійного взаємозв'язку оцінюють за наступною таблицею 2.

Таблиця 2

Значення $\|r_{xy}\|$	**Щільність лінійного зв'язку**
0,9…1,0	Тісна
0,6…0,9	Достатня
0,3…0,6	Слабка
< 0,3	Немає зв'язку

Слід мати на увазі, що величина лінійного коефіцієнта кореляції r_{xy} оцінює щільність тільки *лінійного* зв'язку. Тому близькість до нуля $|r_{xy}|$ не означає відсутності взагалі зв'язку між ознаками. При $|r_{xy}| > 0$ відсутній саме *лінійний* зв'язок між x і y.

Звичайно будують кореляційну таблицю (кореляційну матрицю) зв'язку між змінними *x* і *y*.

Вона має вигляд таблиці 3.

Таблиця 3

	x	y
x	1	r_{yx}
y	r_{xy}	1

Властивості кореляційної матриці:

1 Кореляція *фактора* з самим собою дорівнює 1: $r_{xx} = r_{yy} = 1$.

2 Матриця симетрична щодо головної діагоналі: $r_{xy} = r_{yx}$.

3 ДОБІР ПАРАМЕТРІВ ПРЯМОЇ РЕГРЕСІЇ ЗА МЕТОДОМ НАЙМЕНШИХ КВАДРАТІВ (МНК)

Парною (однофакторною) лінійною регресією називається лінійна залежність $y = b_0 + b_1 x$ між залежним показником *y* і незалежним фактором *x*.

Лінійний зв'язок між *x* і *y* описують залежністю

$$\widehat{y} = b_0 + b_1 x. \qquad (1)$$

Через випадкові впливи показник y_i є випадковим і може бути записаний:

$$y_i = b_0 + b_1 x_i + e_i, \text{ i=1..n}, \qquad (2)$$

де e_i – випадкове відхилення (рис.6).

Відхилення (помилка) початкових даних y_i від розрахованих за моделлю значень $\widehat{y}_i = y(x_i)$ обчислюється за формулою

$$e_i = \widehat{y}_i - y_i = b_0 + b_1 x_i - y_i.$$

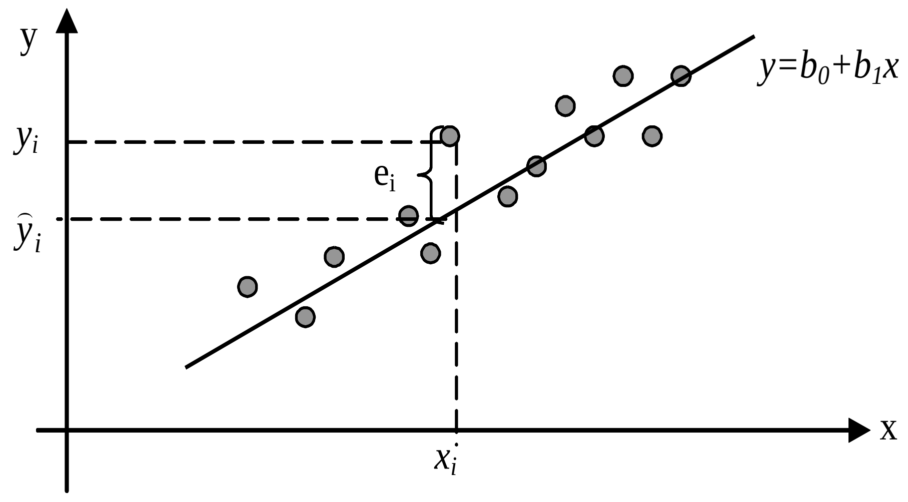


Рисунок 6

3.1 Метод найменших квадратів

Суть МНК полягає в тому, щоб мінімізувати відхилення e_i в сукупності за допомогою правильного підбору коефіцієнтів b_0, b_1.

Оскільки відхилення може мати випадковий знак (+ або –), то розглядають квадрати відхилень і мінімізують суму квадратів відхилень:

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (b_0 + b_1 x_i - y_i)^2 . \qquad (3)$$

Сума S є функцією двох невідомих параметрів b_0, b_1. Необхідна умова мінімуму функції S – рівність нулю її часткових похідних:

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = 0; \\ \dfrac{\partial S}{\partial b_1} = 0. \end{cases} \qquad (4)$$

Отримали систему двох лінійних рівнянь від двох невідомих. Якщо її визначник відмінний від нуля, то така система має єдине рішення.

Зробивши арифметичні перетворення, одержимо формули для визначення коефіцієнтів:

$$b_1 = r_{xy} \frac{\sigma_y}{\sigma_x}; \qquad (5)$$
$$b_0 = \bar{y} - b_1 \bar{x}.$$

Підставляючи формули (5) у рівняння $y = b_0 + b_1 x$, одержимо формулу лінії регресії

$$y - \bar{y} = r_{xy} \frac{\sigma_y}{\sigma_x} (x - \bar{x}) . \qquad (6)$$

Параметр b_1 називають *коефіцієнтом регресії*. Його величина показує середню зміну результату при зміні фактора на одну одиницю. Так, якщо у функції витрат *y(x) = 3 000 + 2x (y*, тис. грн. – витрати, *x* – кількість одиниць продукції), то із збільшенням об'єму продукції х на одну одиницю витрати виробництва зростають у середньому на 2 тис. грн., тобто додатковий приріст продукції на одну одиницю вимагає збільшення витрат у середньому на 2 тис. грн.

Можливість чіткої економічної інтерпретації коефіцієнта регресії b_1 пояснює поширеність лінійного рівняння регресії в економетричних дослідженнях.

Формально b_0 – це значення *y* при *x* = 0. Але правильно інтерпретувати можна лише знак при параметрі b_0. Якщо b_0>0, то відносна зміна показника *y* відбувається повільніше, ніж зміна фактора. Якщо b_0<0, та зміна результату випереджає зміну фактора.

3.2 Властивості лінійної регресії

1 Порівняємо рівняння (6) з рівнянням прямої, що проходить через точку (x_0, y_0)(рис.7):

$$y - y_0 = k(x - x_0). \qquad (7)$$

З порівняння рівнянь (6) і (7) видно, що пряма регресії завжди проходить через центр розсіювання кореляційного поля, тобто через точку ($\bar{x}$, $\bar{y}$).

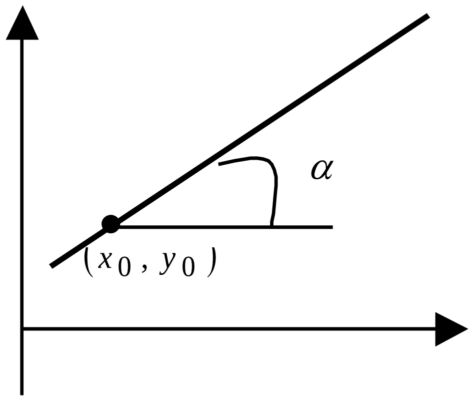


Рисунок 7

2 З виразу $b_1 = r_{xy} \dfrac{\sigma_y}{\sigma_x}$ витікає, що кутовий коефіцієнт b_1 виражається через коефіцієнт кореляції r_{xy} і середнє квадратичне відхилення фактора і відклику, тобто знак b_1, співпадає із знаком коефіцієнта кореляції (оскільки $\sigma_x, \sigma_y > 0$ завжди).

Якщо $r_{xy} > 0$, то $b_1 > 0$, кут α гострий (рис.8), зв'язок між *x* і *y* – прямий, тобто із зростанням *x* зростає *y*.

Якщо $r_{xy} < 0$, то $b_1 < 0$, α тупий, зв'язок між *x* і *y* зворотний (рис.9).

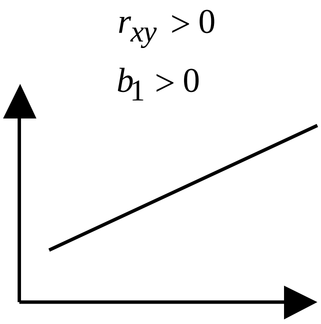


Рисунок 8

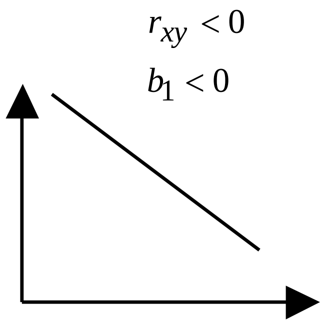


Рисунок 9

4 СТАТИСТИЧНІ КРИТЕРІЇ
ВІДОМОСТІ З МАТЕМАТИЧНОЇ СТАТИСТИКИ

Статистична гіпотеза – це припущення про вид розподілу випадкової величини або про значення числової характеристики випадкової величини.

Наприклад:

1 Висувається гіпотеза: випадкові відхилення e_i вибіркових значень y_i від розрахункових значень $\hat{y}_i = b_0 + b_1 x_i$ розподілені за нормальним законом. Це гіпотеза про вид розподілу.

2 Гіпотеза: дві вибіркові дисперсії $D_1 = \sigma_1^2$ і $D_2 = \sigma_2^2$ рівні між собою, тобто $\dfrac{\sigma_1^2}{\sigma_2^2} \approx 1 \Rightarrow D_1 \approx D_2$. Це гіпотеза про числові характеристики.

4.1 Нульова і конкуруюча гіпотези

Гіпотеза, висунута першою, називається *нульовою* і позначається H_0.

Наприклад, H_0: $b_1 = 0$ означає, що $y(x) = b_0$, тобто між y і x немає залежності.

Гіпотеза, протилежна гіпотезі H_0, називається *конкуруючою,* або альтернативною, і позначається H_1.

Наприклад, H_1: $b_1 \neq 0$.

4.2 Помилки 1- і 2-го роду

При перевірці виконання гіпотез можуть виникнути дві помилки.

Помилка 1-го роду полягає у тому, що буде відкинута правильна гіпотеза. Ймовірність відкинути правильну гіпотезу позначають α і називають *рівнем значущості гіпотези.* Звичайно беруть $\alpha = 0{,}01 \div 0{,}05$.

Наприклад: α = 0,05 означає, що в 5 випадках з 100 буде відкинута правильна гіпотеза.

Величина γ = (1–α) називається *рівнем довіри.*

Помилка 2-го роду полягає у тому, що буде прийнята неправильна гіпотеза.

4.3 Статистичні критерії перевірки нульової гіпотези

Статистичний критерій – це спеціально сконструйована випадкова величина.

Наприклад, для перевірки гіпотези про рівність двох дисперсій використовують критерій Фішера.

Критерій Фішера – це спеціально сконструйована випадкова величина, яка дорівнює відношенню двох дисперсій:

$$F = \frac{D_1}{D_2} = \frac{\sigma_1^2}{\sigma_2^2}.$$

Графік щільності розподілу ймовірності цієї випадкової величини наведений на рисунку 10.

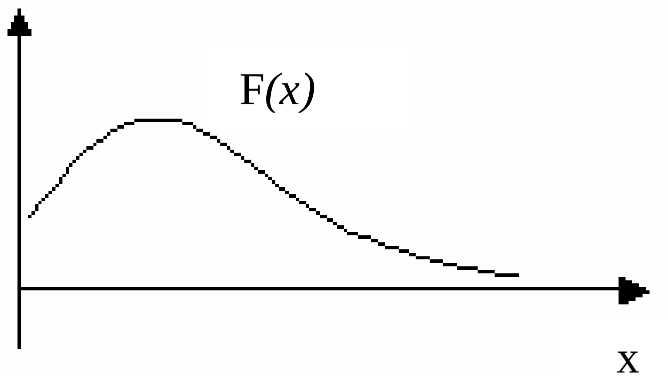


Рисунок 10

При цьому важливу роль виконує поняття кількості ступенів вільності.

4.4 Кількість ступенів вільності

Кількість ступенів вільності – це різниця між об'ємом вибірки, за якою обчислюється вибіркова чисельна характеристика, і кількістю зв'язків, накладених на вибіркові значення.

Приклад. Є вибірка об'єму n: $x_1, x_2, \dots x_n$. За нею обчислюється середнє значення $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Ця величина має n ступенів вільності .

Розглянемо вибіркову дисперсію

$$D = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2. \qquad (8)$$

Вибіркові значення x_i можна змінити (зменшити або збільшити), причому так, що дисперсія не зміниться, але на зміну значень x_i накладений один зв'язок – це вибіркове середнє $\bar{x}$. Воно входить до формули для обчислення дисперсії (8), і значить x_i повинне мінятися так, щоб $\bar{x}$ не змінювалося $\Rightarrow$ дисперсія D має (n – 1) ступенів вільності. Звичайно вибіркову дисперсію обчислюють за формулою

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

Оскільки до критерію Фішера входять 2 дисперсії – D_1 і D_2 – і кожна має свій ступінь вільності k_1 і k_2, то критерій Фішера залежить від двох ступенів вільності – k_1 і k_2.

Одержуємо функцію $F(x, k_1, k_2)$.

При збільшенні k_1 і k_2 розподіл наближається до нормального.

4.5 Значення критерію, що спостерігаються

Значення критерію, що спостерігаються, обчислюється за наявними даними. Припустимо, що перевіряється нульова гіпотеза H_0:

$$H_0 : D_1 = D_2 .$$

За вибіркою знаходять дисперсії D_1 і D_2 і відповідні їм ступені вільності k_1 і k_2. Значення критерію Фішера, що спостерігається: $F_{спост}\ (x, k_1, k_2)$.

4.6 Критичні точки

Для того, щоб прийняти або відкинути гіпотезу H_0, необхідно знати значення, що спостерігається, і критичне значення статистичного критерію.

Критичне значення – наперед розраховане значення критерію з певним рівнем значущості. Це значення визначається як абсциса на графіку щільності розподілу із заданим рівнем значущості α і ступенями вільності k_1 і k_2.

Критична точка має наступний сенс:

$$P\{F > F_{кр}\} = \alpha .$$

Наведемо рисунок, що пояснює критичне значення критерію (рис.11).

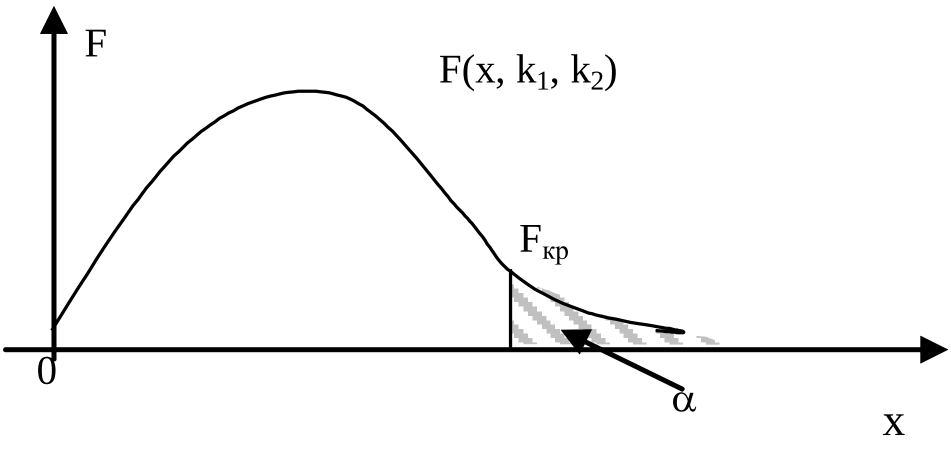


Рисунок 11

На графіку (див. рис.11) α – це площа заштрихованої області. Для критерію Фішера розраховані таблиці критичних точок $F_{кр}$. Кожному значенню α відповідає своя таблиця. Наприклад, для $\alpha = 0{,}05$ таблиця має вигляд таблиці 4.

Таблиця 4

k2\k1	1	…	10	…	∞
1	161		242		254
…					
10	4,97		2,47		2,54
…					
∞	3,84		1,83		1,00

Ці таблиці наводяться у підручниках з математичної статистики.

4.7 Критерій прийняття гіпотези

Прийняти або відкинути гіпотезу H_0 можна, порівнявши критичне і значення критерію, що спостерігається. Якщо воно *менше* за критичне, то гіпотеза H_0 *приймається*. Якщо значення критерію, що спостерігається, *більше* критичного, то гіпотеза H_0 *відкидається*.

Для цього потрібно знати рівень значущості α. У економіці, як правило, беруть $\alpha = 0{,}05$. Якщо зменшувати рівень значущості, то зростає ймовірність зробити помилку 2-го роду.

5 ПЕРЕВІРКА ЛІНІЙНОЇ РЕГРЕСІЇ НА АДЕКВАТНІСТЬ

Після того, як була побудована модель лінійної регресії $y = b_0+b_1x$, необхідно перевірити її на адекватність, тобто чи відповідає побудована модель наявним статистичним даним.

Спочатку розглянемо варіацію (розкид) залежного показника y щодо свого середнього значення. Відхилення дорівнює $y_i - \bar{y}$. Можна записати: $y_i - \hat{y}_i + \hat{y}_i - \bar{y}$, де $\hat{y}_i = b_0 + b_1 x_i$ – розрахункові значення. Тобто варіацію залежного показника У навколо свого середнього значення можна розподілити на два доданки: $\hat{y}_i - \bar{y}$ – варіація розрахункових значень навколо середнього; $y_i - \hat{y}_i$ – варіація розрахункових значень навколо фактичних.

Позначимо:

$\sigma_p^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ – варіація, що пояснюється регресією, з кількістю ступенів вільності $k_1 = 1$;

$\sigma_e^2 = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$ – залишки, непояснений розкид, з кількістю ступенів вільності $k_2 = n - 2$;

$\sigma_y^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2$ – загальна варіація з кількістю ступенів вільності $k_3 = n - 1$.

5.1 Коефіцієнт детермінації

Для аналізу загальної якості оціненої лінійної регресії звичайно використовують коефіцієнт детермінації:

$$R^2 = 1 - \frac{\sigma_e^2}{\sigma_y^2}, \qquad 0 \le R^2 \le 1.$$

У чисельнику стоїть сума квадратів відхилень лінії регресії від фактичних значень, у знаменнику – від середнього значення. Значить, чим менше відхилення розрахункових значень від фактичних, тим менше дріб, і тим ближче значення коефіцієнта детермінації до 1. Тому вважається, що чим ближче значення коефіцієнта детермінації до 1, тим краще модель описує статистичні дані.

Звичайно в економіці для варіаційних рядів величина коефіцієнта детермінації не перевищує 0,6…0,7. Вважається, що загальна якість такої моделі добра. Відповідь на питання про адекватність моделі R^2 не дає.

При додаванні нової змінної до рівняння регресії коефіцієнт R^2 зазвичай збільшується, але це не означає, що модель стала більш якісною. Тому розглядають ще виправлений коефіцієнт R^2_{adj}:

$$R^2_{adj} = 1 - \frac{n-1}{n-p-1}(1-R^2),$$

де p – кількість факторів. Але навіть збільшення R^2_{adj} не завжди означає, що якість моделі поліпшилася. Тому коефіцієнт детермінації використовують як один з показників успіху в специфікації моделі, але не головний.

У більшості математичних пакетів значення R^2 розраховується автоматично.

5.2 Перевірка моделі на адекватність за допомогою критерію Фішера

Якщо розглядається лінійна залежність *у* від фактора *х* вигляду у = $b_0 + b_1 x$, то можуть зустрітися ситуації, показані на рисунку 12.

Перевірка лінійної регресії на адекватність означає з'ясування наявності залежності у від *х*. На рисунку 12, *г* такої залежності немає. Фактично це означає, що кутовий коефіцієнт $b_1 = 0$. На рисунку 12, *а,б,в*, $b_1 \neq 0$.

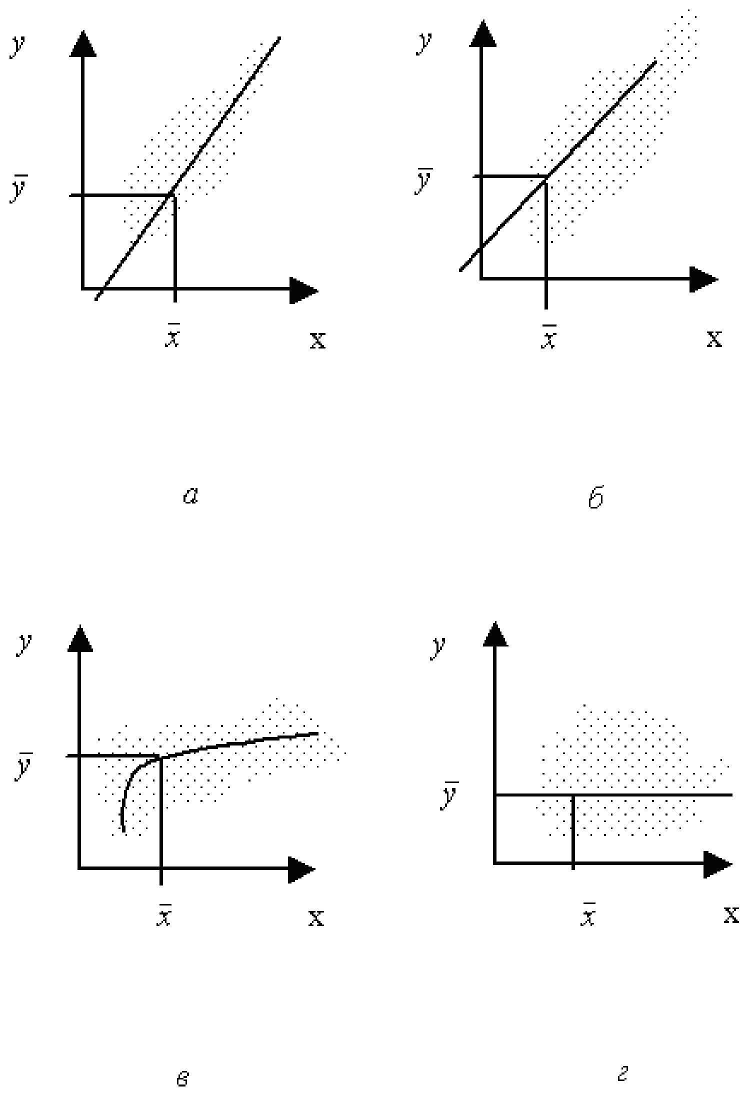


Рисунок 12

Постановка задачі

Висуваємо гіпотезу: H_0: ($b_1 = 0$). Рівняння регресії матиме вигляд $y = b_0 = \bar{y}$. Тобто функціональної залежності між x і y немає.

Для перевірки цієї гіпотези порівнюються між собою дві дисперсії:

$$D_1 = \frac{1}{k_1}\sigma_p^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \text{ і } D_2 = \frac{1}{k_2}\sigma_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 .$$

Тобто обчислюємо дисперсію залишків e_i і дисперсію розрахункових значень $\widehat{y}_i$, узятих з регресійної прямої (рис.13).

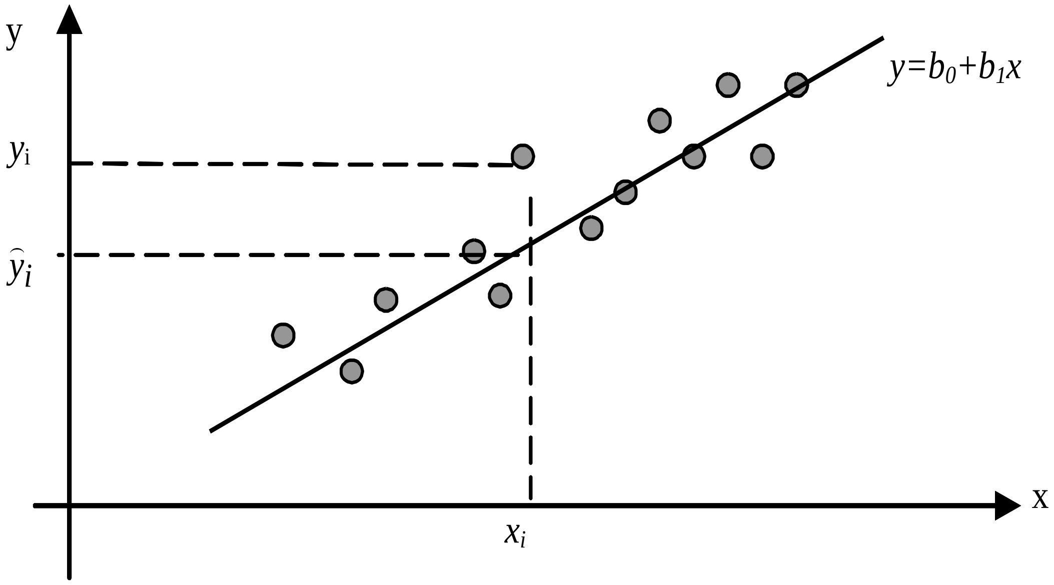


Рисунок 13

Обчислюємо k_1, k_2 – кількості ступенів вільності для статистик D_1 і D_2. Кількість ступенів вільності дисперсії D_2 дорівнює $k_2 = n – 2$ (n – об'єм вибірки).

Кількість ступенів вільності статистики D_1 для однофакторної регресії завжди дорівнює 1, оскільки пряма регресії завжди зобов'язана проходити через центр регресії, для неї можна тільки злегка змінити кут нахилу прямої.

Отже, відношення введених дисперсій є випадковою величиною, розподіленою за законом Фішера із ступенями k_1, k_2:

$$\frac{D_1}{D_2} = F(x, k_1, k_2) = F(x, 1, n-2).$$

Проаналізуємо, що дає відношення дисперсій на рисунку 12,г.

Оскільки $\widehat{y}_i$ береться з регресійної прямої, яка на рисунку 12, *г* – горизонтальна, то $\widehat{y}_i = \overline{y}$, тобто всі доданки у D_1 дорівнюють 0, і значення критерію Фішера, що спостерігається, теж дорівнює 0:

$$F_{спост}=\frac{D_1}{D_2}=0\ .$$

На рисунку 12, *в* $F_{спост} \approx 0$.

Перехід від випадку, коли можна визнати $F_{спост} = 0$ (а, отже, b1= 0, і залежність *y* від *x* відсутня), до випадку, коли слід визнати $F_{спост} \neq 0$ ($b_1 \neq 0$, тобто є залежність *y* від *x*), роблять, порівнюючи $F_{спост}$ з теоретично обчисленим критичним значенням для критерію Фішера $F_{кр}$ (див. п. 4.6).

Розраховують точку $F_{кр}$ при деякому рівні значущості гіпотези α. Якщо $F_{спост} < F_{кр}$, то робимо висновок, що $b_1 = 0$, отже, *y* від *x* *не* залежить, тому, модель неадекватна.

Якщо ж $F_{спост} > F_{кр}$, то гіпотеза H_0 відкидається, значить, $b_1 \neq 0$, *y* залежить від *x*, отже, модель $y = b_0 + b_1 x$ адекватна (з надійністю $(1 - \alpha) \times 100\%$).

Значення критерію Фішера, що спостерігається, можна записати через коефіцієнт детермінації:

$$F_{спост} = \frac{R^2}{1-R^2}(n-2).$$

Для багатофакторної регресії:

$$F_{спост} = \frac{R^2}{1-R^2} \cdot \frac{n-p-1}{p},$$

де p – кількість факторів.

5.3 Статистична значущість коефіцієнтів моделі

Знайдені за МНК параметри моделі b_0 і b_1 є не точними, а випадковими величинами. При зміні навіть однієї точки вибірки одержимо інші коефіцієнти b_0' і b_1' і т.д.

МНК гарантує, що знайдені за його допомогою параметри:

- *не зміщені*. Це означає, що b_0 і b_1 випадкові, оскільки знайдені за вибіркою, але в середньому вони такі, начебто вони були знайдені за генеральною сукупністю;

- *ефективні*. МНК забезпечує швидку збіжність параметрів моделі до точних значень, які можна було б розрахувати за генеральною сукупністю;

- *обґрунтовані*. Із збільшенням об'єму вибірки збільшується точність розрахованих параметрів.

Оскільки b_0 і b_1 є випадковими величинами, то необхідно перевірити їх статистичну значущість. Це можна зробити за допомогою спеціально сконструйованої статистики, розподіленої за законом Стьюдента.

Розподіл Стьюдента T(x, k) виникає кожного разу, коли порівнюються два математичні очікування (два середніх). Розподіл Стьюдента симетричний щодо початку координат (рис.14).

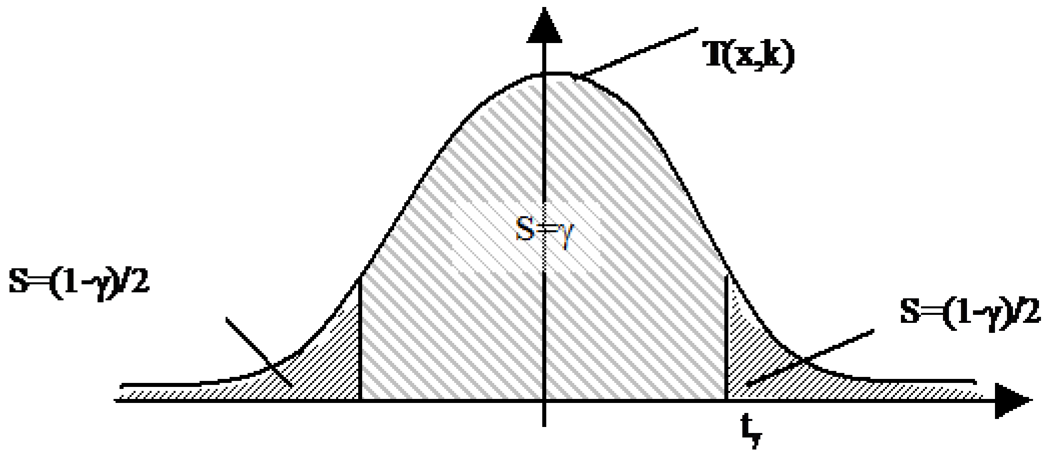


Рисунок 14

Кількість ступенів вільності для критерію Стьюдента k = n -2.

Розраховують для кожного коефіцієнта значення, що спостерігається, $t_{спост} = \frac{b_i}{\sigma_{b_i}}$, для всіх i = 0 ... p і порівнюють з критичним значенням критерію $t_{\text{кр}}$. Якщо $t_{спост} > t_{\text{кр}}$, то відповідний коефіцієнт статистично значущий. Якщо $t_{спост} < t_{\text{кр}}$, то відповідний коефіцієнт статистично не значущий (є статистичним "нулем").

Для даної однофакторної регресії:

$$t_{b_1} = \frac{b_1}{\sigma_{b_1}}, \text{ де } \sigma_{b_1} = \sqrt{\frac{\sigma_e^2/(n-2)}{\sum(x_i-\bar{x})^2}},$$

$$t_{b_0} = \frac{b_0}{\sigma_{b_0}}, \text{ де } \sigma_{b_0} = \sqrt{\frac{\sigma_e^2}{n-2}.\frac{\sum x_i^2}{n\cdot\sum(x_i-\bar{x})^2}}.$$

5.4 Перевірка статистичної значущості коефіцієнта кореляції

Коефіцієнт кореляції r_{xy}, розрахований за вибіркою, сам є випадковою величиною. Значить, необхідно перевірити статистичну значущість коефіцієнта кореляції. Цю перевірку виконують аналогічно перевірці статистичної значущості параметрів моделі b_0 і b_1 за допомогою критерію Стьюдента. Фактичне значення t – критерію Стьюдента визначається за формулою

$$t_r = \frac{r_{xy}}{\sqrt{1-r_{xy}^2}}.\sqrt{n-2},$$

якщо вибірка велика і $|r_{xy}|$ неблизька до 0, або за формулою

$$t_r = \frac{\sqrt{n-3}}{2}.\ln\left(\frac{1+r}{1-r}\right), \text{ якщо } |r_{xy}| \to 1.$$

Якщо $t_r > t_{kp}(\alpha, n-2)$, то коефіцієнт кореляції r_{xy} статистично значущий.

6 ПРОГНОЗ НА ПІДСТАВІ ЛІНІЙНОЇ РЕГРЕСІЇ

Якщо побудована модель $y = b_0 + b_1 x$ адекватна початковим статистичним даним, то за цією моделлю можна розрахувати прогноз у будь-якій точці x_{np} з області прогнозів. *Областю прогнозів* називається відрізок прямої між x_{min} і x_{max} (рис.15).

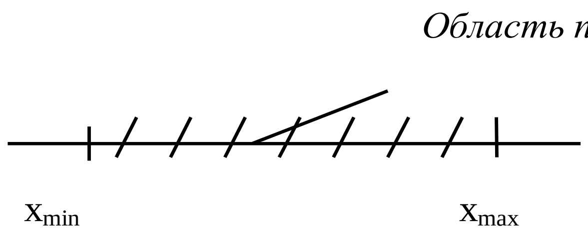


Рисунок 15

Такий прогноз $\hat{y}(x_{np}) = b_0 + b_1 x_{np}$ називається *точковим*.

6.1 Поняття про довірчий інтервал

Якби були дані зі всієї генеральної сукупності (X), то можна було б досить точно знайти статистичні характеристики, наприклад $\bar{x}^*$. Але, як правило, є вибірка, в якій біля десятка точок. За вибіркою розраховують вибіркове середнє $\bar{x}$.

Істинне значення $\bar{x}^*$ може бути як більше, так і менше за вибіркове $\bar{x}$, тобто точне значення $\bar{x}^*$ потрапляє в деякий інтервал, центром якого є вибіркове значення $\bar{x}$.

Якщо задатися ймовірністю γ (наприклад: 0,9; 0,99; 0,95) попадання $\bar{x}^*$ у інтервал, то чим більше буде задана ймовірність, тим ширше виходитиме інтервал. Якщо почати зменшувати γ, інтервал звужуватиметься.

Описаний інтервал називається *довірчим інтервалом*, а γ – коефіцієнтом довіри. Частіше всього на практиці беруть γ = 0,95. Це означає, що в 95% випадків точне значення параметру потрапить до інтервалу.

Довірчий інтервал – це інтервал, в який із заданою ймовірністю потрапить істинне значення невідомого параметра.

Коефіцієнт довіри – це ймовірність, з якою до довірчого інтервалу потрапить невідомий параметр.

6.2 Алгоритм знаходження напівширини довірчого інтервалу

За генеральною сукупністю для конкретного x можна було б досить точно знайти прогноз $y(x) = \beta_0 + \beta_1 x$. За вибіркою будується лінійна регресія $\hat{y} = b_0 + b_1 x$, і за $y(x)$ беруть $\hat{y}(x)$, зняте з прямої регресії.

Довірчий інтервал, до якого потрапляє невідоме $y(x)$ з деяким коефіцієнтом довіри γ, у разі лінійної регресії виявляється симетричним відносним $\hat{y}(x)$ (рис.16). Тому достатньо знайти напівширину довірчого інтервалу δ.

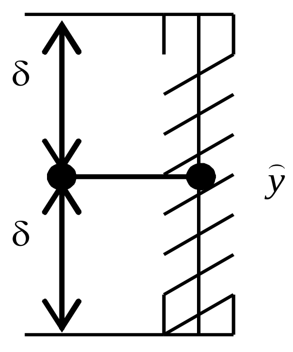


Рисунок 16

Напівширина довірчого інтервалу в точці прогнозу $x_{пр}$ обчислюється за формулою

$$\delta = \sigma_e \cdot t_\gamma \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{np} - \bar{x})^2}{\sum\limits_i (x_i - \bar{x})^2}},$$

де σ_e – середньоквадратичне відхилення вибіркових точок від лінії регресії $\sigma_e = \sqrt{\frac{1}{n-2}\sum e_i^2}$, тут $e_i = y_i - \hat{y}_i$;

t_γ – критична точка розподілу Стьюдента з кількістю ступенів вільності $k_2 = n-2$;

n – об'єм вибірки;

x_{np} – точка з області прогнозів.

Прогнозований довірчий інтервал для будь-якого x з області прогнозів записується $(\hat{y} - \delta, \hat{y} + \delta)$.

Сукупність довірчих інтервалів для всіх x з області прогнозів утворює довірчу область. Для лінійної однофакторної регресії вона симетрична щодо лінії регресії (рис.17). Найвужче місце довірчої області в точці $(\bar{x}, \bar{y})$.

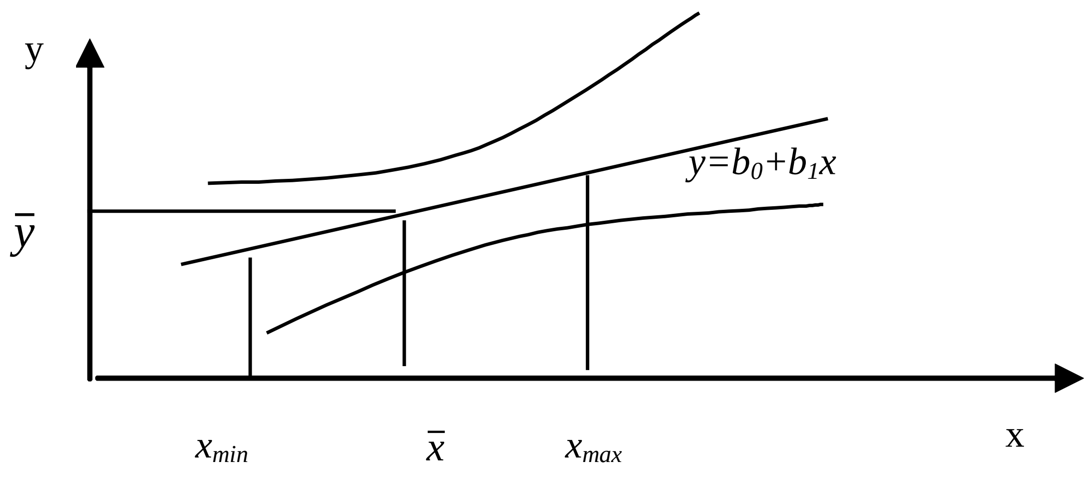


Рисунок 17

Прогноз для довільного x дає інтервал, в який з ймовірністю γ потрапляє невідоме значення $y(x)$. Тобто прогноз при заданому x складе від $\hat{y}-\delta$ до $\hat{y}+\delta$ з надійністю $\gamma \cdot 100\%$. Це прогноз з урахуванням довірчого інтервалу.

7 НЕЛІНІЙНА ОДНОФАКТОРНА МОДЕЛЬ

Багато економічних процесів не можуть бути адекватно описані лінійною залежністю вигляду $y = b_0 + b_1x$.

Прикладом таких економічних процесів можуть служити: життєвий цикл товарів, процес накопичення капіталу, маркетингові зусилля фірм та ін.

Найчастіше використовуються 5 нелінійних залежностей, які відрізняються від інших залежностей тим, що їх вдається лінеарізувати (звести до лінійних).

7.1 Види нелінійних залежностей

1 *Степенева залежність:* $y = Ax^b$.

Криві можуть мати вигляд, показаний на рисунках 18, *а* і 18, *б*.

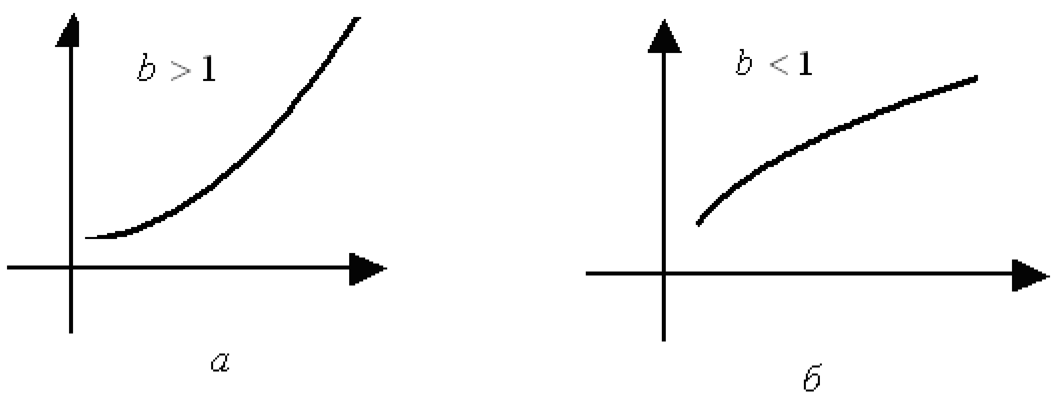


Рисунок 18

Прологарифмуємо y = Ax^{b}:

Lny=ln(Ax^{b}) $\Rightarrow$ Lny=LnA+Lnx^{b} $\Rightarrow$ Lny=LnA+bLnx.

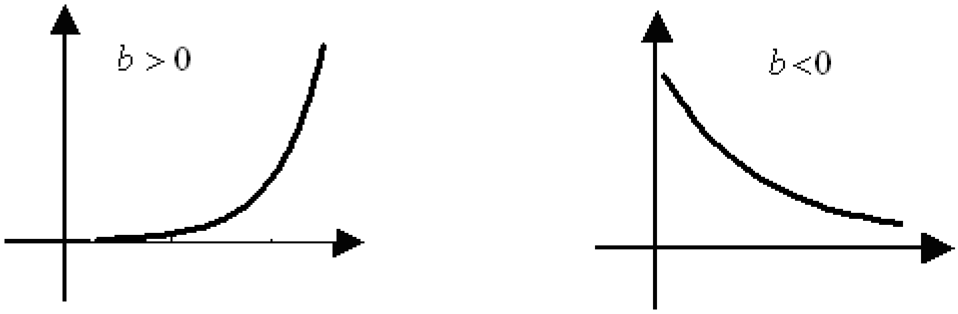
Позначимо V = Lny; u = Lnx; b_0 = LnA і b_1 = b. Одержимо лінійну модель від нових змінних: V = b0 + b1u.

Зворотне перетворення: V = Lny $\Rightarrow$ y = e^{v} $\Rightarrow$ e^{b0+b1u} $\Rightarrow$ $e^{b0}e^{b1u}$ $\Rightarrow$ $e^{b0}(e^{lnx})^{b1}$ $\Rightarrow$ $e^{b0}x^{b1}$.

Значить, A = e^{b0}, b = b1, y=e^{v}.

2 *Експоненційна залежність*: y = A·e^{bx}.

Експоненційні криві можуть мати вигляд, показаний на рисунках 19, *а* і 19, *б*.



а *б*

Рисунок 19

Щоб її лінеарізувати, прологарифмуємо рівняння $y = \text{A}e^{bx}$.

Lny = ln(e^{bx}) $\Rightarrow$ Lny = LnA + Lne^{bx} $\Rightarrow$ Lny = LnA + bxLne $\Rightarrow$ Lny = LnA + bx.

Позначимо V = Lny; u = x; b_0 = LnA і b_1 = b. Одержимо V=b_0 + b_1u.

Зворотне перетворення:

V = Lny $\Rightarrow$ y = e^{v} $\Rightarrow$ e^{b0+b1u} $\Rightarrow$ $e^{b0}e^{b1u}$ $\Rightarrow$ $e^{b0}e^{xb1}$.

Значить, A = e^{b0}, b = b_1, y = e^{v}.

3 *Логарифмічна залежність*: y = A + Bln$x$.

Крива може мати вигляд, показаний на рисунку 20.

Зробимо заміну: $V = y$; $u = \mathrm{Ln}x$; $b_0 = A$ і $b_1 = B$. Одержали $V = b_0 + b_1 u$.

Прикладом використовування такої функції може служити взаємозв'язок частки витрат на товари тривалого користування і загальних сум витрат або доходів.

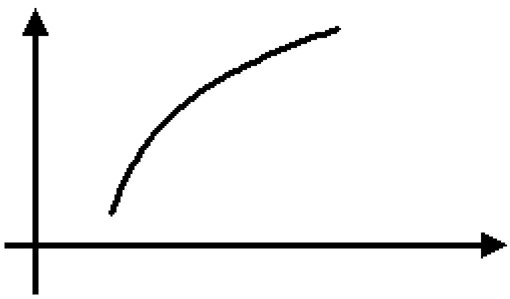


Рисунок 20

Математичний опис подібного роду взаємозв'язків одержав назву кривих Енгеля.

4 *Зворотна залежність*: $y = A + B\frac{1}{x}$.

Крива може мати вигляд, показаний на рисунку 21.

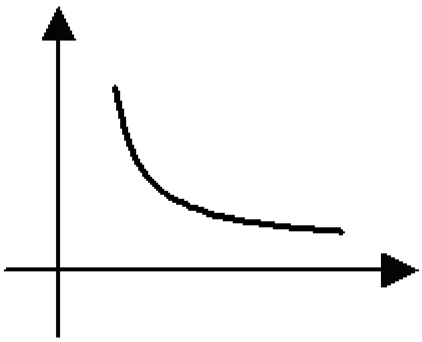


Рисунок 21

Зробимо заміну: $V = y$; $u = \frac{1}{x}$; $b_0 = A$ і $b_1 = B$. Одержимо $V = b_0 + b_1 u$.

Ця модель достатньо відома в економетриці. Класичним її прикладом є крива Філіпса, що характеризує нелінійне співвідношення між нормою безробіття x і відсотком приростом заробітної плати y: $y(x) = 0{,}00679 + \frac{0{,}1842}{x}$. Величина параметра $b_0 = 0{,}00679$ означає, що із зростанням рівня безробіття темп приросту заробітної плати в межі прагне до нуля.

5 *Логістична крива*: $y = \frac{1}{A + Be^{-x}}$.

Крива може мати вигляд, показаний на рисунку 22.

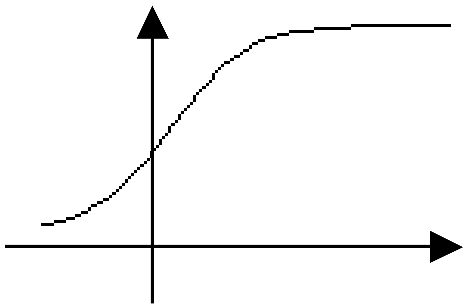


Рисунок 22

Зробимо заміну: V $= \frac{1}{y}$; u $= e^{-x}$; b_0 =A і b_1=B. Одержимо V=b_0 + b_1u.

Нелінійні залежності можна розподілити на:

- регресії, нелінійні за оцінюваними параметрами: y =Ax^b; y =A·e^{bx}; $y = \frac{1}{A + Be^{-x}}$;

- регресії, нелінійні за оцінюваними факторами: y = A + Bln$x$; $y = A + B\frac{1}{x}$.

7.2 Алгоритм побудови нелінійних економетричних моделей

1 Є вибірка (табл. 5), щодо якої є економічні міркування про вид залежності між x і y: y=f(x).

Таблиця 5

X	x_1	x_2	…	x_n
У	y_1	y_2	…	y_n

2 Якщо вид залежності відомий, то, використовуючи відповідну заміну, перераховуємо значення вибірки і одержуємо нову вибірку (табл. 6), за якою можна побудувати лінійну модель:

$$v = b_0 + b_1 u. \tag{9}$$

Таблиця 6

U	u_1	u_2	…	u_n
V	v_1	v_2	…	v_n

3 Знайдену модель (9) перевіряємо на адекватність. Якщо вона адекватна, то і початкова нелінійна модель адекватна. Якщо лінеарізована модель неадекватна, то початкова модель вибрана невірно, і потрібно підібрати іншу нелінійну модель (наприклад, замість степеневої моделі спробувати експонеційну).

4 Якщо лінеарізована модель (9) адекватна, то в тих точках, в яких потрібно порахувати прогноз, розраховуємо напівширину довірчого інтервалу δ_u для лінеарізованої моделі.

5 Перераховуємо прогноз і довірчий інтервал для точки прогнозу з лінеарізованого вигляду в початковий нелінійний. Для цього знаходимо межі довірчого інтервалу для лінеарізованої моделі (v_{min}, v_{max}) і для них, а також для значень $\hat{v}$, за допомогою зворотного перетворення знаходимо y_{min}, y_{max} і $\hat{y}$.

Для нелінійної регресії довірчий інтервал може бути несиметричним щодо лінії регресії.

7.3 Гетероскедастичність

Розглянемо лінію регресії, побудовану за кореляційними полями різного вигляду (рис.23).

На рисунку 23, *а* дисперсія залишків e_i *не* залежить від *x*, тобто $D(e_i) = const$. Такий розподіл називається *гомоскедастичним*. У випадках, поданих на рисунку 23, *б...г*, дисперсія залишків непостійна. Таке явище називається *гетероскедастичністю***.**

Основною властивостю гетероскедастичності є дуже широкий довірчий інтервал і, як наслідок, широка довірча область, що призводить до втрати економічного значення прогнозування на підставі такої моделі (рис.23).

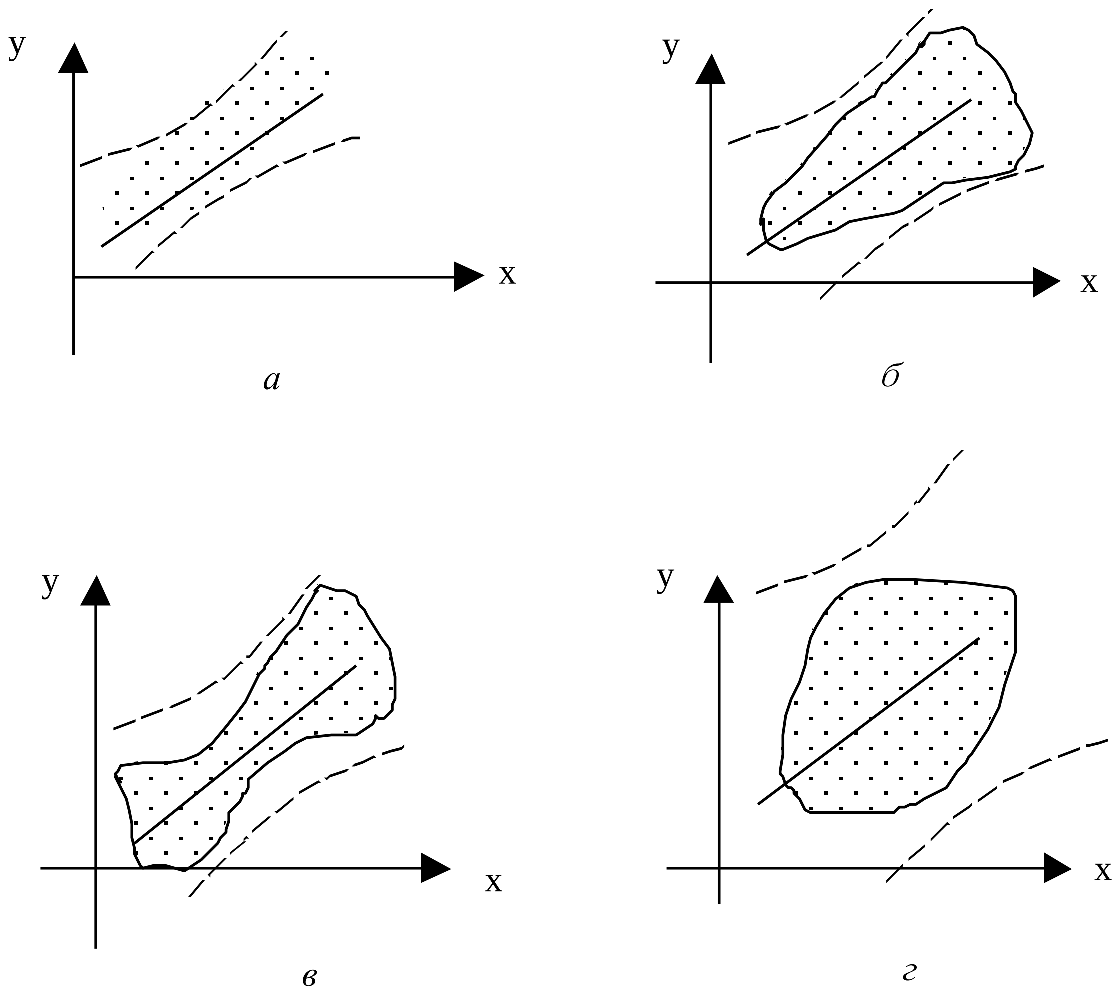


Рисунок 23

7.4 Узагальнений метод найменших квадратів

За наявності гетероскедастичності в залишках рекомендується традиційний метод найменших квадратів (МНК) замінювати узагальненим методом найменших квадратів (УМНК).

Припускатимемо, що:

- середнє значення залишкових величин дорівнює нулю;
- дисперсія залишкових величин не залишається незмінною для різних значень фактора, а пропорційна деякій величині K_i, тобто

$$\sigma_{\varepsilon_i}^2 = \sigma^2 \cdot K_i ,$$

де $\sigma_{\varepsilon_i}^2$ – дисперсія помилки на конкретному (і-му) значенні фактора;

σ^2 – постійна дисперсія помилки при дотриманні передумови про гомоскедастичність залишків;

K_i – коефіцієнт пропорційності, що змінюється із зміною величини фактора, що і обумовлює неоднорідність дисперсії.

При цьому вважається, що величина σ^2 невідома, а відносно величини K_i висуваються певні гіпотези, що характеризують структуру гетероскедастичності.

У загальному вигляді рівняння регресії набуде вигляду:

$$\frac{y_i}{\sqrt{K_i}} = \frac{\alpha}{\sqrt{K_i}} + \beta \frac{x_i}{\sqrt{K_i}} + \varepsilon_i .$$

Початкові дані для цього рівняння набудуть вигляду:

$$y=\begin{vmatrix}\frac{y_1}{\sqrt{K_1}}\\ \frac{y_2}{\sqrt{K_2}}\\ \dots\\ \frac{y_n}{\sqrt{K_n}}\end{vmatrix},\quad x=\begin{vmatrix}\frac{x_1}{\sqrt{K_1}}\\ \frac{x_2}{\sqrt{K_2}}\\ \dots\\ \frac{x_n}{\sqrt{K_n}}\end{vmatrix}.$$

За відношенням до звичайної регресії рівняння з новими, перетвореними змінними, є зваженою регресією, в якій змінні x і y узяті з вагами $1/\sqrt{K}$.

Оцінка параметрів нового рівняння з перетвореними змінними приводить до методу найменших квадратів, для якого необхідно мінімізувати суму квадратів відхилень вигляду

$$S=\sum_{i=1}^{n}\frac{1}{K_i}\cdot\left(y_i-a-b\cdot x_i\right)^2\rightarrow\min .$$

8 БАГАТОФАКТОРНА РЕГРЕСІЯ

8.1 Поняття багатофакторної моделі та етапи її побудови

Реальні економічні процеси, як правило, залежать не від одного, а від декількох факторів:

$$y = f\,(x_1, x_2, \ldots, x_p). \qquad (10)$$

Процес знаходження залежності (10) і отримання корисних висновків з цієї залежності такий:

1 Знайомство з економічною теорією, висунення гіпотез про вид взаємозв'язку. Постановка задачі.

2 Специфікація моделі – теоретичні уявлення і ухвалення гі-

потези у вигляді математичних рівнянь, які встановлюють зв'язок між незалежними змінними.

3 Формування масивів вхідної інформації згідно з задачами дослідження.

4 Дослідження початкових статистичних даних на наявність залежності між факторами (автокореляція), впливу відклику y на фактори x_i (авторегресія), ефектів, пов'язаних із запізнюванням реакції ринку (лагові ефекти).

5 Оцінка параметрів моделі на підставі статистичних даних.

6 Перевірка одержаної моделі на адекватність.

7 Прогноз на підставі рівняння регресії і розрахунок довірчого інтервалу для прогнозу.

8.2 Специфікація моделі

Економетрична модель базується на об'єднанні двох аспектів – теоретичного, якісного аналізу і дослідницькій інформації. Теоретична інформація знаходить своє відображення у специфікації моделі. Специфікація моделі – це аналітична форма економетричної моделі.

З досвіду економетричних досліджень можна навести приклади функцій, які можуть описувати взаємозв'язок між показником У і факторами X_i:

1 Лінійна функція $Y = b_0 + b_1 X_1 + \dots + b_p X_p$.

2 Ступенева функція

$$Y = a_0 X_1^{a_1} \cdot \dots \cdot X_p^{a_p} \to \ln y = \ln a_0 + a_1 \ln X_1 + \dots + a_p X_p \quad \text{або}$$

$$V = b_0 + b_1 U_1 + \dots + b_p U_p.$$

3 Гіперболічна функція

$$Y = a_0 + \frac{a_1}{X_1} + \dots + \frac{a_p}{X_p} \to y = a_0 + a_1 U_1 + \dots + a_p U_p, U_i = \frac{1}{X_i}.$$

4 Квадратична функція

$$Y = b_0 + b_1 X_1^2 + \cdots + b_p X_p^2 \to Y = b_0 + b_1 U_1 + \cdots + b_p U_p, \quad U_i = X_i^2.$$

У всіх моделях У – залежна змінна, показник; X_i – незалежні змінні, фактори; b_i – параметри моделі.

Лінійні функції є достатньо поширеними, і наведені вище нелінійні функції також зводяться до лінійних, тому далі розглядатимемо побудову лінійної моделі.

Маючи на увазі, що вибір аналітичної форми економетричної моделі не може розглядатися без переліку конкретних факторів, специфікація моделі передбачає передусім відбір факторів для економетричного аналізу.

На цьому етапі потрібне добре розуміння економічного процесу. Заздалегідь аналізується, від чого залежить *у*. Наприклад, об'єм продажів може залежати від реклами, іміджу, середнього заробітку людей у регіоні і т.ін. Збираються статистичні дані. Вибіркові відомості оформляються у вигляді таблиці 7. Після виявлення всіх факторів x_i потрібно виключити ті фактори, для яких мало спостережень.

Таблиця 7

Номер спостереження	**У**	$\mathbf{X_1}$	$\mathbf{X_2}$	**…**	$\mathbf{X_p}$
1	y_1	x_{11}	x_{12}		x_{1p}
2	y_2	x_{21}	x_{22}		x_{2p}
			…		
n	y_n	x_{n1}	x_{n2}		x_{np}

Помилки специфікації можуть бути наступними:

1) Ігнорування важливого фактора при побудові економетричної моделі;

2) введення в модель фактора, який істотно не впливає на показник;

3) вибір невідповідної математичної форми залежності.

Зважаючи на чітку інтерпретацію параметрів найбільш широко використовуються лінійна і степенева функції. У лінійній багатофакторній моделі $Y = b_0 + b_1 X_1 + \cdots + b_p X_p$ параметри b_i $(i = 1...p)$ називають коефіцієнтами “чистої” регресії. Вони характеризують середню зміну результату із зміною відповідного фактора на одиницю при незмінному значенні інших факторів, закріплених на середньому рівні.

*Приклад***.** Хай залежність витрат на продукти харчування за сукупністю сімей характеризується наступним рівнянням:

$$\hat{y}(x_1, x_2) = 0.5 + 0.35 x_1 + 173 x_2,$$

де у – витрати сім'ї за місяць на продукти харчування, грн.;

x_1 – місячний дохід на 1 члена сім'ї, грн.;

x_2 – розмір сім'ї (кількість людей).

Аналіз цього рівняння дозволяє зробити висновки:

- із зростанням доходу на одного члена сім'ї на 100 грн. витрати на продукти харчування зростають у середньому на 35 грн. при цьому ж середньому складі сім'ї;

- збільшення розміру сім'ї на 1 людину при тих же її доходах припускає додаткове зростання витрат на харчування на 173 грн.

Ступеневі функції $\hat{Y} = a_0 X_1^{a_1} \cdot \cdots \cdot X_p^{a_p}$ набули найбільшого поширення у виробничих функціях, в дослідженнях попиту і споживання.

*Приклад***.** Хай при дослідженні попиту на м'ясо одержана модель: $\hat{y} = 0.82\, x_1^{-2.63} x_2^{1.11}$, де $\hat{y}$ – попит на м'ясо, x_1 – ціна, x_2 – дохід. Звідси витікає, що зростання цін на 1% при тому ж доході викликає зниження попиту в середньому на 2,63%. Збільшення доходу на 1% при незмінній ціні зумовлює зростання попиту на 1,11%.

Якщо ці функції не влаштовують дослідника, то сучасні комп'ютерні програми дозволяють перебирати інші лінеаризовані функції і вибирати найбільш відповідну з них. При цьому слід пам'ятати, що, чим складніша функція, тим менше можливість інтерпретувати її параметри.

8.3 Аналіз факторів на мультиколінеарність

Однією з умов застосування МНК є те, що фактори мають бути незалежними один від одного. Проте очевидно, що в економіці дуже важко виділити такий масив даних, які були б зовсім незалежні один від одного. Тому кожного разу необхідно з'ясовувати, чи не впливає залежність факторів на оцінку параметрів моделі.

Фактори X_1, X_2, …, X_p інтерпретуються як вектори розмірності n:

$$\vec{X}_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \cdots \\ x_{ni} \end{pmatrix}, \ \text{i = 1...p.}$$

Два вектори колінеарні, якщо їх координати пропорційні:

$$\vec{X}_1 \text{II} \vec{X}_2 \Leftrightarrow \frac{x_{11}}{x_{12}} = \frac{x_{21}}{x_{22}} = \cdots \frac{x_{n1}}{x_{n2}}. \qquad (11)$$

Якщо виконується умова (11), то коефіцієнт кореляції між факторами X_1 і X_2 дорівнює 1:

$$\vec{X}_1 \text{II} \vec{X}_2 \Leftrightarrow r_{X_1 X_2} = 1. \qquad (12)$$

Фактори X_1 і X_2 строго колінеарні.

На практиці строга колінеарність зустрічається рідко. Більш поширений випадок, коли колінеарність не строга:

$$r_{X_i X_J} \approx 1.$$

Оскільки однією з умов побудови множинної регресії є незалежність дії факторів, тобто $r_{X_i X_J} \neq 1$, колінеарність факторів чинників порушує цю умову. Якщо фактори явно колінеарні, то вони дублюють один одного, і один з них рекомендується виключити з регресії. Перевага при цьому віддається не фактора, тісніше пов'язаному з результатом, а тому фактора, який при достатньо тісному зв'язку з результатом має якнайменшу тісноту зв'язку з іншими факторами. У цій вимозі виявляється специфіка множинної регресії як методу дослідження комплексної дії чинників в умовах їх незалежності один від одного.

Хай, наприклад, при вивченні залежності $Y = f(X_1, X_2, X_3)$матриця парних коефіцієнтів кореляції виявилася наступною таблицею 8.

Таблиця 8

	У	X_1	X_2	X_3
У	1			
X_1	0,8	1		
X_2	0,7	**0,8**	1	
X_3	0,6	0,5	0,2	1

Очевидно, що фактори X_1 і X_2 дублюють один одного. До подальшого аналізу доцільно включити фактор X_2, а не фактор X_1, оскільки кореляція X_2 з третім фактором X_3 слабкіша: $r_{X_2 X_3} < r_{X_1 X_3}$. Тому в даному випадку до рівняння множинної регресії включаються фактори X_2, X_3: $Y = f(X_2, X_3)$.

За величиною парних коефіцієнтів кореляції виявляється лише явна колінеарність факторів. Найбільші труднощі у використовуванні апарату множинної регресії виникають за наявності мультиколінеарності факторів, коли має місце сукупна дія факторів один на одного.

Мультиколінеарністю називається лінійна залежність між факторами X_1, X_2, …, X_p. Вектор є лінійно залежним, якщо його можна записати як лінійну комбінацію решти векторів:

$$\vec{X}_1 = \lambda_2 \vec{X}_2 + \lambda_3 \vec{X}_3 + \cdots + \lambda_p \vec{X}_p, \qquad (13)$$

причому хоча б один множник $\lambda_i \neq 0$.

Колінеарність є окремим випадком мультиколінеарності.

Однією з можливостей виявлення міжфакторної кореляції є перехід до суміщених рівнянь регресії, які відображають не тільки вплив факторів, але і їх взаємодію, наприклад, не

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3,$$

а

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{123} x_1 x_2 x_3.$$

Це рівняння враховує взаємодію всіх факторів. Частина доданків може бути відкинута, якщо взаємодія між деякими факторами виявиться неістотною.

8.4 Наслідки мультиколінеарності та способи її усунення

Якщо в умові задачі присутня строга мультиколінеарність, тобто формула (13) виконується точно, то при визначенні коефіцієнтів рівняння лінійної регресії

$$\widehat{Y} = b_0 + b_1 X_1 + \cdots + b_p X_p$$

за МНК виникає невизначеність типу $\frac{0}{0}$, тобто коефіцієнти визначити неможливо – задача не має рішення.

Якщо мультиколінеарність не досконала, тобто $\vec{X}_1 \approx \lambda_2 \vec{X}_2 + \lambda_3 \vec{X}_3 + \cdots + \lambda_p \vec{X}_p$, то невизначеність $\frac{0}{0}$ не виникає. Коефіцієнти $b_0, b_1, \ldots b_p$ за МНК можливо визначити, але середньоквадратичні погрішності дуже великі. Як наслідок: довірчі інтервали дуже широкі, та прогнози втрачають практичну цінність.

Найпростіший спосіб усунення мультиколінеарності в економетричній моделі – виключення одного з факторів. Проте якщо усліпу усувати мультиколінеарність за запропонованим алгоритмом, не замислюючись над економічним значенням задачі, можна вихолостити модель настільки, що вона втратить цінність.

Найпростіший спосіб усунення мультиколінеарністі – це перетворення даних:

1 Узяти $u_i = x_i - \bar{x}$ – відхилення від середнього.

2 Замість абсолютних значень узяти відносні.

3 Нормалізувати фактор $u_{ik} = \dfrac{x_{ik} - \bar{x}_k}{\sigma_{x_k}}, i = 1...n.$

4 Замінити одну змінну на іншу.

Складніші способи усунення мультиколінеарністі див.[5].

Якщо жоден із способів не дає можливості усунути мультиколінеарність, то МНК використовувати не рекомендується.

8.5 Знаходження регресійної моделі

Лінійна багатофакторна модель має вигляд:

$$Y = b_0 + b_1 X_1 + \cdots + b_p X_p. \tag{14}$$

Параметри лінійної багатофакторної моделі визначаються за МНК аналогічно тому, як визначають параметри лінійної однофакторної моделі. Тільки одержуємо систему *p* рівнянь. Цю систему вирішують спеціальними методами матричної алгебри.

На цьому етапі аналізу з'являються наступні *відомості*:

1 Коефіцієнт множинної кореляції

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}}. \tag{15}$$

За допомогою цього коефіцієнта оцінюється практична значущість рівняння багатофакторної регресії. Він оцінює щільність сумісного впливу факторів на результуючий показник.

2 Коефіцієнт детермінації R^2.

3 Уточнений коефіцієнт детермінації з урахуванням ступенів вільності

$$R_{adj}^2 = 1 - \left(1 - R^2\right)\frac{(n-1)}{(n-p)}.$$

4 Значення критерію Фішера, що спостерігається, і ступеня вільності для нього

Критерій Фішера у разі багатофакторної регресії має той же сенс, що і у разі однофакторної регресії, тобто відношення двох дисперсій, проте кількість ступенів вільності міняється: $k_1 = p, k_2 = n - p - 1$. Тут p – кількість факторів у моделі, n – об'єм вибірки. Для багатофакторної регресії гіпотеза H_0: всі коефіцієнти $b_i = 0$, окрім b_0.

5 Статистична значущість коефіцієнтів b_i.

Отримуємо модель:

$$Y = b_0 + b_1 X_1 + \cdots + b_p X_p.$$

В одержаній моделі $Y = b_0 + b_1 X_1 + \cdots + b_p X_p$ коефіцієнти b_i – випадкові величини. Їх математичні очікування при виконанні деяких умов рівні, відповідно, точним значенням β_i. При цьому оцінки тим надійніші, чим менше їх розкид навколо точних значень, тобто дисперсія. Можна довести, що

$$D(b_i) = \sigma_{b_i}^2 = \frac{S^2}{\sum_i (x_i - \bar{x})^2}, i = 1, \ldots, p,$$

$$D(b_0) = \sigma_{b_0}^2 = \frac{S^2 \sum_i x_i^2}{n \sum_i (x_i - \bar{x})^2},$$

де $S^2 = \frac{\sum_i e_i^2}{n-p-1}$ – дисперсія залишків.

Формально значущість (відмінність від нуля) коефіцієнта b_i може бути перевірена за допомогою критерію Стьюдента. Обчислюють $t_{спост} = \frac{b_i}{\sigma_b}$. Якщо $t_{спост} < t_{кр}$, то коефіцієнт b_i є статистичним нулем.

У багатьох пакетах програм статистична значущість коефіцієнтів перевіряється автоматично.

Однак, може статися, що в багатофакторній моделі кілька коефіцієнтів виявляться незначущими, у той час як сама модель є адекватною. Це може означати, що внесок до пояснення кожної з цих змінних незначущий, а спільна пояснююча здатність змінних є високою. Тобто ми не можемо оцінити вплив кожної змінної окремо, але можемо оцінити їхній сумарний ефект.

8.6 Прогноз на підставі лінійної моделі

Одна з найважливіших цілей моделювання полягає в прогнозуванні поведінки досліджуваного об'єкта. Звичайно термін "прогнозування" використовується в тих ситуаціях, коли вимагається передбачити стан системи в майбутньому. Для регресійних моделей він має ширше значення. Дані, як в даному випадку, можуть не мати часової структури, але може ставитися задача оцінити значення залежної змінної для деякого набору незалежних змінних, яких немає в початкових спостереженнях.

Так само, як і у разі однофакторної регресії, розрізняють точкове та інтервальне прогнозування. У першому випадку оцінка – це конкретне число, в другому – це інтервал, в якому із заданим рівнем довіри міститься істинне значення шуканої змінної.

Прогноз робиться за рівнянням регресії

$$\widehat{Y} = b_0 + b_1 X_1 + \cdots + b_p X_p. \tag{16}$$

Точка прогнозу $\vec{X}$ з координатами $(x_1;\ x_2;\ \ldots\ ; x_p)$ з *p-мірного* простору вибирається з області прогнозу, яка визначається системою нерівностей

$$\begin{cases} x_{1_{\min}} \le x_1 \le x_{1_{\max}}; \\ x_{2_{\min}} \le x_2 \le x_{2_{\max}}; \\ \ldots \\ x_{p_{\min}} \le x_p \le x_{p_{\max}}. \end{cases}$$

Якщо, наприклад, модель двофакторна $\widehat{Y} = b_0 + b_1 X_1 + b_2 X_2$, то область прогнозу визначається прямокутником, показаним на рисунку 24.

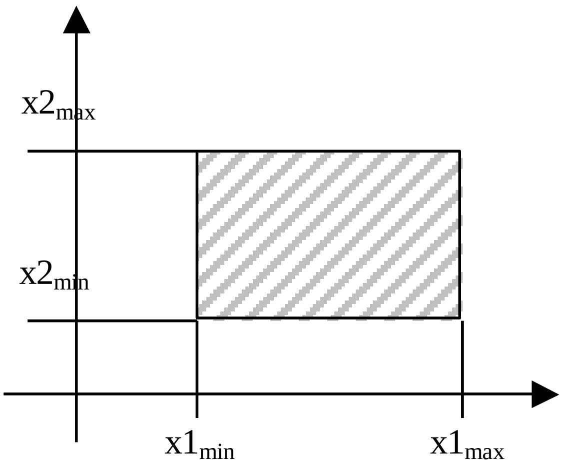


Рисунок 24

Прогноз, обчислений за моделлю, так, як і для однофакторної регресії, називається точковим.

Для того, щоб отримати інтервальну оцінку, треба перейти до матричної форми запису моделі.

Розглянемо рівняння (16). Воно побудоване за даними таблиці 7. Підставляючи ці дані в рівняння, отримаємо систему:

$$\begin{cases} y_1 = b_0 + b_1 x_{11} + b_2 x_{12} + ... + b_p x_{1p} + e_1, \\ \quad . \quad . \quad . \quad . \quad . \\ y_n = b_0 + b_1 x_{n1} + b_2 x_{n2} + ... + b_p x_{np} + e_n. \end{cases} \qquad (17)$$

Кожне з цих рівнянь є рівнянням площини. $e_i \ (i = 1..n)$ - випадкові відхилення, наявність яких пояснюється тим, що точки вибірки не лягають точно на площину, а випадковим чином розкидані навколо неї:

$$e_i = y_i - \hat{y}_i = y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2} + ... + b_p x_{ip}), (i = 1..n).$$

Для того, щоб записати систему (17) у матричній формі, вводимо:

- матрицю X розмірності $n \times (p+1)$, складену з множників при коефіцієнтах $b_0, b_1 ... b_p$

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & ... & x_{1p} \\ 1 & x_{21} & x_{22} & ... & x_{2p} \\ . & . & . & . & . \\ 1 & x_{n1} & x_{n2} & ... & x_{np} \end{pmatrix};$$

- вектори-стовпці розмірності $n \times 1$

$$\bar{Y} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}, \ \bar{E} = \begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ e_n \end{pmatrix};$$

- вектор-стовпець коефіцієнтів розмірності $(p+1) \times 1$

$$\overline{B} = \begin{pmatrix} b_0 \\ b_1 \\ . \\ . \\ b_p \end{pmatrix}.$$

Тепер рівняння лінійної багатофакторної регресії можна записати так:

$$\overline{Y} = X \cdot \overline{B} + E. \qquad (18)$$

Напівширина довірчого інтервалу для прогнозованого значення показника $\widehat{Y}$ буде розраховуватися за формулою

$$\delta = \sigma_e \cdot e_\gamma \cdot \sqrt{1 + \overline{X}^T \cdot (X^T \cdot X)^{-1} \cdot \overline{X}},$$

де $t_\gamma = t_\gamma(\alpha, k_2)$, $k_2 = n - p - 1$, $\overline{X} = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_p \end{pmatrix}$ - точка з області прогнозів.

9 ПОНЯТТЯ ПРО ЕЛАСТИЧНІСТЬ ЕКОНОМІЧНИХ МОДЕЛЕЙ

Еластичністю в економіці називають здатність показника відкликуватися на зміну того або іншого фактора.

9.1 Коефіцієнт еластичності для однофакторної моделі

Хай в точці x_0 фактор зміниться на величину Δx_0. Відповідно, y зміниться на величину $y_0+\Delta y_0$. Відносна зміна x дорівнює $\frac{\Delta x_0}{x_0}$. У відсотках (темп зростання) $T_{x_0} = \frac{\Delta x_0}{x_0} \cdot 100\%$. Відповідно, для y: $T_{y_0} = \frac{\Delta y_0}{y_0} \cdot 100\%$. Щоб взнати, на скільки відсотків зміниться y при зміні x на величину Δx_0, візьмемо відношення

$$\frac{T_{y_0}}{T_{x_0}} = \frac{\frac{\Delta y_0}{y_0} \cdot 100\%}{\frac{\Delta x_0}{x_0} \cdot 100\%} = \frac{\Delta y_0}{\Delta x_0} \cdot \frac{x_0}{y_0}.$$

Оскільки (x_0, y_0) – довільна точка, то можна записати: $\frac{T_y}{T_x} = \frac{\Delta y}{\Delta x} \cdot \frac{x}{y}$.

При $\Delta x \to 0$ $\frac{\Delta y}{\Delta x} = y'_x$.

Звідси $E_x = \frac{x}{y} y'_x$ – формула для розрахунку коефіцієнта еластичності.

Коефіцієнт еластичності показує, на скільки відсотків збільшиться (якщо $E_x > 0$) або зменшиться (якщо $E_x < 0$) показник y, якщо фактор x зміниться на 1% .

Обчислимо коефіцієнт еластичності для деяких моделей:

1 Лінійна:

$$y = b_0 + b_1x \Rightarrow y' = b_1 \Rightarrow E_x = \frac{b_1 x}{b_0 + b_1 x}.$$

Оскільки коефіцієнт еластичності для лінійної функції не є величиною постійною, а залежить від значення фактора *x*, то звичайно розраховують середній показник еластичності за формулою

$$E_x = b_1 \cdot \frac{\bar{x}}{\bar{y}}.$$

2 Ступенева:

$$y = Ax^b \Rightarrow y' = A \cdot bx^{b-1} \Rightarrow E_x = \frac{xAbx^{b-1}}{Ax^b} = b(const).$$

Для ступеневої моделі коефіцієнт еластичності постійний і дорівнює показнику ступеня.

3 Експонерційна:

$$y = Ae^{bx} \Rightarrow y' = A \cdot be^{bx} \Rightarrow E_x = \frac{xAbe^{bx}}{Ae^{bx}} = bx.$$

Оскільки коефіцієнти еластичності викликають економічний інтерес, наведемо формули розрахунку коефіцієнтів еластичності для найбільш поширених типів рівнянь регресії (табл. 9).

Таблиця 9

Вид функції y(x)	**Похідна** $y'(x)$	**Коефіцієнт еластичності** $E_x = \frac{x}{y(x)} y'_x$
$y = b_0 + b_1 x$	b_1	$\frac{b_1 x}{y(x)}$
$y = a\sqrt{x} + b$	$\frac{a}{2\sqrt{x}}$	$\frac{a\sqrt{x}}{2 \cdot y(x)}$
$y = ax^b$	abx^{b-1}	b

Продовження таблиці 9

Вид функції y(x)	**Похідна** $y'(x)$	**Коефіцієнт еластичності** $E_x = \frac{x}{y(x)} y'_x$
$y = \frac{a}{x} + b$	$\frac{-a}{x^2}$	$\frac{-a}{a+bx}$
$y = \frac{1}{ax+b}$	$\frac{-a}{(ax+b)^2}$	$-ax \cdot y(x)$
$y = ax^2 + b$	$2ax$	$\frac{2ax^2}{y(x)}$
$y = be^{ax}$	abe^{ax}	ax
$y = a\ln x + b$	$\frac{a}{x}$	$\frac{a}{y(x)}$

Наприклад: зв`язок між середньою заробітною платою за місяць (y, грн.) і питомою вагою робітників із технічною підготовкою (x, %) описаний моделлю: y=100+1,1x. Знайдемо, як зміниться середня заробітна плата при збільшенні питомої ваги з 40% на 1%.

Рішення: $y(x=40)=100+1{,}1\cdot 40=144$(грн.).

Для лінійної моделі

$$E_x = \frac{b_1 x}{y(x)} = \frac{1{,}1 \cdot 40}{100 + 1{,}1 \cdot 40} = 0{,}306.$$

Це означає, що при збільшенні питомої ваги робітників з технічною підготовкою з 40% на 1% (40%+40×1%=40,4%) середня заробітна плата за місяць збільшиться на 0,306%. Тобто, від 144грн на 144×0,306%=0,44грн.

Не дивлячись на широке використовування в економетриці коефіцієнтів еластичності, можливі випадки, коли їх розрахунок економічного значення не має. Це відбувається у тих випадках, коли для даних ознак не має сенсу визначення зміни значень у відсотках (наприклад, якщо x – стаж роботи, вимірюваний у роках, або x – якість ґрунту, вимірювана в балах).

9.2 Коефіцієнт еластичності для багатофакторних моделей

Якщо показник залежить від декількох факторів, то, використовуючи коефіцієнт еластичності, можна визначити ступінь впливу кожного фактора на показник.

$E_{x_i} = \dfrac{x_i}{y} y'_{x_i}$ – формула для розрахунку коефіцієнта часткової еластичності.

Коефіцієнт часткової еластичності показує, на скільки відсотків зміниться y при зміні фактора x_i на 1% при інших незмінних значеннях факторів.

Наприклад: для моделювання залежності між об'ємом випущеної продукції y, трудовитратами x_1 і об'ємом основних засобів виробництва x_2 використовується функція Кобба-Дугласа $y = Ax_1^{\alpha_1} x_2^{\alpha_2}$. Припустимо, що ми отримали модель $y = 0{,}1x_1^{1{,}62} x_2^{0{,}46}$. Так як ця функція ступенева, то коефіцієнти часткової еластичності дорівнюють ступеню відповідного фактора: $E_{x_1} = \alpha_1 = 1{,}62,\ E_{x_2} = \alpha_2 = 0{,}46$. Перший коефіцієнт 1,62>1, тому фактор x_1 є еластичним. Він показує, що при зміні фактора x_1(праця) на 1% відклику(обсяг виробленої продукції) збільшиться на 1,62% при незмінному значенні фактора x_2(капітал). Другий коефіцієнт 0,46<1, тому фактор x_2 є нееластичним. Він показує, що при зміні обсягу основних засобів виробництва (x_2) на 1% значення обсягу виробленої продукції збільшиться на 0,46% при незмінному значенні трудовитрат.

Ми бачимо, що $\alpha_1 > \alpha_2$ (1,62>0,46). Це означає, що x_1 більше впливає на y, тобто, якщо збільшувати трудовитрати, то обсяг виробленої продукції зростає швидше, ніж при збільшенні основних засобів виробництва.

Сума коефіцієнтів $\alpha_1 + \alpha_2 = 1{,}62 + 0{,}46 = 2{,}08 > 1$. Це означає, що виробництво, для якого побудована ця модель, має зростаючі масштаби, тобто його виробництво є перспективним.

10 СИСТЕМИ ОДНОЧАСНИХ РІВНЯНЬ

При моделюванні складних економічних об'єктів часто доводиться вводити не одне, а декілька зв'язаних між собою рівнянь, тобто описувати модель системою рівнянь.

Наприклад, проста макроекономічна кейнсіанська модель споживання може бути подана в наступному вигляді:

$$\begin{cases} C = a + by + \varepsilon, \\ y = C + I, \end{cases}$$

де C – особисте споживання в постійних цінах,

y – національний дохід в постійних цінах,

I – інвестиції в постійних цінах,

ε – випадкова складова.

Наявність зв'язку між C і y, що входять до рівнянь, вимагає коректування методу найменших квадратів для оцінювання параметрів моделі a і b.

Є два види таких моделей: система зовні не зв'язаних між собою рівнянь (вони зв'язані тільки наявністю кореляції між помилками) і система одночасних рівнянь. Система сумісних, одночасних рівнянь звичайно містить ендогенні та екзогенні змінні.

Ендогенні змінні – це залежні змінні, які визначаються внутрішньою структурою досліджуваного економічного процесу.

Екзогенні змінні – це змінні, які не залежать від внутрішньої структури досліджуваного економічного процесу. Вони впливають на ендогенні змінні, але не залежать від них.

Розподілення змінних на ендогенні та екзогенні залежить від теоретичної концепції прийнятої моделі. Екзогенні змінні можуть виступати в одних моделях як ендогенні, а в інших як екзогенні змінні.

Система одночасних, взаємозалежних рівнянь вигляду

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \ldots + b_{1n}y_n + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2n}y_n + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ y_n = b_{n1}y_1 + b_{n2}y_2 + b_{n3}y_3 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n, \end{cases} \tag{19}$$

називається також *структурною формою моделі.*

У системі одні й ті ж змінні (y) одночасно розглядаються як залежні змінні в одних рівняннях і як незалежні в інших. Кожне рівняння системи не може розглядатися самостійно, тому традиційний МНК не можна застосувати.

Коефіцієнти b_{ij} і $a_{ik}(i = 1..n, j = 1..n, k = 1..m)$ називають *структурними коефіцієнтами моделі.*

Використовування МНК для оцінювання структурних коефіцієнтів моделі дає зміщені та необґрунтовані оцінки. Тому звичайно для визначення структурних коефіцієнтів модель перетворюють у наведену форму моделі, яка є системою лінійних функцій ендогенних змінних від екзогенних:

$$\begin{cases} \hat{y}_1 = \delta_{11}x_1 + \delta_{12}x_2 + \ldots + \delta_{1m}x_{m,} \\ \hat{y}_2 = \delta_{21}x_1 + \delta_{22}x_2 + \ldots + \delta_{2m}x_{m,} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \hat{y}_n = \delta_{n1}x_1 + \delta_{n2}x_2 + \ldots + \delta_{nm}x_{m,} \end{cases} \tag{20}$$

де $\delta_{ik}(i = 1...n, k = 1...m)$ – коефіцієнти наведеної моделі.

Наведена форма моделі за своїм виглядом схожа на систему незалежних рівнянь, параметри якої визначаються за МНК. Коефіцієнти моделі (20) є нелінійними функціями коефіцієнтів структурної форми моделі.

Приступати до оцінювання коефіцієнтів того або іншого рівняння в системі (19) має сенс лише після того, як встановлена його ідентифікованість.

Ідентифікованість – це єдність відповідності між приведеною і структурною формами моделі.

Для цього необхідно, щоб виконувалася нерівність

$$n_s - 1 \leq m - m_s, \tag{21}$$

де n_s – кількість ендогенних змінних, які входять до s-е рівняння структурної форми моделі;

m – загальна кількість екзогенних змінних моделі, які входять до s-е рівняння структурної форми моделі.

Нерівність (21) називається *порядковою (рахунковою) умовою* і є необхідною умовою ідентифікованості рівняння.

Якщо у виразі (21) стоїть знак рівності, то модель точно *ідентифікована*, її структурні коефіцієнти визначаються однозначно, єдиним чином за коефіцієнтами наведеної форми моделі. Кількість параметрів структурної моделі дорівнює кількості параметрів наведеної форми моделі.

Якщо нерівність (21) виконується не строго, то модель *надідентифікована*. У цьому випадку кількість коефіцієнтів у наведеній моделі більше кількості коефіцієнтів структурної моделі, тобто один і той же структурний коефіцієнт допускає різні вирази через коефіцієнти наведеної форми. Це звужує область його застосування як з теоретичної, так і з практичної точки зору.

Якщо нерівність (21) не виконується, то модель *неідентифікована* і структурні коефіцієнти не можуть бути оцінені через коефіцієнти наведеної форми моделі.

Виконання умови ідентифікації перевіряється для кожного рівняння системи. Для оцінки параметрів структурної моделі система має бути ідентифікована або надідентифікована.

Порядкова умова (21) є необхідною, але не достатньою умовою ідентифікації. Рівняння ідентифіковано, якщо за відсутніми у ньому змінними (ендогенними і екзогенними) можна з коефіцієнтів при них в інших рівняннях системи одержати матрицю, визначник якої відмінний від нуля, а ранг матриці не менше, ніж кількість ендогенних змінних у системі без одного. Ця умова називається *ранговою* і є достатньою умовою ідентифікації. Докладніше про ідентифікацію див.[7].

Коефіцієнти структурної моделі можуть бути оцінені різними способами залежно від виду системи одночасних рівнянь. Найпоширенішими є наступні два методи:

1 Непрямий метод найменших квадратів. При ідентифікованості рівняння оцінки структурних коефіцієнтів можна знайти, оцінивши методом найменших квадратів наведену форму моделі. Потім коефіцієнти наведеної форми перетворюють у параметри структурної моделі.

2 Для оцінки параметрів надідентифікованої моделі використовується двокроковий метод найменших квадратів (2МНК). Основна ідея 2МНК – на основі наведеної форми моделі набути для надідентифікованого рівняння теоретичні значення ендогенних змінних, що містяться у правій частині рівняння. Далі, підставивши їх замість фактичних значень, застосувати МНК до структурної форми надідентифікованого рівняння.

11 ЧАСОВІ РЯДИ

11.1 Складові ряду

Часовим (динамічним) рядом називається послідовність спостережень за одним економічним процесом через деякі (бажано рівні) проміжки часу.

Наприклад: курс $, рівень інфляції, будь-яка звітність.

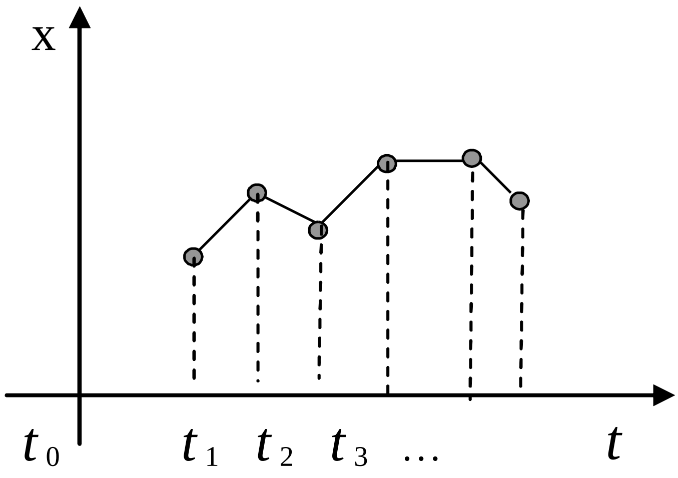


Рисунок 25

Часовий ряд (рис.25) неможливо досліджувати як регресійну модель за методом найменших квадратів з двох причин.

По-перше, в регресійній моделі ми припускали, що показник у залежить від поточного значення *x*. З точки зору економіки це означає, що показник миттєво реагує на зміну фактора. Але це не завжди так. Показник часто на зміну фактора реагує з запізненням (наприклад, через інертність самої маси людей і т.ін.).

Розглянемо такий приклад [8].

Кількість рекламацій за місяцями на деталі А і В, що поступили в автомобілебудівну фірму, наведено в таблиці 10.

Таблиця 10

Кількість рекламацій на:	Місяць											
	1-й	2-й	3-й	4-й	5-й	6-й	7-й	8-й	9-й	10-й	11-й	12-й
деталь A(x)	105	102	100	108	112	115	118	116	120	125	125	128
деталь B(y)	68	71	69	66	65	70	75	76	78	77	79	82

Якщо побудувати кореляційне поле для *x* і *y*, отримаємо (рис.26).

$r_{xy} = 0.81$

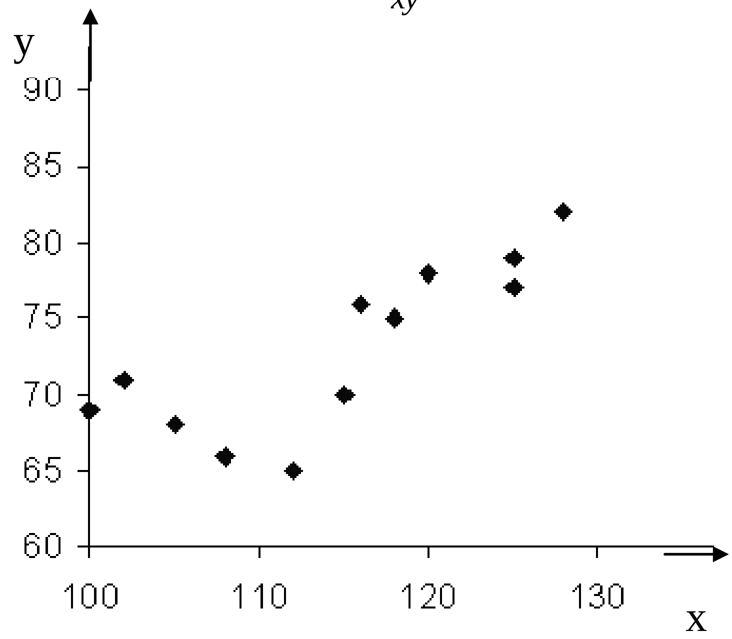


Рисунок 26

Подивимось, що трапиться, якщо здвинути відповідність, тобто розглянути пари точок зі зсувом на 1 місяць: $(x_1, y_2), (x_2, y_3) \ldots (x_{11}, y_{12})$. Отримаємо інше кореляційне поле (рис.27).

$r_{xy} = 0.8978$

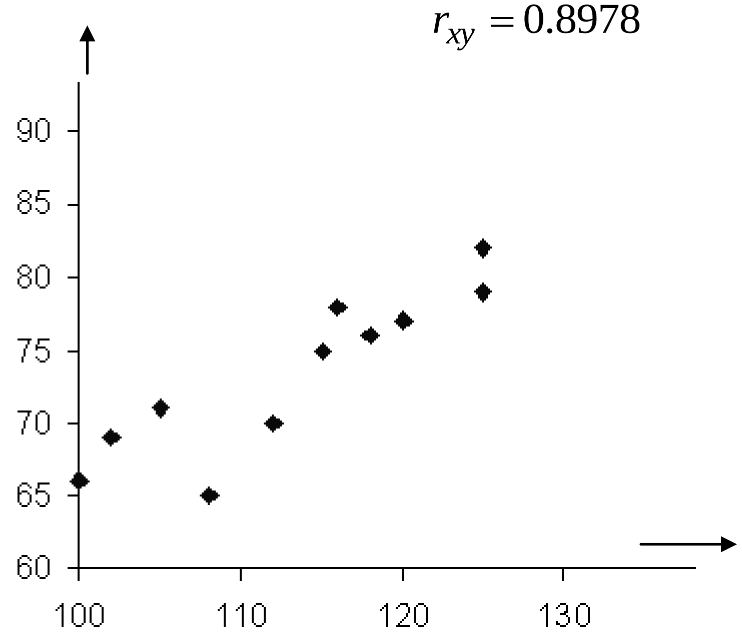


Рисунок 27

Такий «зсув» називається лагом. Якщо задати лаг у 2 і 3 місяці, то отримаємо кореляційні поля, зображені на рисунках 28 і 29.

$r_{xy} = 0.993$

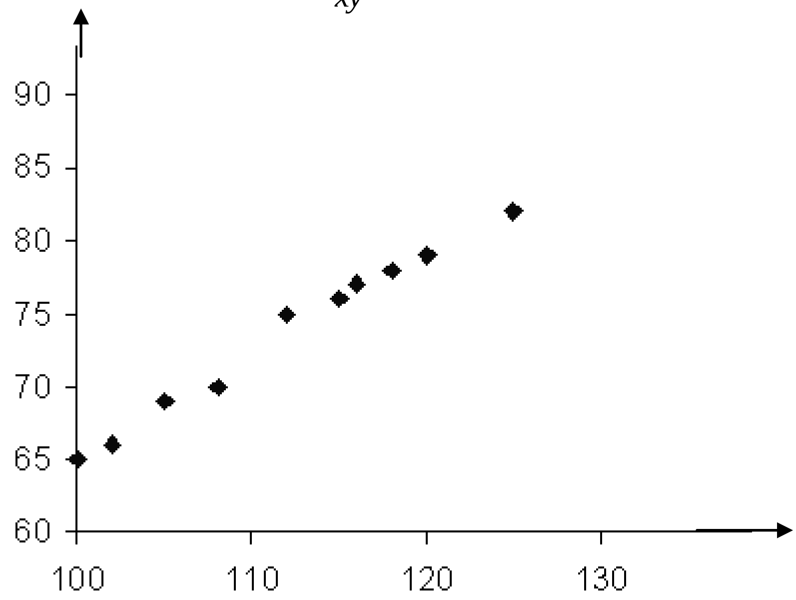


Рисунок 28

$r_{xy} = 0.8971$

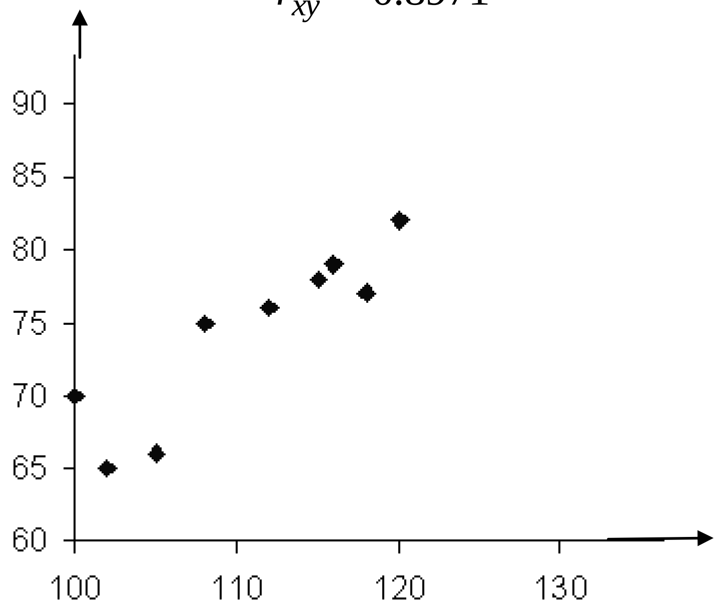


Рисунок 29

Із порівняння цих чотирьох діаграм розсіювання і відповідних коефіцієнтів кореляції можна зробити висновок, що найбільша кореляція буде при часовому лагу за 2 місяці. Іншими словами: рекламації на деталь В добре корелюють з рекламаціями на деталь А, що надійшли за 2 місяці до них.

Лаг – економічний показник, що відбиває відставання у часі одного економічного показника у порівнянні з іншим, зв`язаним з ним. Якщо показник x відстає на s періодів, то він записується x_{t-s}. Такі моделі економічних процесів, що протікають у часі і включають x_t і x_{t-s}, називають *дистрибутивно-лаговими*. Вони мають вигляд:

$$y_t = a_0 + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \ldots + e_t,$$

де e_t - випадковий член.

Якщо модель містить відгук із запізненням, то модель є *авторегресійною*. Такі моделі можуть мати вигляд:

$$y_t = a_0 + b_0 y_t + b_1 y_{t-1} + b_2 x_{t-2} + e_t,$$
$$y_t = a_0 + b_1 y_{t-1} + b_2 x_{t-2} + e_t,$$
$$y_t = a_0 + b_0 y_t + b_1 y_{t-1} + b_2 y_{t-2} + \ldots + e_t.$$

Як правило, члени рівняння, що відстоять у часі далі, на показник впливають менше.

Друга причина, через яку неможливо досліджувати часовий ряд як регресійну модель, полягає в тому, що часові ряди, крім основної тенденції мають ще й так звані сезонні коливання, коливання розвитку. Їх не можна відкидати, бо вони потрібні для прогнозування. Якщо описати такий ряд регресійною моделлю, то разом з помилками можна погасити й коливання.

Часовий ряд містить регулярну частину R_t, яку треба виділити, і випадкову e_t, яку треба погасити.

Регулярна частина, в свою чергу, має головну тенденцію (тренд T_t), сезонні коливання (S_t) і тенденцію глобальних циклів (C_t).

Тренд – тривала тенденція зміни економічних показників, основна складова прогнозованого часового ряду, на яку накладаються сезонні коливання. Ці тенденції можуть взаємодіяти один з одним по-різному. У теорії часових рядів розглядають 2 моделі:

1) дистрибутивну: $x_t = T_t + S_t + C_t + e_t$,

2) мультиплікативну: $x_t = T_t \cdot S_t \cdot C_t \cdot e_t$.

11.2 Методи аналізу рядів

При дослідженні часового ряду потрібно по-перше, згладити ряд. По-друге, зазначити регулярні складові, задати їх рівняннями, одержати теоретичний ряд. Ціллю згладжування є "гасіння" помилок.

Два найбільш поширених методи згладжування рядів – це метод ковзких середніх і метод експоненційного згладжування.

Метод ковзних середніх полягає в тому, що кілька значень часового ряду, які йдуть одне за одним, замінюються їхнім середнім значенням.

Нехай ми маємо часовий ряд: $x_1, x_2, \ldots, x_n$. Задаємо деяке значення k (k<n). Для кожних k послідовних значень ряду можна розрахувати середнє значення: $y_1 = \dfrac{x_1 + x_2 + \ldots + x_k}{k}$. Далі переходимо до розрахунку середнього для наступних значень ряду: $y_2 = \dfrac{x_2 + x_3 + \ldots + x_{k+1}}{k}$ і т.д. Кількість даних k називають порядком ковзної середньої.

Основна ідея такого згладжування – замінити фактичні дані часового ряду розрахунковими даними. Ряд стає коротшим, але більш гладким.

Ковзні середні можна використовувати не тільки для згладжування ряду, але й для прогнозування: k останніх даних і буде прогнозом для наступної дати.

З підвищенням порядку згладжування k ряд стає все більш гладким. Так як часові ряди містять сезонні коливання, то для більшості рядів найбільш гладким він стає при k=12.

Цей метод має свої недоліки:

1) якщо ряд має викиди, то вони погано гасяться;
2) відбувається втрата (k-1) значень ряду;
3) якщо порядок k є парним, то дати розрахункових значень не співпадають з датами реальних спостережень;
4) вплив кожного доданку в розрахунку середнього значення вважається однаковим, хоч зрозуміло, що останні значення ряду несуть більше інформації, ніж початкові.

Метод експоненційного згладжування також полягає в тому, що кілька послідовних значень часового ряду усереднюються ковзною середньою, але з вагою $(w_1 x_t + w_2 x_{t-1} + w_3 x_{t-2})$, де $w_1 + w_2 + w_3 = 1$. Вагові коефіцієнти w_1, w_2, w_3 беруть у вигляді експоненційних функцій. Це дозволяє надавати більшої ваги останнім значенням часового ряду, ніж початковим.

ЧАСТИНА 2

ВСТУП

При обробці вибірок кожного виду використовується специфічний математичний апарат (методи математичної статистики і методи аналізу випадкових процесів). Але у будь-якому випадку така обробка зв'язана з громіздкими і трудомісткими обчисленнями. Тому необхідно використовувати математичні пакети, спеціально призначені для обробки статистичних даних. У середовищі Excel for Windows є спеціальні можливості для таких розрахунків.

1 КОРОТКІ ВІДОМОСТІ ПРО ЕКОНОМЕТРИЧНИЙ АНАЛІЗ В ПАКЕТІ EXCEL

1.1 Настройка пакета аналізу

Для проведення економетричного аналізу в пакеті Excel має бути встановлений «Пакет аналізу» (рис.30). Шлях: **Сервіс – Настройки – Пакет аналізу – Ок.**

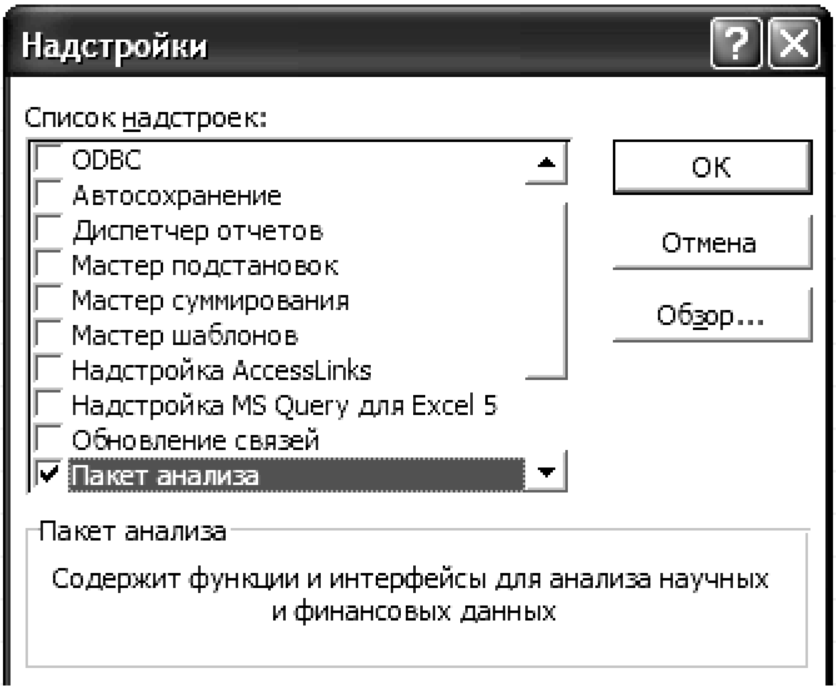


Рисунок 30

Після цього в меню **Сервіс** додасться рядок **Аналіз даних**.

1.2 Введення даних

Початкові дані вводяться на робочий лист пакету Excel (табл. 11).

Таблиця 11

Рядки Excel	**Стовпці Excel**	
	A	**B**
1	x	y
2	6630	39,4
3	2340	42,79
4	8492	38,93
5	8540	38,34
6	2901	46,96
7	7224	40,69
8	1920	46,05
9	2569	43,5
10	3520	56,11
11	2911	44,69
12	6921	40,15
13	7671	40,44
14	1586	69,76
15	3223	42,99
16	5410	39,48

1.3 Побудова діаграми розсіювання (кореляційного поля)

За початковими даними будується діаграма розсіювання за допомогою «Майстра діаграм», тип діаграми – точкова.

Діаграма форматується так, щоб чіткіше уявлялися початкові дані (рис.31).

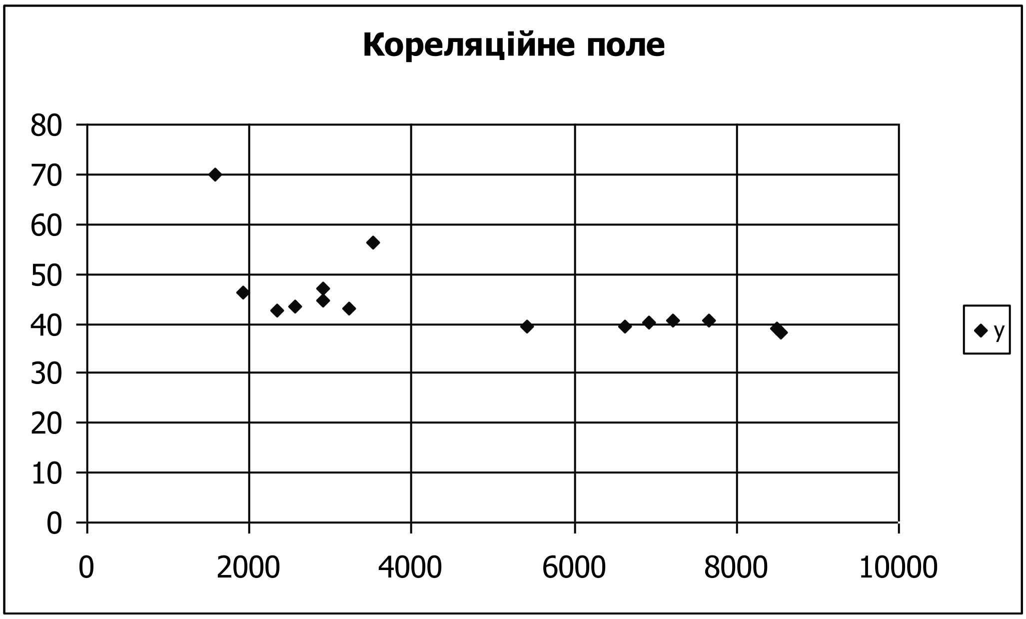


Рисунок 31

1.4 Знаходження коефіцієнта кореляції

Вибирається пункт меню **Сервіс – Аналіз даних – Кореляція**.

Задається вхідний інтервал для X і У – A1:B16 (групування даних – за стовпцями), встановлюється прапорець у віконці «Мітки» (це означає, що в першому рядку – імена даних – X і У), «Вихідний діапазон» – на новий лист вказується вихідний інтервал на початковому листі (рис.32).

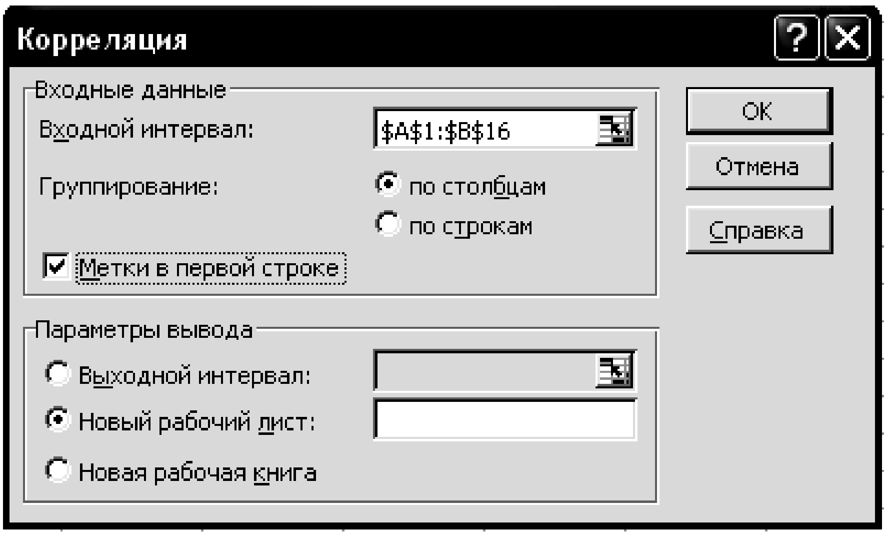


Рисунок 32

Зауваження:

1 Для двофакторної регресії виділяється весь діапазон даних (Х1, Х2, У).

2 Для однофакторної регресії «Вихідний діапазон» – виділити блок 3 на 3, для двофакторної – 4 на 4 осередки.

3 Одержана матриця симетрична щодо головної діагоналі.

Одержуємо матрицю наступного вигляду (табл.12) для *однофакторної* регресії.

Таблиця 12

	X	Y
X	1	
Y	-0,61975	1

Відповідність:

	X	Y
X	r_{xx}	
Y	r_{xy}	r_{yy}

1.5 Знаходження основних числових характеристик

Щоб знайти основні числові характеристики, вибираємо пункт меню **Сервіс – Аналіз даних – Описова статистика**.

Задаємо вхідний інтервал для Х і У – А1:В16; встановлюємо прапорець у віконцях «Мітки» і «Підсумкова статистика»; «Вихідний діапазон» – на новий лист або вказати вихідний інтервал (блок з 15 рядків і 4 стовпців для однофакторної регресії) на початковому листі (рис.33).

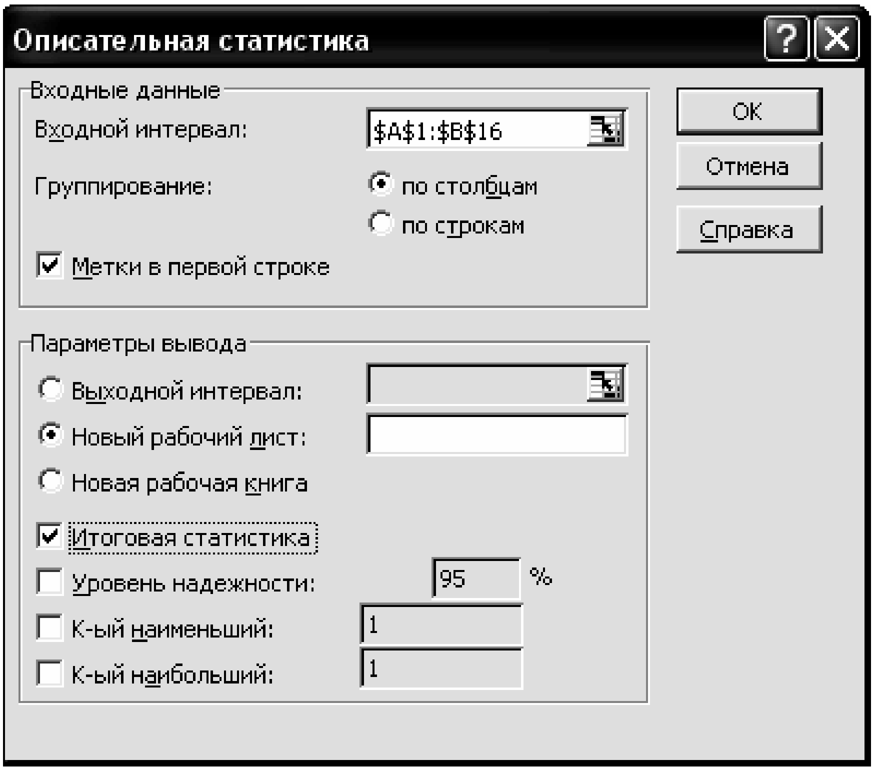


Рисунок 33

Отримуємо наступну таблицю 13 для однофакторної регресії.

Таблиця 13

	X	У	Пояснення
Середнє	4790,533333	44,68533333	Середнє значення
Стандартна помилка	657,9484194	2,134904897	
Медіана	3520	42,79	
Мода	#N/A	#N/A	
Стандартне відхилення	2548,223271	8,268451113	Середньоквадратичне відхилення
Дисперсія вибірки	6493441,838	68,36728381	Дисперсія вибірки
Ексцес	-1,712674833	6,049645013	
Асиметричність	0,280614344	2,374906795	
Інтервал	6954	31,42	
Мінімум	1586	38,34	Мінімальне значення
Максимум	8540	69,76	Максимальне значення
Сума	71858	670,28	
Рахунок	15	15	Об'єм вибірки

1.6 Знаходження параметрів лінійної регресії

Щоб знайти параметри регресії, вибираємо пункт меню **Сервіс – Аналіз даних – Регресія.** Тут задаємо діапазони окремо для У, окремо – для Х (для багатофакторної регресії в полі «Вхідний інтервал Х» виділяємо всі значення чинників), встановлюємо прапорець у віконці «Мітки», «Вихідний діапазон» – на новий лист. Ок (рис.34).

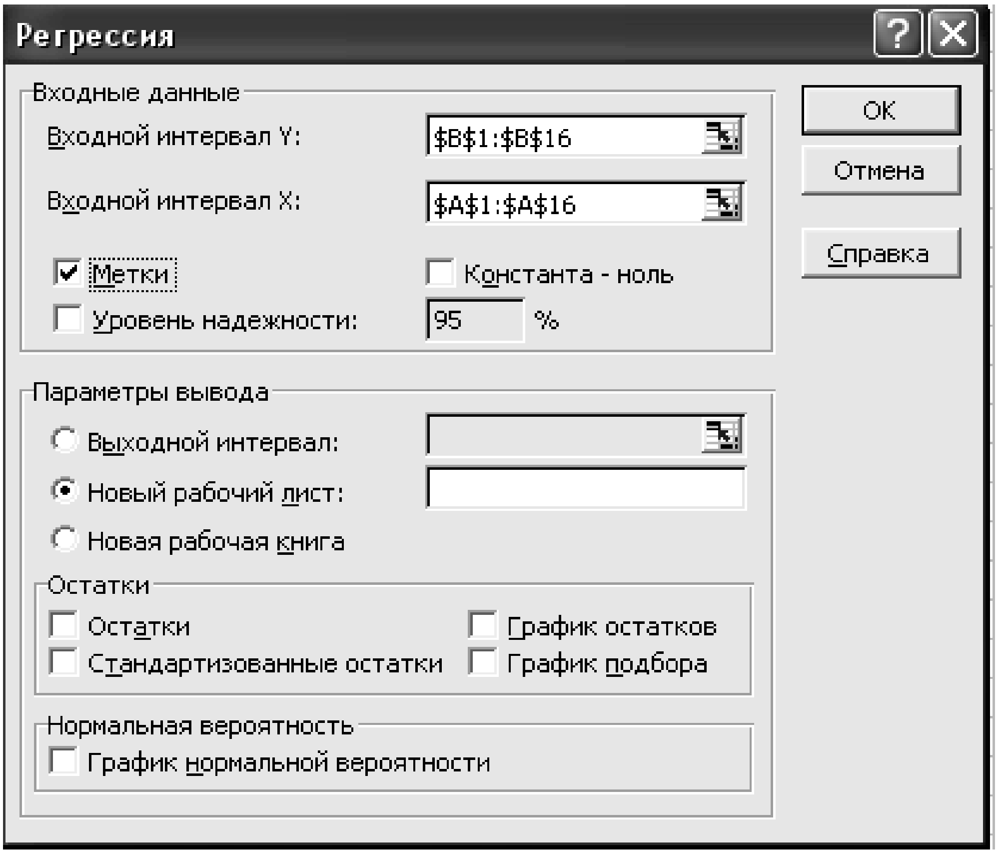


Рисунок 34

Результат одержали у вигляді таблиці (рис.35).

ВЫВОД ИТОГОВ								
Регрессионная статистика								
Множественный R	0,61974714							
R-квадрат	0,384086518							
Нормированный R-квадрат	0,336708557							
Стандартная ошибка	6,734050364							
Наблюдения	15							
Дисперсионный анализ								
	df	*SS*	*MS*	*F*	*Значимость F*			
Регрессия	1	367,6253273	367,6253273	8,106860574	0,013729018			
Остаток	13	589,516646	45,34743431					
Итого	14	957,1419733						
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	54,31885511	3,804055967	14,27919452	2,53059E-09	46,10069341	62,5370168	46,1006934	62,537017
x	-0,00201095	0,000706277	-2,847254919	0,013729018	-0,00353677	-0,00048513	-0,0035368	-0,0004851

Рисунок 35

З цієї таблиці вибираємо наступні величини (табл.14).

Таблиця 14

Назва в Excel	**Значення**	**Для даного прикладу**
1	***2***	***3***
У - перетин	Коефіцієнт b0	54,31885511
X	Коефіцієнт b1	-0,00201095
R - квадрат	Коефіцієнт детермінації R^2	0,384086518
Множинний R	Коефіцієнт кореляції (за модулем)	0,61974714
Стандартна помилка в регресійній статистиці	Середнє квадратичне відхилення залишків	6,734050364
Спостереження	Об'єм вибірки	15
F	Fспост	8,106860574
Значущість F	Рівень значущості для критерію Фішера	0,013729018
Df регресія залишок	Кількість ступенів вільності: k1 k2	 1 13

Продовження таблиці 14

1	*2*	*3*
Стандартна помилка (поряд із значенням коефіцієнтів)	Дисперсія коефіцієнтів	3,804055967 0,000706277
Стовпець t-статистика	Значення критерію Стьюдента, що спостерігається	14,27919452 -2,847254919
Стовпець Р-значення	Значущість коефіцієнтів за критерієм Стьюдента	$2{,}53059 \cdot 10^{-9}$ 0,013729018

1.7 Знаходження критичної точки розподілу Стьюдента

Вибираємо команду «Вставка функції», категорію «Статистичні», функцію СТЬЮДРАСПОБР. Вводимо необхідну ймовірність (0,05) і кількість ступенів вільності (k2 = n – 2). Одержимо для даного прикладу 2,16.

1.8 Додаткові можливості Excel

Пакет «Аналіз даних» використовується для обробки та екстраполяції складних та нелінійних даних. Для більш простих розрахунків або для доповнення регресійного аналізу можна використовувати декотрі додаткові можливості Excel.

Ряди даних, що зображені на двовимірних діаграмах (точкових, графіках, з областями, лінійчатих, гістограмах, біржових, пузиркових) можна доповнити лініями тренду. Лінії тренду дозволяють графічно відображувати тенденції даних і прогнозувати їх подальші зміни. Це також відноситься до регресійного аналізу. Використовуючи його, можна продовжити лінію тренда у діаграмі за межі реальних даних для завбачення наступних значень. Наприклад, наведена нижче діаграма (рис.36) використовує просту лінійну лінію тренду, яка є прогнозом на 4 квартали вперед, для демонстрації тенденції збільшення доходу (табл.15).

Таблиця 15

Квартал	**Доход, млн грн.**
Квартал1	2
Квартал2	2,24
Квартал3	2,49
Квартал4	2,51

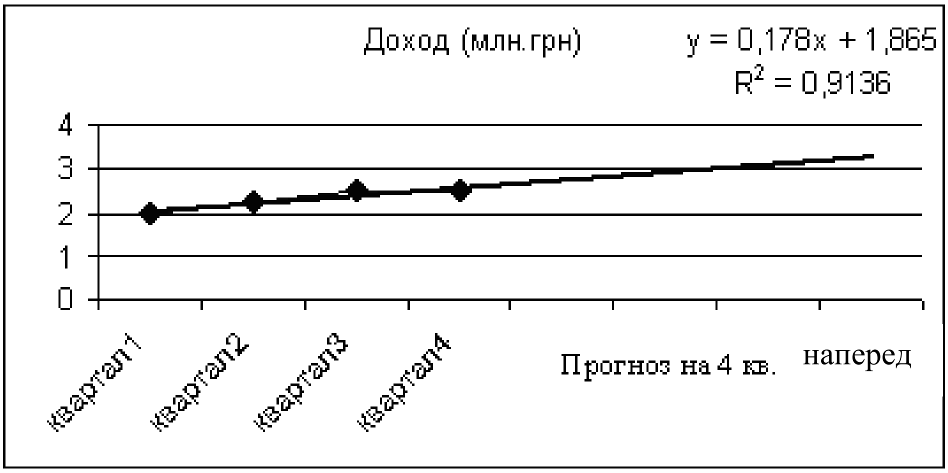


Рисунок 36

Існує шість різних видів лінії тренду (апроксимація і згладжування), які можуть бути додані на діаграму Microsoft Excel. Засіб треба обирати в залежності від типу даних. Лінія тренду найкраще наближається до зображеної на діаграмі залежності, якщо значення R^2 (R-квадрат) дорівнює або близький до одиниці. В Excel значення R^2 розраховується автоматично, і його можна вивести на діаграму.

1 Лінійна апроксимація — це пряма лінія, яка використовується у простих випадках, коли точки даних наближені до прямої. Тобто лінійна апроксимація підходить для величини, що зростає або убуває з постійною швидкістю. На наведеній нижче діа-

грамі (рис.37) пряма лінія описує стабільне зростання продаж холодильників на протязі 13 років. Значення R^2=0,9036, це близько до одиниці, що підтверджує добре співпадання розрахованої лінії з даними.

Рисунок 37

2 Логарифмічна апроксимація використовується для опису величини, яка спочатку швидко зростає або зменшується, а потім стабілізується. На наведеній нижче діаграмі (рис.38) логарифмічна крива описує прогнозоване зростання популяції тварин, що живуть ареалі з фіксованими межами.

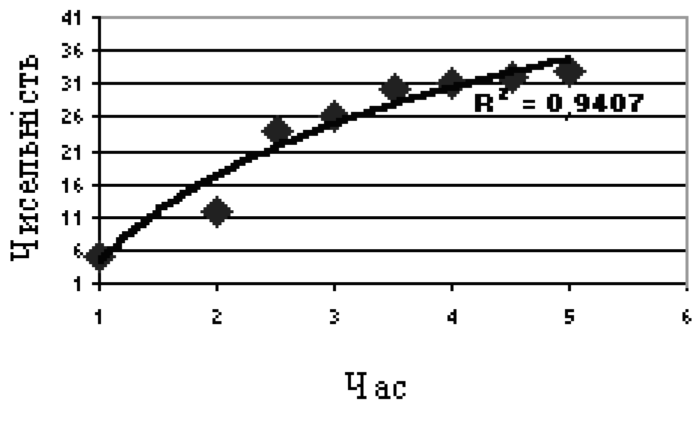


Рисунок 38

3 Поліноміальна апроксимація використовується для опису нестабільної величини, яка то зростає, то зменшується. Ступінь поліному визначається кількістю екстремумів кривої. У наступ-

ному прикладі поліном другого ступеня (один максимум) описує залежність розходу бензину від швидкості автомобілю. Значення R^2 дорівнює 0,9474 – близьке до одиниці (рис.39).

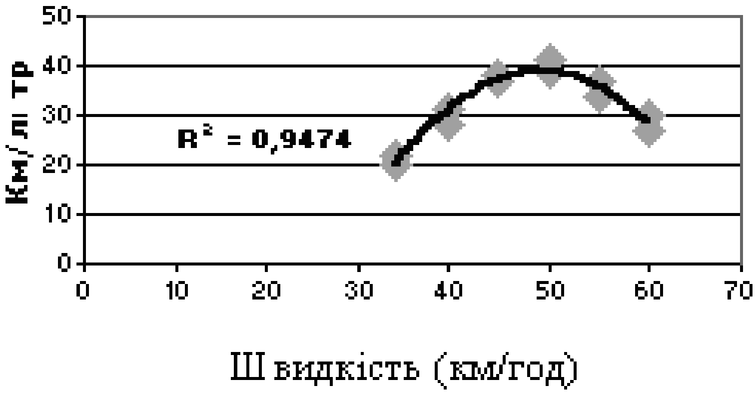


Рисунок 39

4 Ступенева апроксимація використовується для опису величини, яка або монотонно зростає, або монотонно зменшується. Дані не можуть бути нульовими, або від'ємними.

5 Експоненційна апроксимація використовується у тих випадках, коли швидкість зміни даних зростає. Як і для ступеневої апроксимації дані не можуть бути нульовими, або від'ємними. У наведеній нижче діаграмі (рис.40) експоненційна лінія тренду описує вміст радіоактивного вуглецю – 14 у залежності від віку органічного об'єкта.

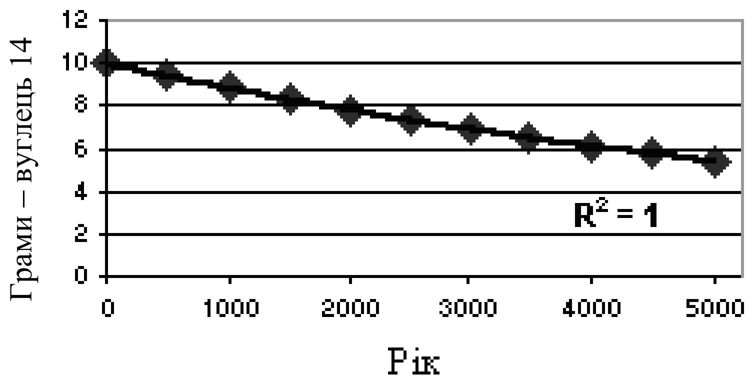


Рисунок 40

6 Використання у якості апроксимації ковзного середнього дозволяє згладити коливання і таким чином більш наглядно показати характер залежності. Така лінія тренду будується за заданою кількістю точок. Дані усереднюються, і отриманий результат використовується як середнє значення для наближення. У наведеній нижче діаграмі показана залежність кількості продаж на протязі 26 тижнів, отримана методом ковзного середнього(рис.41).

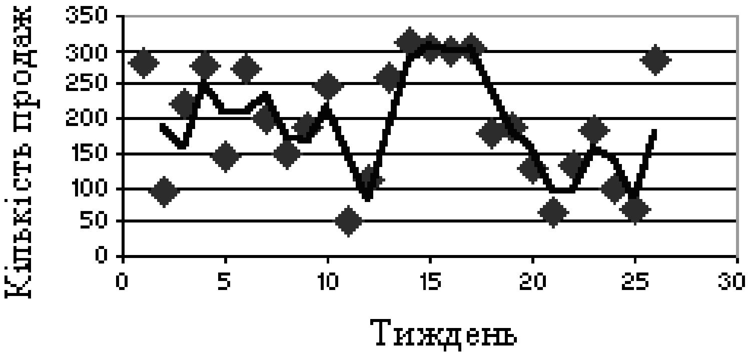


Рисунок 41

Щоб додати лінію тренду на діаграму, треба:

1 Виділити область діаграми.

2 У меню **Діаграма** вибрати команду ***Додати лінію тренда***.

3 У двосторінковому вікні, що з'явиться, на сторінці **Тип** вибрати тип лінії тренда (один з тих, що були описані вище), на сторінці **Параметри** відмітити ***Помістити*** на діаграму величину достовірності апроксимації (R^2).

1.9 Ковзні середні

Ковзні середні використовуються для розрахунку значень у прогнозованому періоді на базі середнього значення змінної для вказаної кількості попередніх періодів. Ковзні середні несуть у собі дані про тенденції зміни даних. Цей метод може використовуватися для прогнозу збуту, запасів та інших економічних процесів. Розрахунок прогнозованих значень у Microsoft Excel виконується за формулою

$$F_{(t+1)} = \frac{1}{k}\sum_{i=1}^{k} x_{t-j+1},$$

де k - кількість попередніх періодів, що входять до ковзного середнього;

x_j - фактичне значення у момент часу j;

F_j - прогнозоване значення у момент часу j.

Вибираємо команду: **Сервіс – Аналіз даних – Ковзні середні**. У полі «Вхідний діапазон» задаємо діапазон даних. Це повинний бути стовпець або рядок, що має не менш 4 клітин. Якщо дані виділені разом із заголовком, то треба відмітити віконце «Мітки у першій строчці/стовпці». У полі «Інтервал» треба задати кількість періодів, необхідних для розрахунку ковзного середнього (автоматично береться 3). У полі «Вихідний діапазон» треба вказати ліву верхню клітинку для друку розрахункових значень на тому ж листі, де знаходяться початкові дані. Якщо початкових даних недостатньо для побудування прогнозу, Excel повертає значення #Н/Д. Щоб побачити поряд із розрахунковими значеннями значення стандартної помилки, треба відмітити вікно «Стандартні помилки». Відмітити вікно «Друк графіка» для автоматичного створення вбудованої діаграми (рис.42).

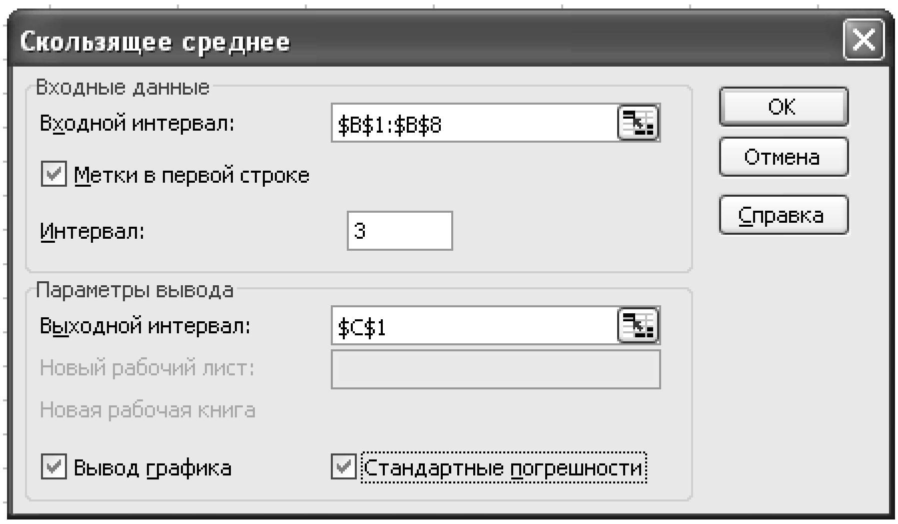


Рисунок 42

1.10 Експоненційне згладжування

Експоненційне згладжування використовується для завбачення значень на основі прогнозу для попереднього періоду, скорегованого з урахуванням помилок у цьому прогнозі. При аналізі використовують константу згладжування α, за величиною якої визначають ступінь впливу на прогнози помилок у попередньому періоді.

Вибираємо команду: **Сервіс – Аналіз даних – Експоненційне згладжування.**

У полі «Вхідний діапазон» задаємо діапазон даних. Це має бути стовпець або рядок, що має не менш 4 клітин. Якщо дані виділені разом із заголовком, то треба відмітити віконце «Мітки». У полі «Фактор затухання» ввести значення константи експоненційного згладжування α. Це той фактор, що буде мінімізувати нестабільність даних генеральної сукупності.

Зазвичай α беруть $0{,}2 \div 0{,}3$. Це означає, що помилка поточного прогнозу встановлена на рівні від 20 до 30% від помилки попереднього прогнозу. Більші значення α прискорюють відклик, але можуть призвести до непередбачуваним викидам. Низькі значення α можуть призвести до великих розривів між розрахунковими значеннями.

Автоматично встановлюється фактор затухання, який дорівнює 0,3.

У полі «Вихідний діапазон» треба вказати ліву верхню клітинку для друку розрахункових значень на тому ж листі, де знаходяться початкові дані. Якщо початкових даних недостатньо для побудування прогнозу, Excel повертає значення #Н/Д. Щоб побачити поряд із розрахунковими значеннями значення стандартної помилки, треба відмітити вікно «Стандартні помилки».

Щоб побудувати діаграму для фактичних і прогнозованих значень, встановіть «прапорець» у вікні «Друк графіка» (рис.43).

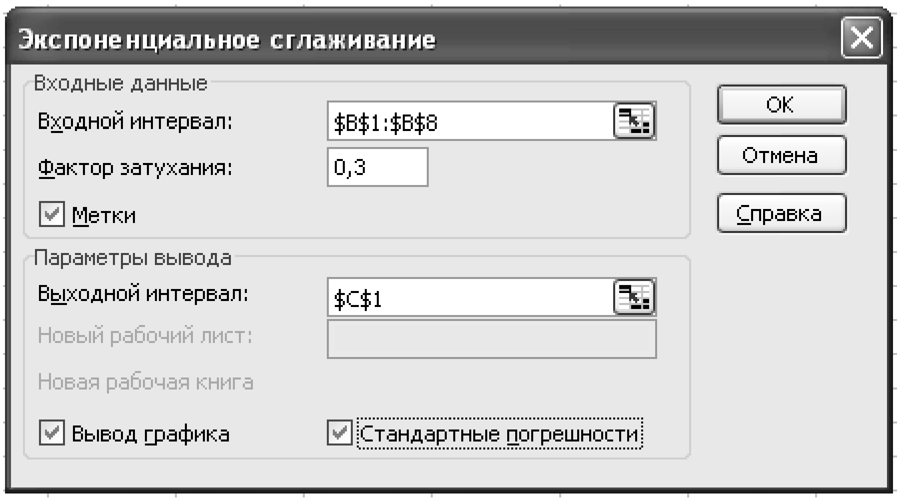


Рисунок 43

2 ПОСЛІДОВНІСТЬ ПОБУДОВИ МОДЕЛЕЙ

2.1 План побудови лінійної однофакторної моделі

Будуємо лінійну модель вигляду $y = b_0 + b_1 x$.

Послідовність дій:

1 Вводимо дані. Визначаємо основні статистики.

2 Будуємо діаграму розсіювання (кореляційне поле).

3 Визначаємо тісноту лінійного зв'язку за коефіцієнтом кореляції.

4 Записуємо лінійну модель у вигляді $y = b_0 + b_1 x$.

5 Визначаємо загальну якість моделі за коефіцієнтом детермінації R^2. Перевіряємо одержану модель на адекватність за допомогою критерію Фішера. Усі подальші розрахунки виконуються тільки за умови адекватності моделі початковим статистичним даним.

6 Перевіряємо статистичну значущість коефіцієнтів моделі.

7 За одержаною моделлю розраховуємо значення показника y для всіх точок вибірки і в точці прогнозу.

8 Розраховуємо напівширину довірчого інтервалу.

9 Розраховуємо довірчий інтервал для всіх точок вибірки і в точці прогнозу: $(\hat{y}-\delta, \hat{y}+\delta)$.

10 Будуємо довірчу область.

11 Розраховуємо коефіцієнт еластичності.

12 Використовуючи одержані дані та теоретичні відомості, робимо економетричний аналіз – описуємо процес побудови моделі і всі супутні розрахунки.

Приклад виконання

Економічні дані

Продуктивність праці та рівень рентабельності по плодоовочевих консервних заводах області за рік задані в таблиці 16.

Таблиця 16

Номер	**Продуктивність праці, грн**	**Рівень рентабельності, %**
1	7224	31,3
2	3223	23,7
3	1586	10,3
4	7671	30,2
5	6921	32,2
6	2340	24,8
7	3520	13,4
8	2569	23,4
9	1920	20,1
10	9410	37,9
11	2901	20
12	8492	35,1
13	6630	37,2
14	2911	23,2
15	11540	39,4

Виконання завдання

1 Вводимо дані у вигляді таблиці 16. Позначимо: x - продуктивність праці, грн, y - рівень рентабельності, %. Визначаємо

основні статистики. Для цього в меню **Сервіс** вибираємо команду ***Аналіз даних***, із списку **Інструменти** аналізу вибираємо ***Описова статистика***, Ок (див п1.5). Результат отримаємо у вигляді таблиці (рис.44). У ній жирним шрифтом виділені значення, які знадобляться далі.

x		y	
Среднее	**5257,2**	**Среднее**	**26,81333**
Стандартная ошибка	816,7708	Стандартная ошибка	2,289788
Медиана	3520	Медиана	24,8
Мода	#Н/Д	Мода	#Н/Д
Стандартное отклонени	**3163,34**	**Стандартное отклонени**	**8,868312**
Дисперсия выборки	**10006718**	**Дисперсия выборки**	**78,64695**
Эксцесс	-0,97862	Эксцесс	-0,76191
Асимметричность	0,55141	Асимметричность	-0,27517
Интервал	9954	Интервал	29,1
Минимум	**1586**	**Минимум**	**10,3**
Максимум	**11540**	**Максимум**	**39,4**
Сумма	78858	Сумма	402,2
Счет	15	Счет	15

Рисунок 44

Отримали, що середнє значення продуктивності праці дорівнює 5257,2грн, вона змінюється у межах від 1586грн (мінімум) до 11540грн (максимум), що задає область прогнозів – діапазон, з якого припустимо вибирати значення фактора x(грн.), для прогнозу рівня рентабельності y(%). $x_{ср} = 5257{,}2$ задає центр області прогнозів. Середнє квадратичне відхилення 33163,34 характеризує середнє значення розсіювання значень продуктивності праці по плодоовочевих консервних заводах області за рік.

2 Для виявлення зв’язку між фактором x і показником y побудуємо кореляційне поле: **Майстер діаграм, тип – Точкова, вид - Точкова** (рис.45).

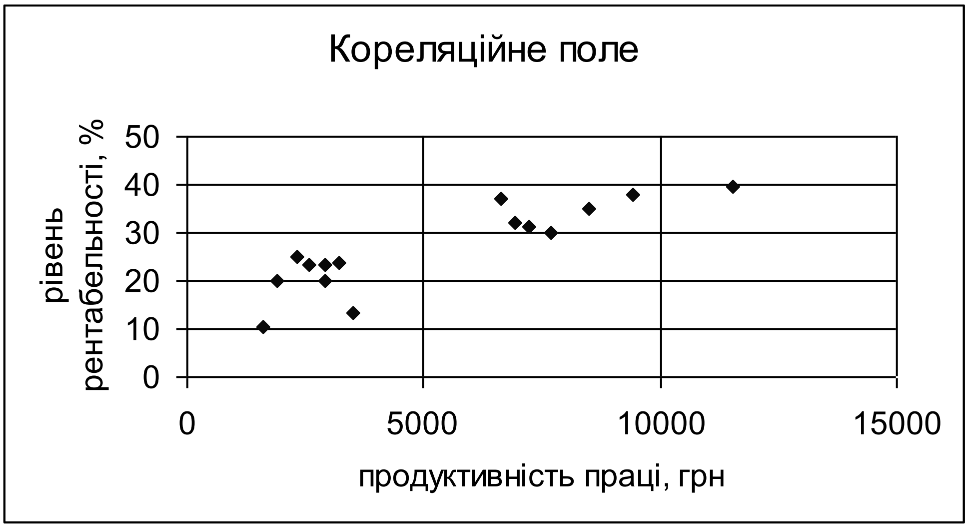


Рисунок 45

За виглядом кореляційного поля можна висунути гіпотезу, що між фактором x і показником y є лінійний зв'язок.

3 Для підтвердження знайдемо коефіцієнт кореляції: **Сервіс – Аналіз даних – Кореляція, Ок** (див п1.4). Результат отримаємо у вигляді таблиці (рис.46).

	x	y
x	1	
y	0,876868	1

Рисунок 46

Коефіцієнт кореляції $r_{xy} \approx 0{,}877$. Це означає, що між x і y існує достатній лінійний зв'язок. Формулу цього зв'язку будемо шукати у вигляді $y = b_0 + b_1 x$.

4 Параметри моделі b_0 і b_1 знайдемо за методом найменших квадратів: **Сервіс – Аналіз даних – Регресія, Ок** (див п1.6). Результат отримаємо у вигляді таблиці (рис.47).

Щоб додатково отримати прогнозовані значення $y(x)$(рис.48) і графік (рис.49) у діалоговому вікні «Регресія» треба встановити прапорець у вікні «Графік підбору».

ВЫВОД ИТОГОВ						
Регрессионная статистика						
Множественный R	**0,876867607**					
R-квадрат	**0,7688968**					
Нормированный R-квадрат	0,751119631					
Стандартная ошибка	**4,424215473**					
Наблюдения	15					
Дисперсионный анализ						
	df	*SS*	*MS*	*F*	***Значимость F***	
Регрессия	1	846,5994602	846,5994602	**43,25192554**	**0,000018**	
Остаток	13	254,4578732	19,57368255			
Итого	14	1101,057333				
	Коэффициенты	*Стандартная ошибка*	***t-статистика***	***P-Значение***	*Нижние 95%*	*Верхние 95%*
Y-пересечение	**13,88972928**	2,272985648	**6,110786177**	**0,000037**	8,979242339	18,80021622
x	**0,002458268**	0,000373789	**6,576619613**	**0,000018**	0,001650746	0,003265789

Рисунок 47

Наблюдение	*Предсказанное y*	*Остатки*
1	31,64825391	-0,348253909
2	21,81272553	1,887274474
3	17,78854158	-7,488541581
4	32,74709949	-2,547099494
5	30,90339885	1,296601152
6	19,6420753	5,157924702
7	22,54283098	-9,142830982
8	20,20501856	3,194981438
9	18,60960294	1,490397064
10	37,02202673	0,877973273
11	21,02116338	-1,021163382
12	34,76533714	0,334662864
13	30,188043	7,011957003
14	21,04574606	2,154253943
15	42,25813656	-2,858136564

Рисунок 48

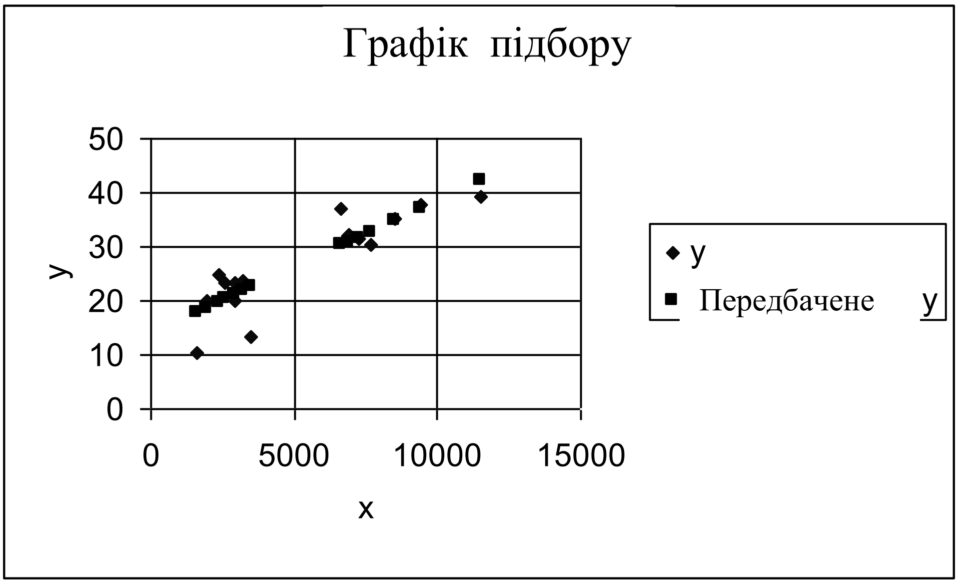


Рисунок 49

Параметри моделі b_0 i b_1 беремо у стовпці «Коефіцієнти» (рис.47): $b_0 = 13{,}89, b_1 = 0{,}0025.$ Модель буде мати вигляд: $y = 13{,}89 + 0{,}0025x$.

5 Визначаємо загальну якість моделі за коефіцієнтом детермінації $R^2 = 0{,}769$ (рис.47). Це означає, що 76,9% початкових даних пояснюється побудованою моделлю, а 23,1% – це непояснений розклад. Загальна якість моделі – добра.

Перевіряємо модель на адекватність за допомогою критерію Фішера.

Значущість $F = 0{,}000018$ (рис.47). Це менше 0,05, тобто, з надійністю 0,95 можна стверджувати, що модель адекватна початковим даним.

6 Перевіряємо статистичну значущість коефіцієнтів моделі b_0 і b_1 за допомогою критерію Стьюдента. Для коефіцієнта b_0 $p-$значення дорівнює 0,000037 (рис.47). Це менше 0,05$\Rightarrow$ коефіцієнт b_0 статистично значущий. Для коефіцієнта b_1 $p-$значення дорівнює 0,000018 < 0,05$\Rightarrow$ коефіцієнт b_1 статистично значущий.

Ми отримали адекватну модель зв’язку продуктивності праці (x) і рівня рентабельності (y): $y = 13{,}89 + 0{,}0025x$.

7 За одержаною моделлю розраховуємо значення показника y для всіх точок вибірки і в точці прогнозу. Точку прогнозу вибираємо із області прогнозу $1586 \le x \le 11540$. Наприклад, $x_{пр} = 5000$. Дописуємо $x_{пр}$ у таблицю (рис.51) і виконуємо розрахунок за моделлю так, як це показано в таблицях (рис.50, 51).

x	y	y(x)	Delta	y_min	y_max	Ex
7224	31,3	31,648	9,998345	21,650	41,647	0,56
3223	23,7	21,813	10,00714	11,806	31,820	0,36
1586	10,3	17,789	10,30695	7,482	28,095	0,22
7671	30,2	32,747	10,062	22,685	42,809	0,58
6921	32,2	30,903	9,962408	20,941	40,866	0,55
2340	24,8	19,642	10,14858	9,493	29,791	0,29
3520	13,4	22,543	9,970574	12,572	32,513	0,38
2569	23,4	20,205	10,10726	10,098	30,312	0,31
1920	20,1	18,610	10,23263	8,377	28,842	0,25
9410	37,9	37,022	10,42546	26,597	47,447	0,62
2901	20	21,021	10,05309	10,968	31,074	0,34
8492	35,1	34,765	10,21116	24,554	44,977	0,60
6630	37,2	30,188	9,933446	20,255	40,121	0,54
2911	23,2	21,046	10,05157	10,994	31,097	0,34
11540	39,4	42,258	11,09887	31,159	53,357	0,67
5000		**26,181**	**9,873579**	**16,307**	**36,055**	**0,47**
b0=	13,89					
b1=	0,0025					
Стандартная ошибка	4,424215473					
t_кр=	2,160368652					
Среднее (x)	5257,2					
Дисперсия выборки (x)	10006718,03					

Рисунок 50

	A	B	C	D	E	F	G
1	x	y	y(x)	Delta	y_min	y_max	Ex
2	7224	31,3	=$B$20+$B$21*A2	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A2-$B$24)^2)/14/$B$25)	=C2-D2	=C2+D2	=$B$21*A2/C2
3	3223	23,7	=$B$20+$B$21*A3	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A3-$B$24)^2)/14/$B$25)	=C3-D3	=C3+D3	=$B$21*A3/C3
4	1586	10,3	=$B$20+$B$21*A4	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A4-$B$24)^2)/14/$B$25)	=C4-D4	=C4+D4	=$B$21*A4/C4
5	7671	30,2	=$B$20+$B$21*A5	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A5-$B$24)^2)/14/$B$25)	=C5-D5	=C5+D5	=$B$21*A5/C5
6	6921	32,2	=$B$20+$B$21*A6	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A6-$B$24)^2)/14/$B$25)	=C6-D6	=C6+D6	=$B$21*A6/C6
7	2340	24,8	=$B$20+$B$21*A7	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A7-$B$24)^2)/14/$B$25)	=C7-D7	=C7+D7	=$B$21*A7/C7
8	3520	13,4	=$B$20+$B$21*A8	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A8-$B$24)^2)/14/$B$25)	=C8-D8	=C8+D8	=$B$21*A8/C8
9	2569	23,4	=$B$20+$B$21*A9	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A9-$B$24)^2)/14/$B$25)	=C9-D9	=C9+D9	=$B$21*A9/C9
10	1920	20,1	=$B$20+$B$21*A10	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A10-$B$24)^2)/14/$B$25)	=C10-D10	=C10+D10	=$B$21*A10/C10
11	9410	37,9	=$B$20+$B$21*A11	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A11-$B$24)^2)/14/$B$25)	=C11-D11	=C11+D11	=$B$21*A11/C11
12	2901	20	=$B$20+$B$21*A12	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A12-$B$24)^2)/14/$B$25)	=C12-D12	=C12+D12	=$B$21*A12/C12
13	8492	35,1	=$B$20+$B$21*A13	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A13-$B$24)^2)/14/$B$25)	=C13-D13	=C13+D13	=$B$21*A13/C13
14	6630	37,2	=$B$20+$B$21*A14	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A14-$B$24)^2)/14/$B$25)	=C14-D14	=C14+D14	=$B$21*A14/C14
15	2911	23,2	=$B$20+$B$21*A15	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A15-$B$24)^2)/14/$B$25)	=C15-D15	=C15+D15	=$B$21*A15/C15
16	11540	39,4	=$B$20+$B$21*A16	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A16-$B$24)^2)/14/$B$25)	=C16-D16	=C16+D16	=$B$21*A16/C16
17	5000		=$B$20+$B$21*A17	=$B$22*$B$23*КОРЕНЬ(1+1/15+((A17-$B$24)^2)/14/$B$25)	=C17-D17	=C17+D17	=$B$21*A17/C17
18							
19		*Коэффициенты*					
20	b0=	13,8897292806605					
21	b1=	0,0024582675288505					
22	Стандартная ошибка	4,42421547283548					
23	t_кр=	=СТЬЮДРАСПОБР(0,05;15-2)					
24	Среднее (x)	5257,2					
25	Дисперсия выборки (x)	10006718,0285714					

Рисунок 51

Перевірте: значення $y(x)$ із таблиці (рис.50) і значення «Передбачені у» із таблиці (рис.48) повинні співпадати.

8 Для побудування довірчого інтервалу для прогнозу розраховуємо його напівширину за формулою

$$\delta = \sigma_e \cdot t_\gamma \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{np} - \overline{x})^2}{(n-1)D(x)}},$$

де σ_e – стандартна помилка, σ_e=4,424 (рис.47);

t_γ – критична точка розподілу Стьюдента для надійності γ = 0,95 і кількості ступенів вільності $k2 = n - 2 = 13$; n =15 – кількість спостережень; $D(x) = 1{,}0006716$ – дисперсія вибірки (рис.44); $\overline{x} = x_{cp} = 5257{,}2$ – середнє значення (рис.44); $x_{np} = 5000$ – точка з області прогнозів.

Додаємо ще один стовпець *Delta* і виконуємо розрахунок так, як показано у таблицях (див. рис.50, 51).

9 Далі розраховуємо ліву межу довірчого інтервалу за формулою $y_{\min} = y(x) - delta$, праву – $y_{\max} = y(x) + delta$ (див рис.50, рис.51).

У 17-му рядку таблиці (рис.50) знаходиться прогноз.

Точковий: для заданого значення продуктивності праці 5000грн (x_{np}) прогнозоване значення рівня рентабельності ($y(x_{np})$) становить 26,181%.

Прогноз з урахуванням довірчого інтервалу: для заданого значення продуктивності праці 5000грн прогнозоване значення рівня рентабельності з надійністю 95% лежить у межах від 16,307% ($y_{\min}$) до 36,055% ($y_{\max}$).

10 Будуємо довірчу область: **Діаграма – Точкова – Точкова**. У якості вхідних даних виділяємо стовпці x, y, $y(x)$, $y_{\min}$, $y_{\max}$ (рис.52).

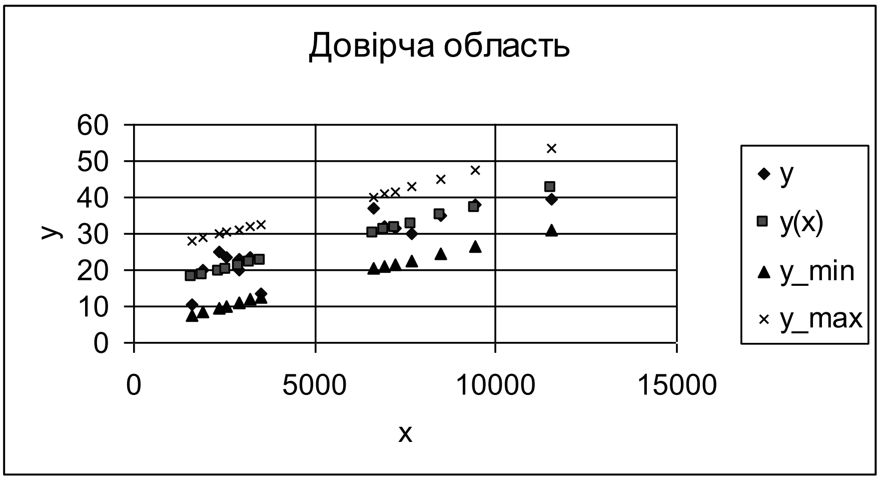


Рисунок 52

11 Розраховуємо коефіцієнт еластичності за формулою

$$E_x = \frac{x}{y(x)} \cdot y'_x.$$

Для лінійної моделі –

$$y'_x = b_1.$$

Одержимо: $$E_x = \frac{b_1 x}{y(x)}.$$

Виконуємо розрахунок, як це показано в таблицях (див.рис.50, 51): $E_x = 0{,}47$.

Коефіцієнт еластичності показує, що при зміні значення фактора продуктивності праці $x_{пр} = 5000$ грн на 1% значення рівня рентабельності ($y(x)$) збільшиться від 26,181% на 0,47%.

Висновки

Виконаний статистичний аналіз показує, що залежність рівня рентабельності по плодоовочевих заводах області за рік y(%) від продуктивності праці x(грн) може бути описана лінійною моделлю $y = 13{,}89 + 0{,}0025 \cdot x$. При цьому всі коефіцієнти моделі статистично значущі, а сама модель адекватна.

Для заданого значення $x_{пр} = 5000$ грн прогнозоване значення рівня рентабельності становить 26,181%, або, з надійністю 0,95 лежить у межах від 16,307% до 36,055%.

Коефіцієнт еластичності $E_x = 0{,}47$ показує, що при зміні продуктивності праці від 5000грн на 1% відповідний рівень рентабельності збільшується на 0,47%.

Примітка. Лист з розрахунками в Excel виглядає так, як показано на рисунку 50, а відповідний лист з формулами Excel – на рисунку 51.

2.2 План побудови нелінійної однофакторної моделі

Послідовність дій

1 Вводимо дані. Визначаємо основні статистики.

2 Будуємо кореляційне поле. За його виглядом висуваємо гіпотезу про нелінійну залежність між *x* і *y*.

3 Лінеаризуємо нелінійну модель за допомогою формул переходу (табл.17).

Таблиця 17

Вигляд залежності	**Лінеаризована підстановка**		**Зворотне перетворення**		
	$u =$	$v =$	$a =$	$b =$	$y =$
$y = \frac{a}{x} + b$	$\frac{1}{x}$	y	b_1	b_0	v
$y = a\sqrt{x} + b$	$\sqrt{x}$	y	b_1	b_0	v
$y = ax^b$	$\ln x$	$\ln y$	e^{b_0}	b_1	e^v
$y = a \cdot \ln x + b$	$\ln x$	y	b_1	b_0	v
$y = e^{ax} \cdot b$	x	$\ln y$	b_1	e^{b_0}	e^v
$y = a \cdot x^2 + b$	x^2	y	b_1	b_0	v

Одержуємо лінійну модель щодо нових змінних $v = b_0 + b_1 u$.

4 Визначаємо тісноту лінійного зв'язку між u та v за коефіцієнтом кореляції.

5 Записуємо лінійну модель вигляду $v = b_0 + b_1 u$.

6 Визначаємо загальну якість моделі за коефіцієнтом детермінації R^2. Перевіряємо одержану модель на адекватність за допомогою критерію Фішера. Усі подальші розрахунки виконуються тільки за умови адекватності моделі початковим статистичним даним.

7 Перевіряємо статистичну значущість коефіцієнтів моделі.

8 За одержаною моделлю розраховуємо значення показника v для всіх точок вибірки і в точці прогнозу (точку прогнозу вибираємо довільно з області прогнозу).

9 Розраховуємо напівширину довірчого інтервалу.

10 Розраховуємо довірчий інтервал для всіх точок вибірки і в точці прогнозу: ($\hat{v}-\delta$, $\hat{v}+\delta$).

11 Якщо лінеаризована модель $v = b_0 + b_1 u$ адекватна (за критерієм Фішера), то і початкова нелінійна модель буде адекватна.

12 За формулами зворотного переходу (див. табл.17) перераховуємо значення y, y_{min} (ліва межа довірчого інтервалу), y_{max} (права межа довірчого інтервалу).

13 Будуємо довірчу область.

14 Розраховуємо коефіцієнт еластичності.

15 Використовуючи одержані дані та теоретичні відомості, робимо економетричний аналіз – описуємо процес побудови моделі та всі супутні розрахунки.

Приклад виконання

1 Початкові дані візьмемо ті, що у попередньому прикладі. Тому основні статистики будуть ті ж самі.

2 Будуємо кореляційне поле (рис.53). Можна побачити, що із зростанням x значення y спочатку зростає, а потім стабілізу-

ється. Тому можна висунути гіпотезу, що зв'язок між y та x логарифмічний: $y = a\ln x + b$.

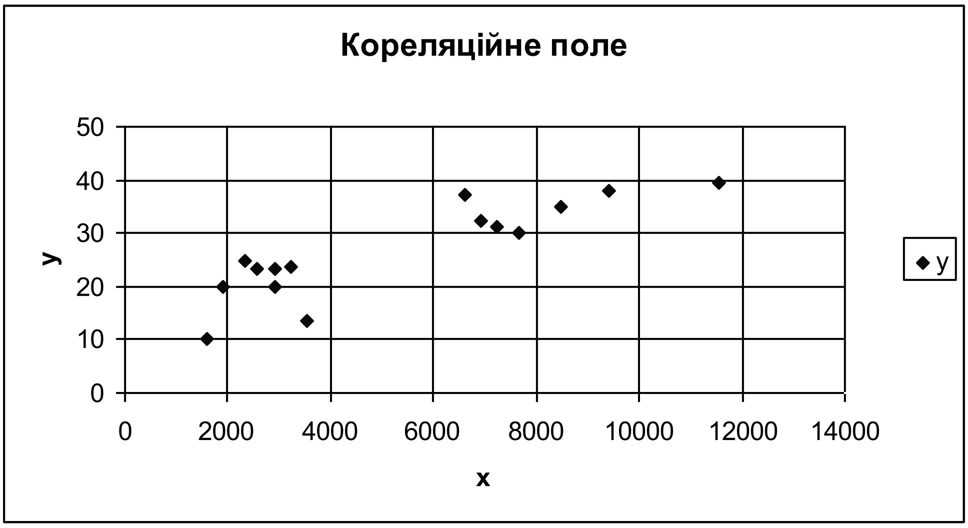


Рисунок 53

3 Зведемо цю модель до лінійної за допомогою формул переходу (табл.17): $u = \ln x$, $v = y$. Додаємо 2 стовпця до таблиці (рис. 60) і виконуємо розрахунки так, як показано у таблицях (див. рис.60, 61). Спробуємо отримати для нових змінних лінійну модель у вигляді: $v = b_0 + b_1 u$.

4 Знайдемо коефіцієнт кореляції: **Сервіс – Аналіз даних – Кореляція, Ок** (див п1.4). «Вхідний інтервал» треба виділити стовпці «u» і «v». Результат отримаємо у вигляді таблиці (рис.54).

	u	*v*
u	1	
v	0,885786	1

Рисунок 54

$r_{uv} \approx 0{,}886$. Це означає, що між u та v існує достатній лінійний зв'язок.

5 Параметри моделі b_0 і b_1 знайдемо за методом найменших квадратів: **Сервіс – Аналіз даних – Регресія, Ок** (див п1.6). У яко-

сті «Вхідного інтервалу y» треба виділити стовпець v, а «Вхідний інтервал x» – стовпець u. Результат отримаємо у вигляді таблиці (рис.55).

ВЫВОД ИТОГОВ						
Регрессионная статистика						
Множественный R	**0,885786392**					
R-квадрат	**0,784617533**					
Нормированный R-квад	0,768049651					
Стандартная ошибка	**4,271087458**					
Наблюдения	15					
Дисперсионный анализ						
	df	*SS*	*MS*	***F***	***Значимость F***	
Регрессия	1	863,9088884	863,9088884	**47,35774486**	**0,000011**	
Остаток	13	237,1484449	18,24218807			
Итого	14	1101,057333				
	Коэффициенты	*Стандартная ошибка*	***t-статистика***	***P-Значение***	*Нижние 95%*	*Верхние 95%*
Y-пересечение	**-75,7208957**	14,94031299	**-5,068226865**	**0,000215**	-107,9974795	-43,4443
u	**12,23037296**	1,777232285	**6,881696365**	**0,000011**	8,39089604	16,06985

Рисунок 55

Щоб додатково отримати прогнозовані значення $v(u)$(рис.56) і графік (рис.57) у діалоговому вікні «Регресія» треба встановити прапорець у вікні «Графік підбору».

ВЫВОД ОСТАТКА		
Наблюдение	*Предсказанное v*	*Остатки*
1	32,94797498	-1,647974984
2	23,07688728	0,623112715
3	14,40436065	-4,104360647
4	33,68226418	-3,482264181
5	32,42392126	-0,223921259
6	19,16119061	5,638809388
7	24,15497493	-10,75497493
8	20,30308903	3,09691097
9	16,74170799	3,358292009
10	36,18124698	1,718753018
11	21,78955372	-1,789553724
12	34,92581865	0,174181353
13	31,89855992	5,301440083
14	21,83164039	1,368359611
15	38,67680942	0,723190577

Рисунок 56

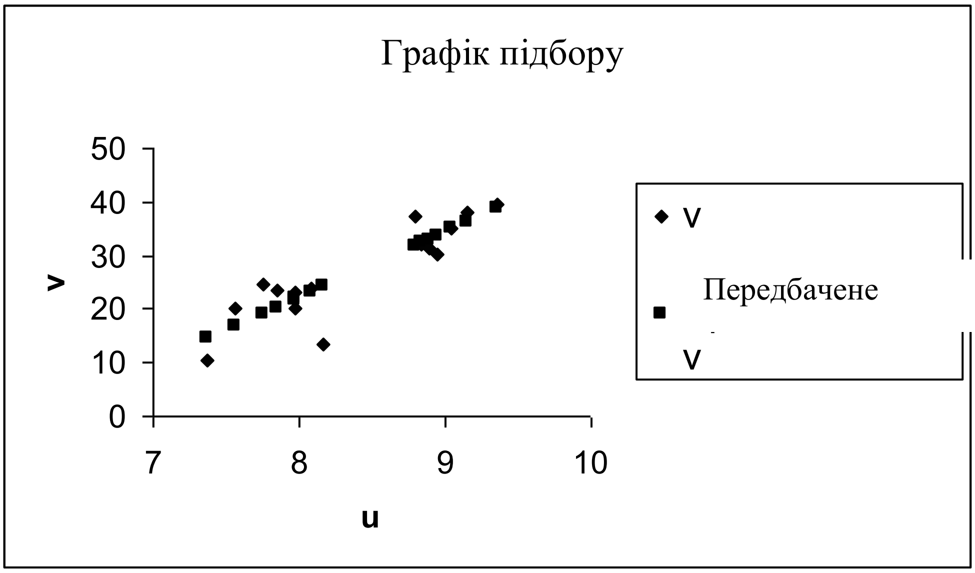


Рисунок 57

Параметри моделі b_0 *і* b_1 беремо у стовпці «Коефіцієнти» (рис.55): $b_0 = -75{,}72,\ b_1 = 12{,}23.$ Модель буде мати вигляд:

$$v = -75{,}72 + 12{,}23 \cdot u\,.$$

6 Визначаємо загальну якість моделі за коефіцієнтом детермінації $R^2 = 0{,}785$ (див.рис.55). Це означає, що 78,5% початкових даних пояснюється моделлю. Загальна якість моделі добра.

Перевіряємо модель на адекватність за допомогою критерію Фішера. Значущість $F = 0{,}000011 < 0{,}05$. Тому можна стверджувати з рівнем довіри 0,95, що модель адекватна.

7 Перевіряємо статистичну значущість коефіцієнтів моделі b_0 і b_1 за допомогою критерію Стьюдента. Для коефіцієнта b_0 $p-$значення дорівнює 0,000215 (див. рис.55). Це менше 0,05$\Rightarrow$ коефіцієнт b_0 статистично значущий. Для коефіцієнта b_1 $p-$значення дорівнює 0,000011 < 0,05$\Rightarrow$ коефіцієнт b_0 статистично значущий.

8 За моделлю $v = -75{,}72 + 12{,}23 \cdot u$ розраховуємо значення показника $v(u)$ для всіх точок вибірки і в точці прогнозу. Точку прогнозу вибираємо із області прогнозу $1586 \le x \le 11540$. Наприклад, $x_{np} = 5000$. Дописуємо x_{np} у таблицю (рис. 60), додаємо

стовпець $v(u)$ і виконуємо розрахунок так, як показано в таблицях (див.рис.60, 61).

9 Для подальших розрахунків треба знати деякі статистики: Сервіс – Аналіз даних – Описова статистика, Ок. «Вхідний інтервал» - «u» і «v». Результат отримаємо у вигляді таблиці (рис.58).

u		v	
Среднее	8,314288	Среднее	25,91429
Стандартная ошибка	0,161847	Стандартная ошибка	2,262107
Медиана	8,122142	Медиана	24,25
Мода	#Н/Д	Мода	#Н/Д
Стандартное отклонение	0,605575	Стандартное отклонение	8,464029
Дисперсия выборки	0,366721	Дисперсия выборки	71,63978
Эксцесс	-1,59632	Эксцесс	-0,66544
Асимметричность	0,013846	Асимметричность	-0,2665
Интервал	1,780558	Интервал	27,6
Минимум	7,36897	Минимум	10,3
Максимум	9,149528	Максимум	37,9
Сумма	116,4	Сумма	362,8
Счет	14	Счет	14

Рисунок 58

Для побудування довірчого інтервалу для прогнозу розраховуємо його напівширину за формулою

$$\delta = \sigma_e \cdot t_\gamma \cdot \sqrt{1 + \frac{1}{n} + \frac{(u_{np} - \bar{u})^2}{(n-1)D(u)}},$$

де σ_e – стандартна помилка, σ_e=4,271 (рис.55);

t_γ – критична точка розподілу Стьюдента для рівня довіри γ = 0,95 і кількості ступенів вільності $k2 = n - 2 = 13$(див п1.7);

n =15 – кількість спостережень; u_{np} відповідає x_{np};

$\bar{u}$ – середнє значення (рис.58);

$D(u) = 0{,}3667$ – дисперсія вибірки (див. рис.58);

Додаємо ще один стовпець *Delta* і виконуємо розрахунок так, як показано у таблицях (див. рис.60, 61).

10 Далі розраховуємо ліву межу довірчого інтервалу за формулою $v_{\min} = v(u) - delta$, праву – $v_{\max} = v(u) + delta$ (див.рис.60, 61).

11 Так як лінеаризована модель адекватна, то й нелінійна модель $y = a \ln x + b$ вважається адекватною. Параметри a і b знаходимо за формулами зворотного перетворення (табл.17): $a = b_1 = 12{,}23;\ b = b_0 = -75{,}72.$

Отримали модель зв'язку продуктивності праці (x) і рівня рентабельності (y) по плодоовочевих консервних заводах області за рік: $y = -75{,}72 + 12{,}23 \cdot \ln x$.

12 За формулами зворотнього перетворення (див.табл.17) перераховуємо значення $y(x),\ y_{\min},\ y_{\max}$. Для нашої моделі $y(x) = v(u),\ y_{\min} = v_{\min},\ y_{\max} = v_{\max}$. Додаємо 3 відповідних стовпця до таблиці (див. рис. 60 і 61).

У 17-му рядку таблиці (рис. 60) знаходиться прогноз.

Точковий: для заданого значення продуктивності праці 5000грн (x_{np}) прогнозоване значення рівня рентабельності ($y(x_{np})$) становить 28,45%.

Прогноз з урахуванням довірчого інтервалу: для заданого значення продуктивності праці 5000грн прогнозований рівень рентабельності лежить у межах від 18,88% ($y_{\min}$) до 38,01% ($y_{\max}$).

13 Будуємо довірчу область: **Діаграма – Точкова – Точкова**, вхідний інтервал – стовпці $x,\ y,\ y(x),\ y_{\min},\ y_{\max}$ (рис 59).

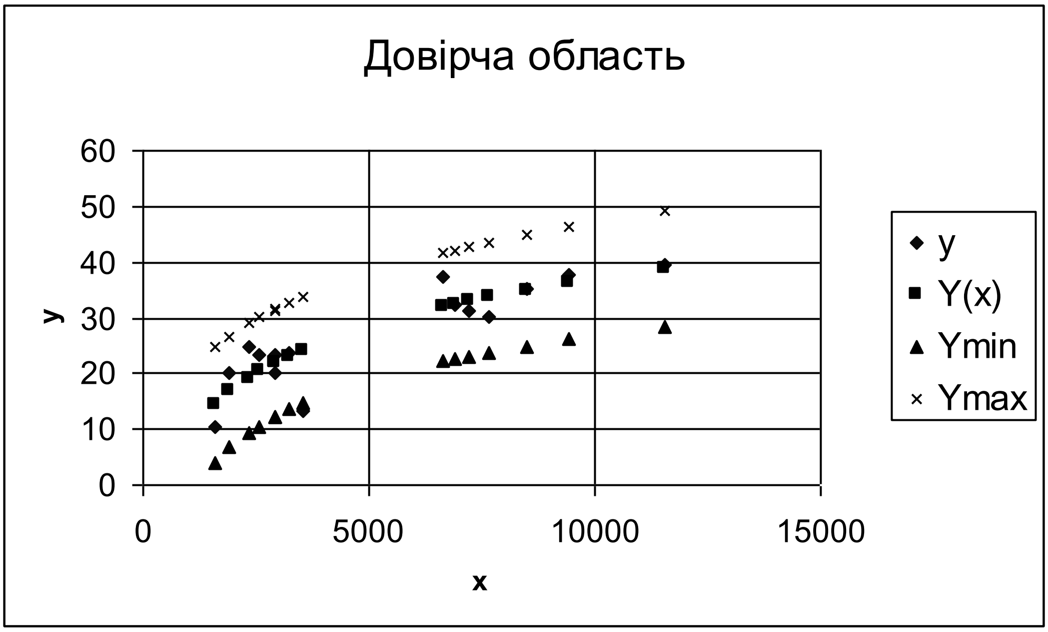


Рисунок 59

x	y	u	v	V(u)	Delta	Vmin	V_max	Y(x)	Ymin	Ymax	Ex
7224	31,3	8,885164	31,3	32,94797	9,809193	23,13878	42,75717	32,95	23,13878	42,75717	0,371
3223	23,7	8,078068	23,7	23,07689	9,578159	13,49873	32,65505	23,08	13,49873	32,65505	0,530
1586	10,3	7,36897	10,3	14,40436	10,27789	4,126474	24,68225	14,40	4,126474	24,68225	0,849
7671	30,2	8,945202	30,2	33,68226	9,869995	23,81227	43,55226	33,68	23,81227	43,55226	0,363
6921	32,2	8,842316	32,2	32,42392	9,76931	22,65461	42,19323	32,42	22,65461	42,19323	0,377
2340	24,8	7,757906	24,8	19,16119	9,795371	9,365819	28,95656	19,16	9,365819	28,95656	0,638
3520	13,4	8,166216	13,4	24,15497	9,54879	14,60619	33,70376	24,15	14,60619	33,70376	0,506
2569	23,4	7,851272	23,4	20,30309	9,714472	10,58862	30,01756	20,30	10,58862	30,01756	0,602
1920	20,1	7,56008	20,1	16,74171	10,01243	6,729276	26,75414	16,74	6,729276	26,75414	0,731
9410	37,9	9,149528	37,9	36,18125	10,11853	26,06272	46,29978	36,18	26,06272	46,29978	0,338
2901	20	7,972811	20	21,78955	9,630654	12,1589	31,42021	21,79	12,1589	31,42021	0,561
8492	35,1	9,04688	35,1	34,92582	9,985783	24,94004	44,9116	34,93	24,94004	44,9116	0,350
6630	37,2	8,79936	37,2	31,89856	9,732304	22,16626	41,63086	31,90	22,16626	41,63086	0,383
2911	23,2	7,976252	23,2	21,83164	9,628641	12,203	31,46028	21,83	12,203	31,46028	0,560
11540	39,4	9,353575	39,4	38,67681	10,42725	28,24956	49,10406	38,68	28,24956	49,10406	0,316
5000		**8,517193**		**28,44755**	**9,565487**	**18,88207**	**38,01304**	**28,45**	**18,88207**	**38,01304**	**0,430**

b0=	-75,72		a=	12,23
b1=	12,23		b=	-75,72
Стандартная ошибка	4,271087			
t_кр=	2,160369			
Среднее (u)	8,314288			
Дисперсия выборки (u)	0,366721			

Рисунок 60

14 Розраховуємо коефіцієнт еластичності за формулою

$$E_x = \frac{x}{y(x)} \cdot y'_x.$$

Для моделі $y = a\ln x + b$ (див. табл.9) –

$$E_x = \frac{a}{y(x)} = \frac{12,23}{y(x)}.$$

Додаємо ще один стовпець E_x до таблиці (див.рис.60, 61). Одержали $E_x = 0,43$.

Коефіцієнт еластичності показує, що при зміні значення фактора продуктивності праці $x_{пр} = 5000$ грн на 1% значення рівня рентабельності ($y(x)$) збільшиться від 28,45% на 0,43%.

Висновки

Виконаний статистичний аналіз показує, що залежність рівня рентабельності по плодоовочевих заводах області за рік $y\%$ від продуктивності праці x(грн) може бути описана логарифмічною моделлю $y = -75,72 + 12,23 \cdot \ln x$. При цьому всі коефіцієнти моделі статистично значущі, а сама модель адекватна.

Для заданої продуктивності праці $x_{пр} = 5000$ грн прогноз-ване значення рівня рентабельності становить 28,45%, або з н-дійністю 0,95 лежить у межах від 18,88% до 38,01%.

Коефіцієнт еластичності $E_x = 0,43$ показує, що при зміні продуктивності праці від 5000грн на 1% відповідний рівень рентабельності збільшується на 0,43%.

2.3 Лінійна двофакторна модель. План побудови моделі

Будуємо лінійну двофакторну модель у вигляді $y = b_0 + b_1x_1 + b_2x_2$ для показника y і факторів x_1 і x_2.

Послідовність дій:

1 Вводимо дані.

2 Визначаємо основні статистики.

3 Перевіряємо фактори на колінеарність.

4 Записуємо лінійну модель вигляду $y = b_0 + b_1x_1 + b_2x_2$.

5 Визначаємо загальну якість моделі за коефіцієнтом детермінації R^2. Перевіряємо одержану модель на адекватність за допомогою критерію Фішера. Усі подальші розрахунки виконуються тільки за умови адекватності моделі початковим статистичним даним.

6 Перевіряємо статистичну значущість коефіцієнтів моделі.

7 За одержаною моделлю розраховуємо значення показника у для всіх точок вибірки і в точці прогнозу (точку прогнозу вибираємо довільно з області прогнозу).

8 Розраховуємо часткові коефіцієнти еластичності.

9 Використовуючи одержані дані та теоретичні відомості, робимо економетричний аналіз – описуємо процес побудови моделі та всі супутні розрахунки.

Приклад виконання

Економічні дані

Рівень рентабельності (y), % і значення факторів, його формуючих, - продуктивності праці (x_1), грн. і фондовіддачі (x_2), % по плодоовочевих консервних заводах області за рік задані в таблиці (див. рис.62).

№ п/п	x1	x2	y
1	7224	0,9	31,3
2	3223	0,67	23,7
3	1586	0,16	10,3
4	7671	0,89	30,2
5	6921	1,03	32,2
6	2340	0,75	24,8
7	3520	0,26	13,4
8	2569	0,65	23,4
9	1920	0,48	20,1
10	9410	1,19	37,9
11	2901	0,44	20
12	8492	1,12	35,1
13	6630	1,18	37,2
14	2911	0,63	23,2
15	11540	1,24	39,4

Рисунок 62

Виконання завдання:

1 Вводимо дані у вигляді таблиці (рис. 62).

2 Визначаємо основні статистики. Результат отримаємо у вигляді таблиці (рис.63).

x1		*x2*		*y*	
Среднее	**5257,2**	**Среднее**	**0,772667**	**Среднее**	**26,81333**
Стандартная ошибка	816,7708	Стандартная ошибка	0,088809	Стандартная ошибка	2,289788
Медиана	3520	Медиана	0,75	Медиана	24,8
Мода	#Н/Д	Мода	#Н/Д	Мода	#Н/Д
Стандартное отклонени	**3163,34**	**Стандартное отклонени**	**0,343957**	**Стандартное отклонени**	**8,868312**
Дисперсия выборки	**10006718**	**Дисперсия выборки**	**0,118307**	**Дисперсия выборки**	**78,64695**
Эксцесс	-0,97862	Эксцесс	-0,97996	Эксцесс	-0,76191
Асимметричность	0,55141	Асимметричность	-0,2582	Асимметричность	-0,27517
Интервал	9954	Интервал	1,08	Интервал	29,1
Минимум	**1586**	**Минимум**	**0,16**	**Минимум**	**10,3**
Максимум	**11540**	**Максимум**	**1,24**	**Максимум**	**39,4**
Сумма	78858	Сумма	11,59	Сумма	402,2
Счет	15	Счет	15	Счет	15

Рисунок 63

Отримали, що середнє значення продуктивності праці – 5257,2грн, вона лежить у межах від 1586грн (мінімум) до 11540грн (максимум), що задає область прогнозів – діапазон, з якого припустимо вибирати значення фактора x_1 для прогнозу рівня рентабельності *y*. $x_{1cp} = 5257{,}2$ грн. задає центр області прогнозу. Стандартне відхилення 3163,34 характеризує середнє значення розсіювання значень x_1 відносно x_{1cp}.

Середнє значення фондовіддачі 0,772667%, вона лежить у межах від 10,3% (мінімум) до 39,4% (максимум), що задає область прогнозів – діапазон, з якого припустимо вибирати значення фактора x_2 для прогнозу рівня рентабельності *y*. $x_{2cp} = 26{,}81333$ грн. задає центр області прогнозу. Стандартне відхилення 8,868312 характеризує середнє значення розсіювання значень x_2 відносно x_{2cp}.

3 Для того, щоб перевірити фактори на колінеарність, знайдемо коефіцієнт кореляції: **Сервіс – Аналіз даних – Кореляція, Ок**. Результат отримаємо у вигляді таблиці (рис.64).

	x1	x2	y
x1	1		
x2	0,863508	1	
y	0,876868	0,995152	1

Рисунок 64

На перетині стовпця x_1 і рядка x_2: $r_{x_1 x_2} = 0{,}86$. Визначимо за допомогою критерію Стьюдента, чи є це значення коефіцієнта статистично значущим.

Знаходимо фактичне значення t_r за формулою

$$t\_r = \frac{\sqrt{n-3}}{2}\ln\left(\frac{1+r_{x_1 x_2}}{1-r_{x_1 x_2}}\right).$$

У Excel ця формула виглядає так:

=корень(15-3)×ln((1+0,86)/(1-0,86))/2.

Отже t_r=4,527.

Критичне значення $t_{кр}$ визначається за допомогою стандартної функції Excel (див.1.7): = СТЬЮДРАСПОБР(0,05;15-2), $t_{кр}$=2,16.

Отже $|t\_r| > t_{кр}$, тому коефіцієнт кореляції статистично значущий. Отже, між факторами x_1 і x_2 є зв'язок, але він не тісний.

4 Будемо шукати модель зв’язку між статистичними даними у вигляді $y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2$.

Параметри моделі b_0, b_1, b_2 знайдемо за МНК: **Сервіс – Аналіз даних – Регресія, Ок** (див.п 1.6). Результат отримаємо у такому вигляді (рис.65).

ВЫВОД ИТОГОВ						
Регрессионная статистика						
Множественный R	**0,995760018**					
R-квадрат	**0,991538014**					
Нормированный R-квад	0,990127683					
Стандартная ошибка	0,881151296					
Наблюдения	15					
Дисперсионный анализ						
	df	*SS*	*MS*	*F*	***Значимость F***	
Регрессия	2	1091,740202	545,870101	703,0534416	**0,000000000004**	
Остаток	12	9,31713128	0,776427607			
Итого	14	1101,057333				
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	***P-Значение***	*Нижние 95%*	*Верхние 95%*
Y-пересечение	**7,158106925**	0,59030655	12,1260842	**0,0000000430**	5,871939443	8,444274407
x1	**0,000193391**	0,000147612	1,310134936	**0,2146747585**	-0,000128227	0,00051501
x2	**24,12234187**	1,357569913	17,76876582	**0,0000000006**	21,16445113	27,08023261

Рисунок 65

Параметри моделі беремо у стовпці «коефіцієнти»: $b_0 = 7{,}158, b_1 = 0{,}00019, b_2 = 24{,}12$. Можна записати модель:

$$y = 7{,}158 + 0{,}00019 \cdot x_1 + 24{,}12 \cdot x_2 .$$

5 Визначаємо загальну якість моделі за коефіцієнтом детермінації: $R^2 = 0{,}99$. Це означає, що 99% початкових даних пояснюються моделлю – загальна якість моделі добра. Множинний R=0,996. Це аналог коефіцієнту кореляції – зв’язок між y та x_1 і x_2 тісний.

Перевіряємо модель на адекватність за допомогою критерію Фішера. Значущість $F = 0{,}4 \cdot 10^{-12} < 0{,}05$ - модель адекватна початковим даним.

6 Перевіряємо статистичну значущість коефіцієнтів моделі за допомогою крітерію Стьюдента.

Для коефіцієнта b_0: P – значення дорівнює $0{,}4 \cdot 10^{-7} < 0{,}05 \Rightarrow$ коефіцієнт b_0 статистично значущий.

Для коефіцієнта b_1: P – значення дорівнює $0{,}2147 > 0{,}05 \Rightarrow$ коефіцієнт b_1 статистично не значущий.

Для коефіцієнта b_2: P – значення дорівнює $0{,}6 \cdot 10^{-9} < 0{,}05 \Rightarrow$ коефіцієнт b_2 статистично значущий.

7 За одержаною моделлю розраховуємо прогнозовані значення показника y для всіх точок вибірки і в точці прогнозу. Точку прогнозу так, як і в попередньому прикладі, вибираємо з області прогнозів:

$$\begin{cases} 1586 \le x_1 \le 11540, \\ 0{,}16 \le x_2 \le 1{,}24. \end{cases}$$

Візьмемо $x_{1пр} = 5000, x_{2пр} = 1,$ додамо в таблицю 18 ще один стовпець $y(x_1, x_2)$ і виконаємо розрахунок так, як це показано у таблиці 19.

Отримаємо прогноз: для заданого значення продуктивності праці 5000грн і фондовіддачі 1%. Прогнозоване значення рівня рентабельності становить 32,25% (табл.18).

Таблиця 18

	A	B	C	D	E	F	G
1	**x1**	**x2**	**y**	**y(x1,x2)**	**Ex1**	**Ex2**	
2	7224	0,9	31,3	30,27	0,046	0,717	
3	3223	0,67	23,7	23,94	0,026	0,675	
4	1586	0,16	10,3	11,32	0,027	0,341	
5	7671	0,89	30,2	30,11	0,049	0,713	
6	6921	1,03	32,2	33,34	0,040	0,745	
7	2340	0,75	24,8	25,70	0,018	0,704	
8	3520	0,26	13,4	14,11	0,048	0,444	
9	2569	0,65	23,4	23,33	0,021	0,672	
10	1920	0,48	20,1	19,11	0,019	0,606	
11	9410	1,19	37,9	37,68	0,048	0,762	
12	2901	0,44	20	18,33	0,031	0,579	
13	8492	1,12	35,1	35,82	0,046	0,754	
14	6630	1,18	37,2	36,90	0,035	0,771	
15	2911	0,63	23,2	22,92	0,025	0,663	
16	11540	1,24	39,4	39,30	0,057	0,761	
17	**5000**	**1**		**32,25**	**0,030**	**0,748**	
18							
19	b0=	7,158107					
20	b1=	0,000193			*x1*	*x2*	*y*
21	b2=	24,12234		x1	1		
22				x2	0,863508	1	
23				y	0,876868	0,995152	1
24							
25							
26					t_r=	4,52749	
27					t_кр=	2,160369	

8 Розраховуємо часткові коефіцієнти еластичності:

- для фактора x_1:

$$E_{x_1} = \frac{x_1}{y(x_1, x_2)} \cdot y'_{x_1} = \frac{b_1 x_1}{y(x_{1,} x_2)};$$

- для фактора x_2:

$$E_{x_2} = \frac{x_2}{y(x_1, x_2)} \cdot y'_{x_2} = \frac{b_2 x_2}{y(x_{1,} x_2)}.$$

За цими формулами виконуємо розрахунок так, як показано у таблиці 19. Отримали: $E_{x_1} = 0{,}03, E_{x_2} = 0{,}748$. Це означає, що при збільшенні $x_1 = 5000$грн. на 1% значення $y(x_1, x_2) = 32{,}25\%$ збільшиться на 0,03%, а при збільшенні $x_2 = 1\%$ на 1% значення $y(x_1, x_2) = 32{,}25\%$ збільшиться на 0,748%.

$E_{x_1} < E_{x_2}$, тобто зміна значення першого фактора менше впливає на зміну рівня рентабельності, ніж зміна другого фактора.

Незначний вплив продуктивності праці (x_1) на рівень рентабельності (y) у порівнянні з впливом фондовіддачі (x_2) випливає також з результатів статистичного аналізу коефіцієнтів моделі: коефіцієнт b_1 при змінній x_1 статистично не значущий.

Висновки

Проведений статистичний аналіз показує, що вибрані фактори: продуктивність праці ($x_1, грн$) і фондовіддача ($x_2, \%$) не колінеарні. Побудована лінійна модель зв'язку між цими факторами і рівнем рентабельності (y,%) по плодоовочевих консервних заводах області за рік: $y = 7{,}158 + 0{,}00019 \cdot x_1 + 24{,}12 x_2$.

Ця модель адекватна. Статистично значущими є два коефіцієнти: $b_0 = 7{,}158\, i\, b_2 = 24{,}12$.

Для заданого значення продуктивності праці 5000грн і фондовіддачі 1% прогнозоване значення рівня рентабельності становить 32,25%.

Часткова еластичність відносно кожного з факторів дорівнює: $E_{x_1} = 0{,}03$, $E_{x_2} = 0{,}748$. Ці коефіцієнти показують на скільки відсотків збільшиться прогнозоване значення рівня рентабельності при збільшенні значення відповідного фактора на 1%.

2.4 Ступенева двофакторна модель. План побудови моделі

Ступенева двофакторна модель має вигляд $y = Ax_1^{a_1} x_2^{a_2}$ і доволі часто використовується. Наведений нижче план може бути використаний і для інших багатофакторних нелінійних моделей. Формули для заміни ті ж, що й у п.2.2.

Послідовність дій:

1 Вводимо дані.

2 Визначаємо основні статистики.

3 Перевіряємо фактори на колінеарність.

4 Перераховуємо дані вибірки за формулами заміни. Записуємо лінійну модель для нових змінних.

5 Визначаємо загальну якість моделі за коефіцієнтом детермінації R^2. Перевіряємо одержану модель на адекватність за допомогою критерію Фішера. Усі подальші розрахунки виконуються тільки за умови адекватності моделі початковим статистичним даним.

6 Перевіряємо статистичну значущість коефіцієнтів моделі.

7 За одержаною моделлю розраховуємо значення показника y для всіх точок вибірки і в точці прогнозу (точку прогнозу вибираємо довільно з області прогнозу).

8 За допомогою формул зворотного перетворення повертаємось до ступеневої моделі.

9 Розраховуємо часткові коефіцієнти еластичності.

10 Використовуючи одержані дані та теоретичні відомості, робимо економетричний аналіз – описуємо процес побудови моделі і всі супутні розрахунки.

Приклад виконання

Економічні дані

Побудуємо ступеневу модель на тих даних, що й у попередньому прикладі.

Виконання завдання

Пункти плану 1-3 повністю співпадають з попереднім прикладом.

4 Перераховуємо дані вибірки за формулами заміни: $u_1 = \ln(x_1)$, $u_2 = \ln(x_2)$, v=ln(y).

Будемо шукати формулу зв'язку між ними у вигляді лінійної моделі: $v = b_0 + b_1 \cdot u_1 + b_2 u_2$. Параметри моделі b_0, b_1, b_2 знайдемо за МНК: **Сервіс – Аналіз даних – Регресія, Ок,** «Вхідний параметр Y» - виділяємо стовпець v, «Вхідний параметр X» - виділяємо стовпці u_1, u_2 (див.п1.6). Результат отримаємо у такому вигляді (рис.66).

ВЫВОД ИТОГОВ						
Регрессионная статистика						
Множественный R	**0,996705178**					
R-квадрат	**0,993421211**					
Нормированный R-квадрат	0,992324746					
Стандартная ошибка	0,03383272					
Наблюдения	15					
Дисперсионный анализ						
	df	*SS*	*MS*	*F*	***Значимость F***	
Регрессия	2	2,074161401	1,037080701	906,0219795	**8,10729E-14**	
Остаток	12	0,013735835	0,001144653			
Итого	14	2,087897236				
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	***P-Значение***	*Нижние 95%*	*Верхние 95%*
Y-пересечение	**2,856659241**	0,199942465	14,28740632	**6,77183E-09**	2,421022033	3,292296449
u1	**0,071465194**	0,022904697	3,120110902	**0,008852992**	0,021560147	0,121370242
u2	**0,58787276**	0,024875769	23,63234482	**1,97185E-11**	0,533673114	0,642072405

Рисунок 66

Параметри моделі беремо у стовпці «коефіцієнти»: $b_0 = 2{,}86, b_1 = 0{,}07, b_2 = 0{,}59.$ Можна записати модель:

$$v = 2{,}86 + 0{,}07 \cdot u_1 + 0{,}59 u_2 .$$

5 Визначаємо загальну якість моделі за коефіцієнтом детермінації: $R^2 = 0{,}99$. Це означає, що 99% початкових даних пояснюються моделлю – загальна якість моделі добра.

Множинний R=0,997, це означає, що зв'язок між показником v і факторами u_1, u_2 тісний.

Перевіряємо модель на адекватність за допомогою критерію Фішера. Значущість $F = 8{,}1 \cdot 10^{-14} < 0{,}05$ - модель адекватна початковим даним.

6 Перевіряємо статистичну значущість коефіцієнтів моделі за допомогою крітерію Стьюдента.

Для коефіцієнта b_0: P – значення дорівнює $6{,}77 \cdot 10^{-9} < 0{,}05 \Rightarrow$ коефіцієнт b_0 статистично значущий.

Для коефіцієнта b_1: P – значення дорівнює $0{,}009 < 0{,}05 \Rightarrow$ коефіцієнт b_1 статистично значущий.

Для коефіцієнта b_2: P – значення дорівнює $1{,}97 \cdot 10^{-11} < 0{,}05 \Rightarrow$ коефіцієнт b_2 статистично значущий.

7 За одержаною моделлю розраховуємо прогнозовані значення показника v для всіх точок вибірки і в точці прогнозу. Точку прогнозу так, як і в попередньому прикладі, вибираємо з області прогнозів:

$$\begin{cases} 1586 \le x_1 \le 11540, \\ 0{,}16 \le x_2 \le 1{,}24. \end{cases}$$

Візьмемо $x_{1пр} = 5000, x_{2пр} = 1,$ додамо ще один стовпець $v(u_1, u_2)$ і виконаємо розрахунок так, як це показано у таблиці 20.

8 За допомогою формул зворотного перетворення отримаємо:

$$y = e^{v} = \exp(v),\ A = e^{b_0} = \exp(b_0),\ a_1 = b_1,\ a_2 = b_2.$$

Додамо ще один стовпець $y(x_1 x_2)$ і виконаємо розрахунок так, як це показано у таблиці 20. Перераховуємо коефіцієнти:

$$A = 17{,}403,\ a_1 = 0{,}071,\ a_2 = 0{,}588.$$

Степенева формула залежності рівня рентабельності по плодоовочевих консервних заводах області за рік від продуктивності праці і фондовіддачі: $y = 17{,}403 x_1^{0{,}071} x_2^{0{,}588}$.

Для заданих $x_{1пр} = 5000\, грн$ і $x_{2пр} = 1\%$ прогнозований рівень рентабельності становить 31,99%(табл.21):

Таблиця 21

	A	B	C	D	E	F	G	H
1	**x1**	**x2**	**y**	**u1**	**u2**	**v**	**v(u1,u2)**	**y(x1,x2)**
2	7224	0,9	31,3	8,885	-0,105	3,444	3,430	30,87
3	3223	0,67	23,7	8,078	-0,400	3,165	3,199	24,50
4	1586	0,16	10,3	7,369	-1,833	2,332	2,306	10,03
5	7671	0,89	30,2	8,945	-0,117	3,408	3,427	30,80
6	6921	1,03	32,2	8,842	0,030	3,472	3,506	33,31
7	2340	0,75	24,8	7,758	-0,288	3,211	3,242	25,58
8	3520	0,26	13,4	8,166	-1,347	2,595	2,648	14,13
9	2569	0,65	23,4	7,851	-0,431	3,153	3,165	23,68
10	1920	0,48	20,1	7,560	-0,734	3,001	2,965	19,40
11	9410	1,19	37,9	9,150	0,174	3,635	3,613	37,07
12	2901	0,44	20	7,973	-0,821	2,996	2,944	18,99
13	8492	1,12	35,1	9,047	0,113	3,558	3,570	35,51
14	6630	1,18	37,2	8,799	0,166	3,616	3,583	35,97
15	2911	0,63	23,2	7,976	-0,462	3,144	3,155	23,45
16	11540	1,24	39,4	9,354	0,215	3,674	3,652	38,54
17	**5000**	**1**		8,517	0,000		3,465	31,99
18								
19	b0=	2,857		A=	17,403			
20	b1=	0,071		a1=	0,071			
21	b2=	0,588		a2=	0,588			
22								
23		*x1*	*x2*	*y*				
24	x1	1						
25	x2	0,863508	1					
26	y	0,876868	0,995152	1				
27								
28		t_r=	4,52749					
29		t_кр=	2,160369					

9 Для ступеневої моделі часткова еластичність відносно кожного з факторів є величиною постійною і дорівнює показнику ступеня при відповідній змінній: $E_{x_1} = 0{,}071$, $E_{x_2} = 0{,}588$.

Це означає, що при збільшенні фактора $x_{1пр} = 5000\,грн$ на 1% значення показника y=31,99% збільшиться на 0,071%, а при збільшенні фактора $x_{2пр} = 1\%$ на 1% значення показника збільшиться на 0,588%.

Висновки

Проведений статистичний аналіз показує, що вибрані фактори: продуктивність праці ($x_1, грн$) і фондовіддача (x_2,%) не колінеарні. Побудована ступенева модель зв'язку між цими факторами і рівнем рентабельності (y,%) по плодоовочевих консервних заводах області за рік: $y = 17{,}403 x_1^{0{,}071} x_2^{0{,}588}$.

Для заданої продуктивності 5000грн і фондовіддачі 1% прогнозований рівень рентабельності становить 31,99%.

Часткова еластичність відносно кожного з факторів дорівнює: $E_{x_1} = 0{,}071$, $E_{x_2} = 0{,}588$. Ці коефіцієнти показують на скільки відсотків збільшиться прогнозоване значення рівня рентабельності при збільшенні значення відповідного фактора на 1%. $E_{x_1} < E_{x_2}$, тобто зміна значення фондовіддачі більше впливає на рівень рентабельності, ніж продуктивність праці.

2.5 Система одночасних рівнянь

Розглянемо побудування систему одночасних рівнянь на прикладі статичної моделі Кейнса для опису народного господарства країни, яка в найпростішому варіанті має наступний вигляд:

$$\begin{cases} C = a + b \cdot y + \varepsilon, \\ \quad y = C + I, \end{cases}$$

де C – особисте споживання в постійних цінах,

y – національний дохід в постійних цінах,

I – інвестиції в постійних цінах,

ε – випадкова складова.

Завдання: За даними, наведеними в таблиці 22, побудуйте систему одночасних рівнянь. Визначте параметри рівнянь за допомогою непрямого методу найменших квадратів (НМНК). Проаналізуйте модель, яку отримали.

Приклад виконання

Економічні дані задані в таблиці 22.

Таблиця 22

Рівень виробництва і доходу, млрд дол.	**Споживання С, млрд дол.**	**Інвестиції І, млрд дол.**
370	365	5
415	397	18
430	409	21
456	430	26
486	450	36
490	455	35
505	467	38
520	479	41
546	500	46
567	516	51

Виконання завдання

По суті задачі, y і C – ендогенні змінні, а I – екзогенна змінна. Визначимо ідентифікованість 1-го рівняння:

$n_1 = 2$ – кількість ендогенних змінних (y і C), що входять до 1-го рівняння,

$m = 1$ – загальна кількість екзогенних змінних у системі,

$m_1 = 0$ – кількість екзогенних змінних, що входять до 1-го рівняння.

Перевіряємо ідентифікованість за допомогою рахункового правила:

$$n_s - 1 \le m - m_s.$$

Одержимо $2-1=1-0$, $1=1$. Перше рівняння точно ідентифіковане. Отже, система точно ідентифікована, і для визначення параметрів рівнянь можна використовувати НМНК.

Для запису наведеної форми моделі перетворимо рівняння структурної форми моделі.

Підставимо значення y з 2-го рівняння в 1-е. Одержимо:

$$C = \frac{a}{1-b} + \frac{b}{1-b} \cdot I + \frac{\varepsilon}{1-b}.$$

Тепер підставимо значення C з 1-го рівняння в 2-е. Одержимо:

$$y = \frac{a}{1-b} + \frac{1}{1-b} \cdot I + \frac{\varepsilon}{1-b}.$$

Позначимо ендогенні змінні C і y через $Y1$і $Y2$, а екзогенну змінну I через $X1$. Тоді систему одночасних рівнянь у наведеній формі можна записати у вигляді

$$Y1 = b_{10} + b_{11}X1 + \varepsilon_1,$$
$$Y2 = b_{20} + b_{21}X1 + \varepsilon_2.$$

Тут $b_{10} = b_{20} = \dfrac{a}{1-b}$, $b_{11} = \dfrac{b}{1-b}$, $b_{21} = \dfrac{1}{1-b}$.

Параметри кожного з рівнянь цієї системи знаходимо за допомогою МНК.

Для першого рівняння одержимо наступні результати (рис.67).

ВИСНОВОК ПІДСУМКІВ

Регресійна статистика	
Множинний R	**0,993323**
R-квадрат	0,986692
Нормований R-квадрат	**0,985028**
Стандартна помилка	5,771513
Спостереження	10

Дисперсійний аналіз

	df	*SS*	*MS*	F	Значущість F
Регресія	1	19757,12	19757,12	593,1223	**8,62366E-09**
Залишок	8	266,4829	33,31036		
Разом	9	20023,6			

	Коефіцієнти	*Стандартна помилка*	*t-статистика*	*P-значення*
Y-перетин	**341,1915**	4,704804	72,51982	**1,46E-12**
X1	**3,331498**	0,136794	24,3541	**8,62E-09**

Рисунок 67

Аналіз одержаних результатів проводиться таким же чином, як і при побудові лінійної однофакторної моделі.

Множинний R = 0,993323, це значить, що між $X1$ і $Y1$ існує тісний лінійний зв'язок. Коефіцієнт детермінації $R^2 = 0{,}985$.

Коефіцієнти для 1-го рівняння $b_{10} = 341{,}19$, $b_{11} = 3{,}33$. При перевірці їх статистичної значущості за критерієм Стьюдента набудуть наступних значень рівня значущості: $\alpha_{b10} = 1{,}46 \cdot 10^{-12}$, $\alpha_{b11} = 8{,}62 \cdot 10^{-9}$. Оскільки ці значення менше 0,05, то з імовірністю 0,95 можна стверджувати, що обидва коефіцієнти статистично значущі і можуть бути включені в модель.

Таким чином, перше рівняння наведеної форми моделі набуде вигляду: $Y1 = 341{,}19 + 3{,}33X1 + \varepsilon_1$. При перевірці цього рівняння на адекватність за критерієм Фішера одержимо $\alpha_{F_{спост}} = 8{,}62 \cdot 10^{-9}$. Оскільки $\alpha_{F_{спост}} < 0{,}05$, то з імовірністю 0,95 можна стверджувати, що рівняння адекватне початковим даним.

Здійснемо аналогічний розрахунок для другого рівняння наведеної форми моделі (рис. 68).

ВИСНОВОК
ПІДСУМКІВ

Регресійна статистика	
Множинний R	**0,9960342**
R-квадрат	0,9920842
Нормований R-квадрат	**0,9910947**
Стандартна помилка	5,771513
Спостереження	10

Дисперсійний аналіз

	df	*SS*	*MS*	F	Значущість F
Регресія	1	33398,02	33398,02	1002,631	**1,07703E-09**
Залишок	8	266,4829	33,31036		
Разом	9	33664,5			

	Коефіцієнти	*Стандартна помилка*	*t-статистика*	*Р-значення*
Y-перетин	**341,19151**	4,704804	72,51982	**1,46E-12**
X1	**4,3314982**	0,136794	31,66436	**1,08E-09**

Рисунок 68

Множинний R = 0,9960342, це значить, що між $X1$ і $Y1$ існує тісний лінійний зв'язок. Коефіцієнт детермінації $R^2 = 0{,}991$.

Коефіцієнти для 2-го рівняння $b_{20} = 341{,}19$, $b_{21} = 4{,}33$. При перевірці їх статистичної значущості за критерієм Стьюдента набудуть наступних значень рівня значущості: $\alpha_{b20} = 1{,}46 \cdot 10^{-12}$, $\alpha_{b21} = 1{,}08 \cdot 10^{-9}$. Оскільки ці значення менше

0,05, то з імовірністю 0,95 можна стверджувати, що обидва коефіцієнти статистично значущі та можуть бути включені до моделі.

Одержали друге рівняння наведеної форми моделі $Y2 = 341{,}19 + 4{,}33X1 + \varepsilon_2$. При перевірці цього рівняння на адекватність за критерієм Фішера одержимо $\alpha_{F_{спост}} = 1{,}077 \cdot 10^{-9}$. Оскільки $\alpha_{F_{спост}} < 0{,}05$, то з імовірністю 0,95 можна стверджувати, що рівняння адекватне початковим даним.

Запишемо приведену форму моделі:

$$Y1 = 341{,}19 + 3{,}33x1 + \varepsilon_1,$$
$$Y2 = 341{,}19 + 4{,}33x1 + \varepsilon_2.$$

Перейдемо від наведеної форми моделі до структурної.

Як було записано раніше:

$$b_{10} = b_{20} = \frac{a}{1-b}, \quad b_{11} = \frac{b}{1-b}, \quad b_{21} = \frac{1}{1-b}.$$

Підставимо набуті значення b_{10}, b_{20}, b_{11} і b_{21}:

$$341{,}19 = \frac{a}{1-b},$$
$$4{,}33 = \frac{1}{1-b},$$
$$3{,}33 = \frac{b}{1-b}$$

Звідси:

$$a = 78{,}81,$$
$$b = 0{,}769.$$

Запишемо структурну форму моделі:

$$\begin{cases} C = 78{,}81 + 0{,}769 \cdot y + \varepsilon, \\ \qquad y = C + I. \end{cases}$$

Висновки

Проведений статистичний аналіз економічних даних дозволив записати статичну модель Кейнса для опису народного господарства країни:

$$\begin{cases} C = 78{,}81 + 0{,}769 \cdot y + \varepsilon, \\ y = C + I. \end{cases}$$

У ній структурний коефіцієнт b характеризує *граничну схильність до споживання*. Тут $b = 0{,}769$, тобто, з кожної додаткової тисячі доходу на споживання витрачається 769 гр.од., а 231 гр.од. інвестується.

Інвестиційний мультиплікатор споживання – це коефіцієнт $b_{11} = \dfrac{b}{1-b} = M_c = 3{,}33$. Ця величина означає, що додаткові вкладення у розмірі 1 тис. гр.од. викличуть за інших рівних умов додаткове збільшення споживання на 3,33 тис. гр.од.

Інвестиційний мультиплікатор національного доходу – це коефіцієнт $b_{21} = \dfrac{1}{1-b} = M_y = 4{,}33$. Ця величина означає, що додаткові інвестиції у розмірі 1 тис.гр.од. приведуть за інших рівних умов до додаткового доходу на 4,33 тис.гр.од.

2.6 Метод ковзних середніх для згладжування часових рядів

Економічні дані

Зібрані відомості про щомісячні перевезення вугілля за період із січня 1995р. по січень 2000р. (млн т/місяць). Дані задані в таблиці 23.

Таблиця 23

A	B
Дата	**X**
01.01.1995	14
01.02.1995	28
01.03.1995	25
01.04.1995	17
01.05.1995	31
01.06.1995	44

Продовження таблиці 23

Дата	X
01.07.1995	44
01.08.1995	32
01.09.1995	15
01.10.1995	14
01.11.1995	14
01.12.1995	11
01.01.1996	22
01.02.1996	37
01.03.1996	31
01.04.1996	21
01.05.1996	45
01.06.1996	66
01.07.1996	66
01.08.1996	54
01.09.1996	29
01.10.1996	10
01.11.1996	36
01.12.1996	41
01.01.1997	46
01.02.1997	74
01.03.1997	59
01.04.1997	68
01.05.1997	74
01.06.1997	95
01.07.1997	95
01.08.1997	80
01.09.1997	58
01.10.1997	42
01.11.1997	62
01.12.1997	67
01.01.1998	76
01.02.1998	89
01.03.1998	77
01.04.1998	79
01.05.1998	114
01.06.1998	125
01.07.1998	138
01.08.1998	106
01.09.1998	87
01.10.1998	68
01.11.1998	90
01.12.1998	92
01.01.1999	92
01.02.1999	132
01.03.1999	131
01.04.1999	125
01.05.1999	139
01.06.1999	160
01.07.1999	168
01.08.1999	133
01.09.1999	107
01.10.1999	76
01.11.1999	97
01.12.1999	100
01.01.2000	84

Згладити ряд, виявити загальну тенденцію, зробити прогноз на наступний місяць.

Виконання завдання

1 Уводимо дані на лист Excel у стовпець (стовпець А – «дата», стовпець В – *х*, млн т/місяць). Щоб не вводити кожну дату окремо, треба: визначити для стовпця А формат «Дата», набрати кілька дат (наприклад 3), виділити блок з цими датами, скопіювати (протягти) вниз на стільки рядків, на скільки задано (для цього прикладу – до 01.01.2000). При правильному копіюванні дата змінюється автоматично.

2 Побудуємо графік змінної *х*: **Діаграма – Тип: Точкова – вид Точкова.** З'явиться графік (рис.69).

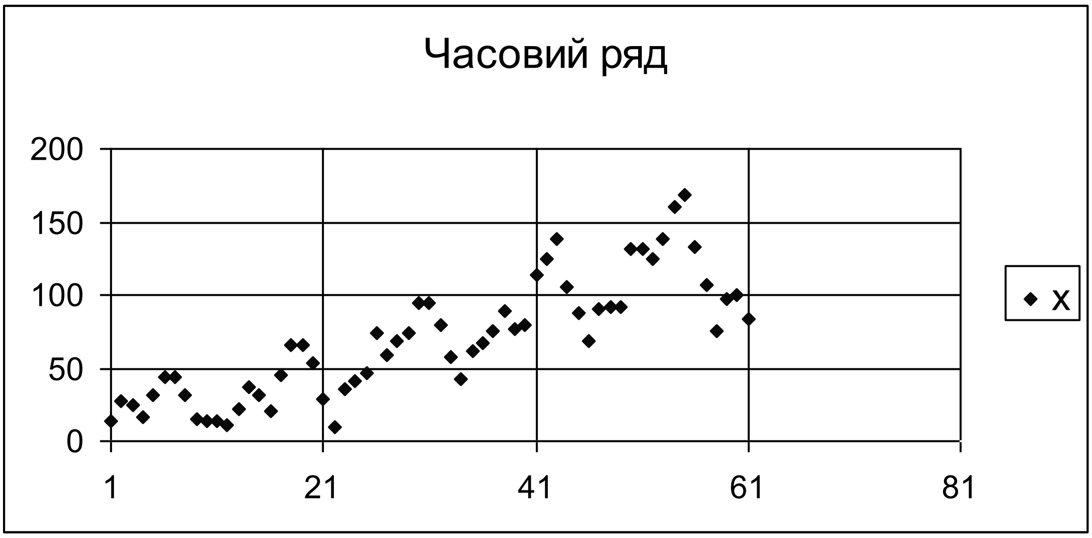


Рисунок 69

3 У меню **Сервіс** вибираємо **Аналіз даних – Ковзні середні**, **Ок.** Задаємо: вхідний інтервал – дані стовпця В, встановлюється прапорець у віконці «Мітки у першій строці», якщо у вхідному діапазоні разом з даними виділили букву «х», Інтервал – 12(лаг), Вихідний інтервал – виділяємо першу клітинку для виводу результату ($C$2), встановлюється прапорець у віконці «Друк графіка», встановлюється прапорець у віконці «Стандартні помилки», Ок.

Результат отримали у стовпцях С і D (табл.24) і у вигляді графіка (рис 70).

Таблиця 24

A	B	C	D
Дата	**X**	**Прогноз**	Залишки
01.01.1995	14	#Н/Д	#Н/Д
01.02.1995	28	#Н/Д	#Н/Д
01.03.1995	25	#Н/Д	#Н/Д
01.04.1995	17	#Н/Д	#Н/Д
01.05.1995	31	#Н/Д	#Н/Д
01.06.1995	44	#Н/Д	#Н/Д
01.07.1995	44	#Н/Д	#Н/Д
01.08.1995	32	#Н/Д	#Н/Д
01.09.1995	15	#Н/Д	#Н/Д
01.10.1995	14	#Н/Д	#Н/Д
01.11.1995	14	#Н/Д	#Н/Д
01.12.1995	11	24,08333	#Н/Д
01.01.1996	22	24,75	#Н/Д
01.02.1996	37	25,5	#Н/Д
01.03.1996	31	26	#Н/Д
01.04.1996	21	26,33333	#Н/Д
01.05.1996	45	27,5	#Н/Д
01.06.1996	66	29,33333	#Н/Д
01.07.1996	66	31,16667	#Н/Д
01.08.1996	54	33	#Н/Д
01.09.1996	29	34,16667	#Н/Д
01.10.1996	10	33,83333	#Н/Д
01.11.1996	36	35,66667	18,85099344
01.12.1996	41	38,16667	18,48687285
01.01.1997	46	40,16667	18,54642573
01.02.1997	74	43,25	20,29152407
01.03.1997	59	45,58333	20,60735777
01.04.1997	68	49,5	21,23237286
01.05.1997	74	51,91667	21,58546482
01.06.1997	95	54,33333	22,17453651
01.07.1997	95	56,75	22,63887185
01.08.1997	80	58,91667	22,64532543
01.09.1997	58	61,33333	22,61663443
01.10.1997	42	64	22,46129644
01.11.1997	62	66,16667	22,49327317
01.12.1997	67	68,33333	22,48169265
01.01.1998	76	70,83333	22,46809724
01.02.1998	89	72,08333	21,21005232
01.03.1998	77	73,58333	20,87674643
01.04.1998	79	74,5	20,22387718
01.05.1998	114	77,83333	21,8487541
01.06.1998	125	80,33333	22,49027568
01.07.1998	138	83,91667	25,05276838
01.08.1998	106	86,08333	24,97308737
01.09.1998	87	88,5	24,95829855
01.10.1998	68	90,66667	25,0079617
01.11.1998	90	93	24,99402706
01.12.1998	92	95,08333	25,00690877
01.01.1999	92	96,41667	24,99493004
01.02.1999	132	100	26,19603138
01.03.1999	131	104,5	27,272332

Продовження таблиці 24

Дата	Х	Прогноз	Залишки
01.04.1999	125	108,3333	27,66298503
01.05.1999	139	110,4167	26,91323723
01.06.1999	160	113,3333	27,19456858
01.07.1999	168	115,8333	26,88073797
01.08.1999	133	118,0833	26,60940032
01.09.1999	107	119,75	26,85925465
01.10.1999	76	120,4167	29,0346018
01.11.1999	97	121	29,83719326
01.12.1999	100	121,6667	30,47271282
01.01.2000	84	121	32,26521318

Рисунок 70

У стовпці С – прогноз, D – залишки. На графіку (рис.70) зображені фактичні значення (*x)* і прогнозовані. Як видно з цього графіка (рис.70), ряд добре згладжується ковзною середньою 12-го порядку. Це пов’язано з наявністю явища сезонності в даному часовому ряді, причому період сезонних коливань збігається з інтервалом згладжування.

Прогнозованих значень на $k-1=11$ менше, ніж початкових.

4 Здійснюємо контроль якості прогнозу. Для цього використовуються непрямі методи.

1) будуємо графік залишків **Діаграма – тип Точкова – вид Точкова** (рис.71).

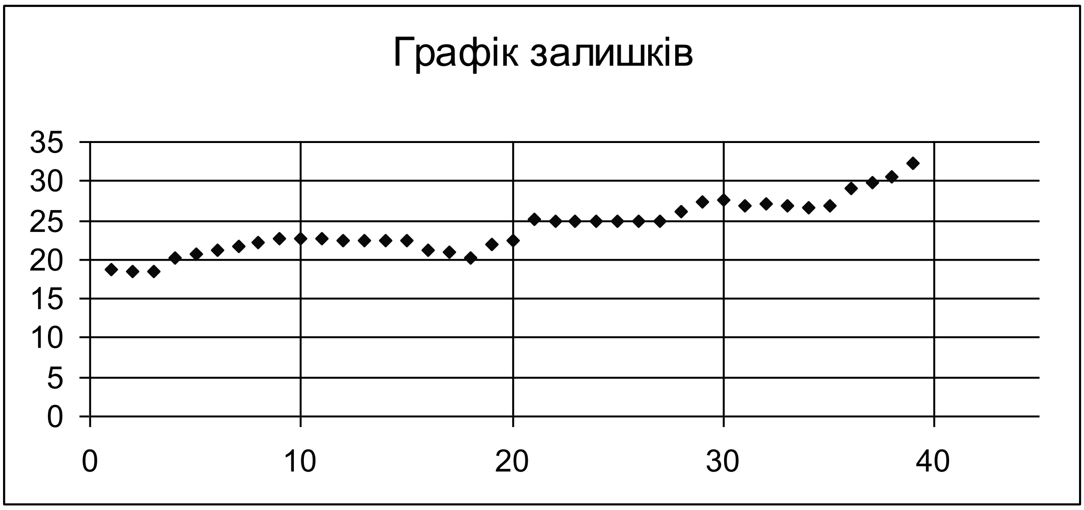


Рисунок 71

У графіку залишків не має бути систематичного зростання амплітуди. Можна бачити, що на рисунку 71 такого немає.

2) Залишки повинні мати нормальний розподіл. Будуємо графік: **Діаграма – Логарифмічна** (рис.72). Можна бачити, що залишки майже лягають на пряму лінію.

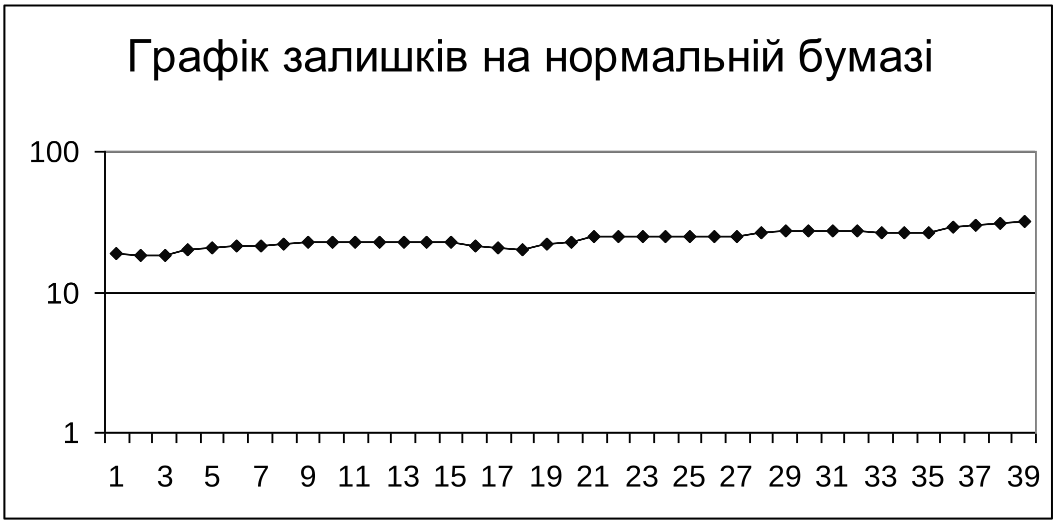


Рисунок 72

5 Ковзні середні можна використати також для прогнозування. При цьому k останніх даних згладжених ковзною середньою k–го порядку, і є прогнозом для наступної дати.

Виділяємо 12 останніх даних у стовпці C (прогнозовані значення) і знаходимо середнє значення (функція СРЗНАЧ). Отримаємо 114,5278. Це прогноз на лютий 2000р. Відомо, що для часових рядів довірчий інтервал практично симетричний щодо прогнозу. Його напівширина для рівня довіри 90% становить 20…30% від прогнозу. Візьмемо напівширину довірчого інтервалу 25% від прогнозу:

$$25\% \times 114{,}5278 = 28{,}6313;$$

114,5278-28,6313=85,8958;

114,5278+28,6313=143,1597.

Тому прогнозоване перевезення вугілля на лютий 2000р. буде знаходиться у межах від 85,8958 до 143,1597млн т.

6 Для виявленої загальної тенденції ряду можна записати рівняння, якщо додати на графік (рис.70), прогнозовані значення) лінію тренда. Для цього виділяємо на графіку ряд даних, в меню **Діаграма** вибираємо команду додати лінію тренда, на вкладці **Тип** відзначаємо **Лінійна**, на вкладці **Параметри** установлюємо прапорець у віконці «Додати рівняння на діаграму».

Отримали рівняння: y=2,2498x-12,254 (рис.73).

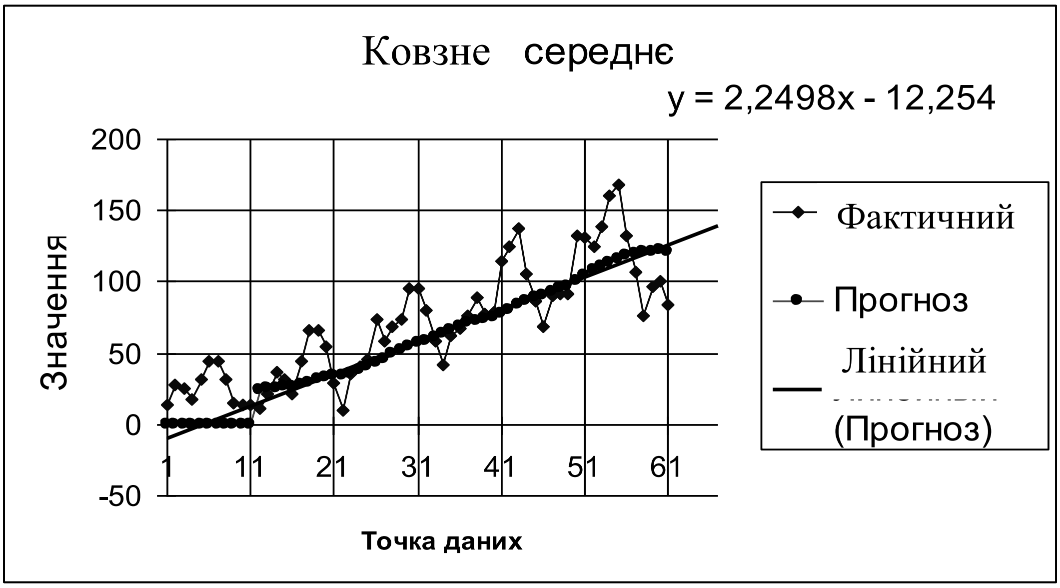


Рисунок 73

Це рівняння відтворює тільки загальну тенденцію часового ряду, виключив сезонну компоненту.

Висновки

Для згладжування даного часового ряду була використана ковзна середня 12-го порядку (лаг=12). Отримали, що загальна тенденція може бути описана лінійною моделлю: y=2,2498x-

12,254. Прогнозоване значення перевезення вугілля на лютий 2000р становить 114,5278млн т, 90%-ий довірчий інтервал для цього прогнозу: від 85,8958млн т до 143,1597млн т.

2.7 Метод експоненційного згладжування для часових рядів

Економічні дані

Зібрані відомості про щомісячні перевезення вугілля за період із січня 1995р. по січень 2000р. (млн т/місяць). Дані задані в таблиці 25. Згладити ряд, виявити загальну тенденцію, зробити прогноз на наступний місяць.

Виконання завдання

1 Вводимо дані на лист Excel у стовпець (стовпець А – «дата», стовпець В – *x*, млн т/місяць), так, як у попередньому прикладі.

2 Побудуємо графік змінної *x*: **Діаграма – Тип: Точкова – вид Точкова**. З’явиться графік (рис.74).

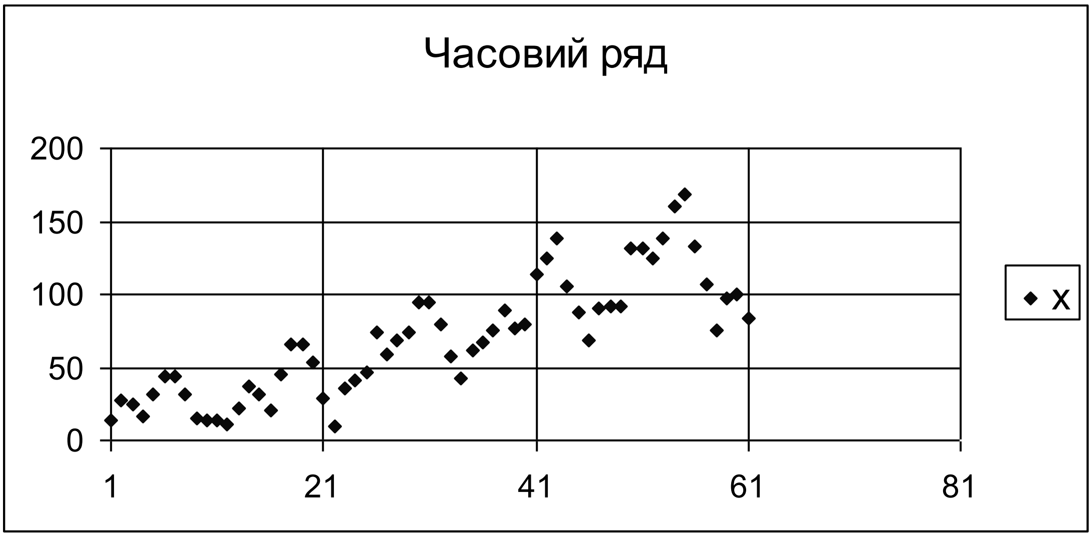


Рисунок 74

3 У меню Сервіс вибираємо **Аналіз даних – Експоненційне згладжування, Ок**. Задаємо: вхідний інтервал – дані стовпця В,

фактор затухання – 0,3, встановлюється прапорець у віконці «Мітки у першій строці», якщо у вхідному діапазоні разом з даними виділили букву «x», Вихідний інтервал – виділяємо першу клітинку для виводу результату ($C$2), встановлюється прапорець у віконці «Друк графіку», встановлюється прапорець у віконці «Стандартні помилки», Ок.

Результат отримали в стовпцях С (прогноз) і D (залишки) (табл. 25) і у вигляді графіка (рис.75).

Таблиця 25

A	B	C	D
Дата	**x**	**Прогноз**	**Залишки**
01.01.95	14	#Н/Д	#Н/Д
01.02.95	28	14	#Н/Д
01.03.95	25	23,8	#Н/Д
01.04.95	17	24,64	#Н/Д
01.05.95	31	19,292	9,234168
01.06.95	44	27,4876	8,10117
01.07.95	44	39,04628	12,49141
01.08.95	32	42,51388	12,03157
01.09.95	15	35,15417	11,65819
01.10.95	14	21,04625	13,43219
01.11.95	14	16,11387	13,74023
01.12.95	11	14,63416	12,38693
01.01.96	22	12,09025	4,737272
01.02.96	37	19,02707	6,215003
01.03.96	31	31,60812	12,03379
01.04.96	21	31,18244	11,85466
01.05.96	45	24,05473	11,93144
01.06.96	66	38,71642	13,4506
01.07.96	66	57,81493	20,71055
01.08.96	54	63,54448	20,41318
01.09.96	29	56,86334	17,34442
01.10.96	10	37,359	17,64897
01.11.96	36	18,2077	23,20904
01.12.96	41	30,66231	24,7753
01.01.97	46	37,89869	19,76486
01.02.97	74	43,56961	12,76799
01.03.97	59	64,87088	19,13555
01.04.97	68	60,76126	18,49421
01.05.97	74	65,82838	18,37458
01.06.97	95	71,54851	7,156394
01.07.97	95	87,96455	14,93482
01.08.97	80	92,88937	14,90241
01.09.97	58	83,86681	15,97504
01.10.97	42	65,76004	17,1729
01.11.97	62	49,12801	21,60067

Продовження таблиці 25

Дата	Х	Прогноз	Залишки
01.12.97	67	58,1384	21,59722
01.01.98	76	64,34152	16,41905
01.02.98	89	72,50246	11,25664
01.03.98	77	84,05074	12,736
01.04.98	79	79,11522	12,35316
01.05.98	114	79,03457	10,35849
01.06.98	125	103,5104	20,59375
01.07.98	138	118,5531	23,69528
01.08.98	106	132,1659	26,22065
01.09.98	87	113,8498	22,5436
01.10.98	68	95,05493	24,38407
01.11.98	90	76,11648	26,69292
01.12.98	92	85,83494	23,42102
01.01.99	92	90,15048	17,91396
01.02.99	132	91,44514	8,835171
01.03.99	131	119,8335	23,70742
01.04.99	125	127,6501	24,30917
01.05.99	139	125,795	24,33385
01.06.99	160	135,0385	10,10089
01.07.99	168	152,5116	16,3755
01.08.99	133	163,3535	18,59516
01.09.99	107	142,106	24,38785
01.10.99	76	117,5318	28,24688
01.11.99	97	88,45954	35,95673
01.12.99	100	94,43786	31,78189
01.01.00	84	98,33136	24,68986
01.02.00		*88,29941*	*10,15324*

Рисунок 75

Можна бачити, що ряд прогнозованих значень більш гладкий, ніж початковий. Це пов’язано з тим, що процедура згладжування дозволяє фільтрувати локальні флуктуації даних часових рядів для вияву чіткої тенденції та прогнозу майбутніх значень.

4 Здійснюємо контроль якості прогнозу. Для цього використовуються непрямі методи.

1) будуємо графік залишків **Діаграма – тип Точкова – вид Точкова** (рис.76).

Рисунок 76

На рисунку 76 можна бачити, що зростання амплітуди незначне.

2) Залишки повинні мати нормальний розподіл. Будуємо графік: **Діаграма – Логарифмічна** (рис.77). Можна бачити, що відхилення залишків від нормального розподілу незначне.

Рисунок 77

Метод експоненційного згладжування дозволяє ряд спостережень згладжувати будь-яку кількість разів. Тому, якщо вище наведені методи показують, що якість згладжування недостатня, то операцію експоненційного згладжування можна повторити двічі або тричі.

5 Для отримання прогнозу подовжимо стовпець прогнозованих значень ще на одну клітинку. Це прогноз на наступний місяць(лютий 2000р): 88,2994млн т.

Візьмемо на півширину довірчого інтервалу, яка дорівнює 25% від прогнозу: $88{,}2994 \times 25\% = 22{,}0749$. Тоді ліва межа довірчого інтервалу 88,2994-22,0749=66,2245, права межа 88,2994+22,0749=110,3743. Тому прогнозоване значення перевезення вугілля у лютому 2000р становить від 66,2245 до 110,3743млн т.

Висновки

Даний часовий ряд був згладжений за допомогою простого експоненційного згладжування з параметром згладжування $\alpha = 0{,}3$.

Прогнозоване значення перевезення вугілля на лютий 2000р становить 88,2994млн т, 90%-ий довірчий інтервал для цього прогнозу: від 66,2245млн т до 110,3743млн т.

ЧАСТИНА 3

ВСТУП

Дисципліна «Економетрика» належить до циклу дисциплін природничонаукової загальноекономічної підготовки і дає необхідну теоретичну базу для підготовки фахівців економічного напряму навчання.

Дисципліна базується на: вищій математиці (лінійна алгебра та математична статистика), статистиці, інформатиці та прикладах задач, що враховують специфіку майбутньої спеціальності студента.

На базі економетричних ідей ґрунтується викладання таких дисциплін, як «Економічний аналіз», «Державне регулювання економіки», «Фінанси». Крім того, студенти можуть використовувати здобуті знання при підготовці курсових робіт та написані дипломних робіт: як при оформлені робіт, так і при виконані розрахунків.

Дисципліна викладається у обсязі 81 навчальної години, розподіл яких між видами навчальних занять здійснюється відповідно до робочих навчальних планів.

Мета викладання дисципліни – навчити студента використовувати комп'ютер для розв'язку економетричних задач, використовувати прикладні системи (Пакет аналізу в Excel); дати студентам необхідну теоретичну базу для подальшого самостійного освоєння економетричної літератури.

Після вивчення дисципліни студенти мають ***знати:***

- основні етапи процесу економіко-математичного моделювання;

- основні типи економетричних моделей та методику їх одержання за допомогою комп'ютера;

- основні етапи комп'ютерного аналізу економетричної моделі;

- методику економетричного прогнозування;

- методику одержання одночасних структурних рівнянь та їх ідентифікацію.

Після вивчення дисципліни студенти повинні ***вміти***:

- вилучати фактори та показники, які достатньо повно характеризують економічну систему;
- створювати файл даних для подальшого моделювання;
- висувати та підтверджувати гіпотезу про наявність зв'язку, визначати силу та форму зв'язку;
- визначати параметри лінійної парної та множинної регресії, визначити інтервали довіри для параметрів моделі та для прогнозу;
- зводити до лінійних однофакторні та багатофакторні нелінійні моделі, визначати параметри нелінійних моделей;
- аналізувати економетричні моделі, оцінювати їх параметри та здійснювати прогнози за допомогою комп'ютера.

Після вивчення дисципліни студенти мають здобути ***навики*** роботи з Пакетом аналізу в Excel.

САМОСТІЙНА РОБОТА студентів над курсом містить:

- вивчення матеріалу за лекціями та навчальною літературою;
- підготовку до практичних занять;
- виконання і оформлення контрольної роботи;
- вивчення додаткових питань, які розширюють кругозір та знання.

Контрольні роботи та їх захист призначені для контролю знань студентів згідно з діючою в академії системою оцінки знань студентів.

Студент, який виконав і захистив контрольну роботу, допускається до складання заліку, на якому перевіряється рівень засвоєння знань та оволодіння уміннями, що передбачені програмою дисципліни.

Оцінка знань студентів з дисципліни «Економетрика» здійснюється згідно з діючою в ДДМА системою оцінки знань студентів.

Далі будуть наведені: перелік питань для перевірки засвоєння теоретичного матеріалу, завдання до контрольної роботи, завдання для самостійної роботи.

1 КОНТРОЛЬНІ ПИТАННЯ ТА ЗАВДАННЯ

1 Основні задачі економетрики. Етапи економетричного аналізу.

2 Класифікація економетричних моделей. Інформаційна база економетрики.

3 Генеральна сукупність. Вибірка. Обсяг вибірки. Середнє значення. Дисперсія. Середнє квадратичне відхилення.

4 Кореляційне поле. Центр розсіювання.

5 Коефіцієнт кореляції і його властивості.

6 Метод найменших квадратів для лінійної однофакторної регресії.

7 Властивості лінійної регресії.

8 Статистична гіпотеза. Нульова і конкуруюча гіпотези. Помилки 1 і 2 роду.

9 Критерій Фішера. Спостережне і критичне значення критерію.

10 Перевірка лінійної регресії на адекватність.

11 Коефіцієнт детермінації.

12 Перевірка моделі на адекватність за допомогою критерію Фішера.

13 Область прогнозу для однофакторної і двофакторної моделей. Довірчий інтервал. Коефіцієнт довіри.

14 Прогноз за лінійною однофакторною моделлю з урахуванням довірчого інтервалу.

15 Коефіцієнт еластичності для однофакторної моделі.

16 Види нелінійних однофакторних моделей. Спосіб їх лінеаризації

17 Алгоритм побудови нелінійних економетричних моделей.

18 Поняття багатофакторної моделі та етапи її побудови.

19 Колінеарность і мультиколінеарність.

20 Коефіцієнт еластичності для багатофакторних моделей.

21 Економічна інтерпретація параметрів лінійної однофакторної моделі $b_i\ (i = 1..p)$.

22 Система одночасних рівнянь. Ендогенні та екзогенні змінні.

23 Часовий ряд, лаг, тренд, дистрибутивні та мультиплікативні моделі.

24 Метод ковзних середніх.

25 Метод експоненційного згладжування.

26 Знайти коефіцієнт еластичності для вказаної моделі в заданій точці *x* (табл.26). Зробити економічний висновок.

Таблиця 26

Номер варіанта	**Модель**	**x**
1	**2**	**3**
1	$y = \frac{2}{x} + 5$	0,2
2	$y = \frac{1}{2x+1}$	1
3	$y = 3x^2 + 1$	1
4	$y = 6x^5$	1
5	$y = 2\sqrt{x} + 4$	4
6	$y = 3e^{2x}$	2
7	$y = \frac{2e^{5x}}{5}$	1

Продовження таблиці 26

1	2	3
8	$y = 3\ln x + 2$	1
9	$y = 2x^3 + 1$	1
10	$y = \frac{e^x}{2}$	2
11	$y = -\frac{x}{4}$	1
12	$y = \frac{1}{x} + 1$	1
13	$y = \frac{e^{x+1}}{6}$	3
14	$y = \ln\frac{x}{4} + 1$	4
15	$y = \sqrt{2x + 4}$	1
16	$y = 5x + 5$	1
17	$y = \frac{x^5}{2}$	1
18	$y = \frac{\sqrt{3x}}{3}$	2
19	$y = \frac{e^{3x}}{3}$	1
20	$y = \frac{4}{4x + 3}$	1
21	$y = 5x^3 + 1$	1
22	$y = \sqrt{x^3} + 1$	1
23	$y = \frac{3}{x^3}$	2
24	$y = -3x + 3$	2

Продовження таблиці 26

1	2	3
25	$y=\frac{\sqrt{x}}{2}$	1
26	$y=\frac{2x^2+2}{3}$	1
27	$y=\frac{2}{3x^3}$	2
28	$y=x^4+1$	1
29	$y=3\ln 2x+1$	0,5
30	$y=\frac{4}{x^2+1}$	1

2 ВИБІР ВАРІАНТА

Завдання вибирають згідно з таблицею 27.

Таблиця27

Передостання цифра залікової книжки	**Остання цифра залікової книжки**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0**
0	1	2	3	4	5	6	7	8	9	25
1	11	12	13	14	15	16	17	18	19	10
2	21	22	23	24	25	1	2	3	4	20
3	6	7	8	9	10	11	12	13	14	5
4	16	17	18	19	20	21	22	23	24	15
5	1	2	3	4	5	6	7	8	9	25
6	11	12	13	14	15	16	17	18	19	10
7	21	22	23	24	25	1	2	3	4	20
8	6	7	8	9	10	11	12	13	14	5
9	16	17	18	19	20	21	22	23	24	15

3 ВИМОГИ ДО ВИКОНАННЯ КОНТРОЛЬНОЇ РОБОТИ

1 Контрольна робота виконується у зошиті або на листах формату А4, жорстко скріплених між собою.

2 Кожне завдання повинне містити умову, економетричний аналіз поданих даних, роздрукування листів пакету Excel з розрахунками і формулами.

3 Економетричний аналіз (допускається друкарський і рукописний варіанти) включає докладний опис побудови моделі, перевірку її адекватності та знаходження прогнозів з використанням зроблених розрахунків на основі початкових даних. У аналізі використовуються необхідні визначення і формули.

3.1 Завдання 1

Зробити економетричний аналіз лінійної залежності показника Y від заданого фактора X1. Зробити прогноз для будь-якої точки з області прогнозу, побудувати довірчу область. Знайти коефіцієнт еластичності для всіх точок вибірки та в точці прогнозу.

Економічні дані наведені у таблицях 28-52.

3.2 Завдання 2

Зробити економетричний аналіз нелінійної залежності показника Y від заданого фактора X2. Зробити прогноз для будь-якої точки з області прогнозу, побудувати довірчу область. Знайти коефіцієнт еластичності для всіх точок вибірки та в точці прогнозу.

Економічні дані наведені у таблицях 28-52. Вибрати тип залежності можна самостійно або використати підказку для вашого варіанта.

3.3 Завдання 3

Зробити економетричний аналіз лінійної залежності показника Y від двох заданих факторів X1 і X2. Перевірити фактори на колінеарність. Зробити прогноз для будь-якої точки з області прогнозу. Знайти часткові коефіцієнти еластичності для всіх точок вибірки та в точці прогнозу.

Економічні дані наведені у таблицях 28-52.

Варіант 1

Продуктивність праці, фондовіддача і рівень рентабельності по плодоовочевих консервних заводах області за рік характеризуються наступними даними (табл. 28).

Таблиця 28

Номер заводу	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Продуктивність праці, грн	**Фондовіддача, грн**	**Рівень рентабельності, %**
1	8540	1,24	38,34
2	2911	0,63	44,69
3	6630	1,18	39,4
4	8492	1,12	38,93
5	2901	0,44	46,96
6	5410	1,19	39,48
7	1920	0,48	46,07
8	2569	0,65	43,5
9	3520	0,26	50,11
10	2340	0,75	42,79
11	6921	1,03	40,15
12	7671	0,89	40,44
13	1586	0,16	60,76
14	3223	0,67	42,99
15	7224	0,9	40,69

Нелінійну залежність прийняти $y = \frac{a}{x} + b$.

Варіант 2

Відомі наступні дані (табл. 29) про збитковість виробництва яловичини за КСП адміністративних районів області за рік.

Таблиця 29

Номер району	Фактор X_1	Фактор X_2	Показник Y
	Середньодобовий приріст, грн	**Собівартість 1 ц, грн**	**Рівень збитковості, %**
1	249	138,99	37,7
2	231	105,86	29,7
3	245	114,19	26,8
4	242	131,73	28,4
5	250	139,86	43,2
6	190	141,52	48
7	283	118,9	33,9
8	273	133,26	29,1
9	290	143,7	29,8
10	150	221,88	66
11	294	102,4	19,6
12	196	149,06	48,8
13	241	135,5	27,4
14	214	178,17	53,6
15	188	229,36	62,1

Нелінійну залежність прийняти $y = a\sqrt{x} + b$.

Варіант 3

Продуктивність праці, фондовіддача і рівень рентабельності по плодоконсервних заводах області за рік характеризуються наступними даними (табл. 30).

Таблиця 30

Номер заводу	Фактор X_1	Фактор X_2	Показник Y
	Фондовіддача, грн	**Продуктивність праці, грн**	**Рівень рентабельності, %**
1	***2***	***3***	***4***
1	1,12	7343	20,1
2	1,05	3991	20
3	0,99	5760	18

Продовження таблиці 30

1	**2**	**3**	**4**
4	0,7	3000	11,7
5	0,98	5241	17,9
6	1,04	4500	16,8
7	1,03	4300	15,6
8	1,35	7500	24,3
9	1,03	6743	18,1
10	0,89	5234	17,8
11	0,78	2500	13
12	0,87	3930	14,2
13	1,43	7433	24,2
14	1,03	6980	20
15	1,05	6740	19,3

Нелінійну залежність прийняти $y = a\ln x + b$.

Варіант 4

Продуктивність праці, фондовіддача і рівень рентабельності по хлібозаводах області за рік характеризуються наступними даними (табл. 31).

Таблиця 31

Номер заводу	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Фондовіддача, грн	**Продуктивність праці з розрахунку на 1 працівника, грн**	**Рівень рентабельності, %**
1	33,4	3447	14,3
2	29,1	3710	14,7
3	25,3	2827	11,9
4	27,1	2933	12,1
5	43,3	5428	22,3
6	47,2	5001	23,1
7	49,3	6432	24,3
8	35,7	4743	18,3
9	45,8	7321	27,6
10	52,4	6432	25,3
11	42,1	6003	25,1
12	40,1	5342	20,2
13	33,3	4341	15,7
14	41,2	5040	19,9
15	39,7	4493	17,2

Нелінійну залежність прийняти $y = a\ln x + b$.

Варіант 5

У таблиці 32 наведені дані про питому вагу робітників із спеціальною технічною підготовкою, питому вагу механізованих робіт і продуктивність праці по плодоовочевих заводах області за рік.

Таблиця 32

Номер заводу	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Питома вага робітників з технічною підготовкою, %	**Питома вага механізованих робіт, %**	**Продуктивність праці, грн**
1	64	84	4300
2	61	83	4150
3	49	68	3000
4	52	67	3300
5	53	69	3300
6	54	78	4300
7	57	77	4280
8	61	81	4100
9	56	77	3700
10	52	72	3500
11	60	74	4000
12	59	85	4450
13	63	83	4270
14	50	70	3300
15	65	81	4500

Нелінійну залежність прийняти $y = e^{ax} \cdot b$.

Варіант 6

У таблиці 33 наведені дані про відносний рівень витрат обігу, продуктивність праці і рівень рентабельності по магазинах промислових товарів за рік:

Таблиця 33

Номер магазина	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Відносний рівень витрат обігу, %	**Продуктивність праці, грн**	**Рівень рентабельності, %**
1	***2***	***3***	***4***
1	9,17	14789	6,9
2	6,5	21100	11,1
3	6,81	19343	10,2
4	7,89	17646	8,9

Продовження таблиці 33

1	*2*	*3*	*4*
5	7,01	18172	8,3
6	8,91	17477	7,8
7	6,17	22110	13,1
8	10,11	14331	4,9
9	5,98	24111	13,3
10	6,1	19393	10,7
11	5,9	25445	13,4
12	6,13	19378	10,8
13	9,01	13137	4,7
14	10,41	13177	3,9
15	8,13	17010	7,6

Нелінійну залежність прийняти $y = a\ln x + b$.

Варіант 7

У таблиці 34 наведені дані про рівень технічної підготовки робітників, стаж їх роботи і рівень заробітної плати по цукрових заводах області за рік.

Таблиця 34

Номер заводу	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Питома вага робітників з технічною підготовкою, %	**Питома вага робітників зі стажем понад 10 років, %**	**Заробітна плата за місяць, грн**
1	35	37	152,2
2	33	40	180,33
3	37	43	204,2
4	39	57	229,95
5	37	42	204,37
6	41	42	199,8
7	49	44	220,11
8	38	48	218,33
9	58	67	263,3
10	43	49	222,72
11	56	63	239,39
12	47	46	217,01
13	44	47	223,4
14	55	62	237,87
15	54	62	234,2

Нелінійну залежність прийняти $y = e^{ax} \cdot b$.

Варіант 8

Продуктивність праці, фондовіддача і рівень рентабельності по плодоконсервних заводах області за рік характеризуються наступними даними (табл. 35).

Таблиця 35

Номер заводу	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Фондовіддача, грн	**Продуктивність праці, грн**	**Рівень рентабельності, %**
1	5,46	3842,9	37,6
2	5,53	3457,7	37,9
3	7,05	3066,4	32,1
4	7,29	3011,9	32,1
5	7,4	3013,3	31,9
6	7,1	3164,3	33,4
7	6,25	3289,1	31,3
8	8,64	4320,3	39,3
9	5,18	2829,3	29,8
10	1,81	2562,2	20
11	2,3	2402,6	25,5
12	5,53	3636,7	37,6
13	2,22	2227,8	20,3
14	3,54	2725,8	29,1
15	3,23	2710,8	27,7

Нелінійну залежність прийняти $y = a\ln x + b$.

Варіант 9

У таблиці 36 наведені дані про питому вагу рілля, луків і пасовищ в сільськогосподарських угіддях і рівень збитковості продукції тваринництва по районах області за рік.

Таблиця 36

Номер району	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Питома вага рілля в сільськогосподарських угіддях, %	**Питома вага луків і пасовищ, %**	**Рівень збитковості продукції тваринництва, %**
1	***2***	***3***	***4***
1	80	20	20
2	87,2	12,8	37,5
3	90,8	9,2	43,4

Продовження таблиці 36

1	2	3	4
4	94,7	11,3	45,6
5	81,4	18,6	23,4
6	79,2	10,8	25
7	71,3	28,7	17,2
8	86,2	13,8	33,3
9	71,4	28,6	15
10	77,7	22,9	18,7
11	75,4	14	24,8
12	77,9	13	34,5
13	87,2	12,8	33,1
14	68,1	25	19,2
15	86,2	13,8	31,8

Нелінійну залежність прийняти $y = \frac{a}{x} + b$.

Варіант 10

У таблиці 37 наведені дані про питому вагу в товарообігу споживацької кооперації продукції власного виробництва, питому вагу переробленої продукції і рівень рентабельності підприємств області за рік.

Таблиця 37

Номер підприємства	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Питома вага продукції власного виробництва, %	**Питома вага переробленої продукції, %**	**Рівень рентабельності, %**
1	25,2	20,5	11,8
2	58,2	28,4	19,8
3	42,2	20,4	14,8
4	46,8	29,1	19,4
5	60,5	30,9	21,4
6	66,1	31,4	20,4
7	26,5	24,1	15,4
8	59,9	28,1	20,7
9	43,2	24,6	16,4
10	47,8	25,7	18,4
11	61,8	28,7	19,7
12	68,1	32,4	22,4
13	32	20,1	13,7
14	60,2	27,1	22,4
15	44,2	23,4	16,7

Нелінійну залежність прийняти $y = ax^2 + b$.

Варіант 11

Продуктивність праці, фондовіддача і рівень рентабельності по м'ясокомбінатах області за рік характеризуються наступними даними (табл. 38).

Таблиця 38

Номер заводу	Фактор X_1	Фактор X_2	Показник Y
	Фондовіддача, грн	**Продуктивність праці, грн**	**Рівень рентабель-ності, %**
1	1,25	5396	9,2
2	2,32	10583	13,7
3	1,71	8675	11,3
4	1,64	7392	10
5	1,38	3088	6,1
6	1,18	5138	9,1
7	1,44	5867	9,8
8	1,17	4154	6,4
9	1,72	13182	14,2
10	2,21	12351	13,8
11	1,64	13000	13,2
12	1,73	9519	11,4
13	1,17	4286	8,1
14	1,39	5000	9
15	2,07	7419	11,1

Нелінійну залежність прийняти $y = a\ln x + b$.

Варіант 12

Збитковість вирощування овочів в сільськогосподарських підприємствах і рівні факторів (збір овочів з 1 га і собівартість 1 ц), її формуючих, характеризуються наступними даними за рік (табл. 39).

Таблиця 39

Номер району	Фактор X_1	Фактор X_2	Показник Y
	Збір овочів з 1га, ц	**Собівартість 1 ц, грн**	**Рівень збитковості,%**
1	***2***	***3***	***4***
1	52,8	31,84	31,4
2	72,6	32,3	30,9
3	50,4	32,21	37,1
4	33,4	48,95	45,7

Продовження таблиці 39

1	2	3	4
5	31,5	42,48	57,7
6	54,6	35,38	46,7
7	54,3	29,11	33,3
8	36,6	67,06	63,8
9	15,6	65,52	68,8
10	73,2	21,26	29,8
11	65,9	31,29	39,4
12	44,6	33,63	46,2
13	23,7	73,35	68,8
14	64,6	40,12	34
15	25,6	43,63	47,6

Нелінійну залежність прийняти $y = ax^b$.

Варіант 13

Збитковість вирощування овочів в сільськогосподарських підприємствах і рівні факторів (збір овочів з 1 га, ц, і витрати праці, людино-годин на 1 ц), її формуючих, характеризуються наступними даними за рік (табл. 40).

Таблиця 40

Номер району	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Збір овочів з 1га, ц	**Витрати праці на 1 ц, люд-год**	**Рівень збитковості,%**
1	93,2	2,3	8,8
2	65,9	26,8	39,4
3	44,6	22,8	26,2
4	18,7	56,6	78,8
5	64,6	16,4	34
6	25,6	26,5	47,6
7	47,2	26	43,7
8	48,2	12,4	23,6
9	64,1	10	19,9
10	30,3	41,7	50
11	28,4	47,9	63,1
12	47,8	32,4	44,2
13	101,3	20,2	11,2
14	31,4	39,6	52,8
15	67,6	18,4	20,2

Нелінійну залежність прийняти $y = e^{ax} \cdot b$.

Варіант 14

Збитковість вирощування овочів в сільськогосподарських підприємствах і рівні факторів, її формуючих, характеризуються наступними даними за рік (табл. 41).

Таблиця 41

Номер району	Фактор X_1	Фактор X_2	Показник Y
	Збір овочів з 1га, ц	Собівартість 1 ц, грн	Рівень збитковості,%
1	52,8	31,84	31,4
2	72,6	32,3	20,9
3	50,4	32,21	37,1
4	33,4	48,95	45,7
5	31,5	42,48	57,7
6	54,6	35,38	46,7
7	54,3	29,11	33,3
8	36,6	67,06	63,8
9	15,6	65,52	68,8
10	73,2	21,26	12,8
11	65,9	31,29	39,4
12	44,6	33,63	26,2
13	23,7	73,35	68,8
14	64,6	40,12	34
15	25,6	43,63	47,6

Нелінійну залежність прийняти $y = a\sqrt{x} + b$.

Варіант 15

Рівень збитковості вирощування овочів в сільськогосподарських підприємствах і фактори, її формуючі, характеризуються наступними даними за рік (табл. 42).

Таблиця 42

Номер району	Фактор X_1	Фактор X_2	Показник Y
	Собівартість 1 ц, грн	Збір овочів з 1га, ц	Рівень збитковості,%
1	***2***	***3***	***4***
1	21,26	73,2	10,8
2	31,29	65,9	29,4
3	33,63	44,6	26,2

Продовження таблиці 42

1	2	3	4
4	73,35	23,7	68,8
5	40,12	64,6	31,1
6	43,63	25,6	47,6
7	32,2	47,2	43,7
8	49,85	38,2	43,6
9	39,02	64,1	25,9
10	41,7	30,3	50
11	49,53	28,4	43,1
12	38	47,8	34,2
13	17,14	101,3	8,2
14	44,17	41,4	52,8
15	31,4	67,6	20,2

Нелінійну залежність прийняти $y = \frac{a}{x} + b$.

Варіант 16

Збитковість вирощування овочів в сільськогосподарських підприємствах і рівні чинників, її формуючих, характеризуються наступними даними за рік (табл. 43).

Таблиця 43

Номер району	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Собівартість 1 ц, грн	**Витрати на 1 га посівів, грн**	**Рівень збитковості, %**
1	31,84	1549	31,4
2	32,3	1694	40,9
3	32,21	1807	47,1
4	48,95	1615	45,7
5	52,48	1926	57,7
6	35,38	1542	46,7
7	20,11	1309	13,3
8	67,06	2093	63,8
9	63,52	1836	58,8
10	21,26	1449	22,8
11	31,29	1601	39,4
12	23,63	1560	26,2
13	73,35	2213	68,8
14	40,12	2028	63
15	65,52	2136	68,8

Нелінійну залежність прийняти $y = a\ln x + b$.

Варіант 17

Рівень рентабельності і показники господарської діяльності торгових підприємств характеризуються наступними даними за рік (табл. 44).

Таблиця 44

Номер підприємства	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Товарообіг на душу населення, грн	**Відносний рівень витрат обігу, %**	**Рівень рентабельності, %**
1	27	17,2	3,62
2	29	20,74	2,02
3	24	17,73	2,77
4	21	21,2	2,01
5	33	16,56	4,33
6	28	17,01	4,01
7	23	19,77	2,12
8	28	17,1	3,73
9	30	16,35	3,92
10	25	18,34	2,87
11	22	22,2	2,11
12	34	16,06	4,39
13	31	16,1	4,11
14	22	18,7	2,13
15	29	17,4	3,2

Нелінійну залежність прийняти $y = ax^b$.

Варіант 18

Збитковість вирощування овочів у сільськогосподарських підприємствах і рівні факторів, її формуючих, характеризуються наступними даними за рік (табл. 45).

Таблиця 45

Номер району	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Собівартість 1 ц, грн	**Ціна реалізації 1 ц, грн**	**Рівень збитковості,%**
1	***2***	***3***	***4***
1	31,84	21,83	31,4
2	32,3	19,09	40,9
3	32,21	20,26	37,1
4	48,95	20,57	45,7

Продовження таблиці 45

1	2	3	4
5	42,48	17,96	57,7
6	35,38	15,32	46,7
7	29,11	29,19	13,3
8	67,06	11,26	63,8
9	65,52	10,47	68,8
10	21,26	29,67	12,8
11	31,29	18,95	39,4
12	33,63	24,81	26,2
13	73,35	12,92	68,8
14	40,12	26,49	34
15	43,63	22,83	47,6

Нелінійну залежність прийняти $y = e^{ax} \cdot b$.

Варіант 19

У таблиці 46 наведені дані про питому вагу робітників із спеціальною технічною підготовкою, питому вагу механізованих робіт і продуктивність праці по плодоовочевих заводах області за рік.

Таблиця 46

Номер заводу	Фактор X_1	Фактор X_2	Показник Y
	Питома вага робітників з технічною підготовкою, %	**Питома вага механізованих робіт, %**	**Продуктивність праці, грн**
1	52	67	3300
2	65	81	4500
3	49	68	3000
4	64	84	4300
5	53	69	3300
6	50	70	3300
7	57	77	4280
8	61	81	4100
9	56	77	3700
10	52	72	3500
11	60	74	4000
12	59	85	4450
13	63	83	4270
14	54	78	4300
15	61	83	4150

Нелінійну залежність прийняти $y = e^{ax} \cdot b$.

Варіант 20

У таблиці 47 наведені дані про відносний рівень витрат обігу, продуктивність праці і рівень рентабельності по магазинах промислових товарів за рік.

Таблиця 47

Номер магазина	Фактор X_1	Фактор X_2	Показник Y
	Відносний рівень витрат обігу, %	Продуктивність праці, грн	Рівень рентабельності, %
1	7,89	17646	8,9
2	10,41	13177	3,9
3	6,81	19343	10,2
4	9,17	14789	6,9
5	7,01	18172	8,3
6	8,91	17477	7,8
7	6,17	22110	13,1
8	10,11	14331	4,9
9	5,98	24111	13,3
10	6,1	19393	10,7
11	5,9	25445	13,4
12	8,13	17010	7,6
13	9,01	13137	4,7
14	6,5	21100	11,1
15	6,13	19378	10,8

Нелінійну залежність прийняти $y = a \ln x + b$.

Варіант 21

Відомі наступні дані про збитковість виробництва яловичини по КСП адміністративних районів області за рік (табл. 48).

Таблиця 48

Номер району	Фактор X_1	Фактор X_2	Показник Y
	Середньодобовий приріст, грн	Собівартість 1 ц, грн	Рівень збитковості, %
1	***2***	***3***	***4***
1	249	138,99	37,7
2	231	105,86	29,7
3	245	114,19	26,8
4	242	131,73	28,4

Продовження таблиці 48

1	2	3	4
5	250	139,86	43,2
6	190	141,52	48
7	283	118,9	33,9
8	273	133,26	29,1
9	290	143,7	29,8
10	150	221,88	66
11	294	102,4	19,6
12	196	149,06	48,8
13	241	135,5	27,4
14	214	178,17	53,6
15	188	229,36	62,1

Нелінійну залежність прийняти $y = a\sqrt{x} + b$.

Варіант 22

Продуктивність праці, фондовіддача і рівень рентабельності по плодоконсервних заводах області за рік характеризуються наступними даними (табл. 49).

Таблиця 49

Номер заводу	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Фондовіддача, грн	**Продуктивність праці, грн**	**Рівень рентабельності, %**
1	1,12	7343	20,1
2	1,05	3991	20
3	0,99	5760	18
4	0,7	3000	11,7
5	0,98	5241	17,9
6	1,04	4500	16,8
7	1,03	4300	15,6
8	1,35	7500	24,3
9	1,03	6743	18,1
10	0,89	5234	17,8
11	0,78	2500	13
12	0,87	3930	14,2
13	1,43	7433	24,2
14	1,03	6980	20
15	1,05	6740	19,3

Нелінійну залежність прийняти $y = a\ln x + b$.

Варіант 23

Продуктивність праці, фондовіддача і рівень рентабельності по плодоконсервних заводах області за рік характеризуються наступними даними (табл. 50).

Таблиця 50

Номер заводу	Фактор X_1	Фактор X_2	Показник Y
	Фондовіддача, грн	**Продуктивність праці, грн**	**Рівень рентабельності, %**
1	5,46	3842,9	37,6
2	5,53	3457,7	37,9
3	7,05	3066,4	32,1
4	7,29	3011,9	32,1
5	7,4	3013,3	31,9
6	7,1	3164,3	33,4
7	6,25	3289,1	31,3
8	8,64	4320,3	39,3
9	5,18	2829,3	29,8
10	1,81	2562,2	20
11	2,3	2402,6	25,5
12	5,53	3636,7	37,6
13	2,22	2227,8	20,3
14	3,54	2725,8	29,1
15	3,23	2710,8	27,7

Нелінійну залежність прийняти $y = a \ln x + b$.

Варіант 24

Продуктивність праці, фондовіддача і рівень рентабельності по м'ясокомбінатах області за рік характеризуються наступними даними (табл. 51).

Таблиця 51

Номер заводу	Фактор X_1	Фактор X_2	Показник Y
	Фондовіддача, грн	**Продуктивність праці, грн**	**Рівень рентабельності, %**
1	***2***	***3***	***4***
1	1,25	5396	9,2
2	2,32	10583	13,7
3	1,71	8675	11,3

Продовження таблиці 51

1	**2**	**3**	**4**
4	1,64	7392	10
5	1,38	3088	6,1
6	1,18	5138	9,1
7	1,44	5867	9,8
8	1,17	4154	6,4
9	1,72	13182	14,2
10	2,21	12351	13,8
11	1,64	13000	13,2
12	1,73	9519	11,4
13	1,17	4286	8,1
14	1,39	5000	9
15	2,07	7419	11,1

Нелінійну залежність прийняти $y = a\ln x + b$.

Варіант 25

Збитковість вирощування овочів в сільськогосподарських підприємствах і рівні факторів (збір овочів з 1 га, ц, і витрати праці, людино-годин на 1 ц), її формуючих, характеризуються наступними даними за рік (табл. 52).

Таблиця 52

Номер району	**Фактор X_1**	**Фактор X_2**	**Показник Y**
	Збір овочів з 1га, ц	**Витрати праці на 1 ц, люд-год**	**Рівень збитковості,%**
1	93,2	2,3	8,8
2	65,9	26,8	39,4
3	44,6	22,8	26,2
4	18,7	56,6	78,8
5	64,6	16,4	34
6	25,6	26,5	47,6
7	47,2	26	43,7
8	48,2	12,4	23,6
9	64,1	10	19,9
10	30,3	41,7	50
11	28,4	47,9	63,1
12	47,8	32,4	44,2
13	101,3	20,2	11,2
14	31,4	39,6	52,8
15	67,6	18,4	20,2

Нелінійну залежність прийняти $y = e^{ax} \cdot b$.

4 ЗАВДАННЯ ДЛЯ ІНДИВІДУАЛЬНОЇ ТА САМОСТІЙНОЇ РОБОТИ

Нижче наведені завдання, що виконуються студентом на практичних заняттях або самостійно. Варіанти та графік виконання завдань кожен студент отримує індивідуально під час настановної сесії.

4.1 Завдання 1

Зробити економетричний аналіз нелінійної залежності показника від двох заданих факторів (ступенева модель Коба-Дугласа). Перевірити фактори на колінеарність. Зробити прогноз для будь-якої точки з області прогнозу. Знайти часткові коефіцієнти еластичності в точці прогнозу.

Економічні дані взяти із завдання 3 контрольної роботи.

4.2 Завдання 2

За даними, наведеними у таблицях 53-62, побудуйте систему одночасних рівнянь у вигляді статичної моделі Кейнса:

$$\begin{cases} C = a + b \cdot y + \varepsilon, \\ \quad y = C + I, \end{cases}$$

де C – особисте споживання в постійних цінах,

y – національний дохід в постійних цінах,

I – інвестиції в постійних цінах,

ε – випадкова складова.

Визначить параметри рівнянь за допомогою непрямого методу найменших квадратів (НМНК). Проаналізуйте отриману модель.

Варіанти

Варіант 1

Таблиця 53

Рівень виробн. і доходу (у), млрд дол.	Споживання (С), млрд дол.	Інвестиції (І), млрд дол.
370	362	8
415	396	18
430	409	21
456	430	26
486	450	33
490	455	34
505	467	37
520	479	40
546	500	46
567	516	50

Варіант 2

Таблиця 54

Рівень виробн. і доходу (у), млрд дол.	Споживання (С), млрд дол.	Інвестиції (І), млрд дол.
370	365	5
415	380	35
430	410	20
456	428	28
486	435	51
490	455	35
505	465	40
520	490	30
546	500	46
567	523	44

Варіант 3

Таблиця 55

Рівень виробн. і доходу (y), млрд дол.	**Споживання (С), млрд дол.**	**Інвестиції (І), млрд дол.**
381	376	5
426	391	35
441	421	20
467	439	28
497	446	51
501	466	35
516	476	40
531	501	30
557	511	46
578	534	44

Варіант 4

Таблиця 56

Рівень виробн. і доходу (y), млрд дол.	**Споживання (С), млрд дол.**	**Інвестиції (І), млрд дол.**
381	375	6
426	390	36
441	420	21
467	438	29
497	445	52
501	465	36
516	475	41
531	500	31
557	510	47
578	533	45

Варіант 5

Таблиця 57

Рівень виробн. і доходу (у), млрд дол.	**Споживання (С), млрд дол.**	**Інвестиції (І), млрд дол.**
370	360	10
415	396	19
430	407	23
456	428	28
486	451	35
490	454	36
505	466	39
520	478	42
546	498	48
567	514	53

Варіант 6

Таблиця 58

Рівень виробн. і доходу (у), млрд дол.	**Споживання (С), млрд дол.**	**Інвестиції (І), млрд дол.**
370	360	10
415	395	20
430	405	25
456	428	28
486	450	36
490	454	36
505	465	40
520	478	42
546	497	49
567	514	53

Варіант 7

Таблиця 59

Рівень виробн. і доходу (y), млрд дол.	**Споживання (C), млрд дол.**	**Інвестиції (I), млрд дол.**
370	360	10
415	395	20
430	405	25
455	428	27
485	450	35
490	454	36
505	465	40
520	478	42
545	497	48
570	514	56

Варіант 8

Таблиця 60

Рівень виробн. і доходу(y), млрд дол.	**Споживання (C), млрд дол.**	**Інвестиції (I), млрд дол.**
384	372	11
429	407	22
444	417	26
469	440	29
499	462	38
504	466	39
519	477	42
534	490	45
559	509	50
584	526	59

Варіант 9

Таблиця 61

Рівень виробн. і доходу (y), млрд дол.	**Споживання (C), млрд дол.**	**Інвестиції (I), млрд дол.**
391	375	16
436	390	46
451	420	31
477	438	39
507	445	62
511	465	46
526	475	51
541	500	41
567	510	57
588	533	55

Варіант 10

Таблиця 62

Рівень виробн. і доходу (y), млрд дол.	**Споживання (C), млрд дол.**	**Інвестиції (I), млрд дол.**
306	311	0
351	346	5
366	358	8
392	378	14
422	401	20
426	405	21
441	418	25
456	429	28
482	449	33
503	465	38

4.3 Завдання 3

Проаналізувати часовий ряд. Згладити його за допомогою ковзної середньої. Спробувати виявити головну тенденцію. Зробити прогноз на 1 період уперед.

Економічні дані наведені у таблицях 63-87.

Варіант 1

Попит на бензин (січень 1990 - січень 1995 р.), тис. т/міс.

Таблиця 63

№ п/п	Попит	№ п/п	Попит	№ п/п	Попит	№ п/п	Попит
1	112	16	135	31	199	46	191
2	118	17	125	32	199	47	172
3	132	18	149	33	184	48	194
4	129	19	170	34	162	49	196
5	121	20	170	35	146	50	196
6	135	21	158	36	166	51	236
7	148	22	133	37	171	52	235
8	148	23	114	38	180	53	229
9	136	24	140	39	193	54	243
10	119	25	145	40	181	55	264
11	104	26	150	41	183	56	272
12	118	27	178	42	218	57	237
13	115	28	163	43	230	58	211
14	126	29	172	44	242	59	180
15	141	30	178	45	209	60	201
						61	204

Варіант 2

Витрата електроенергії (січень1995 - січень 2000р.), тис.кВт/год.

Таблиця 64

№ п/п	Витрата	№ п/п	Витрата	№ п/п	Витрата	№ п/п	Витрата
1	148	16	133	31	171	46	235
2	148	17	114	32	180	47	229
3	136	18	140	33	193	48	243
4	119	19	145	34	181	49	264
5	104	20	150	35	183	50	272
6	118	21	178	36	218	51	237
7	115	22	163	37	230	52	211
8	126	23	172	38	242	53	180
9	141	24	178	39	209	54	201
10	135	25	199	40	191	55	204
11	125	26	199	41	172	56	188
12	148	27	184	42	194	57	235
13	170	28	162	43	196	58	227
14	170	29	145	44	196	59	234
15	158	30	166	45	236	60	264
						61	302

Варіант 3

Попит на карамель (серпень 1995 – серпень 2000 р.), т/міс.

Таблиця 65

№ п/п	Попит	№ п/п	Попит	№ п/п	Попит	№ п/п	Попит
1	118	16	178	31	218	46	273
2	115	17	163	32	230	47	211
3	126	18	172	33	242	48	180
4	141	19	178	34	209	49	201
5	135	20	199	35	191	50	204
6	125	21	199	36	172	51	188
7	149	22	184	37	194	52	235
8	170	23	162	38	196	53	227
9	170	24	146	39	196	54	234
10	158	25	166	40	236	55	264
11	133	26	171	41	235	56	302
12	114	27	180	42	229	57	293
13	140	28	193	43	243	58	259
14	145	29	181	44	264	59	229
15	150	30	183	45	272	60	203
						61	229

Варіант 4

Обсяг залізничних перевезень (січень 1995 – січень 2000 р.), млн.т/міс.

Таблиця 66

№ п/п	Обсяг	№ п/п	Обсяг	№ п/п	Обсяг	№ п/п	Обсяг
1	278	16	356	31	435	46	463
2	284	17	348	32	491	47	407
3	277	18	355	33	505	48	362
4	317	19	422	34	404	49	405
5	313	20	465	35	359	50	417
6	318	21	467	36	310	51	391
7	374	22	404	37	337	52	419
8	413	23	347	38	360	53	461
9	405	24	305	39	342	54	472
10	355	25	336	40	406	55	535
11	306	26	340	41	396	56	622
12	271	27	318	42	420	57	606
13	306	28	362	43	472	58	508
14	315	29	348	44	548	59	461
15	301	30	363	45	559	60	390
						61	432

Варіант 5

Витрата газу в котельні (листопад 1995-листопад 2000 р.), тис.м^3/міс.

Таблиця 67

№ п/п	Витрата	№ п/п	Витрата	№ п/п	Витрата	№ п/п	Витрата
1	274	16	315	31	348	46	548
2	237	17	301	32	363	47	559
3	278	18	356	33	435	48	463
4	284	19	348	34	491	49	407
5	277	20	355	35	505	50	362
6	317	21	422	36	404	51	405
7	313	22	465	37	359	52	417
8	318	23	467	38	310	53	391
9	374	24	404	39	337	54	419
10	413	25	347	40	360	55	461
11	405	26	305	41	342	56	472
12	355	27	336	42	406	57	535
13	306	28	340	43	396	58	622
14	271	29	318	44	420	59	606
15	306	30	362	45	472	60	508
						61	461

Варіант 6

Обсяг автомобільних перевезень між двома містами (травень 1995 - травень 2000р.), т·км/міс.

Таблиця 68

№ п/п	Обсяг	№ п/п	Обсяг	№ п/п	Обсяг	№ п/п	Обсяг
1	270	16	405	31	305	46	342
2	315	17	355	32	336	47	406
3	364	18	306	33	340	48	396
4	347	19	271	34	318	49	420
5	312	20	306	35	362	50	472
6	274	21	315	36	348	51	548
7	237	22	301	37	363	52	559
8	278	23	356	38	435	53	463
9	284	24	348	39	491	54	407
10	277	25	355	40	505	55	362
11	317	26	422	41	404	56	405
12	313	27	465	42	359	57	417
13	318	28	467	43	310	58	391
14	374	29	404	44	337	59	419
15	413	30	347	45	360	60	461
						61	472

Варіант 7

Виробництво молока молочними фермами області (липень 1992 – липень 1997р.), т/міс.

Таблиця 69

№ п/п	**Виробництво**	**№ п/п**	**Виробництво**	**№ п/п**	**Виробництво**	**№ п/п**	**Виробництво**
1	302	16	274	31	315	46	348
2	293	17	237	32	301	47	363
3	259	18	278	33	356	48	435
4	229	19	284	34	348	49	491
5	203	20	277	35	355	50	505
6	229	21	317	36	422	51	404
7	242	22	313	37	465	52	359
8	233	23	318	38	467	53	310
9	267	24	374	39	404	54	337
10	269	25	413	40	347	55	360
11	270	26	405	41	305	56	342
12	315	27	355	42	336	57	406
13	364	28	306	43	340	58	396
14	347	29	271	44	318	59	420
15	312	30	306	45	362	60	472
						61	548

Варіант 8

Витрати на будівництво і модернізацію автодорожних об'єктів (січень 1995 - січень 2000р.), тис.грн/міс.

Таблиця 70

№ п/п	**Витрата**	**№ п/п**	**Витрата**	**№ п/п**	**Витрата**	**№ п/п**	**Витрата**
1	229	16	317	31	422	46	404
2	242	17	313	32	465	47	359
3	233	18	318	33	467	48	310
4	267	19	374	34	404	49	337
5	269	20	413	35	347	50	360
6	270	21	405	36	305	51	342
7	315	22	355	37	336	52	406
8	364	23	306	38	340	53	396
9	347	24	271	39	318	54	420
10	312	25	306	40	362	55	472
11	274	26	315	41	348	56	548
12	237	27	301	42	363	57	559
13	278	28	356	43	435	58	463
14	284	29	348	44	491	59	407
15	277	30	355	45	505	60	362
						61	405

Варіант 9

Замовлення на цеглу (серпень 1995 - серпень 2000 р.), млн шт./міс.

Таблиця 71

№ п/п	Замов-лення	№ п/п	Замов-лення	№ п/п	Замов-лення	№ п/п	Замов-лення
1	293	16	237	31	301	46	363
2	259	17	278	32	356	47	435
3	229	18	284	33	348	48	491
4	203	19	277	34	355	49	505
5	229	20	317	35	422	50	404
6	242	21	313	36	465	51	359
7	233	22	318	37	467	52	310
8	267	23	374	38	404	53	337
9	269	24	413	39	447	54	360
10	270	25	405	40	305	55	342
11	315	26	355	41	336	56	406
12	364	27	306	42	340	57	396
13	447	28	271	43	318	58	420
14	312	29	306	44	362	59	472
15	274	30	315	45	348	60	548
						61	559

Варіант 10

Попит на лісоматеріали (січень1995 - січень 2000р.), тис.$м^3$/міс.

Таблиця 72

№ п/п	Попит	№ п/п	Попит	№ п/п	Попит	№ п/п	Попит
1	204	16	269	31	412	46	347
2	188	17	270	32	405	47	305
3	235	18	315	33	355	48	336
4	227	19	364	34	306	49	340
5	234	20	347	35	271	50	318
6	264	21	312	36	306	51	362
7	302	22	274	37	315	52	348
8	293	23	237	38	301	53	363
9	259	24	278	39	256	54	435
10	229	25	284	40	348	55	491
11	203	26	277	41	355	56	505
12	229	27	317	42	422	57	404
13	242	28	313	43	465	58	359
14	233	29	318	44	467	59	310
15	267	30	374	45	404	60	337
						61	360

Варіант 11

Споживання газу в місті (серпень 1994 - серпень 1999 р.), тис. $м^3$/міс.

Таблиця 73

№ п/п	Спожи-вання	№ п/п	Спожи-вання	№ п/п	Спожи-вання	№ п/п	Спожи-вання
1	293	16	273	31	301	46	363
2	259	17	278	32	356	47	435
3	229	18	284	33	348	48	491
4	203	19	277	34	355	49	505
5	229	20	317	35	422	50	404
6	242	21	313	36	465	51	359
7	233	22	318	37	467	52	310
8	267	23	374	38	404	53	337
9	269	24	413	39	347	54	360
10	270	25	405	40	305	55	342
11	315	26	355	41	336	56	406
12	364	27	306	42	340	57	396
13	347	28	271	43	318	58	420
14	312	29	306	44	362	59	472
15	274	30	315	45	348	60	548
						61	559

Варіант 12

Попит на телевізійний кабель (травень 1995 - травень 2000р.), тис.м/міс.

Таблиця 74

№ п/п	Попит	№ п/п	Попит	№ п/п	Попит	№ п/п	Попит
1	229	16	293	31	237	46	301
2	243	17	259	32	278	47	356
3	264	18	229	33	284	48	348
4	272	19	203	34	277	49	355
5	237	20	229	35	317	50	422
6	211	21	242	36	313	51	465
7	180	22	233	37	318	52	467
8	201	23	267	38	374	53	404
9	204	24	269	39	413	54	347
10	188	25	270	40	405	55	305
11	235	26	315	41	355	56	336
12	227	27	364	42	306	57	340
13	234	28	347	43	271	58	318
14	264	29	312	44	306	59	362
15	302	30	274	45	315	60	348
						61	363

Варіант 13

Попит на мазут (січень 1995 - січень 2000 р.), тис. т/міс.

Таблиця 75

№ п/п	Попит	№ п/п	Попит	№ п/п	Попит	№ п/п	Попит
1	196	16	227	31	364	46	306
2	196	17	234	32	347	47	271
3	236	18	264	33	312	48	306
4	235	19	302	34	274	49	315
5	229	20	293	35	237	50	301
6	243	21	259	36	278	51	356
7	264	22	229	37	284	52	348
8	272	23	203	38	277	53	355
9	237	24	229	39	317	54	422
10	211	25	242	40	313	55	465
11	180	26	233	41	318	56	467
12	201	27	267	42	374	57	404
13	204	28	269	43	413	58	347
14	188	29	270	44	405	59	305
15	235	30	315	45	355	60	336
						61	340

Варіант 14

Попит на цукор (липень1994 - липень 1999р.), тис.т/міс.

Таблиця 76

№ п/п	Попит	№ п/п	Попит	№ п/п	Попит	№ п/п	Попит
1	264	16	229	31	284	46	348
2	272	17	203	32	277	47	355
3	237	18	229	33	317	48	422
4	211	19	242	34	313	49	465
5	180	20	233	35	318	50	467
6	201	21	267	36	374	51	404
7	204	22	269	37	413	52	347
8	188	23	270	38	405	53	305
9	235	24	315	39	355	54	336
10	227	25	364	40	306	55	340
11	234	26	347	41	271	56	318
12	264	27	312	42	306	57	362
13	302	28	274	43	315	58	348
14	293	29	237	44	301	59	363
15	259	30	278	45	356	60	435
						61	491

Варіант 15

Попит на солярку (січень 1995 - січень 2000 р.), тис. т/міс.

Таблиця 77

№ п/п	Попит	№ п/п	Попит	№ п/п	Попит	№ п/п	Попит
1	113	16	135	31	199	46	191
2	119	17	125	32	199	47	172
3	133	18	149	33	184	48	194
4	130	19	170	34	162	49	196
5	121	20	170	35	146	50	196
6	135	21	158	36	166	51	236
7	148	22	133	37	171	52	235
8	148	23	114	38	180	53	229
9	136	24	140	39	193	54	243
10	119	25	145	40	181	55	264
11	104	26	150	41	183	56	272
12	118	27	178	42	218	57	237
13	115	28	163	43	230	58	211
14	126	29	172	44	242	59	180
15	141	30	178	45	209	60	201
						61	204

Варіант 16

Витрата електроенергії (січень 1996 - січень 2001р.), тис.кВт/год.

Таблиця 78

№ п/п	Витрата	№ п/п	Витрата	№ п/п	Витрата	№ п/п	Витрата
1	149	16	133	31	171	46	235
2	148	17	114	32	180	47	229
3	137	18	140	33	193	48	243
4	120	19	145	34	181	49	264
5	104	20	150	35	183	50	272
6	118	21	178	36	218	51	237
7	115	22	163	37	230	52	211
8	126	23	172	38	242	53	180
9	141	24	178	39	209	54	201
10	135	25	199	40	191	55	204
11	125	26	199	41	172	56	188
12	148	27	184	42	194	57	235
13	170	28	162	43	196	58	227
14	170	29	145	44	196	59	234
15	158	30	166	45	236	60	264
						61	302

Варіант 17

Попит на цукор (серпень 1995 – серпень 2000 р.), т/міс.

Таблиця 79

№ п/п	Попит	№ п/п	Попит	№ п/п	Попит	№ п/п	Попит
1	118	16	178	31	218	46	273
2	115	17	163	32	230	47	211
3	126	18	172	33	242	48	180
4	141	19	178	34	209	49	201
5	135	20	199	35	191	50	204
6	125	21	199	36	172	51	188
7	149	22	184	37	194	52	235
8	170	23	162	38	196	53	227
9	170	24	146	39	196	54	234
10	158	25	166	40	236	55	264
11	133	26	171	41	235	56	302
12	114	27	180	42	229	57	293
13	140	28	193	43	243	58	259
14	145	29	181	44	264	59	229
15	150	30	183	45	272	60	203
						61	229

Варіант 18

Обсяг залізничних перевезень (січень 1996 – січень 2001 р.), млн.т/міс.

Таблиця 80

№ п/п	Обсяг	№ п/п	Обсяг	№ п/п	Обсяг	№ п/п	Обсяг
1	279	16	356	31	435	46	463
2	285	17	348	32	491	47	407
3	278	18	355	33	505	48	362
4	316	19	422	34	404	49	405
5	313	20	465	35	359	50	417
6	318	21	467	36	310	51	391
7	374	22	404	37	337	52	419
8	413	23	347	38	360	53	461
9	405	24	305	39	342	54	472
10	355	25	336	40	406	55	535
11	306	26	340	41	396	56	622
12	271	27	318	42	420	57	606
13	306	28	362	43	472	58	508
14	315	29	348	44	548	59	461
15	301	30	363	45	559	60	390
						61	432

Варіант 19

Витрата газу в котельні (листопад 1996-листопад 2001 р.), тис.м^3/міс.

Таблиця 81

№ п/п	**Витрата**	**№ п/п**	**Витрата**	**№ п/п**	**Витрата**	**№ п/п**	**Витрата**
1	273	16	315	31	348	46	548
2	236	17	301	32	363	47	559
3	279	18	356	33	435	48	463
4	285	19	348	34	491	49	407
5	277	20	355	35	505	50	362
6	317	21	422	36	404	51	405
7	313	22	465	37	359	52	417
8	318	23	467	38	310	53	391
9	374	24	404	39	337	54	419
10	413	25	347	40	360	55	461
11	405	26	305	41	342	56	472
12	355	27	336	42	406	57	535
13	306	28	340	43	396	58	622
14	271	29	318	44	420	59	606
15	306	30	362	45	472	60	508
						61	461

Варіант 20

Обсяг автомобільних перевезень між двома містами (травень 1996 - травень 2001р.), т•км/міс.

Таблиця 82

№ п/п	Обсяг	№ п/п	Обсяг	№ п/п	Обсяг	№ п/п	Обсяг
1	271	16	405	31	305	46	342
2	316	17	355	32	336	47	406
3	365	18	306	33	340	48	396
4	348	19	271	34	318	49	420
5	312	20	306	35	362	50	472
6	274	21	315	36	348	51	548
7	237	22	301	37	363	52	559
8	278	23	356	38	435	53	463
9	284	24	348	39	491	54	407
10	277	25	355	40	505	55	362
11	317	26	422	41	404	56	405
12	313	27	465	42	359	57	417
13	318	28	467	43	310	58	391
14	374	29	404	44	337	59	419
15	413	30	347	45	360	60	461
						61	472

Варіант 21

Виробництво молока молочними фермами області (липень 1995 – липень 2000 р.), т/міс.

Таблиця 83

№ п/п	Виробництво	№ п/п	Виробництво	№ п/п	Виробництво	№ п/п	Виробництво
1	300	16	274	31	315	46	348
2	295	17	237	32	301	47	363
3	258	18	278	33	356	48	435
4	229	19	284	34	348	49	491
5	203	20	277	35	355	50	505
6	229	21	317	36	422	51	404
7	242	22	313	37	465	52	359
8	233	23	318	38	467	53	310
9	267	24	374	39	404	54	337
10	269	25	413	40	347	55	360
11	270	26	405	41	305	56	342
12	315	27	355	42	336	57	406
13	364	28	306	43	340	58	396
14	347	29	271	44	318	59	420
15	312	30	306	45	362	60	472
						61	548

Варіант 22

Витрати на будівництво і модернізацію автодорожних об'єктів (березень1996 - березень 2001р.), тис.грн/міс.

Таблиця 84

№ п/п	Витрата	№ п/п	Витрата	№ п/п	Витрата	№ п/п	Витрата
1	228	16	317	31	422	46	404
2	241	17	313	32	465	47	359
3	235	18	318	33	467	48	310
4	267	19	374	34	404	49	337
5	269	20	413	35	347	50	360
6	270	21	405	36	305	51	342
7	315	22	355	37	336	52	406
8	364	23	306	38	340	53	396
9	347	24	271	39	318	54	420
10	312	25	306	40	362	55	472
11	274	26	315	41	348	56	548
12	237	27	301	42	363	57	559
13	278	28	356	43	435	58	463
14	284	29	348	44	491	59	407
15	277	30	355	45	505	60	362
						61	405

Варіант 23

Замовлення на цеглу (липень 1995 – липень 2000 р.), млн. шт./міс.

Таблиця 85

№ п/п	Замов-лення	№ п/п	Замов-лення	№ п/п	Замов-лення	№ п/п	Замов-лення
1	292	16	237	31	301	46	363
2	260	17	278	32	356	47	435
3	229	18	284	33	348	48	491
4	203	19	277	34	355	49	505
5	229	20	317	35	422	50	404
6	242	21	313	36	465	51	359
7	233	22	318	37	467	52	310
8	267	23	374	38	404	53	337
9	269	24	413	39	447	54	360
10	270	25	405	40	305	55	342
11	315	26	355	41	336	56	406
12	364	27	306	42	340	57	396
13	447	28	271	43	318	58	420
14	312	29	306	44	362	59	472
15	274	30	315	45	348	60	548
						61	559

Варіант 24

Попит на лісоматеріали (лютий1995 - лютий 2000р.), тис.м^3./міс.

Таблиця 86

№ п/п	Попит	№ п/п	Попит	№ п/п	Попит	№ п/п	Попит
1	205	16	269	31	412	46	347
2	189	17	270	32	405	47	305
3	235	18	315	33	355	48	336
4	227	19	364	34	306	49	340
5	234	20	347	35	271	50	318
6	264	21	312	36	306	51	362
7	302	22	274	37	315	52	348
8	293	23	237	38	301	53	363
9	259	24	278	39	256	54	435
10	229	25	284	40	348	55	491
11	203	26	277	41	355	56	505
12	229	27	317	42	422	57	404
13	242	28	313	43	465	58	359
14	233	29	318	44	467	59	310
15	267	30	374	45	404	60	337
						61	360

Варіант 25

Споживання газу в місті (вересень 1995 - вересень 2000р.), тис. $м^3$/міс.

Таблиця 87

№ п/п	**Спожи-вання**	**№ п/п**	**Спожи-вання**	**№ п/п**	**Спожи-вання**	**№ п/п**	**Спожи-вання**
1	292	16	273	31	301	46	363
2	260	17	278	32	356	47	435
3	229	18	284	33	348	48	491
4	203	19	277	34	355	49	505
5	229	20	317	35	422	50	404
6	242	21	313	36	465	51	359
7	233	22	318	37	467	52	310
8	267	23	374	38	404	53	337
9	269	24	413	39	347	54	360
10	270	25	405	40	305	55	342
11	315	26	355	41	336	56	406
12	364	27	306	42	340	57	396
13	347	28	271	43	318	58	420
14	312	29	306	44	362	59	472
15	274	30	315	45	348	60	548
						61	559

4.4 Завдання 4

Проаналізувати часовий ряд. Згладити його за допомогою методу експоненційного згладжування. Зробити прогноз на 1 період уперед.

Економічні дані взяти із попереднього завдання.

5 ВАРІАНТИ ЗАВДАНЬ ДЛЯ ТЕСТУВАННЯ

1 Які з наведених нижче статистичних даних відносяться до варіаційних рядів:

а) дані про динаміку рівня інфляції;

б) дані про продуктивність праці по хлібозаводах області?

2 Які з наведених нижче статистичних даних відносяться до часових рядів:

а) дані про прибуток громадян за останні 10 років;

б) дані про середньодобовий приріст виробництва м’яса по совхозах області?.

3 Вказати, які з коефіцієнтів кореляції є можливими:

а) 0,01;

б) -0,33;

в) -1,59.

4 Хай при вивченні залежності $Y = f(X_1, X_2, X_3)$ матриця парних коефіцієнтів кореляції виявилася наступною (табл 88).

Таблиця 88

	У	X1	X2	X3
У	1			
X1	0,8	1		
X2	0,7	0,91	1	
X3	0,6	0,5	0,2	1

Які фактори доцільно включити до подальшого аналізу:

а) X1 і X2;

б) X1 і X3;

в) X2 і X3?

5 Нижче наведені коефіцієнти кореляції. Вказати, чи є залежність прямою:

а) 0,01;

б) -0,33;

в) -1,59.

6 Нижче наведені коефіцієнти кореляції. Вказати, чи є залежність зворотною:

а) 0,01;

б) -0,33;

в) -1,59.

7 Як зміниться ширина довірчого інтервалу при збільшенні обсягу вибірки:

а) збільшиться;

б) зменшиться?

8 Як зміниться ширина довірчого інтервалу при збільшенні рівня довіри:

а) збільшиться;

б) зменшиться?

9 Результати спостережень занесені до таблиці 89.

Таблиця 89

x	y
10	4
12	5
14	7
16	8

Визначити координати центру розсіювання:

а) (12; 6);

б) (11;7);

в) (13;6).

СПИСОК ЛІТЕРАТУРИ

1 Боровиков В.П. STATISTICA/ В.П.Боровиков, И.П.Боровиков. – М.: Информационно-издательский дом “Филинъ”, 1997. – 592с.

2 Доугерти К. Введение в эконометрику. – М.: Инфра-М, 2001. – 402с.

3 Лук'яненко І. Економетріка: Практикум / І.Лук'яненко, Л.Краснікова. – Київ: Знання, 1998. – 217с.

4 Лук'яненко І. Економетрика/ І.Лук'яненко, Л.Краснікова. – Київ: Знання, 1998. – 493с.

5 Магнус Я.Р. Эконометрика. Начальный курс: Учебник/ Я.Р.Магнус, П.К.Катышев, А.А.Пересецкий. – 4-е изд. – М.: Дело, 2000. – 400с.

6 Эконометрика: Учебник/ Под ред. И.И.Елисеевой. – М.: Финансы и статистика, 2002. – 344с.

7 Фишер Ф. Проблемы идентификации в эконометрии. – М.: Статистика, 1978. – 245 с.

8 Контроль качества с помощью персональных компьютеров/ Т.Макино, М.Охаси, Х.Док, К.Макино. – М.: Машиностроение, 1991. – 224с.

9 Назаренко О.М. Основи економетрики: Підручник. – Київ: Центр навчальної літератури, 2004. – 392 с.

10 Толбатов Ю.А. Економетрика: Підручник. – К.:ТП Пресс, 2003. – 320с.

Навчальне видання

ВАСИЛЬЄВА Людмила Володимирівна
КЛЬОВАНИК Олена Анатоліївна

РЕГРЕСІЙНІ МОДЕЛІ ТА АНАЛІЗ ЧАСОВИХ РЯДІВ

Навчальний посібник
для студентів вищих навчальних закладів

Редактор І. І. Дьякова

Комп'ютерна верстка О. С. Орда

Вз.28/2006. Підп. до друку 14.10.10. Формат 60х84/16.
Папір офсетний. Ум. друк. арк. 11,0. Обл.-вид. арк. 6,85.
Тираж 65 прим. Зам. № 103.

Донбаська державна машинобудівна академія
84313, м. Краматорськ, вул. Шкадінова, 72
Свідоцтво про внесення суб'єкта видавничої справи
до державного реєстру
серія ДК № 1633 від 24.12.2003