

**Тема:** Елементи кореляційного аналізу з використанням MS Excel

**Мета роботи:** Освоїти принцип обчислення коефіцієнта кореляції за допомогою вбудованих функцій та аналізу даних в Excel

### **Основні поняття**

Кореляційною називають таку статистичну залежність, при якій якщо змінюється одна з величин, то змінюється середнє значення іншої.

Лінія регресії показує як залежить середнє значення одного емпіричного ряду від значень другого ряду. Якщо побудувати точки  $(x_i, \bar{y}_{x_i})$  в Декартовій системі координат, то за характером розміщення цих точок можна зробити припущення про форму лінії регресії та відповідну форму кореляційного зв'язку – лінійну, квадратичну, логарифмічну тощо.

*Коефіцієнт кореляції (Пірсона)*

Для оцінювання тісноти кореляційного зв'язку між випадковими величинами  $X$  та  $Y$  використовують коефіцієнт кореляції, який визначається за формулою:

$$r = \frac{M(XY) - M(X) \cdot M(Y)}{\delta(X) \cdot \delta(Y)} \quad (6.1)$$

Коефіцієнт кореляції, який знаходиться за даним рівнянням називається *емпіричним коефіцієнтом кореляції (коефіцієнтом кореляції Пірсона)*.

*Властивості коефіцієнту кореляції:*

1. Значення коефіцієнта кореляції змінюється в межах від  $-1$  до  $+1$ .

2. Якщо  $r=0$ , то між ознаками, що вивчаються немає лінійної кореляційної залежності, але ця умова не виключає існування якого-небудь іншого виду кореляційного зв'язку (параболічного, поліноміального і т.п.)

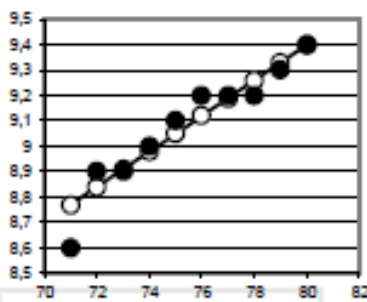
3. Чим більше  $r$ , тим тісніший зв'язок (сильніша спряженість) між ознаками, які вивчаються.



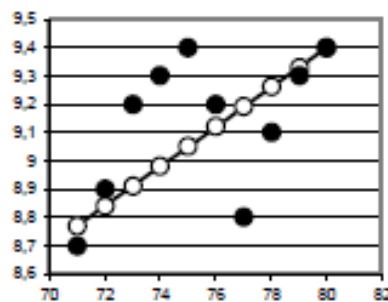
4. Якщо  $r=1$ , то кореляційна залежність стає лінійною функціональною залежністю.

Залежно від значення коефіцієнта кореляції, кореляційні зв'язки поділяють на:

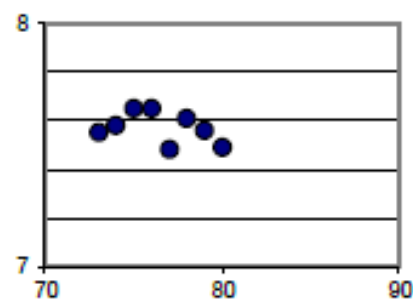
- сильну (тісну) кореляцію -  $0,7 < r < 1$ ;
- середню кореляцію -  $0,5 < r < 0,7$ ;
- помірну кореляцію -  $0,3 < r < 0,5$ ;
- слабку кореляцію -  $0,2 < r < 0,3$ ;
- відсутність кореляції -  $r < 0,19$ .



*тісна кореляція*



*слабка кореляція*



*відсуття кореляція*

Рис. 6.1. Приклади зв'язків у кореляційних залежностях

### *Побудова прямих регресії*

Кореляційна залежність між  $X$  та  $Y$  називається лінійною, якщо обидві функції регресії є лінійними. В такому випадку обидві лінії регресії є прямими, їх називають прямими регресії. Рівняння прямих регресії мають вигляд:

$$y = \rho \left( \frac{X}{Y} \right) \left[ -M(X) \right] + M(Y) - \text{рівняння прямої регресії } X \text{ на } Y$$

$$y = \rho \left( \frac{X}{Y} \right) \left[ -M(Y) \right] + M(X) - \text{рівняння прямої регресії } Y \text{ на } X$$

Кутовий коефіцієнт прямої регресії  $X$  на  $Y$

$$\rho \left( \frac{X}{Y} \right) = \frac{M(XY) - M(X)M(Y)}{D(Y)}$$

Кутовий коефіцієнт прямої регресії  $Y$  на  $X$

$$\rho \left( \frac{Y}{X} \right) = \frac{M(XY) - M(X)M(Y)}{D(X)}$$



Якщо при збільшенні однієї величини збільшується інша це вказує на додатній зв'язок між цими величинами, тобто має пряму кореляційну залежність. І навпаки, якщо збільшення однієї змінної супроводжується зменшенням значень іншої, це вказує на від'ємний зв'язок, тобто обернену кореляційну залежність.

Для прямої кореляції  $\rho\left(\frac{Y}{X}\right) > 0$ , для оберненої -  $\rho\left(\frac{Y}{X}\right) < 0$ .

На практиці для знаходження кутових коефіцієнтів використовують метод найменших квадратів.

В результаті рівняння прямих регресії приймуть вигляд:

$$y - M(Y) = \frac{M(XY) - M(X)M(Y)}{S(X)S(Y)} \cdot \frac{S(Y)}{S(X)} (x - M(X))$$

або

$$y - \bar{y} = \frac{\overline{xy} - \bar{x}\bar{y}}{S^2(X)} \cdot (x - \bar{x})$$

і відповідно

$$x - \bar{x} = \frac{\overline{xy} - \bar{x}\bar{y}}{S^2(Y)} \cdot (y - \bar{y})$$

Перевірка значущості вибіркового коефіцієнту кореляції

На практиці для оцінки тісноти кореляційного зв'язку використовують вибіркового коефіцієнту кореляції:

$$r = \frac{M(XY) - M(X)M(Y)}{S^2(X)S^2(Y)}$$

Після його обчислення перевіряють гіпотезу про значущість вибіркового коефіцієнту кореляції. З цією метою обчислюють експериментальне значення критерію  $t$ , який має розподіл Стюдента:

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

де  $n$  – об'єм вибірки. Потім задавшись рівнем значущості  $p$ , за таблицею критичних точок розподілу Стюдента знаходять критичне значення  $t_{кр}(p; n-2)$  для достовірної критичної області.





Якщо  $|t| > t_{кр}$ , то вважають коефіцієнт кореляції значущим, якщо ж  $|t| < t_{кр}$  - то незначущим.

Відповідно до рівня значущості розглядається часткова класифікація кореляційних зв'язків:

1) висока значима кореляція – при  $r$ , який відповідає рівню статистичної значущості  $p \leq 0,01$ ;

2) значима кореляція - при  $r$ , який відповідає рівню статистичної значущості  $p \leq 0,05$ ;

3) тенденція достовірного зв'язку - при  $r$ , який відповідає рівню статистичної значущості  $p \leq 0,1$ ;

4) незначима кореляція - при  $r$ , який не досягає рівня статистичної значущості.

При роботі з числовим редактором *Excel* коефіцієнт кореляції розраховують за допомогою вбудованої функції КОРРЕЛ (категорія - статистичні).

КОРРЕЛ

Массив1  = массив

Массив2  = массив

Возвращает коэффициент корреляции между двумя множествами данных.

Массив1 первый диапазон значений. Значениями могут быть числа, имена, массивы или ссылки с именами.

Значение:

OK Отмена

### Синтаксис

#### КОРРЕЛ (массив1; массив2)

Массив1 - комірка інтервалу значень.

Массив2 - другий інтервал комірок зі значенням.

**Примітка:** Аргументи повинні бути числами або іменами, масивами або посиланнями, які містять числа.

### Приклад

КОРРЕЛ({3;2;4;5;6};{9;7;12;15;17}) дорівнює 0,997054

Для роботи з інструментами аналізу дані слід представити у вигляді рядків або стовбців листа Excel. Сукупність комірок, які містять дані, що аналізуються, називається вхідним діапазоном.



Якщо команда **Аналіз даних** відсутня в меню **Сервіс**, у *Microsoft Excel* необхідно встановити статистичний пакет аналізу даних.

Щоб встановити пакет аналізу даних

1. У меню **Сервіс** виберіть команду **Надбудови**.

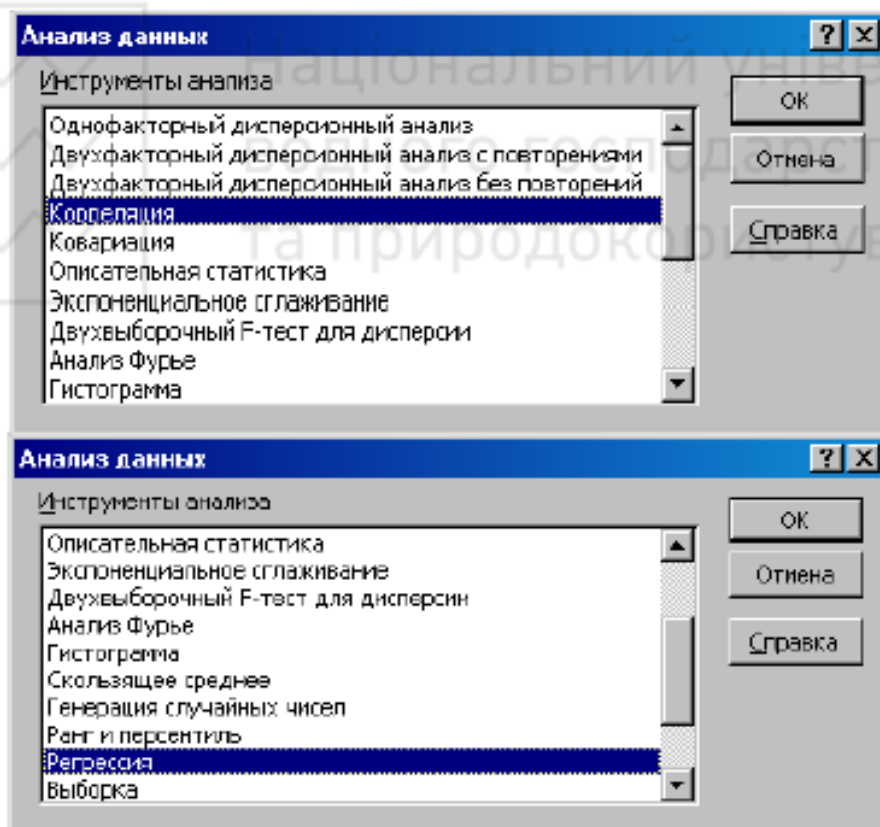
Якщо список у діалоговому вікні **Надбудови** не містить **пакету аналізу даних**, натисніть кнопку **Огляд** і вкажіть диск, папку та ім'я файлу для цієї надстройки, *Analys32.xll* (як правило, папка *Microsoft Office\Office\Library\Analysis*) або запусить програму *Setup*, щоб встановити цю надбудову.

2. Встановіть прапорець **Пакет аналізу**.

Щоб запусити пакет аналізу:

1. У меню **Сервіс** виберіть команду **Аналіз даних**.

2. У списку **Інструменти аналізу** виберіть потрібний рядок.



3. Введіть вхідний та вихідний діапазони, потім оберіть необхідні параметри.

### Завдання

Згідно варіанту завдань побудувати графік та встановити кореляційні залежності для масивів даних. Оцінити тісноту зв'язку за коефіцієнтом Пірсона.

Варіанти завдань до практичної роботи №6

номер варіанту																			
1		2		3		4		5		6		7		8		9		10	
ряд 1	ряд 2	ряд 1	ряд 2	ряд 1	ряд 2	ряд 1	ряд 2	ряд 1	ряд 2	ряд 1	ряд 2	ряд 1	ряд 2	ряд 1	ряд 2	ряд 1	ряд 2	ряд 1	ряд 2
8,5	43,0	22,8	4,5	18,0	4,2	14,2	3,9	11,2	3,6	6,3	2,7	34,7	7,1	39,8	12,8	48,2	23,4	58,3	2,5
10,4	24,2	22,9	4,6	18,1	4,3	14,3	4,0	11,3	3,7	7,2	2,8	34,9	7,3	42,2	13,2	49,2	24,0	59,5	3,6
8,9	33,5	23,0	5,1	18,2	4,8	14,4	4,5	11,3	4,2	6,3	3,1	35,0	8,1	42,4	14,8	51,3	26,9	62,0	8,9
6,8	52,9	23,2	5,1	18,3	4,8	14,5	4,5	16,4	5,8	6,4	3,5	35,3	8,8	42,7	16,0	51,7	29,1	62,6	2,9
7,8	49,8	23,4	5,4	18,5	5,0	14,3	4,7	11,3	4,4	6,3	3,2	34,9	8,5	42,2	11,2	51,1	20,4	61,8	7,0
6,8	47,5	23,6	5,9	12,2	5,5	14,7	5,1	11,6	4,8	6,5	3,6	35,9	9,4	43,5	17,0	52,6	30,9	63,7	6,2
6,1	51,9	24,2	6,2	19,1	5,8	15,1	5,4	22,3	5,1	6,6	4,4	36,9	9,9	44,6	17,9	54,0	32,6	65,3	9,3
7,1	48,7	24,7	6,2	19,5	5,8	15,4	5,5	12,2	5,1	6,8	3,8	37,6	9,9	45,5	18,0	55,1	32,8	66,6	9,6
9,4	50,2	24,9	6,5	19,7	6,1	15,5	5,7	12,3	5,3	6,8	3,9	37,9	10,3	52,0	18,7	62,9	34,1	76,1	1,9
8,8	47,3	25,1	6,8	19,8	6,3	15,7	5,9	12,4	5,5	6,9	4,1	38,2	9,9	46,2	18,0	56,0	32,7	67,7	9,5
6,6	68,0	25,7	7,0	20,3	6,6	16,0	6,1	12,7	5,7	7,1	4,3	39,1	11,1	47,4	17,8	57,3	32,4	69,3	8,8
10,4	32,3	26,4	7,4	20,9	6,9	16,5	7,2	4,4	0,2	7,2	5,0	40,2	13,0	48,6	23,6	58,9	43,0	71,2	8,1
7,9	33,5	26,9	7,5	21,3	7,0	16,8	6,6	13,3	6,1	7,0	4,6	41,0	11,9	51,1	21,7	61,8	39,5	74,8	1,8
9,56	51,3	27,0	7,6	21,3	4,9	16,9	6,7	13,3	6,2	7,4	4,6	41,1	12,1	49,8	22,1	60,2	40,1	72,8	2,9
7,8	49,8	27,5	7,8	21,7	7,3	17,2	6,8	13,6	6,4	7,6	4,7	41,9	12,4	50,7	22,5	61,3	40,8	74,2	4,3
6,9	45,2	27,8	7,9	22,0	7,4	21,2	6,9	16,7	6,5	9,3	4,8	51,7	11,7	62,6	21,3	75,7	38,7	91,6	0,3
6,3	49,6	28,0	8,1	15,4	7,6	17,5	7,1	13,8	6,6	7,7	4,9	42,6	12,8	44,5	23,3	61,1	42,4	73,9	7,1
7,1	48,6	28,1	8,1	22,2	7,6	17,5	4,4	13,9	4,1	7,7	3,1	42,8	8,0	51,8	14,5	62,6	26,4	75,8	8,1
11,7	51,1	28,6	8,4	22,6	7,8	17,8	7,3	14,1	6,8	7,9	3,2	43,6	13,3	52,7	24,1	63,8	43,8	77,2	9,7
7,6	47,3	30,0	8,5	22,8	8,4	18,7	7,4	14,8	6,9	8,2	5,1	45,7	14,5	55,3	19,1	66,9	34,7	80,9	2,1
7,9	7,1	31,2	8,5	22,9	8,1	19,5	7,4	15,4	6,9	8,6	5,1	47,5	13,5	57,5	24,5	69,6	44,6	84,2	3,0



водного господарства  
та природокористування

**Контрольні запитання:**

1. Що виражає кореляційна залежність?
2. Назвіть основні властивості коефіцієнта кореляції.
3. Як ранжується тіснота кореляційних зв'язків?
4. За яких умов коефіцієнт кореляції вважається значущим?
5. За допомогою якої вбудованої функції програмного пакету MS Excel визначається коефіцієнт кореляції?