

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ
УНІВЕРСИТЕТ**

**Розрахунково-графічні завдання
для самостійної роботи
з дисципліни
“ Економетрія ”**

Частина 1 (парна та множинна регресії)

**для студентів економічних спеціальностей
(усіх форм навчання)**

2013

Розрахунково-графічні завдання для самостійної роботи з дисципліни “ Економетрія ” Частина1 (парна та множинна регресії) для студентів економічних спеціальностей усіх форм навчання / Укл. Щолокова М.О. Коротунова О. В. – Запоріжжя: ЗНТУ, 2013. – 50 с.

Укладачі: Щолокова М.О., доцент, к.т.н
Коротунова О. В., доцент, к.т.н

Експерт: Корольков В.В., доцент, к.е.н

Рецензент: Мастиновський Ю.В, доцент, к.т.н.

Відповідальний за випуск: Щолокова М.О., доцент, к.т.н

Затверджено на засіданні
кафедри прикладної математики

Протокол № 6 від 06.03.13

ЗМІСТ

ВСТУП. ВИДИ ЗАЛЕЖНОСТЕЙ МІЖ ЕКОНОМІЧНИМИ ЯВИЩАМИ ТА ПРОЦЕСАМИ	4
1 ВИДИ РІВНЯНЬ ПАРНОЇ РЕГРЕСІЇ Й ВИЗНАЧЕННЯ ПАРАМЕТРІВ	5
1.1 Лінійна модель парної регресії	5
1.1.1 Отримання рівняння регресії	5
1.1.2 Оцінка тісноти й значимості зв'язку між змінними в рівняннях парної регресії.	7
1.1.3 Використання рівняння регресії для складання прогнозу	13
1.2 Нелінійні моделі парної регресії й кореляції	14
2 МНОЖИННА РЕГРЕСІЯ І КОРЕЛЯЦІЯ	22
2.1 Вимоги до чинників множинної регресії	22
2.2 Дослідження на мультиколінеарність	23
2.3 Лінійна модель множинної регресії	25
3 ВАРІАНТИ ЗАВДАНЬ	32
ЛІТЕРАТУРА	44
ДОДАТОК А	48
ДОДАТОК Б	50

ВСТУП

ВИДИ ЗАЛЕЖНОСТЕЙ МІЖ ЕКОНОМІЧНИМИ ЯВИЩАМИ ТА ПРОЦЕСАМИ

Ознака, яка характеризує причину чи умову, є факторною (**x**), а ознака, яка характеризує наслідок, - результативною (**y**).

Розрізняють такі види залежностей між економічними явищами та процесами:

- **функціональна** – характеризується повною відповідністю між причиною і наслідком;
- **стохастична** – кожному значенню ознаки **x** відповідає певна множина значень ознаки **y**;
- **кореляційна** – зі зміною факторної ознаки **x** змінюються групові середні результативної ознаки **y**.

Одностороння стохастична залежність виражається за допомогою функції, яка називається **регресією**. Відносно кількості змінних розрізняють просту (парну, однофакторну) та множинну (багатофакторну) регресії. Відносно форми залежності – лінійна та нелінійна регресії. Види рівнянь регресії зручно представити у вигляді таблиці:

Парна регресія

Множинна регресія

лінійна залежність

$$y = a_0 + a_1 x \qquad \hat{o} = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_m x_m \qquad (1)$$

параболічна залежність

$$y = a_0 + a_1 x^2 \qquad y = a_0 + a_1 x_1^2 + a_2 x_2^2 + \dots + a_m x_m^2 \qquad (2)$$

гіперболічна залежність

$$y = a_0 + a_1 \frac{1}{x} \qquad y = a_0 + a_1 \frac{1}{x_1} + a_2 \frac{1}{x_2} + \dots + a_m \frac{1}{x_m} \qquad (3)$$

степенева залежність

$$y = a_0 x^{a_1} \qquad y = a_0 x_1^{a_1} x_2^{a_2} \dots x_m^{a_m} \qquad (4)$$

Рівняння (2)-(4) є нелінійними, але за допомогою перетворень їх можна звести до лінійної форми.

1 ВИДИ РІВНЯНЬ ПАРНОЇ РЕГРЕСІЇ Й ВИЗНАЧЕННЯ ПАРАМЕТРІВ

1.1 Лінійна модель парної регресії

1.1.1 Отримання рівняння регресії

Лінійна регресія зводиться до знаходження рівняння виду

$$\hat{y} = a + b \cdot x$$

Побудова лінійної регресії зводиться до оцінки її параметрів – a і b . Класичний підхід до оцінювання параметрів лінійної регресії заснований на методі найменших квадратів (МНК).

Система лінійних рівнянь для оцінки параметрів a і b :

$$\begin{cases} na + b \sum x = \sum y, \\ a \sum x + b \sum x^2 = \sum x \cdot y, \end{cases} \quad (1.1)$$

Параметр b називається **коефіцієнтом регресії**. Його величина показує середню зміну результату y зі зміною фактору x на одну одиницю. Можливість чіткої економічної інтерпретації коефіцієнта регресії зробила лінійне рівняння регресії досить розповсюдженим в економетричних дослідженнях.

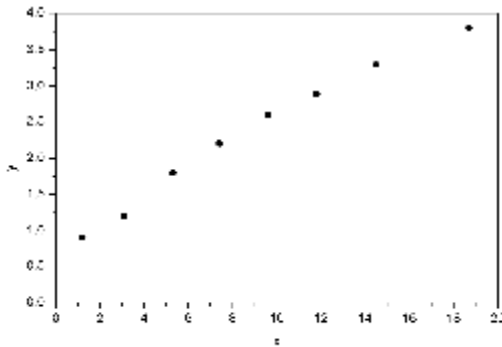
Знаючи коефіцієнт регресії можна обчислити **коефіцієнт еластичності** (відносний ефект впливу фактору на результат) - на скільки відсотків у середньому зміниться результат y зі зміною фактору x на 1%)

$$K_e = \frac{b \cdot \bar{x}}{\bar{y}}. \quad (1.2)$$

Приклад 1. За даними проведеного опитування восьми груп родин відомі дані зв'язку витрат населення на продукти харчування з рівнем прибутків родини.

Витрати на продукти харчування, у, тис. грн.	0,9	1,2	1,8	2,2	2,6	2,9	3,3	3,8
Прибутки родини, х, тис. грн.	1,2	3,1	5,3	7,4	9,6	11,8	14,5	18,7

Припустимо, що зв'язок між прибутками родини й витратами на продукти харчування лінійний. Для підтвердження нашого припущення побудуємо поле кореляції.



По малюнку видно, що точки вибудовуються в деяку пряму лінію.

Для зручності подальших обчислень складемо таблицю (заповнимо поки стовпці 1-4).

	x	y	$x \cdot y$	x^2
	1	2	3	4
1	1,2	0,9	1,08	1,44
2	3,1	1,2	3,72	9,61
3	5,3	1,8	9,54	28,09

4	7,4	2,2	16,28	54,76
5	9,6	2,6	24,96	92,16
6	11,8	2,9	34,22	139,24
7	14,5	3,3	47,85	210,25
8	18,7	3,8	71,06	349,69
Разом	71,6	18,7	208,71	885,24
Середнє значення	8,95	2,34	26,09	110,66

Розрахуємо параметри лінійного рівняння парної регресії $\hat{y} = a + b \cdot x$. Для розв'язання системи (1.1) зручно подати цю систему за допомогою матричного апарату: $A \cdot X = B$,

$$\text{де } A = \begin{pmatrix} n & \sum x \\ \sum x & \sum x^2 \end{pmatrix}, \quad B = \begin{pmatrix} \sum y \\ \sum xy \end{pmatrix}, \quad X = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Тоді коефіцієнти парної регресії визначаються за формулою:

$$X = A^{-1} \cdot B.$$

Одержали рівняння: $\hat{y} = 0,836 + 0,168 \cdot x$. ($a = 0,836$, $b = 0,168$). Такім чином, зі збільшенням прибутку родини на 1000 грн. витрати на харчування збільшуються на 168грн.

$$\text{Визначимо коефіцієнт еластичності (1.2) } K_e = \frac{0,168 \cdot 8,95}{2,34} = 0,642.$$

На основі коефіцієнта еластичності можна зробити висновок, що зі збільшенням прибутків родини на 1% витрати на харчування збільшаться на 0,64%.

1.1.2 Оцінка тісноти й значимості зв'язку між змінними в рівняннях парної регресії.

Тісноту зв'язку оцінюють за допомогою таких характеристик: коефіцієнт кореляції й коефіцієнт детермінації.

Якщо зв'язок між показниками лінійний, то використовується **лінійний коефіцієнт кореляції**, що характеризує не тільки тісноту, але й напрямок.

Лінійний коефіцієнт кореляції r_{xy} можна розрахувати як:

$$r_{xy} = b \cdot \frac{S_x}{S_y} \quad (1.3)$$

Оцінка щільності зв'язку здійснюється за такою шкалою:

$$\begin{array}{ll} r_{xy} = 0 & \text{– зв'язок відсутній;} \\ 0,1 \leq |r_{xy}| < 0,3 & \text{– зв'язок слабкий;} \\ 0,3 \leq |r_{xy}| < 0,5 & \text{– зв'язок помірний;} \\ 0,5 \leq |r_{xy}| < 0,7 & \text{– зв'язок помітний;} \\ 0,7 \leq |r_{xy}| < 0,9 & \text{– зв'язок сильний;} \\ 0,9 \leq |r_{xy}| < 0,99 & \text{– зв'язок достатньо сильний;} \\ |r_{xy}| = 1 & \text{– зв'язок функціональний.} \end{array} \quad (1.4)$$

У випадку вже готового значення лінійного коефіцієнта кореляції, **коефіцієнт детермінації** розраховується як квадрат лінійного коефіцієнта кореляції r_{xy}^2 . Коефіцієнт детермінації показує, якою мірою варіація залежної змінної y визначається варіацією незалежної змінної x .

Коефіцієнт детермінації приймає значення від 0 (відсутній лінійний зв'язок між показниками) до 1 (відсутній кореляційний зв'язок між показниками).

Щоб мати загальне уявлення про якість моделі з відносних відхилень за кожним спостереженням, визначають **середню помилку апроксимації**:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}_x}{y} \right| \cdot 100\% . \quad (1.5)$$

Середня помилка апроксимації не повинна перевищувати 8-10%.

Оцінка значимості рівняння регресії в цілому провадиться на основі F - критерію Фішера. Величина F - критерію пов'язана з

коефіцієнтом детермінації r_{xy}^2 , і її можна розрахувати по наступній формулі:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2). \quad (1.6)$$

Фактичне значення F - критерію Фішера порівнюється з табличним значенням $F(a; k_1; k_2)$ при рівні значимості a й степенях свободи $k_1 = m$ й $k_2 = n - m - 1$ (для парної лінійної регресії $m = 1$). При цьому, якщо фактичне значення F - критерію більше за табличне, то визнається статистична значимість рівняння в цілому.

У парній лінійній регресії оцінюються **значимість** не тільки рівняння в цілому, але й окремих його **параметрів**. Із цією метою по кожному з параметрів визначається його стандартна помилка: m_b і m_a .

Стандартна помилка коефіцієнта регресії визначається по формулі:

$$m_b = \frac{S_{\hat{y}\hat{x}}}{s_x \cdot \sqrt{n}},$$

де $S_{\hat{y}\hat{x}}^2 = \frac{\sum (y - \hat{y}_x)^2}{n - 2}$ – залишкова дисперсія на одну степінь свободи.

Величина стандартної помилки разом з t - розподілом Стюдента при $n - 2$ степенях свободи застосовується для перевірки істотності коефіцієнта регресії й для розрахунку його довірчого інтервалу.

Для оцінки істотності коефіцієнта регресії його величина порівнюється з його стандартною помилкою, тобто визначається

фактичне значення t - критерію Стюдента: $t_b = \frac{b}{m_b}$, яке потім

порівнюється з табличним значенням при певному рівні значимості a й числі степенів свободи $n - 2$.

Довірчий інтервал для коефіцієнта регресії визначається як $b \pm t_{\text{табл}} \cdot m_b$.

Стандартна помилка параметра a визначається по формулі:

$$m_a = S_{\text{ост}} \cdot \frac{\sqrt{\sum x^2}}{s_x \cdot n}$$

Процедура оцінювання істотності даного параметра не відрізняється від розглянутої вище для коефіцієнта регресії.

Обчислюється t -критерій: $t_a = \frac{a}{m_a}$, його величина порівнюється з

табличним значенням при $n - 2$ степенях свободи.

Довірчий інтервал для параметра a визначається як $a \pm t_{\text{табл}} \cdot m_a$.

Значимість лінійного коефіцієнта кореляції перевіряється на основі величини помилки коефіцієнта кореляції m_r :

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

Обчислюється t -критерій: $t_r = \frac{r}{m_r}$, його величина порівнюється з

табличним значенням при $n - 2$ степенях свободи.

Розглянемо **приклад 1** Одержали рівняння: $\hat{y} = 0,836 + 0,168 \cdot x$.
Заповнимо стовпці 5-9.

	y^2	y_x	$y - y_x$	$(y - y_x)^2$	$A_i, \%$
	5	6	7	8	9
1	0,81	1,038	-0,138	0,0190	15,33

2	1,44	1,357	-0,157	0,0246	13,08
3	3,24	1,726	0,074	0,0055	4,11
4	4,84	2,079	0,121	0,0146	5,50
5	6,76	2,449	0,151	0,0228	5,81
6	8,41	2,818	0,082	0,0067	2,83
7	10,89	3,272	0,028	0,0008	0,85
8	14,44	3,978	-0,178	0,0317	4,68
Разом	50,83	18,717	-0,017	0,1257	52,19
Середнє значення	6,35	2,34	–	0,0157	6,52
S	–	–	–	–	–
S^2	–	–	–	–	–

Як було зазначено вище, рівняння лінійної регресії завжди доповнюється показником тісноти зв'язку – лінійним коефіцієнтом кореляції r_{xy} (1.3):

$$r_{xy} = b \cdot \frac{s_x}{s_y} = 0,168 \cdot \frac{5,53}{0,935} = 0,994.$$

Близькість коефіцієнта кореляції до 1 указує на те, що зв'язок між ознаками достатньо сильний.

Коефіцієнт детермінації $r_{xy}^2 = 0,987$ показує, що рівнянням регресії пояснюється 98,7% дисперсії результативної ознаки, а на частку інших факторів доводиться лише 1,3%.

Оцінимо якість рівняння регресії в цілому за допомогою F - критерію Фішера. Обчислимо фактичне значення F - критерію (1.6):

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2) = \frac{0,987}{1 - 0,987} \cdot 6 = 455,54.$$

Табличне значення $F_{табл} = 5,99$. $F > F_{табл}$, то визнається статистична значимість рівняння в цілому.

Для оцінки статистичної значимості коефіцієнтів регресії й кореляції розрахуємо t - критерій Стюдента й довірчі інтервали кожного з показників.

Розрахуємо випадкові помилки параметрів лінійної регресії й коефіцієнта кореляції

$$S_{\text{інд}}^2 = \frac{\sum (y - \hat{y}_x)^2}{n - 2} = \frac{0,1257}{8 - 2} = 0,021$$

$$m_b = \frac{S_{\text{оцт}}}{s_x \cdot \sqrt{n}} = \frac{\sqrt{0,021}}{5,53 \cdot \sqrt{8}} = 0,0093$$

$$m_a = S_{\text{оцт}} \cdot \frac{\sqrt{\sum x^2}}{s_x \cdot n} = \frac{\sqrt{0,021 \cdot 885,24}}{5,53 \cdot 8} = 0,0975$$

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - 0,987}{6}} = 0,0465$$

Фактичні значення t - статистик:

$$t_b = 18,065, t_a = 8,574, t_r = 21,376.$$

Табличне значення t критерію Стюдента при $\alpha = 0,05$ й числі ступенів свободи $n - 2 = 6$ є $t_{\text{табл}} = 2,447$. Так як $t_b > t_{\text{табл}}$, $t_a > t_{\text{табл}}$ і $t_r > t_{\text{табл}}$, то визнаємо статистичну значимість параметрів регресії й показника тісноти зв'язку.

Розрахуємо довірчі інтервали для параметрів регресії a й b : $a \pm t \cdot m_a$ і $b \pm t \cdot m_b$. Одержимо, що $a \in [0,597; 1,075]$ й $b \in [0,145; 0,191]$.

Середня помилка апроксимації (1.5) (знаходимо за допомогою стовпця 9) $\bar{A} = 6,52\%$ говорить про високу якість рівняння регресії, тобто свідчить про добрий підбір моделі до вихідних даних.

1.1.3 Використання рівняння регресії для складання прогнозу

У прогнозних розрахунках по рівнянню регресії визначається прогнозоване $\hat{y}_p$ значення, як точковий прогноз $\hat{y}_x$ при $x = x_p$. Однак точковий прогноз явно не реальний. Тому він доповнюється розрахунком стандартної помилки $\hat{y}_p$, тобто $m_{\hat{y}_p}$, і відповідно інтервальною оцінкою прогнозного значення $\hat{y}_p$:

$$\hat{y}_p - \Delta_{\hat{y}_p} \leq \hat{y}_p \leq \hat{y}_p + \Delta_{\hat{y}_p}$$

де $\Delta_{\hat{y}_p} = m_{\hat{y}_p} \cdot t_{\alpha/2; n-2}$, а $m_{\hat{y}_p}$ – середня помилка прогнозованого індивідуального значення:

$$m_{\hat{y}_p} = S_{\text{інд}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot s_x^2}}.$$

Знайдемо прогнозне значення результативного фактору $\hat{y}_p$ при значенні ознаки-фактору, що становить 110% від середнього рівня $x_p = 1,1 \cdot \bar{x} = 1,1 \cdot 8,95 = 9,845$, тобто знайдемо витрати на харчування, якщо прибутки родини складуть 9,85 тис. гр.

$$\hat{y}_p = 0,836 + 0,168 \cdot 9,845 = 2,490 \text{ (тис. гр.)}$$

Виходить, якщо прибутки родини складуть 9,845 тис. гр., то витрати на харчування будуть 2,490 тис. гр.

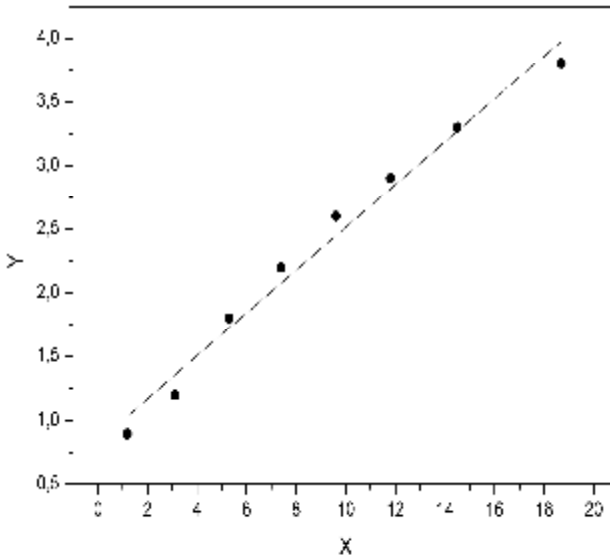
Знайдемо довірчий інтервал прогнозу. Помилка прогнозу

$$m_{\hat{y}_p} = S_{\text{інд}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot s_x^2}} = 0,154,$$

а довірчий інтервал: $2,113 \leq \hat{y}_p \leq 2,867$.

Прогноз є статистично надійним.

Тепер на одному графіку зобразимо вихідні дані й лінію регресії:



1.2 Нелінійні моделі парної регресії

Рівняння нелінійної регресії, так само, як і у випадку лінійної залежності, доповнюється показником тісноти зв'язку. У цьому випадку це **індекс кореляції**:

$$r_{xy} = \sqrt{1 - \frac{s_{\text{оч}}^2}{s_y^2}}, \quad (1.7)$$

$$\text{де } s_{\text{ін}}^2 = \frac{1}{n} \sum (y - \hat{y}_x)^2 \quad (1.8)$$

– залишкова дисперсія.

Чим ближче значення індексу кореляції до одиниці, тим тісніше зв'язок розглянутих ознак, тим більш надійне рівняння регресії (аналогічно (1.2)).

Квадрат індексу кореляції зветься **індексом детермінації** й характеризує частку дисперсії результативної ознаки, що пояснюється регресією, у загальній дисперсії результативної ознаки.

Індекс детермінації використовується для перевірки істотності в цілому рівняння регресії по F - критерію Фішера:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot \frac{n - m - 1}{m}, \quad (1.9)$$

де r_{xy}^2 – індекс детермінації, n – число спостережень, m – число параметрів при змінній x . Фактичне значення F критерію порівнюється з табличним при рівні значимості α й числі ступенів свободи $k_2 = n - m - 1$ (для залишкової суми квадратів) і $k_1 = m$ (для факторної суми квадратів).

Про якість нелінійного рівняння регресії можна також свідчити й по середній помилці апроксимації.

Розглянемо **приклад 1**, припустивши, що зв'язок між ознаками носить нелінійний характер, і знайдемо параметри наступних нелінійних рівнянь:

$$y = a + b \cdot \ln x + e,$$

$$y = a + b \cdot \sqrt{x} + e$$

$$y = a \cdot x^b + e.$$

Для знаходження параметрів регресії $\hat{y} = a + b \cdot \ln x$ робимо заміну $z = \ln x$ й складаємо допоміжну таблицю ($e = y - \hat{y}_x$).

	x	z	y	$z \cdot y$	z^2
	1	2	3	4	5
1	1,2	0,182	0,9	0,164	0,033
2	3,1	1,131	1,2	1,358	1,280
3	5,3	1,668	1,8	3,002	2,781
4	7,4	2,001	2,2	4,403	4,006
5	9,6	2,262	2,6	5,881	5,116

6	11,8	2,468	2,9	7,157	6,092
7	14,5	2,674	3,3	8,825	7,151
8	18,7	2,929	3,8	11,128	8,576
Разом	71,6	15,315	18,7	41,918	35,035
Середнє значення	8,95	1,914	2,34	5,240	4,379
S	–	0,846	0,935	–	–
S^2	–	0,716	0,874	–	–

	y^2	y_x	e	e^2	A_i
	6	7	8	9	10
1	0,81	0,499	0,401	0,1610	44,58
2	1,44	1,508	-0,308	0,0947	25,64
3	3,24	2,078	-0,278	0,0772	15,43
4	4,84	2,433	-0,233	0,0541	10,57
5	6,76	2,709	-0,109	0,0119	4,20
6	8,41	2,929	-0,029	0,0008	0,99
7	10,89	3,148	0,152	0,0232	4,62
8	14,44	3,418	0,382	0,1459	10,05
Разом	50,83	18,720	-0,020	0,5688	116,08
Середнє значення	6,35	–	–	0,0711	14,51
S	–	–	–	–	–
S^2	–	–	–	–	–

Знайдемо рівняння регресії: $\hat{y} = 0,305 + 1,063 \cdot \ln x$.

Індекс кореляції (1.7):

$$r_{xy} = \sqrt{1 - \frac{s_{ocm}^2}{s_y^2}} = \sqrt{1 - \frac{0,0711}{0,874}} = 0,958,$$

індекс детермінації $r^2_{xy} = 0,918$ який показує, що 91,8% варіації результативної ознаки пояснюється варіацією ознаки-фактору, а 8,2% доводиться на частку інших факторів.

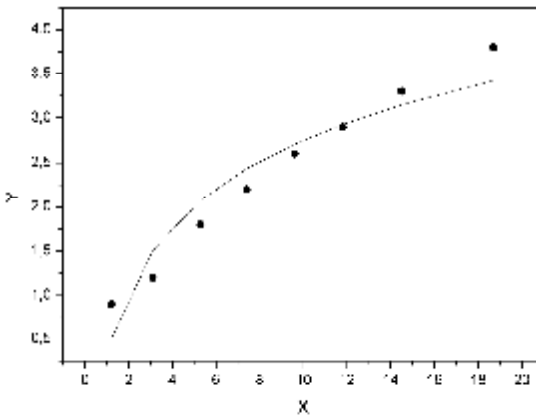
Середня помилка апроксимації (1.5): $\bar{A} = 14,51\%$, що неприпустимо багато.

F - критерій Фішера (1.9):

$$F = \frac{r^2_{xy}}{1 - r^2_{xy}} \cdot \frac{n - m - 1}{m} = \frac{0,919}{1 - 0,919} \cdot \frac{8 - 1 - 1}{1} = 68,07,$$

значно перевищує табличне $F_{табл} = 5,99$.

Зобразимо на графіку вихідні дані й лінію регресії:



Для знаходження параметрів регресії $\hat{y} = a + b \cdot \sqrt{x}$ робимо заміну $z = \sqrt{x}$ й складаємо допоміжну таблицю.

	x	z	y	$z \cdot y$
	1	2	3	4
1	1,2	1,10	0,9	0,99
2	3,1	1,76	1,2	2,11
3	5,3	2,30	1,8	4,14

4	7,4	2,72	2,2	5,98
5	9,6	3,10	2,6	8,06
6	11,8	3,44	2,9	9,96
7	14,5	3,81	3,3	12,57
8	18,7	4,32	3,8	16,43
Разом	71,6	22,54	18,7	60,24
Середнє значення	8,95	2,82	2,34	7,53
S	–	1,00	0,935	–
S²	–	1,00	0,874	–

Зайдемо рівняння регресії:

$$\hat{y} = -0,286 + 0,931 \cdot \sqrt{x}.$$

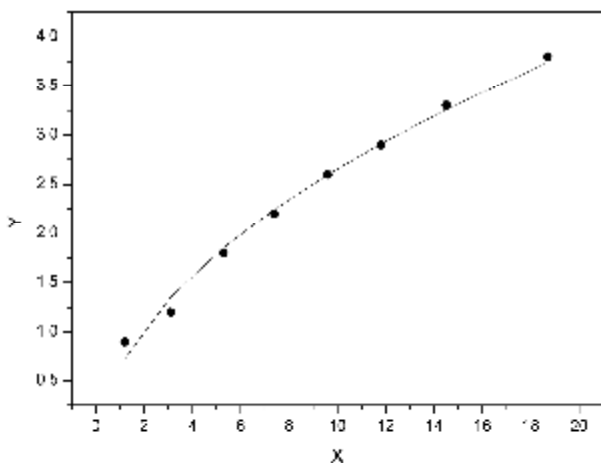
Індекс кореляції (1.7): $r_{xy} = 0,996$, індекс детермінації $r^2_{xy} = 0,991$, який показує, що 99,1% варіації результативної ознаки пояснюється варіацією ознаки-фактору, а 0,9% доводиться на частку інших факторів.

	z^2	y^2	y_x	e	e^2	A_i
	5	6	7	8	9	10
1	1,2	0,81	0,734	0,166	0,0276	18,46
2	3,1	1,44	1,353	-0,153	0,0235	12,77
3	5,3	3,24	1,857	-0,057	0,0033	3,19
4	7,4	4,84	2,247	-0,047	0,0022	2,12
5	9,6	6,76	2,599	0,001	0,0000	0,05
6	11,8	8,41	2,912	-0,012	0,0001	0,42
7	14,5	10,89	3,259	0,041	0,0017	1,20
8	18,7	14,44	3,740	0,060	0,0036	1,58
Разом	71,6	50,83	18,700	-0,001	0,0619	39,82
Середнє значення	8,95	6,35	–	–	0,0077	4,98
S	–	–	–	–	–	–
S²	–	–	–	–	–	–

Середня помилка апроксимації (1.5): $\bar{A} = 4,98\%$ показує, що лінія регресії добре наближає вихідні дані.

F - критерій Фішера (1.9) $F = 660,67$ значно перевищує табличне $F_{\text{табл}} = 5,99$.

Зобразимо на графіку вихідні дані й лінію регресії:



Для знаходження параметрів регресії $y = a \cdot x^b + e$ необхідно провести її лінеаризацію:

$$Y = A + b \cdot X + E,$$

$$\text{де } Y = \ln y, \quad X = \ln x, \quad A = \ln a, \quad E = \ln e.$$

Складаємо допоміжну таблицю для перетворених даних:

	X	Y	$X \cdot Y$	X^2
	1	2	3	4
1	0,182	-0,105	-0,019	0,033
2	1,131	0,182	0,206	1,280
3	1,668	0,588	0,980	2,781

4	2,001	0,788	1,578	4,006
5	2,262	0,956	2,161	5,116
6	2,468	1,065	2,628	6,092
7	2,674	1,194	3,193	7,151
8	2,929	1,335	3,910	8,576
Разом	15,315	6,002	14,637	35,035
Середнє значення	1,914	0,750	1,830	4,379
S	0,846	0,470	–	–
S^2	0,716	0,221	–	–

	Y^2	y_x	e	e^2	A_i
	5	6	7	8	9
1	0,011	0,8149	0,0851	0,0072	9,46
2	0,033	1,3747	-0,1747	0,0305	14,56
3	0,345	1,8473	-0,0473	0,0022	2,63
4	0,622	2,2203	-0,0203	0,0004	0,92
5	0,913	2,5627	0,0373	0,0014	1,43
6	1,134	2,8713	0,0287	0,0008	0,99
7	1,425	3,2165	0,0835	0,0070	2,53
8	1,782	3,7004	0,0996	0,0099	2,62
Разом	6,266	18,608	0,0919	0,0595	35,14
Середнє значення	0,783	–	–	0,0074	4,39
S	–	–	–	–	–
S^2	–	–	–	–	–

Знайдемо рівняння регресії: $\hat{Y} = -0,305 + 0,551 \cdot X$. Після потенціювання знаходимо шукане рівняння регресії:

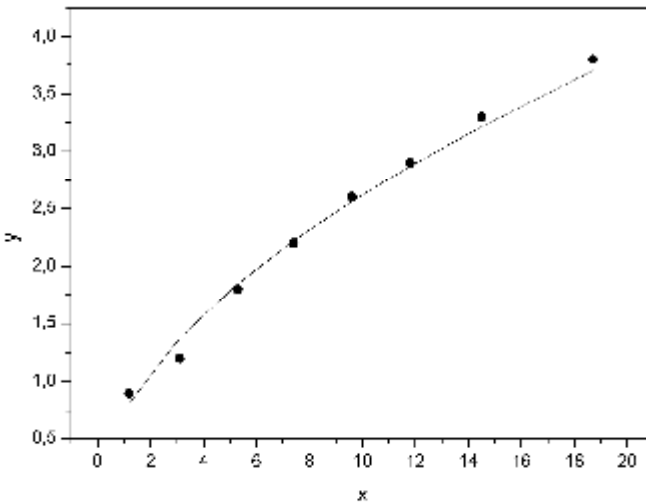
$$\hat{y}_x = 0,737 \cdot x^{0,551}.$$

Індекс кореляції (1.7): $r_{xy} = 0,983$, індекс детермінації $r^2_{xy} = 0,967$, який показує, що 96,7% варіації результативної ознаки пояснюється варіацією ознаки-фактору, а 3,3% доводиться на частку інших факторів.

Середня помилка апроксимації (1.5): $\bar{A} = 4,39\%$ показує, що лінія регресії добре наближає вихідні дані.

F -критерій Фішера (1.9) $F = 175,82$ значно перевищує табличне $F_{табл} = 5,99$.

Зобразимо на графіку вихідні дані й лінію регресії:



Порівняємо побудовані моделі за індексом детермінації й середній помилці апроксимації:

Модель	Індекс детермінації	Середня помилка апроксимації, %
Лінійна модель, $\hat{y} = a + b \cdot x$	0,987	6,52
Напівлогарифмічна модель, $\hat{y} = a + b \cdot \ln x$	0,918	14,51
Модель із квадратним коренем, $\hat{y} = a + b \cdot \sqrt{x}$	0,991	4,98
Степенева модель, $\hat{y} = a \cdot x^b \cdot e$	0,967	4,39

Найбільш добре вихідні дані апроксимує модель із квадратним коренем. Але в цьому випадку, так як індекси детермінації лінійної моделі й моделі із квадратним коренем відрізняються всього на 0,004, то цілком можна обійтися більш простою лінійною функцією.

2 МНОЖИННА РЕГРЕСІЯ І КОРЕЛЯЦІЯ

2.1 Вимоги до чинників множинної регресії

Парна регресія може дати добрий результат при моделюванні, якщо впливом інших чинників, що впливають на об'єкт дослідження, можна нехтувати. Якщо ж цим впливом нехтувати не можна, то в цьому випадку слід спробувати виявити вплив інших чинників, ввівши їх в модель, тобто побудувати рівняння множинної регресії

$$y = f(x_1, x_2, \dots, x_n),$$

де y - залежна змінна (результативна ознака), x_i - незалежні змінні (ознаки-чинники).

Основна мета множинної регресії - побудувати модель з великим числом чинників, визначивши при цьому вплив кожного з них окремо, а також сукупна їх дія на модельований показник.

Побудова рівняння множинної регресії починається з рішення питання про специфікацію моделі. Він включає два круги питань: відбір чинників і вибір виду рівняння регресії.

Чинники, що включаються в множинну регресію, повинні пояснити варіацію незалежної змінної. Якщо будується модель з набором m чинників, то для неї розраховується показник детермінації R^2 , який фіксує частку поясненої варіації результативної ознаки за рахунок m чинників, що розглядаються в регресії. Вплив інших, не врахованих в моделі чинників, оцінюється $1 - R^2$ з відповідною залишковою дисперсією S^2 .

При додатковому включенні в регресію $m + 1$ чинника коефіцієнт детермінації повинен зростати, а залишкова дисперсія зменшуватися. Якщо ж цього не відбувається і дані показники практично не відрізняються один від одного, то чинник x_{m+1} , що включається в аналіз, не покращує модель і практично є зайвим чинником.

Коефіцієнти інтеркореляції (тобто кореляція між пояснюючими змінними) дозволяють виключати з моделі дублюючі чинники. Вважається, що дві змінні явно колінеарні, тобто знаходяться між собою в лінійній залежності, якщо $r_{x_i, x_j} \geq 0,7$. Якщо чинники явно колінеарні, то вони дублюють один одного і один з них рекомендується виключити з регресії. Перевага при цьому віддається тому чиннику, який при достатньо тісному зв'язку з результатом має найменшу тісноту зв'язку з іншими чинниками.

Найбільші труднощі у використанні апарату множинної регресії виникають за наявності **мультиколінеарності** чинників, коли більш ніж два чинники зв'язані між собою лінійною залежністю, тобто має місце сукупна дія чинників один на одного. Наявність мультиколінеарності чинників може означати, що деякі чинники завжди діятимуть в унісон. В результаті варіація у початкових даних перестає бути повністю незалежною і не можна оцінити дію кожного чинника окремо.

2.2 Дослідження на мультиколінеарність

Для оцінки мультиколінеарності чинників може використовуватися визначник **матриці парних коефіцієнтів кореляції** між чинниками. Якщо чинники не корелюють між собою, то матриця парних коефіцієнтів кореляції між чинниками є

одиначною матрицею, визначник якої дорівнює одиниці. Якщо ж, навпаки, між чинниками існує повна лінійна залежність і всі коефіцієнти кореляції рівні одиниці, то визначник такої матриці дорівнює нулю.

В основу алгоритму дослідження на мультиколінеарність (алгоритм Фаррара - Глобера) покладено три види статистичних критеріїв, за якими перевіряється мультиколінеарність: критерій Пірсона c^2 , F - критерій Фішера, і t - критерій Стюдента.

1. Критерій Пірсона c^2

$$c^2 = -(n-1 - \frac{1}{6}(2m+5)) \ln |r|$$

де $|r|$ - визначник матриці $\|r_{x_i x_j}\|$.

Значення цього критерію порівнюється з даними таблиць c^2 при $\frac{1}{2}m(m-1)$ степенях свободи і рівні значущості α . Якщо $c^2 > c^2_{\text{табл}}$, то у масиві пояснюючих змінних існує мультиколінеарність.

2. t - критерій Стюдента

$$t_{x_i x_j} = \frac{r_{x_i x_j} \sqrt{n-m-1}}{\sqrt{1-r_{x_i x_j}^2}}$$

Фактичне значення порівнюється з табличним $k = n - m - 1$ степенях свободи і рівні значущості α .

Якщо $t_{x_i x_j} > t_{\alpha, k}$, то між чинниками X_i і X_j існує мультиколінеарність.

3. F - критерій Фішера

$$F_k = (C_{kk} - 1) \frac{n - m}{m - 1}$$

де C_{kk} - діагональні елементи матриці $C = \|r_{x_i x_j}\|^{-1}$. Фактичне значення порівнюють з табличним при $k_1 = n - m$ і $k_2 = m - 1$ степенях свободи і рівні значущості α . Якщо $F_k > F_{табл}$, тоді k -а пояснююча змінна мультиколінеарна з іншими.

2.3 Лінійна модель множинної регресії

Розглянемо лінійну модель множинної регресії

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m + e.$$

У лінійній множинній регресії параметри при x називаються **коефіцієнтами «чистої» регресії**. Вони характеризують середню зміну результату зі зміною відповідного чинника на одиницю при незміненому значенні інших чинників, закріплених на середньому рівні.

Класичний підхід до оцінювання параметрів лінійної моделі множинної регресії заснований на методі найменших квадратів (МНК). Отже, для знаходження параметрів лінійного рівняння множинної регресії приходимо до системи лінійних нормальних рівнянь (для моделі двох чинників):

$$\begin{cases} na + b_1 \sum x_1 + b_2 \sum x_2 = \sum y, \\ a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 = \sum y x_1, \\ a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 = \sum y x_2. \end{cases}$$

Метод найменших квадратів застосовний і до **рівняння множинної регресії в стандартизованому масштабі**:

$$t_y = b_1 t_{x_1} + b_2 t_{x_2} + \dots + b_m t_{x_m} + e$$

де $t_y, t_{x_1}, \dots, t_{x_m}$ - стандартизовані змінні:

$$t_y = \frac{y - \bar{y}}{s_y}, \quad t_x = \frac{x - \bar{x}}{s_x}, \quad (2.1)$$

для яких середнє значення рівне нулю: $\bar{t}_y = \bar{t}_x = 0$, а середнє квадратичне відхилення рівне одиниці: ; b_i - стандартизовані коефіцієнти регресії.

Стандартизовані коефіцієнти регресії показують, на скільки одиниць зміниться в середньому результат, якщо відповідний чинник зміниться на одну одиницю при незмінному середньому рівні інших чинників. Внаслідок того, що всі змінні задані як центровані і нормовані, стандартизовані коефіцієнти регресії b_i можна порівнювати між собою. Порівнюючи їх один з одним, можна ранжувати чинники по силі їх дії на результат. У цьому основна перевага стандартизованих коефіцієнтів регресії на відміну від коефіцієнтів «чистої» регресії, які незрівнянні між собою.

Коефіцієнти «чистої» регресії b_i пов'язані із стандартизованими коефіцієнтами регресії b_i таким чином:

$$b_i = b_i \frac{s_y}{s_{x_i}}$$

Тому можна переходити від рівняння регресії в стандартизованому масштабі до рівняння регресії в натуральному масштабі змінних, при цьому параметр a визначається як

$$a = \bar{y} - b_1 \bar{x}_1 - \dots - b_m \bar{x}_m .$$

Розглянута ідея стандартизованих коефіцієнтів регресії дозволяє використовувати їх при відсіві чинників - з моделі виключаються чинники з найменшим значенням b_i .

На основі лінійного рівняння множинної регресії

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m + e$$

можуть бути знайдені середні по сукупності показники еластичності:

$$\bar{\varepsilon}_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}_{x_i}}$$

які показують на скільки відсотків в середньому зміниться результат, при зміні відповідного чинника на 1%. Середні показники еластичності можна порівнювати один з одним і відповідно ранжувати чинники за силою їх дії на результат.

Приклад 2 (для скорочення об'єму обчислень обмежимося тільки десятима спостереженнями).

Хай є наступні дані (умовні) про змінний видобуток вугілля на одного робочого y (т), потужності пласта x_1 (м) і рівні механізації робіт x_2 (%), що характеризують процес видобутку вугілля у 10 шахтах

№	1	2	3	4	5	6	7	8	9	10
x_1	8	11	12	9	8	8	9	9	8	12
x_2	5	8	8	5	7	8	6	4	5	7
y	5	10	10	7	5	6	6	5	6	8

Досліджуємо на мультиколінеарність за допомогою алгоритму Фаррара - Глобера

1. Критерій Пірсона c^2

$$c^2 = -(n-1 - \frac{1}{6}(2m+5)) \ln|r| = (10-1 - \frac{2 \cdot 2 + 5}{6}) \ln 0,76 = 2,05$$

де $|r|$ - визначник матриці $\|r_{x_i x_j}\|$: $\begin{vmatrix} 1 & 0,488 \\ 0,488 & 1 \end{vmatrix} = 0,76$.

Отримане значення критерію порівнюється з даними таблиць c^2 при $k = \frac{1}{2}m(m-1) = \frac{1}{2} \cdot 2 \cdot (2-1) = 1$ степенях свободи і рівні значущості $\alpha = 0,05$ ($c^2_{табл} = 3,84$). Оскільки $c^2 < c^2_{табл}$, то в масиві пояснюючих змінних мультиколінеарності немає.

2. t - критерій Стьюдента

$$t_{x_1 x_2} = \frac{r_{x_1 x_2} \sqrt{n-m-1}}{\sqrt{1-r_{x_1 x_2}^2}} = \frac{0,488 \sqrt{10-2-1}}{\sqrt{1-0,488^2}} = 1,48$$

Фактичне значення порівнюється з табличним при $k = n - m - 1 = 10 - 2 - 1 = 7$ степенях свободи і рівні значущості $\alpha = 0,05$ ($t_{табл} = 2,36$).

Оскільки $t_{x_1 x_2} < t_{табл}$, то між чинниками x_1 і x_2 відсутня мультиколінеарність.

3. F критерій Фішера

$$F_k = (C_{kk} - 1) \frac{n-m}{m-1}$$

$$F_1 = (C_{11} - 1) \frac{10 - 2}{2 - 1} = (1,23 - 1) \cdot 8 = 0,028$$

$$F_2 = (C_{22} - 1) \frac{10 - 2}{2 - 1} = (1,29 - 1) \cdot 8 = 0,036$$

де C_{kk} - діагональні елементи матриці

$$C = \|r_{x_i x_j}\|^{-1} = \begin{pmatrix} 1,23 & -0,62 \\ -0,62 & 1,29 \end{pmatrix}$$

Фактичне значення порівнюють з табличним при i і $k_2 = m - 1 = 2 - 1 = 1$ степенях свободи і рівні значущості $\alpha = 0,05$ ($F_{табл} = 239$). Оскільки $F_1 < F_{табл}$, $F_2 < F_{табл}$ мультиколінеарність між x_1 і x_2 відсутній.

Припускаючи, що між змінними y, x_1, x_2 існує лінійна кореляційна залежність, знайдемо рівняння регресії y по x_1 і x_2 .

Для зручності подальших обчислень складаємо таблицю:

№	x_1	x_2	y	x_1^2	x_2^2	y^2
	1	2	3	4	5	6
1	8	5	5	64	25	25
2	11	8	10	121	64	100
3	12	8	10	144	64	100
4	9	5	7	81	25	49
5	8	7	5	64	49	25
6	8	8	6	64	64	36
7	9	6	6	81	36	36
8	9	4	5	81	16	25
9	8	5	6	64	25	36
10	12	7	8	144	49	64
Сума	94	63	68	908	417	496

Середнє значення	9,4	6,3	6,8	90,8	41,7	49,6
S^2	2,44	2,01	3,36	–	–	–
S	1,56	1,42	1,83	–	–	–

№	$x_1 \cdot x_2$	$x_1 \cdot y$	$x_2 \cdot y$	$\frac{y}{x}$	e^2
	7	8	9	10	11
1	40	40	5	5,13	0,016
2	88	110	0	8,79	1,464
3	96	120	0	9,64	0,127
4	45	63	5	5,98	1,038
5	56	40	5	5,86	0,741
6	64	48	8	6,23	0,052
7	54	54	36	6,35	0,121
8	36	45	0	5,61	0,377
9	40	48	0	5,13	0,762
10	84	96	6	9,28	1,631
Сума	603	664	445	68	6,329
Середнє значення	60,3	66,4	44,5	–	
S^2	–	–	–	–	–
S	–	–	–	–	–

Для знаходження параметрів рівняння регресії в даному випадку необхідно вирішити наступну систему нормальних рівнянь:

$$\begin{cases} 10a + 94b_1 + 63b_2 = 68, \\ 94a + 908b_1 + 603b_2 = 664, \\ 63a + 603b_1 + 417b_2 = 445. \end{cases}$$

Звідки отримуємо, що $a = -3,54$, $b_1 = 0,854$, $b_2 = 0,367$. Тобто отримали наступне рівняння множинної регресії:

$$\hat{y} = -3,54 + 0,854x_1 + 0,367x_2.$$

Воно показує, що при збільшенні тільки потужності пласта x_1 (при незмінному x_2) на 1 м видобуток вугілля на одного робочого y збільшиться в середньому на 0,854 т, а при збільшенні тільки рівня механізації робіт x_2 (при незмінному x_1) на 1% - в середньому на 0,367 т.

Знайдемо рівняння множинної регресії у стандартизованому масштабі:

$$t_y = b_1 t_{x_1} + b_2 t_{x_2} + e$$

при цьому стандартизовані коефіцієнти регресії будуть

$$b_1 = b_1 \frac{s_{x_1}}{s_{y_i}} = 0,854 \cdot \frac{1,56}{1,83} = 0,728$$

$$b_2 = b_2 \frac{s_{x_2}}{s_{y_i}} = 0,367 \cdot \frac{1,42}{1,83} = 0,285.$$

Тобто рівняння виглядатиме таким чином:

$$\hat{t}_y = 0,728t_{x_1} + 0,285t_{x_2}$$

Оскільки стандартизовані коефіцієнти регресії можна порівнювати між собою, то можна сказати, що потужність пласта надає більший вплив на змінний видобуток вугілля, чим рівень механізації робіт.

3 ВАРІАНТИ ЗАВДАНЬ

Завдання 1.

За територіями регіону наведено данні за деякий період.

Необхідно:

- 1). Побудувати лінійне рівняння парної регресії y від x
- 2). Розрахувати лінійний коефіцієнт парної кореляції и середню помилку апроксимації.
- 3). Оцінити статистичну значущість параметрів регресії та кореляції за допомогою:
 F - критерію Фішера і
 t - критерію Стюдента.
- 4). Виконати прогноз заробітної плати при прогнозному значенні середньодушового прожиткового мінімуму, який складає 107% від середнього рівня.
- 5). Оцінити точність прогнозу, виконавши розрахунок помилки прогнозу та його довірчий інтервал.
- 6). На одному графіку побудувати вихідні данні та теоретичну пряму.
- 7). Побудувати нелінійні моделі парної регресії

$$y = a + b \cdot \ln x + e$$

$$y = a + b \cdot \sqrt{x} + e$$

$$y = a \cdot x^b + e$$
- 8). З'ясувати, яка з моделей найбільше відповідає вихідним даним

Варіант 1

Номер регіону	Середньодушовий прожитковий мінімум в день одного працездатного, грн., x	Середньоденна заробітна плата, грн., y
1	81	124
2	77	131
3	85	146
4	79	139
5	93	143
6	100	159
7	72	135
8	90	152
9	71	127
10	89	154
11	82	127
12	111	162

Варіант 2

Номер регіону	Середньодушовий прожитковий мінімум в день одного працездатного, грн., x	Середньоденна заробітна плата, грн., y
1	74	122
2	81	134
3	90	136
4	79	125
5	89	120
6	87	127
7	77	125
8	93	148
9	70	122
10	93	157
11	87	144
12	121	165

Варіант 3

Номер регіону	Середньодушовий прожитковий мінімум в день одного працездатного, грн., x	Середньоденна заробітна плата, грн., y
1	77	123
2	85	152
3	79	140
4	93	142
5	89	157
6	81	181
7	79	133
8	97	163
9	73	134
10	95	155
11	84	132
12	108	165

Варіант 4

Номер регіону	Середньодушовий прожитковий мінімум в день одного працездатного, грн., x	Середньоденна заробітна плата, грн., y
1	83	137
2	88	142
3	75	128
4	89	140
5	85	133
6	79	153
7	81	142
8	97	154
9	79	132
10	90	150
11	84	132
12	112	166

Варіант 5

Номер регіону	Середньодушовий прожитковий мінімум в день одного працездатного, грн., x	Середньоденна заробітна плата, грн., y
1	79	134
2	91	154
3	77	128
4	87	138
5	84	133
6	76	144
7	84	160
8	94	149
9	79	125
10	98	163
11	81	120
12	115	162

Варіант 6

Номер регіону	Середньодушовий прожитковий мінімум в день одного працездатного, грн., x	Середньоденна заробітна плата, грн., y
1	92	147
2	78	133
3	79	128
4	88	152
5	87	138
6	75	122
7	81	145
8	96	141
9	80	127
10	102	151
11	83	129
12	94	147

Варіант 7

Номер регіону	Середньодушовий прожитковий мінімум в день одного працездатного, грн., x	Середньоденна заробітна плата, грн., y
1	75	133
2	78	125
3	81	129
4	93	153
5	86	140
6	77	135
7	83	141
8	94	152
9	88	133
10	99	156
11	80	124
12	112	156

Варіант 8

Номер регіону	Середньодушовий прожитковий мінімум в день одного працездатного, грн., x	Середньоденна заробітна плата, грн., y
1	69	124
2	83	133
3	92	146
4	97	153
5	88	138
6	93	159
7	74	145
8	79	152
9	105	168
10	99	154
11	85	127
12	94	155

Варіант 9

Номер регіону	Середньодушовий прожитковий мінімум в день одного працездатного, грн., x	Середньоденна заробітна плата, грн., y
1	78	133
2	94	139
3	85	141
4	73	127
5	91	154
6	88	142
7	73	122
8	82	135
9	99	142
10	113	168
11	69	124
12	83	130

Варіант 10

Номер регіону	Середньодушовий прожитковий мінімум в день одного працездатного, грн., x	Середньоденна заробітна плата, грн., y
1	97	161
2	73	131
3	79	135
4	99	147
5	86	139
6	91	151
7	85	135
8	77	132
9	89	161
10	95	159
11	72	120
12	115	160

Завдання 2

За даними 20 підприємств регіону вивчається залежність виробітку продукції на одного робітника y (тис. грн.) від введення в дію нових основних фондів x_1 (% від вартості фондів на кінець року) і від удільної ваги робочих високої кваліфікації в загальній чисельності робочих x_2 (%).

Необхідно:

- 1). Знайти коефіцієнти парної кореляції та проаналізувати їх.
- 2). Провести дослідження на мультиколінеарність за допомогою алгоритму Фаррара – Глобера.
- 3). Побудувати лінійну модель множинної регресії.
- 4). Записати стандартизоване рівняння множинної регресії.
- 5). На основі стандартизованих коефіцієнтів регресії та середніх коефіцієнтів еластичності ранжувати фактори за ступенем їх впливу на результат.
- 6). Скласти рівняння лінійної парної регресії, лишивши тільки один значущий фактор.

Варіант 1

Номер підприємства	y	x_1	x_2	Номер підприємства	y	x_1	x_2
1	6	3,6	9	11	9	6,3	21
2	6	3,6	12	12	11	6,4	22
3	6	3,9	14	13	11	7	24
4	7	4,1	17	14	12	7,5	25
5	7	3,9	18	15	12	7,9	28
6	7	4,5	19	16	13	8,2	30
7	8	5,3	19	17	13	8	30
8	8	5,3	19	18	13	8,6	31
9	9	5,6	20	19	14	9,5	33
10	10	6,8	21	20	14	9	36

Варіант 2

Номер підприємства	y	x_1	x_2	Номер підприємства	y	x_1	x_2
1	6	3,5	10	11	10	6,3	21
2	6	3,6	12	12	11	6,4	22
3	7	3,9	15	13	11	7	23
4	7	4,1	17	14	12	7,5	25
5	7	4,2	18	15	12	7,9	28
6	8	4,5	19	16	13	8,2	30
7	8	5,3	19	17	13	8,4	31
8	9	5,3	20	18	14	8,6	31
9	9	5,6	20	19	14	9,5	35
10	10	6	21	20	15	10	36

Варіант 3

Номер підприємства	y	x_1	x_2	Номер підприємства	y	x_1	x_2
1	7	3,7	9	11	11	6,3	22
2	7	3,7	11	12	11	6,4	22
3	7	3,9	11	13	11	7,2	23
4	7	4,1	15	14	12	7,5	25
5	8	4,2	17	15	12	7,9	27
6	8	4,9	19	16	13	8,1	30
7	8	5,3	19	17	13	8,4	31
8	9	5,1	20	18	13	8,6	32
9	10	5,6	20	19	14	9,5	35
10	10	6,1	21	20	15	9,5	36

Варіант 4

Номер підприємства	y	x_1	x_2	Номер підприємства	y	x_1	x_2
1	7	3,5	9	11	10	6,3	22
2	7	3,6	10	12	10	6,5	22
3	7	3,9	12	13	11	7,2	24
4	7	4,1	17	14	12	7,5	25
5	8	4,2	18	15	12	7,9	27
6	8	4,5	19	16	13	8,2	30
7	9	5,3	19	17	13	8,4	31
8	9	5,5	20	18	14	8,6	33
9	10	5,6	21	19	14	9,5	35
10	10	6,1	21	20	15	9,6	36

Варіант 5

Номер підприємства	y	x_1	x_2	Номер підприємства	y	x_1	x_2
1	7	3,6	9	11	10	6,3	21
2	7	3,6	11	12	11	6,9	23
3	7	3,7	12	13	11	7,2	24
4	8	4,1	16	14	12	7,8	25
5	8	4,3	19	15	13	8,1	27
6	8	4,5	19	16	13	8,2	29
7	9	5,4	20	17	13	8,4	31
8	9	5,5	20	18	14	8,8	33
9	10	5,8	21	19	14	9,5	35
10	10	6,1	21	20	14	9,7	34

Варіант 6

Номер підприємства	y	x_1	x_2	Номер підприємства	y	x_1	x_2
1	7	3,5	9	11	10	6,3	21
2	7	3,6	10	12	10	6,8	22
3	7	3,8	14	13	11	7,2	24
4	7	4,2	15	14	12	7,9	25
5	8	4,3	18	15	12	8,1	26
6	8	4,7	19	16	13	8,3	29
7	9	5,4	19	17	13	8,4	31
8	9	5,6	20	18	13	8,8	32
9	10	5,9	20	19	14	9,6	35
10	10	6,1	21	20	14	9,7	36

Варіант 7

Номер підприємства	y	x_1	x_2	Номер підприємства	y	x_1	x_2
1	7	3,8	11	11	10	6,8	21
2	7	3,8	12	12	11	7,4	23
3	7	3,9	16	13	11	7,8	24
4	7	4,1	17	14	12	7,5	26
5	7	4,6	18	15	12	7,9	28
6	8	4,5	18	16	12	8,1	30
7	8	5,3	19	17	13	8,4	31
8	9	5,5	20	18	13	8,7	32
9	9	6,1	20	19	13	9,5	33
10	10	6,8	21	20	14	9,7	35

Варіант 8

Номер підприємства	y	x_1	x_2	Номер підприємства	y	x_1	x_2
1	7	3,8	9	11	11	7,1	22
2	7	4,1	14	12	11	7,5	23
3	7	4,3	16	13	12	7,8	25
4	7	4,1	17	14	12	7,6	27
5	8	4,6	17	15	12	7,9	29
6	8	4,7	18	16	13	8,1	30
7	9	5,3	20	17	13	8,5	32
8	9	5,5	20	18	14	8,7	32
9	11	6,9	21	19	14	9,6	33
10	10	6,8	21	20	15	9,8	36

Варіант 9

Номер підприємства	y	x_1	x_2	Номер підприємства	y	x_1	x_2
1	7	3,9	12	11	11	7,1	22
2	7	4,2	13	12	12	7,5	25
3	7	4,3	15	13	13	7,8	26
4	7	4,4	17	14	12	7,9	27
5	8	4,6	18	15	13	8,1	30
6	8	4,8	19	16	13	8,4	31
7	9	5,3	19	17	13	8,6	32
8	9	5,7	20	18	14	8,8	32
9	10	6,9	21	19	14	9,6	34
10	10	6,8	21	20	14	9,9	36

Варіант 10

Номер підприємства	y	x_1	x_2	Номер підприємства	y	x_1	x_2
1	7	3,6	12	11	10	7,2	23
2	7	4,1	14	12	11	7,6	25
3	7	4,3	16	13	12	7,8	26
4	7	4,4	17	14	11	7,9	28
5	7	4,5	18	15	12	8,2	30
6	8	4,8	19	16	12	8,4	31
7	8	5,3	20	17	12	8,6	32
8	8	5,6	20	18	13	8,8	32
9	9	6,7	21	19	13	9,2	33
10	10	6,9	22	20	14	9,6	34

ЛІТЕРАТУРА

1. Економетрія: Навч. посіб. для студ. вищ. навч. заклад./ О.Л.Лещинський, В.В. Рязанцева, О.О. Юнькова – К.:МАУП, 2003. – 208 с.
2. Корольов О.А. Економетрія: Навч. посіб. – К.: Європейський ун-т, 2002. – 660с.
3. Кулинич О.І. Економетрія: Навч. посіб. – Хмельницький : Поділля, 1997. – 120с.
4. Лук'яненко І.Г., Краснікова Л.І. Економетрія: Підручник. – К.: Знання, КОО, 1998.-494с.
5. Вентцель Е. С. Теория вероятностей. – М.: Наука, 1999.
6. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высш. шк., 1975.
7. Гмурман В. Е. Теория вероятностей и математическая статистика. М.: Высш. шк., 1977.
8. Гнеденко В. В. Курс теории вероятностей. – М.: Наука, 1965.
9. Гурский Е. И. Теория вероятностей с элементами математической статистики. – М.: Высш. шк., 1971.
10. Емельянов Г. В., Скитович В. П. Задачник по теории вероятностей и математической статистике. – Л.: Изд-во ЛГУ, 1967.
11. Карасев А. И. Теория вероятностей и математическая статистика. М.: Статистика, 1970.

Додаток А

Для підключення пакету «Анализ данных» необхідно

Файл → Параметры Excel → Надстройки → Пакет анализа
→

Перейти → Пакет анализа  → ОК.

Після чого у розділі « Данные » з'явиться «Анализ данных». Для побудови регресійної моделі за допомогою «Анализа данных» необхідно

**Анализ данных → Регрессия → ОК (График остатков ,
График подбора ).**

Витрати на продукти харчування, у , тис. грн.	0,9	1,2	1,8	2,2	2,6	2,9	3,3	3,8
Прибутки родини, х , тис. грн.	1,2	3,1	5,3	7,4	9,6	11,8	14,5	18,7

ВЫВОД ИТОГОВ

<i>Регрессионная статистика</i>	
Множественный R	0,9912
R-квадрат	0,9824
Нормированный R-квадрат	0,9795
Стандартная ошибка	0,1443
Наблюдения	8,0000

Дисперсионный анализ

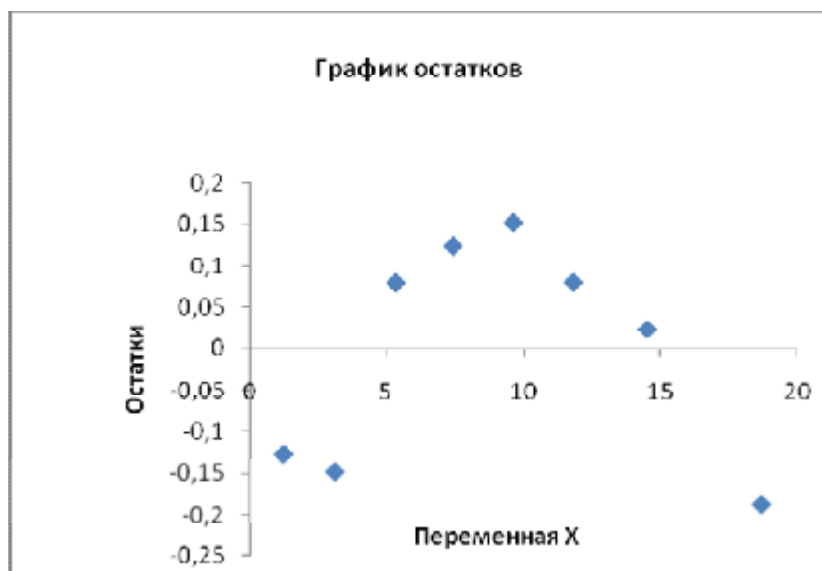
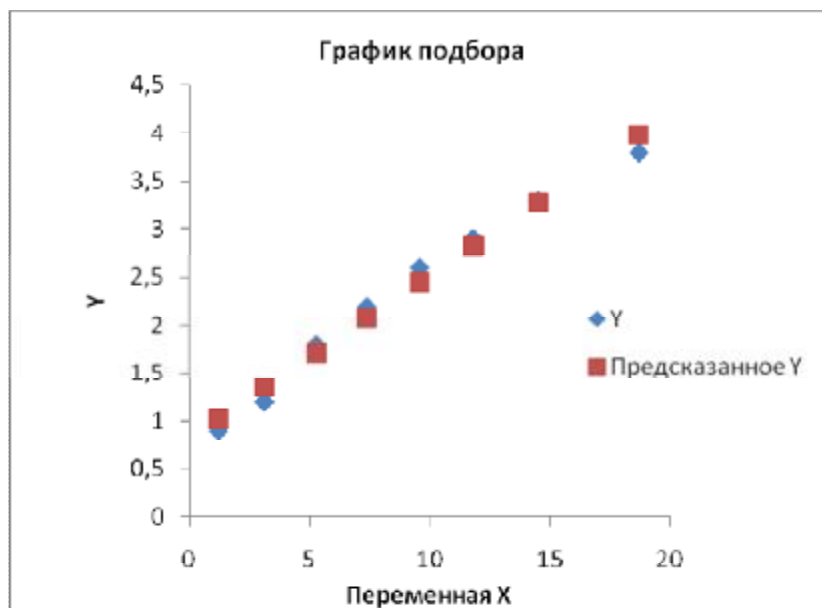
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	1	6,9937	6,9937	335,6625	0,00000170

Остаток	6	0,1250	0,0208
Итого	7	7,1188	

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y	0,8236	0,0971	8,4796	0,0001	0,5859	1,0612	0,5859	1,0612
X	0,1692	0,0092	18,3211	0,0000	0,1466	0,1917	0,1466	0,1917

ВЫВОД ОСТАТКА

<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>
1	1,026544473	-0,126544473
2	1,347940021	-0,147940021
3	1,720082235	0,079917765
4	2,075308895	0,124691105
5	2,447451109	0,152548891
6	2,819593323	0,080406677
7	3,276313313	0,023686687
8	3,986766631	-0,186766631



ВЫВОД ИТОГОВ

Регрессионная статистика

Множественный R – індекс кореляції (коефіцієнт кореляції);

R-квадрат – індекс детермінації (коефіцієнт детермінації);

Стандартная ошибка – корень з залишкової дисперсії;

Наблюдения – кількість значень змінної **Y**;

Дисперсионный анализ

df – значення ступеня свободи;

SS – сума квадратів;

MS – дисперсії;

F – значення критеріальної статистики для перевірки значущості регресії;

Значимость F – ймовірність отриманого значення критеріальної статистики. Коли ця ймовірність менше рівня значущості регресія значима;

Кoeffициенты - в рядку Y вказаний параметр **a**, а в рядку X – коефіцієнт регресії **b**;

Стандартная ошибка - в рядку Y вказана стандартна похибка параметра **a**, в рядку X – стандартна похибка коефіцієнту регресії **b**;

t-статистика – значення критеріальних статистик (відношення значення параметру до величини відповідної стандартної похибки) для перевірки гіпотези про значущість параметрів регресії;

P-Значение – розраховуються рівні значущості, відповідні значенням критеріальних статистик. Коли розрахунковий рівень менше за рівень значущості, відповідний коефіцієнт значущий;

Нижние 95%, Верхние 95% - границі довірчих інтервалів: в рядку Y вказані за параметром **a**, в рядку X – за коефіцієнтом регресії **b**;

ВЫВОД ОСТАТКА

Предсказанное Y - значення Y, розраховане за рівнянням регресії;

Остатки – різниця між теоретичним та експериментальним значенням для відповідного значення X.

Завдання для самоконтролю

1. Найбільш наглядним засобом відбору рівняння парної регресії є:

- а) аналітичний;
- б) графічний;
- в) експериментальний (табличний).

2. Розрахувати параметри рівняння парної лінійної регресії можливо, коли маємо:

- а) не менше 5 спостережень;
- б) не менше 7 спостережень;
- в) не менше 10 спостережень.

3. Задача метода найменших квадратів є:

- а) мінімізація суми залишкових величин;
- б) мінімізація дисперсії результативного признаку;
- в) мінімізація суми квадратів залишкових величин.

4. Коефіцієнт кореляції лінійного парного рівняння регресії:

- а) вказує середнє змінєння результату зі змінєю фактора на одну одиницю;
- б) оцінює статистичну значущість рівняння регресії;
- в) вказує, на скільки відсотків зміниться у середньому результат, коли фактор зміниться на 1%.

5. Співпадають чи ні знаки коефіцієнта регресії з теоретичним уявленням? (відповідь обґрунтувати)

- а) так;
- б) ні.

6. Коефіцієнт детермінації r_{xy}^2 ...

- а) оцінює якість моделі з відносних відхилень за кожним спостереженням;
- б) показує, якою мірою варіація залежної змінної у визначається варіацією незалежної змінної x ;
- в) характеризує не тільки тісноту, але й напрямок зв'язку.

7. Якість моделі з відносних відхилень за кожним спостереженням характеризує:

- а) коефіцієнт детермінації r_{xy}^2 ;

б) F - критерії Фішера;

в) середня помилка апроксимації $\bar{A}$.

8. Значущість рівняння регресії у цілому оцінює:

а) F - критерії Фішера;

б) t критерій Стюдента;

в) коефіцієнт детермінації r_{xy}^2 .

9. Залишкова сума квадратів дорівнює нулю:

а) якщо правильно підібрана регресійна модель;

б) коли між ознаками має місце функціональний зв'язок;

в) ніколи.

10. Для оцінки значущості коефіцієнтів регресії розраховують:

а) F - критерії Фішера;

б) t критерій Стюдента;

в) коефіцієнт детермінації r_{xy}^2 .

11. Коефіцієнт кореляції r_{xy} може приймати значення:

а) від -1 до 1 ;

б) від 0 до 1 ;

в) усі натуральні.

12. Фінансова ситуація на фірмі склалася таким чином, що витрати на рекламу були різко зменшені. У скорому часі об'єми продаж скоротилися, але в меншій мірі ніж очікувалось. Якому значенню коефіцієнта кореляції між витратами на рекламу та об'ємами продаж може відповідати дана ситуація?

а) $-0,6$;

б) $0,9$;

в) $0,5$

г) $-0,3$.