

Лабораторна робота №2

ОБЧИСЛЕННЯ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ ТА ПЕРЕВІРКА ЙОГО СТАТИСТИЧНОЇ ЗНАЧУЩОСТІ

Мета роботи – навчитися обчислювати і визначати статистичну значущість кореляції між факторами.

Після виконання роботи студент повинен:

ЗНАТИ визначення коефіцієнта кореляції, межі зміни коефіцієнта кореляції, тести для перевірки статистичної значущості коефіцієнта кореляції.

УМІТИ розраховувати значення коефіцієнта кореляції, визначати характер залежності між факторами моделі, користуватися статистичними таблицями, визначати статистичну значущість коефіцієнта кореляції.

МАТИ УЯВЛЕННЯ про перевірку статистичних гіпотез.

Завдання:

1. Побудувати шаблон кореляційної таблиці.
2. Визначити коефіцієнт кореляції за допомогою кореляційної таблиці.
3. Визначити характер зв'язку між параметрами моделі.
4. Визначити статистичну значущість коефіцієнта кореляції за допомогою t-критерію Стюдента.

Хід роботи:

1. Побудуємо шаблон кореляційної таблиці (табл. 2.1):

Таблиця 2.1 – Кореляційна таблиця

n	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1					
2					
...					
n					
$\sum_{i=1}^n$	$\sum_{i=1}^n x_i$	$\sum_{i=1}^n y_i$	$\sum_{i=1}^n x_i y_i$	$\sum_{i=1}^n x_i^2$	$\sum_{i=1}^n y_i^2$

2 Розрахуємо оцінку коефіцієнта кореляції \bar{r} для визначення характеру залежності між параметрами моделі за формулою

$$\bar{r} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left(n \sum x_i^2 - \left(\sum x_i \right)^2 \right) \left(n \sum y_i^2 - \left(\sum y_i \right)^2 \right)}}$$

де n – обсяг вибірки.

3 Значення коефіцієнта кореляції знаходиться в межах від -1 до 1 . Якщо $r \rightarrow 0$, то лінійний стохастичний зв'язок між ознаками відсутній, якщо $|r| \rightarrow 1$, то спостерігається лінійний стохастичний зв'язок між ознаками, причому якщо $r \rightarrow 1$, то зв'язок є прямим, а якщо $r \rightarrow -1$, то зворотним.

4 Розрахуємо t – статистику за формулою

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

Розрахункове значення t – статистики за модулем порівнюється з табличним значенням $t_{q,n-2}$ (додаток Б), де q – рівень значущості, а $(n-2)$ – ступені свободи. Якщо $|t| > t_{q,n-2}$, то на рівні значущості q нульова гіпотеза відкидається і приймається альтернативна гіпотеза про те, що коефіцієнт кореляції значно відрізняється від нуля і між факторами існує лінійний зв'язок. Якщо $|t| < t_{q,n-2}$, то з довірчою ймовірністю α приймається нульова гіпотеза про те, що коефіцієнт кореляції незначно відрізняється від нуля, тому лінійного зв'язку між параметрами моделі не існує.

Приклад

Задана вибірка з генеральної сукупності витрат обігу (тис. грн.)

$$X = (2,7; 3; 2,8; 2,9; 2,6; 2,5; 2,8; 2,6; 2,5)$$

і відповідна до неї вибірка з генеральної сукупності вантажообігу (тис. т)

$$Y = (15,6; 15,3; 15,6; 15,2; 15,9; 16,1; 15,5; 16; 16,2).$$

З'ясувати, чи існує зв'язок між даними показниками, якщо існує, то наскільки він статистично значущий із довірчою ймовірністю 95%, який характер він має.

Розв'язання:

1 Зведемо вихідні дані до кореляційної таблиці і здійснимо відповідні розрахунки (табл. 2.2).

Таблиця 2.2 – Кореляційна таблиця до прикладу 2

n	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	2,7	15,6	42,12	7,29	243,36
2	3	15,3	45,9	9	234,09
3	2,8	15,6	43,68	7,84	243,36
4	2,9	15,2	44,08	8,41	231,04
5	2,6	15,9	41,34	6,76	252,81
6	2,5	16,1	40,25	6,25	259,21
7	2,8	15,5	43,4	7,84	240,25
8	2,6	16	41,6	6,76	256
9	2,5	16,2	40,5	6,25	262,44
$\sum_{i=1}^9$	24,4	141,4	382,87	66,4	2222,56

2 Розрахуємо коефіцієнт кореляції:

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

$$= \frac{9 \cdot 382,87 - 24,4 \cdot 141,4}{\sqrt{[9 \cdot 66,4 - 24,4^2] \cdot [9 \cdot 2222,56 - 141,4^2]}} = -0,96011.$$

Коефіцієнт кореляції між параметрами моделі $\bar{r} = -0,96$, тобто досить близький до -1 . Можна висунути гіпотезу, що існує зворотний зв'язок.

Для перевірки статистичної значущості коефіцієнта кореляції обчислимо значення t -статистики:

$$t = \frac{\bar{r} \sqrt{n-2}}{\sqrt{1-\bar{r}^2}} = \frac{-0,96 \sqrt{9-2}}{\sqrt{1-(0,96)^2}} = -9,08.$$

Для $q = 0,05$ і $n - 2 = 7$ знайдемо за таблицями розподілу Стюдента $t_{0,05} = 2,36$ або скористаємося вбудованою функцією Excel СТЬЮДРАСПОБР (рис. 2.1).

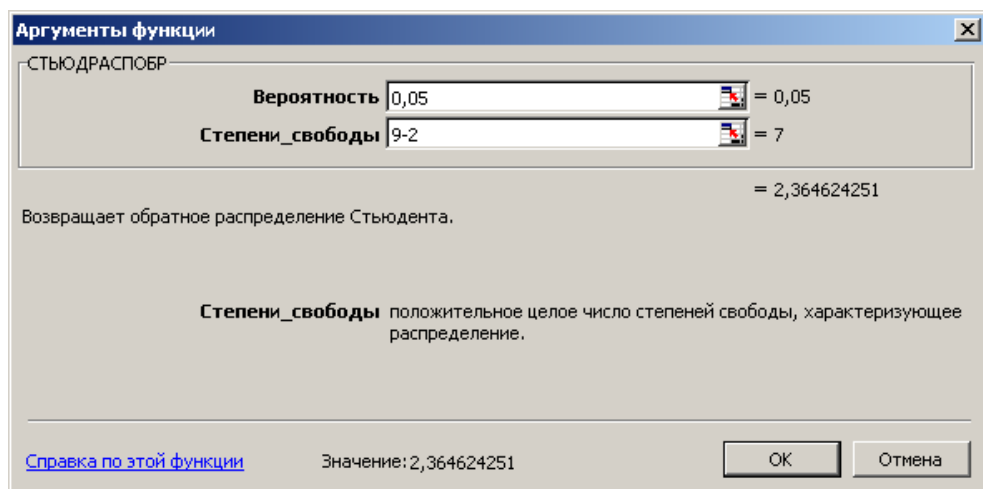


Рисунок 2.1 – Структура вбудованої функції Excel СТЬЮДРАСПОБР

Порівняємо

$$|t_p| > t_{q,n-2}$$

$$|-9,08| > 2,36$$

На рівні значущості 0,05 коефіцієнт кореляції значно відрізняється від нуля, тобто лінійний зв'язок між досліджуваними ознаками існує. Оскільки коефіцієнт кореляції є від'ємним, зв'язок між признаками є зворотним, тобто зі зростанням вантажообігу падають витрати обігу і, навпаки.

Вихідні дані для самостійного виконання лабораторної роботи №2 подано у додатку. Номер варіанту обирається за номером студента у журналі.

ДОДАТОК

Вихідні дані для лабораторних робіт №2

Варіант 1		
n	X _i	Y _i
1	1,8	36,1
2	2,4	38,3
3	2,5	30,6
4	2,3	32,1
5	2,3	37,6
6	2,5	34,8
7	2,4	34,2
8	2,5	34,2
9	2,1	32,5

Варіант 2		
n	X _i	Y _i
1	2	38,2
2	2,1	36,9
3	2,3	39,7
4	2,6	37,2
5	2,8	31,7
6	2,8	30,1
7	2,2	39,9
8	2,3	38,8
9	2,6	38,4

Варіант 3		
n	X _i	Y _i
1	3	29,4
2	2,6	35,4
3	2,3	39,7
4	2,5	37,1
5	2,2	35,7
6	2,4	40,2
7	2,2	39,4
8	2,6	43,7
9	2,6	38,4

Варіант 4		
n	X _i	Y _i
1	2,8	14
2	2,4	17,1
3	2,3	18,2
4	2,5	17,4
5	2,7	16,1
6	2,4	18,8
7	2,3	32,2
8	1,9	31
9	2,3	32,4

Варіант 5		
n	X _i	Y _i
1	2	35
2	2,3	43,7
3	2,7	31,9
4	2,2	37,3
5	2,4	40,9
6	2,3	38,8
7	2,3	35,7
8	2,6	43,2
9	2,7	30,5

Варіант 6		
n	X _i	Y _i
1	2,4	40,2
2	2,2	39,4
3	2,6	43,7
4	2,6	38,4
5	2,3	38,8
6	2,2	39,9
7	2,8	30,1
8	2,8	31,7
9	2,8	37,2

Варіант 7		
n	X _i	Y _i
1	2,3	32,1
2	1,9	31
3	2,3	32,4
4	2,5	33,2
5	2,6	31,2
6	2	34,8
7	1,9	35,4
8	2,4	33
9	2,2	34,8

Варіант 8		
n	X _i	Y _i
1	2,8	13,8
2	2,7	14,8
3	2,4	16,9
4	2,3	16,8
5	2,5	14,8
6	2,5	17,9
7	2,5	17,6
8	2,4	15,7
9	2,3	15,2

Варіант 9		
n	X _i	Y _i
1	2,7	14,9
2	2,5	16,1
3	2,1	19,7
4	2,8	14
5	2,4	17,1
6	2,3	18,2
7	2,5	17,4
8	2,7	16,1
9	2,4	18

Варіант 10		
n	X _i	Y _i
1	30,2	5000
2	32	5200
3	32	5350
4	37	5880
5	30	5430
6	30	5430
7	30	5350
8	29	5740
9	33	5570

Варіант 11		
n	X _i	Y _i
1	29	5350
2	33	2740
3	31	5570
4	30	5530
5	34	6020
6	38	7010
7	31	6420
8	39	7150
9	39,5	7190

Варіант 12		
n	X _i	Y _i
1	5,1	15,7
2	7,2	17
3	2	5,2
4	5,1	17,5
5	10	22,1
6	12	25,8
7	9,4	23
8	6,9	17
9	3,4	9,1

Варіант 13		
n	X _i	Y _i
1	5,4	3,7
2	7,6	8
3	2,3	3,2
4	5,9	2,5
5	11	4,9
6	12,6	4,8
7	10,4	5,1
8	4,9	2,2
9	2,4	1,1

Варіант 14		
n	X _i	Y _i
1	1,4	12,7
2	2,6	18
3	2,3	16,2
4	5,1	25,5
5	6	24,1
6	5,6	24,8
7	10,4	55
8	4,9	33
9	2,4	18,1

Варіант 15		
n	X _i	Y _i
1	0,4	13,7
2	0,6	18
3	0,3	6,2
4	0,9	15,5
5	1	24,1
6	1,6	24,8
7	1,4	25
8	0,6	13
9	0,3	8,1

Варіант 16		
n	X _i	Y _i
1	2,4	3,7
2	7,6	8,2
3	2,3	3,2
4	4,9	2,5
5	10,7	4,9
6	12,6	7,8
7	10,4	5,1
8	4,9	2,2
9	2,4	1,1

Варіант 17		
n	X _i	Y _i
1	1,4	12,7
2	3,6	18
3	2,3	16,2
4	4,1	25,5
5	6	24,1
6	5,6	23,8
7	11,4	55
8	4,9	33
9	2,4	18,1

Варіант 18		
n	X _i	Y _i
1	1,4	13,7
2	0,6	18
3	0,3	6,2
4	0,9	15,5
5	1	22,1
6	1,6	24,8
7	1,4	25
8	0,6	13
9	0,3	8,1

Варіант 19		
n	X _i	Y _i
1	35,2	5000
2	32	5200
3	33	5350
4	37	5880
5	30	5431
6	34	5437
7	30	5359
8	29	5740
9	33	5570

Варіант 20		
n	X _i	Y _i
1	29	1350
2	33	2740
3	31	5570
4	30	5530
5	34	6020
6	35	7010
7	31	4420
8	39	7150
9	39,5	7190

Варіант 21		
n	X _i	Y _i
1	3,1	15,7
2	7,2	25
3	2	5,2
4	5,1	17,5
5	10	22,1
6	11	25,8
7	9,4	23
8	6,9	17
9	3,4	9,1