

ПРАКТИЧНА РОБОТА №2

ПАРНА ЛІНІЙНА РЕГРЕСІЯ ТА ЇЇ АНАЛІЗ

Завдання. На основі вхідних статистичних даних фактору X та результативного показника Y , табл. 1-2, знайти:

1. Значення коефіцієнту кореляції між ними;
2. Параметри a та b лінійної регресії $Y_{\text{розн}} = a + bx$;
3. Використовуючи t -статистику, з надійністю $P = 0,95$ оцінити значущість коефіцієнта кореляції;
4. Використовуючи критерій Фішера, з надійністю $P = 0,95$ оцінити адекватність побудованої лінійної економетричної моделі вхідним статистичним даним;
5. Оцінити надійні інтервали параметрів a та b парної лінійної регресії $Y_{\text{розн}} = a + bx$;
6. Якщо побудована лінійна економетрична модель адекватна вхідним статистичним даним, то знайти:
 - прогноз результативного показника $Y_{\text{розн}}$ для заданого значення X_p ;
 - надійні інтервали для базисних значень результативного показника Y ;
 - надійні інтервали прогнозних значень результативного показника $Y_{\text{розн}}$;
 - коефіцієнт еластичності результативного показника $Y_{\text{розн}}$ від фактору x .
7. Побудувати графіки вхідних статистичних даних, лінії регресії та її довірчих інтервалів, коефіцієнта еластичності.
8. На основі одержаної економетричної моделі зробити висновки.

Таблиця 1 – Значення фактору X по варіантах

№	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	2,06	2,53	2,17	3,65	3,22	2,16	4,57	2,25	6,15	1,86	2,07	3,11
2	2,58	3,54	2,90	3,82	3,87	2,65	5,42	2,98	5,66	1,91	3,22	3,15
3	3,14	3,84	3,29	3,76	4,95	3,49	5,29	2,15	7,50	2,14	3,04	3,85
4	3,54	4,84	4,13	5,24	5,10	3,16	6,33	2,71	6,90	3,39	3,42	4,84
5	4,13	5,25	5,24	5,98	5,53	3,85	7,63	3,70	8,31	3,95	5,23	4,62
6	4,78	4,81	5,52	5,62	7,28	4,58	7,53	4,59	8,25	4,30	5,70	4,87
7	5,11	5,53	5,79	5,62	6,90	5,33	7,73	4,77	9,39	5,10	6,53	6,09
8	5,48	5,82	5,99	6,87	6,98	5,89	8,44	5,34	9,73	5,47	6,41	7,06
9	6,02	6,43	6,87	6,91	7,91	6,20	9,49	5,45	9,33	5,97	6,68	6,23
10	6,65	7,73	7,04	7,95	8,40	6,39	9,18	6,00	10,50	6,16	7,46	6,83
11	7,05	8,19	8,14	7,24	8,14	6,95	10,14	6,25	11,10	6,46	6,83	8,01
12	7,65	8,06	9,27	8,76	9,25	6,79	11,51	6,07	11,51	6,67	6,93	8,26
13	8,03	9,31	8,57	9,67	10,08	7,80	10,92	8,24	12,42	6,71	8,19	9,37
14	8,56	9,26	9,45	10,28	10,47	8,47	11,89	8,51	12,40	7,16	9,19	9,02
15	9,03	9,69	10,72	11,04	11,32	8,92	11,43	9,14	12,23	7,52	9,61	9,32
Xp	9,52	9,69	10,30	10,05	11,58	9,32	11,73	9,78	12,56	8,07	8,71	10,28

Таблиця 2 – Значення результативного показника Y по варіантах

№	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12
1	7,2	10,8	16,2	12,1	15,2	16,6	10,2	12,5	19,6	14,8	22,6	10,6
2	8,0	11,9	17,7	12,3	15,4	17,0	10,5	13,4	20,5	15,7	23,8	11,8
3	9,2	12,4	18,3	13,1	16,0	18,1	12,0	14,0	21,5	17,0	25,0	12,3
4	10,1	13,2	18,8	14,8	18,5	20,5	13,2	16,0	22,5	18,0	25,9	13,4
5	11,1	14,1	19,6	15,8	18,5	20,5	13,2	16,6	23,2	18,5	26,9	15,1
6	12,1	15,2	21,6	16,4	19,8	21,9	15,1	17,7	24,1	19,9	27,8	15,7
7	13,0	16,0	21,8	17,8	20,7	22,4	15,8	18,4	25,7	20,3	28,1	16,2
8	14,1	17,4	23,0	18,6	21,3	23,5	17,0	19,5	26,7	22,1	29,8	18,0
9	15,2	18,6	24,4	19,5	22,2	24,9	17,9	20,5	27,3	22,4	30,3	18,4
10	16,2	19,4	25,3	21,2	24,1	25,5	18,3	21,7	28,3	23,4	31,5	19,5
11	17,0	20,5	25,5	21,0	24,1	26,1	19,4	22,1	29,1	24,0	32,9	20,4
12	18,0	21,3	27,1	22,9	25,6	28,0	20,5	23,8	30,5	25,5	33,7	21,7
13	19,1	22,5	27,9	23,4	26,5	28,3	21,3	24,7	31,2	27,0	34,6	23,1
14	20,2	23,7	28,9	24,6	27,4	29,4	23,2	25,5	32,5	27,5	35,9	23,5
15	21,2	25,0	30,8	25,4	29,0	30,4	24,2	27,1	33,6	28,4	36,9	24,4

Рекомендації до виконання.

1. Коефіцієнт кореляції між фактором X та результативним показником Y розраховується за формулою:

$$r[X, Y] = \frac{n \cdot \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \cdot \sqrt{n \cdot \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}}$$

Коефіцієнт кореляції також можна розрахувати за допомогою вбудованих функцій Microsoft Excel:

= *CORREL*($x; y$), або

= *KORPEЛ*($x; y$).

Обчислене значення коефіцієнту кореляції інтерпретується наступним чином:

– $r \in [-1; -0,7)$, або $Corr \in (+0,7; +1]$ – сильна зворотна, або пряма лінійна залежність;

– $r \in [-0,7; -0,3)$, або $Corr \in (+0,3; +0,7]$ – середня зворотна, або пряма лінійна залежність. Зв'язок існує, але він менш сильний;

– $r \in [-0,3; -0,1)$, або $Corr \in (+0,1; +0,3]$ – слабка зворотна, або пряма лінійна залежність. Зв'язок слабкий;

– $r \in [-0,1; +0,1]$ – дуже слабка, або відсутня лінійна залежність.

– $r \rightarrow 0$ – кореляція відсутня, лінійного статистичного взаємозв'язку між змінними не спостерігається.

2. Параметри a та b лінійної регресії $Y_{\text{розн}} = a + bx$ знаходяться за формулами:

$$\begin{cases} a = \frac{n \cdot \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}, \\ b = \bar{y} - a\bar{x} \end{cases}$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

Параметри a та b лінійної регресії $Y_{\text{розн}} = a + bx$ у Microsoft Excel також можна знайти за допомогою вбудованої функції =LINEST(), або =ЛИНЕЙН(). Дана функція має наступні аргументи:

- I аргумент – вектор-стовпець відомих значень результативного показника Y;
- II аргумент – матриця відомих значень факторів X.
- III, IV аргументи – логічні параметри, повинні приймати значення «істина», або 1.

Після введення даної функції, виділити діапазон комірок, куди повинні повернутись розраховані значення. Розмірність цього діапазону $5 \times (n + 1)$. Далі слід натиснути <F2> та одночасно <CTRL>+<SHIFT>+<ENTER>. Розраховані значення параметрів регресії, які необхідно було знайти, будуть розташовані у першому рядку:

b_n	...	b_2	b_1	b_0
...	...	...	...	...
$F_{\text{розр}}$				
...	...	...	...	...

3. Оцінимо значущість коефіцієнта кореляції з надійністю $P = 0,95$, використовуючи t-статистику.

Критерій Ст'юдента про значущість коефіцієнта кореляції передбачає виконання умови:

$$|t| \geq t_{pk},$$

Розрахункове значення t-статистики обчислюється за формулою:

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

p – заданий рівень ймовірності ($p = 0,95$);

n – кількість спостережень;

m – кількість факторів x , що впливають на y ($m = 1$);

$k = (n - m - 1)$ – кількість ступенів вільності.

Для знаходження табличного значення критерія Ст'юдента t_{pk} , слід використовувати вбудовану функцію Microsoft Excel:

$= TINV(1-p; k)$, або

$= СТБЮДРАСПОБР(1-p; k)$.

4. Використовуючи критерій Фішера, з надійністю $P = 0,95$ оцінимо адекватність побудованої лінійної економетричної моделі вхідним статистичним даним.

Достовірність всіх побудованих прогнозів в даному завданні будемо перевіряти за допомогою критерію Фішера. Для цього спочатку слід обчислити вибіркового коефіцієнт детермінації R^2 :

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

Тоді, критерій Фішера про значущість характеристик парної лінійної регресії буде мати вигляд:

$$F \geq F_{pk},$$

Розрахункове значення критерію Фішера обчислюється за формулою:

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

p – заданий рівень ймовірності ($p = 0,95$);

n – кількість спостережень;

m – кількість факторів x , що впливають на y ($m = 1$);

$k = (n - m - 1)$ – кількість ступенів вільності.

Табличне значення F_{pk} можна розрахувати в Microsoft Excel за допомогою функції:

$= FINV(1-p; m; n-m-1)$, або

$= FPACPOBP(1-p; m; n-m-1)$.

5. Оцінимо надійні інтервали параметрів a та b парної лінійної регресії

$Y_{\text{розр}} = a + bx$:

– для параметра a будемо мати:

$$a - \Delta a \leq \alpha \leq a + \Delta a$$

$$\Delta a = t_{pk} \cdot S \cdot \frac{1}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

– для параметра b будемо мати:

$$b - \Delta b \leq \beta \leq b + \Delta b$$

$$\Delta b = t_{pk} \cdot S \cdot \sqrt{\frac{\sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2}}$$

Де

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

6. Прогноз результативного показника $Y_{\text{розр}}$ для заданого значення X_p виконується на основі отриманої регресії $Y_{\text{прогноз}} = a + bx$.

Надійні інтервали для базисних значень результативного показника Y визначаються за формулами:

$$\hat{y}_i - \Delta \hat{y}_i \leq y_i \leq \hat{y}_i + \Delta \hat{y}_i$$

$$\Delta \hat{y}_i = t_{pk} \cdot S[\hat{y}_i]$$

$$S[\hat{y}_i] = S \cdot \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Надійні інтервали прогнозних значень результативного показника $Y_{\text{прогноз}}$ визначаються за формулами:

$$\hat{y}_{\text{прогноз}} - \Delta \hat{y}_{\text{прогноз}} \leq y_{\text{прогноз}} \leq \hat{y}_{\text{прогноз}} + \Delta \hat{y}_{\text{прогноз}}$$

$$\Delta \hat{y}_{\text{прогноз}} = t_{pk} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Коефіцієнт еластичності результативного показника $Y_{\text{розр}}$ від фактору x для базисних даних й прогнозу розраховується за формулою:

$$K_{x_i} = \frac{a \cdot x_i}{\hat{y}(x_i)}$$

7. Побудувати графіки вхідних статистичних даних, лінії регресії та її довірчих інтервалів, коефіцієнта еластичності.

ДОДАТКИ

Додаток А

Табличні значення коефіцієнта Ст'юдента

Кількість ступенів свободи $f = n - 1$	Довірча вірогідність			
	0,90	0,95	0,99	0,999
1	6,3138	12,7062	63,6567	636,6192
2	2,9200	4,3027	9,9248	31,5991
3	2,3534	3,1824	5,8409	12,9240
4	2,1318	2,7764	4,6041	8,6103
5	2,0150	2,5706	4,0321	6,8688
6	1,9432	2,4469	3,7074	5,9588
7	1,8946	2,3646	3,4995	5,4079
8	1,8595	2,3060	3,3554	5,0413
9	1,8331	2,2622	3,2498	4,7809
10	1,8125	2,2281	3,1693	4,5869
11	1,7959	2,2010	3,1058	4,4370
12	1,7823	2,1788	3,0545	4,3178
13	1,7709	2,1604	3,0123	4,2208
14	1,7613	2,1448	2,9768	4,1405
15	1,7531	2,1314	2,9467	4,0728
16	1,7459	2,1199	2,9208	4,0150
17	1,7396	2,1098	2,8982	3,9651
18	1,7341	2,1009	2,8784	3,9216
19	1,7291	2,0930	2,8609	3,8834
20	1,7247	2,0860	2,8453	3,8495
21	1,7207	2,0796	2,8314	3,8193
22	1,7171	2,0739	2,8188	3,7921
23	1,7139	2,0687	2,8073	3,7676
24	1,7109	2,0639	2,7969	3,7454
25	1,7081	2,0595	2,7874	3,7251
26	1,7056	2,0555	2,7787	3,7066
27	1,7033	2,0518	2,7707	3,6896
28	1,7011	2,0484	2,7633	3,6739
29	1,6991	2,0452	2,7564	3,6594
30	1,6973	2,0423	2,7500	3,6460
40	1,6839	2,0211	2,7045	3,5510
60	1,6706	2,0003	2,6603	3,4602
∞	1,6449	1,9600	2,5758	3,2905