

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВАСИЛЯ СТУСА

*Г. В. Тарадіна, О. І. Доценко, А. М. Міщенко*

**СТАТИСТИЧНІ МЕТОДИ В БІОЛОГІЇ ТА ЕКОЛОГІЇ:  
ТЕОРІЯ І ПРАКТИКА**

Навчальний посібник

Вінниця  
«ТВОРИ»  
2022

УДК 57.087.1:574[:519.22/.25](075.8)

Т 19

Рекомендовано до друку Вченою радою ДонНУ імені Василя Стуса  
(протокол № 9 від 28.01.2022)

Рецензенти:

- Д. Х. Штофель* – кандидат технічних наук, доцент, доцент кафедри біомедичної інженерії факультету інфокомунікацій, радіоелектроніки та наносистем Вінницького національного технічного університету
- О. В. Іващук* – кандидат біологічних наук, доцент, доцент кафедри біологічної фізики, медичної апаратури та інформатики Вінницького національного медичного університету ім. М. І. Пирогова

**Т 19**      Статистичні методи в біології та екології: теорія і практика: навч. посібник / Г. В. Тарадіна, О. І. Доценко, А. М. Міщенко. Вінниця : ТВОРИ, 2022. 180 с.

ISBN 978-617-552-094-9

Навчальний посібник містить теоретичні відомості про основні статистичні методи, які найбільш часто використовуються в прикладних дослідженнях у галузі біології та екології, та шляхи їх реалізації засобами Microsoft Excel та Statistica. Теоретичний матеріал супроводжується великою кількістю професійно спрямованих прикладів і завдань для самостійної роботи.

Навчальний посібник розрахований на здобувачів вищої освіти спеціальностей 091 Біологія та 101 Екологія і використовується під час викладання дисциплін «Математичні методи в біології», «Інформаційні технології в екології», «Моделювання та прогнозування стану довкілля».

© Тарадіна Г. В., 2022

© Доценко О. І., 2022

© Міщенко А. М., 2022

© ДонНУ імені Василя Стуса, 2022

ISBN 978-617-552-094-9

**ЗМІСТ**

Вступ .....	4
<b>Тема 1.</b> Способи угруповання первинних даних. Варіаційний ряд.....	5
<b>Тема 2.</b> Характеристики випадкової величини .....	18
<b>Тема 3.</b> Перевірка даних на відповідність нормальному розподілу. Перевірка даних на відповідність нормальному розподілу в програмі Excel.....	25
<b>Тема 4.</b> Порівняння середніх показників вибірок. Параметричні критерії перевірки гіпотез про середні та дисперсії. Критерій Фішера. Перевірка гіпотези про рівність дисперсій.....	29
<b>Тема 5.</b> Порівняння середніх показників вибірок. Непараметричні критерії ....	41
<b>Тема 6.</b> Перевірка наявності зв'язку між змінними. Кореляційний аналіз .....	60
<b>Тема 7.</b> Аналіз таблиць спряженості .....	72
<b>Тема 8.</b> Дисперсійний аналіз .....	77
<b>Тема 9.</b> Регресійний аналіз (множинна лінійна регресія).....	92
<b>Тема 10.</b> Кластерний аналіз .....	102
<b>Тема 11.</b> Дискримінантний аналіз.....	113
Завдання для самостійної роботи .....	119
Список рекомендованої літератури.....	179

## ВСТУП

Широке використання статистичних методів у процесі аналізу даних біологічного експерименту та екологічного моніторингу вимагає знання основ математичної статистики, підходів до обробки матеріалу, вміння правильно формулювати конкретну задачу аналізу даних і вибирати відповідні статистичні процедури для її вирішення, а також правильно інтерпретувати отримані результати.

Метою посібника є виклад основних статистичних методів для вирішення типових задач аналізу даних у біологічному та екологічному експериментальному дослідженні та їх реалізація в програмних пакетах Excel 2016 і Statistica 10; формування у здобувачів вищої освіти базових фундаментальних знань з основ аналізу, обробки результатів експерименту, створення математичних моделей для опису і прогнозування різних явищ, їх якісного і кількісного аналізу, з комп'ютерних методів обробки даних біологічних досліджень та екологічного моніторингу.

Посібник містить велику кількість практичних завдань із біології та екології і докладні приклади їх розв'язання з використанням інформаційних технологій.

Навчальний посібник розрахований на здобувачів вищої освіти зі спеціальностей 091 Біологія та 101 Екологія; матеріали посібника використовуються під час викладання дисциплін «Математичні методи в біології», «Інформаційні технології в екології», «Моделювання та прогнозування стану довкілля», а також під час виконання кваліфікаційних робіт – для представлення, обробки, аналізу даних експериментальних досліджень.

## Тема 1. Способи угруповання первинних даних. Варіаційний ряд

Зафіксовані в документах обліку відомості про досліджуваний об'єкт (або об'єкти) є тим первинним фактичним матеріалом, який потребує відповідної обробки. У більшості випадків ці масиви даних біологічних та екологічних досліджень мають великий обсяг. Для організації та узагальнення великих масивів даних використовують таблиці частот, що робить процес обробки даних досить зручним та ефективним.

Варіаційним рядом, або *рядом розподілу*, називають подвійний ряд чисел, що показує, скільки разів певна ознака повторюється в даній статистичній сукупності. Наприклад, досліджували кількість стегнових пір у живородних ящірок. Випадково було відібрано 30 тварин з кількістю пір: 8 9 11 10 11 10 8 9 9 7 9 9 7 11 10 10 11 11 9 9 9 10 9 11 10 10 10 12 8.

Підрахуємо, скільки разів (*число варіантів,  $f_i$* ) в числовій послідовності зустрічаються певні значення (*варіанти,  $x_i$* ) досліджуваного параметру, розташували їх у порядку зростання (або зменшення). Наприклад, кількість стегнових пір 7 мали 2 тварини; кількість пір 8 – 3 ящірки і т. д. У результаті маємо такий варіаційний ряд:

Варіанти, $x_i$	7	8	9	10	11	12
Число варіантів, $f_i$	2	3	10	8	6	1

Число варіантів називають *частотами* або *вагами* варіант. Загальна сума частот варіаційного ряду дорівнює обсягу даної сукупності, в нашому випадку 30.

У наведеному прикладі число варіантів підраховують безпосередньо для ранжированих значень ознаки. Такий варіаційний ряд називають *безінтервальним*.

У випадку, коли підраховують частоти, які належать до інтервалів, на які розбивається діапазон від найменшого до найбільшого значень ознаки, варіаційний ряд називають *інтервальним*. Ці інтервали називають *класовими інтервалами*. Класові інтервали можуть бути однаковими за шириною, тоді варіаційний ряд називають *рівноінтервальним*, або різними – варіаційний ряд називають *нерівноінтервальним*. Для даних біологічних досліджень та екологічного моніторингу здебільшого використовують рівноінтервальні ряди, що полегшує виявлення закономірності варіювання і зіставлення рядів розподілу один з одним.

### Техніка побудови варіаційних рядів

1. Приступаючи до побудови варіаційного ряду, потрібно у вихідних даних відшукати мінімальне  $x_{min}$  і максимальне  $x_{max}$  значення варіант. Величина  $x_{max} - x_{min}$  називається розмахом варіювання ознаки

2. Визначити кількість інтервалів  $K$ . Якщо в сукупності менше 100 спостережень ( $n$ ), то кількість інтервалів розраховується за формулою  $K = 1 + 3,32 \times \lg(n)$ , в інших випадках  $K = 5 \times \lg(n)$ .

3. Розрахувати величину інтервалу  $\lambda$  за формулою:

$$\lambda = \frac{x_{\max} - x_{\min}}{K}.$$

Якщо  $\lambda = 1$ , зібраний матеріал розподіляється в безінтервальний варіаційний ряд; якщо  $\lambda \neq 1$ , вихідні дані необхідно розподіляти в інтервальний ряд. Значення величини інтервалу округлюється до точності, прийнятій при вимірюванні ознаки.

Програмні засоби для статистичної обробки даних зазвичай мають за замовчуванням алгоритми для вибору границь інтервалів, але часто виникає необхідність визначити інтервали самостійно, щоб вибрати ті, що є репрезентативними для ваших даних. Наприклад, розглянемо гістограми розподілу частот для одних і тих же вихідних даних із різною величиною класового інтервалу (рис. 1.1).

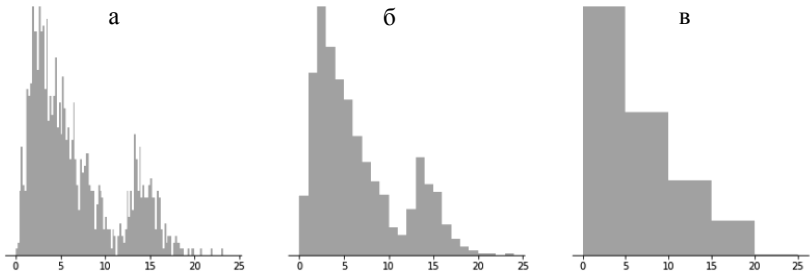


Рис. 1.1. Приклад гістограми розподілу частот з величиною класового інтервалу: а – 0,1; б – 1; в – 5

Розмір інтервалів має зворотну залежність від їхньої кількості. Чим більші розміри інтервалу, тим менша їх кількість буде для покриття всього діапазону даних. Варто витратити деякий час на тестування різних розмірів класових інтервалів, щоб побачити, як виглядає розподіл у кожному з них, а потім вибрати графік, який найкраще відображає дані. Якщо інтервалів занадто багато, то розподіл даних виглядатиме грубим, буде важко відрізнити похибки даних від основної тенденції; з іншого боку, при занадто малій кількості інтервалів на гістограмі не вистачає деталей, необхідних для визначення моделі або правильної інтерпретації даних.

4. При побудові інтервального варіаційного ряду мінімальна варіанта сукупності повинна потрапляти приблизно в середину першого класового інтервалу. Визначити нижню межу першого класового інтервалу можна за допомогою формули:

$$x_i = x_{\min} - \frac{\lambda}{2},$$

де  $x_i$  – нижня межа першого класового інтервалу;

$x_{\min}$  – мінімальна варіанта досліджуваної сукупності;

$\lambda$  – величина класового інтервалу.

5. Після визначення величини і границь інтервалів розраховують число варіантів, які потрапили до кожного класу.

Коли значення варіанти знаходиться на межі двох інтервалів, воно додається до інтервалу праворуч або ліворуч. Яка сторона буде обрана, залежить від програми візуалізації, що використовується для обробки даних; наприклад, у програмі Excel за замовчуванням значення на межі додається до інтервалу праворуч.

### **Двовимірний візуальний аналіз даних. Гістограми**

Двовимірний (2М) візуальний аналіз – це графічне представлення експериментальних даних на площині. У ньому використовуються різноманітні гістограми, діаграми розсіювання, імовірнісні, лінійні і послідовні графіки (графіки послідовних значень), діаграми діапазонів, розмахів, кругові і стовпчасті діаграми тощо, що дозволяє побачити специфіку даних.

Термін «гістограма» ввів Карл Пірсон у 1895 році. Гістограми дають змогу побачити, як розподілені значення змінних за інтервалами угруповань, тобто як часто змінні приймають значення з різних інтервалів.

Особливо корисний цей графік для великої кількості спостережень, наприклад, понад 100.

Гістограма допомагає візуально перевірити виконання певних припущень про вид розподілу спостережень – чи є розподіл симетричним; чи має діапазон спостережуваних значень одну моду або кілька, тобто є мультимодальним. Останнє може вказувати на неоднорідність даних, їх приналежність до різних генеральних сукупностей.

*Примітка: вибірка – частина всієї сукупності експериментальних даних, які підлягають дослідженню за певними ознаками; генеральна сукупність – усі досліджувані об'єкти, частиною яких є вибірка.*

Гістограму використовують для зображення інтервального ряду: по осі абсцис відкладають інтервали і на цих відрізках будують прямокутники, висота

яких дорівнює частоті (кількості реалізацій) або частотам (частка реалізацій від загальної кількості) відповідного інтервалу (рис. 1.2).

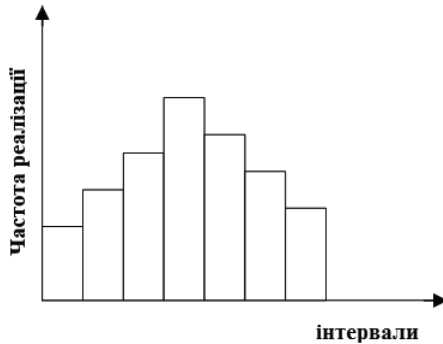


Рис. 1.2. Гістограма

Під час побудови гістограм важливим є вибір інтервалів, оскільки форма розподілу буде залежати від цього вибору.

Існують формули розрахунку розміру інтервалу (див. вище), але зазвичай вважається, що інтервалів має бути 12–15, водночас у кожен інтервал має потрапити не менше 5–6 реалізацій величини (інтервали можуть бути різної величини).

### Полігон частот

Полігон частот є аналогом гістограми для дискретного розподілу. По осі абсцис відкладаються значення величин, а по осі ординат – частоти. Отримані точки з'єднуються ламаною лінією (рис. 1.3). Полігон може бути побудований для безперервного розподілу, тоді ламаною з'єднуються середини верхніх відрізків прямокутників гістограми.

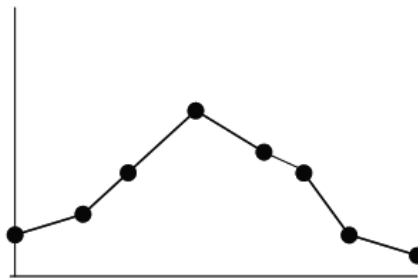


Рис. 1.3. Полігон частот

### Кумулятивна крива

Під час побудови кумулятивної кривої по осі абсцис відкладаються інтервали, а по осі ординат – число або частка елементів сукупності, що менше або дорівнює заданому числу значень (рис. 1.4).

При збільшенні до нескінченності розміру вибірки гістограма перетворюється у графік щільності розподілу, а кумулятивна крива – у графік функції розподілу.

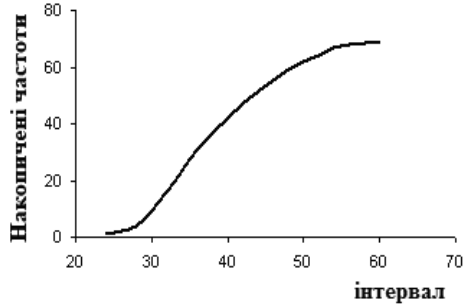


Рис. 1.4. Кумулятивна крива

### Побудова варіаційного ряду і гістограми в Excel

Припустимо, ми маємо показники мертвонароджених у 75 містах, за якими необхідно побудувати гістограму (рис. 1.5).

	A	B	C	D	E	F	G	H	I
1	Показники мертвонароджених в 75 містах					Границі інтервалів	Частота	Накопичення частоти	
2	27	36	34	46	43	26	=ЧАСТОТА(A2:E16;F2:F10)		
3	28	29	37	40	43	31			
4	40	33	50	37	41	36			
5	32	27	43	34	32	41			
6	30	41	54	42	47	46			
7	35	49	49	54	36	51			
8	36	51	36	24	35	56			
9	25	33	38	38	36	61			
10	29	51	32	36	53	66			
11	30	55	44	46	38				
12	29	44	48	30	34				
13	46	47	36	37	36				
14	30	58	42	46	46				
15	29	38	44	40	30				
16	35	35	63	47	37				
17	Макс		63						
18	Мін		24						
19	n		75						
20	λ		5,397772						
21	x <sub>n</sub>		21						

Рис. 1.5. Приклад розрахунку частот для побудови гістограми

1. Знаходимо межі інтервалів. Максимальне значення масиву даних визначаємо за допомогою формули: =MAX(A2:E16), мінімальне значення: =MIN(A2:E16), кількість даних у масиві (n) – за допомогою формули: =COUNT(A2:E16).

2. Знаходимо величину класового інтервалу  $\lambda$ , використовуючи формулу 
$$\lambda = \frac{x_{\max} - x_{\min}}{1 + 3,321g \cdot n}$$
. Оскільки експериментальні дані являють собою цілі числа, округляємо отримане  $\lambda$  до цілого значення. У нашому випадку  $\lambda = 5$ .

3. Знаходимо нижню межу першого класового інтервалу і також округляємо до цілого значення. Отже, отримуємо перший інтервал 21–26.

4. Верхні межі інтервалів – 26, 31, 36, 41, 46, 51, 56, 61, 66 поміщаємо у вільний стовпець (F). У нашому прикладі значеннями меж інтервалів заповнені комірки F2:F10.

5. Виділяємо комірки G2:G10. У верхній комірці стовпця (G2) набираємо формулу: =FREQUENCY(A2:E16; F2:F10). Перший параметр описує поля, в яких знаходяться вихідні дані, другий – поля, в яких записані верхні межі інтервалів. Для введення використовуємо комбінацію клавіш Ctrl-Shift-Enter (оскільки ми маємо справу з формулою масиву). У результаті в комірках G2:G10 повинні роз'являтися частоти, відповідні інтервалам, що знаходяться в діапазоні F2:F10.

6. Розраховуємо накопичені частоти. Для цього в комірку H2 вводимо: = G2, в комірку H3 = H2 + G3. Копіюємо формулу з комірки H3 в комірки H4–H10.

7. Будемо за совпцями G і H графічне зображення гістограми і кумулятивної кривої. Для цього в Excel виконуємо такі операції:

**Крок 1.** Виділяємо дані в стовпцях G, H (рис. 1.6) і натискаємо на кнопку *Вставити стовпчасту діаграму або гістограму* на вкладці *Діаграми* пункту меню *Вставлення*. З переліку типів гістограм вибираємо тип *Плоска стовпчаста діаграма*.

	A	B	C	D	E	F	G	H
1	Показники метвонародженних в 75 містах					Границі інтервалів	Частота	Накопичення частоти
2	27	36	34	46	43	26	2	2
3	28	29	37	40	43	31	12	14
4	40	32	50	37	41	36	20	34
5	32	27	43	34	32	41	13	47
6	30	41	54	42	47	46	13	60
7	35	49	49	54	36	51	9	69
8	36	51	36	24	35	56	4	73
9	25	33	38	38	36	61	1	74
10	29	51	32	36	53	66	1	75
11	30	55	44	46	38			

Рис. 1.6. Побудова гістограми в програмі Excel (Крок 1).

**Крок 2.** На отриманій гістограмі змінюємо тип діаграми для другого ряду. Для цього потрібно правою кнопкою миші натиснути на будь-якому стовпці ряду 2, у контекстному меню вибрати команду *Змінити тип діаграми для рядів*; у діалоговому вікні *Змінення типу діаграми* вибрати тип діаграми *Комбінована*; в розділі *Виберіть тип діаграми та осі для рядів даних* поміняти тип діаграми для ряду 2 на *Лінійчатую діаграму з маркерами*, поставити прапорець у полі *Допоміжна ось* і натиснути кнопку ОК – у правій частині діаграми з'явиться додаткова вертикальна вісь (рис. 1.7).

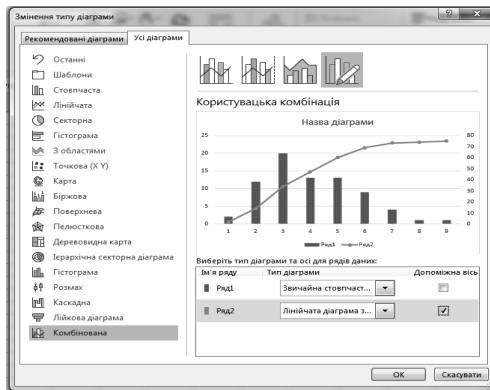


Рис. 1.7. Діалогове вікно зміни типу діаграми

**Крок 3.** Зазначаємо, що підписи осі X знаходяться в комірках F2:F10. Для цього натискаємо правою кнопкою миші в будь-якому місці діаграми, з контекстного меню вибираємо команду *Вибрати дані*. В діалоговому вікні *Вибір джерела даних*, в полі *Підписи горизонтальної осі (категорії)* натискаємо кнопку *Змінити* і на робочому полі відзначаємо потрібні комірки. Натискаємо кнопку ОК діалогового вікна *Підписи осі* для повернення в попереднє діалогове вікно (рис. 1.8).

**Крок 4.** Даємо назви рядам: Ряд 1 – Гістограма, Ряд 2 – Кумулятивна крива. Для цього в діалоговому вікні *Вибір джерела даних* в полі *Записи легенди (ряди)* вибираємо Ряд 1 і натискаємо кнопку *Редагувати*. У діалоговому вікні *Редагування ряду* в полі *Ім'я ряду* вводимо текст «Гістограма» та натискаємо ОК для повернення в попереднє діалогове вікно. Аналогічно змінюємо назву другого ряду. Натискаємо кнопку ОК діалогового вікна *Вибір джерела даних*.

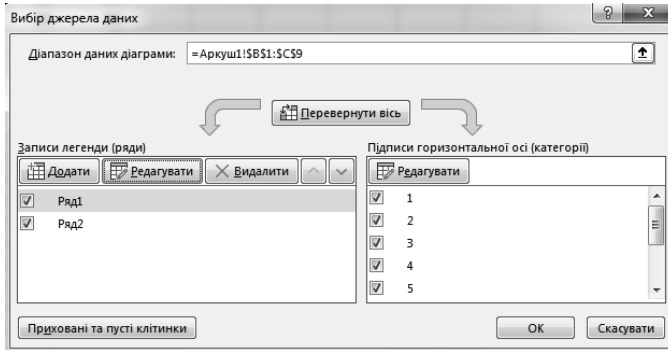


Рис. 1.8. Діалогове вікно вибору вихідних даних для гістограми

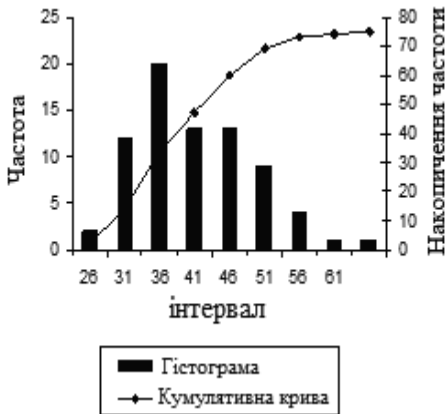


Рис. 1.9. Гістограма і кумулятивна крива

**Крок 5.** Підписуємо горизонтальні та вертикальні осі. Для цього в пункті меню *Знаряддя для діаграм / Конструктор* на вкладці *Додати елемент діаграми* вибираємо пункт *Назви осей*. З випадаючого списку вибираємо команду *Назва основної горизонтальної осі* і вводимо текст *Інтервал* (набраний текст відображається при цьому в рядку формул). Натискаємо клавішу *Enter*. Аналогічно підписуємо основну і допоміжну вертикальні осі. Діаграма повинна мати вигляд, як на рис. 1.9.

Можлива побудова гістограми за допомогою функції, що є в пакеті *Data Analysis (Аналіз даних)*, який знаходиться в пункті меню *Дані*, вкладка *Analysis (Аналіз)*. Якщо це налаштування відсутнє за вказаним шляхом, його можна встановити. Для цього натисніть пункт меню *Файл*, а потім – *Параметри*. В діалоговому вікні виберіть пункт *Надбудови*, в розділі *Керування* натисніть кнопку *Перейти*. У полі *Наявні надбудови* встановіть прапоречок для надбудови *Пакет аналізу* та натисніть кнопку *ОК*.

У діалоговому вікні *Data Analysis (Аналіз даних)* виберіть пункт *Histogram (Гістограма)*. Після цього з'являється вікно, в якому необхідно задати початкові дані для побудови діаграми.

**Параметри діалогового вікна Histogram (Гістограма):**

- *Input Range (Вхідний діапазон)*. Задається посилання на діапазон комірок, в яких знаходяться вихідні дані.
- *Bin Range (Інтервал кишень)*. Вводиться діапазон клітинок із граничними значеннями, які визначають відрізки інтервалів і повинні бути введені у зростаючому порядку.
- *Labels (Мітки)*. Ставиться відмітка, якщо перший рядок або стовпець вхідного інтервалу містить заголовки.
- *Output Range (Вихідний діапазон)*. Вводиться посилання на верхню ліву комірку вихідного діапазону.
- *New Worksheet Ply (Новий робочий аркуш)*. Встановлюється перемикач, щоб вставити результати аналізу на новий аркуш книги з комірки A1.
- *New Workbook (Нова робоча книга)* – вибирається в разі, якщо результати необхідно розмістити в нову робочу книгу.
- *Pareto (sorted histogram) (Парето (відсортована гістограма))*. У вигляді діаграми відображаються дані за спаданням частоти.
- *Cumulative Percentage (Кумулятивний відсоток)*. Розраховуються значення, будується гістограма і графік накопичених частот (рис. 1.10).
- *Chart Output (Вивід діаграми)*. Відображається гістограма.

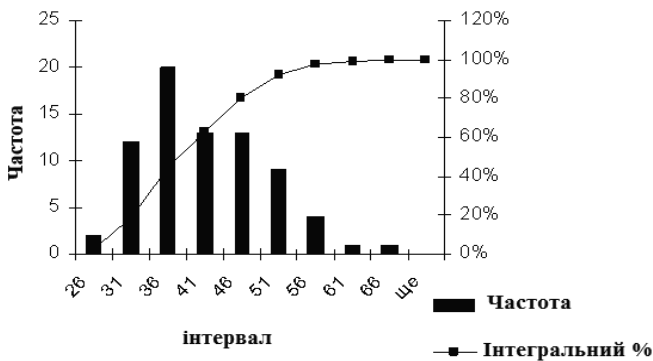


Рис. 1.10. Гістограма з інтегральним відсотком

**Побудова таблиці частот і гістограм у програмі Statistica**

Виконаємо побудову таблиці частот і гістограми для даних завдання, що наведено вище.

1. Завантажуємо програму Statistica. У перший стовпець таблиці переносимо дані попередньої задачі, використовуючи буфер обміну.

2. Вибираємо послідовно: пункт меню: *Statistic*, кнопка *Basic Statistics*, у діалоговому вікні *Basic Statistics and Tables (Основна статистика і таблиці)* обираємо пункт *Frequency tables (Таблиці частот)*. Натискаємо *OK*.

3. У діалоговому вікні *Frequency tables* натискаємо кнопку *Variables (Змінні)*, відмічаємо *Var1* (стовпець, в якому знаходяться дані), натискаємо *OK*.

4. Переходимо на вкладку *Advanced (розширений аналіз)*, встановлюємо параметри розбиття на інтервали. Для цього потрібно активувати пункт *Step Size (Розмір інтервалу)*, клацнувши по ньому мишею і ввести розмір інтервалу розбивки. Знімаємо «галочку» з пункту *at minimum* (мінімум) і у вікні *starting at* (починаючи з) вводимо нижню межу першого інтервалу розбивки.

Для даних прикладу «Кількість мертвонароджених у 75 містах» вікно *Frequency tables* повинно мати вигляд, як на рис. 1.11. Для цієї задачі розмір інтервалу розбиття буде дорівнювати 5, а нижня межа першого інтервалу розбиття становитиме 21.

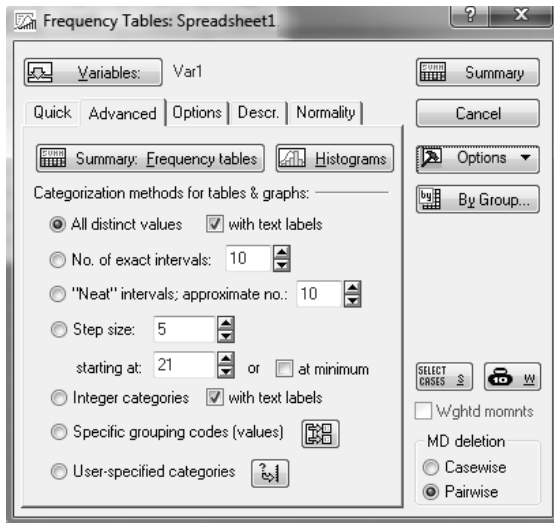


Рис. 1.11. Вид діалогового вікна *Frequency tables*

5. Переходимо на вкладку *Advanced*. Виводимо на екран таблицю щільності розподілу (*Summary: frequency table*) і гістограму (*Histograms*), натиснувши відповідні кнопки вікна *Frequency tables* (Таблиці щільності розподілу).

## Порівняння вибірок за допомогою гістограм розподілу в програмі Statistica

**Приклад 1.1.** Миші піддавалися впливу п'ятиденної низькочастотної вібрації з частотою 24 Гц. Встановити вплив вібрації на діаметр (мкм) еритроцитів.

Діаметр еритроцитів мишей, що не піддавалися вібрації: 5,00 5,00 6,36 10,00 6,36 10,00 5,00 5,45 5,45 5,91 6,82 5,45 4,55 6,36 8,18 6,36 6,82 5,91 5,45 5,00 4,55 7,27 5,91 3,18 8,64 6,36 6,36 5,45 7,73 5,91 8,18 5,91 6,82 6,82 5,45 6,82 5,91 7,73 5,00 4,09 5,00 5,91 7,73 6,36 6,36 5,00 7,27 6,82 5,91 5,45 6,82 5,00 5,00 7,27 6,36 5,45 5,45 5,00 5,91 8,18 6,82 4,55 5,45 5,91 7,73 6,36 5,00 6,82 2,73 5,91 4,55 7,27 5,00 5,45 5,45 6,36 5,45 5,00 5,91 5,45 5,00 5,91 6,36 5,45 6,36 4,55 5,00 6,36 5,91 5,45 4,55 5,91 5,45 5,91 8,18 6,36 4,55 7,27 6,82

Діаметр еритроцитів мишей, що піддавалися вібрації протягом трьох днів: 4,55 5,00 3,64 5,00 4,55 6,82 6,36 4,09 6,82 4,55 6,36 5,91 4,55 4,55 5,00 6,36 5,91 5,45 7,27 6,82 5,00 5,45 4,55 5,00 5,45 6,82 6,36 5,00 4,55 5,45 6,82 5,91 7,27 6,82 5,00 5,45 4,55 5,00 5,45 4,55 5,00 2,73 5,45 4,09 4,55 5,91 4,55 5,00 5,45 4,55 6,36 5,00 5,91 5,00 5,45 4,55 3,64 5,00 6,82 6,82 5,91 6,36 6,82 4,09 5,45 5,00 4,55 5,45 5,00 5,91 6,36 7,27 5,45 4,55 3,64 6,82 5,91 4,55 6,36 3,64 4,55 3,18 5,00 4,55 7,27 5,00 5,45 4,55 5,00 4,55 5,45 3,64 5,45 4,09 5,45 4,55 5,00 4,55 5,91

Діаметр еритроцитів мишей, що піддавалися вібрації протягом п'яти днів: 6,36 6,82 4,55 9,09 5,00 4,55 5,91 4,55 5,00 5,91 4,55 5,45 4,55 5,00 5,45 4,55 5,91 5,45 4,55 5,45 4,55 5,45 4,55 5,00 5,45 5,91 5,45 4,55 3,64 4,09 4,55 6,36 4,55 4,09 4,55 6,36 4,55 7,27 4,55 5,45 4,55 4,55 5,45 7,27 4,55 7,73 4,55 5,45 5,45 7,73 4,55 6,82 5,45 3,18 5,45 4,55 5,00 5,91 8,18 7,73 6,36 4,55 5,00 5,45 6,82 5,45 5,91 6,82 4,55 5,00 6,36 4,55 5,91 6,36 5,45 4,55 5,00 5,45 6,36 6,82 5,00 5,45 6,82 4,09 3,18 6,82 4,55 4,55 5,00 5,45 5,91 5,00 4,55 5,45 5,91 4,55 6,36 4,55 5,45

1. Відкриємо новий робочий аркуш і введемо дані в три перших стовпчики.

2. Перейменуємо заголовки стовпців. Для цього потрібно двічі клацнути лівою кнопкою миші по назві стовпця і в діалоговому вікні у полі *Name* замість *Var1* ввести слово «контроль», замість *Var2* ввести «третій день» тощо.

3. Виберемо послідовно пункти меню: *Statistic, Basic statistics / Tables, Frequency tables (Статистика, Основна статистика / таблиці, Таблиці частот)*. Натиснемо *OK*, після чого має відкритися вікно *Frequency tables* (Таблиці щільності розподілу).

4. Натиснемо кнопку *Variables*, відзначимо заголовки трьох перших стовпчиків з даними.

5. Перейдемо на вкладку *Advanced*. Встановимо параметри розбиття на інтервали: інтервал розбиття – 1, починати побудову (нижня межа першого інтервалу розбивки) – 2.

6. Виведемо на екран таблиці щільності розподілу для аналізованих даних і оцінимо характер впливу вібрації на діаметр еритроцитів мишей.

7. Покажемо вплив вібрації на діаметр еритроцитів за допомогою діаграм. Для цього виберемо послідовно такі пункти меню: *Graphs, Histograms, 2D Histograms (Графіки, Гістограми, Двовимірні гістограми)*. На вкладці *Quick*

(Швидкий аналіз) зробимо наступні налаштування: натиснемо кнопку *Variables* і відзначимо заголовки перших трьох стовпців даних, у вікні Тип графіка встановимо – *Multiple (Множинний)*, у вікні *Intervals (Інтервали)* – цілочисельний режим. Натиснемо *OK*.

8. Побудована діаграма повинна мати вигляд, як на рис. 1.12.

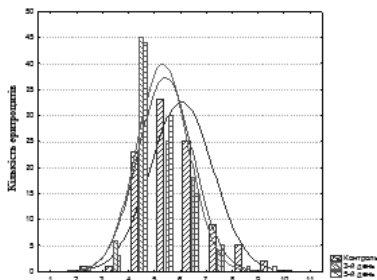


Рис. 1.12. Характер зміни діаметру еритроцитів у процесі вібрації з частотою 24 Гц

9. Для зміни назви осі ординат двічі клацнемо по назві осі і у вікні введемо потрібну назву.

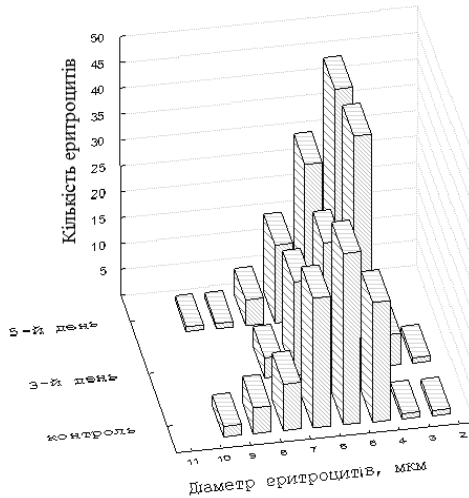
10. Аналіз рис. 1.12 показує, що вібрація з частотою 24 Гц призводить до зміщення величини діаметра основної фракції еритроцитів у бік менших значень.

11. Покажемо вплив вібрації на діаметр еритроцитів за допомогою об'ємної діаграми.

12. Скопіюємо дані стовпця «Контроль» у стовпець Var5. У першу комірку стовпця Var4 введемо слово «контроль» і скопіюємо його в усі інші 99 комірок. Потім скопіюємо дані стовпця «третій день» і помістимо їх у стовпець Var5 під даними «Контроль». Навпроти даних для третього дня вібрації в стовпці Var4 пишемо «третій день». Цю процедуру виконаємо з даними стовпчика «п'ятий день».

13. Відновимо вікно *Frequency tables* (воно згорнуте у вигляді відповідного прямокутника внизу робочого поля вікна програми Statistica). Натиснемо кнопку *Variables (Змінні)*, зазначимо заголовки стовпців Var4 і Var5.

14. Перейдемо на вкладку *Descr. (Onuc)* і виберемо останній пункт меню *3D histograms, bivariate distributions (Об'ємні гістограми, двомірний розподіл)*. Отримана діаграма повинна мати вигляд, як на рис. 1.13. Підписи на осях підправимо, як описано в пункті 9.



*Рис. 1.13. Об'ємна діаграма, що показує зміну розмірів еритроцитів під впливом вібрації*

### **Завдання для самостійної роботи**

1. Побудувати варіаційний ряд, гістограми розподілу і кумулятивні криві для даних з розділу «Варіанти завдань для закріплення тем 1–3» у програмах Excel і Statistica.
2. Проаналізувати, чи існують відмінності у даних, наведених у розділі «Варіанти завдань для закріплення теми 1, підрозділу «Порівняння вибірок за допомогою гістограми розподілу», використовуючи метод побудови гістограм розподілу.

### **Контрольні запитання до теми 1**

1. Що таке варіаційний ряд? Які типи варіаційних рядів Ви знаєте?
2. Як визначити величину класового інтервалу?
3. Для чого використовуються гістограма, полігон частот, кумулятивна крива?

## Тема 2. Характеристики випадкової величини

У статистиці під час аналізу даних стикаються з такими проблемами:

- 1) кількість даних, які необхідно вибрати, і спосіб їх відбору;
- 2) правомірність поширення висновків, зроблених на підставі вибіркового даних, на всю генеральну сукупність;
- 3) вибір оптимальних способів оцінювання;
- 4) вибір способів узагальнення, класифікація та подання даних.

### Властивості оцінок параметрів

Оцінки параметрів повинні відповідати таким основним вимогам:

1. *Незміщеність*. При проведенні великої кількості випробувань із вибірками однакового розміру середнє значення кожної вибірки прагне до істинного значення генеральної сукупності. Зміщеність зазвичай обумовлена наявністю систематичної помилки.

2. *Спроможність*. З ростом розміру вибірки оцінка повинна прагнути до значення відповідного параметра генеральної сукупності з імовірністю, що прямує до одиниці.

3. *Ефективність*. Обрана оцінка для вибірки рівного об'єму повинна мати мінімальну дисперсію.

4. *Достатність*. Оцінка параметрів вибірки повинна містити всю необхідну інформацію, яка відображається на генеральну сукупність, і не вимагати доповнень.

Під час розробки оцінок зазвичай висувають деякі передумови. Тому оцінки здебільшого відповідають наведеним вимогам тільки при виконанні цих передумов. Про це необхідно пам'ятати і при використанні оцінок.

Для оцінювання параметрів використовуються різні методи, особливе місце серед яких належить методу максимальної правдоподібності. Він застосовується в тих випадках, коли відомий закон розподілу даних. Суть його в тому, що оцінки повинні бути рівні значенням, за яких вибірка має максимальну ймовірність появи.

До характеристик одновимірного розподілу належать:

- 1) *міри положення* (середнє, медіана, мода та ін.);
- 2) *міри розсіювання* (розмах, коефіцієнт варіації, дисперсія, середньо-квадратичне відхилення);
- 3) *міри форми* (асиметрія, ексцес, моменти третього і четвертого порядку).

### Середнє арифметичне (вибіркове)

$$\bar{X} = \frac{\sum_{i=1}^N x_i}{N} ,$$

де  $x_i$  – значення досліджуваного параметру,  
 $N$  – загальна кількість спостережень.

Якщо ряд спостережень складається з  $k$  груп, то середнє арифметичне всього ряду дорівнює зваженому груповому середньому, вагами при цьому є обсяги груп.

$$\bar{X} = \frac{\sum_{i=1}^k X_i n_i}{\sum_{i=1}^k n_i},$$

де  $n_i$  – розмір  $i$ -ї групи,  
 $X_i$  – середнє  $i$ -ї групи.

### Середнє геометричне

Геометричне середнє застосовується, якщо

- значення параметру змінюється в часі з постійним співвідношенням між його вимірами:

$$\frac{X_{i-1}}{X_i} = \frac{X_i}{X_{i+1}} = const$$

(наприклад, збільшення числа бактерій, зростання капіталу на рахунку, експлуатаційні витрати тощо);

- окремі значення у вибірці знаходяться далеко один від одного (наприклад, відрізняються на порядок).

Геометричне середнє розраховують за формулою:

$$\bar{X}_G = \sqrt[n]{\prod_{i=1}^n X_i}.$$

### Середнє гармонічне

У ряді випадків (розрахунок середньої тривалості життя, визначення середньої швидкості) використовується середнє гармонічне:

$$\bar{X}_H = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}}.$$

### Мода

Це значення з вибірки експериментальних даних, яке спостерігається найбільше число разів.

Для інтервального варіаційного ряду мода розраховується за формулою:

$$Mo = X_{Mo} + \frac{h(m_{Mo} - m_{Mo-1})}{2m_{Mo} - m_{Mo+1} - m_{Mo-1}},$$

де  $X_{Mo}$  – початок класового інтервалу, якому відповідає найбільша частота;

$h$  – величина модального інтервалу;

$m_{Mo}$  – частота модального інтервалу;

$m_{Mo-1}$  – частота класового інтервалу, який розташований перед модальним;

$m_{Mo+1}$  – частота інтервалу, наступного за модальним.

Мода не може бути застосована, якщо розподіл мультимодальний.

### Медіана (вибіркова)

Це значення, яке ділить варіаційний ряд, значення якого впорядковані за збільшенням або за зменшенням на дві рівні за обсягом групи.

Для ряду з непарною кількістю членів медіаною є елемент з номером  $(N+1)/2$ . Для ряду з парною кількістю членів медіана дорівнює середньому значенню членів ряду з номерами  $\frac{N}{2}$  і  $\frac{N}{2} + 1$ . Наприклад, для ряду 4; 5; 6,7; 8; 12 медіаною буде значення 6,7.

Для інтервального ряду медіана обчислюється за формулою:

$$Me = X_{Me} + \frac{h(\sum m_x - m_x^{\max})}{m_m},$$

де  $X_{Me}$  – початок медіанного інтервалу;

$h$  – величина медіанного інтервалу;

$m_x$  – частоти за всіма інтервалами;

$m_x^{\max}$  – частота, накопичена до початку медіанного інтервалу;

$m_m$  – частота медіанного інтервалу.

Медіанним називається інтервал, в якому знаходиться значення медіани.

**Приклад 2.1.** Обчислити середнє і медіану.

У таблиці наведені ціни на певний препарат у різних організаціях.

Організація	1	2	3	4	5	6	7	8
Ціна	100	110	115	125	140	145	145	150

Середнє значення ціни  $(100 + 110 + 115 + 125 + 140 + 145 + 145 + 150) / 8 = 128,75$ .

Ранжируваний ряд  $\{100;110;115;125;140;145;145;150\}$

Медіана ціни  $(125 + 140) / 2 = 132,5$  (число членів ряду парне).

**Приклад 2.2.** Розрахувати моду, середнє і медіану віку хворих.

Вік	21–30	31–40	41–50	51–60	61–70
Кількість хворих	45	36	175	361	825
Накопичення частоти	45	81	256	617	1442

Середнє значення віку хворих

$$X_{\text{середнє}} = \frac{(24,5 \times 45 + 34,5 \times 36 + 44,5 \times 175 + 54,5 \times 361 + 64,5 \times 825)}{(45 + 36 + 175 + 361 + 825)} = 58,57$$

Медіана

Початок інтервалу, який містить медіану,  $X_{Me} = 60$ . Оскільки сума частот 1442, то половина (медіана ділить навпіл) – 721. До початку інтервалу 60–69 накопичено 617, тому зрозуміло, що медіана знаходиться саме в цьому інтервалі.

Ширина медіанного інтервалу  $h = 69 - 59 = 10$ ;

$m_x^{\max}$  – частота, накопичена до початку медіанного інтервалу – 617;

$$\frac{\sum m_x}{2} = 1442/2 - \text{загальна накопичена частота ділиться на 2};$$

$m_m$  – частота медіанного інтервалу – 825.

$$Me = 60 + \frac{10 \times (\frac{1442}{2} - 617)}{825} = 61,26.$$

Мода

Початок модального інтервалу  $X_{Mo} = 60$ . Модальним є той інтервал, в якому частота найбільша.

Ширина модального інтервалу  $h = 69 - 59 = 10$ .

Частота модального інтервалу  $m_{Mo} = 825$ .

Частота інтервалу, що передєє модальному  $m_{Mo-1} = 361$ , а частота інтервалу, наступного за модальним,  $m_{Mo+1} = 0$ , оскільки модальний інтервал є останнім.

$$M_o = 60 + \frac{10 \times (825 - 361)}{2 \times 825 - 0 - 361} = 65.$$

### Показники варіації. Варіаційний розмах

$$R_B = X_{\max} - X_{\min}$$

Цей показник ненадійний, оскільки на нього впливають крайні значення, він не змінюється при змінах варіаційного ряду, що не зачіпають крайні значення.

### Емпірична дисперсія

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{N - 1}.$$

$\sqrt{S^2} = S$  – середньоквадратичне відхилення (в англійській літературі – стандартна помилка середнього).

### Коефіцієнт варіації (відносне стандартне відхилення)

$$V = \frac{S}{\bar{X}} \times 100 \%$$

Це статистична міра відносної дисперсії даних навколо свого середнього значення. Коефіцієнт варіації застосовується при порівнянні мінливості ознак об'єкта, виражених у різних одиницях вимірювання, і розраховується у відсотках. Мінливість вважається значною, якщо  $V > 30 \%$ . Якщо коефіцієнт варіації понад 100 %, то дані неоднорідні.

### Довірчий інтервал

*Довірчий інтервал* – це такий інтервал, в якому з наперед заданою ймовірністю  $p = 1 - \alpha$  знаходяться показники генеральної сукупності. Тут  $1 - \alpha$  – довірна ймовірність,  $\alpha$  – рівень значущості.

#### *Властивості довірчого інтервалу*

Під час збільшення числа вимірювань точність підвищується. Це справедливо тільки в тому випадку, якщо немає систематичних помилок і спостереження незалежні.

Збільшення довірчої ймовірності при фіксованій вибірці веде до збільшення довірчого інтервалу і зниження точності.

Якщо збільшувати кількість вимірювань, то оцінка параметру стає більш точною, і довірчий інтервал зменшується. Це не стосується тих ситуацій, коли вимірювання залежні або в них присутні систематичні помилки. У цих випадках точність при збільшенні кількості вимірювань не тільки не збільшується, але навіть може знижуватися.

Довірчий інтервал означає не ймовірність попадання значення оцінюваного параметра в інтервал певних меж, а те, що якщо взяти достатню кількість вибірок, то в  $(100-P)$  % випадків параметр буде перебувати в заданому інтервалі, і тільки у  $(100-\alpha)$  % випадках значення досліджуваного параметра можуть вийти за межі довірчого інтервалу.

Рівень значущості вибирається в інтервалі від 0,01 до 0,05. Водночас 0,05 – звичайні вимоги надійності, 0,01 – підвищені, 0,001 – дуже високі, а 0,1 – знижені.

*Довірчий інтервал для середнього* визначають за формулою

$$\left[ \bar{X} - t_{n,p} \frac{S}{\sqrt{v}}, \bar{X} + t_{n,p} \frac{S}{\sqrt{v}} \right],$$

де  $S$  – середньоквадратичне відхилення;

$n$  – число дослідів;

$t_{n,p}$  – табличне значення розподілу Стюдента з числом ступенів вільності  $v$  і довірчою ймовірністю  $p$  (пояснення далі в тексті).

### **Визначення характеристик випадкової величини за допомогою програми Excel**

В Excel для визначення характеристик випадкової величини використовують такі функції:

AVERAGE(перелік комірок або інтервал імен комірок) – середнє значення;

MEDIAN(перелік комірок або інтервал імен комірок) – медіана;

MODE.SNGL(перелік комірок або інтервал імен комірок) – мода;

STDEV.S(перелік комірок або інтервал імен комірок) – стандартне відхилення;

VAR.S(перелік комірок або інтервал імен комірок) – дисперсія;

CONFIDENCE.T(альфа; стандартне відхилення; число дослідів) – довірчий інтервал.

### **Описова статистика в програмі Statistica**

Щоб визначити середню моду, медіану та інші параметри за допомогою програми Statistica, потрібно скористатися пунктом *Statistic, Basic statistics / Tables (Статистика, Основна статистика / Таблиці)*. У вікні скористатися пунктом *Descriptive statistics (Описова статистика)*, де слід натиснути кнопку

*Variables* і вказати, для яких стовпців даних будуть обчислюватися параметри описової статистики. Далі у вікні *Basic statistics / Tables* на вкладці *Advanced* зазначити, клацаючи мишею по маленькому віконцю, які параметри описової статистики будуть обчислюватися: *Mean* (Середнє), *Mode* (Мода), *Median* (Медіана), *Geom. Mean* (Геом. середнє), *Harmonic mean* (Гармонійне середнє), *Standard deviation* (Стандартне відхилення) (рис. 2.1). Після цього натиснути кнопку *Summary* (Підсумок).

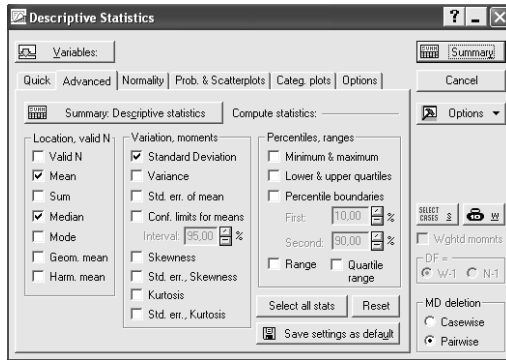


Рис. 2.1. Діалогове вікно завдання характеристик описової статистики

### Завдання для самостійної роботи

1. Для даних, наведених у розділі «Варіанти завдань для закріплення тем 1–3», розрахувати середнє, моду, медіану, дисперсію, стандартне відхилення, довірчий інтервал, використовуючи стандартні функції програми Excel.
2. Для даних, наведених у розділі «Варіанти завдань для закріплення тем 1–3», побудувати варіаційний ряд і розрахувати середнє, моду, медіану для даних, представлених у вигляді інтервального ряду.
3. Для даних, наведених у розділі «Варіанти завдань для закріплення теми 1, підрозділу «Порівняння вибірок за допомогою гістограми розподілу», розрахувати параметри описової статистики для кількох вибірок у програмі Statistica.

### Контрольні запитання до теми 2

1. Які властивості оцінок параметрів Ви знаєте?
2. Назвіть, які статистичні показники належать до характеристик одновимірного розподілу.
3. Перерахуйте відомі Вам показники варіації.
4. Назвіть властивості довірчого інтервалу. Як розраховується довірчий інтервал для середнього?

### Тема 3. Перевірка даних на відповідність нормальному розподілу. Перевірка даних на відповідність нормальному розподілу в програмі Excel.

Однією з важливих характеристик вибірки є характер розподілу даних всередині неї. Розподіл даних залежить від того, скільки разів у вибірці зустрічається те або інше значення випадкової величини.

Випадкова величина розподілена за нормальним законом, якщо виконуються умови, які є наслідком нормального закону розподілу:

- 1) майже всі (99,7 %) відхилення від середнього менше  $3S$  ( $\varepsilon_i < 3S$ );
- 2) дві третини (68,3 %) відхилень менше ніж  $S$ ;
- 3) половина відхилень менше ніж  $0,625S$ .

Якщо хоча б одна з цих умов не виконується, можна вважати, що гіпотеза про нормальний розподіл суперечить наявним даним.

Від того, який розподіл мають дані, залежить подальше визначення способів їх обробки. Якщо дані нормально розподілені, то для їх аналізу застосовують параметричні методи, в інших випадках (незалежно від типу розподілу, наприклад, рівномірний, біноміальний, Пуассона тощо) використовують непараметричні методи аналізу.

**Приклад 3.1.** Подано показники мертвнонароджених у 75 містax. Визначити, чи розподілені ці дані за нормальним законом (рис. 3.1)

	A	B	C	D	E	F	G	H	I	J
1	Масив даних					Обчислення відхилень від середнього				
2	27	36	34	46	43	=ABS(\$E\$17-A2)	5.23	6.77	3.77	
3	28	29	37	40	43	ABS(число)	10.23	2.23	0.77	3.77
4	40	33	30	37	41	0.77	6.23	10.77	2.23	1.77
5	32	27	43	34	32	7.23	12.23	3.77	5.23	7.23
6	30	41	54	42	47	9.23	1.77	14.77	2.77	7.77
7	35	49	49	54	36	4.23	9.77	9.77	14.77	3.23
8	36	51	36	24	35	3.23	11.77	3.23	15.23	4.23
9	25	33	38	38	36	14.23	6.23	1.23	1.23	3.23
10	29	51	32	36	53	10.23	11.77	7.23	3.23	13.77
11	30	55	44	46	38	9.23	15.77	4.77	6.77	1.23
12	29	44	48	30	34	10.23	4.77	8.77	9.23	5.23
13	46	47	36	37	36	6.77	7.77	3.23	2.23	3.23
14	30	58	42	46	46	9.23	18.77	2.77	6.77	6.77
15	29	38	44	40	30	10.23	1.23	4.77	0.77	9.23
16	35	35	63	47	37	4.23	4.23	23.77	7.77	2.23
17	Середнє	39.22667								
18	S	8.362049								
19	3S	25.14616								
20	0.625S	5.23878								
21	Кільк даних	75								

Рис. 3.1. Дані для прикладу 3.1 і його розв'язання

AVERAGE(перелік комірок або інтервал імен комірок) – середнє значення;  
 MEDIAN(перелік комірок або інтервал імен комірок) – медіана;  
 MODE.SNGL(перелік комірок або інтервал імен комірок) – мода;  
 STDEV.S(перелік комірок або інтервал імен комірок) – стандартне відхилення.

Для знаходження середнього використано функцію AVERAGE(A2:E16); для визначення середньоквадратичного відхилення – STDEV.S(A2:E16); для підрахунку кількості даних – COUNT(A2:E16).

Заповнюємо підсумкову таблицю:

Умова	Необхідно	Вибірка
Майже всі (99,7 %) відхилення від середнього менше 3S	75	= {COUNTIF(F2:J16; «<25,146»)} (Ctrl–Shift–Enter) 75
Дві третини (68,3 %) відхилень менше S	= (2/3) × 75 (50)	= {COUNTIF(F2:J16; «<8,38»)} (Ctrl–Shift–Enter) 49
Половина відхилень менше ніж 0,625S	= (1/2) × 75 (37,5)	= {COUNTIF(F2:J16; «<5,239»)} (Ctrl–Shift–Enter) 36

Бачимо, що друга і третя умови не виконуються. Ці вибірки не відповідають нормальному закону розподілу.

### ***Перевірка даних на відповідність нормальному розподілу в програмі Statistica***

*Перевірка нормальності за допомогою оцінок коефіцієнтів асиметрії та ексцесу*

*Асиметрія, або коефіцієнт асиметрії, є мірою несиметричності розподілу. Якщо цей коефіцієнт значно відрізняється від 0, розподіл є асиметричним. Формально маємо:*

$$g_1 = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left[ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{\frac{3}{2}}}$$

*Ексцес, або коефіцієнт ексцесу, вимірює гостроту піку розподілу. Оцінка ексцесу, або вибіркового ексцесу, обчислюється за формулою*

$$b_2 = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{\left[ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2} - 3$$

де  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ .

Асиметрія та ексцес корисні для перевірки нормальності даних. Нормальний розподіл симетричний, отже, коефіцієнт асиметрії дорівнює 0. Ексцес нормального

розподілу також дорівнює 0, тому за відхиленнями вибіркового ексцесу й асиметрії від 0 можна говорити про близькість розподілу спостережуваної змінної до нормального. Відомо, що розподіл із більш гострою вершиною, ніж нормальний, має позитивний ексцес, а з більш закругленою – негативний.

**Приклад 3.2.** Ввести дані прикладу 3.1 в один стовпець програми Statistica. Перевірку на нормальний закон розподілу можна виконати кількома способами.

**1. Оцінка коефіцієнтів асиметрії та ексцесу.** У діалоговому вікні *Descriptive statistics (Описова статистика)* натиснемо кнопку *Variables (Змінні)* і виберемо змінну, яка вказує на стовпець, в якому знаходяться дані. Перейдемо на сторінку *Advanced (Розширений аналіз)*. Дамо вказівку системі, що потрібно обчислити асиметрію (skewness) й ексцес (kurtosis), а також їхні стандартні помилки (stand. errors). Натиснемо кнопку *Summary (Підсумок)*, і таблиця з результатами з'явиться на екрані (рис. 3.2).

Basic statistics / Tables Spreadsheet 1						
	Valid N	Mean	Skewness	Std. error Skewness	Kurtosis	Std. error Kurtosis
Var5	75	39,22667	0,500943	0,277400	-0,193523	0,548211

Рис. 3.2. Результати розрахунків коефіцієнтів асиметрії та ексцесу

Із результатів розрахунків бачимо, що абсолютна величина оцінок асиметрії й ексцесу та їхні помилки мають один той же порядок і близькі до нуля. Отже, можна вважати, що дані узгоджені з гіпотезою нормальності.

**2. Аналіз нормальних імовірнісних графіків.** У вікні *Descriptive statistics (Описова статистика)* перейдемо на вкладку *Prob.&Scatterplots (Імовірнісні графіки і діаграми розсіювання)*. Натиснемо кнопку *Normal probability plot (Графіки нормального розподілу)* (рис. 3.3).

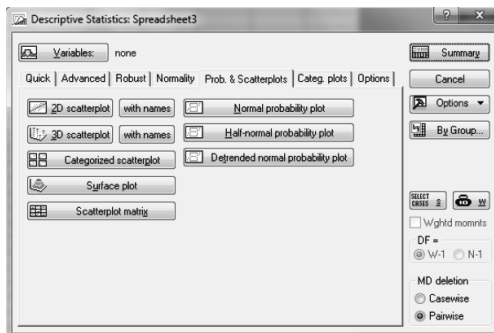


Рис. 3.3. Діалогове вікно модуля Описова статистика / Імовірнісні графіки і діаграми розсіювання

Отриманий графік (рис. 3.4) дає змогу візуально досліджувати, наскільки розподіл даних близький до нормального. Якщо дані експериментальних досліджень розподілені за нормальним законом, то всі значення на графіку мають розташуватися на прямій лінії. Якщо значення не є нормально розподіленими, то буде спостерігатися відхилення від прямої. За графіком бачимо, що закон розподілу відрізняється від нормального.

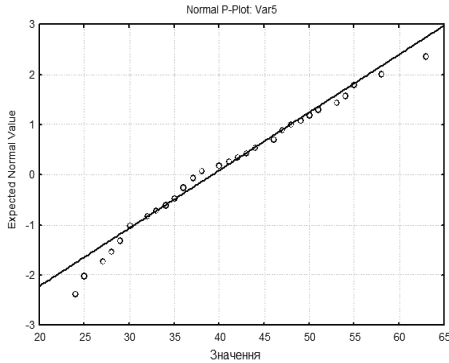


Рис. 3.4. Графік нормального розподілу

### Завдання для самостійної роботи

1. Визначити, чи розподілені дані, наведені у розділі «Варіанти завдань для закріплення тем 1–3», за нормальним законом. Виконати обчислення в програмах Excel і Statistica.

### Контрольні запитання до теми 3

1. Як визначити, чи розподілена випадкова величина за нормальним законом?
2. Для чого потрібна оцінка коефіцієнтів асиметрії та ексцесу?
3. Наведіть формули для розрахунку асиметрії та ексцесу.

**Тема 4. Порівняння середніх показників вибірок.  
Параметричні критерії перевірки гіпотез про середні та дисперсії.  
Критерій Фішера. Перевірка гіпотези про рівність дисперсій**

*Призначення.* Перевірка гіпотези про приналежність двох дисперсій одній генеральній сукупності і, отже, про їх рівність.

*Нульова гіпотеза.*  $S_1^2 = S_2^2$

*Передумови.* Дані незалежні і розподілені за нормальним законом.

Якщо відношення більшої дисперсії до меншої менше критичного (табличного) значення розподілу Фішера, то вважається, що дисперсії двох нормальних генеральних сукупностей однакові:

$$F_{розр} = \frac{s_1^2}{s_2^2}, \rightarrow F_{розр} < F_{\alpha, \nu_1, \nu_2},$$

де  $\alpha$  – рівень значущості;

$\nu_1, \nu_2$  – ступені вільності для дисперсії в чисельнику і знаменнику відповідно.

За описуваного способу перевірки значення обов'язково має бути більшим за одиницю.

**Приклад 4.1.** Наведено дані за двома незалежними вибірками розміру пухлини карциноми Герена на четвертий день захворювання. Було проведено дослідження впливу магнітними полями низької частоти на новоутворення.

Номер досліджу	Номер вибірки	
	1	2
1	0,027	0,075
2	0,036	0,4
3	0,1	0,08
4	0,12	0,105
5	0,32	0,075
6	0,45	0,12
7	0,049	0,06
8	0,105	0,075
Дисперсія	0,02323	0,01283
$F_{розр}$	1,81144	
$F_{крит}$	3,78705	
Розрахунковий рівень значущості	0,22566	

Перевірити гіпотезу про рівність дисперсій.

**Примітка:** Вибірки називаються *незалежними*, якщо процедура відбору даних в першу вибірку ніяк не пов'язана з процедурою відбору спостережень у другу вибірку. Прикладом двох незалежних вибірок можуть бути вибірки чоловіків і жінок, що беруть участь в експерименті, або групи пацієнтів, що приймали певні ліки, і контрольна група, де приймалися нешкідливі ліки. Незалежність вибірок означає відсутність схожості цих вибірок (їх однорідності).

Вибірки називають *залежними*, якщо для однієї і тієї ж групи об'єктів проводять вимірювання деякого параметру за різних умов експерименту (наприклад, вимірюється частота серцевих скорочень у здорових людей до і після фізичних навантажень).

1. Визначаємо, чи можна вважати закон розподілу першої і другої вибірок нормальним. Якщо ні, то необхідно використовувати непараметричний критерій.

2. Розрахуємо дисперсії для першого і другого стовпчика. Для цього використовуємо функцію  $=\text{VAR.S}(\text{інтервал комірок})$ .

3. Знаходимо розрахункові значення для критерію Фішера (див. вище).

4. Обчислюємо критичне (табличне) значення розподілу Фішера, задаючи рівень значущості 0,05. У комірку вводимо функцію  $=\text{F.INV.RT}(0,05;7;7)$ . Число ступенів вільності дорівнює числу експериментів мінус один. Результат 3,787051. Оскільки розрахункове значення критерію Фішера менше критичного (табличного), приймаємо гіпотезу про рівність дисперсій.

5. Розраховуємо для отриманого розрахункового значення критерію Фішера відповідну ймовірність  $=\text{F.DIST.RT}(F13;7;7)$ . У комірці F13 знаходиться розрахункове значення критерію Фішера. Оскільки отримане значення  $0,2256 > 0,05$ , то приймаємо гіпотезу про рівність дисперсій.

Розрахунки, наведені вище, можна виконати спеціальною функцією, яку викликають за допомогою пакету аналізу даних. У вікні *Data analysis (Аналіз даних)* вибираємо *F-test Two-Sample for variances (Двовибірковий F-тест для дисперсій)*.

У діалоговому вікні, що з'явилося, вводимо:

1) інтервали першої та другої вибірок (*Variable Range*);

2) рівень значущості (*alpha*);

3) в поле *Output Range (Вихідний інтервал)* вводимо ім'я комірки, починаючи з якої буде виводитися результат.

### **Перевірка гіпотези про рівність середніх при рівних дисперсіях**

**Призначення.** Перевірка рівності середніх двох генеральних сукупностей, з яких вилучені дві вибірки.

**Передумови.** Обидві вибірки витягнуті з сукупності, що має нормальний розподіл; дані незалежні; дисперсії вибірок однакові.

*Короткі теоретичні відомості.* Розрахункове значення t-критерію обчислюється за такою формулою:

$$t_{\text{розрах}} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(S_1^2(n_1 - 1) + S_2^2(n_2 - 1))}{n_1 + n_2 - 2}}},$$

де  $n_1$  і  $n_2$  – розміри першої і другої вибірок;

$S_1^2$  і  $S_2^2$  – емпіричні дисперсії;

$\bar{X}_1$  і  $\bar{X}_2$  – оцінки середніх значень.

Гіпотеза про рівність середніх відкидається, якщо за абсолютною величиною розрахункове значення t-критерію більше верхньої  $\alpha/2$  % точки t-розподілу, взятого з  $\nu$  ступенями вільності ( $\nu = n_1 + n_2 - 2$ ), тобто при  $|t_{\text{розрах}}| > t_{\alpha/2, \nu}$ .

**Приклад 4.2.** Наведено дані за двома незалежними вибірками розміру пухлини карциноми Герена на четвертий день захворювання. Було проведено дослідження впливу магнітними полями низької частоти на новоутворення. Визначити, чи впливає обробка магнітним полем на зростання пухлини.

Номер досліджу	Номер вибірки	
	1	2
1	0,027	0,075
2	0,036	0,4
3	0,1	0,08
4	0,12	0,105
5	0,32	0,075
6	0,45	0,12
7	0,049	0,06
8	0,105	0,075
Дисперсія	0,02323	0,01283
Середнє	0,15	0,12
$F_{\text{розрах}}$	1,81144	
$F_{\text{табл}}$	3,78705	
Розрахунковий рівень значущості	0,22566	
Число ступенів вільності	14	
Розрахункове значення t-критерію	0,404	
Табличне значення t-критерію	2,5	

Вплив магнітного поля на розмір пухлини визначатимемо за допомогою *t*-критерію Стьюдента. Розрахункове значення критерію Стьюдента розраховується за формулою, наведеною вище; критичне (табличне) значення – з використанням функції =T.INV.2T(0,05/2; число ступенів вільності). *Оскільки розрахункове значення t-критерію менше критичного (табличного) значення, середні показники вибірок рівні. Отже, обробка новоутворень магнітними полями низької частоти не призводить до достовірних змін розміру пухлини.*

Це завдання вирішується безпосередньо спеціальними засобами Excel. Попередньо з'ясуємо, однакові дисперсії чи ні. Припустимо, вони значуще не відрізняються. Вибираємо пункт меню *Дані*, вкладку *Analysis / Data analysis*; у діалоговому вікні, що відкрилося, вибираємо тип аналізу *t-Test: Two-Sample Assuming Equal Variances* (*Двовибірковий t-тест з однаковими дисперсіями*).

У діалоговому вікні задаємо інтервали першої і другої вибірок, рівень значущості (0,025) і місце, де буде знаходитися результат. Порівняйте результати, отримані різними способами.

### **Перевірка гіпотези про рівність середніх при нерівних дисперсіях вибірок**

*Призначення.* Перевірка рівності середніх двох генеральних сукупностей, з яких вилучені дві вибірки.

*Передумови.* Обидві вибірки витягнуті із сукупності, що має нормальний розподіл, дисперсії вибірок відрізняються, вибірки незалежні.

*Короткі теоретичні відомості.* Розрахункове значення *t*-критерію значення визначається за формулою:

$$t_{\text{розрах}} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

При цьому

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 + 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 + 1}} - 2,$$

де  $n_1$  і  $n_2$  – розміри першої і другої вибірок;

$S_1^2$  і  $S_2^2$  – емпіричні дисперсії;

$\bar{X}_1$  і  $\bar{X}_2$  – оцінки середніх значень;

$\nu$  – число ступенів вільності для t-критерію.

При обчисленнях передбачається, що  $\bar{X}_1 > \bar{X}_2$ .

Гіпотеза про рівність середніх відкидається, якщо за абсолютною величиною розрахункове значення t-критерію більше верхньої  $\alpha/2$  % точки t-розподілу, взятого з  $\nu$  ступенями вільності, тобто при  $|t_{розр}| > t_{\alpha/2, \nu}$ .

**Приклад 4.3.** Перша вибірка – маса тварин контрольної групи, друга вибірка – маса експериментальних тварин, які отримували  $\alpha$ -токоферол ацетат. Визначити, чи значуще ці дві вибірки відрізняються одна від одної, тобто, чи впливає введення  $\alpha$ -токоферолу на масу тварин.

1. Перевіряємо закон розподілу на нормальність.
2. Визначаємо середні значення для обох вибірок.
3. Знаходимо дисперсії.

4. Знаходимо розрахункове значення критерію Фішера  $F_{розр} = \frac{S_1^2}{S_2^2}$ .

5. Визначаємо критичне значення критерію Фішера  $=F.INV.RT(0,05;9;8)$ . 0,05 – рівень значущості, 9 і 8 – ступені вільності (кількість даних у вибірках мінус один).

6. Оскільки розрахункове значення більше критичного ( $4,27 > 3,3$ ), то гіпотеза про рівність дисперсій відкидається. Дисперсії нерівні, отже, застосовуємо метод порівняння середніх при нерівних дисперсіях.

7. Знаходимо розрахункове значення t-критерію і числа ступенів вільності за формулами, наведеними вище (1,57 і 14,25).

8. Обчислюємо критичне (табличне) значення t-критерію для розподілу Стьюдента  $=T.INV.2T(0,025, N17)$ , де 0,025 –  $\alpha/2$ , при 5 % рівні значущості, N17 – адреса комірки, в якій знаходиться розраховане значення числа ступенів вільності. Оскільки *розрахункове значення t-критерію менше табличного* ( $1,57 < 2,5$ ), гіпотеза про рівність середніх приймається. *Відмінності між масами тварин контрольної групи і тварин групи, які отримували  $\alpha$ -токоферол ацетат, не достовірні* (рис. 4.1).

	A	B	C
1		Група 1	Група 2
2		1,85	2,27
3		1,87	2,09
4		1,87	2,09
5		2,3	2,41
6		2,52	2,31
7		1,89	2,17
8		2,37	2
9		1,7	2,1
10		1,7	2,02
11		1,94	
12	Середнє	2,00	2,16
13	Дисперсія	0,08	0,02
14	Розрахункове значення критерію Фішера	4,28	
15	Критичне значення критерію Фішера	3,39	
16	t розрахункове	1,57	
17	Число ступенів свободи	13,28	
18	t табличне	2,53	

Рис. 4.1. Дані для прикладу 4.3 і його розв'язання

Це завдання вирішується безпосередньо спеціальними засобами Excel. Попередньо з'ясуємо, однакові дисперсії чи ні. Припустимо, вони значуще відрізняються. Вибираємо пункт меню *Дані*, вкладку *Analysis / Data analysis*; у діалоговому вікні, що відкрилося, вибираємо тип аналізу *t-Test: Two-Sample Assuming Unequal Variances* (*Двовибірковий t-тест з різними дисперсіями*). У вікні задаються інтервали першої і другої вибірок, рівень значущості (0,025) і місце, де буде знаходитися результат. Порівняйте результати, отримані різними способами.

### Парний t-тест Стьюдента для залежних вибірок

*Призначення.* Перевірка рівності середніх двох залежних вибірок, які належать до однієї сукупності нормально розподілених даних.

*Короткі теоретичні відомості.* Розрахункове значення t-критерію обчислюється за такою формулою:

$$t_{розр} = \frac{\bar{X}}{S_{\bar{X}}},$$

де  $\bar{X} = \frac{\sum X}{N}$  – оцінка середнього значення змін між парами даних до ( $X_1$ ) і після ( $X_2$ ) впливу фактора;

$$X = (X_1 - X_2);$$

$N$  – кількість пар спостережень;

$$S_{\bar{X}} = \frac{S_X}{\sqrt{N}} \text{ – стандартна помилка;}$$

$$S_X = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}} \text{ – стандартне відхилення змін значень досліджуваного параметру до і після експерименту.}$$

Гіпотеза про рівність середніх відкидається, якщо  $|t_{\text{розр}}| > t_{\alpha, \nu}$ , де  $\alpha$  – рівень значущості,  $\nu$  – число ступенів вільності ( $N - 1$ ).

**Приклад 4.4.** Було проведено дослідження впливу низькочастотної вібрації на діаметр еритроцитів (мкм) мишей. Зразки крові у досліджуваних тварин відбирали до і після експерименту. Встановити вплив вібрації на діаметр еритроцитів.

У стовпчиках А, В (рядках 3–8) знаходяться вихідні дані. Для перевірки рівності середніх двох залежних вибірок за парним t-критерієм Стьюдента виконаємо такі операції (рис. 4.2):

	A	B	C	D
1	Діаметр еритроцитів, мкм		$X$	$(X - \bar{X})^2$
2	До вібрації	Після вібрації		
3	4,55	7,27	-2,72	0,59
4	5,15	7,33	-2,18	0,05
5	5,15	7,88	-2,73	0,60
6	6,12	6,82	-0,7	1,57
7	5,65	7,18	-1,53	0,18
8	6,36	8,22	-1,86	0,01
9	Сума		-11,72	3,00
10	Середнє		-1,95	
11	Стандартне відхилення		0,77	
12	Стандартна помилка		0,32	
13	t розрахункове		-6,18	
14	t табличне		2,57	

Рис. 4.2. Дані для прикладу 4.4 і його розв'язання

1. Визначимо, чи можна вважати закон розподілу вибірок нормальним.
2. У стовпці С (рядки 3–8) розрахуємо різницю між значеннями пар ознак до і після впливу вібрації.

3. У комірці C9 розрахуємо суму значень, які знаходяться в комірках C3:C8.

4. У комірці C10 розрахуємо середнє значення  $\bar{X}$  для діапазону комірок C3:C8: =AVERAGE(C3:C8).

5. Для розрахунку стандартного відхилення  $S_x$  в комірку C11 введемо формулу: =STDEV.S(C3:C8) (це відповідає формулі для розрахунку, яка наведена вище).

6. Розрахуємо стандартну помилку  $S_{\bar{X}}$ . Для цього в комірку C12 введемо формулу: =C11/SQRT(6).

7. Розрахункове значення t-критерію знайдемо, поділивши середнє значення  $\bar{X}$  на стандартну помилку  $S_{\bar{X}}$ .

8. Знаходимо табличне значення t-критерію за формулою: =T.INV.2T(0,05;5) для ступеня значущості 0,05 та числа ступенів вільності 5.

Оскільки за модулем розраховане значення t-критерію більше за табличне, то гіпотеза про рівність середніх відкидається – низькочастотна вібрація впливає на діаметр еритроцитів.

### Аналіз однорідності дисперсій за Кохреном

*Призначення.* Перевірка гіпотези про приналежність кількох дисперсій до однієї генеральної сукупності.

*Нульова гіпотеза.*  $S_1^2 = S_2^2 = \dots = S_i^2 = S^2$ .

*Передумови.* Дані незалежні і розподілені за нормальним законом. Кількість спостережень однакова в кожній вибірці.

Розрахункове значення критерію Кохрена розраховується за формулою:

$$G = \frac{S_{\max}^2}{\sum_{i=1}^k S_i^2},$$

де  $S_{\max}^2$  – максимальна дисперсія;

$S_i^2$  – дисперсії кожної вибірки.

Якщо розраховане значення критерію Кохрена менше за табличне ( $G_{розр} < G_{табл.\alpha,k,n-1}$ , де  $k$  – кількість вибірок,  $n$  – кількість даних у вибірках), то гіпотеза про однорідність приймається.

Відхилення нульової гіпотези може означати відхилення закону розподілу даних від нормального, а не однорідність дисперсій.

**Приклад 4.5.** Під час дослідження комбінованої дії трьох лікарських препаратів на культурі тканин було проведено дев'ять експериментів (по три рази кожен). Необхідно перевірити, чи можна вважати дисперсії у всіх дев'яти експериментах однорідними. Якщо вони такими не є, то розсіювання залежить від доз будь-яких із досліджуваних препаратів, або наявні грубі помилки в результатах спостережень, пов'язані, наприклад, з особливостями тварин, яких використовували в експериментах.

Номер досліджу	Пригнічення синтезу білка		
	Серійний номер		
	1	2	3
1	30,2	43,6	56
2	4	1	3,5
3	3,5	2,7	3,5
4	31,5	28,7	30,1
5	13,6	16,1	21,1
6	56	28,7	36,2
7	16,1	16,1	28,7
8	11,1	16,1	13,6
9	26,1	35,2	16
Максимальна дисперсія	198,93		
Сума дисперсій	536,17		
Розрахункове значення критерію Кохрена	0,37		
Число вибірок (k)	3		
Розмір вибірки (n)	9		
Рівень значущості	0,05		
$\alpha/(n-1)$	0,00625		
$F_{\alpha/n-1, k, (n-2)k}$ (=F.INV.2T())	5,4488		
$G_{\alpha, k, n-1}$	0,4377		

У стовпчиках С, D, E (рядках 2–10) зазначено вихідні дані. Для проведення аналізу однорідності за Кохреном виконаємо такі операції:

1. Визначимо, чи можна вважати закон розподілу вибірок нормальним.
2. У стовпці G (рядках 2–10) розрахуємо дисперсії для кожного досліджу за формулою = VAR.S(C2:E2).
3. Знайдемо максимальну з дисперсій (=MAX(G2: G10)).
4. Знайдемо суму дисперсій.
5. Знайдемо розрахункове значення критерію Кохрена (0,37).
6. Порівняємо розрахункове значення критерію Кохрена з табличним. Такої функції в Ексел немає, але відомо, що розподіл Кохрена можна апроксимувати розподілом Фішера. Тут розподіли пов'язані співвідношенням:

$$G_{\alpha,k,n-1} = \frac{F_{\alpha/n-1,k,(n-2)k}}{F_{\alpha/n-1,k,(n-2)k} + n - 2} .$$

Оскільки табличне значення критерію Кохрена більше розрахункового, то гіпотеза про однорідність вибірок приймається; різниці між дією препаратів немає.

### Порівняння середніх у двох групах даних у програмі Statistica

Найбільш використовуваним методом, що дає змогу виявити розходження між середніми двох вибірок є t-критерій. Застосування t-критерію має деякі обмеження: він може застосовуватися, якщо змінні *нормально* розподілені (всередині груп), а дисперсії спостережень у групах не дуже відрізняються.

Перевірити відповідність даних нормальному розподілу можна візуально (за допомогою гістограм) або застосовуючи критерії нормальності. Для перевірки рівності дисперсій у групах використовують F-критерій Фішера, який вміщений в таблицю висновку t-критерію в програмі Statistica.

**Приклад 4.6.** Наведено дані за двома незалежними вибірками розміру пухлини карциноми Герена на четвертий день захворювання. Було проведено дослідження впливу магнітними полями низької частоти на новоутворення. Виявити, чи існують відмінності між вибірками.

Номер досліджу	Номер вибірки	
	1	2
1	0,027	0,075
2	0,036	0,4
3	0,1	0,08
4	0,12	0,105
5	0,32	0,075
6	0,45	0,12
7	0,049	0,06
8	0,105	0,075

1. Введемо дані в програму Statistica. Нехай у стовпці 1 знаходяться дані першої вибірки, а в стовпці 2 – другої вибірки даних.

2. Для порівняння середніх показників із використанням параметричного t-критерію відкриємо пункт меню *Statistic, Basic statistics / Tables* (*Статистика, Основна статистика / Таблиці*). У вікні *Основна статистика і таблиці* виберемо пункт *t-test, independent, by variable* (*t-критерій для незалежних вибірок*).

3. У діалоговому вікні натискаємо кнопку *Variables (Змінні)* і вказуємо, які змінні будуть попарно порівнюватися (рис. 4.3).

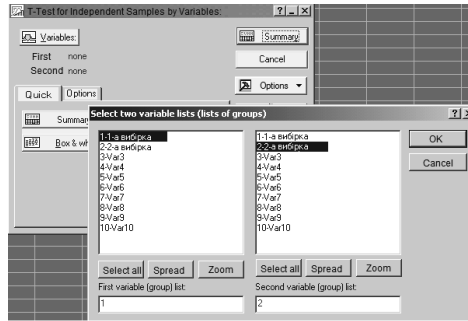


Рис. 4.3. Діалогове вікно вибору змінних

4. Для виводу результатів натискаємо кнопку *Summary: T-tests (Підсумок: T-тести)*. Результати розрахунку показані на рис. 4.4.

		T-test for Independent Samples (Spreadsheet1)										
		Note: Variables were treated as independent samples										
Group 1 vs. Group 2		Mean Group 1	Mean Group 2	t-value	df	p	Valid N Group 1	Valid N Group 2	Std. Dev. Group 1	Std. Dev. Group 2	F-ratio Variances	p Variances
1-а вибірка	vs. 2-а вибірка	0,150875	0,123750	0,404009	14	0,692310	8	8	0,152430	0,113255	1,811442	0,451327

Рис. 4.4. Результати розрахунку за t-критерієм

Звертаємо увагу на рівень достовірності  $p$ . Розрахований рівень  $p$  (0,69) більше 0,05. Отже, відмінності між середніми двох вибірок недостовірні.

5. Натискаємо кнопку *Box & whisker plot («Коробка з вусами»)*. З діаграми розмаху (рис. 4.5) бачимо, що області помилок середнього (великі прямокутники) перекриваються для першої і другої вибірок. Відмінності середніх для двох вибірок знаходяться в межах помилки і, отже, недостовірні.

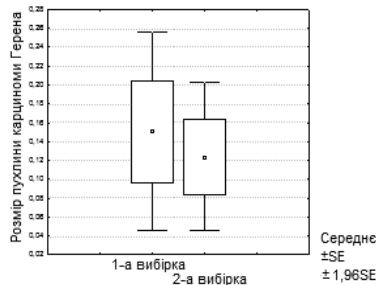


Рис. 4.5. Графічне представлення результатів порівняння вибірок

### **Завдання для самостійної роботи**

1. Порівняти вибірки експериментальних даних, які наведено у розділі «Варіанти завдань для закріплення теми 4», використовуючи параметричні критерії в програмах Excel і Statistica.

### **Контрольні запитання до теми 4**

1. Дайте визначення залежних та незалежних вибірок.
2. Для чого використовується критерій Фішера? За якою формулою визначається розрахункове значення критерію Фішера?
3. Який метод найбільш часто використовується для виявлення розходження між середніми двох вибірок? Які обмеження має цей критерій?
4. За яким критерієм здійснюється перевірка гіпотези про приналежність кількох дисперсій до однієї генеральної сукупності? Наведіть формулу для визначення розрахункового значення цього критерію.

## Тема 5. Порівняння середніх показників вибірок. Непараметричні критерії

Непараметричні критерії застосовують у тих випадках, якщо закон розподілу даних відрізняється від нормального, або дані вимірюють у дискретних шкалах вимірювання.

У всіх випадках використання наведених непараметричних критеріїв приймають такі припущення:

- усі випадкові величини взаємно незалежні;
- аналізуються вибірки, розподілені за одним і тим же законом розподілу.

Під час використання непараметричних методів досліджень часто перевіряють гіпотезу про рівність медіан, оскільки для даних, закон розподілу яких відрізняється від нормального, медіана є більш точною оцінкою положення центру розподілу.

Варто пам'ятати, що непараметричні критерії для випадку нормального розподілу мають меншу потужність і точність, ніж відповідні параметричні критерії. Тому не варто використовувати непараметричні критерії при нормальному розподілі випадкових величин досліджуваних вибірок.

### Поняття про ранги, їх побудова

На практиці досить часто зустрічають числові дані у вибірці, які носять у певному сенсі умовний характер. Це можуть бути експертні оцінки, тестові бали, дані про які-небудь переваги досліджуваної групи людей. У процесі аналізу таких даних часто неможливо дотриматися всіх передумов застосування класичних статистичних методів, які передбачають приналежність робочої вибірки одному з відомих законів розподілу (наприклад, нормальному). Іноді це дуже важко перевірити чи довести (через недостатню кількість спостережень). У таких випадках робити науково обґрунтовані висновки, застосовуючи методи прикладної статистики, можна, використовуючи не самі значення (наприклад, бали), а їхній порядок, заснований на співвідношенні «менше – більше».

Порядок значень називають *рангами*, а *рангом спостереження* – позицію елемента сукупності (номер) у впорядкованій послідовності даних після розташування їх згідно з певним правилом, наприклад, від меншого значення до більшого або навпаки.

*Ранжирування* – це процедура переходу від сукупності спостережень до послідовності їх рангів.

Розглянемо процес ранжирування на прикладі. Припустимо, у нас є вибірка, що складається з п'яти чисел 54, 25, 42, 3, 1. Цим значенням будуть присвоєні відповідні ранги 5, 3, 4, 2, 1 – це позиції елементів наведеної вибірки, якщо її відсортувати за зростанням.

Якщо дані у вибірці збігаються – повторюється одне й те ж саме значення кілька разів, то кожному спостереженню присвоюється той самий ранг, отриманий шляхом підсумовування наявних рангів і ділення на кількість прив'язаних

спостережень. Отже, якби три спостереження були ідентичними і мали б ранги 3, 4, 5, їм усім було б присвоєно *середній ранг* 4  $((3 + 4 + 5) / 3)$ . Сукупність спостережень, які мають однакові значення, називають *зв'язкою*, а кількість спостережень у зв'язці – *розміром зв'язки*.

Статистичні методи, які використовують ранги для отримання науково обґрунтованих висновків з аналізованих даних, називають *ранговими*. Ці методи широко застосовують там, де дуже важко (або неможливо) з'ясувати, яким законам розподілу відповідають аналізовані дані. Однак варто зазначити, що коли в аналізованих даних присутні великі зв'язки (або їх багато), застосування цих методів викликає сумнів. Якщо ж зв'язків небагато (і вони не дуже великі), то їх необхідно враховувати в конкретних розрахункових формулах.

Різні статистичні програми мають процедури, що дають змогу виконати процес ранжирування вибірки, тобто отримати ранги елементів досліджуваної вибірки. Є така функція RANK.AVG(x;R;t) і в електронних таблицях Microsoft Excel, яка враховує середні ранги елементів вибірки за наявності зв'язок.

Параметри цієї функції: x – посилання на комірку з числом, ранг якого необхідно визначити; R – абсолютне посилання на діапазон комірок, який необхідно проранжувати, t – параметр, який вказує напрям ранжування (значення 0 або не вказано – ранжування від більшого до меншого, будь-яке ненульове значення – ранжування від меншого до більшого).

### **Критерій Манна–Уїтні (U-критерій Уїлкоксона–Манна–Уїтні)**

Існує непараметричний тест для порівняння середніх значень із двох незалежних груп, де припущення про нормальність не є виправданим, U-критерій Манна–Уїтні. Цей тест можна застосовувати тільки в тому разі, коли дисперсії вибірок однакові. Нульова гіпотеза цього тесту полягає в тому, що немає різниці між середніми значеннями двох незалежних вибірок, які належать до однієї генеральної сукупності. U-критерій є непараметричним аналогом t-критерію Стьюдента для перевірки середніх значень.

Як і для всіх непараметричних тестів, статистика тесту розраховується після ранжирування спостережень.

Формуємо єдину вибірку, для її значень визначаємо ранги (за зростанням).

Для невеликих вибірок розрахункове значення критерію Манна–Уїтні отримують за допомогою формул:

$$U_1 = n_1 n_2 + 0,5 n_1 (n_1 + 1) - R_1$$

$$U_2 = n_1 n_2 + 0,5 n_2 (n_2 + 1) - R_2$$

Тут  $R_1$  і  $R_2$  – суми рангів першої та другої вибірок розміром  $n_1$  та  $n_2$  відповідно. Зауважимо, що  $n_1$  повинно бути менше ніж  $n_2$ . Якщо ця умова не виконується, потрібно поміняти вибірки місцями.

Правильність розрахунку перевіряємо співвідношенням

$$U_1 + U_2 = n_1 n_2$$

Розрахункове значення визначасмо за формулою:

$$U_{розр} = \min(U_1, U_2)$$

Якщо розрахункове значення U-критерію більше критичного ( $U_{розр} > U(n_1, n_2, \alpha)$ ), то припущення про рівність середніх значень двох вибірок відкидають.

Критичне значення статистики Манна–Уїтні розраховують за формулою:

$$U(\alpha; n_1; n_2) = w(\alpha; n_1; n_2) - n_1(n_1 + 1) / 2,$$

де  $w(\alpha; n_1; n_2)$  – нижнє критичне значення статистики Уїлкоксона W, обчислення якого проводять за апроксимуючою формулою й округляють до цілого значення:

$$w(\alpha, n_1, n_2) = \left[ \frac{n_1 \cdot (n_1 + n_2 + 1) - 1}{2} + \psi(\alpha) \sqrt{\frac{n_1 \cdot n_2 \cdot (n_2 + n_2 + 1)}{12}} \right],$$

де  $\alpha$  – рівень значущості для одностороннього критерію;

$\psi(\alpha)$  – значення оберненої функції нормального розподілу з нульовим середнім і одиничним стандартним відхиленням, яке в програмі Excel розраховується за формулою NORM.S.INV( $\alpha$ ). Отже, для  $\alpha = 0,05$   $\psi(\alpha)$  дорівнює  $-1,64$ .

**Приклад 5.1.** Досліджується ембріотоксичність. Показником є маса плаценти тварин двох груп з умовними номерами 4 і 6, які піддавалися різній обробці.

	<b>A</b>	<b>B</b>
<b>1</b>	Маса плаценти	
<b>2</b>	Група 4	Група 6
<b>3</b>	0,58	0,6
<b>4</b>	0,52	0,6
<b>5</b>	0,51	0,57
<b>6</b>	0,5	0,62
<b>7</b>	0,48	0,52
<b>8</b>	0,5	0,52
<b>9</b>	0,51	0,5
<b>10</b>	0,52	0,5
<b>11</b>	0,51	0,47
<b>12</b>	0,48	0,47
<b>13</b>	NO_NORM*	0,45
<b>14</b>		0,45
<b>15</b>		NORM

**\*Примітка:** якщо під час перевірки розподіл даних відповідає нормальному закону, то далі по тексту в прикладах вирішення завдань це позначатиметься *NORM*, в іншому випадку – *NO\_NORM*

1. Насамперед, необхідно перевірити, чи є закон розподілу вибірки нормальним.
2. Оскільки одна з вибірок має закон розподілу, відмінний від нормального, ми повинні використовувати непараметричний критерій. U-критерій застосовується при рівності дисперсій вибірок. Робимо перевірку за критерієм Фішера, застосовуючи *Двовибірковий F-тест для дисперсій*. Якщо розрахункове значення критерію Фішера менше критичного, дисперсії вибірок однакові.
3. У стовпці С формуємо об'єднану вибірку.
4. У стовпці D розраховуємо ранги для об'єднаної вибірки. Для цього в комірку D3 вводимо функцію `=RANK.AVG(D3;$D$3:$D$24;1)`, а потім множи-мо її для всього стовпця.
5. Виконуємо розрахунки всіх необхідних значень для перевірки гіпотези (табл. 5.1).

Таблиця 5.1. Формули для перевірки критерію Манна–Уїтні

Номер рядка	Код стовпчика		Коментар
	E	F	
7	$R_1=$	<code>=SUM(D3:D12)</code>	сума значень рангів, що належать першій вибірці
8	$R_2=$	<code>=SUM(D13:D24)</code>	сума значень рангів, що належать другій вибірці
9	$n_1$	<code>=COUNT(A3:A12)</code>	розмір першої вибірки
10	$n_2$	<code>=COUNT(B3:B14)</code>	розмір другої вибірки
11	$U_1=$	<code>=F9×F10+0,5×F9×(F9+1)–F7</code>	розрахункове значення критерію для першої вибірки
12	$U_2=$	<code>=F9×F10+0,5×F10×(F10+1)–F8</code>	розрахункове значення критерію для другої вибірки
13	$U_1+U_2=$	<code>=F11+F12</code>	перевірка правильності розрахунків (значення повинні збігатися)
14	$n_1×n_2$	<code>=F9×F10</code>	
15	$U_{розр}$	<code>=MIN(F11:F12)</code>	розрахункове значення
16	$w(\alpha; n_1; n_2)$	$= \frac{10 \cdot (10 + 12 + 1) - 1}{2} - 1.64 \times \sqrt{\frac{10 \cdot 12 \cdot (12 + 12 + 1)}{12}} = 89$	нижня межа критерію Уїлкоксона (отримане значення округлюється до цілого числа)
17	$U_{критич}$	<code>=F16–10×(10+1)/2=34</code>	критичне (табличне) значення критерію Манна–Уїтні

Оскільки розраховане значення  $U$  (59) більше критичного (34), нульова гіпотеза про рівність середніх досліджуваних вибірок відкидається, між порівнюваними вибірками існують достовірні відмінності.

*Варто зауважити, що при великій кількості спостережень для зіставлення середніх двох незалежних вибірок, розподіл яких відрізняється від нормального, краще застосовувати двовибірковий критерій Уїлкоксона.*

### **Двовибірковий критерій Уїлкоксона**

Критерій Уїлкоксона призначений для перевірки гіпотези про рівність середніх значень двох незалежних вибірок, що належать до однієї генеральної сукупності і мають однаковий тип розподілу.

Варто зазначити, що розміри вибірок  $n_1$  і  $n_2$  можуть відрізнитися. Якщо  $n_1 > n_2$ , то вибірки необхідно поміняти місцями. Подальший аналіз даних проводиться аналогічно до попереднього прикладу (Приклад 5.2, пп.1–4). Якщо розміри вибірок однакові, то розрахункове критеріальне значення  $W_{розр}$  дорівнюватиме одній із сум рангів ( $R_1$  або  $R_2$ ), в іншому випадку  $W_{розр} = R_1$ .

Обчислення значення  $W_{розр}$  порівнюємо з верхнім  $W(\alpha, n_1, n_2)$  і нижнім  $w(\alpha; n_1; n_2)$  критичними значеннями статистики Уїлкоксона. Розрахунок нижнього критичного значення статистики Уїлкоксона наведено в попередньому пункті.

Верхнє критичне значення розраховується за формулою:

$$W(\alpha, n_1, n_2) = n_1 \times (n_1 + n_2 + 1) - w(\alpha; n_1; n_2).$$

Якщо розраховане значення  $W_{розр} \leq w(\alpha; n_1; n_2)$  або  $W_{розр} \geq W(\alpha, n_1, n_2)$ , то гіпотеза про рівність середніх двох вибірок відхиляється.

### **Критерій Краскела–Уолліса (H-критерій)**

Для перевірки рівності середніх значень  $k$  вибірок використовують H-критерій Краскела–Уолліса.

*Нульова гіпотеза.* Всі  $k$  вибірок мають однаковий розподіл.

*Передумови.* Всі випадкові величини взаємно незалежні.

Для визначення розрахункового значення H-критерію всі вибірки об'єднують в одну. Для об'єднаної вибірки виконують ранжирування. Після цього визначають суми рангів у кожній вибірці.

Розрахунок значення H-критерію визначають за формулою:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1),$$

де  $N$  – загальна кількість спостережень, яка визначається так:  $N = \sum_{i=1}^k n_i$ ;

$n_i$  – число спостережень в  $i$ -тій вибірці;  
 $R_i$  – сума рангів в  $i$ -тій вибірці;  
 $k$  – кількість вибірок.

Виконуємо порівняння розрахункового значення критерію з критичним (табличним). У тому разі, якщо розрахункове значення Н-критерію більше табличного, то гіпотеза про рівність середніх відхиляється.

При малому загальному числі спостережень ( $N < 15$ ) використовують спеціальні таблиці розподілу Краскела–Уолліса; в іншому випадку для критичного значення Н-критерію використовують верхню 5-відсоткову точку розподілу  $\chi^2$  с  $(k - 1)$  числом ступенів вільності.

**Примітка.** У тих випадках, коли вибірки мають різний розподіл, цей критерій непридатний. Замість нього слід використовувати критерій Фрідмана. Коли в стовпці спостережень є зв'язки, значення критерію коригується за формулою:

$$H' = \frac{H}{1 - \frac{\sum_{i=1}^L T_i}{N^3 - N}}$$

де  $L$  – кількість зв'язок;

$T_i$  – кількість співпадаючих елементів в  $i$ -й зв'язці.

**Приклад 5.2.** Досліджуємо ембріотоксичність. У стовпці В, С і D наведені значення розмірів плаценти тварин трьох груп (1, 2, 3). Група 1 – інтактні вагітні тварини; тваринам групи 2 та 3 вводили розчин  $\alpha$ -токоферолу ацетату з різною концентрацією. Чи можна вважати, що середні значення всіх груп однакові й ефект обробки відсутній?

	А	В	С	Д	Ф
1		Розмір плаценти (мм)			
2	Група	1	2	3	
3		14	14	12	
4		12	16	14	
5		14	16	14	
6		13	15	15	
7		13	12	15	
8		16	16	15	
9		12	12	14	
10		13	12	13	
11		12	12	16	
12		13	11	13	
13				14	
14				14	

1. Необхідно перевірити, чи можна вважати, що дані розподілені за нормальним законом. Відповідь у всіх трьох випадках негативна, тому застосовуємо ранговий критерій.

2. Об'єднуємо всі три вибірки в одну. Об'єднана вибірка займає комірки з F3 по F34.

3. Для об'єднаної вибірки будуємо ранги. Для цього в комірці G3 вводим функцію  $=\text{RANK.AVG}(F3;\$F\$3:\$F\$34;1)$ , потім копіюємо її в комірки G4–G34.

4. Знаходимо суми рангів для кожної вибірки.

I12	$=\text{SUM}(G3:G12)$
I13	$=\text{SUM}(G13:G22)$
I14	$=\text{SUM}(G23:G34)$

5. Перевіряємо правильність розрахунків за рангами. Сума рангів у стовпці G (G3:G34) повинна дорівнювати сумі рангів у стовпці I (I12:I14). Для наведеного прикладу – 528.

6. Розраховуємо квадрати сум рангів

J12	$=I12*I12$
J13	$=I13*I13$
J14	$=I14*I14$

7. Знаходимо квадрати рангів, поділені на кількість експериментів

K12	$=J12/\text{COUNT}(B3:B12)$
K13	$=J13/\text{COUNT}(C3:C12)$
K14	$=J14/\text{COUNT}(D3:D14)$
K15	$=\text{SUM}(K12:K14)$

8. У комірку I16 поміщаємо значення N (загальне число спостережень):  $=\text{COUNT}(B3:B12)+\text{COUNT}(C3:C12)+\text{COUNT}(D3:D14)$ .

9. У комірці I17 знаходимо розрахункове значення H-критерію за формулою  $=(12/(I16 \times (I16+1))) \times K15 - 3 \times (I16+1)$ . Має вийти позитивне значення ( $H_{\text{розн}} = 2,306203$ ).

10. Оскільки загальна кількість дослідів понад 15, для отримання табличного значення можна скористатися відсотковою точкою розподілу  $\chi^2$ . Для цього в комірку I18 вводим функцію:  $=\text{CHIQ.INV.RT}(0,05; 3-1)$ , 3 – кількість вибірок.

11. Оскільки розрахункове значення H-критерію менше табличного, то нульова гіпотеза приймається. Отже, вважаємо, що середні значення цих вибірок рівні й, отже, ефект обробки відсутній.

12. У нашому прикладі багато зв'язок: 5 значень по 16; 4 – по 15; 8 – по 14; 6 – 13; 8 – по 12; 1 – 11. Отже,  $L = 6$ , а масив  $T = \{1, 8, 6, 8, 4, 5\}$ . Тоді скориговане значення критерію визначають так:

$$H' = \frac{3,89}{1 - \frac{(1+8+6+8+4+5)}{32^3 - 32}} = 2,308459$$

При скоригованому значенні висновки не змінюються.

### Медіанний критерій для кількох вибірок

Медіанний критерій використовують для перевірки належності  $k$ -вибірок до однієї генеральної сукупності. Всі випадкові величини взаємно незалежні.

Спочатку знаходять медіану для кожної вибірки і після цього формують таблицю 5.2. Обчислюють очікуване число спостережень для кожної комірки. Для цього елемент рядка «Всього» множать на елемент стовпця і ділять на загальне число спостережень  $n$ .

Таблиця 5.2. Підготовчі розрахунки для медіанного критерію

	Вибірка				Всього
	1	2	...	k	
Кількість спостережень, які більше медіани	$L_1$	$L_2$	...	$L_k$	$\sum_{j=1}^k L_j$
Кількість спостережень, які менше медіани	$n_1 - L_1$	$n_2 - L_2$	...	$n_k - L_k$	$n - \sum_{j=1}^k L_j$
Всього	$n_1$	$n_2$	...	$n_k$	$n$

Очікуване значення для першого рядка (в якому знаходяться значення  $L_j$ ) визначають за формулою:

$$K_{1j} = \frac{L_{1j} \sum_{j=1}^k L_j}{n}, j=1 \dots k.$$

Для другого рядка, в якому знаходяться  $(n_j - L_j)$ , очікувані значення визначають за формулою:

$$K_{2j} = \frac{(n_j - L_j)(n - \sum_{j=1}^k L_j)}{n}$$

Потім обчислюють розрахункове значення критерію:

$$\chi^2 = \sum \frac{(\text{спостережуване значення} - \text{очікуване значення})^2}{\text{очікуване значення}}$$

Сума знаходиться за всіма  $2k$  комірками таблиці. Якщо розраховане значення  $\chi^2$  більше верхнього критичного значення розподілу, то гіпотеза про рівність середніх відхиляється.

**Приклад 5.3.** Є дані про розміри плаценти для трьох вибірок. Встановити, чи є достовірні відмінності між аналізованими вибірками даних.

	A	B	C	D	E	F	G
1		Розмір плаценти					
2	Група	1	2	3	1	2	3
3		15	12	12	1	0	0
4		13	16	14	0	1	0
5		15	16	14	1	1	0
6		14	15	14	1	0	0
7		13	14	15	0	0	1
8		16	15	15	1	0	1
9		15	16	15	1	1	1
10		13	14	14	0	0	0
11		12	16	13	0	1	0
12		13	15	16	0	0	1
13				13			0
14				14			0
15	Середнє	14	13,45	14,08			
17	Медіана	13,5	15	14			

Для перевірки гіпотези про рівність медіан цих вибірок необхідно виконати такі дії:

1. Визначаємо медіани для всіх трьох вибірок. Для цього в комірки B17, C17, D17 поміщаємо формули =MEDIAN(B3:B12), =MEDIAN(C3:C12), =MEDIAN(D3:D14) відповідно.

2. У стовпцях E, F, G формуємо додатковий масив із нулів і одиниць: поміщаємо одиниці, якщо спостережувані значення більше медіани, в іншому випадку – нулі. Це можна зробити за допомогою формул =IF(\$B3>\$B\$17;1;0), =IF(\$C3>\$C\$17;1;0), =IF(\$D3>\$D\$17;1;0), які вводять у комірки E3, F3 і G3 відповідно, а потім розмножують на всі відповідні стовпці.

## 3. Формуємо необхідну нам таблицю 5.3.

Таблиця 5.3. Таблиця розрахунків для перевірки медіанного критерію

Рядки / Стовпці	Вибірка 1 (H)	Вибірка 2 (I)	Вибірка 3 (J)	Всього (K)
Число спостережень, які більше медіани (18)	=SUM(E3:E12)	=SUM(F3:F12)	=SUM(G3:G14)	=SUM(H18:J18)
Число спостережень, які менше медіани (19)	=H20-H18	=I20-I18	=J20-J18	=SUM(H19:J19)
Всього (20)	=COUNT(B3:B12)	=COUNT(C3:C12)	=COUNT(D3:D14)	=SUM(H20:J20)

## 4. Формуємо таблицю з розрахунковими очікуваними значеннями (табл. 5.4).

Таблиця 5.4. Розрахункові очікувані значення

Рядки / Стовпці	L	M	N
18	=H18*\$K\$18/\$K\$20	=I18*\$K\$18/\$K\$20	=J18*\$K\$18/\$K\$20
19	=H19*\$K\$19/\$K\$20	=I19*\$K\$19/\$K\$20	=J19*\$K\$19/\$K\$20
20	=(H18-L18)*(H18-L18)/L18	=(I18-M18)*(I18-M18)/M18	=(J18-N18)*(J18-N18)/N18
21	=(H19-L19)*(H19-L19)/L19	=(I19-M19)*(I19-M19)/M19	=(J19-N19)*(J19-N19)/N19

5. Знаходимо розрахункове значення медіанного критерію за формулою  $=SUM(L20:N21)$ ; введено в комірку L24.

6. За допомогою функції  $=CHIQ.INV.RT(0,05; 3-1)$  обчислюємо розрахункове значення медіанного критерію і порівнюємо з табличним. Результат функції розміщений в комірці L26. Оскільки розрахункове значення (16,56) більше табличного (5,99), то нульова гіпотеза відкидається. Відмінності між вибірками достовірні.

### Непараметрична статистика в програмі Statistica

Одним із факторів, що обмежують застосування критеріїв, заснованих на припущенні нормальності, є обсяг вибірки. До того часу, поки вибірка досить велика (наприклад, 100 або більше спостережень), можна вважати, що вибірковий розподіл нормальний, навіть якщо ви не впевнені, що розподіл змінної в популяції є нормальним. За невеликої кількості спостережень обов'язковою є перевірка на нормальний розподіл.

### Короткий огляд непараметричних процедур у програмі Statistica

Кожний параметричний критерій має хоча б одну непараметричну альтернативу. Загалом, ці процедури потрапляють в одну з таких категорій:

- критерії відмінності для незалежних вибірок;
- критерії відмінності для залежних вибірок;
- оцінка ступеня залежності між змінними.

Стартова панель модуля *Nonparametric Statistics (Непараметрична статистика)* має вигляд (рис 5.1):

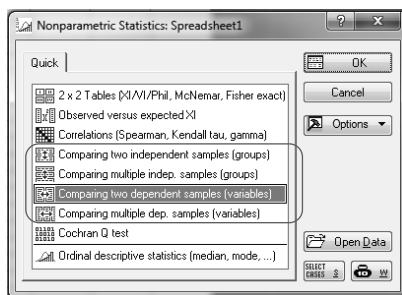


Рис. 5.1. Стартова панель модуля *Nonparametric Statistics (Непараметрична статистика)*

Розглянемо критерії відмінності для незалежних та залежних вибірок, наявні в цьому діалоговому вікні.

Для порівняння двох незалежних вибірок (*Comparing two independent samples (groups)*) застосовуються критерії Манна-Уїтні, Вальда-Вольфовиця і двовибірковий критерій Колмогорова-Смирнова (рис. 5.2).

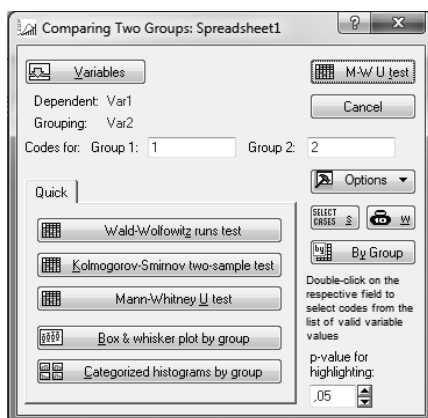


Рис. 5.2. Діалогове вікно порівняння двох незалежних вибірок

Критерії передбачають, що розглянуті змінні є безперервними і виміряні принаймні в порядковій шкалі.

Файл із вихідними даними повинен містити гуртуючу змінну, що має якнайменше два різних коди для ідентифікації приналежності кожного спостереження до певної групи.

Критерії порівняння двох незалежних груп влаштовані так: уявіть, що ви хочете порівняти чоловіків і жінок за певною ознакою. Ви можете впорядкувати дані, наприклад, за зростанням, і знайти ті випадки, коли суб'єкти однієї і тієї ж статі примикають один до одного в побудованому варіаційному ряді (тобто утворюють серію).

Якщо немає відмінності між чоловіками і жінками, то число і довжина «серій», що відносяться до однієї і тієї ж статі, будуть більш-менш випадковими. В іншому випадку дві групи (чоловіки і жінки) відрізняються один від одного, тобто не є однорідними.

Розглянемо зіставлення двох незалежних груп у програмі Statistica на даних із прикладу 5.1:

Маса плаценти	
Група 4	Група 6
0,58	0,6
0,52	0,6
0,51	0,57
0,5	0,62
0,48	0,52
0,5	0,52
0,51	0,5
0,52	0,5
0,51	0,47
0,48	0,47
	0,45
	0,45

1. Із даних першого та другого стовпчика формуємо єдину вибірку, яку розміщуємо в стовпчик Var1. Дасмо назву стовпчику – Маса плаценти. У другий стовпчик навпроти кожного вимірювання прописуємо номер групи досліджуваних тварин (рис. 5.3).

2. Для вибору гуртуючої змінної і списку залежних змінних натискаємо кнопку *Variable* (Змінні). У діалоговому вікні вибору змінних обираємо «Масу плаценти» як залежну змінну (*Dependent variable list*), «№ групи» як гуртуючу змінну (*Independent (grouping) list*). Натискаємо кнопку ОК для повернення в діалогове вікно *Comparing two group* (Порівняння двох груп).

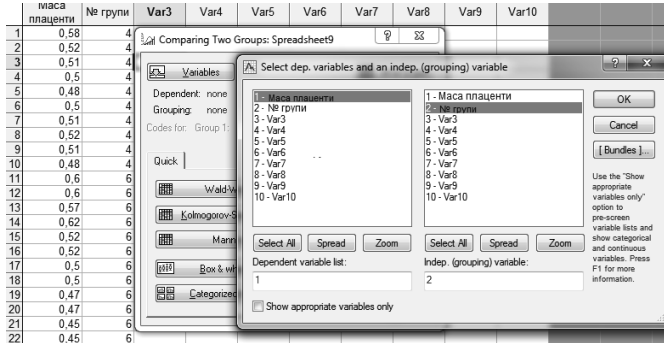


Рис. 5.3. Діалогове вікно вибору залежної і групової змінних у програмі Statistica

3. Після вибору одного з критеріїв (Манна–Уїтні, Вальда–Вольфовиця або Колмогорова–Смирнова) отримуємо результати розрахунку (рис. 5.4). Звертаємо увагу на рівень достовірності  $p$ . Розрахований рівень  $p$  (0,97) більше 0,05. Отже, відмінності між середніми двох вибірок недостовірні.

Mann-Whitney U Test (Spreadsheet9)										
By variable № групи										
Marked tests are significant at $p < .05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Маса пацієнти	114,0000	139,0000	59,00000	-0,032969	0,973699	-0,033233	0,973489	10	12	0,974236

Рис. 5.4. Результати зіставлення двох незалежних груп за  $U$ -критерієм

4. Натискаємо кнопку *Box & whisker plot* («Коробка з вусами»). На діаграмі розмаху (рис. 5.5) бачимо, що відмінності середніх для двох вибірок знаходяться в межах помилки і, отже, недостовірні.

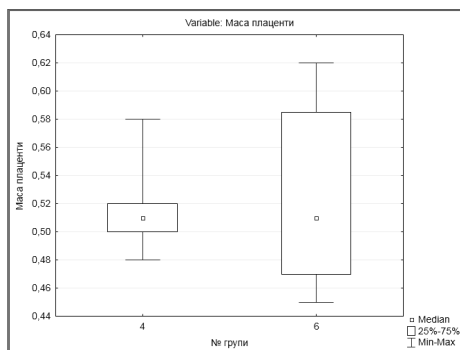


Рис. 5.5. Графічне представлення результатів порівняння вибірок

Для зіставлення *k* незалежних вибірок у програмі Statistica використовуються тест Краскела–Уолліса і медіанний тест (*Kruskal–Wallis ANOVA and Median Test*) (рис. 5.6).

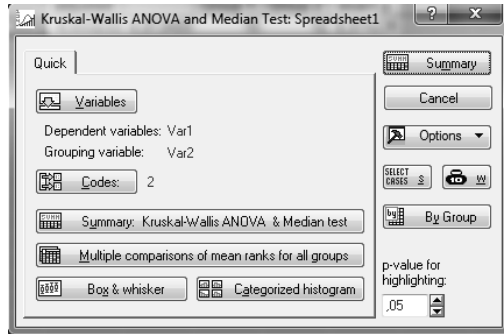


Рис. 5.6. Диалогове вікно порівняння *k* незалежних вибірок

Критерій Краскела–Уолліса, як і інші непараметричні критерії, базується на рангах (а не на вихідних спостереженнях) і передбачає, що розглянута змінна неперервна і виміряна, як мінімум, у порядковій шкалі. Критерій перевіряє гіпотезу: чи мають порівнювані вибірки один і той же розподіл або ж розподіл з однією і тією медіаною.

Нульова гіпотеза (всі вибірки витягнуті з генеральної сукупності з рівними медіанами) приймається, якщо половина спостережень із кожної вибірки більше (або менше) загальної медіани.

Розглянемо порівняння *k* незалежних груп у програмі Statistica на даних з прикладу 5.2:

Розмір плаценти (мм)		
1 група	2 група	3 група
14	14	12
12	16	14
14	16	14
13	15	15
13	12	15
16	16	15
12	12	14
13	12	13
12	12	16
13	11	13
		14
		14

Після запуску модуля *Comparing multiple indep. samples (groups)* (Порівняння кількох незалежних груп) проводимо такі операції:

1. Формуємо єдину вибірку, як у попередньому прикладі, де розмір плаценти – залежна, а номер групи (1, 2, 3) – незалежна (гуртуюча) змінна.

2. Після вибору змінних у діалоговому вікні вибору змінних (*Select dep. Variables and an indep. (grouping) variables*) натискаємо кнопку *Codes (Коду)* і вибираємо всі коди для незалежної змінної (кнопка *All*). Зауважимо, що коли груп багато, то за допомогою кнопки *Codes* можна зіставляти вибірково окремі групи.

3. Натиснемо кнопку *Kruskal–Wallis ANOVA and Median Test (ANOVA Краскела–Уолліса і медіанний тест)*. У діалоговому вікні натискаємо *OK* для початку аналізу.

Результати рангової *ANOVA Краскела–Уолліса* будуть показані в першій таблиці результатів (рис. 5.7), результати медіанного тесту – у другій (рис. 5.8).

Depend.: Розмір плаценти		Kruskal-Wallis ANOVA by Ranks: Розмір плаценти (Spr) Independent (grouping) variable: № групи Kruskal-Wallis test: H (2, N= 32) =2,409083 p = ,2998			
Code	Valid N	Sum of Ranks	Mean Rank		
1	10	135,5000	13,55000		
2	10	158,0000	15,80000		
3	12	234,5000	19,54167		

Рис. 5.7. Результати розрахунку за ранговим критерієм Краскела–Уолліса

Dependent: Розмір плаценти		Median Test, Overall Median = 14,0000: Розмір плаценти (Spr) Independent (grouping) variable: № групи Chi-Square = 2,483736 df = 2 p = ,2888			
	1	2	3	Total	
≤ Median: observed	9,00000	6,00000	8,00000	23,00000	
expected	7,18750	7,18750	8,62500		
obs - exp	1,81250	-1,18750	-0,62500		
> Median: observed	1,00000	4,00000	4,00000	9,00000	
expected	2,81250	2,81250	3,37500		
obs - exp	-1,81250	1,18750	0,62500		
Total: observed	10,00000	10,00000	12,00000	32,00000	

Рис. 5.8. Результати розрахунку за медіанним критерієм

Бачимо, що за обома критеріями розрахований рівень  $p$  (0,29) більше 0,05 і відмінності між середніми трьох вибірок недостовірні. Такий самий висновок можна зробити, побудувавши діаграму розмаху, тип *Median / Quart / Range* (рис. 5.9)

Крім того, в діалоговому вікні зіставлення вибірок є можливість побудувати категоризовані гістограми (*Categorized histogram*), які дають змогу візуально оцінити розподіл даних кожної з вибірок (рис. 5.10).

Розглянемо критерії, які застосовуються в ситуаціях, коли дослідник проводить два виміри (наприклад, при різних умовах) одних і тих же суб'єктів і хоче встановити наявність або відсутність відмінності результатів. Для застосування цих критеріїв достатньо слабких припущень, наприклад, однозначна визначеність

медіани для різниці значень. Не потрібно ніяких припущень про природу або форми розподілу.

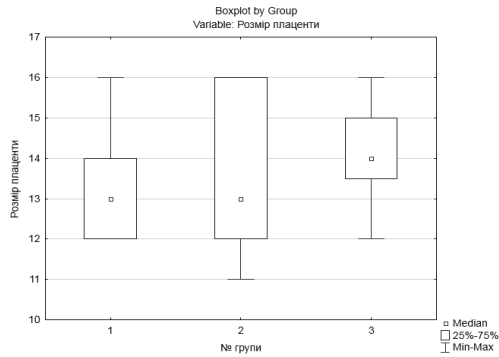


Рис. 5.9. Графічне представлення результатів зіставлення трьох вибірок

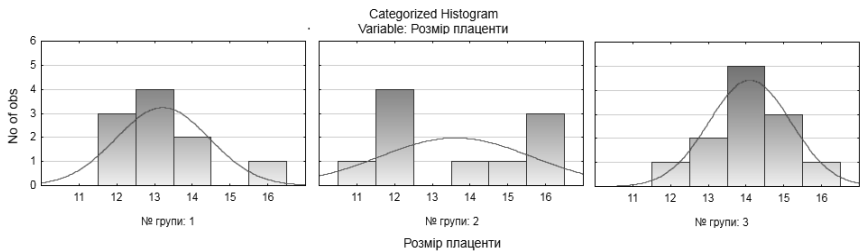


Рис. 5.10. Категоризовані гістограми досліджуваних вибірок

Для зіставлення двох залежних вибірок у програмі Statistica використовуються парний тест Уїлкоксона (*Wilcoxon matched pairs test*) і знаковий тест (*Sign test*) (рис. 5.11).

Критерій Уїлкоксона парних порівнянь є непараметричною альтернативою t-критерію для двох залежних вибірок. Після вибору опції на екрані з'являється діалогове вікно, в якому можна вибрати змінні з двох списків. Змінні списків порівнюються попарно. Попередньо змінні ранжують. W-статистика Уїлкоксона дорівнює сумі рангів елементів другої вибірки у загальному варіаційному ряді двох вибірок. Отже, спостереження двох груп об'єднуються, будується загальний варіаційний ряд і обчислюється сума рангів другої групи в побудованому ряді.

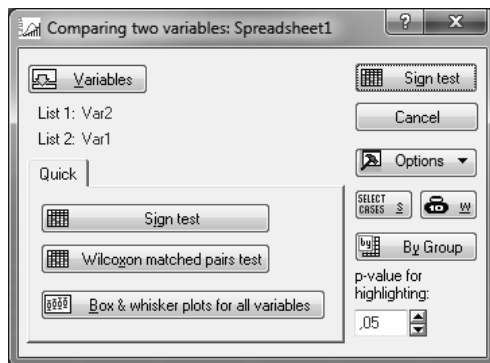


Рис. 5.11. Діалогове вікно зіставлення двох залежних вибірок у програмі Statistica

**Приклад 5.4.** У таблиці наведено дані щодо рівня гемоглобіну в крові хворих на анемію до і після курсу лікування. Встановити, чи призвело лікування до достовірного збільшення рівня гемоглобіну в крові.

Рівень гемоглобіну до лікування, г/л		Рівень гемоглобіну після лікування, г/л	
102	104	121	118
100	104	118	121
102	104	128	118
98	102	122	121
101	102	128	124
104	110	118	120
102	105	128	118
104	102	124	118
100	102	118	123
98	104	128	120
98	98	120	127
98	102	120	118

Вводимо дані в два стовпці програми Statistica. Змінну першого стовпчика Var1 перейменувати «До лікування», змінну другого стовпчика – «Після лікування».

У діалоговому вікні *Comparing two dependent samples (variables)* (Порівняння двох залежних змінних) задаємо змінні для зіставлення: *first variable list* – До лікування, *second variable list* – Після лікування. Натискаємо кнопку *Wilcoxon matched pairs test* (Парний тест Уїлкоксона). У таблиці звертаємо увагу на  $p$ -рівень. Якщо  $p$ -рівень ( $p$ -level) менше 0,05, то відмінності між вибірками достовірні.

Самостійно виконайте аналіз із використанням критерію Уїлкоксона і побудуйте діаграму розмаху, вибравши тип діаграми Mean / SE / SD.

Для зіставлення *k* залежних вибірок у програмі Statistica використовується аналіз ANOVA Фрідмана і коефіцієнт конкордації, або згоди, Кендала (*Friedman ANOVA & Kendall's concordance*) (рис. 5.12).

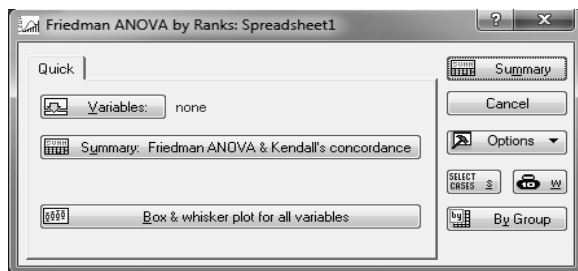


Рис. 5.12. Діалогове вікно зіставлення кількох залежних вибірок

ANOVA Фрідмана – це непараметрична альтернатива однофакторного дисперсійного аналізу з повторними вимірами (буде розглядатися в наступних розділах). Коефіцієнт конкордації (згоди) Кендала – аналог коефіцієнта кореляції Спірмена, коли число змінних більше двох.

**Приклад 5.5.** У таблиці наведено дані вмісту гемоглобіну (г / л) у пацієнтів однієї і тієї ж групі до і після (через один і два тижні) лікування. Встановити, чи привело лікування до достовірного збільшення рівня гемоглобіну в крові.

Рівень гемоглобіну до лікування		Рівень гемоглобіну через тиждень від початку лікування		Рівень гемоглобіну через два тижні від початку лікування	
102	104	105	108	121	118
100	104	102	112	118	121
102	104	102	109	128	118
98	102	102	106	122	121
101	102	106	108	128	124
104	110	108	104	118	120
102	105	109	102	128	118
104	102	108	109	124	118
100	102	109	104	118	123
98	104	102	110	128	120
98	98	108	100	120	127
98	102	108	108	128	118

Вводимо дані в три стовпці програми Statistica. Змінну першого стовпчика Var1 перейменуємо в «До лікування», змінну другого стовпчика – «Тиждень», змінну третього стовпчика – «Два тижні».

У діалоговому вікні *Friedman ANOVA by ranks* (Ранговий аналіз Фрідмана) задаємо змінні для зіставлення (вибираємо в списку всі три вибірки).

Натискаємо кнопку *Friedman ANOVA & Kendall's concordance*. У таблиці звертаємо увагу на р-рівень. Якщо р-рівень менше 0,05, то відмінності між вибірками достовірні (рис. 5.13).

Friedman ANOVA and Kendall Coef. of Concordance (Spre					
ANOVA Chi Sqr. (N = 24, df = 2) = 47,36986 p = ,00000					
Coef. of Concordance = ,98687 Aver. rank r = ,98630					
Variable	Average Rank	Sum of Ranks	Mean	Std.Dev.	
До лікування	1,000000	24,00000	101,9167	2,842483	
Тиждень	2,479167	59,50000	121,6250	3,739478	
Два тижні	2,520833	60,50000	121,9583	3,939534	

Рис. 5.13. Результати розрахунку за аналізом Фрідмана

Самостійно побудуйте діаграму розмаху, вибравши тип діаграми Mean / SE / SD.

### Завдання для самостійної роботи

1. Порівняти вибірки експериментальних даних, наведених у розділі «Варіанти завдань для закріплення теми 5», використовуючи непараметричні критерії в програмах Excel і Statistica.

### Контрольні запитання до теми 5

1. В яких випадках для зіставлення груп використовують непараметричні критерії? Які припущення при цьому приймають?
2. Що таке ранги? У яких випадках застосовується ранжирування даних? Наведіть приклади.
3. Перерахуйте відомі Вам непараметричні критерії зіставлення груп та особливості їх застосування.

## Тема 6. Перевірка наявності зв'язку між змінними. Кореляційний аналіз

Зазвичай у біологічних або екологічних дослідженнях вимірюють більше однієї змінної з кожної вибірки або об'єкта експериментальних досліджень. Наприклад, фізіолог може записати артеріальний тиск і масу тіла у піддослідних тварин, або еколог може записати чисельність певного виду чагарників та рН ґрунту з серії ділянок під час відбору проб рослинності.

У такому випадку часто постає питання, наскільки величина одних досліджуваних параметрів залежить від інших. Для перевірки гіпотез про наявність зв'язків між змінними використовують такі види статистичного аналізу:

- *кореляційний аналіз* застосовується в тих випадках, коли змінні вимірюються в шкалах відносин, інтервалів або порядку;
- *дисперсійний аналіз* застосовується, коли залежна змінна вимірюється в шкалі відносин, інтервалів або порядку, а незалежна змінна має нечислову природу (шкала найменувань);
- *аналіз таблиць спряженості* застосовується, якщо залежна змінна показує кількість (відсоток) спостережень, для яких ознака присутня або відсутня, а незалежні (впливаючі) змінні мають нечислову природу.

### Кореляційний аналіз

Кореляційний аналіз – це статистична процедура для перевірки зв'язку між кількісними або категоріальними змінними. Інакше кажучи, він описує ступінь зв'язку між двома змінними. Це один із найбільш часто використовуваних статистичних прийомів.

### Параметрична кореляція (коефіцієнт кореляції Пірсона)

Лінійна кореляція – це статистичний зв'язок між двома випадковими взаємно незалежними змінними, які мають нормальний закон розподілу. Ступінь кореляції вимірюється статистикою, яка називається коефіцієнтом кореляції і виражає ступінь передбачуваного лінійного зв'язку між двома вибраними змінними X та Y:

$$r_{xy} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \sum_{i=1}^N (Y_i - \bar{Y})^2}}$$

Коефіцієнт кореляції – безрозмірна величина, значення якої може змінюватися від  $-1$  (ідеальна негативна кореляція) до  $+1$  (ідеальна позитивна кореляція). Позитивний коефіцієнт кореляції вказує на те, що змінні мають прямий зв'язок, тобто коли значення однієї змінної зростає, значення іншої змінної також має

тенденцію до зростання (рис. 6.1а). І навпаки, якщо коефіцієнт є від'ємним числом, це означає, що вибрані змінні мають негативний зв'язок, тобто коли значення однієї змінної зростає, значення інших має тенденцію до зменшення (рис. 6.1б). У статистичному плані будь-яка інша форма співвідношення між двома безперервними змінними, яка не є лінійною (рис. 6.1в-г), не розглядається як кореляція. В такому випадку коефіцієнт лінійної кореляції дорівнює нулю.

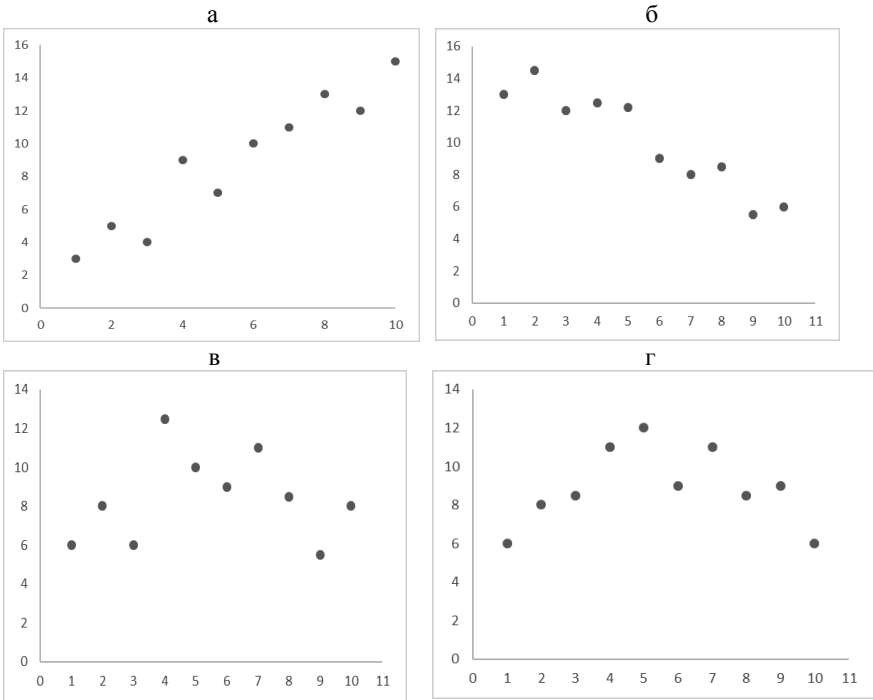


Рис. 6.1. Приклади кореляційних залежностей: а – пряма кореляція; б – зворотна кореляція; в – відсутність зв'язку між змінними; г – нелінійна кореляція

### Значущість коефіцієнта кореляції

Оскільки коефіцієнт кореляції розраховується для випадкових змінних, то і сам коефіцієнт кореляції також є випадковою величиною і потребує перевірки на значущість. Достовірність коефіцієнта кореляції (його відмінність від нуля) перевіряється за допомогою критерію Стьюдента. Розрахункове значення t-критерію знаходимо за формулою:

$$t_{\text{розр}} = \frac{r\sqrt{(N-2)}}{\sqrt{(1-r^2)}} ,$$

де  $r$  – значення коефіцієнта кореляції, а  $N$  – кількість спостережень.

Якщо розрахункове значення  $t_{\text{розр}}$  більше табличного, взятого з  $N-2$  ступенями вільності, нульова гіпотеза (рівність коефіцієнта кореляції нулю) відкидається.

Для розрахунку напівширини довірчого інтервалу для коефіцієнта кореляції використовуємо формулу:

$$\Delta = \frac{t_{N-2,p}(1-r^2)}{\sqrt{N}} ,$$

де  $N$  – число спостережень, за якими розраховується коефіцієнт кореляції;

$r$  – значення коефіцієнта кореляції;

$t_{N-2,p}$  – табличне значення критерію Стьюдента, взятого з  $N-2$  ступенями вільності.

Використовуючи кореляційний аналіз Пірсона, необхідно пам'ятати, що коефіцієнт кореляції показує тісноту тільки лінійного зв'язку. В тому випадку, коли залежності складніші, ніж лінійні, коефіцієнт кореляції буде показувати відсутність зв'язку. Наприклад, на рис. 6.1г добре видно, що між змінними існує залежність другого порядку. Розрахований коефіцієнт кореляції буде близький до нуля, і перевірка покаже, що він статистично незначущий. Тому для визначення складних залежностей між змінними використовуються інші статистичні методи, найбільш часто й ефективно – регресійний аналіз.

Варто також пам'ятати, що при наявності фізичного зв'язку між змінними наслідком зазвичай є наявність і кореляційного зв'язку. Однак, якщо вибірка не репрезентативна, тобто містить не ті реалізації випадкової величини, які дають змогу визначити залежність, коефіцієнт кореляції буде близький до нуля і незначний.

Наприклад, рис. 6.2 повторює рис. 6.1б, на якому присутній зворотний зв'язок із коефіцієнтом кореляції, близьким до  $-1$ . І лише в тому випадку, якщо вибірка буде містити виділені (великі) точки, які є підмножиною наявної вибірки, коефіцієнт парної кореляції буде близький до нуля.

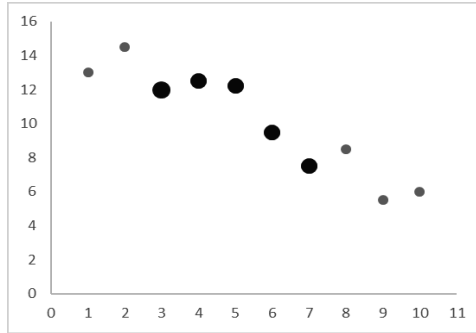


Рис. 6.2. Приклад неправильного формування підмножини даних

З іншого боку, наявність статистичного зв'язку не означає наявність фізичного. Причиною може бути непрезентабельність вибірки, як у попередньому випадку, або той факт, що досліджувана змінна  $X$  не залежить від змінної  $Y$ , але обидві вони залежать від змінної  $Z$ . У цьому випадку при проведенні кореляційного аналізу бачимо зв'язок між  $X$  і  $Y$ , який у фізичному сенсі відсутній.

За наявності кореляційного зв'язку, який є віддзеркаленням дійсно наявного фізичного зв'язку, необхідно робити правильні висновки.

**Приклад 6.1.** Взаємозв'язок між об'ємом циркулюючої крові (ОЦК), об'ємом циркулюючої плазми (ОЦП) і гематокритом (Н) виражається формулою:  $V_k = V_n / (1 - 100H)$ .

Отримано експериментальні дані. Необхідно визначити ступінь залежності між ними, використовуючи коефіцієнт кореляції.

	A	B	C	D
1	2	3	4	5
1		ОЦП	ОЦК	Н*100
2		36,8	52,5	30
3		43,5	62	30
4		33,4	43,4	30
5		34,5	41,5	17
6		46,8	64,1	27
7		31	56,1	45
8		29,6	47,7	38
9		35	55,5	37
10		29,6	35,6	17
11		50	56,8	12
12		46	56,8	19
13		49,3	58	15

1	2	3	4	5
14		27,4	31,9	14
15		27,2	38,3	29
16		37	55,3	33
17		33,6	62,2	46
18		35	45,6	23
19		36,2	53,2	32
20		39,6	59,5	33
21		44,8	61,3	27
22		37,1	48,8	24
23	Коефіцієнти кореляції для пар ознак		Розрахункове значення t-критерію:	
24	ОЦП і ОЦК	0,72		
25	ОЦП і Н	-0,34		
26	ОЦК і Н	0,38		
27	Критичне значення t-критерію:			
28	Напівширина довірчого інтервалу			

Вибираємо пункт меню *Дані*, вкладку *Analysis / Data analysis*; у діалоговому вікні, що відкрилося, вибираємо тип аналізу *Correlation (Кореляція)*, після чого з'являється вікно завдання вихідних даних кореляції. У ньому задаємо початкові дані для кореляційного аналізу:

*Input Range (Вхідний інтервал)* – необхідно виділити таблицю, в якій розміщені вихідні дані.

*Grouped By (Групування за)* – необхідно вказати, як розташовані дані, що належать до одного рівня фактору, – в стовпцях (*columns*) або рядках (*rows*) (в цій ситуації в стовпцях).

*Labels in first row (Мітки в першому рядку)* – ставиться прапорець, якщо разом із вихідними даними виділені назви стовпців.

*Output Range (Вихідний інтервал)* – вводиться посилання на комірку, розташовану в лівому верхньому кутку вихідного діапазону (місця, куди ви хочете помістити результат).

*New Worksheet Ply (Новий робочий лист)* – вибирається в разі, якщо ви хочете помістити результати роботи на інший аркуш.

*New Workbook (Нова робоча книга)* – вибирається в разі, якщо результати необхідно розмістити в нову робочу книгу.

Коефіцієнти парної кореляції можна отримати й іншим способом. Розмістимо коефіцієнти кореляції в комірках B24, B25, B26. Для цього скористаємося функціями:

=CORREL(B2:B22;C2:C22);

=CORREL(B2:B22;D2:D22);

=CORREL(C2:C22;D2:D22).

Для прийняття рішення про наявність значущого зв'язку між змінними необхідно перевірити значущість коефіцієнтів кореляції. Для цього для кожного коефіцієнта розраховуємо значення t-критерію. В комірках C24, C25, C26 поміщаємо формули:

$$\begin{aligned} &= B24 \times \text{SQRT}(\text{COUNT}(B2:B22) - 2) / \text{SQRT}(1 - B24 \times B24); \\ &= B25 \times \text{SQRT}(\text{COUNT}(C2:C22) - 2) / \text{SQRT}(1 - B25 \times B25); \\ &= B26 \times \text{SQRT}(\text{COUNT}(D2:D22) - 2) / \text{SQRT}(1 - B26 \times B26). \end{aligned}$$

Критичне (табличне) значення поміщаємо в комірку C27 за допомогою функції:

$$= \text{T.INV.2T}(0,05; \text{COUNT}(B2:B22) - 2),$$

де 0,05 – рівень значущості;

COUNT(B2:B22) – 2 – число ступенів вільності.

*Оскільки розрахункове значення t-критерію більше критичного тільки для коефіцієнта кореляції між ОЦП і ОЦК (4,78 проти 2,09), то цей коефіцієнт є значущим.*

У низці випадків для оцінки коефіцієнта кореляції необхідно розрахувати довірчий інтервал. Спочатку визначаємо напівширину довірчого інтервалу (комірка C28) за формулою:

$$= C27 \times (1 - B24 \times B24) / \text{SQRT}(\text{COUNT}(B2:B22))$$

Потім знайдемо його нижню і верхню межу.

$$= B23 - C28 - \text{нижня межа};$$

$$= B23 + C28 - \text{верхня межа}.$$

### **Приватна (часткова) кореляція**

У біостатистиці одновимірні статистичні тести, як-от тест  $\chi^2$ , F-критерій Фішера, t-тест та аналіз дисперсії, не дають змогу враховувати під час аналізу вплив інших факторів. Однак методика, яка називається частковою кореляцією, дозволяє досліднику контролювати вплив третьої змінної на зв'язок між двома вибраними змінними. Часткова кореляція розглядає зв'язок між двома змінними, видаляючи при цьому вплив інших змінних. У такому разі коефіцієнт кореляції між змінними X та Y розраховується за формулою:

$$r = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}},$$

де  $r_{12}, r_{13}, r_{23}$  – коефіцієнти парної кореляції між змінними  $X$  і  $Y$ ,  $X$  і  $Z$ ,  $Y$  і  $Z$  відповідно.

У прикладі 6.1 коефіцієнт кореляції між ОЦП і ОЦК не дуже великий, хоча між ними існує функціональний зв'язок. Можливо, це обумовлено впливом на дані третьої змінної. Спробуємо отримати коефіцієнт кореляції між ОЦП і ОЦК, «очистивши» його від впливу  $H$ . Для цього набираємо формулу:

$$= (B23 - C23 \times D23) / \text{SQRT}((1 - C23 \times C23) \times (1 - D23 \times D23)).$$

Внаслідок отримаємо значення приватного коефіцієнта кореляції між ОЦП і ОЦК – 0,98.

### Рангова кореляція

Аналогом парної кореляції Пірсона для даних, які не розподілені за нормальним законом, є *рангова кореляція*.

Коефіцієнти кореляції називаються ранговими, оскільки перед обчисленням значення змінних перетворюють у ранги. Для цього наявні значення змінних ранжирують. У кожній вибірці значенням присвоюються ранги від 1 до  $n$ , де  $n$  – кількість спостережень.

Здебільшого використовують *коефіцієнт кореляції Спірмена ( $r_s$ )* і *коефіцієнт Кендала ( $\tau$ )*.

### Коефіцієнт рангової кореляції Спірмена

Коефіцієнт Спірмена обчислюється за формулою:

$$r_s = 1 - \frac{6 \sum_{i=1}^n (R_{1i} - R_{2i})^2}{n^3 - n},$$

де  $R_{1i}$  і  $R_{2i}$  – ранги  $i$ -го об'єкта для кожної з порівнюваних змінних;

$n$  – кількість пар спостережень. Значення  $r_s$  не залежить від того, як впорядковані ранги – за збільшенням чи за зменшенням.

Перевірку значущості коефіцієнта рангової кореляції здійснюють за допомогою  $t$ -критерія Стьюдента, як і для коефіцієнта парної кореляції Пірсона.

**Приклад 6.2.** Необхідно перевірити, чи існує статистично значущий зв'язок між двома факторами (ознаками), які є параметрами гомеостазу крові, отриманої з ліктьової вени у хворих зі стабільною стенокардією. Візьмемо фактори, що

визначають функціональну активність тромбоцитів: тромбопластиновий ( $\Phi_3$ ) і антигепариновий ( $\Phi_4$ ). Вихідні дані і результати обчислень наведено на рис. 6.3

	A	B	C	D	E
1	$\Phi_3$	$\Phi_4$	R1	R2	$(R1-R2)^2$
2	55	72	11	4	49
3	24	77	2	5,5	12,25
4	40	85	8	11	9
5	60	90	12	13	1
6	39	82	7	9	4
7	28	77	4	5,5	2,25
8	22	60	1	2	1
9	37	79	6	7	1
10	72	88	13	12	1
11	42	80	10	8	4
12	33	66	5	3	4
13	41	83	9	10	1
14	25	46	3	1	4
15	NO	NOR	NORM		
16	Сума квадратів різниці рангів				93,5
17	Розмір вибірки				13
18	Коефіцієнт рангової кореляції				0,74
19	Рівень значущості				0,05
20	Критичне (табличне) значення				0,48

Рис. 6.3. Дані для прикладу 6.2 і його розв'язання

1. Перевіряємо обидві вибірки на відповідність їх значень нормальному закону розподілу.
2. Будуємо ранги для першої і другої вибірок.
3. Отримуємо квадрати різниць рангів першої і другої вибірок (стовпець F).
4. Підсумовуємо квадрати різниць за стовпчиком F.
5. Обчислюємо коефіцієнт рангової кореляції Спірмена.
6. Перевіряємо отриманий коефіцієнт рангової кореляції Спірмена на його статистичну значущість. Якщо обчислений коефіцієнт рангової кореляції більший за критичне (табличне) значення, то нульова гіпотеза відкидається. Тобто аналізовані нами вибірки пов'язані досить тісно.

### Кореляція Кендала

Рангова кореляція Кендала вимірює силу та напрям наявного зв'язку (визначає, чи існує монотонна залежність) між двома змінними. Тому цей вид кореляції застосовується в тому разі, якщо дані змінюються монотонно, тобто зі збільшенням значення однієї змінної зростає інша змінна, або зі збільшенням значення однієї змінної інша зменшується.

Розрахунки коефіцієнта кореляції Кендала ґрунтуються на узгоджених і не-узгоджених парах ранжированих змінних  $(R1_i, R2_i) \dots (R1_n, R2_n)$ . Пара спостережень

вважається узгодженою, якщо для пар  $(R_{1i}, R_{2i})$  і  $(R_{1j}, R_{2j})$  при  $i < j$  виконується умова:  $R_{1i} < R_{1j}$  та  $R_{2i} < R_{2j}$ . І навпаки, пара вважається неузгодженою, якщо  $R_{1i} < R_{1j}$  та  $R_{2i} > R_{2j}$ .

Коефіцієнт кореляції Кендала обчислюється за формулою:

$$\tau = (C - D) / N,$$

де  $C$  – кількість узгоджених пар;

$D$  – кількість неузгоджених пар;

$N = n(n - 1) / 2$  – кількість можливих комбінацій  $i$  та  $j$  для  $n$  спостережень.

Розглянемо розрахунок коефіцієнта кореляції Кендала для даних прикладу 6.2.

1. Розраховуємо ранги для першої і другої вибірок за допомогою функції Excel. У комірку C2 вводимо формулу: RANK.AVG(A2;A\$2:A\$14). Результат розрахунку копіюємо в комірки C3:C14 і D2:D14.

2. Скопіюємо отримані ранжирувані вибірки в стовпці F і G і відсортуємо їх відносно стовпця F (ранг R1) у порядку зростання.

3. Підрахуємо кількість узгоджених пар. Для цього в комірку H2 вводимо формулу: =COUNTIFS(F3:F\$14;">"&F2;G3:G\$14;">"&G2), як показано на рис. 6.4. Результат розрахунку скопіюємо в комірки H3:H14.

H2		=COUNTIFS(F3:F\$14;">"&F2;G3:G\$14;">"&G2)											
	A	B	C	D	F	G	H	I					
	Ф_3	Ф_4	R1	R2	R1 відсортовані	R2 відсортовані	Кількість узгоджених пар	Кількість неузгоджених пар					
1													
2	55	72	3	10	1	2	11						
3	24	77	12	8,5	2	1	11						
4	40	85	6	3	3	10	3						
5	60	90	2	1	4	6	6						
6	39	82	7	5	5	4	7						
7	28	77	10	8,5	6	3	7						
8	22	60	13	12	7	5	6						
9	37	79	8	7	8	7	5						
10	72	88	1	2	9	11	2						
11	42	80	4	6	10	8,5	2						
12	33	66	9	11	11	13	0						
13	41	83	5	4	12	8,5	1						
14	25	46	11	13	13	12	0						
15					Сума узгоджених пар		61						
16					Сума неузгоджених пар			16					
17					Кількість комбінацій для заданої кількості спостережень			78					
18					Коефіцієнт кореляції Кендала			0,576923077					

Рис. 6.4. Розрахунок коефіцієнта кореляції Кендала для даних прикладу 6.2

4. Підрахуємо кількість неузгоджених пар. Для цього в комірку I2 вводим формулу: =COUNTIFS(F3:F\$14;">"&F2;G3:G\$14;"<"&G2). Результат розрахунку скопіюємо в комірку I3:I14.

5. Розрахуємо кількість узгоджених і кількість неузгоджених пар.

6. За формулою, наведеною вище, розрахуємо кількість можливих комбінацій пар для заданої кількості спостережень. У цьому прикладі  $N = 78$ .

7. З кількістю узгоджених та неузгоджених пар розрахуємо коефіцієнт Кендала. У цьому прикладі  $\tau = 0,58$ . Зазвичай, коефіцієнт кореляції Кендала має менші значення, ніж кореляція Спірмена.

### Пошук кореляції в програмі Statistica. Лінійна кореляція

**Приклад 6.3.** Встановити, чи існує взаємозв'язок між об'ємом циркулюючої крові (ОЦК), об'ємом циркулюючої плазми (ОЦП) і гематокритом (H).

Використовуємо дані прикладу 6.1 і вводимо їх у програму Statistica. Переименовуємо заголовки стовпців: Var1 – ОЦК, Var2 – ОЦП, Var3 – H. Для обчислення коефіцієнта кореляції використовуємо команди: *Statistics, Basic statistics / Tables, Correlation matrices (Статистика, Основна статистика / Таблиці, Кореляційні матриці)*. В діалоговому вікні *Product-Moment and Partial Correlations (Парні і приватні кореляції)* натискаємо кнопку *One variable list (Квадратна матриця)*, розташовану у верхньому лівому кутку вікна (рис. 6.4).

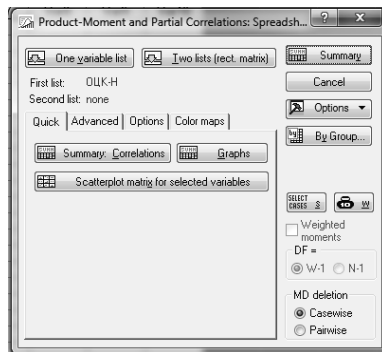


Рис. 6.4. Діалогове вікно розрахунку коефіцієнтів кореляції

У вікні *Select the Variables for the Analysis (Вибір змінних для аналізу)*, відзначаємо, для яких змінних будуть розраховані парні коефіцієнти кореляції, після чого натискаємо *OK*.

Повертаємося у вікно *Product-Moment and Partial Correlations (Парні і приватні кореляції)* і переходимо на вкладку *Advanced*. Для виведення таблиці з

коефіцієнтами кореляції потрібно натиснути кнопку *Summary: Corrs* (Підсумок: Матриця кореляцій) (рис. 6.5).

Correlations (Spreadsheet1)				
Marked correlations are significant at p < .05000				
N=21 (Casewise deletion of missing data)				
Variable	ОЦК	ОЦП	Н*100	
ОЦК	1,000000	0,722218	-0,341688	
ОЦП	0,722218	1,000000	0,385497	
Н*100	-0,341688	0,385497	1,000000	

Рис. 6.5. Результати розрахунку коефіцієнтів парної кореляції в програмі Statistica

Достовірні коефіцієнти кореляції в таблиці виділяються червоним кольором. У такий спосіб встановили, що між обсягом циркулюючої крові і плазми наявна достовірна лінійна залежність.

Щоб представити отримані результати у вигляді лінійної залежності, натиснемо кнопку *2D scatterplots (2М графіки)*. У діалоговому вікні, що відкрилося, задаємо, що по осі абсцис будуть відкладатися значення ОЦК (*first (horizontal) variable list*), а по осі ординат – значення ОЦП (*second (vertical) variable list*). Залежність між ОЦК і ОЦП повинна мати вигляд, як на рис. 6.6.

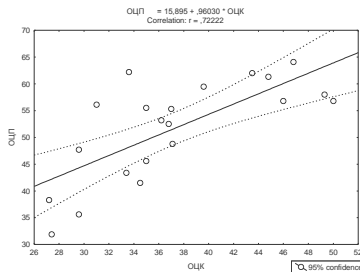


Рис. 6.6. Графік залежності ОЦП від ОЦК із визначеними коефіцієнтами кореляції та довірчим інтервалом

### Непараметрична кореляція

**Приклад 6.4.** Введіть дані прикладу 6.2 у програму Statistica. Поміняйте заголовки стовпців відповідно до умов прикладу. Виконайте послідовно такі команди: *Statistics, Nonparametrics, Correlation Spirman, Kendall tau, gamma* (Статистика, Непараметричний, Кореляція Спірмена, тау Кендала, гамма).

У вікні *Nonparametric correlation* (Непараметрична кореляція) натискаємо кнопку *Variables* і вказуємо змінні, між якими будемо шукати кореляційний зв'язок.

Натискаємо *OK* для того, щоб повернутися у попереднє діалогове вікно. Для знаходження коефіцієнта кореляції Спірмена натискаємо кнопку *Spearman rank R*. Достовірні коефіцієнти кореляції в таблиці виділяються червоним кольором. Порівняйте значення коефіцієнта кореляції Спірмена, розраховані в програмі Excel та Statistica.

### **Завдання для самостійної роботи**

1. Знайти коефіцієнти кореляції Пірсона або Спірмена для даних, наведених в розділі «Варіанти завдань для закріплення теми 6», у програмах Excel і Statistica.
2. У файлі *Аскаридоз.xls* представлені дані про стан імунної системи у хворих аскаридозом та ентеробіозом. Проаналізувати залежність вмісту циклічних імунних комплексів від віку, діагнозу та статі хворого.
3. У файлі *Страх.xls* представлені результати психологічного тестування школярів на предмет впливу страху і тривожності на самооцінку школяра і його статус у колективі. Проаналізувати залежність між цими показниками для всієї вибірки.
4. Відомо, що для здорового організму необхідний тісний взаємозв'язок між активністю окисно-відновних ферментів. Показати, чи впливає свинцева інтоксикація на взаємозв'язок між ферментами. Дані знаходяться у файлі *Ферменти.xls*.

### **Контрольні запитання до теми 6**

1. Що таке кореляційний аналіз і для чого застосовується цей метод аналізу даних? Наведіть приклади кореляційних залежностей. Які значення може приймати коефіцієнт кореляції?
2. В яких випадках застосовують кореляцію Пірсона? Наведіть формулу для розрахунку коефіцієнта кореляції Пірсона.
3. Що є аналогом парної кореляції Пірсона і для яких даних застосовують цей метод аналізу?
4. Для чого використовується приватна кореляція? Наведіть приклади та формулу для розрахунку приватної кореляції.
5. На чому ґрунтуються розрахунки коефіцієнта кореляції Кендала? В якому випадку використовують цей коефіцієнт?
6. За допомогою якого критерію визначається значущість коефіцієнта кореляції?

## Тема 7. Аналіз таблиць спряженості

У біологічних та екологічних дослідженнях велику роль відіграє аналіз таблиць спряженості. Ці методи використовують для аналізу даних, про властивість яких можна лише сказати наявна вона чи ні (наприклад, табл. 7.1).

Таблиця 7.1. Показники смертності від раку легень та ішемічної хвороби серця (на 100 тис. осіб на рік)

Захворювання	Палять	Не палять
Рак легенів	48,33	4,49
Ішемічна хвороба серця	394,67	109,54

Нульова гіпотеза полягає в тому, що показник смертності від зазначених захворювань не залежить від того, палить людина чи ні. Якщо нульова гіпотеза буде відкинута, то це означає, що між курінням і смертністю від вказаних захворювань існує статистично значущий зв'язок.

### Чотириклітинкові таблиці

У загальному вигляді чотириклітинкова таблиця (таблиця  $2 \times 2$ ) має такий вигляд (табл. 7.2).

Таблиця 7.2. Загальний вигляд чотириклітинкової таблиці спряженості

Вибірка	Наявність ознаки	Відсутність ознаки	Всього
Перша вибірка	A	B	$n_1 = A + B$
Друга вибірка	C	D	$n_2 = C + D$
Всього	$A + C$	$B + D$	$n = n_1 + n_2$

Нульова гіпотеза про належність обох вибірок до однієї генеральної сукупності виконується з використанням критерію  $\chi^2$ , який розраховується за формулою:

$$\chi^2 = \frac{n(AD - BC)^2}{(A + B)(C + D)(A + C)(B + D)}.$$

Для маленьких вибірок замість  $n$  беруть  $(n-1)$ . Розраховане значення критерію  $\chi^2$  порівнюється з критичним, узятим з одним ступенем вільності і заданим рівнем значущості. Якщо розрахункове значення більше критичного, то приймається гіпотеза про наявність між досліджуваними ознаками істотного зв'язку.

Правильність отриманих висновків залежить від того, як були вибрані дані. Вибірка повинна бути однорідна щодо аналізованої ознаки. Наприклад, якщо до вибірки, що аналізується, входять одночасно люди, на яких препарат має позитивний вплив, та люди, на яких він діє негативно, можна отримати висновок, що препарат не чинить ніякого впливу. А це не відповідає реальному стану речей.

Варто пам'ятати, що обсяг вибірок не повинен бути занадто маленьким. Наприклад, для рівня значущості 0,05 мінімальна кількість спостережень у кожній вибірці повинна бути 124.

**Приклад 7.1.** Розглянемо аналіз на прикладі табл. 7.1. Для розв'язання задачі про перевірку наявності зв'язку між курінням і смертністю від раку легенів та ішемічної хвороби серця виконаємо такі дії:

1. Знайдемо суми за рядками і стовпцями, для чого введемо формули підсумовування у відповідні клітинки.

Рядок / Стовпець	C	D	E
3	48,33	4,49	=SUM(C3:D3)
4	394,67	109,54	=SUM(C4:D4)
5	=SUM(C3:C4)	=SUM(D3:D4)	=SUM(E3:E4)

1. Визначимо розрахункове значення  $\chi^2$  за допомогою формули:

$$= (E5 \times (C3 \times D4 - D3 \times C4) \times (C3 \times D4 - D3 \times C4)) / (E3 \times E4 \times C5 \times D5).$$

2. Визначимо табличне значення  $\chi^2$ , ввівши в комірку функцію:

$$=CHIQ.INV.RT(0,05; 1).$$

3. Виконаємо порівняння розрахункового і табличного значення. Оскільки розрахункове значення (5,135995) більше критичного (3,841455), з рівнем значущості 0,05 нульова гіпотеза відхиляється. Отже, існує статистично значущий зв'язок між курінням і смертністю від зазначених захворювань.

### Таблиці виду $2 \times k$

Таблиця спряженості типу  $2 \times k$  має загальний вигляд, як таблиця 7.3.

Для перевірки нульової гіпотези про однорідність  $k$  вибірок використовується формула, запропонована Брандтом і Снедекором:

$$\chi^2 = \frac{n^2}{m(n-m)} \left[ \sum_{i=1}^k \frac{m_i^2}{n_i} - \frac{m^2}{n} \right].$$

Таблиця 7.3. Загальний вигляд таблиці спряженості виду  $2 \times k$ 

№ вибірки або ознаки другого рівня	Перша ознака:		Всього
	Наявність ознаки	Відсутність ознаки	
1	$m_1$	$n_1 - m_1$	$n_1$
2	$m_2$	$n_2 - m_2$	$n_2$
...	...	...	...
$k$	$m_k$	$n_k - m_k$	$n_k$
$\Sigma$	$m$	$n - m$	$n$

Розрахункове значення порівнюється з критичним, узятим з  $(k - 1)$  ступеня вільності і заданим рівнем значущості. Якщо розрахункове значення більше критичного, то гіпотезу про однорідність необхідно відкинути і прийняти гіпотезу про наявність між досліджуваними ознаками істотного зв'язку.

**Приклад 7.2.** Розглянемо дані, в яких представлено залежність зараженості населення бруцельозом типу *Suis* від частоти контактів із тваринами (табл. 7.4).

Таблиця 7.4. Зараженість населення бруцельозом типу *Suis* залежно від частоти контактів із тваринами

Група обстежених	Кількість обстежених		Всього
	З позитивною реакцією	З негативною реакцією	
Працювали у свинарнику	19	42	61
Мали епізодичні контакти з тваринами	23	71	94
Без контактів із тваринами	23	227	250
Всього	65	340	405

Необхідно встановити, чи є зв'язок між рівнем зараженості бруцельозом і ступенем контактів населення з тваринами. В табл. 7.5 показано, як за допомогою операцій і функцій Excel розрахувати попередні дані, які необхідні для визначення значення критерію  $\chi^2$ .

Таблиця 7.5. Підготовчі розрахунки

Рядок / Стовець	C	D	E	F
3	19	42	=SUM(C3:D3)	=C3*C3/E3
4	23	71	=SUM(C4:D4)	=C4*C4/E4
5	23	227	=SUM(C5:D5)	=C5*C5/E5
6	=SUM(C3:C5)	=SUM(C3:C5)	=SUM(C6:D6)	=SUM(F3:F5)

Після виконання всіх перерахованих дій маємо необхідну інформацію для обчислення розрахункового значення  $\chi^2$ . Для цього поміщаємо в комірку формулу  $=E6 \times (F6 - C6 \times C6 / E6) / (C6 \times (E6 - C6))$  (23,96).

Далі отримуємо критичне значення (відсоткову точку) розподілу  $\chi^2$ . У наступну комірку поміщаємо виклик функції  $=CHIQ.INV.RT(0,05;3-1)$ , де 0,05 – рівень значущості, 3 – кількість різних вибірок.

Оскільки  $23,96 > 5,99$ , нульова гіпотеза про відсутність зв'язку відхиляється. Отже, можемо стверджувати, що між ступенем контактів населення з тваринами і захворюваністю бруцельозом існує статистично значущий зв'язок.

### Таблиці виду $k \times l$

Таблиці виду  $k \times l$  (табл. 7.6) є найбільш загальним видом таблиць спряженості. При цьому використовуються градації вираженості ознаки.

Таблиця 7.6. Загальний вигляд таблиці спряженості виду  $k \times l$

Друга ознака	Перша ознака						Суми за рядками
	1	2	...	$j$	...	$l$	
1	$n_{11}$	$n_{12}$	...	$n_{1j}$	...	$n_{1l}$	$n_{1.}$
2	$n_{21}$	$n_{22}$	...	$n_{2j}$	...	$n_{2l}$	$n_{2.}$
...	...	...	...	...	...	...	...
$i$	$n_{i1}$	$n_{i2}$	...	$n_{ij}$	...	$n_{il}$	$n_{i.}$
...	...	...	...	...	...	...	...
$k$	$n_{k1}$	$n_{k2}$	...	$n_{kj}$	...	$n_{kl}$	$n_{k.}$
суми за стовпцями	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.l}$	$n$

Наприклад, першою ознакою можуть бути різні види лікування: симптоматичне, специфічне з нормальними дозами, специфічне з підвищеними дозами або специфічне. Другою ознакою – одужання за 2 тижні, одужання за 4 тижні, летальний результат. Друга ознака може бути також сукупністю різних вибірок.

У такому разі застосовують один критерій для перевірки гіпотези про незалежність ознак та однорідності вибірок. Для випадку, коли перший стовпець табл. 7.6 являє собою  $k$  значень рівня другої ознаки, перевіряють гіпотезу про незалежність першої і другої ознак. Якщо ж перший стовпець містить  $k$  різних вибірок, то перевіряють гіпотезу про однорідність цих вибірок, і можна вважати, що ці вибірки витягнуті з однієї генеральної сукупності.

Значення критерію розраховується за формулою:

$$\chi^2 = n \left[ \sum_{i=1}^k \sum_{j=1}^l \frac{n_{ij}^2}{n_i n_j} - 1 \right].$$

Розрахункове значення порівнюється з критичним, узятим з  $(k - 1) \times (l - 1)$  ступенями вільності і заданим рівнем значущості. Якщо розрахункове значення більше критичного, то гіпотезу про однорідність слід відкинути і прийняти гіпотезу про наявність зв'язку між досліджуваними ознаками.

**Приклад 7.3.** Розглянемо задачу про розподіл фракцій ПВП (полівінілпіролідону) у препараті різних виробників:

Виробник препарату	Площа піків фракцій ПВП		
	Високомолекулярні, $M > 160\,000$	З номінальною $M$ , $M > 8\,000 + 2\,000$	Низькомолекулярні $M > 4\,500$
1	15,4	80,1	4,5
2	16,7	78,3	4,8
3	31,9	65,3	2,5
4	32,4	64,3	3,3
5	20,7	77,0	2,4
6	37,9	58,4	3,0
7	29,6	67,8	2,4

- Обчислюємо суми за рядками і стовпчиками.
- Формуємо допоміжну таблицю такого ж розміру, як вихідна. Кожна комірка буде містити квадрат ознаки, поділений на добуток сум рядка та стовпця, в яких знаходиться значення ознаки.
- Отримуємо розрахункове значення критерію  $\chi^2$ .
- Визначаємо критичне значення  $\chi^2$  за допомогою функції  $=\text{CHI.Q.INV.RT}(0,05;(3-1)*(7-1))$ .
- Оскільки розрахункове значення (25,06) більше критичного (21,02), то нульова гіпотеза відхиляється. Це означає, що за якістю між препаратами представлених виробників існують статистично значущі відмінності.

### Завдання для самостійної роботи

- Експериментальні дані представлені у розділі «Варіанти завдань для закріплення теми 7» у вигляді таблиць спряженості. Встановити за допомогою програми Excel, чи існує статистично достовірний зв'язок між зазначеними параметрами.

### Контрольні запитання до теми 7

- В яких випадках використовуються таблиці спряженості? Яка нульова гіпотеза перевіряється і за допомогою якого критерію?
- Наведіть приклади використання різних типів таблиць спряженості.

## ТЕМА 8. ДИСПЕРСІЙНИЙ АНАЛІЗ

Дисперсійний аналіз – це статистичний метод, який використовується для перевірки впливу одного або кількох факторів шляхом дослідження значущості відмінностей у середніх значеннях двох або більше груп. Треба зазначити, що дисперсійний аналіз оперує дисперсіями (а не середніми) для перевірки частки відхилень в наборі даних, які пояснюються груповими відмінностями.

В основі дисперсійного аналізу лежать три важливі припущення:

- вибірки повинні бути випадковими та незалежними;
- набір даних має бути нормально розподілений;
- дисперсія однорідна за групами.

### Параметричний дисперсійний аналіз

#### Однофакторна задача

Для найпростішого випадку таблиця вихідних даних має вигляд (табл. 8.1):

Таблиця 8.1. Загальний вигляд вихідних даних для однофакторної задачі

Номер сукупностей	Номер елементів сукупностей					
	1	2	...	$j$	...	$n$
1	$X_{11}$	$X_{12}$	...	$X_{1j}$	...	$X_{1n}$
2	$X_{21}$	$X_{22}$	...	$X_{2j}$	...	$X_{2n}$
...	...	...	...	...	...	...
$i$	$X_{i1}$	$X_{i2}$	...	$X_{ij}$	...	$X_{in}$
...	...	...	...	...	...	...
$m$	$X_{m1}$	$X_{m2}$	...	$X_{mj}$	...	$X_{mn}$

Це може бути, наприклад,  $m$  розчинів різної концентрації, з кожного з яких взято  $n$  зразків. Необхідно з'ясувати, як змінюються показники розчинів залежно від концентрації.

Можуть розглядатися якісь характеристики лабораторних тварин ( $m$  груп по  $n$  тварин у кожній). Щоб з'ясувати, чи відрізняються їхні характеристики від групи до групи, потрібно порівняти дисперсію, обумовлену випадковими причинами, з дисперсією, що викликається наявністю деякого фактора. Якщо дисперсії значно різняться, то фактор має статистично значущий вплив на досліджувану змінну.

Відмінність вважається значущою, якщо розраховане значення критерію Фішера буде більше табличного, взятого із заданим рівнем значущості і ступенями вільності  $(m - 1)$  і  $(n - m)$ .

Загальна сума квадратів відхилень значень параметру від середнього визначається за формулою:

$$SS_{заг} = \sum_{j=1}^m \sum_{i=1}^n X_{ij}^2 - \frac{\left( \sum_{j=1}^m \sum_{i=1}^n X_{ij} \right)^2}{n},$$

де  $X_{ij}$  – значення відповідного спостереження;

$m$  – кількість вибірок;

$n$  – загальна кількість спостережень.

*Міжгрупова* сума квадратів відхилень (викликана впливом фактору) визначається за формулою:

$$SS_{фактор} = \sum_{j=1}^m \frac{\left( \sum_{i=1}^{n_j} X_{ij} \right)^2}{n_j} - \frac{\left( \sum_{j=1}^m \sum_{i=1}^n X_{ij} \right)^2}{n}.$$

*Залишкова* (внутрішньогрупова) сума квадратів знаходиться як різниця між загальною і факторною сумами

$$SS_{залишк} = SS_{заг} - SS_{фактор}.$$

Потім знаходяться залишкова (внутрішньогрупова) і факторна (міжгрупова) дисперсії і розрахункове значення критерію Фішера:

$$S_{фактор}^2 = SS_{фактор} / (m - 1)$$

$$S_{залишк}^2 = SS_{залишк} / (n - m)$$

$$F_{розр} = \frac{S_{фактор}^2}{S_{залишк}^2}.$$

Критичне (табличне) значення Фішера можна розрахувати за допомогою функції Excel:

$$F_{крит} = F.INV.RT(0,05, m-1, n-m).$$

Якщо  $F_{розр} > F_{крит}$ , то приймається гіпотеза про вплив фактору.

**Приклад 8.1.** Дослідження електропровідності в стандартній точці акупунктури, розташованій в області кисті Тай-юань проводили у чоловіків в положенні «сидячи», з використанням спеціального обладнання. Для оцінки стійкості зміни досліджуваних показників вимірювання проводили повторно на першій, п'ятій,

десятій хвилинах. Встановити, чи залежить величина електропровідності в цій точці від часу вимірювання.

Електропровідність			Електропровідність		
перша хв	п'ята хв	десята хв	перша хв	п'ята хв	десята хв
6	6	6	5	5	5
20	20	8	20	20	17
13	10	11	5	4	8
13	14	7	17	10	7
3	5	4	2	3	5
7	3	5	6	3	4
9	9	10	10	6	11
9	11	10	11	6	7
3	3	4	2	3	4
20	20	11	20	20	17

Вихідні дані і результати розрахунків у програмі Excel представлені на рис. 8.1 і в табл. 8.2.

	A	B	C	D	E	F	G
1		Електропровідність ( $X_{ij}$ )			$(X_{ij})^2$		
2		Перша	П'ята	Десята			
3		5	5	5	25	25	25
4		20	20	17	400	400	289
5		5	4	8	25	16	64
6		17	10	7	289	100	49
7		2	3	5	4	9	25
8		6	3	4	36	9	16
9		10	6	11	100	36	121
10		11	6	7	121	36	49
11		2	3	4	4	9	16
12		20	20	17	400	400	289
13		6	6	6	36	36	36
14		20	20	8	400	400	64
15		13	10	11	169	100	121
16		13	14	7	169	196	49
17		3	5	4	9	25	16
18		7	3	5	49	9	25
19		9	9	10	81	81	100
20		9	11	10	81	121	100
21		3	3	4	9	9	16
22		20	20	11	400	400	121
23	$\Sigma X_{ij}$	201	181	161	→ Загальна сума		543
24	$n_{ij}$	20	20	20	→ Загальна сума		60
25	$(\Sigma X_{ij})^2/n_{ij}$	2020,05	1638,1	1296,1	→ Загальна сума		4954,2
26	$(\Sigma X_{ij})^2$	294849			$(\Sigma X_{ij})^2/n$		4914,2
27	$\Sigma(X_{ij})^2$	6815					

Рис. 8.1. Вихідні дані та допоміжні розрахунки прикладу 8.1.

Таблиця 8.2. Результати розрахунку прикладу 8.1 у програмі Excel

Компоненти дисперсії	Сума квадратів	Число ступенів вільності	Середній квадрат (дисперсія)	$F_{розр}$	$F_{теор}$
Загальна	1900,85	59	32,22	0,62	3,16
Міжгрупова (впливає фактор)	40,00	2	20,00		
Внутрішньогрупова (випадковий вплив)	1860,85	57	32,65		

За даними таблиці бачимо, що величина електропровідності в точці акупунктури залежить від часу вимірювання.

Розрахунки, наведені вище, можна виконати спеціальною функцією, яку викликають за допомогою пакету аналізу даних. У вікні *Data analysis (Аналіз даних)* вибираємо пункт *ANOVA: Single Factor (Однофакторний дисперсійний аналіз)*.

У діалоговому вікні, що з'явилося, вводимо:

*Input Range* – інтервал вхідних даних;

*Grouped By* – групування за стовпцями (*columns*) або рядками (*rows*) даних (у наведеному прикладі – за стовпцями);

*Labels in first row (Мітки в першому рядку)* – ставиться прапорець, якщо разом із вихідними даними виділені назви стовпців;

*Alpha* – рівень значущості (0,05);

*Output Range (Вихідний інтервал)* – вводиться посилання на комірку, починаючи з якої буде виведено результат. Після натискання кнопки ОК результат з'явиться в робочому полі вікна (рис. 8.2).

ANOVA						
Source of Variation	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	40	2	20	0,61	0,55	3,16
Within Groups	1861	57	32,65			
Total	1901	59				

Рис. 8.2. Результат однофакторного дисперсійного аналізу, проведеного за допомогою пакету аналізу даних в Excel

Порівняйте результати, отримані різними способами. Зверніть увагу на рівень значущості (*p-value*). Оскільки  $p = 0,55 (>0,05)$ , різниці між вибірками немає, отже, величина електропровідності в досліджуваній точці акупунктури не залежить від часу вимірювання.

### Двофакторна задача з рівномірним числом спостережень

Двофакторний дисперсійний аналіз передбачає наявність двох факторів різних рівнів, що вимірюються в шкалі найменувань і впливають на спостережуваний параметр. Загальний вигляд представлення даних з однаковим числом повторюваних значень у комірці наведено в табл. 8.3 або з обчисленими середніми значеннями досліджуваного параметру – в табл. 8.4.

Таблиця 8.3. Представлення даних у двофакторному дисперсійному аналізі

Фактор $A$	Фактор $B$			
	$B_1$	$B_2$	...	$B_n$
$A_1$	$X_{111}, X_{112}, \dots, X_{11k}$	$X_{121}, X_{122}, \dots, X_{12k}$		$X_{1n1}, X_{1n2}, \dots, X_{1nk}$
$A_2$	$X_{211}, X_{212}, \dots, X_{21k}$	$X_{221}, X_{222}, \dots, X_{22k}$		$X_{2n1}, X_{2n2}, \dots, X_{2nk}$
...	...	...	$X_{ij1}, X_{ij2}, \dots, X_{ijk}$	...
$A_m$	$X_{m11}, X_{m12}, \dots, X_{m1k}$	$X_{m21}, X_{m22}, \dots, X_{m2k}$		$X_{mn1}, X_{mn2}, \dots, X_{mnk}$

Таблиця 8.4. Представлення даних у дисперсійному аналізі з усередненими значеннями параметру

Фактор $A$	Фактор $B$				
	$B_1$	$B_2$	...	$B_n$	$\bar{X}_{i**}$
$A_1$	$\bar{X}_{11*}$	$\bar{X}_{12*}$		$\bar{X}_{1n*}$	$\bar{X}_{1**}$
$A_2$	$\bar{X}_{21*}$	$\bar{X}_{22*}$		$\bar{X}_{2n*}$	$\bar{X}_{2**}$
...	...	...	$\bar{X}_{ij*}$	...	$\bar{X}_{j**}$
$A_m$	$\bar{X}_{m1*}$	$\bar{X}_{m2*}$		$\bar{X}_{mn*}$	$\bar{X}_{m**}$
$\bar{X}_{*j*}$	$\bar{X}_{*1*}$	$\bar{X}_{*2*}$		$\bar{X}_{*n*}$	$\bar{X}$

Тут  $i = 1 \div m$  – номери рядків;  $j = 1 \div n$  – номери стовпців;  $l = 1 \div k$  – номери спостережуваних значень у межах однієї комірки.  $X_{111}, X_{112}, \dots, X_{mnk}$  – спостережувані значення досліджуваної змінної;

$$\bar{X}_{ij*} = \frac{1}{k} \sum_{l=1}^k X_{ijl} \text{ – середнє значення в комірці;}$$

$$\bar{X}_{i**} = \frac{1}{n} \sum_{j=1}^n \bar{X}_{ij*} \text{ – середнє значення за рядком;}$$

$$\bar{X}_{*j*} = \frac{1}{m} \sum_{i=1}^m \bar{X}_{ij*} \text{ – середнє значення за стовпцем;}$$

$$\bar{X} = \frac{1}{m} \sum_{i=1}^n \sum_{j=1}^m \bar{X}_{ij^*} - \text{загальне середнє.}$$

Формули для подальших розрахунків наведено в таблиці 8.5.

Таблиця 8.5. Форма представлення результатів  
двофакторного дисперсійного аналізу

Компоненти дисперсії	Сума квадратів	Число ступенів вільності	Середній квадрат (дисперсія)
Між середніми за рядками (за фактором А) $S_A^2$	$nk \sum_{i=1}^m (\bar{X}_{i^{**}} - \bar{X})^2$	$n - 1$	$\frac{nk}{n-1} \sum_{i=1}^m (\bar{X}_{i^{**}} - \bar{X})^2$
Між середніми за рядками (за фактором В) $S_B^2$	$mk \sum_{j=1}^m (\bar{X}_{*j^*} - \bar{X})^2$	$m - 1$	$\frac{mk}{m-1} \sum_{j=1}^m (\bar{X}_{*j^*} - \bar{X})^2$
Взаємодія двох факторів $S_{AB}^2$	$k \sum_{j=1}^m \sum_{i=1}^m (\bar{X}_{ij^*} - \bar{X}_{i^{**}} - \bar{X}_{*j^*} + \bar{X})^2$	$(m-1) \times (n-1)$	$\frac{k \sum_{j=1}^m \sum_{i=1}^m (\bar{X}_{ij^*} - \bar{X}_{i^{**}} - \bar{X}_{*j^*} + \bar{X})^2}{(m-1)(n-1)}$
Залишкова $S_{\text{залишок}}^2$	$\sum_{i=1}^m \sum_{j=1}^n \sum_{l=1}^k (X_{ijk} - \bar{X}_{ij^*})^2$	$nm(k-1)$	$\frac{\sum_{i=1}^m \sum_{j=1}^n \sum_{l=1}^k (X_{ijk} - \bar{X}_{ij^*})^2}{nm(k-1)}$
Повна	$\sum_{i=1}^m \sum_{j=1}^n \sum_{l=1}^k (X_{ijk} - \bar{X})^2$	$nmk - 1$	$\frac{\sum_{i=1}^m \sum_{j=1}^n \sum_{l=1}^k (X_{ijk} - \bar{X})^2}{nmk - 1}$

де  $S_A^2$  – вплив фактору А;  $S_B^2$  – вплив фактору В;  $S_{AB}^2$  – спільний вплив обох факторів;  $S_{\text{залишок}}^2$  – обумовлена впливом випадкових факторів.

Необхідно перевірити значущість відмінності дисперсій  $S_A^2$  і  $S_{\text{залишок}}^2$ ,  $S_B^2$  і  $S_{\text{залишок}}^2$ ,  $S_{AB}^2$  і  $S_{\text{залишок}}^2$ . Для цього знаходять розрахункові значення  $F$ -критерію Фішера:

$$F_A = \frac{S_A^2}{S_{\text{залишок}}^2},$$

$$F_B = \frac{S_B^2}{S_{\text{залишок}}^2},$$

$$F_{AB} = \frac{S_{AB}^2}{S_{\text{залишок}}^2}.$$

Якщо виконуються умови

$$F_A > F_{\alpha, n-1, nm(k-1)},$$

$$F_B > F_{\alpha, m-1, nm(k-1)},$$

$$F_{AB} > F_{\alpha, (m-1)(n-1), nm(k-1)},$$

то вплив факторів  $A$  і  $B$  є значущим.

**Приклад 8.2.** Вивчався вплив забруднення ґрунту  $\text{Na}_2\text{S}$  (фактор  $B$ ) і  $\text{NaF}$  (фактор  $A$ ) на зростання паростків горобини звичайної. Показати, чи наявний вплив забруднень на ростові показники рослин.

Вихідні дані і результати розрахунків у програмі Excel представлені на рис. 8.3 та в табл. 8.6.

G4												
=AVERAGE(B2:B7)												
	A	B	C	D	E	F	G	H	I	J	K	
1		B1	B2	B3								
2		23,2	10,8	10,1								
3		18,4	13,1	10,6			B1	B2	B3	Середнє за рядками	$(\bar{X}_{j..} - \bar{X})^2$	
4	A1	20,3	12,8	8,6		A1	21,40	12,98	9,90	14,76	9,31	
5		28,6	12	12,1		A2	13,55	13,53	10,43	12,51	0,63	
6		13,4	15,7	8,6		A3	9,60	9,28	4,70	7,86	14,81	
7		24,5	13,5	9,4		Середнє за стовпцями	14,85	11,93	8,34	11,71		
8		18,4	13	10,6		$(\bar{X}_{.j} - \bar{X})^2$	9,86	0,05	11,32	Заг. Середнє		
9	A2	10,4	15,7	8,6								
10		14,6	12,7	11,4								
11		9,5	12,6	12,3								
12		17,6	12,6	11,1								
13		10,8	14,6	8,6								
14		7	6,5	4,5								
15		11,4	10,7	4,3								
16	A3	11,9	12,2	3,7								
17		11,8	11,8	5,7								
18		10,4	6,2	5,1								
19		5,1	8,3	4,9								

Рис. 8.3. Вихідні дані і допоміжні розрахунки до прикладу 8.3

Таблиця 8.6. Результати розрахунку прикладу 8.2 в програмі Excel

Компоненти дисперсії	Сума квадратів	Число ступенів вільності	Середній квадрат (дисперсія)	$F_{розр}$	$F_{теор}$
Між середніми за рядками (за фактором А)	445,61	2	222,81	29,94	3,20
Між середніми за рядками (за фактором В)	382,26	2	191,13	25,68	3,20
Взаємодія двох факторів	171,82	4	42,96	5,77	2,58
Залишкова	334,90	45	7,44		
Повна	1334,59	53	25,18		

Як бачимо з таблиці, всі розрахункові значення  $F$ -критерію Фішера більше відповідних табличних, що дає змогу зробити висновок про вплив забруднень на ростові показники рослин.

Двофакторний дисперсійний аналіз можна виконати в Excel автоматично за допомогою надбудови *Data analysis (Аналіз даних)*, *ANOVA: Two-Factor With Replication (Двофакторний дисперсійний аналіз з повтореннями)*. Для отримання результату в наведеному вище прикладі потрібно заповнити вікно в такий спосіб (рис. 8.4):

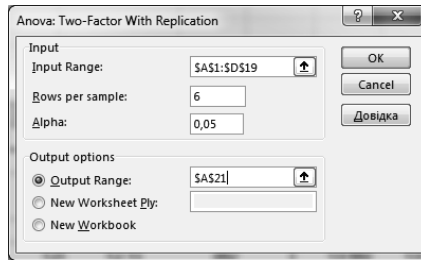


Рис. 8.4. Диалогове вікно Двофакторний дисперсійний аналіз з повтореннями

Результат двофакторного дисперсійного аналізу наведено на рис. 8.5.

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Sample	445,61	2	222,81	29,938	5,396E-09	3,2043
Columns	382,26	2	191,13	25,682	3,625E-08	3,2043
Interaction	171,82	4	42,955	5,7719	0,0007749	2,5787
Within	334,9	45	7,4422			
Total	1334,6	53				

Рис. 8.5. Результат двофакторного дисперсійного аналізу, проведеного за допомогою пакету аналізу даних в програмі Excel

Порівняйте результати, отримані різними способами. Зверніть увагу на рівень значущості ( $p$ -value). Оскільки  $p < 0,05$  для всіх трьох компонентів дисперсії, забруднення ґрунту хімічними речовинами  $\text{Na}_2\text{S}$  (фактор  $B$ ) і  $\text{NaF}$  (фактор  $A$ ) достовірно впливає на зростання паростків горобини звичайної.

### Дисперсійний аналіз у програмі Statistica

У програмі Statistica дисперсійний аналіз (будь-якого виду – одно-, дво-, багатфакторний; з повторностями або без) можна провести за допомогою мо-

дуля *ANOVA* (від англ. Analysis Of Variance – аналіз дисперсій) з пункту меню *Statistics / Base* (*Статистика / Основна*). Стартова панель *General ANOVA / MANOVA* (*Дисперсійний аналіз*) дає змогу обрати тип аналізу (*type of analyses*) (рис. 8.6):

- *One-way ANOVA* (*Однофакторний дисперсійний аналіз*) – дослідження одного незалежного фактора (однофакторний аналіз можна також проводити за допомогою модуля *Breakdown & one-way ANOVA* (*Угруповання і однофакторний дисперсійний аналіз*), розташованого в пункті меню *Statistics, Basic statistics* (*Аналіз, Основні статистики*);
- *Main effects ANOVA* (окремий випадок факторного аналізу) – дослідження одночасного впливу кількох незалежних факторів;
- *Factorial ANOVA* (*Багатофакторний дисперсійний аналіз*) – дослідження взаємопов'язаних ефектів незалежних факторів;
- *Repeated measures ANOVA* (*Дисперсійний аналіз з повторностями*) – зіставлення пов'язаних груп.

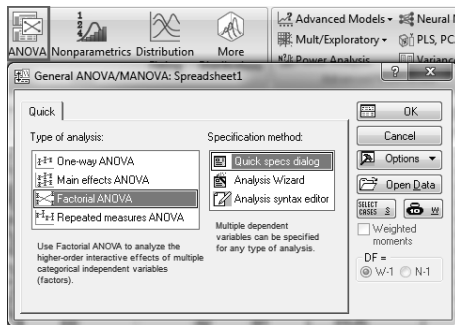


Рис. 8.6. Стартова панель модуля *Дисперсійний аналіз*

Розглянемо можливості модулю *ANOVA* (*Дисперсійний аналіз*) в програмі *Statistica* на модельному прикладі.

**Приклад 8.3.** Для спрощення сприйняття візьмемо частину даних із прикладу 8.2. Вивчався вплив забруднення ґрунту  $\text{Na}_2\text{S}$  (фактор *B*) і  $\text{NaF}$  (фактор *A*) на зростання паростків горобини звичайної. Показати, чи наявний вплив забруднень на ростові показники рослин.

Із залежних параметрів (ростові показники, *h*) формуємо єдину вибірку, яку розміщуємо у стовпчик *Var3*. Дасмо назву стовпчику – *h*. У перший та другий стовпчик навпроти кожного вимірювання прописуємо назви факторів групи дос-

ліджуваних рослин, на перетині яких вони знаходяться. Дасмо назви стовпчикам відповідно до позначення фактора – *A* та *B*. Внаслідок отримуємо таку таблицю, як на рис. 8.7, де перші два стовпці – це незалежні (гуртуючі), а третій стовпчик – залежні змінні.

	B1	B2
A1	23,2	10,8
	18,4	13,1
	20,3	12,8
	28,6	12
	13,4	15,7
A2	24,5	13,5
	18,4	13
	10,4	15,7
	14,6	12,7
	9,5	12,6
	17,6	12,6
	10,8	14,6

1	2	3	4	5	6	7	8	9	10
A	B	h	Var4	Var5	Var6	Var7	Var8	Var9	Var10
1	A1	B1	23,2						
2	A1	B1	18,4						
3	A1	B1	20,3						
4	A1	B1	28,6						
5	A1	B1	13,4						
6	A1	B1	24,5						
7	A2	B1	18,4						
8	A2	B1	10,4						
9	A2	B1	14,6						
10	A2	B1	9,5						
11	A2	B1	17,6						
12	A2	B1	10,8						
13	A1	B2	10,8						
14	A1	B2	13,1						
15	A1	B2	12,8						
16	A1	B2	12						
17	A1	B2	15,7						
18	A1	B2	13,5						
19	A2	B2	13						
20	A2	B2	15,7						
21	A2	B2	12,7						
22	A2	B2	12,6						
23	A2	B2	12,6						
24	A2	B2	14,6						

Рис. 8.7. Формування таблиці з вихідними даними в програмі Statistica

У діалоговому вікні *General ANOVA / MANOVA (Дисперсійний аналіз)* вибираємо *Factorial ANOVA (Багатофакторний дисперсійний аналіз)*, *Quick specs dialog (Діалоговий режим)* і натискаємо кнопку *OK*.

У діалоговому вікні *Select dependent variables and categorical predictors (factors)* вводимо залежні (*Dependent variable list*) та незалежні (*Categorical*

*predictors (factors)*) змінні і натискаємо кнопку *OK*. В цьому ж діалоговому вікні натискаємо кнопку *Factors Codes (Коди факторів)* і напроти кожного рядка натискаємо *All (Всі)*, вказуючи, що всі градації факторів та відповідні їм значення залежного параметра будуть аналізуватися.

Після того як налаштування зроблені, у вікні *ANOVA/MANOVA Factorial ANOVA* натискаємо *OK*. У діалоговому вікні *ANOVA Results (Результати дисперсійного аналізу)* натискаємо кнопку *All effects (Всі ефекти)*. На екран виводиться підсумкова таблиця (рис. 8.8).

Effect	SS	Degr. of Freedom	MS	F	p
A	79.9	1	79.9	6.780	.017*
B	106.7	1	106.7	9.049	.007*
A*B	105.8	1	105.8	8.977	.007*

Рис. 8.8. Підсумкова таблиця з результатами дисперсійного аналізу

Із підсумкової таблиці бачимо, що обидва досліджувані фактори та їх комбінація достовірно впливають на ростові параметри рослин – розраховане значення  $p$  набагато менше 0,05. Знак \* також вказує на достовірний вплив фактора.

Встановивши курсор у підсумковій таблиці на окремий фактор (або їх комбінацію) і натиснувши кнопку *OK*, можна продемонструвати вплив фактора у вигляді графіка. Наприклад, встановимо курсор у підсумковій таблиці на ефекті *A* і натиснемо *OK*. На екран виводиться графік, з якого видно, що збільшення концентрації  $\text{NaF}$  у ґрунті призводить до зниження ростових параметрів рослин (рис. 8.9А). Повернутися до попереднього діалогового вікна для продовження роботи можна, натиснувши на ярличок *ANOVA Results* в нижній частині робочого поля вікна. Аналогічні результати отримані для фактора *B* (рис. 8.9В). За виглядом графіка взаємодії факторів *A* і *B* (рис. 8.9С) можна зробити висновок про її наявність або відсутність: якщо лінії на графіку паралельні, то взаємодії між факторами немає, якщо не паралельні – взаємодія є. Достовірна взаємодія факторів вказує на те, що різниця між рівнями одного з факторів неоднакова для всіх рівнів іншого фактора.

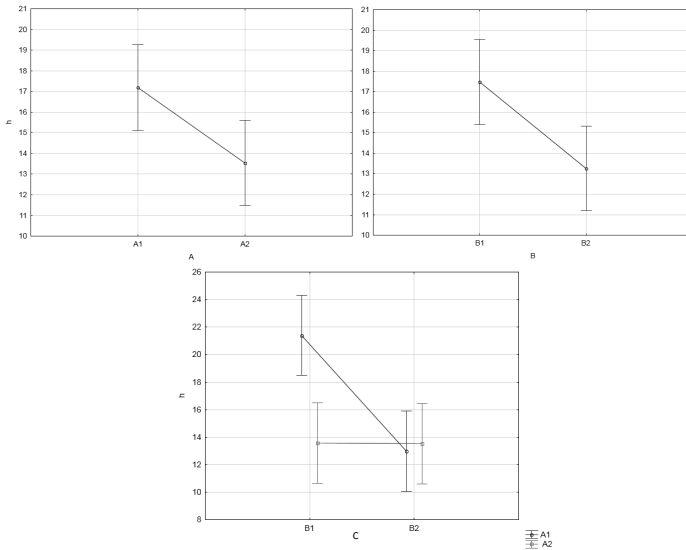


Рис. 8.9. Графічне представлення результатів двофакторного дисперсійного аналізу

Щоб вивести підсумкову таблицю з результатами дисперсійного аналізу (рис. 8.10), потрібно натиснути кнопку *All effects* діалогового вікна *ANOVA results*. У таблиці звертаємо увагу на  $p$ -рівень. Якщо  $p$ -рівень менше 0,05, то наявний достовірний вплив фактора (факторів), в такому разі рядок таблиці виділений червоним кольором.

Univariate Tests of Significance for h (Spreadsheet8)					
Sigma-restricted parameterization					
Effective hypothesis decomposition					
Effect	SS	Degr. of Freedom	MS	F	p
A	79,935	1	79,935	6,7800	0,016985
B	106,682	1	106,682	9,0486	0,006948
A*B	105,840	1	105,840	8,9772	0,007137

Рис. 8.10. Підсумкова таблиця двофакторного дисперсійного аналізу

### Багатофакторний дисперсійний аналіз у Statistica

**Приклад 8.4.** Завантажте файл *SH-group.sta* у програму Statistica. Розчини імуноглобуліну різних концентрацій (1, 0,75, 0,5 г/дл) піддавалися вібрації з частотою 24 Гц і амплітудами 0,5 і 1 мм протягом трьох годин. Кожні 15 хв у розчині визначався вміст SH-груп. Встановити, чи впливає амплітуда, тривалість вібрації і концентрація розчину білка на зміну вмісту SH-груп у розчині.

Вид представлення даних у файлі *SH-grup.sta*: перший стовпець – часові проміжки, через які визначався вміст SH-груп; другий – амплітуда вібрації (0,5 мм та 1 мм); третій – концентрація розчину (1, 0,75 та 0,5 г/дл). Четвертий стовпець – зміна вмісту SH-груп щодо контролю (вміст SH-груп у розчині до вібрації).

Виконуємо послідовно команди: *Statistic (Статистика)*, *ANOVA*, *Factorial ANOVA*, *OK*. У вікні робимо такі налаштування: натискаємо кнопку *Variables* і як залежну змінну вказуємо SH-групи, а як предиктор – час, амплітуду, концентрацію. Натискаємо *OK*.

Натискаємо кнопку *Коди факторів (Factor codes)* і у вікні, напроти кожного рядка, натискаємо *All (Всі)*. Натискаємо кнопку *OK*.

Після того як налаштування зроблені, у вікні *ANOVA / MANOVA Factorial ANOVA* натискаємо *OK*. У вікні натискаємо кнопку *All effects (Всі ефекти)*. На екран виводиться підсумкова таблиця (рис. 8.11).

Effect	SS	Degr. of Freedom	MS	F	p
Час	,400	11	,036	46,5	0,00*
Амплітуда	,300	1	,300	383,5	0,00*
Концентрація	,158	2	,079	101,0	0,00*
Час*Амплітуда	,140	11	,013	16,3	0,00*
Час*Концентрація	,122	22	,006	7,1	0,00*
Амплітуда*Концентрація	,047	2	,024	30,3	0,00*
Час*Амплітуда*Концентрація	,067	22	,003	3,9	0,00*

Рис. 8.11. Підсумкова таблиця багатфакторного дисперсійного аналізу

Із підсумкової таблиці видно, що всі досліджувані фактори та їх комбінації достовірно впливають на зміну вмісту SH-груп у досліджуваних розчинах. Для цього дивимося на стовпець *p*. Розраховане значення *p* набагато менше 0,05. Знак \* також вказує на достовірний вплив фактора.

Встановивши курсор у підсумковій таблиці на цікавий для вас фактор і натиснувши кнопку *OK*, можна продемонструвати вплив фактора у вигляді графіка. Наприклад, встановимо курсор у підсумковій таблиці на ефекті *Амплітуда* і натиснемо *OK*. На екран виводиться графік, з якого видно, що вібрація з амплітудою 0,5 мм призводить до більшого зменшення вмісту SH-груп у розчині (рис. 8.12).

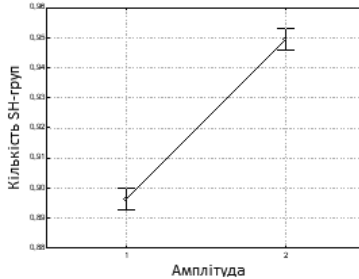


Рис. 8.12. Графічне представлення результатів дисперсійного аналізу

Самостійно побудуйте графіки для інших ефектів, їх поєднань і зробіть висновки про вплив амплітуди, тривалості вібрації і концентрації розчину на зміну вмісту SH-груп.

### Завдання для самостійної роботи

1. Виконати однофакторний дисперсійний аналіз для даних, наведених у розділі «Варіанти завдань для закріплення теми 8. Однофакторний дисперсійний аналіз», у програмах Excel та Statistica.
2. Виконати двофакторний дисперсійний аналіз для даних, наведених у розділі «Варіанти завдань для закріплення теми 8. Двофакторний дисперсійний аналіз», у програмах Excel та Statistica.
3. Виконати дисперсійний аналіз у програмі Statistica.
  - 3.1. Використовуючи файл *Ascaridoz.sta*, оцінити достовірність відмінностей вмісту імунних комплексів залежно від статі хворого і виду захворювання.
  - 3.2. Використовуючи файл *Страх.sta*, оцінити достовірність відмінностей рівнів страху, тривожності, самооцінки і статусу в колективі школярів 1-А та 1-Б класів різної статі.
  - 3.3. Використовуючи файл *Aggression.sta*, встановити, чи розрізняються рівні агресії у представників різної статі.
  - 3.4. Препарат Ферро-Фольгама застосовується для лікування анемії. Встановити, чи привело лікування цим препаратом до поліпшення гематологічних показників крові у вагітних і породіль.
  - 3.5. Файл *Реологія.sta* містить дані з лікування хворих на ревматоїдний артрит. Основна група поряд із лікарською терапією проходила курс магнітотерапії, контрольна група проходила тільки курс лікарської терапії. Встановити,

який зі способів лікування виявився найбільш ефективним. Провести зіставлення з показниками здорових людей.

- 3.6. Діти з ревматизмом отримували різну лікарську терапію. Одна група отримувала традиційне лікування, інша – поряд із традиційними препаратами отримували препарат «Кардонат». Використовуючи дані файлу *Кардонат.sta*, встановити, чи привело лікування до бажаних результатів.

### **Контрольні запитання до теми 8**

1. Який критерій використовується для перевірки значущості відмінності дисперсій? Як визначається цей критерій?
2. Як визначаються та що характеризують загальна, внутрішньогрупова та міжгрупова дисперсії?
3. Для чого використовується однофакторний дисперсійний аналіз? У який спосіб здійснюється перевірка гіпотез у цьому виді аналізу?
4. Для чого використовується двофакторний дисперсійний аналіз? Скільки нульових гіпотез перевіряється у цьому виді аналізу?

## ТЕМА 9. РЕГРЕСІЙНИЙ АНАЛІЗ (МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ)

У біологічних дослідженнях важливо знати, як зміна одного параметра впливає на зміну іншого, наприклад, врожайність пшениці змінюється зі зміною кількості внесених добрив, маса тварин змінюється зі зміною вмісту білка в кормі.

У таких задачах потрібно виділити основну змінну (об'єкт досліджень) та одну або кілька змінних, які впливають на основну. Основна змінна називається залежною змінною, яка позначається  $Y$ , а всі інші – незалежними (або предикторами), які позначаються  $X$ .

Отже, зміна врожайності пшениці є залежною змінною ( $Y$ ), а зміна кількості внесених добрив – незалежною змінною ( $X$ ), яка впливає на ефективність врожайності.

Метою регресійного аналізу є визначення математичної залежності (рівняння регресії), яка описує взаємозв'язок залежної і незалежної змінних, і передбачення значення одного досліджуваного параметра за відомим значенням іншої змінної (або змінних). Якщо зв'язок між змінними є лінійним, то регресійна модель у найпростішому випадку (для однієї залежної та однієї незалежної змінної) має вигляд  $Y = aX + b$ .

У випадку, коли розглядається залежність між однією залежною змінною  $Y$  і кількома незалежними  $X_1, X_2, \dots, X_m$ , говорять про множинну лінійну регресію, яка в загальному вигляді описується рівнянням регресії

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_mX_m,$$

де  $a_0, a_1, a_2, \dots, a_m$  – коефіцієнти регресії, які потребують визначення.

Коефіцієнти рівняння регресії визначаються за допомогою методу найменших квадратів, домагаючись мінімально можливої суми квадратів розбіжностей реальних значень змінної  $Y$  і обчислених за регресійним рівнянням. Наприклад, рівняння лінійної регресії може бути побудоване навіть у тому разі, коли лінійний кореляційний зв'язок відсутній.

Основними результатами множинної регресії є:

– *коефіцієнт множинної кореляції*  $R$  – узагальнення коефіцієнта парної кореляції для випадку, коли число незалежних факторів, включених у рівняння, більше одного.  $R$  є безрозмірною величиною і не змінюється у разі зміни одиниць вимірювання відповідних ознак.  $R$  приймає значення в інтервалі  $[0; 1]$ ;

– *коефіцієнт детермінації*  $R^2$  – міра ефективності регресійної моделі. Цей коефіцієнт є квадратом відповідного коефіцієнта кореляції і виражається у відсотках. Чим більше  $R$ , тим сильніший лінійний зв'язок між сукупністю незалежних факторів і результативною ознакою. Коефіцієнт детермінації  $R^2$  показує,

наскільки зміни залежної ознаки (у відсотках) пояснюються змінами сукупності незалежних ознак. Це частка дисперсії залежної ознаки, яка пояснюється впливом незалежних ознак;

– значущість регресійної моделі, мірою якої є значення  $F$ -критерію – відношення факторної дисперсії до залишкової. Чим краща регресійна модель, тим вища частка факторної і нижча доля залишкової дисперсії. Для значення  $F$  можна обчислити відповідну ймовірність. Якщо значення цієї ймовірності менше прийнятого рівня значущості  $p$ , або ймовірності помилки (зазвичай це 0,05 або 5 %), гіпотеза про відсутність лінійного зв'язку між залежною ознакою і предикторами відхиляється, і регресія визнається значущою;

– коефіцієнти регресії та рівні їх значущості. Коефіцієнт регресії показує, наскільки зміниться значення залежного параметра, якщо незалежний параметр збільшити на одиницю при незмінних значеннях інших предикторів. Коефіцієнт вважається нульовим, якщо рівень значущості  $p > 0,05$ . В такому разі вплив відповідної незалежної змінної на залежну змінну недостовірний, і ця незалежна змінна може бути виключена з моделі. Достовірність відмінності коефіцієнтів від нуля перевіряється за  $t$ -критерієм Стьюдента.

Регресійний аналіз може використовуватися для побудови моделей, що пояснюють механізм впливу факторних ознак на результат; для статистичного прогнозу в межах діапазону значень наявних незалежних змінних; для заповнення пропусків даних.

Розглянемо проведення лінійного регресійного аналізу в програмі Excel.

**Приклад 9.1.** Емісія  $\text{CO}_2$  з ґрунтів (або загальне дихання ґрунту  $\text{SR}$ ,  $\text{г}/(\text{см}^2 \times \text{доба})$ ) поряд із акумуляцією вуглецю внаслідок фотосинтетичної діяльності рослин є основними протилежно спрямованими потоками у вуглецевому циклі наземних екосистем. Основними абіотичними факторами, які впливають на його величину, є температура ( $T$ ,  $^\circ\text{C}$ ) і вологість ґрунту, яку у вегетаційний період часто виражають через кількість опадів ( $P$ , мм). Використовуючи дані багаторічного моніторингу емісії  $\text{CO}_2$  з ґрунтів різних лісових екосистем (на річному рівні усереднення), наведених у таблиці, провести регресійний аналіз, скласти модель, яка показує залежність емісії  $\text{CO}_2$  від абіотичних факторів. Спрогнозувати величину  $\text{SR}$  на 2022 рік.

Розрахунки виконуємо за допомогою пакета аналізу даних. Вибираємо пункт меню *Дані*, вкладку *Analysis / Data analysis*; у діалоговому вікні, що відкрилося, вибираємо тип аналізу *Regression (Регресійний аналіз)* (рис. 9.1).

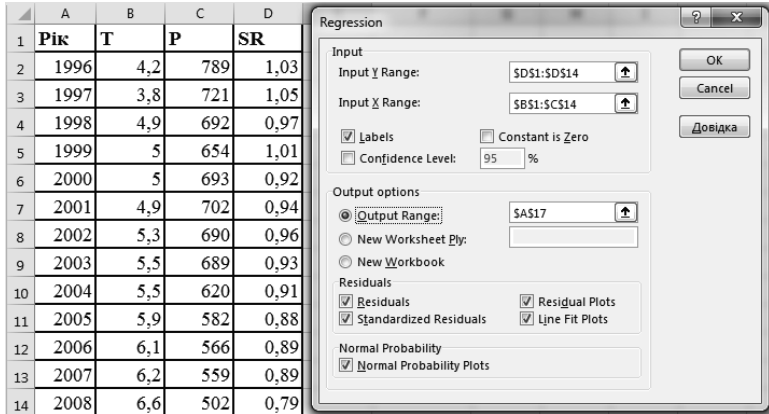


Рис. 9.1. Діалогове вікно Регресійний аналіз

У діалоговому вікні, що з'явилося, вводимо:

- *Input Y Range* (Введення залежної змінної – SR);
- *Input X Range* (Введення незалежних змінних – T, P);
- *Labels* – ставимо прапорець, якщо разом із вихідними даними виділені назви стовпців;
- *Output Range* (*Вихідний інтервал*) вводимо ім'я комірки, починаючи з якої буде виводитися результат;
- у розділі *Residuals* ставимо прапорці навпроти опцій *Residuals*, *Residual Plots*, *Line Fit Plots*, *Normal Probability Plots* для виведення залишків (відхилень спостережуваного значення залежного параметра від розрахованого за рівнянням регресії); графіків залишків; лінійних графіків, побудованих за результатами регресійного аналізу, а також графіків нормального розподілу, які є одним із критеріїв адекватності моделі.

Після натискання кнопки ОК результат буде виведено в робоче поле вікна (рис. 9.2).

Тлумачення кожного з отриманих параметрів буде наведено нижче, при детальному аналізі наведеного прикладу в програмі Statistica.

SUMMARY OUTPUT					RESIDUAL OUTPUT				PROBABILITY OUTPUT		
Regression Statistics					Observation	Predicted SR	Residuals	Standard Residuals	Percentile	SR	
Multiple R	0,925				1	1,029	0,001	0,052	3,846	0,79	
R Square	0,856				2	1,054	-0,004	-0,134	11,538	0,88	
Adjusted R Square					3	0,969	0,001	0,033	19,231	0,89	
Square	0,827				4	0,959	0,051	1,925	26,923	0,89	
Standard Error	0,029				5	0,962	-0,042	-1,570	34,615	0,91	
Observations	13				6	0,970	-0,030	-1,123	42,308	0,92	
ANOVA					7	0,939	0,021	0,789	50,000	0,93	
	df	SS	MS	F	Significance F	8	0,924	0,006	0,226	57,692	0,94
Regression	2	0,050	0,025	29,638	0,000063	9	0,919	-0,009	-0,336	65,385	0,96
Residual	10	0,008	0,001			10	0,886	-0,006	-0,234	73,077	0,97
Total	12	0,059				11	0,870	0,020	0,750	80,769	1,01
						12	0,862	0,028	1,051	88,462	1,03
						13	0,828	-0,038	-1,428	96,154	1,05
	Coef.icients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%					
Intercept	1,285	0,294	4,367	0,001	0,629	1,941					
T	-0,075	0,026	-2,924	0,015	-0,132	-0,018					
P	0,000	0,000	0,289	0,778	0,000	0,001					

Рис. 9.2. Результат регресійного аналізу в програмі Excel

### Регресійний аналіз у програмі Statistica

Вводимо дані в програму Statistica. Перейменовуємо заголовки стовпців:  $Var1 - Pik$ ,  $Var2 - T$ ,  $Var3 - P$ ,  $Var4 - SR$ . Зверніть увагу, параметр  $Pik$  у регресійному аналізі не бере участі.

У рядку з пункту меню *Statistics* вибираємо модуль *Multiple regression* (Множинна регресія). Відкриється стартова панель модуля *Multiple Linear regression* (Множинна лінійна регресія) (рис. 9.3).

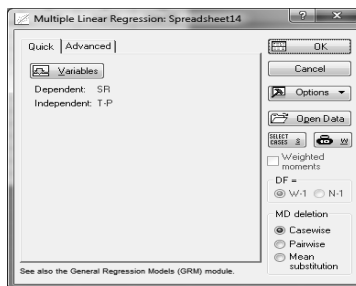


Рис. 9.3. Multiple Linear regression (Множинна лінійна регресія)

У діалоговому вікні *Select dependent and independent variable list* (Виберіть список залежних і незалежних змінних), що відкрилося, натискаємо кнопку *Variables* (Змінні) для вибору залежної та незалежних змінних: у лівій частині вікна вибираємо залежну змінну ( $SR$ ), у правій – незалежні змінні ( $T, P$ ).

Ніяких додаткових налаштувань у стартовій панелі у цьому випадку проводити не потрібно, тому натиснемо кнопку *OK*. Програма зробить оцінювання параметрів моделі і з'явиться діалогове вікно результатів регресійного аналізу (рис. 9.4).

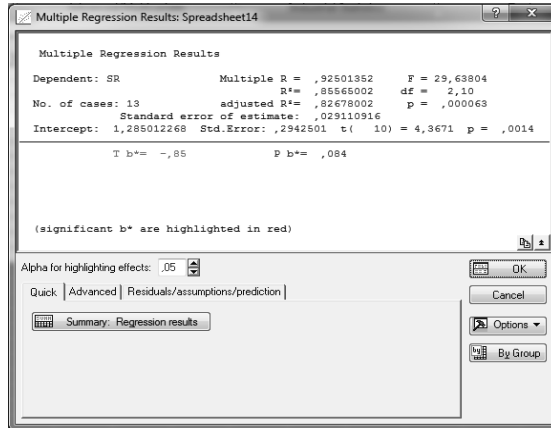


Рис. 9.4. Діалогове вікно результатів множинної регресії

Діалогове вікно *Multiple Regression Results* (*Результати множинної регресії*) має дві частини: верхня панель – інформаційна. Вона складається з двох розділів: перший містить основну інформацію про результати регресійного аналізу, другий – отримані стандартизовані коефіцієнти регресії, значущі з яких виділені червоним кольором. Нижня частина вікна містить функціональні кнопки і вкладки, які дають змогу ознайомитися з результатами оцінювання. Інформаційна частина вікна містить основні відомості про результати регресійного аналізу:

- *Dependent* (*Залежні змінні*) – ім'я залежної змінної. У цьому разі – *SR*.
- *No of Cases* (*Число спостережень*) – число випадків, за якими побудована регресія. У цьому разі – 13.
- *Multiple R* (*Множинне R*) – коефіцієнт множинної кореляції. В нашому прикладі – 0,925.
- $R^2$  – квадрат коефіцієнта множинної кореляції, зазвичай називається коефіцієнтом детермінації. У цьому разі – 0,856.

**Зверніть увагу:** коефіцієнт детермінації є однією з основних статистик у даному вікні, він показує частку загального розкиду (щодо вибіркового середнього залежної змінної), яка пояснюється побудованою регресією. Значення  $R^2$  змінюється від 0 до 1.

У прикладі  $R^2 = 0,86$ . Це досить гарне значення, яке показує, що побудована регресія пояснює більш ніж 86 % розкиду значень змінної *SR* щодо середнього.

- *Adjusted R2* (*Скоригов. R<sup>2</sup>*) – скоригований коефіцієнт детермінації. У цьому разі – 0,827.

Скоригований коефіцієнт детермінації визначається, як  $1 - (1 - R^2) \cdot \frac{n}{n - m}$ ,

де  $n$  – число спостережень в моделі;

$m$  – число параметрів моделі (число незалежних змінних плюс один, оскільки в модель вміщено вільний член).

- *Standard error of estimate (Стандартна помилка оцінки)* – міра розсіювання спостережуваних значень щодо регресійної прямої. У прикладі – 0,029.

- *Intercept (Вільний член)* – оцінка вільного члена регресії; значення коефіцієнта  $b$  в рівнянні регресії  $y = ax + b$ . В прикладі – 1,285.

- *St. Error* – стандартна помилка оцінки вільного члена  $b$  у рівнянні регресії. У прикладі – 0,29.

-  $t$  і  $p$  – значення  $t$ -критерію і рівень значущості  $p$ . У цьому разі  $t(10) = 4,367$ , а  $p = 0,0014$ .  $t$ -критерій використовують для перевірки гіпотези про рівність нулю коефіцієнта детермінації. Рівень значущості  $p$  ( $< 0,05$ ) означає, що коефіцієнт детермінації значно відрізняється від 0 і отримана модель значуща.

Наступний блок – перевірка  $F$ -критерію:

-  $F$  – значення  $F$ -критерію. У прикладі – 29,638.

-  $df$  – число ступенів вільності  $F$ -критерію. У прикладі – 2,10.

-  $p$  – рівень значущості для  $F$ -критерію. У цьому разі – 0,000063.

$F$ -критерій використовується для перевірки гіпотези про значущість регресії і показує, наскільки добре ця модель пояснює загальну дисперсію залежної змінної. У прикладі велике значення  $F$ -критерію і низький рівень значущості для критерію Фішера показують, що побудована регресія значуща.

Нижня частина інформаційного вікна містить розраховані регресійні коефіцієнти:  $T$  beta = -0,85 і  $P$  beta = -0,084, значущі з яких виділено червоним кольором (*significant beta are highlighted in red*). Тут наведені beta – стандартизовані коефіцієнти при незалежних змінних  $T$  і  $P$ .

### Функціональна частина вікна результатів

На вкладці *Quick (Швидкий аналіз)* натиснемо кнопку *Summary: Regression Results (Підсумкова таблиця регресії)*. В робочому полі вікна з'явиться таблиця з підсумковими результатами оцінювання регресійної моделі (рис. 9.5).

Regression Summary for Dependent Variable: SR (Spreadsheet1)						
R= .92501352 RI= .85565002 Adjusted RI= .82678002						
F(2,10)=29.638 p<.00006 Std Error of estimate: .02911						
	b*	Std.Err. of b*	b	Std.Err. of b	t(10)	p-value
Intercept			1,285012	0,294250	4,36707	0,001406
T	-0,8480701	0,290055	-0,074833	0,025594	-2,92382	0,015201
P	0,083821	0,290055	0,000073	0,000254	0,28898	0,778496

Рис. 9.5. Підсумкова таблиця регресії

Таблиця результатів регресійного аналізу містить таку інформацію: стандартизовані коефіцієнти регресійного рівняння ( $b^*$ ) та їхні стандартні помилки (*Std.Err. of  $b^*$* ); коефіцієнти при незалежних змінних  $T$  і  $P$ , вільний член рівняння регресії ( $b$ ) та їхні стандартні помилки (*Std.Err. of  $b$* ), значення статистик  $t$ -критерію і рівень значущості для кожного з параметрів моделі.

З таблиці бачимо, що оцінена модель має вигляд:

$$SR = 1,285 - 0,075 \times T + 0,000073 \times P.$$

Значущими є коефіцієнт при  $T$  ( $p = 0,015 < 0,05$ ) і вільний член рівняння ( $p = 0,001, < 0,05$ ), тобто модель можна записати у вигляді:  $SR = 1,285 - 0,075 \times T$ .

Важливим етапом регресійного аналізу є перевірка адекватності моделі, яка базується на аналізі залишків (різниця між експериментальними даними і значеннями, розрахованими за отриманою регресійною моделлю).

Для аналізу залишків у модулі *Множинна регресія (Multiple regression)* потрібно перейти на вкладку *Residuals / Assumptions / Prediction (Залишки / Припущення / Прогнозування)* і натиснути кнопку *Rerform Residual Analysis (Аналіз залишків)*. У діалоговому вікні *Residual Analysis (Аналіз залишків)*, що відкрився, в його функціональній частині (інформаційна частина вікна не змінилась) є вкладки, за допомогою яких можна в графічному або табличному вигляді отримати інформацію про залишки, що вказує на адекватність моделі (рис. 9.6).

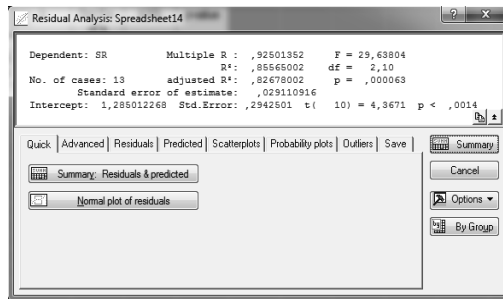


Рис. 9.6. Діалогове вікно аналізу залишків у модулі *Множинна лінійна регресія*

Відкриємо, наприклад, вкладку *Scatterplots (Діаграми розсіювання)* і натиснемо кнопку *Predicted vs. Observed (Передбачені і спостережувані значення)*. На екрані з'явиться графік, який говорить про достатню адекватність моделі (рис. 9.7).

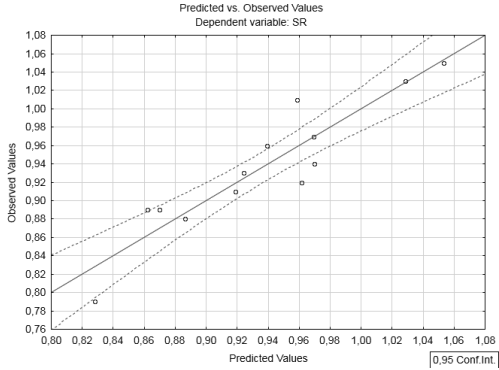


Рис. 9.7. Графік залежності спостережуваних значень від прогнозованих

У цьому ж діалоговому вікні можна побудувати графічні залежності, які вказують на адекватність моделі:

- графік, на якому по одній з осей відкладаються модельні значення, а по інший – значення залишків, дає змогу стверджувати про сталість або мінливість дисперсії помилки залишків. Якщо точки, нанесені на графік, невпорядковані, як на рис. 9.8, то дисперсія помилки – величина стала. А якщо ні, то розкид залишків змінюється зі зміною залежної змінної, що вказує на неадекватність моделі і необхідність перетворення змінних. Перетворення однієї або кількох змінних може значно покращити відповідність регресійної моделі експериментальним даним та зменшити вплив викидів. Як спосіб перетворення експериментальних даних часто використовується логарифмування змінних.

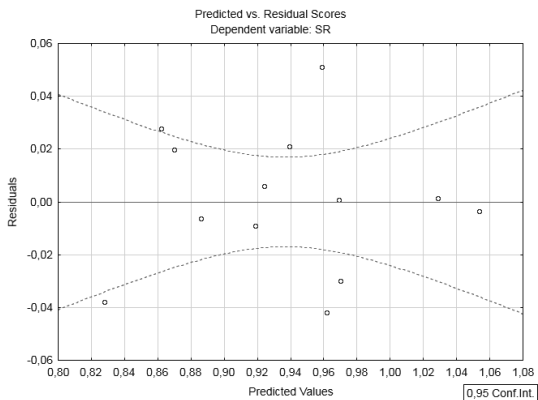


Рис. 9.8. Графік залежності залишків від модельних значень параметру

На вкладці *Probability plots* (Ймовірнісні графіки) можна побудувати графік перевірки нормальності розподілу залишків. Аналіз залишків дає змогу визначити, наскільки добре підібрана модель і наскільки правильно обраний метод оцінки параметрів. Якщо побудована регресійна модель добре описує експериментальну залежність, то залишки повинні бути незалежними нормально розподіленими випадковими величинами з нульовим середнім, що виконується для даних завдання. Графік нормального розподілу залишків представлений на рис. 9.9. Інший спосіб перевірки на нормальний розподіл – перевірка асиметрії і стандартної помилки асиметрії, ексцесу та стандартної помилки ексцесу – також підтверджує нормальний розподіл залишків.

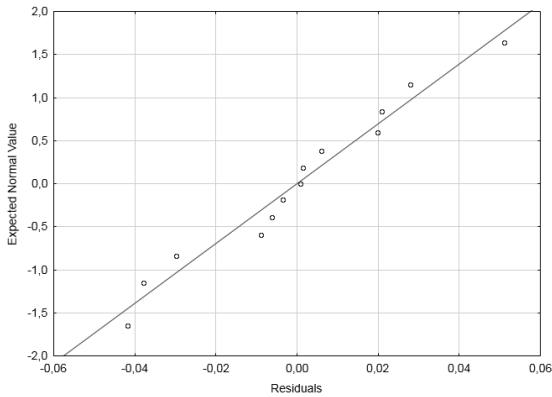


Рис. 9.9. Графік нормального розподілу залишків

Регресійний аналіз показав, що найістотнішим із факторів, які впливають на емісію CO<sub>2</sub> з ґрунтів, є температура. За запропонованими даними кількість опадів виявилася незначним фактором, який у регресійну модель не ввійшов. На основі значення коефіцієнта детермінації, графіків, аналізу залишків можна зробити висновок про адекватність моделі.

Крім того, в діалоговому вікні *Multiple Regression Results* (Результати множинної регресії) на вкладці *Residuals / Assumptions / Prediction* у розділі *Predict value* (Прогнозовані значення) можна спрогнозувати значення залежної змінної для певних незалежних параметрів. Для цього натиснемо на кнопку *Predict dependent variable* (Передбачити залежну змінну) і в діалоговому вікні задаємо значення незалежних змінних (рис. 9.10).

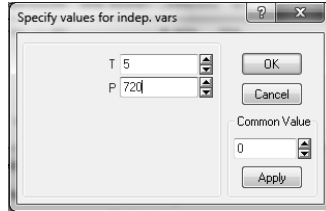


Рис. 9.10. Діалогове вікно прогнозування залежної змінної

Після натискання кнопки *OK* на екрані з'явиться результуюча таблиця з прогнозованим значенням залежної змінної – 0,96369 і довірчим інтервалом із рівнем значущості 0,05 (рис. 9.11).

Predicting Values for (Spreadsheet variable: SR)			
Variable	b-Weight	Value	b-Weight * Value
T	-0,074833	5,0000	-0,374165
P	0,000073	720,0000	0,052843
Intercept			1,285012
Predicted			0,963690
-95,0%CL			0,933144
+95,0%CL			0,994237

Рис. 9.11. Підсумкова таблиця з результатами прогнозування залежної змінної

### Завдання для самостійної роботи

1. У програмах Excel та Statistica знайти рівняння множинної регресії і перевірити отриману модель на адекватність для даних, наведених у розділі «Варіанти завдань для закріплення теми 9».

### Контрольні запитання до теми 9

1. Дайте визначення залежної та незалежної змінної. Наведіть приклади.
2. Наведіть загальне рівняння множинної лінійної регресії. Як визначаються коефіцієнти рівняння регресії?
3. Які параметри є основними результатами множинної регресії? На що вказує коефіцієнт детермінації? Як визначається скоригований коефіцієнт детермінації?
4. За яким критерієм визначається значущість регресійної моделі та рівні значущості коефіцієнтів регресії?

## ТЕМА 10. КЛАСТЕРНИЙ АНАЛІЗ

Кластерний аналіз у статистиці – це набір інструментів та алгоритмів, який використовується для класифікації різних об'єктів у групи так, що подібність між двома об'єктами є максимальною, якщо вони належать до однієї групи – кластеру, і мінімальною в іншому випадку. У біології кластерний аналіз є важливим інструментом систематики (класифікація живих і вимерлих організмів). У клінічній медицині його можна використовувати для виявлення пацієнтів, які мають захворювання із загальною причиною, пацієнтів, які повинні отримувати однакове лікування, або пацієнтів, які повинні мати однаковий рівень реакції на лікування.

В епідеміології кластерний аналіз може бути використаний для пошуку значущих конгломератів регіонів зі схожими епідеміологічними характеристиками. Взагалі, коли потрібно класифікувати великі обсяги інформації на невелику кількість значущих категорій, може бути корисним кластерний аналіз.

Дослідники доволі часто стикаються із завданням сортування спостережуваних даних у значущі структури. Кластерний аналіз – це індуктивна дослідницька техніка в тому сенсі, що вона розкриває структури без пояснення причин їх існування. Це техніка, що генерує гіпотези, а не перевірка гіпотез. На відміну від дискримінантного аналізу (буде розглянуто нижче), коли об'єкти присвоюються попередньо наявним групам на основі статистичних правил розподілу, кластерний аналіз генерує групи або виявляє приховану структуру груп усередині даних.

Головний принцип групування в кластерному аналізі – це відстань між об'єктами на діаграмі розсіювання, яку можна розраховувати різними методами. Один із найбільш поширених методів – це евклідова відстань, яка на площині розраховується для двох точок  $i$  та  $j$  з відомими координатами  $(x, y)$  за формулою:

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}.$$

Тут  $i$  та  $j$  – об'єкти, кожному з яких властиві два досліджувані параметри –  $x$  та  $y$ . В загальному випадку, коли досліджуваних параметрів об'єктів більше двох, евклідова відстань – це квадратний корінь із суми квадратів різниць між усіма координатами векторів, що визначають кожен об'єкт. Розглянемо за допомогою програми Excel принцип розподілу об'єктів на кластери.

**Приклад 10.1.** Візьмемо шість об'єктів спостережень – рослини. У кожної рослини вимірювали два параметри – висота стебла ( $x$ ) та довжина кореня ( $y$ ), які її характеризують. Визначити, чи можна за цими ознаками розподілити рослини на групи.

	A	B	C
1	№	x	y
2	1	2	8
3	2	4	10
4	3	5	7
5	4	12	6
6	5	14	6
7	6	15	4

Побудуємо точкову діаграму розсіювання для досліджуваних об'єктів: по осі  $X$  будемо відкладати висоту стебла, по осі  $Y$  – довжину кореня (рис. 10.1). Візуально можна визначити, що рослини розподілилися на дві групи. Одна група має коротке стебло і довге коріння, інша, навпаки, довге стебло і коротке коріння. Очевидно, що відстань між рослинами, які належать до однієї групи, менша, ніж відстань до другої групи рослин.

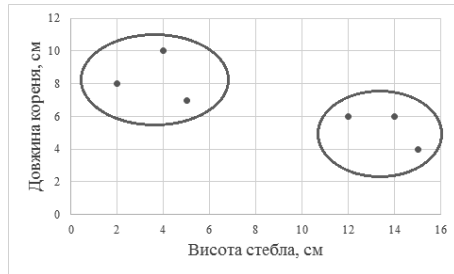


Рис. 10.1. Діаграма розсіювання для даних прикладу 10.1

**Крок 1.** Сформуємо квадратну матрицю, де назвами стовпчиків і рядків будуть номери рослин. На перетині рядків та стовпців розрахуємо евклідові відстані між об'єктами, як показано на рис. 10.2.

		G2									
		= SQRT ((B3-B2)^2+(C3-C2)^2)									
	A	B	C	D	E	F	G	H	I	J	K
1	№	x	y	№	1	2	3	4	5	6	
2	1	2	8	1	0	2,83	3,16	10,20	12,17	13,60	
3	2	4	10	2	2,83	0	3,16	8,94	10,77	12,53	
4	3	5	7	3	3,16	3,16	0	7,07	9,06	10,44	
5	4	12	6	4	10,20	8,94	7,07	0	2,00	3,61	
6	5	14	6	5	12,17	10,77	9,06	2,00	0	2,24	
7	6	15	4	6	13,60	12,53	10,44	3,61	2,24	0	

Рис. 10.2. Процедура проведення кластерного аналізу. Крок 1

**Крок 2.** В отриманій матриці на перетині першого рядка та другого стовпця буде знаходитися відстань між першою та другою рослинами, першого рядка

та третього стовпця – відстань між першою та третьою рослинами тощо. В результаті отримаємо діагональну матрицю ( $X_{ij} = X_{ji}$ ), на головній діагоналі якої знаходяться нульові значення. Серед отриманих відстаней знаходимо мінімальне значення. В нашому прикладі це 2,00, яке знаходиться на перетині четвертого рядка та п'ятого стовпця (а також на перетині п'ятого рядка і четвертого стовпця). Отже, найближчими один до одного об'єктами є рослини з номерами 4 і 5, і ці об'єкти можна об'єднати в одну групу під назвою [4, 5] – об'єднуємо два стовпці (четвертий та п'ятий), залишаючи з двох сусідніх розрахованих відстаней найменше значення. Після об'єднання матриця відстаней буде мати такий вигляд (рис. 10.3):

E	F	G	H	I	J
№ n/п	1	2	3	[4,5]	6
1	0	2,83	3,16	10,20	13,60
2	2,83	0	3,16	8,94	12,53
3	3,16	3,16	0	7,07	10,44
[4,5]	10,20	8,94	7,07	0	2,24
6	13,60	12,53	10,44	2,24	0

Рис. 10.3. Процедура проведення кластерного аналізу. Крок 2

**Крок 3.** З нової матриці видно, що можна об'єднати в один кластер об'єкти [4, 5] і 6 (як найбільш близькі один до одного за значеннями – евклідова відстань дорівнює 2,24). Ми залишаємо найменші значення відстаней з кожної пари об'єднаних стовпців (рядків) і формуємо нову матрицю (рис. 10.4):

E	F	G	H	I
№ n/п	1	2	3	[4,5,6]
1	0	2,83	3,16	10,20
2	2,83	0	3,16	8,94
3	3,16	3,16	0	7,07
[4,5,6]	10,20	8,94	7,07	0

Рис. 10.4. Процедура проведення кластерного аналізу. Крок 3

**Крок 4.** Об'єкти 1 і 2 можна об'єднати в один кластер (як найбільш близькі з наявних). Вибираємо найменше значення і формуємо нову матрицю відстаней. У результаті отримуємо три кластери (рис. 10.5):

E	F	G	H
№ n/п	[1,2]	3	[4,5,6]
[1,2]	0	3,16	8,94
3	3,16	0	7,07
[4,5,6]	8,94	7,07	0

Рис. 10.5. Процедура проведення кластерного аналізу. Крок 4

**Крок 5.** Найближчі об'єкти – 1, 2 і 3. Об'єднаємо їх (рис. 10.6). В результаті отримали два кластери: до першого кластеру потрапили об'єкти 1, 2 і 3 – рослини з коротким стеблом та довгим корінням, до другого кластеру – об'єкти 4, 5 і 6 – рослини з довгим стеблом та коротким корінням. Відстань між кластерами – 7,07.

E	F	G
№ п/п	[1,2,3]	[4,5,6]
[1,2,3]	0	7,07
[4,5,6]	7,07	0

Рис. 10.6. Процедура проведення кластерного аналізу. Крок 5

### Кластерний аналіз у програмі Statistica

Розглянемо реалізацію кластерного аналізу на даних прикладу 10.1.

З пункту меню *Statistics*, вкладка *Advanced / Multivariate* (Розширений / Багатомірний аналіз) натискаємо кнопку *Mult / Exploratory* (Багатомірний / Дослідницький) і вибираємо команду *Cluster* (Кластер). Відкривається діалогове вікно модуля *Clustering Method* (Методи кластеризації) (рис. 10.7)

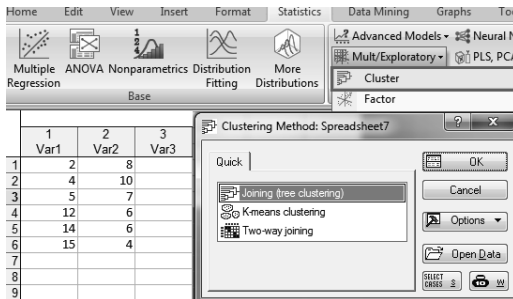


Рис. 10.7. Діалогове вікно модуля методів кластеризації

Діалогове вікно модуля *Clustering Method* передбачає кілька різних методів розподілу об'єктів на кластери:

- *Joining (tree clustering)* – приєднання (деревоподібна кластеризація);
- *K-means clustering* – кластеризація методом k-середніх;
- *Two-way joining* – двохідне приєднання.

Розглянемо основні принципи класифікації методів кластерного аналізу.

У загальному випадку методи кластерного аналізу можуть бути розділені на **ієрархічні**, якщо результуюче угруповання має все більшу кількість вкладених класів, що нагадують філогенетичну класифікацію, та **неієрархічні (ітеративні)**, якщо результати виражаються як унікальний розділ цілої сукупності об'єктів.

*Ієрархічні алгоритми* можуть бути розділювальними або агломеративними. *Метод розділення* починається з усіх випадків в одному кластері. Це скупчення поступово розпадається на все менші та менші скупчення.

*Агломеративні методи* зазвичай починаються з одноступових кластерів, які послідовно зливаються, поки не утворюється один великий кластер. На початковому етапі два об'єкти з найменшою відстанню (або найбільшою подібністю) об'єднуються в кластер. На наступному кроці ідентифікується і вивчається об'єкт, що має найменшу відстань до будь-якого з перших двох. Якщо він ближче до четвертого об'єкта, ніж до будь-якого з перших двох, третій та четвертий об'єкти стають другим кластером із двох випадків; інакше третій об'єкт вноситься до першого кластера. Процес повторюється, додаючи випадки до наявних кластерів, створюючи нові кластери або комбінуючи ті, що виникли, до тих пір, поки кожен об'єкт не буде досліджений і виділений в один кластер або не стане окремим кластером сам по собі. *Joining (tree clustering)* належить до ієрархічного методу кластеризації.

До неієрархічної методики кластеризації належить *алгоритм k-середніх*, який базується на принципах дисперсійного аналізу. Насправді це можна розглядати як аналіз дисперсії у зворотному напрямі: алгоритм розпочинається із визначення  $k$  випадкових кластерів, а потім між кластерами відбувається переміщення об'єктів з метою мінімізації мінливості всередині кластерів та максимізації мінливості між кластерами.

Третій метод – *two-way joining* – застосовується для одночасної кластеризації об'єктів та їх параметрів (змінних). Наприклад, маємо дані про різні показники фізичної підготовленості (змінні параметри) для вибірки пацієнтів із серцево-судинними захворюваннями (об'єкти). За допомогою *two-way joining* можна згрупувати як об'єкти (пацієнтів) для виявлення груп хворих пацієнтів із подібними синдромами, так і змінні (міри придатності), щоб виявити скупчення заходів, які впливають на подібні фізичні здібності.

Розглянемо перший метод – деревоподібну кластеризацію на даних прикладу 10.1. Після вибору методу кластеризації *Joining (tree clustering)* у діалоговому вікні *Cluster Analysis: Joining (Tree Clustering) (Кластерний аналіз: деревовидна кластеризація)* на вкладці *Advanced* вибираємо змінні  $x$  – висота стебла та  $y$  – довжина коріння (кнопка *Variables*). Усі інші параметри виставляємо так, як показано на рис. 10.8.

У полі *Input file (Вхідний файл)* можливий вибір одного з двох пунктів: *Raw data (Рядки даних)*, якщо у файлі з вхідними даними параметри для досліджуваних об'єктів розташовані в рядках, або *Distance matrix (Матриця відстаней)*, якщо файл містить вже розраховану матрицю подібності.

У полі *Cluster* вибираємо *Variables (columns)*, якщо параметри об'єктів знаходяться в стовпцях, і *Cases (rows)* – якщо в рядках.

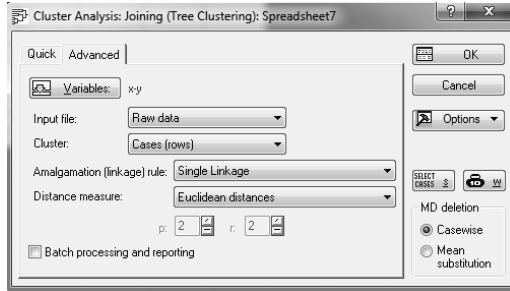


Рис. 10.8. Діалогове вікно *Cluster Analysis: Joining (Tree Clustering)*

Поле *Amalgamation (linkage) rule* (Правило об'єднання (зв'язування)) дає змогу вибрати метод розрахунку відстані між кластерами. Коли кластери утворені двома або більше об'єктами, для обчислення цих відстаней повинні бути визначені певні правила. Одним із популярних методів є *Single linkage* (Одинарний зв'язок, або Метод найближчого сусіда) – був використаний у прикладі 10.1. У цьому методі відстань між двома кластерами визначається мінімальною відстанню двох об'єктів (найближчих сусідів), що належать до різних кластерів.

Поле *Distance measure* (Міра відстані) визначає метод розрахунку відстані між об'єктами в кластері. Вибір типу відстані є важливим для всіх ієрархічних алгоритмів кластеризації і залежить від природи змінних та очікуваної форми кластерів. Серед багатьох типів відстаней найпоширенішою є евклідова відстань, сутність якої було розглянуто вище.

Після встановлення всіх параметрів кластеризації натискаємо кнопку *OK* – на екрані з'явиться діалогове вікно *Joining results*, в інформаційній частині якого вказано параметри кластеризації, а функціональна частина містить кнопки для побудови горизонтальної або вертикальної дендрограми (рис. 10.9) і підсумкової матриці відстаней (рис. 10.10).

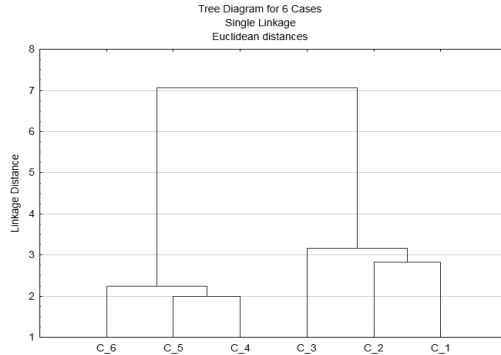


Рис. 10.9. Вертикальна дендрограма кластеризації об'єктів дослідження

Дендрограма (dendrogram) – це графічне зображення кластеру, яке відображає зв'язки між групами об'єктів з довжиною ліній, які представляють «несхожість» цих об'єктів. Дендрограма містить  $n$  рівнів, що відповідають процесу послідовного формування кластерів. Проведіть аналогію з покроковим формуванням кластерів у прикладі 10.1 у програмі Excel.

Case No.	C 1	C 2	C 3	C 4	C 5	C 6
C 1	0,0	2,8	3,2	10,2	12,2	13,6
C 2	2,8	0,0	3,2	8,9	10,8	12,5
C 3	3,2	3,2	0,0	7,1	9,1	10,4
C 4	10,2	8,9	7,1	0,0	2,0	3,6
C 5	12,2	10,8	9,1	2,0	0,0	2,2
C 6	13,6	12,5	10,4	3,6	2,2	0,0

Рис. 10.10. Матриця відстаней (або подібностей) між парами об'єктів

Далі проведемо кластеризацію об'єктів методом  $k$ -середніх. Після вибору методу кластеризації *K-means clustering* у діалоговому вікні *Cluster Analysis: K-Means Clustering* (Кластерний аналіз: кластеризація методом  $k$ -середніх) на вкладці *Advanced* вибираємо змінні  $x$  – висота стебла та  $y$  – довжина коріння (кнопка *Variables*). Всі інші параметри виставляємо так, як показано на рис. 10.11. Це діалогове вікно містить опції:

- *Number of clusters* (Кількість кластерів). Зробимо припущення, що досліджувані об'єкти належать до двох кластерів.

Усі інші параметри можна залишити без змін, а саме:

- *Number of iterations* (Кількість ітерацій) – оскільки метод  $k$ -середніх ітеративний, то вказується максимальна кількість кроків, які необхідно виконати (якщо розділення на кластери не відбудеться за меншу кількість ітерацій).

- *Initial cluster centers (Початкові центри кластерів)* – розділ, в якому задається правило визначення центру кластеру. Цей параметр змінюють, якщо кластери не визначаються за задану кількість ітерацій.

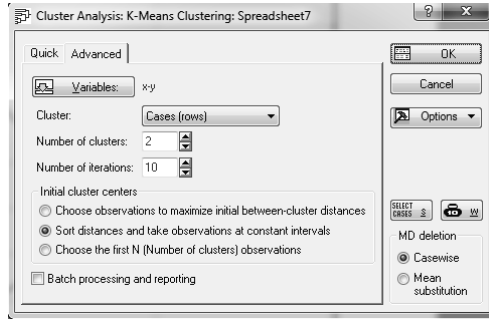


Рис. 10.11. Діалогове вікно Cluster Analysis: K-Means Clustering

Після встановлення всіх параметрів кластеризації натискаємо кнопку *OK*. На екрані з'явиться діалогове вікно *K-means clustering results (результати кластеризації методом k-середніх)*, в інформаційній частині якого вказано параметри кластеризації, а функціональна частина містить кнопки для виводу результатів поділу на кластери. Кнопка *Graph of means (Графіки середніх)* дозволяє переглянути середні значення для кожного кластера на лінійному графіку. Таке розташування координат центрів кластерів вказує на гарну якість кластеризації – рослини розділені на дві групи, які відрізняються своїми параметрами. До першого кластеру віднесено рослини з довгим стеблом (змінна *x*) і коротким корінням (*y*) – синій графік; до другого кластеру – рослини з коротким стеблом та довгим корінням – червона лінія (рис. 10.12)

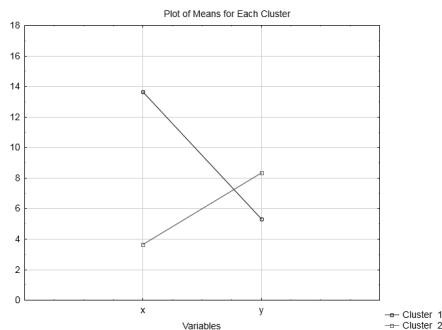


Рис. 10.12. Результати кластеризації методом k-середніх

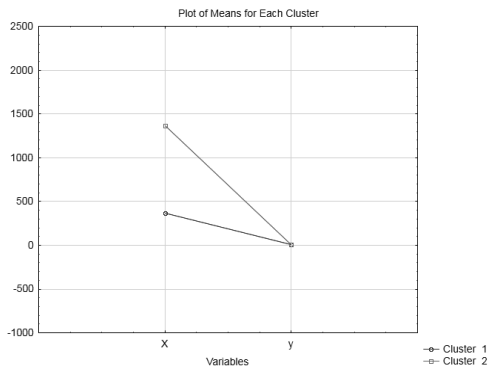
### **Стандартизація параметрів при проведенні кластерного аналізу**

Коли для обчислення відстані використовуються дві або більше змінних, домінуватиме змінна з більшою величиною. Щоб цього уникнути, загальноприйнятою практикою є стандартизація всіх змінних перед початком аналізу. Стандартизувати – означає від значень кожної ознаки відняти його середнє значення і розділити на стандартне відхилення.

Припустимо, що значення параметра  $x$  на кілька порядків більше ніж змінна  $y$ . Для наочності візьмемо дані прикладу 10.1, але значення параметра  $x$  збільшимо у 100 разів (цю мінну позначимо як  $X$ ):

$X$	$y$
200	8
400	10
500	7
1200	6
1400	6
1500	4

У такому разі при кластеризації об'єктів методом  $k$ -середніх отримаємо такий графік (рис. 10.13):



*Рис. 10.13. Результати кластеризації методом  $k$ -середніх для змінних, значення яких відрізняються на кілька порядків*

Візуально параметр  $y$  не відрізняється в першому та другому кластерах і дорівнює 0. Однак такі результати недостовірні з огляду на різницю розмірності даних та їх варіабельності. Щоб уникнути такої ситуації, виконаємо стандартизацію стовпців і ще раз запустимо кластерний аналіз.

Для цього клацнемо правою кнопкою миші на будь-якому стовпчику даних і з контекстного меню виберемо команду *Fill / Standardize Block* (Заповнити / Стандартизувати блок), *Standardize Columns* (Стандартизувати стовпці) (якщо значення параметрів розташовані в стовпцях, як у нашому випадку). Стандартизовані значення параметрів представлені на рис. 10.14.

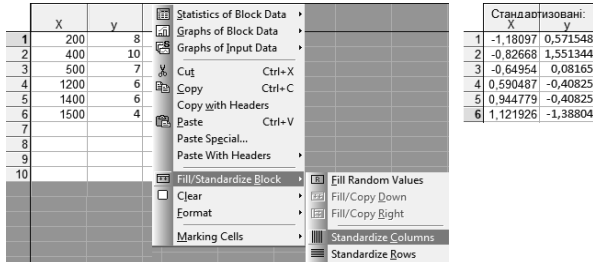


Рис. 10.14. Процедура і результати стандартизації в програмі Statistica

Після стандартизації ще раз проведемо кластерний аналіз методом  $k$ -середніх. Тепер, як і в прикладі 10.1, маємо два кластери, що відрізняються за двома досліджуваними параметрами (рис. 10.15)

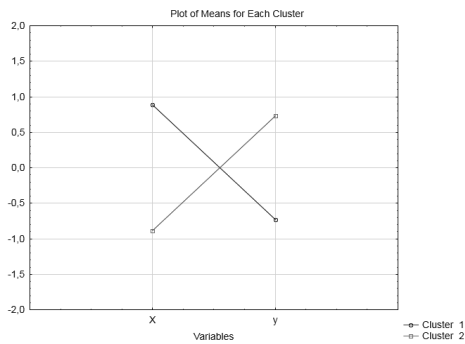


Рис. 10.15. Результати кластерного аналізу після стандартизації

### **Завдання для самостійної роботи**

1. Для даних, наведених у розділі «Варіанти завдань для закріплення теми 10», виконати кластеризацію методом «найближчого сусіда» у програмі Excel.
2. Для даних, наведених у розділі «Варіанти завдань для закріплення теми 10», виконати кластерний аналіз методом деревоподібної дендрограми та методом k-середніх у програмі Statistica.

### **Контрольні запитання до теми 10**

1. Для чого використовується кластерний аналіз? Наведіть приклади використання цього методу в біологічних дослідженнях.
2. Який основний принцип групування в кластерному аналізі? Наведіть приклад визначення відстані між кластерами.
3. Які алгоритми розділення об'єктів використовують у кластерному аналізі?
4. До якої методики кластеризації належить алгоритм k-середніх? На яких принципах базується цей алгоритм?
5. Що таке дендрограма? До яких алгоритмів кластеризації належить деревоподібна кластеризація? На яких принципах ґрунтується цей метод?
6. В якому випадку при проведенні кластерного аналізу використовується стандартизація параметрів?

## ТЕМА 11. ДИСКРИМІНАНТНИЙ АНАЛІЗ

Дискримінантний аналіз – це статистичний метод, призначений для виявлення відмінностей між спостереженнями, віднесеними до заздалегідь визначених двох або більшої кількості груп об'єктів за певними ознаками. Основна задача дискримінантного аналізу – знайти найбільш значущі (тобто дискримінаційні) ознаки, які відрізняють об'єкти, що належать до однієї групи, і визначити дискримінантну функцію, яка найкраще відображає відмінності між досліджуваними групами. Наступний крок дискримінантного аналізу – це класифікація об'єктів, приналежність яких до тієї чи іншої групи заздалегідь невідома, з використанням дискримінантної функції, створеної з набору «навчальних» експериментальних даних.

У біології дискримінантний аналіз використовується для класифікації таксонів, морфометричного аналізу, для ідентифікації та розподілу видів. Для екологів вибір відповідної підмножини змінних також є важливим методологічним питанням. Часто існує інтерес до отримання загальної можливості прогнозування або пошуку причинно-наслідкових висновків із предикторських змінних (параметрів навколишнього середовища).

Наприклад, необхідно дослідити, які параметри (характеристики) відрізняють фрукти, які їдять примати або птахи. З цією метою можна було б збирати дані про численні характеристики плодів тих видів, які їсть кожна із груп тварин. Більшість фруктів природним способом потрапляє в одну із двох категорій. Використовуючи дискримінантний аналіз, можна визначити, які з параметрів найкраще передбачають, чи будуть плід їсти птахи або примати.

Розглянемо проведення дискримінантного аналізу в програмі Statistica на модельному прикладі (постачається разом із програмним забезпеченням і знаходиться в папці *Examples / Datasets*, файл *Irisdat.sta*). Доступ до цієї папки можна отримати, вибравши пункт меню *Home*, розділ *File*, кнопка *Open*, команда *Open Examples*.

Файл базується на класичній множині даних Р. Фішера (засновника методу) і містить дані про довжину і ширину чашолистків і пелюсток трьох типів ірисів. Метою аналізу є вивчення дискримінації між трьома типами квітів, ґрунтуючись на чотирьох вимірах: ширини і довжини чашолистків і пелюсток. Частина файлу наведено на рис. 11.1. Перші дві змінні в цьому файлі – довжина чашолистків

	1	2	3	4	5
	SEPALLEN	SEPALWID	PETALLEN	PETALWID	IRISTYPE
1	5.0	3.3	1.4	0.2	SETOSA
2	6.4	2.8	5.6	2.2	VIRGINIC
3	6.5	2.8	4.6	1.5	VERSICO
4	6.7	3.1	5.6	2.4	VIRGINIC
5	6.3	2.8	5.1	1.5	VIRGINIC
6	4.6	3.4	1.4	0.3	SETOSA
7	6.9	3.1	5.1	2.3	VIRGINIC
8	6.2	2.2	4.5	1.5	VIRGINIC
9	5.9	3.2	4.8	1.8	VERSICO
10	4.6	3.6	1.0	0.2	SETOSA
11	6.1	3.0	4.6	1.4	VERSICO
12	6.0	2.7	5.1	1.6	VERSICO
13	6.5	3.0	5.2	2.0	VIRGINIC
14	5.6	2.5	3.9	1.1	VERSICO
15	6.5	3.0	5.5	1.8	VIRGINIC
16	5.8	2.7	5.1	1.9	VIRGINIC

Рис. 11.1 Файл із вихідними даними для дискримінантного аналізу

(*Sepallen*) і ширина чашолистків (*Sepalwid*); наступні дві змінні – довжина пелюсток (*Petalen*) і ширина пелюсток (*Petalwid*). Остання змінна кодує приналежність до певної сукупності, тобто до видів ірису – *Iris setosa*, *I. versicol* і *I. virginic*. Загалом у цій вибірці представлено 150 квіток, по 50 кожного виду.

Для відображення стартової панелі модуля Дискримінантний аналіз обираємо пункт меню Statistics (Статистика), вкладка *Advanced / Multivariate (Розширений / Багатовимірний аналіз)* натискаємо кнопку *Mult / Exploratory (Багатовимірний / Дослідницький)* і вибираємо команду *Discriminant (Дискримінантний аналіз)*. У діалоговому вікні *Discriminant Function Analysis* натисніть кнопку *Variables (Змінні)* і оберіть гуртуючу змінну – *Iristype* і незалежні змінні – *Sepallen*, *Sepalwid*, *Petalen* і *Petalwid*.

Потім необхідно вказати коди (кнопка *Codes for grouping variable*), щоб обрати, які типи ірисів будуть враховані в дискримінантному аналізі. У діалоговому вікні *Select codes for grouping variable (Вибір кодів для гуртуючої змінної)* натисніть *All* для вибору всіх типів ірисів (рис. 11.2).

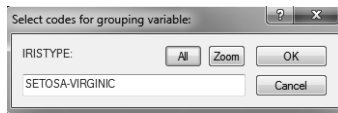


Рис. 11.2. Діалогове вікно для вибору кодів гуртуючої змінної

Якщо коди гуртуючої змінної не обрано, то аналіз буде виконуватися для всіх варіантів гуртуючої змінної за замовчуванням.

Виконаємо дискримінантний аналіз із налаштуваннями за замовчуванням. Його результат показано на рис. 11.3.

Discriminant Function Analysis Summary (Iridat)						
No. of vars in model: 4; Grouping: IRISTYPE (3 grps)						
Wilks' Lambda: .0234386 approx. F (8,288)=199,1453 p<0,0000						
	Wilks' Lambda	Partial Lambda	F-remove (2,144)	p-value	Toler.	1-Toler. (R-Sqr.)
N=150						
SEPALLEN	0.024976	0.938464	4.72145	0.010329	0.347993	0.652007
SEPALWID	0.030580	0.766480	21.93593	0.000000	0.608859	0.391141
PETALLEN	0.035025	0.669206	35.59018	0.000000	0.365126	0.634874
PETALWID	0.031546	0.743001	24.90433	0.000000	0.649314	0.350686

Рис. 11.3. Діалогове вікно результатів дискримінантного аналізу і підсумкова таблиця для прикладу 11.1

Верхня частина діалогового вікна *Discriminant Function Analysis Results* (Результати дискримінантного аналізу) містить інформаційну панель, в якій розміщено основні статистичні дані для інтерпретації результатів аналізу:

- *Number of variables in the model* – кількість змінних у моделі;
- *Wilks' Lambda* – статистика Лямбда Уїлкса з урахуванням усіх задіяних змінних – це відношення міри внутрішньогрупової мінливості до міри загальної мінливості. Внутрішньогрупова мінливість – частина загальної, і це означає, що лямбда Уїлкса може набувати значення від 0 (групи повністю однорідні) до 1 (розподіл об'єктів на групи не призводить до того, що внутрішньогрупова мінливість виявляється менше загальної). Отже, чим менше значення має лямбда Уїлкса, тим якіснішим виявляється поділ на групи при дискримінантному аналізі;
- *approx. F і p* – наближене значення *F*-критерію і *p* – рівень значущості для *F*-критерію; велике значення *F*-критерію і низький рівень значущості показують, що знайдена дискримінантна функція значуща.

Нижня частина діалогового вікна містить вкладки та кнопки для детального представлення результатів аналізу дискримінантних функцій.

Після натискання на кнопку *Summary: Variables in the model* (Підсумки: змінні в моделі) на вкладці *Quick* (Швидкий аналіз) діалогового вікна *Discriminant Function Analysis Results* (Результати дискримінантного аналізу) з'являється підсумкова таблиця (рис. 11.3), яка містить таку інформацію:

- *Wilks' Lambda* – значення лямбда Уїлкса для аналізу, в якому ця змінна *не використовується*. Якщо виключення якоїсь змінної з аналізу призвело до істотного погіршення результату, то ми можемо стверджувати, що ця змінна вносила в нього важливий внесок, то ж, чим вище значення лямбда Уїлкса в першому стовпці вікна *Discriminant Function Analysis Summary*, тим важливіша ця ознака;
- *Partial Lambda* – приватна лямбда. Ця статистика показує відношення лямбди Уїлкса після додавання цієї змінної до лямбди Уїлкса до додавання змінної. Якщо змінна робить хоча б якийсь внесок у розділення груп, після її додавання лямбда Уїлкса повинна зменшитися. Тому, чим меншим виявляється значення приватна лямбди, тим ціннішою є ця ознака;
- *F-remove* – *F*-критерій, пов'язаний із виключенням даної ознаки з аналізу, а *p-level* – це рівень його статистичної значущості. Якщо виключення ознаки призводить до статистично значущої зміни співвідношення дисперсій, то ця ознака робить важливий внесок у дискримінацію груп;
- *Toler.* – толерантність – це міра надмірності ознаки, яка обчислюється як  $1 - R^2$ , де  $R^2$  – коефіцієнт множинної кореляції даної ознаки з усіма іншими ознаками, які використовуються в аналізі. Чим нижчою є толерантність, тим сильніше цей показник пов'язаний з усіма іншими.

Ще однією важливою мірою, що дає змогу оцінити якість поділу об'єктів за групами, є відсоток коректно класифікованих об'єктів. Щоб дізнатися його, треба в діалоговому вікні *Discriminant Function Analysis Result* перейти на вкладку *Classification (Класифікація)* (рис. 11.4).

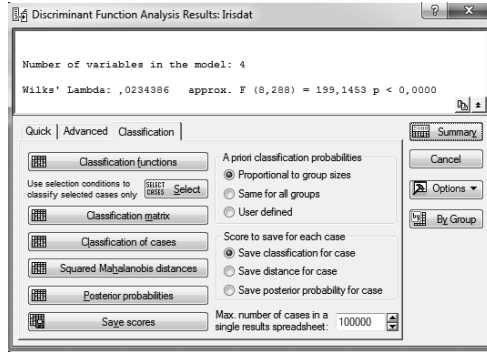


Рис. 11.4. Діалогове вікно *Discriminant Function Analysis Result*, вкладка *Classification*

Щоб оцінити коректність отриманого в результаті аналізу розподілу за групами, треба натиснути на кнопку *Classification Matrix (Матриця класифікації)*. В результаті отримаємо матрицю, в рядках якої вказано, до яких груп насправді належать розглянуті рослини; стовпці – відображення того, як ці об'єкти були б класифіковані (за використаними змінними), якби було невідомо, до якої групи рослина належить.

На рис. 11.5 з матриці класифікації можна побачити, що аналіз на 100 % правильно визначив приналежність рослин за ознаками до виду *Setosa*, на 96 % – до виду *Versicol* (2 рослини було віднесено до іншого виду) і на 98 % – до виду *Virginic*.

Classification Matrix (Irisdat)				
Rows: Observed classifications				
Columns: Predicted classifications				
Group	Percent Correct	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
SETOSA	100,0000	50	0	0
VERSICOL	96,0000	0	48	2
VIRGINIC	98,0000	0	1	49
Total	98,0000	50	49	51

Рис. 11.5. Матриця класифікації для даних прикладу 11.1

Зручніше подивитися на розподіл класифікованих об'єктів на графіку. Для цього виконаємо канонічний аналіз: на вкладці *Advanced* діалогового вікна

*Discriminant Function Analysis Results* натискаємо кнопку *Perform canonical analysis* (Виконання канонічного аналізу). У діалоговому вікні *Canonical Analysis*, що відкрилося, потрібно перейти на вкладку *Canonical scores* (Канонічні корені) (рис. 11.6) і натиснути кнопку *Scatterplot of canonical scores* (Графіки розподілу канонічних коренів).

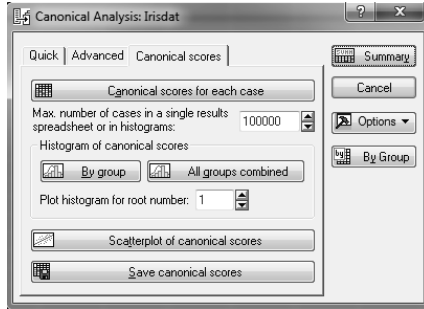


Рис. 11.6. Діалогове вікно *Canonical scores*

Результат представлено на рис. 11.7.

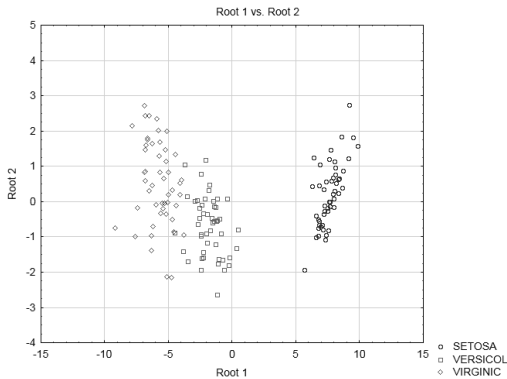


Рис. 11.7. Діаграми розсіювання канонічних коренів

З отриманих результатів можна побачити, що перший канонічний корінь (*Root1*) відображає значні відмінності виду *Setosa* від інших рослин. Розділення видів *Versicol* і *Virginic* не абсолютне і має незначне перекриття, з яким пов'язані випадки, коли класифікація була некоректною.

Канонічні корені – це ті самі канонічні дискримінантні функції, які вибираються так, щоб найкраще відображати відмінності між групами об'єктів. Ці корені є лінійними комбінаціями дискримінантних змінних (у граничному випадку – відповідають якимось із змінних). Щоб подивитися, як пов'язані канонічні корені і дискримінантні змінні, треба натиснути на кнопку *Factor structure* на вкладці *Advanced* у діалоговому вікні *Canonical Analysis*, де можна побачити коефіцієнти кореляції, що зв'язують кожен канонічний корінь із кожною дискримінантною змінною.

### **Завдання для самостійної роботи**

1. Для даних, наведених у розділі «Варіанти завдань для закріплення теми 11», виконати дискримінантний аналіз у програмі Statistica.

### **Контрольні запитання до теми 11**

1. Що таке дискримінантний аналіз? Наведіть приклад використання дискримінантного аналізу в біологічних дослідженнях.
2. Які основні статистичні параметри використовують для інтерпретації результатів дискримінантного аналізу?
3. Які критерії використовують для визначення значущості дискримінантної функції?
4. Що таке канонічні корені дискримінантної функції? На що вказують значення канонічних коренів?

## ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

### Варіанти завдань для закріплення тем 1–3

#### Варіант 1

*Тривалість серцевого циклу в кардіограмі здорових дітей:* 0,91 0,71 0,73 0,82 0,67 0,89 0,90 1,00 0,77 0,78 0,72 0,71 0,90 0,68 0,52 0,58 0,59 0,56 0,74 0,54 0,72 0,76 0,74 0,79 0,66 0,84 0,85 0,81 1,11 0,77 0,84 0,59 0,71 0,74 0,65 0,96 0,83 0,78 0,94 0,93 0,62 0,60 0,57 0,69 0,56 0,45 0,93 0,74 0,86 0,55 0,66 0,49 0,71 0,68 0,96 0,78 0,85 0,54 0,64 0,82 0,60 0,71 0,76 0,56 0,79 0,66 0,76 0,65 0,66 0,72 0,92 0,74 0,66 0,61 0,67 0,60 0,69 0,63 0,80 0,76 0,65 0,85 0,68 0,80 0,86 0,56 0,64 0,69 0,49 0,56 0,74 0,87 0,69 0,62 0,61 0,97 0,83 0,66 0,78 0,89 0,75 0,90 0,81 0,59 0,68 0,53 0,59 0,75 0,81 0,73 0,85 0,72 0,79 0,86 0,54 0,62 0,94 0,57 0,75 0,73 0,52 0,68 0,85 0,75 0,79 0,72

#### Варіант 2

*Тривалість серцевого циклу в кардіограмі хворих дітей:* 0,91 0,86 0,74 0,51 1,07 0,79 0,98 1,16 0,74 0,88 0,84 0,66 0,73 0,91 1,22 1,43 0,81 0,72 0,79 0,98 0,62 1,37 0,94 0,77 0,71 0,85 0,85 0,57 0,83 1,27 1,06 0,73 0,61 0,67 0,75 0,91 1,05 0,54 0,85 0,82 0,79 0,64 0,97 1,17 1,06 0,82 0,58 0,93 0,85 0,96 1,08 0,67 0,74 0,84 1,03 0,97 0,89 0,63 0,77 0,86 0,61 0,74 0,72

#### Варіант 3

*Кількість частинок вугілля в одній вакуолі у парамецій при впливі 0,04 % розчину пептону:* 6 5 6 6 6 3 4 4 5 6 6 4 5 5 5 5 5 4 4 3 6 5 5 4 5 4 4 5 5 5 5 6 4 5 3 4 5 5 6 5 5 4 4 6 5 4 4 3 6 5 5 4 3 5 3 4 4 5 5 4 6 5 5 3 4 4 4 5 5 6 6 3 5 3 6 4 5 3 5 6 3 4 3 6 5 3 4 4 6 6

#### Варіант 4

*Кількість частинок вугілля в одній вакуолі у парамецій в контрольній групі:* 4 3 3 0 2 2 2 3 1 0 0 1 2 2 2 2 3 4 2 2 3 0 2 1 3 2 2 3 2 1 2 2 1 2 1 2 4 0 0 2 3 3 2 1 0 1 2 2 3 1 0 0 4 2 2 2 4 2 1 2 2 3 2 3 2 2 2 3 1 2 0 2 0 0 2 1 2 3 3 4 0 1 2 1 2 3 3

#### Варіант 5

*Число стегнових пір у живородних ящірок:* 8 9 11 10 11 10 8 9 9 7 9 9 9 7 11 10 10 11 11 9 10 9 10 9 10 10 10 10 11 8 11 10 9 11 10 10 12 10 9 8 10 10 9 10 8 10 9 9 8 9 8 10 8 9 8 9 9 9 8 10 11 10 11 14 11 11 9 11 9 9 9 9 10 10 11 12 9 12 9 13 9 10 11 11 11 11 11 11 11 10 11 10 10 9 10 9 9 11 12 11 10 11 12 11 10 11 9 11 12 8 14 12 15 12 11 13 12 12 11 10 12 11 10 12 11 10 11 9 5 12 10 11 11 9 11 10 11 10 11 11 12 11 9 11 10 7 7 10 11 9 6 7 8 11

**Варіант 6**

*Розміри крипт в ободовій кишці у щурів (в розподілах окуляр-мікрометра) без операції:* 7 10 8 12 16 9 10 7 9 9 14 10 10 8 9 11 9 8 10 11 9 10 11 10 9 8 13 9 7 9 10 8 9 9 10 9 8 10 10 11 8 9 7 8 9 10 10 10 9 10 10 11 12 9 10 10 8 12 10 9 8 7 10 11 12 13 8 8 8 13 12 10 9 8 7 11 12 14 10 10 8 8 7 9 10 11 9 8 8 7 9 10 11 14 10 7 13 10 9 12 12 10 8 9 8 7 10

**Варіант 7**

*Розміри крипт в ободовій кишці у щурів (в розподілах окуляр-мікрометра) після видалення сліпої кишки:* 10 11 7 14 10 8 17 8 10 11 7 12 9 10 12 13 8 11 8 10 15 8 10 20 13 8 11 12 9 12 9 12 12 11 10 9 11

**Варіант 8**

*Число фрагментів хромосом на тисячу клітин:* 328 521 421 461 410 481 492 501 499 505 401 502 383 470 482 441 540 541 522 459 506 523 518 450 483 385 493 477 440 504 497 449 430 498 499 425 515 439 491 530 530 460 502 413 475 477 543 387 531 478 443 539 457 507 443 417 556 471 489 560 560 418 542 453 508 479 519 419 527 389 469 462 545 426 463 561 451 519 431 528 496 562 484 479 500 438 520 529 453 495 485 447 581 471 503 486 449 461 563 481 570 571 470 488 590 461 515 517 463 500 487 539 490 591 489 530 531 494 599 481

**Варіант 9**

*Кількість хромосомних містків в полі зору:* 17 18 19 17 20 21 20 22 19 18 21 22 18 20 17 21 19 20 22 19 20 18 21 19 20 17 21 19 17 19 20 22 20 17 21 22 22 19 17 18 22 19 20 17 18 19 20 17 21 20 22 19 18 21 22 18 20 17 21 19 20 22 19 20 18 21 19 20 17 21 19 17 20 22 22 20 17 21 22 19 17 18 22 19 20 17 18 19 17 20 21 20 22 19 18 21 22 18 21 18 21 22 19 20 17 21 18 20 19 22 20 18 19 20 21 19 20 17 21 19 17 20 22 22 20 17 21 22 19 17 18 22

**Варіант 10**

*Кількість язичкових квіток у суцвіттях ромену звичайного (Карпати) на сухому лузі:* 23 23 20 18 21 25 19 29 24 23 21 27 21 20 20 8 16 12 24 21 20 15 21 22 18 19 17 20 20 22 21 21 20 20 21 13 15 14 21 20 22 18 21 18 13 15 22 16 21 14 18 21 25 13 12 13 17 14 15 17 20 20 16 17 21 28 21 21 24 19 21 20 14 13 14 16 20 15 14 20 13 20 19 18 13 17 20 21 21 22 21 19 15 13 21 21 10 14 13 13 16 17 13 21 21 19 14 18 12 27 21 13 8 8 16 21 18 13 14 15 13 20 19 21 18 16 22 20 20 16 20 13 16 20 18 12 12 11 17 13 14 20 16 15 14 17 15 21 15 12 8 21 21 13 16 16 12 19 13 21 14 18 14 13 15 17 19 18 15 17 13 11 15 16 14 13 13 20 13 13 12 13 13 13 10 21 18 17 17 21 20 13 10 13 14 13 14 16 12

**Варіант 11**

*Кількість язичкових квіток у суцвіттях ромену звичайного (Карпати) на вологому лузі:* 25 20 25 26 21 24 16 28 23 28 28 32 23 23 26 22 29 34 27 28 25 27 27 25 25 23 29 22 24 23 23 27 21 32 24 36 27 29 22 23 24 22 24 24 30 34 26 22 22 23 27 26 24 27 28 22 21 23 24 32 31 21 23 24 28 25 26 28 21 22 24 22 24 23 28 23 23 24 23 23 23 21 35 34 33 30 28 34 37 32 31 23 28 33 22 32 34 24 32 24 27 23 23 33 27 34 27 32 33 18 28 24 34 30 32 24 23 28 32 24 21 23 30 26 21 33 30 23 27 25 29 32 30 30 34 26 22 32 23 34 32 23 33 22 26 26 21 26 24 33 27 27 24 24 33 35 24 33 21 21 32 33 23 30 25 30 24 27 28 26 33 35 26 31 31 32 24 21 34 23 26 36 40 31 24 22 23 26 26 31 22 26 31 26 32 30 32 33 27

**Варіант 12**

*Кількість язичкових квіток у суцвіттях ромену звичайного (Карпати) на узліссі:* 21 22 24 26 23 22 24 21 28 30 21 20 21 22 23 18 29 24 20 24 20 22 16 23 24 23 22 21 21 30 20 25 18 22 21 19 22 23 23 19 21 22 27 21 21 21 23 26 16 21 21 25 24 17 23 34 24 34 17 24 23 33 23 33 25 23 22 24 20 21 24 23 16 31 16 21 19 21 26 19 19 21 22 21 20 20 22 23 21 21 22 22 25 29 20 24 21 25 25 21 21 21 25 24 26 21 27 19 20 21 19 21 21 24 20 20 24 24 21 29 23 21 30 25 20 19 21 23 21 21 22 21 24 22 20 22 21 24 21 25 22 21 23 21 23 30 21 20 18 28 27 21 17 22 28 15 34 24 22 22 25 22 29 21 29 24 22 21 24 20 19 30 18 18 17 21 21 21 23 21 21 13 21 17 21 22 29 26 23 23 28 15 24 27 26 22 26 20 19 23

**Варіант 13**

*Кількість язичкових квіток у суцвіттях ромену круглолистого (Карпати) на узліссі:* 27 34 29 31 24 20 21 21 21 22 23 29 25 18 19 21 21 20 21 33 19 21 21 21 21 24 21 21 25 31 20 27 34 23 21 21 21 21 22 25 23 23 22 21 23 20 23 21 21 17 20 21 19 22 22 14 18 18 21 21 24 22 21 22 22 20 20 21 21 21 20 20 21 18 19 29 22 21 15 18 21 26 23 22 21 21 21 21 21 23 21 20 21 19 19 29 21 21 24 24 21 22 28 21 20 25 22 21 21 23 24 16 21

**Варіант 14**

*Довжина зерен пшениці (у мм):* 5,39 5,42 5,38 5,47 5,51 5,30 5,40 5,28 5,43 5,46 5,53 5,55 5,47 5,24 5,44 5,54 5,66 5,43 5,42 5,43 5,52 5,45 5,26 5,33 5,42 5,50 5,49 5,47 5,52 5,48 5,34 5,36 5,59 5,45 5,44 5,34 5,33 5,41 5,54 5,49 5,35 5,40 5,26 5,50 5,46 5,41 5,55 5,31 5,45 5,54 5,32 5,32 5,39 5,62 5,40 5,23 5,45 5,47 5,40 5,42 5,45 5,32 5,44 5,58 5,50 5,36 5,44 5,50 5,37 5,47 5,50 5,44 5,28 5,31 5,64 5,46 5,47 5,57 5,58 5,45 5,37 5,48 5,46 5,51 5,29 5,42 5,69 5,60 5,45 5,38 5,46 5,52 5,43 5,18 5,61 5,36 5,39 5,44 5,41

**Варіант 15**

*Час життя клітини (в з):* 504 507 522 507 519 503 502 511 533 501 504 517  
 510 526 501 520 514 514 504 509 522 503 518 503 504 506 508 504 513 515 510  
 521 502 505 526 503 526 509 504 514 554 500 532 509 502 524 521 529 504 535  
 507 528 500 537 510 501 501 504 501 512 505 506 504 520 513 512 504 521 521  
 533 505 511 537 503 506 501 516 544 542 503 506 500 515 506 522 504 503 509  
 503 501 525 503 504 516 532 530 514 503 512 528 501 510 507 501 533 522 506  
 508 500 505 521 514 502 501 521 524 539 510 500 513 503 536 511 503 509 500  
 501 504 519 514 523 514 527 523 547 509 510 507 538 533 504 512 500 514 509  
 522 501 523 517 507

**Варіант 16**

*Товщина волокна (в см):* 38,2 35,0 34,8 41,0 41,8 50,8 36,7 48,3 42,3 40,8 37,7  
 34,9 43,0 38,1 44,6 40,6 35,7 43,0 45,3 48,3 38,0 45,8 44,0 43,2 38,8 42,1 45,7 41,5  
 41,2 40,7 40,0 39,3 42,3 44,4 42,8 35,1 44,1 36,8 41,1 38,2 39,5 45,2 41,7 49,7 43,1  
 40,1 34,6 38,7 39,5 37,1 40,7 46,2 39,1 38,4 39,9 39,3 40,5 41,9 41,7 37,9 38,3 43,9  
 43,9 30,9 42,7 35,9 35,8 41,9 42,3 40,6

**Варіант 17**

*Довжина максимального стеблового листка (в мм) у ромені звичайного:*  
 56 61 66 53 84 82 80 44 68 60 61 54 60 60 62 62 55 55 54 57 47 55 60 63 38 59 45  
 49 57 78 68 43 62 64 48 74 43 40 53 76 72 74 76 86 50 69 69 71 76 47 71 54 63 66  
 56 52 87 44 69 64 64

**Варіанти завдань для закріплення теми 1, підрозділу  
 «Порівняння вибірок за допомогою гістограми розподілу»**

**Варіант 1**

*Тривалість серцевого циклу в кардіограмі здорових дітей:* 0,91 0,71 0,73  
 0,82 0,67 0,89 0,90 1,00 0,77 0,78 0,72 0,71 0,90 0,68 0,52 0,58 0,59 0,56 0,74 0,54  
 0,72 0,76 0,74 0,79 0,66 0,84 0,85 0,81 1,11 0,77 0,84 0,59 0,71 0,74 0,65 0,96 0,83  
 0,78 0,94 0,93 0,62 0,60 0,57 0,69 0,56 0,45 0,93 0,74 0,86 0,55 0,66 0,49 0,71 0,68  
 0,96 0,78 0,85 0,54 0,64 0,82 0,60 0,71 0,76 0,56 0,79 0,66 0,76 0,65 0,66 0,72 0,92  
 0,74 0,66 0,61 0,67 0,60 0,69 0,63 0,80 0,76 0,65 0,85 0,68 0,80 0,86 0,56 0,64 0,69  
 0,49 0,56 0,74 0,87 0,69 0,62 0,61 0,97 0,83 0,66 0,78 0,89 0,75 0,90 0,81 0,59 0,68  
 0,53 0,59 0,75 0,81 0,73 0,85 0,72 0,79 0,86 0,54 0,62 0,94 0,57 0,75 0,73 0,52 0,68  
 0,85 0,75 0,79 0,72

*Тривалість серцевого циклу в кардіограмі хворих дітей:* 0,91 0,86 0,74 0,51  
 1,07 0,79 0,98 1,16 0,74 0,88 0,84 0,66 0,73 0,91 1,22 1,43 0,81 0,72 0,79 0,98 0,62  
 1,37 0,94 0,77 0,71 0,85 0,85 0,57 0,83 1,27 1,06 0,73 0,61 0,67 0,75 0,91 1,05 0,54

0,85 0,82 0,79 0,64 0,97 1,17 1,06 0,82 0,58 0,93 0,85 0,96 1,08 0,67 0,74 0,84 1,03  
0,97 0,89 0,63 0,77 0,86 0,61 0,74 0,72

### Варіант 2

*Кількість частинок вугілля в одній вакуолі у парамецій при впливі 0,04 % розчину пептону:* 6 5 6 6 6 3 4 4 5 6 6 4 5 5 5 5 4 4 3 6 5 5 4 5 4 4 4 5 5 5 5 6 4  
5 3 4 5 5 6 5 5 4 4 6 5 4 4 3 6 5 5 4 3 5 3 4 4 5 5 4 6 5 5 3 4 4 4 5 5 5 6 6 3 5 3 6 4 5 3  
5 6 3 4 3 6 5 3 4 4 6 6

*Кількість частинок вугілля в одній вакуолі у парамецій в контрольній групі:* 4 3 3 0 2 2 2 3 1 0 0 1 2 2 2 2 3 4 2 2 3 0 2 1 3 2 2 3 2 1 2 2 1 2 1 2 4 0 0 2 3 3  
2 1 0 1 2 2 3 1 0 0 4 2 2 2 4 2 1 2 2 3 2 3 2 2 2 3 1 2 0 2 0 0 2 1 2 3 3 4 0 1 2 1 2 3 3

### Варіант 3

*Розміри крипт в ободовій кишці у щурів (в розподілах окуляр-мікрометра) без операції:* 7 10 8 12 16 9 10 7 9 9 14 10 10 8 9 11 9 8 10 11 9 10 11 10 9 8 13 9 7  
9 10 8 9 9 10 9 8 10 10 11 8 9 7 8 9 10 10 10 9 10 10 11 12 9 10 10 8 12 10 9 8 7 10  
11 12 13 8 8 8 13 12 10 9 8 7 11 12 14 10 10 8 8 7 9 10 11 9 8 8 7 9 10 11 14 10 7 13  
10 9 12 12 10 8 9 8 7 10

*Розміри крипт в ободовій кишці у щурів (в розподілах окуляр-мікрометра) після видалення сліпої кишки:* 10 11 7 14 10 8 17 8 10 11 7 12 9 9 10 12 13 8 11 8  
10 15 8 10 20 13 8 11 12 9 12 9 12 12 11 10 9 11

### Варіант 4

*Кількість язичкових квіток у суцвіттях ромену звичайного (Карпати) на сухому лузі:* 21 23 23 20 18 21 25 19 29 24 23 21 27 21 20 20 8 16 12 24 21 20 15  
21 22 18 19 17 20 20 22 21 21 20 20 21 13 15 14 21 20 22 18 21 18 13 15 22 16 21  
14 18 21 25 13 12 13 17 14 15 17 20 20 16 17 21 28 21 21 24 19 21 20 14 13 14 16  
20 15 14 20 13 20 19 18 13 17 20 21 21 22 21 19 15 13 21 21 10 14 13 13 16 17 13  
21 21 19 14 18 12 27 21 13 8 8 16 21 18 13 14 15 13 20 19 21 18 16 22 20 20 16 20  
13 16 20 18 12 12 11 17 13 14 20 16 15 14 17 15 21 15 12 8 21 21 13 16 16 12 19 13  
21 14 18 14 13 15 17 19 18 15 17 13 11 15 16 14 13 13 20 13 13 12 13 13 10 21  
18 17 17 21 20 13 10 13 14 13 14 16 12

*Кількість язичкових квіток у суцвіттях ромену звичайного (Карпати) на сиром лузі:* 25 20 25 26 21 24 16 28 23 28 28 32 23 23 26 22 29 34 27 28 25 27 27  
25 25 23 29 22 24 23 23 27 21 32 24 36 27 29 22 23 24 22 24 24 30 34 26 22 22 23  
27 26 24 27 28 22 21 23 24 32 31 21 23 24 28 25 26 28 21 22 24 22 24 23 28 23 23  
24 23 23 23 23 21 35 34 33 30 28 34 37 32 31 23 28 33 22 32 34 24 32 24 27 23 23  
33 27 34 27 32 33 18 28 24 34 30 32 24 23 28 32 24 21 23 30 26 21 33 30 23 27 25  
29 32 30 30 34 26 22 32 23 34 32 23 33 22 26 26 21 26 24 33 27 27 24 24 33 35 24

33 21 21 32 33 23 30 25 30 24 27 28 26 33 35 26 31 31 32 24 21 34 23 26 36 40 31  
24 22 23 26 26 31 22 26 31 26 32 30 32 33 27

*Кількість язичкових квіток у суцвіттях ромену звичайного (Карпати) на узліссі:* 21 22 24 26 23 22 24 21 28 30 21 20 21 22 23 18 29 24 20 24 20 22 16 23 24  
23 22 21 21 30 20 25 18 22 21 19 22 23 23 19 21 22 27 21 21 21 23 26 16 21 21 25  
24 17 23 34 24 34 17 24 23 33 23 33 25 23 22 24 20 21 24 23 16 31 16 21 19 21 26  
19 19 21 22 21 20 20 22 23 21 21 22 22 25 29 20 24 21 25 25 21 21 21 25 24 26 21  
27 19 20 21 19 21 21 24 20 20 24 24 21 29 23 21 30 25 20 19 21 23 21 21 22 21 24  
22 20 22 21 24 21 25 22 21 23 21 23 30 21 20 18 28 27 21 17 22 28 15 34 24 22 22  
25 22 29 21 29 24 22 21 24 20 19 30 18 18 17 21 21 21 23 21 21 13 21 17 21 22 29  
26 23 23 28 15 24 27 26 22 26 20 19 23

### Варіант 5

*Діаметр еритроцитів у контрольній групі:* 6,36 4,55 6,82 5,45 5,00 6,82  
6,82 5,91 5,45 4,09 6,36 6,82 5,45 5,91 5,91 3,64 4,09 5,91 5,00 6,36 5,00 5,00 3,18  
6,36 4,55 6,00 5,00 6,82 5,45 4,55 3,64 5,00 6,82 5,45 3,18 7,27 5,91 6,36 5,91 6,36  
5,00 5,91 6,36 6,82 4,55 5,00 6,82 5,00 3,64 5,45 6,82 5,45 5,00 6,82 5,45 6,36 6,36  
5,91 5,91 5,91 6,82 5,00 5,00 5,00 8,18 4,55 4,55 4,55 5,00 5,00 5,91 8,64 6,36 5,91  
6,36 5,45 7,27 6,36 6,82 6,36 5,91 5,91 5,45 6,82 5,00 5,91 7,73 5,00 4,55 5,45 9,09  
7,73 5,45 6,36 8,64 4,55 6,36 5,91 6,82 8,18 5,81

*Діаметр еритроцитів на третій день вібрації з частотою 16 Гц:* 5,45 3,64  
2,27 4,55 5,45 6,82 5,91 5,45 4,55 3,18 3,64 4,55 3,64 5,00 5,45 3,64 4,55 5,45 6,36  
4,55 7,27 5,91 4,55 5,45 5,45 4,55 4,55 4,09 3,64 4,55 5,00 4,55 4,55 3,64 5,00 2,73  
3,18 4,55 5,00 3,18 2,27 5,45 4,55 3,64 5,00 5,45 4,55 5,45 4,55 6,36 5,91 6,36 6,36  
5,45 4,55 5,00 5,45 4,55 4,55 5,00 8,18 5,45 4,55 4,09 3,18 3,64 4,55 6,36 5,45 6,36  
5,91 4,55 3,18 3,64 5,00 5,91 6,36 4,55 4,55 5,45 5,00 5,91 6,36 5,45 4,55 6,36 5,91  
6,36 6,82 7,27 6,36 5,91 5,91 4,55 3,64 3,18 3,64 4,55 5,00 5,45

*Діаметр еритроцитів на третій день вібрації з частотою 32 Гц:* 4,09 3,64  
3,64 3,64 4,55 4,09 3,18 4,09 4,09 4,09 4,55 5,00 4,09 4,55 4,55 5,45 5,00 4,55 5,00  
5,45 4,09 3,18 3,18 3,64 4,55 5,00 3,64 4,55 4,09 4,09 3,18 5,45 3,64 4,09 4,55 4,55  
6,36 6,36 5,91 4,55 5,00 5,45 4,09 4,55 4,55 4,09 5,45 5,00 5,00 4,55 4,55 3,64 3,64  
4,55 5,45 3,64 5,00 4,55 4,09 3,64 4,09 4,09 4,55 5,45 3,18 5,91 4,55 4,09 4,09 5,00  
5,45 4,09 3,64 5,45 5,00 5,00 5,45 5,91 5,00 4,55 4,55 5,45 5,00 4,55 5,00 5,00 5,45  
5,91 6,36 4,55 4,09 3,64 3,18 4,09 4,55 5,00 3,64 3,18 4,55 4,55

### Варіант 6

*Діаметр еритроцитів у контрольній групі:* 6,36 4,55 6,82 5,45 5,00 6,82  
6,82 5,91 5,45 4,09 6,36 6,82 5,45 5,91 5,91 3,64 4,09 5,91 5,00 6,36 5,00 5,00 3,18  
6,36 4,55 6,00 5,00 6,82 5,45 4,55 3,64 5,00 6,82 5,45 3,18 7,27 5,91 6,36 5,91 6,36

5,00 5,91 6,36 6,82 4,55 5,00 6,82 5,00 3,64 5,45 6,82 5,45 5,00 6,82 5,45 6,36 6,36  
 5,91 5,91 5,91 6,82 5,00 5,00 5,00 8,18 4,55 4,55 4,55 5,00 5,00 5,91 8,64 6,36 5,91  
 6,36 5,45 7,27 6,36 6,82 6,36 5,91 5,91 5,45 6,82 5,00 5,91 7,73 5,00 4,55 5,45 9,09  
 7,73 5,45 6,36 8,64 4,55 6,36 5,91 6,82 8,18 5,81

*Діаметр еритроцитів на перший день вібрації з частотою 40 Гц:* 5,45 3,18  
 4,55 4,55 5,00 5,00 5,45 3,64 4,09 5,00 4,09 6,36 5,91 5,00 4,55 5,45 3,64 4,09 4,55  
 5,45 5,00 6,36 6,82 5,00 5,45 3,18 4,55 5,45 4,09 3,64 5,00 4,55 3,18 4,55 5,00 5,45  
 3,64 5,91 5,45 6,36 4,55 6,36 5,45 6,36 4,55 5,00 4,55 7,27 5,00 5,45 6,36 5,45 6,36  
 5,45 5,00 4,55 6,36 5,91 3,64 4,55 5,00 5,45 6,36 6,82 5,91 5,45 6,82 4,55 4,55 5,00  
 4,55 4,55 5,91 5,45 5,00 5,00 4,55 5,00 5,45 6,36 5,45 5,91 5,00 4,55 5,45 6,36 4,09  
 3,64 4,55 5,00 5,45 4,55 6,82 5,00 4,55 4,09 3,64 3,18 4,55 5,00

*Діаметр еритроцитів на п'ятий день вібрації з частотою 40 Гц:* 6,82 6,82  
 6,82 7,27 6,36 6,82 4,55 7,27 5,91 5,45 6,82 5,45 4,55 4,55 5,91 6,36 5,00 5,00 6,36  
 5,00 6,82 5,91 5,45 5,45 5,45 5,91 4,55 5,91 5,45 7,27 5,00 6,36 6,36 6,82 5,00 5,00  
 4,55 7,27 6,82 6,82 5,00 6,36 6,82 6,36 5,00 6,82 6,36 8,18 4,55 8,18 8,18 5,91 5,45  
 5,91 5,91 6,36 5,91 7,73 5,91 5,45 6,36 5,45 5,91 5,00 4,55 9,09 5,45 4,55 6,36 6,36  
 7,73 7,27 5,45 4,55 5,00 6,82 6,82 6,82 6,82 6,36 5,91 5,00 5,91 5,91 5,45 6,82 5,45  
 5,45 5,91 5,91 7,73 8,18 5,45 5,45 6,36 5,00 6,36 6,36 5,45 5,45

### Варіант 7

*Діаметр еритроцитів мишей, що не піддаються вібрації:* 5,00 5,00 6,36  
 10,00 6,36 10,00 5,00 5,45 5,45 5,91 6,82 5,45 4,55 6,36 8,18 6,36 6,82 5,91 5,45  
 5,00 4,55 7,27 5,91 3,18 8,64 6,36 6,36 5,45 7,73 5,91 8,18 5,91 6,82 6,82 5,45 6,82  
 5,91 7,73 5,00 4,09 5,00 5,91 7,73 6,36 6,36 5,00 7,27 6,82 5,91 5,45 6,82 5,00 5,00  
 7,27 6,36 5,45 5,45 5,00 5,91 8,18 6,82 4,55 5,45 5,91 7,73 6,36 5,00 6,82 2,73 5,91  
 4,55 7,27 5,00 5,45 5,45 6,36 5,45 5,00 5,91 5,45 5,00 5,91 6,36 5,45 6,36 4,55 5,00  
 6,36 5,91 5,45 4,55 5,91 5,45 5,91 8,18 6,36 4,55 7,27 6,82

*Діаметр еритроцитів мишей на перший день вібрації з частотою 8 Гц:*  
 5,45 5,00 4,55 5,91 5,45 5,00 4,55 3,18 3,64 4,55 4,55 4,55 6,36 4,55 5,91 6,36 5,45  
 5,91 6,36 7,27 6,82 4,55 4,55 5,45 5,91 6,82 7,27 4,55 5,45 5,91 3,64 5,00 6,36 5,91  
 6,36 4,55 5,00 3,18 5,91 4,55 5,45 6,36 5,91 6,82 5,00 4,55 3,18 5,45 6,36 7,27 5,91  
 5,00 5,00 4,55 5,00 5,00 5,00 5,45 4,09 3,18 5,45 5,00 5,45 4,55 5,45 6,36 5,91 5,45  
 5,00 4,55 6,36 5,91 4,55 5,00 5,45 3,64 5,00 4,55 5,00 6,82 3,18 5,00 5,00 5,45 5,91  
 4,55 5,00 5,45 3,18 5,00 4,55 5,45 4,55 5,00 4,09 4,09 3,18 4,55 4,55 5,00

*Діаметр еритроцитів мишей на третій день вібрації з частотою 8 Гц:*  
 6,82 6,82 6,82 6,82 7,73 7,27 6,82 6,82 6,82 7,73 6,36 7,27 7,27 7,73 8,18 6,82 6,82  
 6,82 6,82 6,82 6,82 6,82 6,82 6,82 6,82 6,36 5,91 5,45 5,00 6,82 5,91 5,00 6,82 7,27  
 6,82 7,27 6,82 6,82 5,91 5,45 5,00 6,82 5,45 7,73 7,27 5,91 6,36 4,55 5,45 5,91 6,82  
 5,91 5,45 6,36 5,91 4,55 5,91 5,91 6,82 7,27 6,82 5,91 5,45 7,27 7,27 5,45 6,82 5,91

5,00 6,36 7,27 6,82 5,45 5,45 5,45 4,55 5,00 4,55 3,64 6,82 3,64 4,55 5,00 6,82 6,36 5,91 6,36 6,82 5,45 5,91 5,91 5,45 4,55 6,82 5,00 7,27 7,73 4,55 4,55 6,82

### Варіант 8

*Діаметр еритроцитів на третій день вібрації з частотою 8 Гц:* 6,82 6,82 6,82 6,82 7,73 7,27 6,82 6,82 6,82 7,73 6,36 7,27 7,27 7,73 8,18 6,82 6,82 6,82 6,82 6,82 6,82 6,82 6,82 6,36 5,91 5,45 5,00 6,82 5,91 5,00 6,82 7,27 6,82 7,27 6,82 6,82 5,91 5,45 5,00 6,82 5,45 7,73 7,27 5,91 6,36 4,55 5,45 5,91 6,82 5,91 5,45 6,36 5,91 4,55 5,91 5,91 6,82 7,27 6,82 5,91 5,45 7,27 7,27 5,45 6,82 5,91 5,00 6,36 7,27 6,82 5,45 5,45 5,45 4,55 5,00 4,55 3,64 6,82 3,64 4,55 5,00 6,82 6,36 5,91 6,36 6,82 5,45 5,91 5,91 5,45 4,55 6,82 5,00 7,27 7,73 4,55 4,55 6,82

*Діаметр еритроцитів на третій день вібрації з частотою 16 Гц:* 5,45 3,64 2,27 4,55 5,45 6,82 5,91 5,45 4,55 3,18 3,64 4,55 3,64 5,00 5,45 3,64 4,55 5,45 6,36 4,55 7,27 5,91 4,55 5,45 5,45 4,55 4,55 4,09 3,64 4,55 5,00 4,55 4,55 3,64 5,00 2,73 3,18 4,55 5,00 3,18 2,27 5,45 4,55 3,64 5,00 5,45 4,55 5,45 4,55 6,36 5,91 6,36 6,36 5,45 4,55 5,00 5,45 4,55 4,55 5,00 8,18 5,45 4,55 4,09 3,18 3,64 4,55 6,36 5,45 6,36 5,91 4,55 3,18 3,64 5,00 5,91 6,36 4,55 4,55 5,45 5,00 5,91 6,36 5,45 4,55 6,36 5,91 6,36 6,82 7,27 6,36 5,91 5,91 4,55 3,64 3,18 3,64 4,55 5,00 5,45

*Діаметр еритроцитів на третій день вібрації з частотою 24 Гц:* 6,36 6,82 4,55 9,09 5,00 4,55 5,91 4,55 5,00 5,91 4,55 5,45 4,55 5,00 5,45 4,55 5,91 5,45 4,55 5,45 4,55 5,00 5,45 5,91 5,45 4,55 3,64 4,09 4,55 6,36 4,55 4,09 4,55 6,36 4,55 7,27 4,55 5,45 4,55 4,55 5,45 7,27 4,55 7,73 4,55 5,45 5,45 7,73 4,55 6,82 5,45 3,18 5,45 4,55 5,00 5,91 8,18 7,73 6,36 4,55 5,00 5,45 6,82 5,45 5,91 6,82 4,55 5,00 6,36 4,55 5,91 6,36 5,45 4,55 5,00 5,45 6,36 6,82 5,00 5,45 6,82 4,09 3,18 6,82 4,55 4,55 5,00 5,45 5,91 5,00 4,55 5,45 5,91 4,55 6,36 4,55 5,45

### Варіант 9

*Діаметр еритроцитів мишей, що не піддаються вібрації:* 5,00 5,00 6,36 10,00 6,36 10,00 5,00 5,45 5,45 5,91 6,82 5,45 4,55 6,36 8,18 6,36 6,82 5,91 5,45 5,00 4,55 7,27 5,91 3,18 8,64 6,36 6,36 5,45 7,73 5,91 8,18 5,91 6,82 6,82 5,45 6,82 5,91 7,73 5,00 4,09 5,00 5,91 7,73 6,36 6,36 5,00 7,27 6,82 5,91 5,45 6,82 5,00 5,00 7,27 6,36 5,45 5,45 5,00 5,91 8,18 6,82 4,55 5,45 5,91 7,73 6,36 5,00 6,82 2,73 5,91 4,55 7,27 5,00 5,45 5,45 6,36 5,45 5,00 5,91 5,45 5,00 5,91 6,36 5,45 6,36 4,55 5,00 6,36 5,91 5,45 4,55 5,91 5,45 5,91 8,18 6,36 4,55 7,27 6,82

*Діаметр еритроцитів мишей на третій день вібрації з частотою 50 Гц:* 6,36 7,27 4,55 5,45 4,55 4,55 5,00 5,45 6,36 5,91 4,55 5,45 5,91 6,36 6,36 5,91 5,45 6,36 4,55 5,45 5,00 4,55 5,45 5,91 6,36 5,45 6,82 4,55 5,45 5,00 5,45 5,91 4,55 5,00 5,45 4,55 5,00 5,91 6,36 4,55 5,45 5,91 5,45 6,36 6,82 7,27 7,73 4,55 4,09 3,64 4,55 4,55 5,45 5,00 4,55 5,91 6,36 4,55 5,00 5,45 5,91 4,55 4,09 4,55 4,09 3,64 3,18 6,82

5,45 5,91 5,91 4,55 3,64 5,00 5,00 4,55 4,55 5,45 5,45 5,00 4,55 5,45 4,55 6,82 4,55  
5,00 5,45 5,91 4,55 3,18 3,64 4,55 5,45 5,91 5,00 4,55 5,45 5,00 4,55 5,00

*Діаметр еритроцитів мишей на п'ятий день вібрації з частотою 50 Гц:*

5,91 6,82 5,45 7,73 5,00 5,45 6,36 5,00 6,36 6,82 7,73 5,00 5,91 6,36 6,36 6,82  
5,45 3,64 6,82 6,82 3,64 5,45 8,18 5,45 6,36 6,36 7,27 5,45 5,00 6,82 6,82 5,45 5,45  
7,27 7,27 9,09 8,64 6,36 7,73 6,82 5,45 5,91 5,45 5,00 7,73 3,64 5,45 4,55 6,36 6,82  
5,45 2,73 7,73 3,18 3,64 6,82 5,45 5,00 5,91 6,82 6,36 5,45 5,91 5,00 5,91 6,82 5,45  
4,55 5,00 8,18 4,55 2,73 5,45 6,36 7,27 5,91 5,00 3,18 4,55 6,36 3,64 6,82 8,18 5,45  
7,27 6,36 5,91 6,82 5,00 6,36 6,82 7,73 5,00 6,82 5,45 7,73 4,55 3,64 6,36 5,45

### Варіант 10

*Діаметр еритроцитів на третій день вібрації з частотою 24 Гц:* 3,18 4,55  
4,55 3,64 4,09 4,55 3,18 2,73 4,55 5,00 4,55 4,55 4,55 4,55 4,09 4,55 5,00 4,55 4,09  
4,55 5,00 5,00 4,55 4,09 4,55 3,64 3,64 4,09 4,55 4,55 3,64 5,00 5,00 4,55 4,55  
4,09 4,55 2,73 3,18 4,55 4,55 5,00 4,55 4,55 5,45 3,64 3,64 5,00 4,55 3,64 3,64 4,55  
4,09 4,55 4,09 5,00 5,45 4,55 4,09 3,64 5,00 4,09 3,64 3,64 5,00 4,55 5,91 4,55 5,00  
5,00 4,09 6,36 4,55 5,00 5,00 3,64 4,55 5,00 3,18 4,55 5,00 4,09 3,18 4,55 5,00 6,36  
4,55 5,00 4,09 5,45 4,55 2,73 6,82 4,55 4,55 4,55 4,09 4,09 4,55

*Діаметр еритроцитів на третій день вібрації з частотою 32 Гц:* 4,09 3,64  
3,64 3,64 4,55 4,09 3,18 4,09 4,09 4,09 4,55 5,00 4,09 4,55 4,55 5,45 5,00 4,55 5,00  
5,45 4,09 3,18 3,18 3,64 4,55 5,00 3,64 4,55 4,09 4,09 3,18 5,45 3,64 4,09 4,55 4,55  
6,36 6,36 5,91 4,55 5,00 5,45 4,09 4,55 4,55 4,09 5,45 5,00 5,00 4,55 4,55 3,64 3,64  
4,55 5,45 3,64 5,00 4,55 4,09 3,64 4,09 4,09 4,55 5,45 3,18 5,91 4,55 4,09 4,09 5,00  
5,45 4,09 3,64 5,45 5,00 5,00 5,45 5,91 5,00 4,55 4,55 5,45 5,00 4,55 5,00 5,00 5,45  
5,91 6,36 4,55 4,09 3,64 3,18 4,09 4,55 5,00 3,64 3,18 4,55 4,55

*Діаметр еритроцитів на третій день вібрації з частотою 50 Гц:* 6,36 7,27  
4,55 5,45 4,55 4,55 5,00 5,45 6,36 5,91 4,55 5,45 5,91 6,36 6,36 5,91 5,45 6,36 4,55  
5,45 5,00 4,55 5,45 5,91 6,36 5,45 6,82 4,55 5,45 5,00 5,45 5,91 4,55 5,00 5,45 4,55  
5,00 5,91 6,36 4,55 5,45 5,91 5,45 6,36 6,82 7,27 7,73 4,55 4,09 3,64 4,55 4,55 5,45  
5,00 4,55 5,91 6,36 4,55 5,00 5,45 5,91 4,55 4,09 4,55 4,09 3,64 3,18 6,82 5,45 5,91  
5,91 4,55 3,64 5,00 5,00 4,55 4,55 5,45 5,45 5,00 4,55 5,45 4,55 6,82 4,55 5,00 5,45  
5,91 4,55 3,18 3,64 4,55 5,45 5,91 5,00 4,55 5,45 5,00 4,55 5,00

### Варіант 11

*Діаметр еритроцитів мишей, що не піддавалися вібрації:* 5,00 5,00 6,36  
10,00 6,36 10,00 5,00 5,45 5,45 5,91 6,82 5,45 4,55 6,36 8,18 6,36 6,82 5,91 5,45  
5,00 4,55 7,27 5,91 3,18 8,64 6,36 6,36 5,45 7,73 5,91 8,18 5,91 6,82 6,82 5,45 6,82  
5,91 7,73 5,00 4,09 5,00 5,91 7,73 6,36 6,36 5,00 7,27 6,82 5,91 5,45 6,82 5,00 5,00  
7,27 6,36 5,45 5,45 5,00 5,91 8,18 6,82 4,55 5,45 5,91 7,73 6,36 5,00 6,82 2,73 5,91

4,55 7,27 5,00 5,45 5,45 6,36 5,45 5,00 5,91 5,45 5,00 5,91 6,36 5,45 6,36 4,55 5,00  
6,36 5,91 5,45 4,55 5,91 5,45 5,91 8,18 6,36 4,55 7,27 6,82

*Діаметр еритроцитів мишей на третій день вібрації з частотою 8 Гц:*  
6,82 6,82 6,82 6,82 7,73 7,27 6,82 6,82 6,82 7,73 6,36 7,27 7,27 7,73 8,18 6,82 6,82  
6,82 6,82 6,82 6,82 6,82 6,82 6,82 6,82 6,36 5,91 5,45 5,00 6,82 5,91 5,00 6,82 7,27  
6,82 7,27 6,82 6,82 5,91 5,45 5,00 6,82 5,45 7,73 7,27 5,91 6,36 4,55 5,45 5,91 6,82  
5,91 5,45 6,36 5,91 4,55 5,91 5,91 6,82 7,27 6,82 5,91 5,45 7,27 7,27 5,45 6,82 5,91  
5,00 6,36 7,27 6,82 5,45 5,45 5,45 4,55 5,00 4,55 3,64 6,82 3,64 4,55 5,00 6,82 6,36  
5,91 6,36 6,82 5,45 5,91 5,91 5,45 4,55 6,82 5,00 7,27 7,73 4,55 4,55 6,82 6,23

*Діаметр еритроцитів мишей на п'ятий день вібрації з частотою 8 Гц:*  
5,91 6,82 5,91 5,91 6,82 6,36 5,45 5,45 2,73 7,27 5,91 3,64 7,73 6,36 5,00 4,55 6,36  
7,27 4,55 5,00 4,55 5,91 5,91 5,45 5,91 2,73 6,36 5,91 5,00 5,00 6,82 6,36 5,45 3,18  
6,82 6,36 4,55 5,45 4,55 5,45 7,27 4,55 5,45 6,82 5,91 6,82 5,00 5,91 5,45 6,36 6,36  
5,00 5,00 5,91 5,91 5,91 4,55 5,45 5,45 5,45 5,00 4,55 6,36 6,82 4,55 5,00 5,91  
5,00 5,45 5,00 5,91 5,00 5,91 5,45 5,00 4,55 7,27 5,91 5,00 4,09 6,36 6,82 4,55 3,18  
4,55 6,36 5,45 3,64 6,82 7,27 6,82 4,55 7,27 6,82 7,27 5,45 7,73 6,82 5,91 5,64

## Варіант 12

*Діаметр еритроцитів на третій день вібрації з частотою 32 Гц:* 4,09 5,91  
4,55 5,00 5,45 5,91 4,55 4,09 3,64 7,73 4,55 5,45 4,55 4,09 6,82 6,36 5,91 5,45 4,55  
4,09 3,64 5,91 5,00 5,45 5,45 6,36 5,45 4,55 4,55 5,00 5,45 4,55 4,55 5,00 5,45 4,55  
5,00 3,18 3,64 4,09 5,45 7,27 5,91 6,36 5,45 5,91 6,36 4,55 5,00 4,55 5,91 6,36 5,00  
5,45 4,55 6,82 4,55 5,00 5,45 5,91 6,36 4,55 5,91 5,45 4,55 5,00 5,45 5,91 6,36 4,55  
3,64 4,09 4,55 5,45 6,36 5,91 4,55 4,09 3,64 4,55 5,00 5,45 5,91 6,36 5,45 4,55 8,18  
7,27 5,91 6,36 4,55 5,45 6,36 4,55 5,00 5,45 6,36 4,55 5,00 5,45

*Діаметр еритроцитів на третій день вібрації з частотою 40 Гц:* 6,36 5,00  
5,45 3,64 3,18 4,55 7,27 5,91 4,55 5,00 5,00 4,55 4,55 5,45 5,00 4,55 5,45 4,55  
5,00 5,91 4,09 4,09 3,64 4,55 5,00 5,45 5,91 6,36 4,55 5,45 5,91 3,18 4,55 3,64 4,55  
5,45 5,91 5,45 6,36 4,55 5,00 5,45 6,36 4,55 7,27 4,55 5,45 6,82 4,55 5,45 7,27 4,55  
5,91 5,45 5,91 5,45 5,91 3,18 5,45 5,91 4,55 5,91 4,55 5,91 6,82 5,00 4,55 7,27 6,36  
5,00 5,45 5,91 5,91 5,45 5,00 4,55 6,36 4,55 6,36 5,00 4,55 5,45 5,91 6,36 4,55 5,00  
5,45 5,91 4,55 5,00 4,09 3,18 4,55 5,45 4,55 6,36 6,82 4,55 6,36

*Діаметр еритроцитів на третій день вібрації з частотою 50 Гц:* 6,36 7,27  
4,55 5,45 4,55 4,55 5,00 5,45 6,36 5,91 4,55 5,45 5,91 6,36 6,36 5,91 5,45 6,36 4,55  
5,45 5,00 4,55 5,45 5,91 6,36 5,45 6,82 4,55 5,45 5,00 5,45 5,91 4,55 5,00 5,45 4,55  
5,00 5,91 6,36 4,55 5,45 5,91 5,45 6,36 6,82 7,27 7,73 4,55 4,09 3,64 4,55 4,55 5,45  
5,00 4,55 5,91 6,36 4,55 5,00 5,45 5,91 4,55 4,09 4,55 4,09 3,64 3,18 6,82 5,45 5,91  
5,91 4,55 3,64 5,00 5,00 4,55 4,55 5,45 5,45 5,00 4,55 5,45 4,55 6,82 4,55 5,00 5,45  
5,91 4,55 3,18 3,64 4,55 5,45 5,91 5,00 4,55 5,45 5,00 4,55 5,00

**Варіант 13**

*Діаметр еритроцитів мишей, що не піддавалися вібрації:* 5,00 5,00 6,36 10,00 6,36 10,00 5,00 5,45 5,45 5,91 6,82 5,45 4,55 6,36 8,18 6,36 6,82 5,91 5,45 5,00 4,55 7,27 5,91 3,18 8,64 6,36 6,36 5,45 7,73 5,91 8,18 5,91 6,82 6,82 5,45 6,82 5,91 7,73 5,00 4,09 5,00 5,91 7,73 6,36 6,36 5,00 7,27 6,82 5,91 5,45 6,82 5,00 5,00 7,27 6,36 5,45 5,45 5,00 5,91 8,18 6,82 4,55 5,45 5,91 7,73 6,36 5,00 6,82 2,73 5,91 4,55 7,27 5,00 5,45 5,45 6,36 5,45 5,00 5,91 5,45 5,00 5,91 6,36 5,45 6,36 4,55 5,00 6,36 5,91 5,45 4,55 5,91 5,45 5,91 8,18 6,36 4,55 7,27 6,82

*Діаметр еритроцитів мишей на третій день вібрації з частотою 16 Гц:* 5,45 3,64 2,27 4,55 5,45 6,82 5,91 5,45 4,55 3,18 3,64 4,55 3,64 5,00 5,45 3,64 4,55 5,45 6,36 4,55 7,27 5,91 4,55 5,45 5,45 4,55 4,55 4,09 3,64 4,55 5,00 4,55 4,55 3,64 5,00 2,73 3,18 4,55 5,00 3,18 2,27 5,45 4,55 3,64 5,00 5,45 4,55 5,45 4,55 6,36 5,91 6,36 6,36 5,45 4,55 5,00 5,45 4,55 4,55 5,00 8,18 5,45 4,55 4,09 3,18 3,64 4,55 6,36 5,45 6,36 5,91 4,55 3,18 3,64 5,00 5,91 6,36 4,55 4,55 5,45 5,00 5,91 6,36 5,45 4,55 6,36 5,91 6,82 7,27 6,36 5,91 5,91 4,55 3,64 3,18 3,64 4,55 5,00 5,45

*Діаметр еритроцитів мишей на п'ятий день вібрації з частотою 16 Гц:* 7,73 5,91 7,27 4,55 5,45 7,27 5,91 6,82 4,55 7,73 6,82 5,00 3,64 6,36 5,91 5,00 5,00 4,55 5,45 5,00 3,64 3,64 3,18 5,45 6,36 4,55 5,45 6,36 5,91 4,55 5,91 6,36 5,45 5,00 5,91 5,00 4,55 5,45 5,45 6,36 7,73 6,82 5,91 5,45 5,45 4,55 5,00 6,36 6,82 7,27 7,27 5,45 6,36 8,18 4,55 5,91 6,36 4,55 5,00 5,91 8,18 4,55 5,45 5,91 7,73 6,82 5,91 5,45 4,55 5,00 6,82 6,36 5,00 5,45 5,91 8,18 5,45 5,45 4,55 6,36 5,91 4,55 6,36 5,00 5,91 5,00 5,45 6,36 6,82 7,27 5,45 6,36 4,55 4,55 5,00 6,82 5,00 5,45 5,91 5,45

**Варіант 14**

*Діаметр еритроцитів мишей, що не піддавалися вібрації:* 5,00 5,00 6,36 10,00 6,36 10,00 5,00 5,45 5,45 5,91 6,82 5,45 4,55 6,36 8,18 6,36 6,82 5,91 5,45 5,00 4,55 7,27 5,91 3,18 8,64 6,36 6,36 5,45 7,73 5,91 8,18 5,91 6,82 6,82 5,45 6,82 5,91 7,73 5,00 4,09 5,00 5,91 7,73 6,36 6,36 5,00 7,27 6,82 5,91 5,45 6,82 5,00 5,00 7,27 6,36 5,45 5,45 5,00 5,91 8,18 6,82 4,55 5,45 5,91 7,73 6,36 5,00 6,82 2,73 5,91 4,55 7,27 5,00 5,45 5,45 6,36 5,45 5,00 5,91 5,45 5,00 5,91 6,36 5,45 6,36 4,55 5,00 6,36 5,91 5,45 4,55 5,91 5,45 5,91 8,18 6,36 4,55 7,27 6,82

*Діаметр еритроцитів мишей на третій день вібрації з частотою 32 Гц:* 4,09 3,64 3,64 3,64 4,55 4,09 3,18 4,09 4,09 4,55 5,00 4,09 4,55 4,55 5,45 5,00 4,55 5,00 5,45 4,09 3,18 3,18 3,64 4,55 5,00 3,64 4,55 4,09 4,09 3,18 5,45 3,64 4,09 4,55 4,55 6,36 6,36 5,91 4,55 5,00 5,45 4,09 4,55 4,55 4,09 5,45 5,00 5,00 4,55 4,55 3,64 3,64 4,55 5,45 3,64 5,00 4,55 4,09 3,64 4,09 4,09 4,55 5,45 3,18 5,91 4,55 4,09 4,09 5,00 5,45 4,09 3,64 5,45 5,00 5,00 5,45 5,91 5,00 4,55 4,55 5,45 5,00 4,55 5,00 5,00 5,45 5,91 6,36 4,55 4,09 3,64 3,18 4,09 4,55 5,00 3,64 3,18 4,55 4,55

*Діаметр еритроцитів мишей на п'ятий день вібрації з частотою 32 Гц:*

5,45 4,55 5,91 5,91 6,36 5,91 5,45 5,00 9,09 5,91 6,82 5,45 5,00 5,00 5,45 5,91 6,82  
5,45 5,91 6,36 6,36 5,00 5,00 5,45 6,36 5,91 4,55 5,00 4,09 6,36 3,64 4,55 3,18 4,55  
5,45 5,00 3,64 4,09 6,36 5,45 4,55 5,00 5,00 4,55 3,64 3,18 4,55 6,36 4,09 5,45 5,45  
5,00 6,36 6,82 6,36 4,55 4,09 5,91 4,09 5,45 4,55 5,45 4,09 4,09 3,64 5,00 4,55 4,55  
5,00 4,55 3,64 5,45 6,82 3,64 4,09 5,00 6,82 3,64 4,09 5,00 3,64 2,73 5,00 3,18 5,45  
5,45 4,09 3,64 5,00 5,00 5,45 4,55 4,55 6,36 5,00 5,00 6,36 5,45 4,55 5,45

### Варіант 15

*Діаметр еритроцитів мишей, що не піддавалися вібрації: 5,00 5,00 6,36*

10,00 6,36 10,00 5,00 5,45 5,45 5,91 6,82 5,45 4,55 6,36 8,18 6,36 6,82 5,91 5,45  
5,00 4,55 7,27 5,91 3,18 8,64 6,36 6,36 5,45 7,73 5,91 8,18 5,91 6,82 6,82 5,45 6,82  
5,91 7,73 5,00 4,09 5,00 5,91 7,73 6,36 6,36 5,00 7,27 6,82 5,91 5,45 6,82 5,00 5,00  
7,27 6,36 5,45 5,45 5,00 5,91 8,18 6,82 4,55 5,45 5,91 7,73 6,36 5,00 6,82 2,73 5,91  
4,55 7,27 5,00 5,45 5,45 6,36 5,45 5,00 5,91 5,45 5,00 5,91 6,36 5,45 6,36 4,55 5,00  
6,36 5,91 5,45 4,55 5,91 5,45 5,91 8,18 6,36 4,55 7,27 6,82

*Діаметр еритроцитів мишей на перший день вібрації з частотою 2 Гц:*

4,55 5,00 4,55 4,09 5,00 3,64 4,55 4,55 3,18 5,00 5,45 4,55 3,64 4,09 4,55 5,00 5,00  
4,55 4,55 5,00 3,64 3,18 4,55 6,36 4,55 5,00 5,91 3,18 5,45 5,00 4,55 4,55 3,64 6,36  
4,55 3,64 4,55 5,00 4,55 4,55 5,00 5,45 5,45 5,91 4,55 3,64 3,64 5,00 5,00 5,45 4,55  
4,55 5,00 5,00 4,55 4,55 5,45 6,36 5,00 5,91 4,55 5,45 5,00 3,18 5,91 5,00 6,36 4,55  
3,64 5,45 5,45 4,55 5,00 4,09 4,09 7,27 4,55 3,64 5,00 5,45 4,09 3,18 4,55 5,00 5,45  
2,27 5,00 3,64 3,18 7,27 4,55 5,00 5,00 5,45 6,36 4,55 4,09 6,36 4,55 5,00

*Діаметр еритроцитів мишей на третій день вібрації з частотою 2 Гц:*

6,82 5,91 4,55 5,45 5,45 4,55 6,36 5,00 4,55 5,45 5,00 4,55 5,45 4,55 4,09 5,00 5,45  
5,91 5,45 4,55 5,91 4,09 5,91 4,55 3,64 4,55 4,55 5,45 5,00 5,91 6,36 5,91 5,45 5,45  
5,91 4,55 5,00 5,00 4,55 5,91 6,36 4,55 3,64 6,36 5,91 5,91 5,00 6,36 3,64 3,18 5,91  
6,82 5,45 6,36 4,55 6,36 4,55 5,00 5,45 5,91 6,82 4,55 5,00 4,55 5,91 5,91 4,55 5,00  
5,91 5,45 4,55 4,55 5,00 5,91 5,45 5,45 6,82 4,55 3,18 3,18 3,64 4,55 5,45 6,82 5,45  
5,91 4,55 6,36 4,55 5,45 5,91 6,36 4,09 3,64 4,55 5,45 5,45 6,36 5,45 4,55

*Діаметр еритроцитів мишей на п'ятий день вібрації з частотою 2 Гц:*

6,82 8,18 6,36 4,55 7,27 6,82 5,91 5,45 6,36 5,91 5,45 6,36 5,45 5,91 5,00 6,82 4,09  
4,09 4,09 5,00 6,36 4,55 8,18 7,27 5,91 5,00 4,55 5,45 3,64 5,91 5,00 5,91 4,09 3,64  
6,82 4,55 5,00 4,09 6,36 5,00 5,45 5,91 5,00 6,36 4,55 6,36 5,45 2,73 4,55 6,82 6,36  
6,82 5,45 4,55 5,91 5,45 2,73 5,00 8,18 4,55 4,55 5,45 5,00 5,00 5,00 3,64 4,55 3,18  
6,82 3,18 3,64 6,36 6,36 4,55 5,45 5,91 5,45 5,00 6,82 5,45 5,91 5,45 7,27 5,00 5,00  
6,36 6,36 6,82 6,82 6,82 5,91 7,27 7,27 7,27 5,45 7,73 7,27 5,45 6,82 6,82

## Варіанти завдань для закріплення теми 4 «Порівняння середніх показників вибірок. Параметричні критерії перевірки гіпотез про середні і дисперсії»

### Варіант 1

Дослідження електропровідності в стандартній точці акупунктури, розташованої в області кистей Тай-юань (р 9), проводили у чоловіків в положенні «сидячи», з використанням спеціального обладнання. Для оцінки стійкості зміни досліджуваних показників вимірювання проводили повторно на першій, п'ятій, десятій хвилинах. Встановити, чи залежить величина електропровідності в даній точці від часу вимірювання.

Стать	Рука	Точка акупунктури	Електропровідність		
			перша хв	п'ята хв	десята хв
m	R	р 9	5	5	5
m	R	р 9	20	20	17
m	R	р 9	5	4	8
m	R	р 9	17	10	7
m	R	р 9	2	3	5
m	R	р 9	6	3	4
m	R	р 9	10	6	11
m	R	р 9	11	6	7
m	R	р 9	2	3	4
m	R	р 9	20	20	17
m	L	р 9	6	6	6
m	L	р 9	20	20	8
m	L	р 9	13	10	11
m	L	р 9	13	14	7
m	L	р 9	3	5	4
m	L	р 9	7	3	5
m	L	р 9	9	9	10
m	L	р 9	9	11	10
m	L	р 9	3	3	4
m	L	р 9	20	20	11

### Варіант 2

Вимірювання температури в стандартній точці акупунктури, розташованої в області кистей Тай юань (р 9), проводили у чоловіків в положенні «сидячи», з використанням спеціального обладнання. Для оцінки стійкості зміни досліджуваного показника вимірювання повторно проводили на першій, п'ятій, десятій хвилинах. Встановити, чи залежить величина температури в даній точці від часу вимірювання.

Стать	Рука	Точка акупунктури	Температура		
			перша хв	п'ята хв	десята хв
М	Р	р 9	27	27	27
М	Р	р 9	27	27	28
М	Р	р 9	29	30	30
М	Р	р 9	27	29	30
М	Р	р 9	22	22	21
М	Р	р 9	23	22	21
М	Р	р 9	27	26	27
М	Р	р 9	26	27	27
М	Р	р 9	21	21	21
М	Р	р 9	30	30	31
М	Л	р 9	27	28	27
М	Л	р 9	28	28	28
М	Л	р 9	29	30	30
М	Л	р 9	28	28	28
М	Л	р 9	21	21	21
М	Л	р 9	22	22	22
М	Л	р 9	26	26	26
М	Л	р 9	28	28	27
М	Л	р 9	21	21	21
М	Л	р 9	31	32	32

### Варіант 3

Показники кров'яного тиску у здорових людей після прийому кофеїну (X1) і помилкової пігулки (X2) наведені в таблиці. Встановити, чи впливає кофеїн на кров'яний тиск.

Кров'яний тиск			
X1	X2	X1	X2
126	121	153	141
145	143	135	138
137	115	163	147
116	106	133	129
137	135	143	160
157	157	129	123
126	117	144	132
139	130	130	121
143	135	138	134
145	123	130	129

### Варіант 4

Діаметр суцвіття (в мм) у ромену на початку і кінці цвітіння наведено в таблиці. Встановити, чи залежить діаметр суцвіття ромену від терміну цвітіння.

Діаметр суцвіття, мм			
Початок цвітіння	Кінець цвітіння	Початок цвітіння	Кінець цвітіння
13	14	13	14
9	12	9	12
12	11	12	11
9	14	9	14
9	11	9	11
8	10	8	10
10	11	10	11
11	13	11	13
7	12	7	12
9	12	9	12

### Варіант 5

Дослідження електропровідності в стандартній точці акупунктури, розташованій в області кистей Шеєн-мен (с 7), проводили у жінок в положенні «сидячи», з використанням спеціального обладнання. Для оцінки стійкості зміни досліджуваних показників вимірювання проводили повторно на першій, п'ятій, десятій хвилині. Встановити, чи залежить величина електропровідності в даній точці від часу вимірювання.

Стать	Рука	Точка акупунктури	Електропровідність		
			перша хв	п'ята хв	десята хв
W	R	с 7	12	7	11
W	R	с 7	10,5	8	7
W	R	с 7	7	6	10
W	R	с 7	13	10	7
W	R	с 7	10	8	10
W	R	с 7	9	12	12
W	R	с 7	10	9	9
W	R	с 7	10	13	11
W	R	с 7	9	8	7
W	R	с 7	13	13	12
W	L	с 7	8,5	13	7
W	L	с 7	8	6,5	8
W	L	с 7	9	8	10
W	L	с 7	10	13	14
W	L	с 7	11	5	5
W	L	с 7	14	9	12
W	L	с 7	6	5	5
W	L	с 7	9	8	9
W	L	с 7	6	5	5
W	L	с 7	14	13	14

**Варіант 6**

Число стегових пір у живородної ящірки наведено в таблиці.

Число стегових пір у живородної ящірки		
з Бельгії	зі Швейцарії	з Сахаліну
9	12	11
13	9	10
9	11	8
10	12	7
11	11	9
12	10	11
11	11	10
11	12	8
11	11	7
11	12	9
11	12	11
10	11	10
9	10	8
13	11	7
9	12	9
10	10	11
11	11	10
12	10	8
11	11	7
11	10	9
	9	

Встановити, чи є відмінності для груп ящірок за цими показниками.

**Варіант 7**

Результати визначення швидкості кровотоку (с), який вимірюється у дітей різними методами, наведено в таблиці. Визначте, чи залежать отримані величини швидкості кровотоку від методу його вимірювання.

Швидкість кровотоку, с					
метод 1	метод 2	метод 3	метод 1	метод 2	метод 3
1	2	3	4	5	6
9	11	9	9	10	6
5	4	5	5	11	11
6	11	6	6	4	4
12	9	12	12	11	11
8	13	8	8	9	9
7	8	7	7	13	13
5	4	5	5	8	8
9	12	9	9	4	4

## Закінчення таблиці

1	2	3	4	5	6
11	14	11	11	12	12
8	9	8	8	14	14
11	7	11	11	9	9
5	9	5		7	7
6	9	9		10	10

**Варіант 8**

Вимірювання температури в стандартній точці акупунктури, розташованій в області кистей Шеєн-мен (з 7), проводили у чоловіків в положенні «сидячи», з використанням спеціального обладнання. Для оцінки стійкості зміни досліджуваних показників вимірювання проводили повторно на першій, п'ятій, десятій хвилинах. Встановити, чи залежить величина температури даної точки від часу вимірювання.

Стать	Рука	Точка акупунктури	Температура		
			перша хв	п'ята хв	десята хв
М	Р	с 7	27	27	27
М	Р	с 7	30	28	28
М	Р	с 7	30	30	30
М	Р	с 7	29	29	29
М	Р	с 7	21	22	21
М	Р	с 7	23	22	21
М	Р	с 7	27	27	27
М	Р	с 7	27	28	28
М	Р	с 7	21	22	21
М	Р	с 7	30	30	30
М	Л	с 7	28	28	27
М	Л	с 7	29	29	29
М	Л	с 7	31	32	32
М	Л	с 7	29	29	28
М	Л	с 7	21	22	22
М	Л	с 7	22	22	21
М	Л	с 7	27	27	27
М	Л	с 7	28	28	28
М	Л	с 7	21	22	21
М	Л	с 7	31	32	32

**Варіант 9**

Вимірювання електропровідності й електричного опору шкіри в стандартній точці акупунктури, розташованій в області кистей Тай-юань (р 9), проводили у жінок в положенні «сидячи», з використанням спеціального обладнання. Вста-

новити, чи залежать значення електропровідності й електричного опору шкіри від того, на якій руці проводили вимірювання.

Стать	Рука	Точка	Електропровідність	Електричний опір шкіри
W	R	p 9	11	0,632
W	R	p 9	6,5	0,921
W	R	p 9	9	0,092
W	R	p 9	13	0,132
W	R	p 9	6	0,224
W	R	p 9	7	0,184
W	R	p 9	5	0,211
W	R	p 9	7	0,053
W	R	p 9	5	0,053
W	R	p 9	13	0,921
W	L	p 9	12	0,75
W	L	p 9	14,5	0,987
W	L	p 9	6	0,158
W	L	p 9	10	0,197
W	L	p 9	10	0,316
W	L	p 9	7	0,158
W	L	p 9	4	0,303
W	L	p 9	6	0,158
W	L	p 9	4	0,158
W	L	p 9	10	0,987

### Варіант 10

Вимірювання електропровідності, температури й електричного опору шкіри в стандартній точці акупунктури, розташованій в області кистей Тай-юань (p 9), проводили у чоловіків в положенні «сидячи», з використанням спеціального обладнання. Встановити, чи залежать значення електропровідності, температури й електричного опору шкіри від того, на якій руці проводили вимірювання.

Стать	Рука	Точка	Електропровідність	Температура	Електричний опір шкіри
1	2	3	4	5	6
M	R	p 9	20	27	0,079
M	R	p 9	5	29	0,368
M	R	p 9	17	27	0,566
M	R	p 9	2	22	0,75
M	R	p 9	6	23	0,105
M	R	p 9	10	27	0,118
M	R	p 9	11	26	0,013
M	R	p 9	2	21	0,013
M	R	p 9	20	30	0,75
M	L	p 9	6	27	0,316
M	L	p 9	20	28	0,092

## Закінчення таблиці

1	2	3	4	5	6
М	L	p 9	13	29	0,447
М	L	p 9	13	28	0,671
М	L	p 9	3	21	0,829
М	L	p 9	7	22	0,118
М	L	p 9	9	26	0,329
М	L	p 9	9	28	0,026
М	L	p 9	3	21	0,026
М	L	p 9	20	31	0,829

## Варіант 11

Швидкість кровотоку до і після присідань представлена в таблиці. Встановити, чи викликали присідання збільшення швидкості кровотоку.

Швидкість кровотоку, с			
до присідань	після присідань	до присідань	після присідань
4	3,5	4	3,5
5	4	5	4
5	4,2	5	4,2
5,1	4,2	5,1	4,2
5,2	4	5,2	4
6	4,7	6	4,7
6	4,3	6	4,3
5,8	4	5,8	4
5	3	5	3
5,7	3,7	5,7	3,7
5	3,6	5	3,6
4,5	3,1	4,5	3,1
5,9	4,3	5,9	4,3

## Варіант 12

Вимірювання електропровідності, температури й електричного опору шкіри в стандартній точці акупунктури, розташованій в області кистей Тай-юань (p 9), проводили у чоловіків і жінок на правій руці в положенні «сидячи», з використанням спеціального обладнання. Встановити, чи залежать значення електропровідності, температури й електричного опору шкіри від статі пацієнта.

Стать	Рука	Точка	Електропровідність	Температура	Електричний опір шкіри
1	2	3	4	5	6
W	R	p 9	11	28	0,632
W	R	p 9	6,5	29	0,921
W	R	p 9	9	27	0,092
W	R	p 9	13	27	0,132
W	R	p 9	6	24	0,224

Закінчення таблиці

1	2	3	4	5	6
W	R	p 9	7	25	0,184
W	R	p 9	5	22	0,211
W	R	p 9	7	23	0,053
W	R	p 9	5	21	0,053
W	R	p 9	13	29	0,921
M	R	p 9	5	27	0,329
M	R	p 9	20	27	0,079
M	R	p 9	5	29	0,368
M	R	p 9	17	27	0,566
M	R	p 9	2	22	0,75
M	R	p 9	6	23	0,105
M	R	p 9	10	27	0,118
M	R	p 9	11	26	0,013
M	R	p 9	2	21	0,013
M	R	p 9	20	30	0,75

**Варіант 13**

Встановити, чи є відмінність у вмісті циркулюючих імунних комплексів (ЦИК) у плазмі крові у хворих на аскаридоз і ентеробіоз.

Вміст ЦИК (ум. од.)			
Аскаридоз		Ентеробіоз	
67	97,3	58,4	37,5
63,2	104,9	159	82,2
66,41	31,9	68,3	60,1
61	60,8	45,6	53,8
169	33,8	90	115,01
108,7	67,8	69,7	64,3
67,2	61	65	61,9
35,4	111,01	50,4	62,1
90,5	70,2	51,2	90,5
109,8	69,5	62	158,6
92,4	65,4	110,4	117,2
68,3	107,2	114,1	53,1
110,2	66	84,6	112,4
62,9	67,2	87,3	62,2
66,2	67	107,3	183
89,9	72	66,4	65,7
105,1		63	85,6
61		65,3	115
98,6		69	43

**Варіант 14**

Проаналізувати, чи змінюється вміст малонового діальдегіду (МДА) в плазмі крові у здорових пацієнтів різних вікових груп.

Вміст МДА (мкмоль / л)					
30–50 років		51–70 років		71–80 років	
7,9	7,4	12,8	12,8	15,5	15,7
6,8	8,6	13,6	11,7	15,3	15,7
6,2	5,6	13,8	10,8	15,2	16
6	6,2	14	12,5	15,1	16,1
5,8	6,4	11,8	13,3	15,9	16,3
5,4	6	10,4	14,7	16,6	16,3
5,1		13,7	13,4	16,8	15,4
5,6		14	13,9	16	15,6
5,8		15,1	13,6	16,1	15,2
6,8		14,9	15,4	15,8	15,4
9,4		14,8	14,9	15,5	15,2
10,1		13,7	13	15,1	14,9
8,4		12,8	11,4	14,4	15
9,1		11,6		16,4	15,8

**Варіант 15**

Проаналізувати, чи змінюється активність ферменту каталази в крові у здорових пацієнтів різних вікових груп.

Активність каталази (мккат / мг білка)					
30–50 років		51–70 років		71–80 років	
420	356	296	301	278	296
385	361	295	320	275	299
340	355	260	345	267	300
465	345	300	354	278	304
469	389	315	298	265	276
456	320	320	291	260	287
454	337	298	290	275	296
425	335	296	298	280	285
403	418	294	300	286	276
356	365	295	304	280	279
415	360	291	325	295	289
385	395	356	320	291	277
389	452	335	315	289	270
389	378	350	310	267	255

### Варіант 16

Вивчали вплив кобальту на масу тіла кроликів. Дослід проводили на двох групах тварин: дослідній і контрольній. Були досліджені кролики у віці від півтора до двох місяців, масою тіла 500–600 г. Тварин обох груп утримували на одному і тому ж кормовому раціоні. Однак дослідні кролики, на відміну від контрольних, щодня отримували добавку до раціону у вигляді водного розчину по 0,06 г хлористого кобальту на 1 кг живої маси тіла. За час дослідження тварини дали наступні надбавки живої маси тіла. Встановити, чи достовірно впливає добавка кобальту до раціону на збільшення маси тіла.

	Приріст, г								
Піддослідна група	580	692	700	621	640	561	680	630	
Контрольна група	504	560	420	600	580	530	490	580	470

### Варіанти завдань для закріплення теми 5

#### «Порівняння середніх показників вибірок. Непараметричні критерії»

##### Варіант 1

Вивчалася полярографічна активність сироватки крові у десяти випробовуваних при різній довжині хвилі. Результати вимірювань наведені в таблиці. Встановити, чи впливає тривалість захворювання на активність сироватки.

Довжина хвилі, нм	Полярографічна активність фільтрату сироватки крові											
340	122	104	110	72	140	104	80	170	86	110	103	124
370	90	79	101	86	171	142	82	104	108	64	72	106
420	69	71	116	60	110	64	120	106	92	76	53	102
450	51	121	122	94	64	66	72	135	75	114	75	122

##### Варіант 2

Вивчалася кількість сульфгідрильних груп сироватки крові у десяти випробовуваних при різній тривалості захворювання. Результати вимірювань наведені в таблиці. Встановити, чи впливає тривалість захворювання на кількість сульфгідрильних груп сироватки крові.

Період захворювання	Кількість сульфгідрильних груп сироватки крові, мМ/мл											
1	41	31	38	39	38	37	34	38	40	42	44	
2	31	35	32	39	37	37	36	31	32	37	30	
3	10	11	7	14	10	8	17	8	10	11	10	

##### Варіант 3

Вивчався вплив зростаючих доз добрив на врожайність. Дослід проводився у 18 повторностях. Результати вимірювань наведені в таблиці. Встановити, чи впливає доза добрив на урожайність.

Повторності	Дози добрива					Повторності	Дози добрива				
1	30	35	38	37		10	30	35	36	35	
2	35	36	39	36		11	34	35	36	35	
3	34	37	39	38		12	35	33	39	38	
4	34	36	40	37		13	35	35	46	39	
5	33	33	37	35		14	30	37	42	33	
6	30	34	38	39		15	31	37	41	38	
7	32	35	38	36		16	31	39	36	39	
8	33	34	36	34		17	35	35	48	40	
9	32	33	41	35		18	34	33	39	35	

#### Варіант 4

У 12 популяціях ромену круглолистного, зібраних із різної висоти над рівнем моря, вимірювалася величина придатку на сім'яках – «коронки». Результати вимірювань наведені в таблиці. Встановити, чи впливає висота над рівнем моря на величину придатка на сім'яках.

Висота над рівнем моря, м	Величина придатка на сім'яках											
120	3	5	4	1	6	6	24	16	7	3	5	4
260	2	16	21	8	8	1	8	6	4	6	6	2
750	5	8	9	8	10	15	9	14	7	6	2	6

#### Варіант 5

Вимірювали швидкість кровотоку у дітей до дев'яти місяців різними методами. Результати вимірювань наведені в таблиці. Встановити, чи залежить величина швидкості кровотоку від методу вимірювання?

Метод вимірювання	Швидкість кровотоку у дітей у віці до дев'яти місяців															
1	11	9	7	5	6	7	8	4	8	6	7	9	1	12	7	5
2	13	12	9	13	11	9	8	10	13	8	7	11	10	11	9	11

#### Варіант 6

Швидкість кровотоку при отруєнні оцтовою кислотою вимірювали до і після вливання кровозамінника. Результати вимірювань наведені в таблиці. Встановити, чи впливає кровозамінник на швидкість кровотоку.

	Швидкість кровотоку												
до	8	2	2	4	4	1	2	3	2	5	7	4	4
після	3	6	1	1	1	2	3	0	4	3	6	4	0

#### Варіант 7

На пухлину карциноми Герена у восьми випробовуваних впливали магнітним полем різної частоти (А). Встановити, чи здійснює вплив частота поля на розмір пухлини.

Частота поля	Розмір пухлини карциноми Герена							
	A1	0,027	0,036	0,01	0,12	0,32	0,045	0,05
A2	0,75	0,4	0,8	0,15	0,75	0,082	0,06	0,75

### Варіант 8

У дев'яти повторях підраховували число пелюсток водозбору при різній освітленості. Результати вимірювань наведені в таблиці. Встановити, чи впливає освітленість на досліджувану ознаку.

	Число пелюсток водозбору								
	у тіні	1	0	3	3	41	8	1	6
на сонці	0	3	3	1	30	4		1	0

### Варіант 9

Вимірювали частоту пульсу (% відхилення від вікової норми) при отруєнні оцтовою кислотою в станах різного ступеня тяжкості. Результати вимірювань наведені в таблиці. Встановити, чи залежить частота пульсу від тяжкості отруєння.

Ступінь тяжкості стану	Відхилення від норми частоти пульсу, %							
вкрай важкий	28	35	25	40	28	29	7	6
важкий	24	9	30	44	30	44	0	4
середній	13	14	20	2	2	14	44	4
початкова стадія	24	73	42	90	37	28	40	42

### Варіант 10

Вивчали рівень смертності від ішемічної хвороби серця в різних містах України за період 1992–1999 рр. Проаналізувати залежність смертності від її структури.

Населений пункт	Рік							
	1992	1993	1994	1995	1996	1997	1998	1999
Полтава	37,1	67,9	78,4	67,5	69,0	78,9	122,0	114,5
Дніпро	78,0	56,0	111,0	89,0	96,5	198,8	225,2	314,2
Львів	113,7	67,2	82,2	89,8	76,4	134,5	165,2	147,0
Київ	88,4	123,0	167,6	123,7	45,0	176,8	259,7	188,2

### Варіант 11

Був проведений експеримент з підгодівлі 32 свиноматок препаратом афоромом, що містив залізо і мідь, з метою зменшення відсотка мертвороджених поросят. Були отримані такі результати:

Номера маток	Мертвонароджені, %		Номера маток	Мертвонароджені, %	
	з афоромом	без афорому		з афоромом	без афорому
1	1,0	8,3	19	0,0	9,1
2	1,0	12,5	20	0,0	10,0
3	2,0	9,1	21	2,0	40,0
4	18,2	22,2	22	0,0	15,9
5	0,0	10,0	23	0,0	18,0
6	25,0	33,3	24	0,0	13,5
7	10,0	0,0	25	1,0	34,9
8	11,1	0,0	26	1,0	26,1
9	0,5	16,7	27	2,0	32,0
10	0,4	28,6	28	11,0	0,0
11	1,0	25,0	29	14,0	1,0
12	11,1	0,0	30	18,0	0,0
13	11,1	0,0	31	0,0	2,4
14	0,0	25,0	32	1,0	12,0
15	1,0	9,1	33	2,0	8,0
16	2,0	14,3	34	16,9	4,1
17	2,0	35,7	35	24,0	2,0
18	1,0	63,6	36	0,0	40,2

Визначити вплив афорому на народжуваність поросят.

### Варіант 12

На трьох групах щурів був поставлений дослід із порівняння впливу раціону харчування на збільшення ваги. Щури кожної групи отримували раціон із різним вмістом білка. За 56 днів досліді приріст для кожної групи становив:

Раціон	Середній приріст (г)														
	Високобілковий	134	136	104	120	124	161	–	83	113	129	70	123	45	13
З нормальним вмістом білка	–	–	–	104	88	69	92	81	59	76	15	89	6	4	97
Низькобілковий	70	118	101	85	107	132	94	67	54	58	9	–	–	–	–

Провести аналіз впливу раціону харчування на середній приріст тварин.

### Варіант 13

Дві групи вагітних із пізнім гестозом та анемією отримували лікування за різними методиками: перша – за загальноприйнятим способом, друга – за допомогою комплексної терапії, що містить препарат «Хофітол». Встановити, чи впливає спосіб лікування на вміст феритину в крові. Оцінити ефективність двох способів терапії шляхом порівняння з показниками до лікування.

група 1	Вміст феритину, нг / мл															
до лікування	2,4	3,5	5,6	5,9	6,8	6,1	8,3	4,3	8,1	5,9	6,1	7,8	8,1	8	6	5,3
після лікування	4,4	5,5	6,6	6,4	6,9	8,2	9,5	8,8	5,9	6,7	8,5	8,8	9,3	9,5	7,1	6,3
група 2	Вміст феритину, нг / мл															
до лікування	4,7	5,1	5,6	6,7	7,2	9,1	8,3	8,8	7,4	7,6	4,9	5	6	7,1	7,6	6,8
після лікування	2,8	1,5	0,4	2,8	3,1	0,1	3,8	3,1	3,9	1,9	7,1	2,3	1,3	0,7	8,9	2,1

#### Варіант 14

Дві групи вагітних із пізнім гестозом та анемією отримували лікування за різними методиками: перша – за загальноприйнятим способом, друга – за допомогою комплексної терапії, що містить препарат «Хофітол». Встановити, чи впливає спосіб лікування на вміст гемоглобіну в крові.

група 1	Вміст гемоглобіну, г/л															
до лікування	108	103	105	109	101	108	110	106	106	108	110	102	105	113	107	109
після лікування	110	105	109	112	107	109	111	108	109	110	110	107	108	112	109	110
група 2	Вміст гемоглобіну, г/л															
до лікування	109	107	111	108	106	108	109	109	110	112	116	110	112	110	107	106
після лікування	120	119	115	110	109	125	112	120	115	120	118	119	118	119	111	109

#### Варіант 15

Дві групи вагітних із пізнім гестозом та анемією отримували лікування за різними методиками: перша – за загальноприйнятим способом, друга – за допомогою комплексної терапії, що містить препарат «Хофітол». Встановити, чи впливає спосіб лікування на вміст сироваткового заліза в крові. Оцінити ефективність двох способів терапії шляхом порівняння з показниками до лікування.

Сироваткове залізо, мкг %			
перша група		друга група	
до лікування	після лікування	до лікування	після лікування
1	2	3	4
0,23	0,25	0,241	0,427
0,24	0,266	0,24	0,415
0,238	0,248	0,245	0,41
0,239	0,269	0,343	0,416
0,24	0,25	0,238	0,42
0,239	0,27	0,239	0,422
0,241	0,261	0,24	0,409
0,245	0,25	0,238	0,41
0,242	0,262	0,239	0,413
0,241	0,256	0,243	0,425
0,243	0,249	0,241	0,41
0,238	0,253	0,24	431
0,238	0,272	0,243	0,413
0,236	0,248	0,245	0,44

Закінчення таблиці

1	2	3	4
0,237	0,252	0,246	0,409
0,24	0,26	0,24	0,41
0,244	0,271	0,239	0,408

**Варіант 16**

Вивчали рівень смертності від гіпертонічної хвороби серця в різних містах України за період 2002–2009 рр. Проаналізувати залежність смертності від місця проживання.

Населений пункт	Рік							
	2002	2003	2004	2005	2006	2007	2008	2009
Дніпро	45,9	48,0	54,3	67,9	46,1	13,9	87,0	82,1
Київ	32,1	37,4	42,0	46,0	56,0	78,2	72,1	75,4
Львів	12,6	22,5	25,0	20,0	24,4	27,0	32,4	30,0

**Варіант 17**

Дві групи вагітних із пізнім гестозом та анемією отримували лікування за різними методиками: перша – за загальноприйнятим способом, друга – за допомогою комплексної терапії, що містить препарат «Хофітол». Встановити, чи впливає спосіб лікування на утримання трансферину в крові. Оцінити ефективність двох способів терапії шляхом порівняння з показниками до лікування.

Вміст трансферину, %			
перша група		друга група	
до лікування	після лікування	до лікування	після лікування
9,31	9,3	8,91	8,21
8,9	8,95	8,96	8,57
8,9	9,11	8,71	8,34
9,13	9,22	8,82	7,98
8,91	9,3	8,98	8,22
9,12	9,1	8,93	8,56
9,15	9,16	9,26	8,49
9,09	9,22	9,12	8,15
9,61	9,23	9,03	7,98
9,6	9,45	9,93	8,73
9,15	9,4	8,99	8,23
8,95	9,14	9,09	8,34
9,18	8,99	8,92	8,36
8,93	9,17	9,04	8,1
8,97	9,95	9,2	8,51
8,9	8,96	9,07	8,05
9,24	8,9	9,09	8,04

**Варіант 18**

Вивчали вплив туберкуліну на склад периферичної крові нижчих мавп.  
Підтвердити достовірність впливу туберкуліну на склад периферичної крові.

Номери підслідних тварин	Еозинофіли крові	
	до введення туберкуліну	після введення туберкуліну
1	2	2
2	3	2
3	2	1
4	2	1
5	3	2
6	2	3
7	2	1
8	1	2
9	3	2
10	2	1
11	2	2
12	3	1
13	2	1
14	2	1

## Варіанти завдань для закріплення теми 6 «Кореляційний аналіз»

### Варіант 1

Одночасно з вимірюванням довжини максимального стеблового листа проводилися вимірювання висоти рослин (в см). Знайти і проаналізувати (якщо така існує) залежність (кореляцію) між зазначеними ознаками за допомогою пакету аналізу даних і стандартної статистичної функції. Порівняти ці значення.

Довжина стеблового листка	Висота рослин	Довжина стеблового листка	Висота рослин	Довжина стеблового листка	Висота рослин
56	38	57	60	38	59
61	40	78	61	59	60
66	43	68	61	45	60
67	43	43	62	49	60
53	44	62	62	56	60
84	44	64	62	54	60
82	45	48	63	52	60
80	47	74	63	66	76
44	47	43	64	56	76
68	48	40	64	52	78
60	49	53	64	87	80
61	50	76	66	44	82
54	52	72	66	69	84
60	53	74	67	64	86
60	53	76	68	64	87
62	54	86	68	63	57
62	54	50	69	63	76
55	54	69	69	60	57
55	55	69	69	54	74
54	55	71	71	55	56
57	55	76	71	71	74
47	56	47	72	69	69

### Варіант 2

У процесі вивчення максимального артеріального тиску у дітей різного віку з різною вагою були отримані такі дані:

Вага дитини, г	Максимальний артеріальний тиск	Вік, міс.	Вага дитини, г	Максимальний артеріальний тиск	Вік, міс.
1	2	3	4	5	6
9 680	120	10	12 500	96	30
8 760	115	10	13 000	95	35
8 650	105	9	14 000	98	36
9 250	108	9	16 200	96	50
7 600	110	7	14 500	90	42
9 850	115	1	15 900	95	48

Закінчення таблиці

1	2	3	4	5	6
11 250	105	13	14 700	95	42
10 200	108	14	16 000	102	48
10 050	110	11	16 300	95	48
9 300	110	8	17 500	105	54
9 850	105	5	18 700	100	60
12 300	110	8	19 100	102	60
7 070	108	5	19 000	105	60
6 500	98	6	18 000	102	54
13 100	105	9	11 000	120	13
12 100	110	3	10 250	107	14
13 000	115	4	12 300	107	23
5 700	110	4	13 000	98	24
7 400	108	6	12 900	96	24
6 450	100	5	14 100	95	30
9 600	115	0	12 700	98	24
11 100	105	5	13 200	100	24
7 350	115	7	13 900	105	36
10 500	120	4	13 700	94	36
14 000	98	6	7 400	108	6

Знайти і проаналізувати (якщо існує) залежність (кореляцію) між зазначеними ознаками за допомогою пакету аналізу даних і стандартної статистичної функції. Порівняти ці значення.

### Варіант 3

Вимірювали частоту пульсу і максимальний артеріальний тиск у здорових дітей у віці від 3 місяців до 5 років. Знайти і проаналізувати (якщо така існує) залежність (кореляцію) між зазначеними ознаками за допомогою пакету аналізу даних і стандартної статистичної функції. Порівняти ці значення.

Частота Пульсу	Артеріальний тиск	Частота пульсу	Артеріальний тиск	Частота пульсу	Артеріальний тиск
1	2	3	4	5	6
123	100	127	105	109	110
117	92	121	107	111	108
122	105	99	111	115	109
109	92	118	108	117	105
119	90	113	98	110	100
95	101	112	108	107	96
115	105	108	110	113	103
120	100	117	95	97	107
128	105	111	105	118	110
131	98	109	115	96	103

Закінчення таблиці

1	2	3	4	5	6
129	102	105	98	98	102
127	104	107	96	118	96
122	106	100	100	107	98
128	100	102	95	123	110
121	105	90	97	107	105
119	100	100	98	115	102
123	106	78	94	110	105
108	105	87	98	105	99
120	103	90	100	85	95
116	100	103	105	94	102
118	108	99	100	83	95
120	100	110	102	92	100
112	105	92	96	95	102
108	103	88	90	100	105
109	100	100	95		

**Варіант 4**

У процесі підрахунку числа міжвузлів у батьківських рослин пшениці і в потомстві були отримані такі дані:

Число міжвузлів					
у батьківських рослин			у потомстві		
1	2	3	4	5	6
6	9	6	8	8	9
6	9	7	8	7	8
7	8	8	8	7	7
9	7	8	7	7	8
8	8	6	8	7	7
6	8	6	7	8	8
7	6	7	8	8	7
9	7	7	8	9	8
8	9	6	9	8	8
6	6	8	9	7	7
6	6	9	6	8	9
6	6	9	9	8	6
8	8	6	8	8	8
8	8	8	8	7	7
7	7	9	7	7	7
9	7	7	7	8	7
7	8	9	7	8	8
9	8	7	9	7	7
7	7	8	9	7	8
7	7	9	6	8	7
6	9	8	9	6	8

## Закінчення таблиці

1	2	3	4	5	6
8	9	6	8	8	9
7	9	8	9	7	9
9	7	6	9	9	8
9	8	6	6	9	7
6	7	7	8	7	8
8	9	8	8	8	7
9	7	6	9	8	8
6	8	7	9	7	9
8	6	6	6	8	8
7	9	6	6	8	9
6	9	8	9	7	9
9	7	9	9	9	6

Знайти і проаналізувати (якщо існує) залежність (кореляцію) між зазначеними ознаками за допомогою пакету аналізу даних і стандартної статистичної функції. Порівняти ці значення.

**Варіант 5**

Вивчали довжину крила польового жайворонка (в мм) в різних географічних широтах (у градусах). Знайти і проаналізувати (якщо існує) залежність (кореляцію) між зазначеними ознаками за допомогою пакету аналізу даних і стандартної статистичної функції. Порівняти ці значення.

Довжина Крила	Географічна частота	Довжина крила	Географічна частота	Довжина крила	Географічна частота
1	2	3	4	5	6
104	67	102	49	105	52
101	56	101	52	100	43
98	38	99	34	105	67
111	47	101	38	102	53
103	60	102	45	102	48
96	44	97	45	112	47
95	45	105	53	121	57
94	52	86	39	100	36
105	52	112	59	102	53
110	50	106	60	88	39
101	48	97	33	104	52
100	60	103	68	96	39
92	39	102	48	101	47
100	52	107	53	95	45
91	37	106	67	89	38
84	42	97	46	107	44
95	43	100	54	89	36

Закінчення таблиці

1	2	3	4	5	6
102	62	106	49	102	56
115	44	89	41	102	49
115	57	98	53	101	53
112	65	96	51	111	59
118	53	87	41	87	39
109	57	102	44	109	55
116	59	85	38	111	49
107	51	115	59	120	44
104	45	99	43	99	33
95	43	101	53	98	50
97	33	104	57	99	43
88	36	102	49	104	57

**Варіант 6**

Досліджували довжину пластинки листка (в мм) і число зубців на ній. Знайти і проаналізувати (якщо існує) залежність (кореляцію) між зазначеними ознаками за допомогою пакету аналізу даних і стандартної статистичної функції. Порівняти ці значення.

Довжина пластинки	Число зубців на пластинці	Довжина пластинки	Число зубців на пластинці	Довжина пластинки	Число зубців на пластинці
1	2	3	4	5	6
56	63	37	36	59	56
55	52	50	50	55	42
66	56	47	57	72	49
58	46	52	65	52	52
55	55	74	55	57	47
54	33	57	52	44	48
53	41	50	49	46	62
55	63	35	37	45	39
64	55	36	32	64	46
62	62	63	44	45	36
53	58	44	57	48	42
52	43	48	38	42	45
50	52	54	37	82	66
70	58	34	47	48	48
53	40	69	62	34	39
57	41	57	42	63	39
69	55	44	51	33	46
51	52	41	35	63	47
54	45	41	49	47	38
63	36	49	52	41	58
35	38	28	33	36	34

Закінчення таблиці

1	2	3	4	5	6
28	30	39	31	54	35
66	39	31	36	43	47
20	38	58	42	55	36
60	44	45	41	74	59
50	41	44	47	36	36
39	41	56	40	52	45
55	48	30	40	43	34
63	34	40	54	60	47
30	33	39	31	54	35

**Варіант 7**

Знайти і проаналізувати (якщо існує) залежність (кореляцію) між парами досліджуваних ознак за допомогою пакету аналізу даних і стандартної статистичної функції. Порівняти ці значення.

Вага телят, г	Коефіцієнт зрілості	Кількість ікринок	Вага ооциту	Вага телят, г	Коефіцієнт зрілості	Кількість ікринок	Вага ооциту
13,1	5,1	319	1,6	9,4	6,2	185	2,6
14,5	6,3	307	2,2	12,5	5,5	260	2
14,6	5,8	239	2,3	29	6,3	490	3
17,4	6,1	290	3,1	29,6	6	365	2,7
27,5	7,6	429	3,9	22,3	7,7	397	3,7
17,5	5,2	270	2,6	23,9	6,1	546	1,9
12,6	6,4	252	2,6	36,5	5,8	647	2,6
11,9	6,3	247	2,1	25,1	5,4	305	2,8
11,3	6,2	232	2,2	27,2	5,6	310	2,9
22,3	5	201	3,3	30,7	5	380	3,3
22,9	5,9	383	2,7	23	6,5	400	2,8
10,1	5	280	1,5	7,8	6,1	142	2,8
10,2	5,7	254	1,5	32,6	5,4	390	3,4
10,4	5,5	145	2,6	23,6	5,6	428	2,3
20,4	5,5	266	3,4	11,7	5,8	219	2,4

**Варіант 8**

Знайти і проаналізувати (якщо існує) залежність (кореляцію) між парами досліджуваних ознак за допомогою пакету аналізу даних і стандартної статистичної функції. Порівняти ці значення.

Вік хворих, міс.	Групи сироватки крові, М/мл	Висота фільтратної хвилі	Вік хворих, міс.	Групи сироватки крові, М/мл	Висота фільтратної хвилі
57	45	35	54	50,75	38
6	32,25	51	24	46,5	86
36	30,25	108	138	47,5	42
56	34,25	68	138	42,75	32
156	34,25	168	30	48,25	44
120	31,25	103	60	38,5	66
29	52,25	48	24	40,25	40
60	24,5	89	48	40,75	32
101	33,25	92	72	45	49
168	31	90	42	50,25	335
144	37,75	44	144	53,5	48
136	12,75	102	36	42,25	437
108	41,75	64	48	49,25	35
72	37,25	62	60	57,5	47
84	22	77	96	47,65	44
65	36,75	100	84	53,75	47
36	28,25	74	108	54	36
54	41,75	89	72	49	51
84	30,25	124	120	49,25	32
156	32,25	119	49	57,25	92
168	41	62	30	50,75	61
66	22,25	109	42	46,5	107
24	27,25	59	108	47,5	69
48	20,75	109	60	42,75	60
132	27,75	72	96	38,25	40
42	40	46	108	17,25	102
72	41	92	60	31,25	36
90	44,75	98	30	48,25	34
120	40,25	82	144	41,5	42
100	36,75	52	78	30,75	38
12	24,5	44	102	49	28
18	39,25	74	54	49,25	36
90	46	152	24	47,25	114
156	49,25	162	24	46,5	80
144	50	184	84	40,25	68
72	43,5	148	24	38,25	36
132	47,75	124	108	17,25	34
108	22,75	163	144	31,25	43
120	48,25	76	90	48,5	36
54	49	56	782	30,75	82
132	50,75	45	66	49,25	41
120	46,75	100	138	57,25	25
90	42,75	74	138	42,75	32

**Варіант 9**

Знайти і проаналізувати (якщо існує) залежність (кореляцію) між парами досліджуваних ознак за допомогою пакету аналізу даних і стандартної статистичної функції. Порівняти ці значення.

Рік	Відхилення від норми рівня лейкоцитів, % на тис. населення	Радіаційний фон	Рік	Відхилення від норми рівня лейкоцитів, % на тис. населення	Радіаційний фон
1971	8,3	77,25	1986	14,3	133,10
1972	12,5	116,35	1987	35,7	332,28
1973	9,1	84,70	1988	63,6	591,97
1974	22,2	206,63	1989	9,1	84,70
1975	10,0	93,08	1990	10,0	93,08
1976	33,3	310,00	1991	40,0	372,31
1977	11,2	104,25	1992	15,9	147,69
1978	35,6	331,35	1993	18,0	167,54
1979	16,7	155,44	1994	13,5	125,65
1980	28,6	266,20	1995	34,9	324,85
1981	25,0	232,69	1996	26,1	242,93
1982	47,0	437,46	1997	32,0	297,85
1983	14,5	134,96	1998	27,8	258,75
1984	25,0	232,69	1999	14,0	130,31
1985	9,1	84,70	2000	26,1	242,73

**Варіант 10**

Знайти і проаналізувати (якщо існує) залежність (кореляцію) між парами досліджуваних ознак за допомогою пакету аналізу даних і стандартної статистичної функції. Порівняти ці значення

Дата	Вміст пігменту «а»	Вміст каротиноїдів	Середньодобова температура води	Рівень освітленості
1	2	3	4	5
10.03.2008	0,138	0,09	12	56
17.03.2008	0,138	0,056	14	52
24.03.2008	0,140	0,17	15	53
07.04.2008	0,142	0,052	15	53
14.04.2008	0,144	0,1	16	56
21.04.2008	0,146	0,004	16	56
02.05.2008	0,144	0,002	17,5	53
13.05.2008	0,144	0,01	18	53
27.05.2008	0,142	0,09	20	53
03.06.2008	0,152	0,08	21,6	51
11.06.2008	0,140	0,07	24	57
27.06.2008	0,140	0,05	24	50
08.07.2008	0,146	0,002	26	55

Закінчення таблиці

1	2	3	4	5
15.07.2008	0,142	0,003	23	50
21.07.2008	0,142	0,004	27	54
30.07.2008	0,136	0,104	26	53
12.08.2008	0,134	0,105	24	50
19.08.2008	0,138	0,132	25,5	50
26.08.2008	0,140	0,056	26	56
01.09.2008	0,140	0,052	21	52
09.09.2008	0,136	0,1	22	51
19.09.2008	0,136	0,052	21	56
27.09.2008	0,138	0,049	18	54
05.10.2008	0,144	0,012	17	51
14.10.2008	0,140	0,004	18	54
25.10.2008	0,144	0,4	18	52
03.11.2008	0,142	0,0072	17	50
11.11.2008	0,140	0,132	16	53
28.11.2008	0,138	0,004	17	50

**Варіант 11**

У 36 аналізах крові визначали кількість еритроцитів (у млн), вміст гемоглобіну (у %) і осідання крові за 24 години (в мм).

Кількість еритроцитів, млн	Вміст гемоглобіну, %	Осідання крові за 24 години, мм	Кількість еритроцитів, млн	Вміст гемоглобіну, %	Осідання крові за 24 години, мм
0,8	54	32	4,33	82	35
1,71	48	26	3,8	79	36
2,63	61	24	3,82	87	37
3,19	66	26	3,81	87	37
2,8	72	28	4,2	87	38
3,14	83	29	4,47	90	40
3,21	73	30	3,71	97	40
3,28	82	30	4,22	96	40
3,63	78	30	3,9	92	44
3,3	82	32	4,36	94	20
4,1	81	32	1,3	52	20
3,29	82	32	2,5	50	26
3,46	77	33	2,8	63	28
3,32	80	33	3,1	71	29
3,11	82	34	2,87	70	30
3,28	79	34	3,68	72	30
3,66	84	34	3,59	76	30

Визначити, чи існує залежність між парами досліджуваних ознак.

### Варіант 12

У процесі вивчення максимального артеріального тиску у дітей різного віку з різною вагою були отримані дані, наведені в таблиці. Визначити, чи існує залежність між зазначеними ознаками.

Вага дитини, г	Максимальний артеріальний тиск	Вага дитини, г	Максимальний артеріальний тиск	Вага дитини, г	Максимальний артеріальний тиск
5 700	110	12 100	110	10 200	108
7 070	108	12 300	107	10 500	120
6 450	100	13 000	115	10 250	107
6 500	98	13 000	98	9 850	105
7 400	108	12 900	96	11 100	105
7 600	110	12 700	98	12 300	110
7 350	115	13 200	100	13 100	105
9 300	110	14 100	95	16 000	102
8 650	105	12 500	96	16 300	95
9 250	108	13 000	95	16 200	96
9 680	120	13 900	105	17 500	105
8 760	115	13 700	94	18 000	102
9 600	115	14 000	98	18 700	100
9 850	115	14 000	98	19 100	102
10 050	110	14 500	90	19 000	105
11 250	105	14 700	120	13 700	94
11 000	120	15 900	95	14 000	98

### Варіант 13

Встановити, чи залежить вміст пігментів і каротиноїдів від рівня освітленості і температури води. Чи існує достовірний зв'язок між вмістом каротиноїдів і пігментів?

Дата	Вміст пігменту «а»	Вміст каротиноїдів	Середньодобова температура води	Рівень освітленості
1	2	3	4	5
10.03.2009	0,138	0,09	12	56
17.03.2009	0,138	0,056	14	52
24.03.2009	0,140	0,17	15	53
07.04.2009	0,142	0,052	15	53
14.04.2009	0,144	0,1	16	56
21.04.2009	0,146	0,004	16	56
02.05.2009	0,144	0,002	17,5	53
13.05.2009	0,144	0,01	18	53
27.05.2009	0,142	0,09	20	53
03.06.2009	0,152	0,08	21,6	51
11.06.2009	0,140	0,07	24	57

Закінчення таблиці

1	2	3	4	5
27.06.2009	0,140	0,05	24	50
08.07.2009	0,146	0,002	26	55
15.07.2009	0,142	0,003	23	50
21.07.2009	0,142	0,004	27	54
30.07.2009	0,136	0,104	26	53
12.08.2009	0,134	0,1056	24	50
19.08.2009	0,138	0,132	25,5	50
26.08.2009	0,140	0,056	26	56
01.09.2009	0,140	0,052	21	52
09.09.2009	0,136	0,1	22	51
19.09.2009	0,136	0,052	21	56
27.09.2009	0,138	0,049	18	54
05.10.2009	0,144	0,012	17	51
14.10.2009	0,140	0,004	18	54
25.10.2009	0,144	0,4	18	52
03.11.2009	0,142	0,0072	17	50
11.11.2009	0,140	0,132	16	53
28.11.2009	0,138	0,004	17	50

**Варіант 14**

Встановити, чи існує достовірний зв'язок між параметрами, зазначеними в таблиці.

Дата	Біологічне споживання O <sub>2</sub> (БСК)	Вміст розчиненого O <sub>2</sub>	Окиснюваність (S)	Ступінь забрудненості гідрооб'єкта (P)
1	2	3	4	5
31.03.2010	6,6	98,160	5,4	0,067
01.04.2010	6,7	93,567	5,8	0,072
02.04.2010	6,6	89,636	6,3	0,074
03.04.2010	5,8	100,000	8	0,058
04.04.2010	6,5	67,797	10,8	0,096
05.04.2010	6,8	66,946	10,2	0,102
06.04.2010	7,2	63,745	7,8	0,113
07.04.2010	6,0	63,241	7,4	0,095
08.04.2010	7,3	65,173	8,9	0,112
09.04.2010	6,3	66,667	12	0,095
10.04.2010	6,9	68,085	13	0,101
11.04.2010	6,3	65,306	11	0,096
12.04.2010	9,0	66,667	10	0,135
13.04.2010	6,5	66,667	10	0,098
14.04.2010	7,6	86,022	4,8	0,088
15.04.2010	8,0	96,970	5	0,083
16.04.2010	7,6	86,022	4,8	0,088

Закінчення таблиці

1	2	3	4	5
17.04.2010	9,0	96,970	5	0,093
18.04.2010	8,0	86,957	5,2	0,092
19.04.2010	6,2	108,108	5,4	0,057
20.04.2010	8,0	93,023	5,6	0,086
21.04.2010	6,4	90,909	5,8	0,070
22.04.2010	7,9	61,776	6,2	0,128
23.04.2010	8,4	63,492	7,6	0,132
24.04.2010	8,0	60,952	5,5	0,131
25.04.2010	8,0	66,667	10	0,120
26.04.2010	5,0	66,667	10	0,075
27.04.2010	7,0	66,667	10	0,105
28.04.2010	6,0	92,754	5,5	0,065
29.04.2010	5,0	91,954	5,2	0,054
30.04.2010	6,0	95,808	6,6	0,063

**Варіант 15**

Встановити, чи існує достовірний зв'язок між параметрами, зазначеними в таблиці.

Діаметр фолікула (D)	Обсяг фолікула (V)	Площа фолікула (S)	Фактор форми (F)
0,45	0,67	2,78	0,05
0,45	0,69	2,80	0,08
0,45	0,69	2,96	0,09
0,45	0,73	2,89	0,09
0,45	0,78	2,80	0,09
0,45	0,97	2,93	0,09
0,45	1,00	2,89	0,09
0,45	1,00	2,89	0,10
0,45	1,02	2,71	0,08
0,45	1,05	2,96	0,06
0,45	1,09	2,91	0,06
0,45	1,09	2,91	0,06
0,45	1,12	3,07	0,06
0,45	1,16	2,93	0,08
0,45	1,19	2,82	0,08
0,45	1,21	2,78	0,08
0,45	1,27	2,71	0,10
0,45	1,27	2,73	0,10
0,45	1,27	2,84	0,10
0,45	1,27	2,80	0,10
0,45	1,30	2,80	0,10

**Варіант 16**

Встановити залежить вмісту фікоеритрину, R-фікоціаніну та зважених частинок від рівня освітленості.

Дата	Вміст фікоеритрину	Вміст R-фікоціаніну	Рівень освітленості	Наявність зважених частинок
10.03.2011	0,038	0,012	25,23	8,3
17.03.2011	0,000	0,067	39,21	12,5
24.03.2011	0,010	0,020	40,99	9,1
07.04.2011	0,030	0,012	40,99	8,1
14.04.2011	0,020	0,130	41,80	10,0
21.04.2011	0,216	0,023	43,64	33,3
02.05.2011	0,035	0,008	44,46	45,0
13.05.2011	0,028	0,015	45,29	34,0
27.05.2011	0,000	0,015	38,23	16,7
03.06.2011	0,020	0,034	26,00	28,6
11.06.2011	0,010	0,017	26,00	54,4
27.06.2011	0,320	0,009	27,52	24,0
08.07.2011	0,008	0,014	29,05	27,3
15.07.2011	0,002	0,011	36,50	25,0
21.07.2011	0,010	0,003	37,46	9,1
30.07.2011	0,005	0,001	37,46	14,3
12.08.2011	0,007	0,014	47,44	35,7
19.08.2011	0,024	0,019	47,44	17,0
26.08.2011	0,006	0,097	47,44	48,8
01.09.2011	0,010	0,087	47,44	33,0
09.09.2011	0,163	0,095	48,52	41,7
19.09.2011	0,011	0,098	49,60	67,0
27.09.2011	0,037	0,025	50,68	24,7
05.10.2011	0,028	0,003	52,52	8,0
14.10.2011	0,006	0,003	52,52	34,9
25.10.2011	1,002	0,006	52,52	67,0
03.11.2011	1,068	0,007	54,58	13,0
11.11.2011	0,095	0,004	51,76	14,0
28.11.2011	0,005	0,060	59,79	92,0

**Варіант 17**

Встановити, чи залежить відхилення лейкоцитів від норми і рівень смертності від рівня сонячної активності і радіаційного фону.

Рік	Відхилення від норми рівня лейкоцитів, % на тис. населення	Смертність, % на тис. населення	Рівень сонячної активності (у числах Вольфа)	Радіаційний фон
1971	8,3	21,43	58	77,25
1972	12,5	25,50	69	116,35
1973	9,1	27,71	75	84,70
1974	22,2	33,63	91	206,63
1975	10,0	4,70	92	93,08
1976	33,3	37,32	101	310,00
1977	11,2	32,89	89	104,25
1978	35,6	32,15	87	331,35
1979	16,7	20,69	56	155,44
1980	28,6	15,52	42	266,20
1981	25,0	11,46	31	232,69
1982	47,0	12,56	34	437,46
1983	14,5	12,56	34	134,96
1984	25,0	20,69	56	232,69
1985	9,1	34,00	92	84,70
1986	14,3	34,37	93	133,10
1987	35,7	36,21	98	332,28
1988	63,6	27,90	109	591,97
1989	9,1	32,15	87	84,70
1990	10,0	17,00	46	93,08
1991	40,0	13,67	37	372,31
1992	15,9	11,82	32	147,69
1993	18,0	12,19	33	167,54
1994	13,5	12,93	35	125,65
1995	34,9	15,15	41	324,85
1996	26,1	18,11	49	242,93
1997	32,0	21,80	59	297,85
1998	27,8	9,93	81	258,75
1999	14,0	41,39	112	130,31

## Варіанти завдань для закріплення теми 7 «Таблиці спряженості»

### Варіант 1

Встановити, чи існує взаємозв'язок між віком і кількістю курців.

№ з/п	Вік групи	Палять	№ з/п	Вік групи	Палять	№, з/п	Вік групи	Палять
1	20–40	Так	18	< 20	Ні	35	> 40	Так
2	< 20	Ні	19	< 20	Ні	36	> 40	Ні
3	> 40	Ні	20	20–40	Так	37	> 40	Так
4	> 40	Ні	21	> 40	Так	38	> 40	Так
5	< 20	Ні	22	< 20	Ні	39	< 20	Ні
6	20–40	Так	23	< 20	Ні	40	> 40	Ні
7	> 40	Так	24	> 40	Так	41	< 20	Так
8	20–40	Ні	25	< 20	Ні	42	< 20	Так
9	20–40	Так	26	20–40	Так	43	< 20	Ні
10	20–40	Ні	27	< 20	Ні	44	20–40	Ні
11	20–40	Так	28	> 40	Так	45	< 20	Так
12	20–40	Так	29	> 40	Так	46	20–40	Ні
13	> 40	Так	30	> 40	Ні	47	20–40	Ні
14	< 20	Ні	31	< 20	Так	48	20–40	Так
15	20–40	Так	32	> 40	Так	49	> 40	Ні
16	< 20	Так	33	20–40	Так	50	> 40	Ні
17	> 40	Так	34	20–40	Ні	51	20–40	Ні

### Варіант 2

Частині із 2 663 пацієнтів були зроблені щеплення проти холери (щеплені пацієнти), а частині – ні (нещеплені пацієнти). Встановити, чи значуще впливають профілактичні заходи на зростання захворювання.

	Не захворіли	Хворі
Щеплені	1 625	5
Нещеплені	1 022	11

### Варіант 3

Вивчали вплив ендотоксину на виживання опромінених тварин. Результати дослідження наведені в таблиці. Встановити достовірність впливу ендотоксину на виживаність тварин.

Група тварин	Вижило	Загило
контрольна	3	11
дослідна	23	13

**Варіант 4**

Три інтерв'юера (Інтр\_1, Інтр\_2 і Інтр\_3,) опитували населення про те, чи хочуть вони користуватися продукцією фірми N. Визначити, чи існують значущі відмінності в отриманих відповідях.

№ з/п	Інтр_1	Інтр_2	Інтр_3	№ з/п	Інтр_1	Інтр_2	Інтр_3
1	НІ	НІ	НІ	10	НІ	НІ	НІ
2	ТАК	ТАК	НІ	11	ТАК	ТАК	ТАК
3	НІ	НІ	НІ	12	ТАК	ТАК	ТАК
4	НІ	НІ	НІ	13	ТАК	ТАК	НІ
5	ТАК	НІ	НІ	14	ТАК	ТАК	НІ
6	ТАК	ТАК	НІ	15	ТАК	ТАК	НІ
7	ТАК	ТАК	НІ	16	ТАК	ТАК	ТАК
8	НІ	ТАК	НІ	17	ТАК	ТАК	НІ
9	ТАК	НІ	НІ	18	ТАК	ТАК	НІ

**Варіант 5**

Встановити, чи залежить виявлення антитіл у сироватці крові онкохворих від локалізації пухлини.

Локалізація	Наявність антитіл	Відсутність антитіл
Шийка матки	12	4
Молочна залоза	37	17
Ліва гомілка	3	0
Кишечник	16	5
Лімфогранулематоз	3	0
Ротова порожнина	2	4
Яечник	35	28
Шкіра	5	2
Шлунок	14	12
Легеня	11	12
Підшлункова залоза	6	2
Матка	4	0

**Варіант 6**

Встановити, чи впливає тяжкість онкозахворювання на наявність антитіл до ДНК у сироватці крові онкохворих.

Ступінь тяжкості захворювання	Наявність антитіл	Відсутність антитіл
1	5	8
2	27	31
3	66	22
4	27	22

### Варіант 7

Досліджували вплив препарату різної концентрації на фібриноліз. Встановити, чи існує зв'язок між тривалістю лікування фібринолізу та концентрацією препарату, що застосовується для лікування.

Концентрація	Тромбоцитарний фібриноген		
	лікування протягом п'яти днів	лікування протягом двох тижнів	лікування протягом одного місяця
C <sub>1</sub>	5,6	5,9	4,9
C <sub>2</sub>	3,1	4,7	6
C <sub>3</sub>	4,8	5,2	5,7
C <sub>4</sub>	3,2	4,8	3,4
C <sub>5</sub>	5	5,3	4,7

### Варіант 8

Досліджували полярографічну активність фільтрату сироватки крові у груп із різним ступенем вираженості спленомегалії при різній тривалості захворювання. Встановити, чи існує статистично значущий зв'язок між тривалістю захворювання та ступенем вираженості спленомегалії.

Ступінь вираженості спленомегалії (A)	Тривалість захворювання:			
	2 тижні	4 тижні	7 тижнів	12 тижнів
A1	98	90	69	51
A2	152	79	71	51
A3	168	101	116	71
A4	143	86	102	94
A5	128	171	110	64
A6	162	142	64	66
A7	96	82	120	72
A8	157	104	141	135
A9	162	108	92	75
A10	184	64	76	124

### Варіант 9

Досліджували ембріотоксичність у різних груп тварин, що мешкають у місцевості з різним радіоактивним фоном.

Група 1 – інтактні вагітні тварини; група 2 – на п'ятий день хвороби вводився водний розчин ембіхіна 0,2 мг/кг; група 3 – на п'ятий день хвороби вводився водний розчин ембіхіна 0,4 мг/кг; група 5 – на п'ятий день хвороби вводився водний розчин ембіхіна 0,6 мг/кг. Проаналізувати, чи існує зв'язок між концентрацією ембіхіна і рівнем радіоактивності (РА).

Концентрація ембіхіна	Середній розмір плаценти, мм		
	РА в межах норми	РА незначно перевищує норму	РА значно вище допустимого рівня
група 1	14	12	7
група 2	13	15	18
група 3	14	14	6
група 4	12	9	19

### Варіант 10

Встановити, чи існує зв'язок між ступенем вираженості хвороби та екологічною забрудненістю місця проживання випробуваних.

	Ознаки захворювання відсутні	Перенесли захворювання в дитячому віці	Хворіють у поточний момент	Хронічний перебіг хвороби
Екологічно чистий район	114	37	13	24
Рівень екологічного забруднення незначно перевищує норму	56	34	22	53
Рівень екологічного забруднення значно перевищує ГДР	78	56	34	127

### Варіант 11

Під впливом опромінення різної інтенсивності і тривалості спостерігали розмноження вірусу Аукуба-мозаїки (в тис.). Встановити, чи існує зв'язок між інтенсивністю і часом опромінення.

Інтенсивність опромінення	Тривалість опромінення, хв								
	0	3	7,5	10	15	20	30	40	60
A 1	344	290	234	202	184	122	74	42	29
A 2	280	245	212	199	107	89	76	34	17
A 3	271	226	209	160	108	92	59	29	12
A 4	268	200	156	159	145	104	43	45	26
A 5	197	202	187	134	128	46	34	28	19

### Варіант 12

Спостерігали пари одностатевих близнят, один з яких був туберкульозним. Встановити, чи існує залежність між захворюваністю на туберкульоз та умовами середовища, в яких жили близнюки.

A1 – повністю подібні умови; A2 – умови подібні, але схожість досить слабка; A3 – умови несхожі, але відмінність виражена не дуже різко; A4 – умови різко різні.

Стан туберкульозу	Умови середовища			
	A1	A2	A3	A4
Повністю подібні	44	16	12	4
Подібні, але слабо	40	4	12	0
Несхожі, але виражені в слабкому ступені	48	16	28	4
Різко відрізняються	4	4	48	16

### Варіант 13

Досліджували вплив різних раціонів харчування і температури навколишнього середовища на зростання шурів. Визначити, чи існує залежність між вмістом білка в їжі і температурними умовами середовища.

Раціон	Температура, °С							
	11	15	19	23	27	31	34	38
Високобілковий	134	136	104	120	124	161	107	83
З нормальним вмістом білка	62	66	81	104	88	69	92	81
Низькобілковий	70	118	101	85	107	132	94	67

### Варіант 14

Досліджували відхилення від норми рівня лейкоцитів (% на тис. населення) при різному рівні сонячної активності (в числах Вольфа) і радіаційного фону.

Радіаційний фон	Рівень сонячної активності (в числах Вольфа)					
	58	69	75	91	92	101
77	8,3	12,5	9,1	22,2	10,0	33,3
84	11,2	35,6	16,7	28,6	25,0	47,0
93	14,5	25,0	9,1	14,3	35,7	63,6
116	9,1	10,0	40,0	15,9	18,0	13,5
206	34,9	26,1	32,0	27,8	14,0	16,7

Визначити, чи існує статистично значущий зв'язок між рівнем радіаційного фону і сонячною активністю.

### Варіант 15

Проводили лікування випробовуваних різними способами. Дані досліджень наведені в таблиці. Встановити, чи існують статистично значущі відмінності періоду одужання від способу лікування.

Лікування	Тривалість одужання:		
	п'ять днів	два тижні	один місяць
Симптоматичне	34	56	23
Специфічне	342	456	176
Із застосуванням вітамінного комплексу	234	346	78
Із застосуванням анаболіків	127	97	34

**Варіант 16**

Досліджували вплив різних антикоагулянтів на фібриноліз. Визначити, чи існує зв'язок між тривалістю лікування фібринолізу і типом коагулянту, що застосовується для лікування.

Антикоагулянти (А)	Тривалість лікування		
	п'ять днів	два тижні	один місяць
A1	5,6	5,9	4,9
A2	3,1	4,7	6
A3	4,8	5,2	5,7
A4	3,2	4,8	3,4
A5	5,0	5,3	4,7

**Варіант 17**

Під час дослідження швидкості кровотоку у дітей різного віку були отримані такі дані:

	Вік дітей, міс.					
	3–6	6–9	9–12	12–18	18–24	24–48
Здорові	11	6	6	9	7	5
Ішемічна хвороба серця	9	7	5	5	8	10
Серцево-судинна дистонія	7	9	7	6	11	9

Визначити, чи існує статистично значуща залежність між наявністю захворювання і віком випробуваних.

**Варіанти завдань для закріплення теми 8 «Дисперсійний аналіз»**  
*Однофакторний дисперсійний аналіз*

**Варіант 1**

На навчально-дослідній ділянці агростанції вивчали вплив різних способів внесення в ґрунт органічних добрив на урожай зеленої маси кукурудзи. Дослід проводили на десятиметрових ділянках. Результати дослідження приведені в таблиці. Визначити, чи впливає спосіб внесення добрив на урожайність зеленої маси кукурудзи.

Спосіб введення добрива	Врожайність, кг						
	23,6	22,6	28	23	27,5		
Нижче насіння на 4 см	24	30	29,2	25,6	24,5	28	26,5
Осторонь від насіння на 4 см	29,2	28	27	25	25,5	28,2	

**Варіант 2**

Порівнювали вплив раціону харчування на збільшення ваги на трьох групах щурів. Щури кожної групи отримували раціон із різним вмістом білка. За 56 днів дослідження приріст (в г) для кожної групи становив:

Раціон	Середній приріст (г)						
	1	2	3	4	5	6	7
Високобілковий	134	136	104	120	124	161	–
З нормальним вмістом білка	–	–	–	104	88	69	92
Низькобілковий	70	118	101	85	107	132	94

Раціон	Середній приріст (г)						
	8	9	10	11	12	13	14
Високобілковий	83	113	129	97	123	145	113
З нормальним вмістом білка	81	59	76	115	89	86	74
Низькобілковий	67	54	58	89	–	–	–

Провести аналіз впливу раціону харчування на середній приріст тварин.

**Варіант 3**

На одній із дослідних станцій спостерігали врожайність шести місцевих сортів пшениці. Встановити, чи залежить врожайність від сорту пшениці.

Сорт пшениці	Врожайність, ц/га					
1	26,1	29,2	30,0	27,3	28,5	
2	25,0	24,3	23,5	29,0	26,5	
3	27,2	26,4	31,0	26,4	28,4	27,0
4	23,6	26,2	25,2	24,8	24,1	26,8
5	30,0	33,0	26,6	29,9	31,2	
6	23,0	26,0	26,0	24,8		

#### Варіант 4

Досліджували вплив різних доз мінеральних добрив на врожайність озимого жита. Встановити, чи впливає обране дозування на урожайність.

Дози добрив, кг / га	Урожай, ц / га					
15	8,0	8,4	9,0	8,6		
20	8,2	9,0	10	10	9,2	10
25	11,0	13,0		12		
30	7,8	8,5				

#### Варіант 5

У трьох вікових групах дітей зі здоровими зубами визначали гігієнічний індекс (ГІ), який виражається в ум. од. Встановити, чи впливає вік дитини на гігієнічний індекс.

Вік	ГІ, умов. од.					
3 роки	1,0	1,0	1,0	2,6	2,7	
4 роки	3,0	1,2	2,0	1,0	2,7	1,3
5 років	3,0	1,3	2,2	1,2	1,3	3,0

#### Варіант 6

Вивчали вплив кобальту на масу тіла кроликів. Дослід проводили на двох групах тварин – дослідній і контрольній. Були досліджені кролики у віці від півтора до двох місяців, масою тіла 500–600 г. Дослід тривав півтора місяці. Тварин обох груп утримували на одному і тому ж кормовому раціоні. Проте дослідні кролики, на відміну від контрольних, щодня отримували добавку до раціону у вигляді водного розчину по 0,06 г хлористого кобальту на 1 кг живої маси тіла. За час досліду тварини дали надбавки живої маси тіла.

	Приріст, г								
Дослідна група	580	692	700	621	640	561	680	630	
Контрольна група	504	560	420	600	580	530	490	580	470

Встановити, чи достовірно впливає добавка кобальту до раціону на прибавку маси тіла?

#### Варіант 7

На двох групах лабораторних мишей (дослідній і контрольній) вивчали вплив на організм нового препарату. Після місячних випробувань маса тіла тварин (г) варіювала так:

	Приріст, г										
Дослідна група	80	76	75	64	70	68	72	79	83		
Контрольна група	70	78	60	80	62	68	73	60	71	66	69

Встановити, чи впливає випробуваний препарат на надбавку ваги тварин.

### Варіант 8

У таблиці наведено результати впливу кобальту на величину маси тіла кроликів. Підтвердити достовірність впливу кобальту на величину маси тіла.

	Маса тіла, г								
Контрольна група	420	470	490	504	530	560	580	580	600
Дослідна група	561	580	621	630	640	680	692	700	

### Варіант 9

Швидкість кровотоку при отруєнні оцтовою кислотою вимірювали до і після вливання кровозамінника. Результати вимірювань наведені в таблиці. Встановити, чи впливає кровозамінник на швидкість кровотоку.

	Швидкість кровотоку при отруєнні оцтовою кислотою												
до	18	22	22	34	24	31	32	23	22	25	17	24	
після	13	6	11	11	11	12	13	10	14	13	6	14	10

### Варіант 10

У таблиці наведено дані вимірювання швидкості кровотоку у дітей різних вікових груп, виміряні різними методами. Встановити, чи впливає вік і метод вимірювання на швидкість кровотоку.

Метод вимірювання	Швидкість кровотоку											
перший	9	5	6	12	8	7	5	9	11	7	8	11
другий	1	4	11	9	13	8	4	12	14	6	7	10

## *Двофакторний дисперсійний аналіз*

### Варіант 1

У таблиці наведені дані щодо кількості пелюсток водозбору, зібраних на різних типах місцевості і при різній освітленості. Встановити, чи впливає тип місцевості й освітлення на число пелюсток водозбору.

	На сонці								
Рівнина	6	6	7	5	8	6	5	6	8
Плоскогір'я	6	5	5	5	4	6	5	4	3
На схилі гори	4	2	2	1	4	3	3	4	2
На вершині	1	2	1	1	2	1	2	1	3

	У тіні								
Рівнина	7	5	6	4	4	3	2	4	3
Плоскогір'я	3	5	4	4	4	3	4	5	2
На схилі гори	2	3	3	4	3	2	1	1	3
На вершині	1	2	1	1	2	1	2	2	3

### Варіант 2

Встановити, чи достовірний вплив факторів А і В на кількість сульфгід-рильних груп сироватки крові (моль/100 мл).

	B1	B2	B3
A1	41	30	34
	31	30	38
	38	28	40
	39	37	42
A2	38	30	44
	32	31	40
	28	40	35
	40	30	32
	32	40	39
A3	32	20	37
	18	20	27
	27	24	21
	17	23	24
	22	24	21
	22	23	27

### Варіант 3

Вивчали рівень смертності в різних містах України за період 2002–2009 рр. Проаналізувати зв'язок між смертністю від хвороб системи кровообігу та рівнем сонячної активності.

Причина смерті	Населений пункт	Рівень сонячної активності (в числах Вольфа)							
		58	69	75	91	92	101	108	111
1	2	3	4	5	6	7	8	9	10
Ішемічна хвороба серця	Полтава	37,1	67,9	78,4	67,5	69,0	78,9	122,0	114,5
	Дніпро	78,0	56,0	111,0	89,0	96,5	198,8	225,2	314,2
	Лисичанськ	59,0	145,4	160,9	145,0	39,9	200,1	197,4	278,0
	Львів	113,7	67,2	82,2	89,8	76,4	134,5	165,2	147,0
	Київ	88,4	123,0	167,6	123,7	45,0	176,8	259,7	188,2
Гіпертонічна хвороба	Полтава	13,9	24,7	25,9	21,6	29,0	34,9	32,0	39,7
	Дніпро	78,4	67,5	79,0	84,1	112,7	67,2	95,0	98,6
	Лисичанськ	45,9	48,0	54,3	67,9	46,1	13,9	87,0	82,1
	Львів	32,1	37,4	42,0	46,0	56,0	78,2	72,1	75,4
	Київ	12,6	22,5	25,0	20,0	24,4	27,0	32,4	30,0

## Закінчення таблиці

1	2	3	4	5	6	7	8	9	10
Судинні ураження мозку	Полтава	23,3	42,4	49,0	42,2	43,1	49,3	76,3	71,6
	Дніпро	48,8	35,0	69,4	55,6	60,3	124,3	140,8	196,4
	Лисичанськ	36,9	91,0	100,6	90,6	24,9	125,1	123,4	173,8
	Львів	71,1	42,0	51,4	56,1	47,8	84,1	103,3	91,9
	Київ	55,3	76,9	104,8	77,3	28,1	110,5	162,3	117,6

**Варіант 4**

У таблиці наведені дані вимірювання швидкості кровотоку, які проводилися двома методами у дітей різних вікових груп. Визначити, чи впливає вік і метод вимірювання на швидкість кровотоку.

Вік дітей	Метод вимірювання	Швидкість кровотоку								
		11	9	7	5	6	7	8	4	8
3–6 міс.	перший	11	9	7	5	6	7	8	4	8
	другий	13	12	9	13	11	9	8	10	13
6–9 міс.	перший	6	7	9	11	12	7	5	9	8
	другий	8	7	11	10	11	9	11	10	11
9–12 міс.	перший	6	5	7	9	5	11	6	7	11
	другий	8	7	8	8	9	13	7	7	13

**Варіант 5**

Вивчали вміст пігменту фікоеритрину (проби відбирали в різних прибережних містах Азовського моря) протягом року. Результати наведені в таблиці.

Місяць року	Дослідна ділянка			
	Маріуполь	Новоазовськ	Мелекіне	Бердянськ
I	0,004	0,0318	0,0241	0,0119
II	0,104	0,0270	0,0034	0,0077
III	0,1056	0,0062	0,0090	0,0022
IV	0,132	0,0135	0,0047	0,0046
V	0,056	0,0038	0,0025	0,0028
VI	0,052	0,0032	0,0012	0,0029
VII	0,100	0,0031	0,0004	0,0039
VIII	0,052	0,0047	0,0012	0,0029
IX	0,049	0,0035	0,0011	0,0041
X	0,012	0,0034	0,001	0,0036
XI	0,004	0,0038	0,001	0,0032
XII	0,004	0,0022	0,004	0,004

Визначити, чи залежить вміст досліджуваного пігменту від часу відбору проб і місцезнаходження досліджуваної ділянки.

### Варіант 6

Випробовували вплив трьох видів мікроелементів на жирномолочність корів. Експеримент проводили на чотирьох групах одновікових тварин різних порід. Кожне випробування мало шестикратну повторність. Отримані результати наведені в таблиці, де А – фактор впливу, тобто мікроелементи, а В – групи корів різних порід. Необхідно з'ясувати випадкові або достовірні відмінності, що спостерігаються між груповими середніми.

Групи корів	Відсоток жиру в молоці		
	$A_1$	$A_2$	$A_3$
$B_1$	2,1 2,0 3,4	2,8 2,6 3,0	2,4 2,1 2,8
$B_2$	4,0 3,2 4,1	3,9 4,1 4,5	3,0 3,9 4,2
$B_3$	3,0 2,8 2,7	3,5 4,0 2,8	4,8 3,1 2,9
$B_4$	3,4 3,0 2,9	3,0 2,9 3,0	3,3 2,8 3,0

### Варіант 7

В одному з дослідних господарств мали врожайність різних сортів агрусу та їх стійкість проти медоносної дії агрусового пильщика. Отримані результати наведені в таблиці. Встановити достовірні або випадкові відмінності, що спостерігаються між груповими середніми.

Зібрано плодів з окремих кущів агрусу, кг	Сорти агрусу А	
	Англійський жовтий ( $A_1$ )	Малахітовий ( $A_2$ )
Не пошкоджених пильщиком ( $B_1$ )	6,3 6,1 5,3 4,9 5,2 4,2 3,8	4,4 4,3 3,6 4,3 3,5 3,9 4,0
Пошкоджених пильщиком ( $B_2$ )	5,2 4,3 3,5 4,5 5,3 4,0	3,9 3,2 4,3 3,0 3,8 2,7

### Варіант 8

Вивчали дію соку і парів часнику, цибулі та перцю на загоєння гнійних ран. Дослідження проводили в одновіковій групі піддослідних тварин. Ефект оцінювали в умовних одиницях. Результати дослідження наведені в таблиці. Встановити, чи є лікувальний ефект від використання препаратів цибулі, часнику і перцю під час лікування гнійних ран.

Спосіб впливу	Часник	Цибуля	Перець
Сік	7 8 6 7	7 6 5 2	3 2 6 5
Пари	5 4 6 5	5 3 6 4	2 4 3 3

### Варіант 9

Кількість пророслого насіння на різних ділянках із різним вмістом нітратів у ґрунті наведено в таблиці. Встановити, чи впливають місцезростання ділянок та вміст нітратів у ґрунті на кількість пророслого насіння.



## Варіанти завдань для закріплення теми 9 «Регресійний аналіз»

### Варіант 1

Для розробки стохастичної моделі забруднення атмосферного повітря озonom використовували вихідні дані за місяцями 2018 року, представлені в таблиці, де  $C_{\text{макс}}$  – максимальна концентрація озону ( $\text{мкг}/\text{м}^3$ );  $C_{\text{NO}}$  – концентрація оксиду азоту ( $\text{мкг}/\text{м}^3$ ),  $P$  – атмосферний тиск (гПа);  $F$  – відносна вологість повітря (%).

$C_{\text{макс}}$	$C_{\text{NO}}$	$P$	$F$
33,23	5,69	1005	77
38,08	8,75	1006	80
51,09	6,16	1009	89
52,71	5,83	1012	92
52,76	15,3	1012	90
54,43	10,96	1013	98
63,42	8,47	1017	96
65,03	8,41	1020	97
69,98	12,85	1021	98
70,57	12,91	1022	98
74,48	13,4	1024	95
75,65	13,99	1024	101

Використовуючи багатофакторний лінійний регресійний аналіз, показати залежність між концентрацією озону і факторами-предикторами. Спрогнозувати рівень забруднення атмосферного повітря озonom на наступний рік.

### Варіант 2

Визначити залежність частоти захворюваності жителів міста на бронхіальну астму від якості повітря. Очевидно, чим гірше повітря, наприклад, вища концентрація чадного газу ( $C_{\text{CO}}$ ,  $\text{мг} / \text{м}^3$ ) і пилу (СП,  $\text{мг} / \text{м}^3$ ) в повітрі, тим більше хронічних хворих ( $P$ ) на тисячу жителів. Статистичні дані є усередненими і наближеними, їх збирають у різних містах упродовж тривалого часу.

$C_{\text{CO}}$	СП	$P$
2,0	45	19
2,5	47	20
2,9	51	32
3,2	55	34
3,6	59	51
3,9	63	55
4,2	67	90
4,6	74	108
5,0	79	171

Спрогнозувати рівень захворюваності на бронхіальну астму при підвищенні концентрації досліджуваних забруднюючих речовин у повітрі.

### Варіант 3

Для прогнозування концентрації  $\text{NO}_2$  (С) використовувався метод множинної лінійної регресії. Для отримання регресійної моделі були використані Т – температура, °С; F – вологість; V – швидкість вітру, м/с; P – кількість опадів, мм.

С	Т	F	V	P
70,57	20	101	1	730
33,23	8	77	2	240
65,03	18	92	3	680
69,98	15	97	2	670
51,09	11	89	1	340
52,71	12	92	4	320
52,76	12	90	3	460
74,48	18	102	4	840
38,08	9	80	4	290
54,43	12	94	2	520
63,42	15	95	0	590
75,65	11	105	2	830

Вкажіть найбільш значущі фактори, які впливають на концентрацію двоокису азоту в повітрі. Спрогнозуйте концентрацію забруднюючої речовини при підвищенні кількості опадів на 100 одиниць щодо середнього значення цього показника.

### Варіант 4

Скласти лінійну регресійну модель залежності крупноплідності потомства від живої маси і довжини тулуба свиноматок.

№ групи	Жива маса, кг	Довжина тулуба, см	Крупно-плідність, кг
1	95	99,26	0,92
2	97,33	112,13	0,96
3	99,56	130,20	0,97
4	110,40	99	0,95
5	110,40	112,16	0,98
6	111	130,20	1,02
7	128	101,23	1,12
8	128,53	112,17	1,18

Спрогнозуйте розмір плода при живій масі – 140 кг, довжині тулуба – 100 см.

**Варіант 5**

Для розробки стохастичної моделі залежності концентрації NO в атмосфері від широти, довготи і температури повітря використовували вихідні дані, представлені в таблиці.

Широта	Довгота	Температура, °C	$C_{NO}$ , мкг/м <sup>3</sup>
46	36	35	70,57
48	32	28	33,23
52	32	24	65,03
53	30	18	69,98
56	35	16	51,09
58	37	14	52,71
60	38	11	52,76
61	40	9	74,48
66	40	4	38,08

Використовуючи метод множинної лінійної регресії, побудуйте модель вмісту забруднюючої речовини від запропонованих чинників. Спрогнозуйте концентрацію NO при довготі – 30, за температури – 35 °C.

### Варіанти завдань для закріплення теми 10 «Кластерний аналіз»

Визначити за допомогою дендрограми та методом k-середніх групи країн, подібних за показниками забрудненості довкілля, за вказаними ознаками: X1 – рівень розораності території держави, %; X2 – частка природно-заповідного фонду, %; X3 – сумарний показник викидів в атмосферне повітря забруднюючих речовин від стаціонарних джерел, мг/м<sup>3</sup>; X4 – лісистість, %; X5 – фінансування на утилізацію відходів, млн \$; X6 – величина річкового стоку на 1 км<sup>2</sup> площі держави.

Країни	Показники забрудненості					
	X1	X2	X3	X4	X5	X6
Азербайджан	20	4,1	12,4	7,9	52	141
Вірменія	20	3,7	4,3	6,5	72	134
Білорусь	72	3,6	28	5,4	38	120
Великобританія	67	3,5	39	8,8	91	91
Німеччина	78	6,8	35	8,1	138	73
Грузія	21	3,8	36	9,8	55	140
Канада	38	3,1	44	7,4	123	77
Польща	55	3,9	30	5	28	124
Україна	46	4,1	23,5	6,7	20	134

Для виконання кластерного аналізу використовуйте ознаки, наведені для вашого варіанту:

Варіант 1 – за ознаками X1, X2, X3

Варіант 2 – за ознаками X4, X5, X6

Варіант 3 – за ознаками X1, X2, X4

Варіант 4 – за ознаками X1, X3, X5

Варіант 5 – за ознаками X1, X4, X6

Варіант 6 – за ознаками X2, X3, X5

Варіант 7 – за ознаками X2, X4, X6

Варіант 8 – за ознаками X3, X4, X5

Варіант 9 – за ознаками X3, X5, X6

Варіант 10 – за ознаками X2, X3, X4

### Варіанти завдань для закріплення теми 11 «Дискримінантний аналіз»

Файл *pollution.sta* містить показники стану водних об'єктів, розташованих у різних регіонах України: X1 – температура води, °C, X2 – прозорість, см; X3 – водневий показник (рН); X4 – вміст розчиненого кисню, мг O<sub>2</sub>/дм<sup>3</sup>; X5 – біологічне споживання кисню (БСК5), мг O<sub>2</sub>/ дм<sup>3</sup>; X6 – вміст хлоридів, мг/дм<sup>3</sup>; X7 – вміст заліза загального (Fe заг.), мг/дм<sup>3</sup>; X8 – рівень поверхнево активних речовин аніонних (ПАР), мг/дм<sup>3</sup>; X9 – вміст амонію, мг/дм<sup>3</sup>; X10 – вміст сульфатів, мг/дм<sup>3</sup>; X11 – вміст важких металів, мг/дм<sup>3</sup>; X12 – вміст фенолу, мг/дм<sup>3</sup>; X13 – вміст нафтопродуктів, мг/дм<sup>3</sup>; X14 – вміст нітратів, мг/дм<sup>3</sup>.

Для проведення дискримінантного аналізу із загальної таблиці виберіть ознаки за варіантами:

Варіант 1 – X1, X2, X3, X4, X5, X6;

Варіант 2 – X2, X3, X4, X6, X7, X11;

Варіант 3 – X3, X4, X5, X7, X8, X11;

Варіант 4 – X3, X4, X7, X8, X9, X11;

Варіант 5 – X3, X4, X8, X9, X11, X14;

Варіант 6 – X4, X5, X6, X7, X11, X12;

Варіант 7 – X4, X7, X9, X11, X12, X14;

Варіант 8 – X1, X8, X9, X10, X11, X12;

Варіант 9 – X1, X2, X8, X9, X11, X13;

Варіант 10 – X1, X3, X9, X10, X13, X15.

На основі вибірових даних побудуйте модель дискримінантного аналізу; дайте оцінку якості моделі, значущості змінних та проведіть канонічний аналіз функцій.

### Список рекомендованої літератури

1. Антомонов М. Ю. Математическая обработка и анализ медико-биологических данных. 2-е изд. Киев: МИЦ «Мединформ», 2018. 579 с.
2. Еріна А. М. Статистичне моделювання і прогнозування. Київ: КНЕУ, 2001. 170 с.
3. Лаврик В. І. Методи математичного моделювання в екології. Київ: Фітосоціоцентр, 1998. 131 с.
4. Лапач С. Н., Чубенко А. В., Бабич П. Н. Статистичні методи в медико-біологічних дослідженнях з використанням Excel. Київ: Моріон, 2001. 408 с.
5. Тарасова В. В., Парфенцева Н. О., Ковалевська І. М. Екологічна статистика. Теоретичні основи та лабораторний практикум на базі комп'ютерних технологій в системі Excel: підруч. 2-ге вид. з грифом МОНУ. Київ: Центр навчальної літератури, 2013. 295 с.
6. Тарасова В. В. Екологічна статистика: підруч. з грифом МОНУ. Київ: Центр навчальної літератури, 2008. 392 с.
7. Чекотовський Е. В. Статистичні методи на основі Microsoft Excel 2016: навч. посіб. Київ: Знання, 2018. 408 с.
8. John H. McDonald Handbook of Biological Statistics / Sparky HousePublishing, 2014. 299 p.
9. Quinn G., Keough M. Experimental Design and Data Analysis for Biologists / Cambridge University Press, 2002. 557 p.

Навчальне видання

*Тарадіна Галина Вікторівна*  
*Доценко Ольга Іванівна*  
*Мищенко Артем Михайлович*

## **СТАТИСТИЧНІ МЕТОДИ В БІОЛОГІЇ ТА ЕКОЛОГІЇ: ТЕОРІЯ І ПРАКТИКА**

Навчальний посібник

Редактор

І. М. Колесникова

Технічний редактор

Т. О. Важеніна-Гопрак

Підписано до друку 21.04.2022

Формат 60 x 84/16. Папір офсетний. Друк цифровий.  
Умовн. друк. арк. 10,46 Тираж 100 прим. Зам. 1935/1

Донецький національний університет імені Василя Стуса  
21021, м. Вінниця, 600-річчя, 21  
Свідоцтво про внесення суб'єкта видавничої справи  
до Державного реєстру  
серія ДК № 5945 від 15.01.2018

Віддруковано з оригіналів замовника.

Видавець та виготовлювач ТОВ «ТВОРИ».  
Свідоцтво про внесення суб'єкта видавничої справи  
до Державного реєстру видавців, виготовлювачів і розповсюджувачів  
видавничої продукції серія ДК № 6188 від 18.05.2018 р.  
21034, м. Вінниця, вул. Немирівське шосе, 62а.  
Тел.: 0 (800) 33-00-90, (096) 97-30-934, (093) 89-13-852, (098) 46-98-043.  
e-mail: [info@tvoru.com.ua](mailto:info@tvoru.com.ua)  
<http://www.tvoru.com.ua>