МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

С.М. Лапач

ТЕОРІЯ ПЛАНУВАННЯ ЕКСПЕРИМЕНТІВ

ЛАБОРАТОРНИЙ ПРАКТИКУМ

*Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для студентів,
які навчаються за спеціальністю 131* «Прикладна механіка»,
спеціалізацією «Технологія машинобудування»

Київ
КПІ ім. Ігоря Сікорського
2020

Рецензенти: Радченко С. Г., док. техн. наук, проф.
Кореньков В. М., канд. техн. наук, доц.

Відповідальний редактор *Лашина Ю.В.* , канд. техн. наук, доц.

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського (протокол № 4 від 10.12.2020 р.)
за поданням Вченої ради Механіко-машинобудівного інституту (протокол № 4 від 23.11. 2020 р.)

Електронне мережне навчальне видання

Лапач Сергій Миколайович

ТЕОРІЯ ПЛАНУВАННЯ ЕКСПЕРИМЕНТІВ

ЛАБОРАТОРНИЙ ПРАКТИКУМ

Теорія планування експериментів: Лабораторний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 131 «Прикладна механіка», спеціалізації «Технологія машинобудування» / С.М. Лапач ; КПІ ім. Ігоря Сікорського. – Електронні текстові данні (1 файл: 2,14 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2020. – 125 с.

В посібнику описується виконання лабораторних робіт за курсом «Теорія планування експериментів», які включають завдання з першої частини курсу, саме математичної статистики. Вони включають наступні теми: числові характеристики випадкових величин; перевірка гіпотез щодо параметрів положення і розсіяння, включаючи множинні перевірки; перевірка наявності зв'язку між випадковими змінними (кореляційний та дисперсійний аналіз і аналіз таблиць часток та пропорцій) ; розв'язання задач класифікації (кластер ний аналіз, включаючи нечіткий, дискримінант ний і покроковий дискримінантний аналіз). Завдання побудовані таким чином, що студент в рамках теми повинен сам вибрати метод обробки даних, який забезпечує виконання завдання.

Зміст

Вступ

В посібнику містяться рекомендації по виконанню лабораторних робіт з першого модулю («Математична статистика».) курсу «Теорія планування експериментів». У вказівках виклад теоретичних положень дуже короткий і не достатній в загальному випадку для повноцінної підготовки до захисту. Це зроблено для спрощення роботи з матеріалом. Необхідні для підготовки матеріали студент може знайти в електронному конспекті лекцій з даного курсу та в спеціальній літературі, список якої надається в навчальному плані (для деяких існують електронні варіанти).

Матеріали для підготовки до лабораторної роботи.

1. Приклади розрахунку: Додаток А1 даних методичних вказівок.
2. Теоретичні відомості: Конспект лекцій по курсу (на сайті кафедри і сервері кафедри)
3. Теоретичні відомості і приклади розв’язання задач: книга Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистические методы в медико-биологических исследованиях с использованием Excel –2 изд. перераб. и доп. –К.: 2001, Морион. – 408с. в електронному вигляді (на сайті кафедри)

Лабораторна робота № 1.

Числові характеристики випадкової величини.

Мета.

Ознайомитися з основними числовими характеристиками випадкової величини та їх властивостями. Поняття про вибірковий метод і зв'язок емпіричних характеристик з теоретичними.

Завдання.

- Виконати розрахунки основних статистичних характеристик випадкової величини: показники положення (середні, медіана, мода), форми (асиметрія, ексцес), розсіяння (розмах, дисперсія, середньоквадратичне відхилення), квартілі, гістограма, закон розподілу). Розрахунки положення додатково виконати також з включенням до даних "викидів".
- Побудувати за допомогою вбудованої функції криві розподілу чи криві густини ймовірності однієї функції розподілу для кількох значень параметрів.
- Визначити необхідний розмір вибірки при заданих умовах.

Склад звіту: завдання, вихідні дані, результати розрахунків.

Короткі теоретичні відомості.

До характеристик одновимірного розподілу відносяться:

Міри положення (середнє, медіана, мода тощо).

Міри розсіяння (розмах, коефіцієнт варіації, дисперсія, середньоквадратичне відхилення).

Міри форми (асиметрія, ексцес, моменти третього та четвертого порядку).

Властивості середнього (вибіркового)

1. Сума відхилень від середнього дорівнює 0.
2. Якщо всі значення вибірки збільшити чи зменшити, помножити чи поділити на одне і теж число, то середнє значення зміниться аналогічно.
3. Зі збільшенням числа вимірювань точність оцінки збільшується і вона наближається до відповідного значення генеральної сукупності, але тільки в тому випадку, якщо немає систематичних похибок і виміри незалежні.
4. Середнє суми двох вибірок дорівнює сумі їх середніх, якщо вибірки однакових розмірів (аналогічно для різниці). $\overline{X+Y} = \overline{X} + \overline{Y}$

Декілька несподіваних зауважень

Середнє зовсім ***не є типовим***. Наприклад, середній прибуток ні в якому разі не є типовим для більшості країн.

Середнє ***не співпадає з математичним сподіванням***. За винятком випадку нормального розподілу арифметичне середнє навіть ***не є незміщеною оцінкою математичного сподівання з найменшою дисперсією***. Більш того, навіть для нормального розподілу можна вказати оцінку, яка буде ближче до математичного сподівання, правда без деяких корисних властивостей середнього.

Формули розрахунку різних видів середніх значень приведені в табл..1.1.

Таблиця 1.1. Формули розрахунку середнього значення

Характеристика	Обчислення	Обчислення для згрупованих даних[1]
Середнє арифметичне	$\overline{X} = \frac{\sum_{i=1}^{N} x_i}{N}$	$\overline{X} = \sum_{i=1}^{k} X_i n_i / \sum_{i=1}^{k} n_i$
Середнє геометричне	$\overline{X}_G = \sqrt[n]{\prod_{i=1}^{n} X_i}$	–
Середнє гармонійне	$\overline{X}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$	–
Медіана	$\begin{cases} Me = x_{(N+1)/2}, N - \text{непарне} \\ Me = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}, N - \text{парне} \end{cases}$	$Me = X_{Me} + h\,(\Sigma m_X/2 - m^{max}_X)/\,m_m$
Мода		$Mo = X_{Mo} + h\,(m_{Mo} - m_{Mo-1})/(2m_{Mo} - m_{Mo+1} - m_{Mo-1})$,

Тут X_{Mo} — початок модального інтервалу, тобто такого, котрому відповідає найбільша частота; h — величина модального інтервалу; m_{Mo} — частота модального інтервалу; m_{Mo-1} — частота інтервалу, що передує модальному; m_{Mo+1} — частота інтервалу, наступного за модальним.

X_{Me} — початок медіанного інтервалу; h — величина медіанного інтервалу; m_X — частоти по всім інтервалам; m^{max}_X — частота, накопичена до початку медіанного інтервалу; m_m — частота медіанного інтервалу. Медіанним називається інтервал, в якому міститься значення медіани.

[1] Ці формули використовують у тих випадках, коли ми маємо справу зі згрупованими даними, наприклад, інтервали доходу (від і до) та доля населення з такими доходами. Тоді середина інтервалу вважається середнім і ми можемо по вказаній формулі виконувати розрахунок середнього для всієї вибірки.

Середнє геометричне (вибіркове)

Геометричне середнє використовується в тих випадках, якщо:

- величина змінюється в часі з постійним співвідношенням між її вимірами $\frac{X_{i-1}}{X_i} = \frac{X_{i+1}}{X_i} = const$ (наприклад, ріст числа бактерій, ріст капіталу на рахунку, експлуатаційні витрати тощо);
- окремі значення в варіаційному ряду знаходяться дуже далеко одне від іншого (наприклад, різниця на порядок між значеннями).

Середнє гармонічне

В деяких випадках, наприклад при розрахунку середньої тривалості життя, визначення середньої швидкості при русі з різними швидкостями та інших подібних ситуаціях використовується середнє гармонічне.

Мода

Мода — це значення, котре спостерігається найбільше число разів (найбільш ймовірна величина).

Пам`ятайте, що мода не може використовуватися у випадку, коли закон розподілу мультимодальний (багатовершинний).

Медіана (вибіркова)

Медіана — це значення, котре ділить ранжований варіаційний ряд на дві рівні по обсягу групи.

Властивості медіани

- Сума абсолютних величин відхилень варіантів від медіани, помножених на відповідні частоти є мінімальна, тобто менше, ніж від будь якої іншої величини[2]

$$\sum_{\forall x}|x - Me|\, m_x < \sum_{\forall x}|x - a|\, m_x.$$

[2] Ця властивість медіани може використовуватись, наприклад, для вибору місця зупинки міського транспорту, бензоколонок тощо.

- На значення медіани не впливають зміни крайніх значень варіаційного ряду, якщо тільки менше медіани залишається меншим, а більше продовжує залишатся більшим[3].

Емпірична дисперсія

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

Властивості дисперсії

- Дисперсія постійної величини дорівнює 0.
- Якщо всі результати збільшити або зменшити на одне і то ж число, то дисперсія не зміниться.
- Дисперсія суми і різниці випадкових величин дорівнює сумі дисперсій.
- Якщо всі результати змінити в К разів то дисперсія зміниться в $К^2$ раз[4].

Корінь квадратний з дисперсії це середньоквадратичне відхилення — S. В англомовній літературі цей термін прийнято називати стандартною помилкою.

Деякі зауваження відносно питання N чи N-1

Одне з найбільш частих питань: на що ділити суму квадратів в формулі дисперсії, на N чи на N-1?

В загальному випадку, при великому N різниця між цими оцінками невелика, але все ж може мати значення. В тому випадку, коли математичне сподівання відоме, незміщена оцінка дисперсії розраховується по формулі $S^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - M(X))^2$

Але сума квадратів в чисельнику мінімальна тільки в тому випадку, коли математичне сподівання дорівнює середньому, що взагалі то не так (дивися зауваження до математичного сподівання). В зв`язку з цим необхідна поправка Бесселя, котра приводить після ряду перетворень до появи в знаменнику N-1.

В практичній діяльності часто користуються наступним правилом. Якщо математичне сподівання відоме з теоретичних положень або його оцінка отримана по іншій виборці — використовують формулу з N. В тому випадку, коли оцінка середнього отримана по тій самій виборці, по якій рахується оцінка дисперсії, то використовують формулу з N-1.

---

[3] В зв`язку з цією властивістю, медіану використовують замість середнього в тих випадках, коли крайні значення сильно відрізняються від інших.

[4] Відоме явище, коли при інфляції бідні стають біднішими, а багаті ще багатшими, з точки зору статистики є лише властивістю дисперсії.

Довірчий інтервал

Довірчий інтервал — інтервал, відносно якого з наперед заданою ймовірністю $P = 1 - \alpha$ можна стверджувати, що він містить невідоме значення параметра θ.

$$P[\theta_1 < \theta < \theta_2] = 1 - \alpha$$

При цьому $1 - \alpha$ — довірча ймовірність, α — рівень значущості.

Властивості

- При збільшенні кількості спостережень точність збільшується (тобто інтервал зменшується). Але це має місце тільки в тому випадку, коли немає систематичних похибок і спостереження незалежні.
- Збільшення надійності при фіксованій виборці веде до збільшення довірчого інтервалу та падіння точності.

Довірчий інтервал означає не ймовірність попадання значення оцінюваного параметру всередині певного інтервалу, а то, що якщо ми візьмемо достатню кількість вибірок, то в $100 \times p$ % випадків параметр буде знаходитися в заданому інтервалі.

Рівень значущості при розв`язання практичних задач звичайно приймається в інтервалі від 0,01 до 0,05.

Довірчий інтервал для середнього

Визначається за формулою $\left[\overline{X} - t_{n,p}\frac{S}{\sqrt{n}}, \overline{X} + t_{n,p}\frac{S}{\sqrt{n}}\right]$, де S — середнє квадратичне відхилення, n — число дослідів, $t_{n,p}$ — табличне значення розподілу Стюдента с числом степенів свободи n та довірчою ймовірністю p.

Приведена формула використовується в тому випадку, коли дисперсія невідома і використовується її оцінка по експериментальних даних. Для випадку відомої дисперсії існує інша формула, яка не приводиться в зв'язку з рідкістю такої ситуації.

Довірчий інтервал для середнього квадратичного відхилення

$$P[\sqrt{n}Sc_1 < s < \sqrt{n}Sc_2] < 1 - \alpha,$$

де χ_1 и χ_2 квантилі розподілу χ^2 (хі-квадрат). Квантилі χ_1 и χ_2 мають по (n-1) степенів свободи та рівні значущості $1 - \alpha/2$ та $\alpha/2$ відповідно.

Зауваження.

1. В тих випадках, коли по виборці розраховуються декілька параметрів, довірчі інтервали насправді будуть більше, чим розраховані. Це зв`язано з тим, що статистики, які використовуються для побудови довірчих інтервалів параметрів, наприклад середнього та дисперсії не є незалежними.

2. Формули для довірчих інтервалів отримані в припущенні нормального розподілу.

Одночасне визначення довірчих інтервалів для кількох параметрів

При одночасному оцінюванні середнього та дисперсії нас цікавить ймовірність того, що обидва параметра знаходяться в заданих інтервалах.
Довірча область для математичного сподівання та дисперсії має форму трапеції

$$\Delta_\alpha(\hat{\Theta}) = \left\{ (m_x, D_x) : \left| m_x - \overline{X} \right| < \varepsilon_\alpha \sqrt{\frac{D_x}{n}}, (1-\delta_\alpha)\sqrt{D_x} < \sqrt{D_x} < (1+\delta_\alpha)\sqrt{D_x} \right\}$$

$$P((\overline{X}, D_x) \in D_\alpha(\theta\theta = P(\left| \overline{X} - m_x \right| < \varepsilon_\alpha \sqrt{\frac{D_x}{n}}) \times P(\frac{n-1}{(1+\delta_\alpha)^2} < Z < \frac{n-1}{(1-\delta_\alpha)_+^2}) = 2\Phi(\varepsilon_\alpha) L_{n-1}(\delta_\alpha)$$

$\Phi(\varepsilon_\alpha) = c\frac{\sqrt{\alpha}}{2}; L_{n-1}(\delta_\alpha) = \frac{\sqrt{\alpha}}{c}$, де с – довільно вибране число

Приклад.

Вихідні дані: X={5, 42, 22, –40, 18, 38, 2, –16, 34, 6, 54, 20, 74, 0, 4, –28, 36, 44, 16, 24}
Середнє 17,9; Дисперсія 756,83;Середнє квадратичне відхилення 27,51.

Розрахунок за стандартними формулами дає $17{,}0 < \sqrt{D_x} < 38{,}0$ і $5{,}4 < m_x < 30{,}4$.

Для області складної форми $\Phi(\varepsilon_\alpha) = \frac{\sqrt{0{,}95}}{2}; L_{n-1}(\delta_\alpha) = \sqrt{0{,}95}$; $\Phi(\varepsilon_\alpha) = \frac{\sqrt{0{,}9747}}{2};$

Звідси ε_α=2,24/√20=2,24/4,4721=0,501 і для математичного сподівання залежність приймає вид $17{,}9 - 0{,}501\sqrt{D_x} < m_x < 17{,}9 + 0{,}501\sqrt{D_x}$

Для дисперсії $L_k(\delta_\alpha) = \int\limits_{k(1+\delta_\alpha)^2}^{k(1-\delta_\alpha)_+^2} p_k(z)dz$, де $(1-\varepsilon_\alpha)_+ = \max(1-\varepsilon_\alpha; 0)$. Таблиці 4 δ_α= 0,464. Звідси

$$14{,}7 < \sqrt{D_x} < 40{,}3$$

В зв'язку зі складністю побудови області не прямокутної області, часто шукають прямокутну область, яка б відповідала поставленій умові.

Прямокутна довірча область для кількох параметрів одночасно відповідає довірчій ймовірності, яка визначається за формулою: 1– (1–p)/r.
Для нашого прикладу 1– (1–0,95)/2=0,975.

$14{,}5 < \sqrt{D_x} < 40{,}5$ і $4{,}1 < m_x < 31{,}7$

Контрольні питання.

- Що таке випадкова величина?
- Чим відрізняються випадкові числа від нечітких і інтервальних?

- Що таке закон розподілу та функція густини ймовірності?
- Властивості середнього, медіани та дисперсії.
- Які характеристики описують закон розподілу випадкової величини?
- Для чого використовуються числові характеристики випадкової величини?
- Міри положення і міри розсіяння.
- Чому для характеристики закону розподілу недостатньо показників положення?
- Для яких типів законів розподілу використовується мода?
- Порівняйте стійкість і інформаційну якість середнього та медіани при різних законах розподілу та наявності “викидів”.
- Що таке вибірка і генеральна сукупність.
- Чому всі емпіричні значення в статистиці є тільки оцінками?

Варіанти завдань

Дані для розрахунків вибираються відповідно до завдань по таблиці А.1–А.9 з таблиць Додатку А.

Таблиця 1.2. Варіанти завдань по розрахунку показників

№ варіанту	Завдання	№ варіанту	Завдання	№ варіанту	Завдання
1	X3, X4, X5; Біноміальний	10	X46, X47; Бета	19	X33, X34, X35; χ^2
2	X6. X7, X8; Геометричний	11	X19, X20; Гамма	20	X19, X20; Гауса
3	X10, X11, X12; Гіпергеометричний	12	X55, X56, X57; Показниковий	21	X36, X37, X38; Пуассона
4	X13, X14, X15; Паска ля	13	X26, X36; Логнормальний	22	X39, X48; Паскаля
5	X16, X17, X18; Пуассона	14	X59, X60, X61; Фишера	23	X40, X41, X42; Гіпергеометричн ий
6	X30, X31, X32; Гауса	15	X3, X4, X5; Гауса	24	X43, X44, X45; Геометричний
7	X33, X34, X35; Стьюдента	16	X6. X7, X8; Пуассона	25	X16, X17, X18; Біноміальний
8	X19, X20; Фишера	17	X10, X11, X12; Паска ля	26	X30, X31, X32; Бета
9	X36, X37, X38; χ^2	18	X13, X14, X15; Гіпергеометричн ий	27	X33, X34, X35; Гамма

Лабораторна робота № 2.

Порівняння середніх та дисперсій. Параметричні та непараметричні критерії.

Мета.

Навчитися використовувати критерії по перевірці гіпотез про параметри положення та розсіяння.

Завдання.

- По заданим варіантам провести перевірку гіпотези про рівність параметрів положення та розсіяння.

Склад звіту. Завдання, вихідні дані, результати розрахунків з висновками на кожному криці виконання.

2.1. Параметричні критерії

Це гіпотези, які відносяться до виду чи значення окремих параметрів розподілу випадкової величини. Нехай $f(X, \theta)$ — закон розподілу випадкової величини X з деяким параметром θ. Тоді:

H_0 (нульова гіпотеза): $\theta = \theta_0$,

H_1 (альтернативна або конкуруюча гіпотеза) $\theta = \theta_1$.

Помилка першого роду: — H_0 відкидається, коли вона істинна.

Помилка другого роду: — H_0 приймається, коли істинна H_1.

Будь-які гіпотези перевіряють, висуваючи спочатку комплекс певних допущень про закон розподілу випадкової величини. Невиконання цих допущень робить висновки із перевірок по гіпотезі некоректними.

Нульова гіпотеза відхиляється в тому випадку, коли ймовірність того, шо вона вірна виявляється нижче деякого рівня, який називається рівнем значущості.

Таблиця 2.1. Можливі ситуації при перевірці статистичних гіпотез

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Фактична ситуація</td></tr>
<tr><td>H_0 –вірна</td><td>H_0 – невірна</td></tr>
<tr><td rowspan="2">Дії перевіряючого</td><td>Відкинути гіпотезу H_0</td><td>α</td><td>$1–\beta$</td></tr>
<tr><td>Прийняти гіпотезу H_0</td><td>$1–\alpha$</td><td>β</td></tr>
</table>

Тут α – ймовірність помилки першого роду або рівень значущості; $(1–\alpha)$ – довірча ймовірність; β – ймовірність помилки другого роду; $(1–\beta)$ – потужність критерію.

В технічних дослідженнях, як правило потужність критерію не розраховують. Можливе його значення враховують при виборі критерію. Це зв’язано з значними складностями при визначенні потужності в реальних дослідженнях. На його значення впливає велика кількість факторів і їх необхідно надати певні значення. Ситуація ускладнюється тим, що частина факторів залежить одна від одного.

На величину потужності впливають α, β, кількість експериментів, варіабельність, виконання передумов та допущень критерію (відповідність прийнятої гіпотези реальному стану справ), фактичне значення параметру, який перевіряється..

Зауваження.

Спочатку необхідно визначити, який клас методів (параметричний чи непараметричний) необхідно вибрати. Потім уточнити, якщо необхідно, варіант методу. Всі розрахунки, необхідні для вказаних виборів з відповідними висновками, повинні бути приведені в звіті. Для вибору критерію скористайтесь табл..2.2.

Непараметричні критерії вибираються в тому випадку, коли шкали вимірювань нечислова (порядкова чи найменувань), або коли закон розподілу відмінний від нормального.

Перевірити чи є закон розподілу нормальним можна за допомогою простої нестрогої перевірки. Якщо виконується умова

$$\left|\frac{D_{abs}}{S} - 0{,}7979\right| < \frac{0{,}4}{\sqrt{N}}, \text{ де } D_{abs} = \frac{\sum\left|X_i - \overline{X}\right|}{N}.$$

то закон розподілу можна вважати нормальним.

Для випадку нормального в залежності від деяких вихідних умов можливі наступні варіанти розрахунку емпіричного критеріального значення при перевірці гіпотези про рівність середніх (таблиця 2.3).

Якщо $t < t_{p,v}$, де p — довірча ймовірність, а v — число степенів свободи, гіпотеза про рівність середніх приймається.

Таблиця 2.2. Вибір критерію для перевірки статистичної гіпотези про міри розсіяння та положення

Про яку міру перевіряється гіпотеза?	Кількість вибірок	Припущення чи гіпотеза	Закон розподілу	Додаткова умова	Критерій
Про положення	Одна	Середнє дорівнює А	N	–	Стюдента
			~~N~~	–	Знаковий, Гупта
		Середнє до і після експерименту	N	–	Стюдента
			~~N~~	–	Знаковий, Одно вибірковий Уілкоксона
	Дві	Дисперсії однакові	N	–	Стюдента
			~~N~~		Манна-Уітні
		Дисперсії не однакові	N	–	Стюдента
			~~N~~	–	–
		Без припущень про дисперсії	N	–	Стюдента
			~~N~~	–	Медіанний, двовибірковий Уілкоксона
	Три і більше	–	N	–	Шефе, LSD
			~~N~~		Рангових сум Фрідмана
Про розсіяння	Дві	–	N	–	Фішера
			~~N~~	–	Зігеля-Тюкі
	Три і більше	–	N	Вибірки одного розміру	Кохрена
				Вибірки різного розміру	Бартлета
			~~N~~	–	Краскела-Уолліса

Таблиця 2.3. Формули для розрахунку критеріального значення критерію Стюдента

Кількість вибірок	Гіпотеза чи припущення	Формула	Число степенів свободи
Одна	Середнє дорівнює А	$t = \frac{(\bar{X} - A)\sqrt{N}}{S^2}$	V = N — 1
	Середнє до і після експерименту	$t = \frac{\frac{\sum_{i=1}^{N}(x_i - y_i)}{N}}{\sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2 - (\sum_{i=1}^{N}(x_i - y_i))^2 n}{N(N-1)}}}$	V = N — 1
Дві	Дисперсії однакові	$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{(1N_1 + 1N_2)(S_1^2(N_1 - 1) + S_2^2(N_2 - 1))}{(N_1 + N_2 - 2)}}}$	**V = N₁ + N₂ — 2**
	Дисперсії не однакові	$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$	$V = \left(\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}\right)^2 / \frac{\left(\frac{S_1^2}{N_1}\right)^2}{N_1 + 1} + \frac{\left(\frac{S_2^2}{N_2}\right)^2}{N_2 + 1} + 2$
	Без припущень про дисперсії	$t = \frac{(\bar{X}_1 - \bar{X}_2)\sqrt{N}}{\sqrt{\left(\sum_{i=1}^{N}(X_{1i} - X_{2i} - (\bar{X}_1 - \bar{X}_2))^2\right)/(N-1)}}$	V = N — 1

N1, N2 – розміри першої та другої вибірки відповідно (N для однієї чи при рівних розмірах вибірок); $\bar{X}_1$, $\bar{X}_2$ – середнє значення першої та другої вибірки відповідно ($\bar{X}$ для однієї вибірки); S_1^2, S_2^2 – дисперсії першої та другої вибірки відповідно
X_{1i}, X_{2i} – значення елементів першої та другої вибірки відповідно ;x_i, y_i – значення до і після експерименту
При розрахунках вважається, що $\bar{X}_1 > \bar{X}_2$

Перевірка гіпотез про дисперсії

Формули для параметричних критеріїв перевірки гіпотез щодо параметру розсіяння приведені в табл..2.4.

Таблиця 2.4. Параметричні критерії

Кількість вибірок	Додаткові умови	Критерій і формули	Гіпотеза приймається
Дві		Критерій Фішера $F_{расч}=S_1^2/S_2^2$	$F_{расч}<F_{\alpha,N1-1,N2-1}$, $F_{расч}>1$
Три і більше	Всі вибірки однакового розміру	***Критерій Кохрена*** $G=\dfrac{S_{\max}^2}{\sum\limits_{i=1}^{N} S_i^2}$	$G<G_{\alpha,n-1,k(n-1)}$
	Вибірки різного розміру	Критерій Бартлета $c_{kp}^2 = \dfrac{2}{c}(\,2{,}3026(v\ln S^2 - \sum\limits_{i=1}^{k} v_i \ln S_i^2\,))$ де $c=\dfrac{\sum\limits_{i=1}^{k}\frac{1}{v_i}-\frac{1}{v}}{3(k-1)}+1$; $S^2=\dfrac{\sum\limits_{i=1}^{k} v_i S_i^2}{v}$. $v=\sum\limits_{i=1}^{k} v_i$ k — число груп; $\nu_i=n_i-1$ —n_i – розмір і-ї групи); S^2_i — оцінка дисперсії в кожній групі;	

2.2. Непараметричні критерії перевірки різниці положення середніх значень

Вище приведені критерії можуть використовуватися тільки при нормальному законі розподілу. В тих випадках, коли ця умова не виконується використовуються непараметричні критерії.

Зауваження до непараметричних критеріїв

У всіх випадках використання приведених нижче критеріїв приймаються наступні припущення:

- всі випадкові величини взаємно незалежні;
- вибірки, що вивчаються, розподілені по неперервному закону розподілу одного і того ж виду.

Критерій знаків

Використовується в тих випадках, коли вибірки однакового розміру і природнім способом розбиваються на пари. Наприклад, показники одних і тих же хворих до прийому препарату і після, показники росту чи врожайності пар кущів рослин, які ростуть поряд, але при цьому один із пари отримує додаткову підкормку чи обробку.

Фактично перевіряється гіпотеза про рівність нулю медіани різниці між парами двох вибірок.

Розраховуються значення

$$Z_i = \left\{ \frac{1,\text{ если } Y_i - X_i > 0}{0,\text{ если } Y_i - X_i < 0} \right\},$$

де X_i і Y_i значення елементів відповідних вибірок.

Постільки припускається, що спільний розподіл X і У неперервний, то ймовірність рівності значень X та У вважається рівною нулю. В зв'язку з цим, у випадку співпадання значень вони відкидаються і кількість спостережень відповідно зменшується. Обчислюється число m, яке дорівнює кількості Z_i, котрі рівні 1.

Для вибірок малих розмірів можна обчислити рівень значущості (ймовірність відкинути гіпотезу про рівність нулю медіани різниці, якщо вона істинна) по наступній формулі

$$P = \frac{1}{2^n} \sum_{i=0}^{n-m} \frac{n!}{i!(n-i)!},$$

де n — розмір вибірки.

Для великих вибірок можна користуватися приблизною нормальною апроксимацією

$$P = 1 - F\left(\frac{m-1-0{,}5n+0{,}5}{\sqrt{0{,}5n0{,}5}}\right).$$

Якщо отримане значення P мале (менше 0,05), то гіпотеза про рівність нулю відкидається.

В тих випадках, коли розміри вибірок не рівні, або природне розбиття на пари неможливе, то використовується медіанний критерій для двох вибірок або двохвибірковий критерій Уілкоксона.

Критерій Манна-Уітні (U-критерій Уілкоксона, Манна, Уітні)

Цей критерій — непараметричний аналог t-критерію для перевірки середніх.

Цей критерій є самим строгим із непараметричних критеріїв. Він може використовуватися тільки в тому випадку якщо дисперсії вибірок рівні. Перевіряється гіпотеза: дві незалежні вибірки належать одній генеральній сукупності і їх функції розподілу однакові. Ця гіпотеза включає рівність середніх і медіан.

Формується єдина вибірки і для значень цієї об'єднаної вибірки визначаються ранги (номера у впорядкованій по зростанню послідовності). Для співпадаючих значень присвоюється середній ранг.

Обчислюється розрахункове значення критерію.

Для малих вибірок.

$$U_1 = n_1n_2 + 0{,}5\ n_1(n_1+1) - R_1$$

$$U_2 = n_1n_2 + 0{,}5\ n_2(n_2+1) - R_2$$

Правильність розрахунку U_1 і U_2 перевіряється співвідношенням

$$U_1+U_2 = n_1n_2$$

Критеріальне значення визначається наступним чином

$$U = \min(U_1, U_2)$$

Гіпотеза про рівність вибірок відкидається, якщо $U > U(n_1, n_2, \alpha)$, де $U(n_1, n_2, \alpha)$ — критичне значення статистики Манна-Уітні.

Для вибірок, розмір яких достатньо великий ($n_1+n_2>60$) можна скористатися апроксимацією

$$U(n_1,n_2,\alpha) = \frac{n_1n_2}{2}\tilde{z}_\alpha\sqrt{\frac{n_1,n_2(n_1+n_2+1)}{12}},$$

де $\tilde{z}$ відповідає значенню стандартного нормального розподілу. Для випадку $n_1 \geq 8$ и $n_2 \geq 8$ можна скористатися нормальною апроксимацією:

$$\tilde{z} = \frac{(U - \frac{n_1n_2}{12})}{\sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}}.$$

Якщо розраховане значення $\tilde{z}$ більше табличного значення нормального стандартного розподілу взятого з відповідним рівнем значущості, то гіпотеза відкидається.

Медіанний критерій для кількох вибірок

Використовується для перевірки належності К вибірок до однієї генеральної сукупності.

Знаходиться медіана об`єднаної вибірки і після цього формується таблиця, форма якої відповідає табл. 8.

Обчислюється очікуване число спостережень для кожної клітинки (при виконанні нульової гіпотези). Для цього множиться елемент рядка "всього", що відповідає даній клітинці, на елемент стовпчика "всього" для цієї клітинки і ділиться на загальну кількість спостережень n. Потім обчислюється розрахункове значення критерію.

$\chi^2 = \Sigma$ (спостережене значення — очікуване значення)^2 / очікуване значення.

Таблиця 8

	Вибірка 1	Вибірка 2	...	Вибірка К	Всього
Число спостережень, які більші медіани	L_1	L_2	...	L_K	$\sum_{j=1}^{K} L_j$
Число спостережень, які менші 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

Сума шукається по всім 2К клітинам таблиці. Якщо це значення більше верхнього критичного, взятого з вибраним рівнем значущості α і числом степенів свободи (К-1), то гіпотеза про рівність середніх відхиляється.

Двохвибірковий критерій Уілкоксона

Вибірки, що порівнюються з числом елементів n_1 і n_2, об\`єднуються в одну. Для об\`єднаної вибірки будуються ранги. Розраховуються значення сум рангів по першій та другій вибірках R_1 і R_2. Після цього розраховується критеріальне значення $W = R_2$.

Для малих вибірок гіпотеза про рівність середніх відкидається, якщо $W \geq w(n_1, n_2, \alpha_2)$ або $W \leq (n_1 (n_1+n_2+1) - w(n_1, n_2, \alpha_1))$.

Тут $\alpha_1+\alpha_2 = \alpha$, а $w(n_1, n_2, \alpha_2)$ критичне значення розподілу Уілкоксона.

Для великих вибірок W розраховується по наступній формулі

$$W = \frac{2R_2 - n_2(n_1 + n_2 + 1) + 1}{\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{3}}}.$$

Якщо обчислене значення більше значення стандартного нормального розподілу, взятого з відповідним рівнем значущості, то гіпотеза про рівність відхиляється.

Необхідно пам\`ятати, що нормальну апроксимацію можна використовувати тільки в тому випадку, якщо розмір хоча б однієї з вибірок перевищує 25.

В тому випадку, якщо в вибірках є однакові значення, то знаменник вище приведеного виразу розраховується по формулі

$$\sqrt{\frac{n_1 n_2}{12}\left(n_1 + n_2 + 1 - \frac{\sum_{i=1}^{k} t_i (t_i^2 - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)}\right)}.$$

Де k — число груп зв\`язків, t_i — розмір і-й групи (кількість однакових елементів у групі).

Контрольні питання.

- Що таке статистична гіпотеза?
- Що таке рівень значущості?
- Що таке альфа і бета і зв'язок між ними?
- Які ситуації можуть виникнути при перевірці статистичних критеріїв?
- Чи можна порівнювати середні або дисперсії виключно за їх величиною?
- Різниця між параметричними та непараметричними критеріями. В яких випадках які із них використовувати?
- Що таке одно- та двосторонні критерії?
- Послідовність операцій по прийняттю рішень по вибору критерію для порівняння параметрів положення?
- Послідовність операцій по прийняттю рішень по вибору критерію для порівняння параметрів розсіяння?
- Як перевірити, чи різниця між середніми більша за певну величину?

Варіанти завдань.

Дані для розрахунків вибираються відповідно до завдань по таблиці 2.1 з таблиць Додатку А.

Таблиця 2.1. Варіанти завдань по розрахунку показників

№ варіанту	Змінні	№ варіанту	Змінні
1	X3, X4	14	X19, X20
2	X6. X7, X8	15	X55, X56
3	X10, X11	16	X26, X36
4	X13, X14	17	X59, X60
5	X16, X17	18	X56, X57
6	X30, X31	19	X60, X61
7	X33, X34	20	X55, X57
8	X19, X20	21	X59, X61
9	X36, X37	22	X37, X38
10	X39, X48	23	X36, X38
11	X40, X41	24	X41, X42
12	X43, X44	25	X40, X42
13	X46, X47	26	X44, X45

Лабораторна робота № 3.

Множинне порівняння середніх.

Мета.

Ознайомлення з критеріями, що використовуються для множинного порівняння параметрів положення та їх практичним використанням.

Завдання. По завданням виконати множинне порівняння параметрів положення.

Склад звіту. Завдання, вихідні дані, розрахунки та висновки по них.

Короткі теоретичні відомості

Досить частою є ситуація, коли необхідно порівняти між собою не два значення середніх, а більше. Порівняння їх за допомогою дисперсійного аналізу дозволяє встановити чи можемо ми вважати їх рівними, чи ні. Але в тому випадку, коли вони не рівні, необхідно визначити, які середні рівні між собою, які ні. Використовувати попарні порівняння в цьому випадку не можна, так як виникають парадокси (див. нижче). Крім того, фактичний рівень значущості буде набагато більше ніж встановлений експериментатором. Наприклад, коли ми маємо 5 вибірок, то загальна кількість можливих пар порівнянь k=5!/2!(5–2)!=10, тоді ймовірність отримати хоч один значимий результат при рівні значимості в кожному $P = (1 - (1 - 0{,}05)^{10} \approx 0{,}40$.

Можливі парадокси при перевірці гіпотез про середні

При перевірці гіпотез за допомогою описаних вище критеріїв можливі наступні ситуації:

1. Прийняті гіпотези $\overline{X}_1 = \overline{X}_2$ **і** $\overline{X}_3 = \overline{X}_2$, але відкинута $\overline{X}_1 \neq \overline{X}_3$ **.**

2. Прийняті гіпотези $\overline{X}_1$ =A і $\overline{X}_2$ =A, але відкинута $\overline{X}_1 = \overline{X}_2$ **.**

Щоб прийняти рішення в таких ситуаціях існують спеціальні методи перевірки рівності між собою декількох середніх.

3.1. Метод множинних порівнянь Шеффе

Призначення. Перевірка гіпотези про приналежність кількох середніх до однієї генеральної сукупності, або визначення груп середніх значень, що належать до однієї сукупності.

Нульова гіпотеза.

H_0 : $\sum_{i=1}^{k} C_i \overline{X}_i = 0$, де c_i – певні константи, на яких накладається умова $\sum_{i=1}^{k} C_i = 0$.

Передумови. Дані повинні бути розподілені по нормальному закону і бути незалежними.

Контрастом називається лінійна комбінація середніх значень вибірок. Наприклад ми маємо 5 вибірок, середні значення яких $\overline{X}_1, \overline{X}_2$, $\overline{X}_3, \overline{X}_4$ и $\overline{X}_5$. Припустимо, що ми вважаємо, що ці вибірки належать до двох різних генеральних сукупностей, середні значення в яких $\overline{X}_A$ и $\overline{X}_B$ відповідно. Тоді нульова гіпотеза може бути сформульована у вигляді H_0: $\overline{X}_A - \overline{X}_B$ =0. При цьому в залежності від складу вибірок, із яких можуть складатися групи A і B можливі наступні варіанти цієї гіпотези.

$\frac{1}{2}(\overline{X}_1 + \overline{X}_2) - \frac{1}{3}(\overline{X}_3 + \overline{X}_4 + \overline{X}_5) = 0$ або $\overline{X}_1 - \frac{1}{4}(\overline{X}_2 + \overline{X}_3 + \overline{X}_4 + \overline{X}_5) = 0$, тощо.

Зрозуміло, що коефіцієнти C_i для першого випадку мають значення 1/2, 1/2, –1/3, –1/3, –1/3 , а для другого 1, –1/4, –1/4, –1/4, –1/4. Якщо б ми бажали порівняти першу та четверту вибірки, то коефіцієнти С будуть мати значення 1, 0, 0, -1, 0. Таким чином, змінюючи С, можна перевірити будь-які комбінації пар вибірок Критеріальне значення розраховується по формулі

$$S = \frac{\left(\sum_{i=1}^{k} C_i \overline{X}_i\right)^2}{(k-1) S_{\text{ВН}}^2 \sum_{i=1}^{k} \frac{C_i^2}{n_i}} \qquad (3.1)$$

Де внутрігрупова дисперсія $S_{\text{ВН}}^2 = \frac{1}{n-k} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$, а где $\bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i}$, k – число вибірок, n_i – кількість елементів кожної вибірки, $n = \sum_{i=1}^{k} n_i$ – загальна кількість спостережень. Якщо розраховане значення S буде більше критичного значення $F_{k-1,n-k,\alpha}$, то гіпотеза про рівність середніх відповідних вибірок, чи груп вибірок відкидається.

3.2. Формування груп середніх за допомогою LSD-критерію (least significant difference)

Виконується наступна послідовність дій.

1. Впорядкувати значення середніх по величині.
2. Знайти різницю між кожною сусідньою парою.
3. Для кожної пари, починаючи з першої, виконати перевірки значимості середніх. Для цього розраховується значення LSD. При рівних розмірах вибірок використовується формула:

$$LSD = t_{n-k,\alpha} \sqrt{\frac{2}{n_i} S_{\text{ВН}}^2} = \sqrt{\frac{2}{n_i} S_{\text{ВН}}^2 F_{1,n-k,\alpha}} \qquad (3.2)$$

Якщо, розміри вибірок різні, то користуються формулою, по якій розраховується значення для кожної пара вибірок

$$LSD_{a,b} = t_{n-k,\alpha} \sqrt{\frac{n_a + n_b}{n_a n_b} S_{\text{ВН}}^2} = \sqrt{\frac{n_a + n_b}{n_a n_b} S_{\text{ВН}}^2 F_{1,n-k,\alpha}} ,$$

де $t_{n-k,\alpha}$ – табличне значення критерію Стьюдента; $S^2_{\text{вн}}$ – внутрішньо групова дисперсія; $F_{1,n-k,\alpha}$ – табличне значення критерію Фішера; n_i –кількість спостережень в кожній виборці (якщо вони однакового розміру); n_a и n_b – кількість спостережень в вибірках, які перевіряються; k – кількість вибірок.

Якщо різниця середніх значень сусідньої пари менше значення LSD, то ці середні вважаються однаковими, а відповідні вибірки об’єднуються в однорідну групу.

Зауваження.

При сумісному використання порівнянь по Шеффе і критерию LSD можлива поява уявних протиріч. Наприклад, середні значення вибірок, що входять в *різні* однорідні групи, не відрізняються значимо одне від одного. Справа в тому, що в однорідній вибірці зібрані вибірки, для *сукупності* середніх яких приймається гіпотеза про їх рівність. *сукупності* середніх которых принимается гипотеза об их равенстве. То ж, що деякі з них при попарному порівнянні можуть бути рівними середнім інших вибірок цьому ніяк не протирічить.

Контрольні питання.

- Чому неможливо замінити множинні порівняння попарним порівнянням середніх?
- В чому різниця в висновках при використанні критерію Шеффе і LSD–критерію?
- Які можливі парадокси при використання парного порівняння при факточно множинних?
- Як змінюється величина фактичного довірчого інтервалу при використанні замість множинних критеріїв парних?
- Чи існують непараметричні аналоги множинних порівнянь?

Варіанти завдань.

Дані для розрахунків вибираються відповідно до завдань по таблиці 3.1 з таблиць Додатку А.

Таблиця 3.1. Варіанти завдань по розрахунку показників

№ варіанту	Змінні	№ варіанту	Змінні
1	X2, X39, X48, X51, X58	14	X9, X10, X11, X12, X13, X14, X15
2	X1, X24, X26, X36, X37, X38, X59	15	X1, X24, X26, X36, X37, X38, X59
3	X3, X4, X5, X6, X7, X8, X25, X40, X41,	16	X17, X18, X19, X20, X22, X23, X27, X30, X31, X32, X33
4	X9, X10, X11, X12, X13, X14, X15,	17	X17, X18, X19, X20, X22, X23, X27, X30, X31
5	X13, X14, X15, X16, X17,	18	X36, X37, X38, X59,

	X18, X1		X60,X61, X62
6	X18, X19, X20, X22, X23, X27, X30, X31,	19	X7, X8, X25, X40, X41, X42, X43,
7	X27, X30, X31, X32, X33, X34, X35	20	X41, X42, X43, X44, X45, X55, X56, X57
8	X41, X42, X43, X44, X45, X55, X56	21	X12, X13, X14, X15, X16, X17, X18, X19, X20, X22
9	X19, X20, X22, X23, X27, X30, X31, X32, X33	22	X3, X4, X5, X6, X7, X8, X25, X40, X41
10	X17, X18, X19, X20, X22, X23, X27, X30, X31	23	X41, X42, X43, X44, X45, X55, X56
11	X15, X16, X17, X18, X19, X20, X22, X23, X27, X30, X31	24	X19, X20, X22, X23, X27, X30, X31, X32, X33
12	X18, X19, X20, X22, X23, X27, X30, X31	25	X7, X8, X25, X40, X41, X42, X43,
13	X36, X37, X38, X59, X60,X61, X62	26	X41, X42, X43, X44, X45, X55, X56, X57

Лабораторна робота № 4.

Перевірка наявності зв’язку. Кореляційний аналіз.

Мета.

Вивчити використання кореляційного аналізу.

Завдання. По даному варіанту провести перевірку наявності статистичного зв’язку, використовуючи з учбовою метою параметричні та непараметричні методи кореляції.

Склад звіту. Завдання, вихідні дані, результати розрахунків та висновки по ним.

Короткі теоретичні відомості

4.1. Загальна схема вибору методу визначення наявності зв’язку.

Методи визначення наявності/сили зв’язку між випадковими змінними залежать від характеру зв’язку і від шкал вимірювання випадкових величин.

Вид зв’язку

Таблиця 4.1. Залежність метода визначення зв’язку від кількості змінних

№ п/п	Вид зв’язку	Методи визначення наявності і сили зв’язку
1	Один до одного	парний коефіцієнт кореляції (вид залежить від шкали вимірювань)
2	Багато до одного	множинний коефіцієнт кореляції / регресійна модель / ANOVA
3	Багато між собою	коефіцієнт конкордації
4	Один з багатьох до одного	частковий коефіцієнт кореляції

Таблиця 4.2. Залежність методу оцінювання сили і наявності зв’язку від шкали вимірювання

<table>
<tr><th colspan="2" rowspan="2">Шкали вимірювання змінних</th><th colspan="4">Шкала першої змінної</th></tr>
<tr><th>Номінальна</th><th>дихотомічна</th><th>Рангова</th><th>Відношень / інтервалів</th></tr>
<tr><td rowspan="4">Шкала другої зм</td><td>Номінальна</td><td>Таблиця часток і пропорцій K×L</td><td>Таблиця часток і пропорцій 2×K</td><td>ANOVA непараметрична</td><td>ANOVA</td></tr>
<tr><td>дихотомічна</td><td>Таблиця часток і пропорцій 2×K</td><td>Таблиця часток і пропорцій 2×2</td><td>Рангово-бісеріальний коефіцієнт кореляції</td><td>Точково-бісеріальний коефіцієнт кореляції</td></tr>
<tr><td>Рангова</td><td>ANOVA непараметрична</td><td>Рангово-бісеріальний коефіцієнт кореляції</td><td>Ранговий коефіцієнт кореляції Кендала або Спірмена</td><td>Ранговий коефіцієнт кореляції Кендала або Спірмена</td></tr>
<tr><td>Відношень / інтервалів</td><td>ANOVA</td><td>Точково-бісеріальний</td><td>Ранговий коефіцієнт</td><td>Коефіцієнт кореляції</td></tr>
</table>

і н н о ї			коефіцієнт кореляції	кореляції Кендала або Спірмена	Пірсона

4.2. Кореляція Пірсона.

***Кореляційний зв`язок*** є окремим випадком статистичного зв`язку $M(Y \mid X = x) = \bar{y}(x)$, тобто математичне сподівання змінної Y, при умові, що випадкова величина X приймає значення x. Значення коефіцієнта кореляції обчислюється за формулою:

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

Коефіцієнт кореляції показує силу ***лінійного*** зв`язку між двома вибірками випадкових величин. Його значення змінюється від -1 (рис. 1.7, $r_{xy} \approx -0{,}8$), що відповідає обернено-пропорційному зв`язку до +1 (рис. 1.8, $r_{xy} \approx 0{,}8$), що відповідає прямо пропорційній залежності. Значення 0, означає відсутність зв`язку (рис. 1.9). Оскільки ми маємо справу з випадковими величинами, то одного значення величини коефіцієнта парної кореляції недостатньо для висновків про наявність чи відсутність зв`язку. Необхідно перевірити, чи відрізняється він статистично значимо від нуля, що виконується за допомогою критерію Стюдента.

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

Якщо розрахункове значення t ($t_{розр}$) більше табличного, взятого з *N*–2 степенями свободи, то коефіцієнт кореляції значимий з вибраним рівнем значущості. Напівширина довірчого інтервалу визначається по наступній формулі

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

де N число спостережень, по яким розраховується коефіцієнт кореляції; r — значення коефіцієнта кореляції, $t_{N-2,\alpha}$.— табличне значення критерію Ст’юдента, взятого с N-2 степенями свободи.

4.3. Рангова кореляція

Коефіцієнти кореляції називаються ранговими, постільки перед їх обчисленням значення змінних перетворюють в ранги. Це виконується наступним чином. Спочатку наявні значення змінних розміщують в ранжованому[5] ряду (значення можуть в вихідному стані бути таким рядом). Потім кожному значенню присвоюється ранг від 1 до N, де N — кількість об`єктів чи спостережень. В тому випадку, коли кілька елементів мають один і той же ранг, то кожному з них присвоюється середнє від рангів всіх однакових об`єктів.

Існує кілька різних способів обчислення коефіцієнтів рангової кореляції. Найбільш часто використовують коефіцієнт кореляції Спірмена ρ та коефіцієнт Кендалла τ. Коефіцієнт кореляції Спірмена обчислюється по наступній формулі:

$$\rho(A,B) = 1 - 6/(n^3 - n)\sum_{i=1}^{n}(R_{1i} - R_{2i})^2,$$

де R_{1i} і R_{2i} — ранги i-го об`єкту для кожної з порівнюваних змінних. Значення ρ не залежить від способу впорядкування значень змінних.

Можна побачити, що цей коефіцієнт є повним аналогом коефіцієнта парної кореляції — після перетворення його можливо представити в наступному вигляді

$$\rho(A,B) = \frac{\sum_{i=1}^{n}(R_{1i} - \frac{n+1}{2})(R_{2i} - \frac{n+1}{2})}{\frac{1}{12}(n^3 - n)}.$$

Для перевірки значимості коефіцієнта рангової кореляції Спірмена при n>9 можна користуватися критерієм Стьюдента як і для звичайного коефіцієнта парної кореляції. Перевірка значимості для загального випадку приводиться в [1] в опису табл. 6.10 а.

В тому випадку коли необхідно порівняти не двох, а більшої кількості змінних, наприклад, при визначенні узгодженості думок групи експертів, використовується коефіцієнт конкордації, запропонований Кендаллом:

$$W = \frac{12}{m^3(n^3 - n)}\sum_{j=1}^{n}(\sum_{i=1}^{m}(R_{ij} - \frac{n+1}{2}))^2,$$

де n — кількість досліджуваних об`єктів, $m$ — кількість експертів, $R_{ij}$ — ранг $j$-го об`єкту, який присвоєний йому i-им експертом.

Необхідно звернути увагу на відмінність в значеннях коефіцієнта конкордації від коефіцієнта кореляції. В тому випадку, коли думки експертів повністю протилежні, коефіцієнт конкордації дорівнює нулю (W=0), але коефіцієнт кореляції в даному випадку буде дорівнювати –1.

Значимість коефіцієнта конкордації при малій кількості експертів перевірити важко. Для малих значень існують неповні таблиці в важкодоступній літературі. В тому ж випадку, коли число експертів більше 7, то можливо порівняння значення n(m-1)W з

[5] Впорядкований по величині.

табличним, розподіленим по χ^2 з N-1 степенями свободи.

Перевірка значимості для загального випадку приведена в опису таблиці 6.10 в [1].

4.4. Розрахунок бісеріального коефіцієнта кореляції

Використовується для визначення вили і наявності зв'язку між змінними, одна з яких виміряна в дихотомічній шкалі.

Якщо друга виміряна в шкалі відношень або інтервалів, то використовується точково-бісеріальний коефіцієнт кореляції

$$r_{pb} = \frac{m_1 - m_0}{\sigma_x} \sqrt{\frac{n_1 n_0}{n(n-1)}}$$

де: m_1 и m_0 – середні значення X для для 1 або 0 по Y; σ_x – середнє квадратичне відхилення всіх значень по X; n_1 ,n_0 – кількість значень X з 1 або 0 по Y; n – загальна кількість пар значень

Значимість перевіряється за критерієм хі-квадрат.

В тому випадку, якщо друга змінна виміряна в порядковій шкалі, то використовується рангово-бісеріальний коефіцієнт кореляції.

$r_{rb}=2 * (m_1 - m_0)/ n$

де: n – загальна кількість пар значень

m_1 и m_0 - середні ранги об'єктів з 1 або 0 по дихотомічній змінній.

Значимість перевіряється за критерієм Стюдента.

Коли вважають, що дихотомічна змінна розподілена за нормальним законом, то використовується бісеріальний коефіцієнт кореляції.

$$r_{bis} = \frac{x_1 - x_0}{s_x} \frac{n_1 n_0}{U_n \sqrt{n^2 - n}}$$

Тут всі описані параметри аналогічні точковому коефіцієнту кореляції, за винятком U_n. Це ординати нормального розподілу, які відповідає n_1 /n % площі під кривою. Якщо значення коефіцієнту за абсолютною величиною більше 1, то це означає, що гіпотеза про нормальний розподіл не підтверджується.

Значимість перевіряється за критерієм Стюдента.

4.5.Часткові коефіцієнти кореляції

Для того, щоб «очистити» значення кореляційного зв'язку між двома змінними від можливого впливу третьої, введено поняття часткової кореляції. За ним коефіцієнт кореляції між змінними X і Z визначається за формулою

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

Тут r_{12}, r_{13} і r_{23} — коефіцієнти парної кореляції між змінними X і Y, X і Z, Y і Z відповідно.

Для випадку, коли закорельованих між собою змінних більше трьох, використовується наступна формула

$$r_{ij,X^{i,j}} = \frac{-R_{ij}}{\sqrt{Rii * R_{jj}}}$$

Довірчий інтервал для часткового коефіцієнта кореляції розраховується за формулою $th(z_1) < r < th(z_2)$, де th() – гіперболічний тангенс, а Z розраховується за формулою

$$z_{1,2} = \frac{1}{2}\ln\frac{1+\hat{r}}{1-\hat{r}} \mp \frac{u_{\alpha/2}}{\sqrt{(n-1)-3}} - \frac{\hat{r}}{2((n-1)-1)},$$

В якій $u_{\alpha/2}$ – квантиль стандартного нормального розподілу.

Контрольні питання.

- Як будуються ранги?
- Як перевіряється значимість коефіцієнта кореляції і для чого ця перевірка потрібна?
- Яке співвідношення між фізичною та статистичною залежністю?
- Що таке частковий коефіцієнт кореляції?
- Що таке коефіцієнт конкордації і коли він використовується?
- Які є непараметричні коефіцієнти кореляції і в яких випадках вони використовуються?
- Що таке частковий коефіцієнт кореляції?
- Від чого залежить використання різних видів коефіцієнтів кореляції?
- Силу якого виду залежності перевіряє коефіцієнт кореляції?

Варіанти завдань.

Дані для розрахунків вибираються відповідно до завдань по таблиці 4.1 з таблиць Додатку А.

Таблиця 4.1. Варіанти завдань по розрахунку показників

№ варіанту	Змінні	№ варіанту	Змінні
1	X3, X4	14	X46, X47
2	X6. X7, X8	15	X19, X20
3	X10, X11	16	X55, X56
4	X13, X14	17	X26, X36
5	X16, X17	18	X59, X60
6	X30, X31	19	X56, X57
7	X33, X34	20	X60, X61
8	X19, X20	21	X55, X57

9	X36, X37	22	X59, X61
10	X39, X48	23	X37, X38
11	X40, X41	24	X36, X38
12	X43, X44	25	X41, X42
13	X40, X42	26	X44, X45

Лабораторна робота № 5.

Перевірка наявності зв'язку. Дисперсійний аналіз.

Мета.

Освоїти використання дисперсійного аналізу при обробці даних спостережень чи експерименту в наукових дослідженнях.

Завдання.

По отриманим вихідним даним провести перевірку наявності зв'язку між змінними.

Склад звіту.

Завдання, вихідні дані, результати розрахунків, висновки.

Короткі теоретичні відомості

В задачах, для розв'язання яких використовується дисперсійний аналіз, є відгук числової природи, на який впливають кілька змінних, що мають номінальну природу. Наприклад, кілька різних покриттів для деталі. Вважається, що ми можемо розглядати модель (для двох факторів)

$$(X_i - \overline{X}) = \alpha_i + \beta_i + e_i \,, \qquad (5.1)$$

тобто, розсіяння дорівнює змінам, залежним від одного фактору α_i, плюс розсіяння, яке залежить від другого фактору β_i, плюс випадкова помилка e_i. Тоді загальне розсіяння складається з кількох компонент: $\sigma^2 = \sigma^2_\alpha + \sigma^2_\beta + \sigma^2$. Виділивши відповідні компоненти, за допомогою критерію Фішера можливо перевірити їх значимість.

5.1. Параметричний дисперсійний аналіз

Для найпростішого одно факторного випадку таблиця вихідних даних для дисперсійного аналізу має наступний вигляд:

Таблиця 5.1 Загальний вигляд вихідних даних для дисперсійного аналізу

Номера елементів сукупностей	1	2	…	j	…	n
Номера сукупностей						
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_{jn}
…			…		…	
m	X_{m1}	X_{m2}		X_{mj}		X_{mn}

Це може бути, наприклад, m партій сировини, з кожної з яких взято n зразків. Необхідно визначити, змінюються показники сировини від партії до партії. Суть дисперсійного аналізу в тому, щоб порівнювати дисперсію, обумовлену випадковими величинами, з диперсією, обумовленою впливом деякого фактору (факторів). Якщо вони статистично значимо відрізняються, то це означає, що фактор статистично значимо впливає на досліджувану змінну. Відмінність вважається статистично значимою, якщо розрахункове значення критерію Фішера (відношення міжгрупової дисперсії до внутрішньогрупової) буде більше критичного, взятого з заданим рівнем значущості і числом степенів свободи (m-1) і m(n-1).

Міжгрупова дисперсія розраховується за формулою $S_1^2 = \frac{1}{m-1}\sum_{i-1}^{m}(\bar{X}_i - \bar{X})^2$. (5.2)

Внутрішньогрупова — $S_2^2 = \frac{1}{m(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij} - \bar{X}_i)^2$. (5.3)

Тут $\bar{X}$ — загальне середнє, а $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$.

Для дисперсійного аналізу в англійській мові прийнято скорочення ANOVA (ANalys Of VAriances, що власне і означає дисперсійний аналіз), яке використовується в деяких російсько- і українськомовних джерелах.

Для однофакторного випадку результати дисперсійного аналізу прийнято представляти в формі таблиці (див.. табл. 5.2).

Таблиця 5.2 Представлення результатів однофакторного дисперсійного аналізу

Компоненти дисперсії	Сума квадратів	Число степенів свободи	Дисперсія
Міжгрупова (впливаючий фактор)	$\sum_{i=1}^{m}(\bar{X}_i - \bar{X})^2$	m-1	$S_1^2 = \frac{1}{m-1}\sum_{i-1}^{m}(\bar{X}_i - \bar{X})^2$
Внутрішньогрупова (випадковий вплив)	$\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij} - \bar{X}_i)^2$	m(n-1)	$S_2^2 = \frac{1}{m(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij} - \bar{X}_i)^2$
Загальна	$\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij} - \bar{X})^2$	mn-1	$\frac{1}{mn-1}\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij} - \bar{X})^2$

Тут внутрішньогрупова диспесія відповідає впливу випадкової компоненти, а міжгрупова — вплив фактора, який вивчається.

5.2. Пепараметричний дисперсійний аналіз Фрідмана

Нульова гіпотеза. Середні значення всіх вибірок рівні.

Передумови

- Всі випадкові величина взаємно незалежні.
- Данні кожної вибірки розподілені по одному закону розподілу. Зверніть увагу: закони розподілу в різних вибірках можуть бути різними.

Вихідні дані представляються в наступному вигляді (табл. 5.4).

Таблиця 5.4 Загальний вид представлення для однофакторного дисперсійного аналізу

Номера елементів	1	2	…	j	…	n

сукупностей						
Номера сукупностей						
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_{jn}
…			…		…	
m	X_{m1}	X_{m2}		X_{mj}		X_{mn}

Для цього в кожному стовпці значення X заміняють їх рангами (див. Відповідний розділ). Після цього розраховується значення критерію:

$$c^2 = \frac{12\sum_{j=1}^{n}(\sum_{i=1}^{m} R_{ij})^2}{mn(n+1)} - 3m(n+1), \qquad (5.4)$$

де R_{ij} — відповідні значення рангів.

Якщо розрахункове значення χ^2 буде більше критичного, взятого з заданим рівнем значущості і (n-1) степенями свободи, нульова гіпотеза відкидається і приймається гіпотеза про статистично значиму відмінність між партіями.

Контрольні питання.

- В яких випадках використовується дисперсійний аналіз?
- Яка ідея лежить в основі дисперсійного аналізу?
- Від чого залежить використання конкретної процедури дисперсійного аналізу?
- Якою процедурою можлива заміна дисперсійного аналізу?
- Що використовується у випадку, якщо закон розподілу залежної змінної відмінний від нормального?
- Чи можуть бути в моделі дисперсійного аналізу взаємодії факторів?

Варіанти завдань.

Таблиця 5.5.

№ варіанту	Змінні	№ варіанту	Змінні
1	X68, X69	14	X42, X43
2	X65, X66	15	X44, X45
3	X59, X60	16	X30, X31,
4	X61, X62	17	X32, X33
5	X55, X56	18	X34, X35
6	X51, X54	19	X20, X22
7	X46, X47	20	X22, X23
8	X40, X41	21	X24, X26
9	X12, X13	22	X7, X8

10	X14, X15	23	X40, X42
11	X16, X17	24	X41, X44
12	X17, X18	25	X33, X34
13	X6, X7	26	X8, X9

Пояснення

Кожен стовпчик завдання вважати рядком вихідної дисперсійної таблиці

Лабораторна робота № 6.

Перевірка наявності зв'язку. Таблиці часток та пропорцій.

Мета.

Освоїти використання методів аналізу таблиць часток та пропорцій при обробці даних спостережень чи експерименту в наукових дослідженнях.

Завдання.

По отриманим вихідним даним провести перевірку наявності зв'язку між змінними.

Склад звіту.

Завдання, вихідні дані, результати розрахунків, висновки.

Короткі теоретичні відомості

В загальному випадку чотирьохкліткова таблиця (або таблиця 2Х2) має вигляд представлений в табл. 6.1.

Таблиця 6.1. Загальний вид чотирьохкліткових таблиці часток та пропорцій

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

Перевірка нульової гіпотези про належність обох вибірок до однієї генеральної сукупності виконується застосуванням критерію χ^2, який розраховується за формулою

$$c^2 = \frac{n(AD - BC)^2}{(A+B)(C+D)(A+C)(B+D)} . \qquad (6.1)$$

Для малих вибірок замість n беруть (n–1).

Розрахункове значення порівнюється з критичним, взятим з одним степенем свободи і заданим рівнем значущості. Якщо розрахункове більше критичного, то нульова гіпотеза відкидається і приймається гіпотеза про наявність статистично значущого зв'язку між досліджуваними змінними.

. Таблиці виду 2×К

Таблиця часток і пропорцій 2×К має загальний вигляд, представлений в табл. 6.2.

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

$$c^2 = \frac{n^2}{m(n-m)}\left[\sum_{i=1}^{k} \frac{m_i^2}{n_i} - \frac{m^2}{n}\right] \qquad (6.2)$$

Розрахункове значення порівнюється з критичним, взятим з (k-1) степенями свободи і заданим рівнем значущості. Якщо розрахункове значення більше критичного, то гіпотезу про однорідність слід відкинути і прийняти гіпотезу про наявність зв'язку між змінними.

Таблиця 6.2 Загальний вид таблиці часток та пропорцій типу 2×К

№ вибірки або № рівня 2-гої ознаки	Ознака 1		Σ
	Присутня	Відсутня	
1	m_1	$n_1 – m_1$	n_1
2	m_2	$n_2 – m_2$	n_2
..	...	...	...
I	m_i	$n_i – m_i$	n_i
..	...	...	...
k	m_k	$n_k – m_k$	n_k
Σ	m	n – m	N

Таблиці виду К×L

Таблиці виду К×L (див. табл. 6.3) є найбільш узагальненим видом таблиць часток та пропорцій. В цьому випадку значеннями ознаки 1 можуть бути, наприклад, різні види лікування: симптоматичне, специфічне з нормальними дозами, специфічне з підвищеними дозами або специфічне з доданням інших препаратів тощо. Значеннями ознаки 2 може бути, наприклад, виздоровлення за 2 тижні, виздоровлення за 4 тижні,, летальний результат.

Ознака 2 може також бути сукупністю різних вибірок. Використовується один і той же критерій для перевірки гіпотез про незалежність ознак і про однорідність вибірок. Для випадку, коли перший стовпчик табл. 4.23 є К значеннями рівня другої ознаки, провіряється гіпотеза про незалежність першої та другої ознаки. Якщо ж перший стовпчик складають k різних вибірок, то перевіряється гіпотеза про однорідність цих вибірок, тобто, чи можна вважати, що ці вибірки належать до однієї генеральної сукупності.

Значення критерію розраховується за формулою

$$c^2 = n[\sum_{i=1}^{k}\sum_{j=1}^{m}\frac{n_{ij}2}{n_i.n._j}-1] \qquad (6.3).$$

Розраховане значення порівнюється з критичним, взятим з (k–1)(m–1) степенями свободи і заданим рівнем значущості. Якщо розрахункове значення більше критичного, то гіпотезу про однорідність слід відкинути і прийняти гіпотезу про наявність зв’язку між змінними.

Таблиця 6.3 Загальний вигляд таблиці часток і пропорцій виду К×L

Ознака 2 (k значень рівнів)	Ознака 1 (m значень рівнів)						Суми по рядкам
	1	2		j		m	
1	n_{11}	n_{12}	..	n_{1j}	...	n_{1m}	n_1
2	n_{21}	n_{22}	...	n_{2j}	...	n_{2m}	n_2
...	...	...	...	...	...	...	...
i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{im}	n_i

...	...	...	...	...	...	...	...
k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{km}	n_k
Суми по столпцям	$n._k$	$n._2$	...	$n._j$	...	$n._m$	$n.. = n$

Контрольні питання.

- В яких випадках використовуються таблиці часток та пропорцій?
- Яка задача розв'язується при використанні таблиці часток та пропорцій?
- Які різновиди таблиць ви знаєте?

Варіанти завдань.

Таблиці А10, А11, А12 додатку А. Варіант – в першому стовпчику таблиці. Нумерація наскрізна через всі три таблиці.

Лабораторна робота № 7.

Методи класифікації. Кластерний аналіз.

Мета.

Освоїти використання кластерного аналізу при обробці даних спостережень чи експерименту в наукових дослідженнях.

Завдання.

1. По заданому варіанту розподіл вибірки на кластери. Виконати такий розподіл для трьох варіантів перетворень даних (ізоморфічний, ізотонічний, ще один на вибір) та двох метрик для визначення відстані між кластерами (на власний вибір).
2. Для того ж варіанту виконати нечітке розбиття на кластери.
3. Виконати порівняння.

Склад звіту.

Завдання, вихідні дані, результати розрахунків, висновки.

Короткі теоретичні відомості

Теоретичні відомості.

До задач класифікації відносять задачі розподілу об`єктів чи явищ на однорідні групи при наявності сукупності властивостей, що їх описують.

Типи задач.

Визначення належності об`єкту до однієї з груп, які задані навчальними вибірками (***дискримінантний аналіз***).

Розподіл об`єктів на однорідні групи при відсутності навчальних вибірок (***кластерний аналіз***).

Визначення оптимальної множини змінних для опису об`єкту, яка достатня для розподілу об`єктів на однорідні групи, визначення кількості та значущості дискримінантних функцій.

1 та 2 – це задачі *класифікації*, а 3 – задача *інтерпретац*ії.

Кластерний аналіз – це сукупність багатовимірних статистичних процедур, що дозволяє впорядкувати об`єкти по однорідним групам, які називаються кластерами.

Вихідними даними є матриця "об`єкти–властивості". Кількість груп може бути наперед відома або ні. Існує більш ніж 200 різних алгоритмів кластерного аналізу.

Оскільки різні процедури можуть давати для одних і тих же даних різне розбиття на кластери як по кількості так і по складу, то обов`язково необхідний семантичний аналіз.

7.1. Загальна схема розв`язку задачі для ієрархічної процедури.

1. Формування вибірки для аналізу.
2. Вибір сукупності властивостей, які характеризують об`єкт.

3. Вибір способу нормування вихідних даних.
4. Вибір міри близькості об`єктів
5. Формування початкових кластерів
6. Визначення міри близькості кластерів
7. Формування кластерів

Нормування простору

Таблиця 7.1. Способи нормування простору

Спосіб нормування	Формула
До середнього	$Z_{ij} = \frac{x_{ij}}{\bar{x}_j}, \bar{x}_j = \frac{\sum_{i=1}^{N} x_{ij}}{N}, \forall j \exists (1, M)$
До середнього і середньоквадратичного	$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, s_j = \sqrt{\frac{\sum_{i=1}^{N} (x_{ij} - \bar{x}_j)^2}{N-1}}, \forall j \exists (1, M)$
До максимального	$Z_{ij} = \frac{x_{ij}}{x_{j\max}}, x_{j\max} = \max_{j} x_{*j}, \forall j \exists (1, M)$
До 1	$Z_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}$
Ізотонічне	$V_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}$; $w_i = \sum_{j=1}^{m} V_{ij}$ відстань $d_{ij} = \| w_i — w_j \|$.
Ізоморфічне	$Z_{ij} = \left(x_{ij} \Big/ \sum_{i=1}^{n} x_{ij} \right) \Bigg/ \sum_{j=1}^{m} \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}$

Міри близькості для об`єктів.

Найбільш часто використовуються міри, що базуються на узагальненій відстані Махалонобіса, яке визначається формулою:

$$d_{ij} = \sqrt{(X_i - X_j)^T L^T S^{-1} (X_i - X_j)}$$

де X – вектор спостережень, Λ – симетрична невід`ємно визначена матриця вагових коефіцієнтів (звичайно діагональна), Σ – коваріаційна матриця сукупності, з якої вибрані спостереження.

Реально використовуються наступні окремі види відстаней.

Таблиця 7.2. Формули визначення відстані між об'єктами

Евклідова відстань	Зважена Евклідова відстань	*Хемінгова відстань. (відстань міських кварталів)*
$d_{ij} = \sqrt{\sum_{k=1}^{m}(x_{ik} - x_{jk})^2}$	$d_{ij} = \sqrt{\sum_{k=1}^{m} w_k(x_{ik} - x_{jk})^2}$	$d_{ij} = \sum_{k=1}^{m}\lvert x_{ik} - x_{jk}\rvert$

Таблиця 7.3. Визначення відстані між класами

Назва	***Спосіб визначення***
"Найближчого сусіда".	Мінімальна відстань між парою об`єктів, кожен з яких належить до іншого кластера
"Дальнього сусіда".	Максимальна відстань між парою об`єктів, кожен з яких належить до іншого кластера
"По центрам ваги ".	Відстань визначається між центрами ваги кластерів
"Середнього зв`язку"	Відстань є арифметичним середнім всіх можливих пар комбінацій між об`єктами, що входять в різні кластери

7.2. Нечітка кластерізація методом k-середніх

При нечіткій кластерізації деякі точки можуть належати (частково) до кількох кластерів одночасно. Належність елементів вибірки до певного класу описується матрицею $F = [\mu_{ij}], \mu_{ij} \in [0,1], i = \overline{1,M}; j = \overline{1,k}$. Рядок *i* містить степінь належності об'єкту *i* до кластеру *j*.

При цьому $\sum_{j=1}^{k} \mu_{ij} = 1; 0 < \sum_{i=1}^{M} \mu_{ij} < M$. Кластерізація виконується за наступним алгоритмом:

1. Встановлюються параметри: k – кількість кластерів, m – параметр експоненціальної ваги, ε – критерій зупинки.
2. Випадковим чином генерується матриця нечіткого розбиття F.
3. Розраховуються центри кластерів за формулою $V_j = \dfrac{\sum_{i=1}^{n}(\mu_{ij})^m X_i}{\sum_{i=1}^{n}(\mu_{ij})^m}, j = \overline{1,k}$
4. Розраховуються відстані між об'єктами і центрами кластерів $d_{ij} = \sqrt{\lVert X_i - V_i \rVert^2}$
5. Перерахувати елементи матриці нечіткого розбиття $\mu_{ij} = \begin{cases} \dfrac{1}{\left(d_{ji}^2 \sum_{l=1}^{k} \dfrac{1}{d_{li}^2}\right)^{1/(m-1)}}, d_{ij} > 0 \\ \begin{cases} 1, d_{ij} = 0, i = j \\ 0 d_{ij} = 0, i \neq j \end{cases} \end{cases}$

6. Перевіряється умова зупинки. Якщо $\left\|F - F^{*}\right\| < \varepsilon$, то процедура закінчується, інакше – виконується перехід в п.3

Для оцінки якості розбиття використовують параметри розсіяння. Загальне розсіювання $S = \sum_{i=1}^{M} d_{i\bar{X}}^{2}$, де $\bar{X} = \frac{1}{M}\sum_{i=1}^{M} X_i$ - загальний центр ваги. Між групове розсіювання (між центрами кластерів) $B = \sum_{i=1}^{k}\sum_{j=1}^{k} d^2(\bar{X}_i, \bar{X}_j)$. Внутрігрупове розсіювання $Q = \sum_{i=1}^{k}\sum_{j=1}^{n_i} d^2(X_j, \bar{X}_i)$.

Це формули для класичного кластерного аналізу. Для нечіткого вони відрізняються наявністю множника μ_{ij}.

При використанні евклідової, або зваженої евклідової відстані має місце співвідношення S = Q + B.

Статистика T, яка показує частку загального розсіяння, яка пояснюється між груповим розсіянням T = 1 – Q/S. Кластером вважається множина точок, для якої виконується співвідношення $Q_i/N<S/N$, а згущенням множина точок при виконанні умови $\max(d_i^2) < S/N$.

Контрольні питання.

- На які групи розділяються процедури кластерного аналізу?
- Геометрична інтерпретація кластерного аналізу?
- Які етапи виконання процедури кластерного аналізу?
- В чому різниця при використання ізоморфічного та ізотонічного розбиття?
- Чи зберігається розміщення кластерів у просторів при різних способах нормування?
- Задачі кластерного аналізу?
- Суть методу куль?
- Суть методу дендритів?
- Що таке нечіткий кластерний аналіз?
- Які є міри близькості між об’єктами?
- Які існують міри близькості між кластерами?
- Загальна схема ієрархічної процедури кластерного аналізу.
- Загальна схема послідовної процедури кластерного аналізу.

Варіанти завдань

Таблиця 7.3.

№ варіанту	Дані	№ варіанту	Дані	№ варіанту	Дані

1	X1,X2,X3,X4, X5	10	X61,X62,X63, X64,X65	19	X44,X45,X48, X49,X50
2	X6,X7,X8,X9, X10	11	X66,X67,X68, X69,X70	20	X51,X52,X53, X56,X57
3	X11,X12,X13, X14,X15	12	X71,X72,X73, X74,У1	21	X54,X55,X58, X59,X60
4	X16,X17,X18, X19,X20	13	X1,X2,X3, X6,X7	22	X61,X62,X63, X66,X67
5	X21,X22,X23, X24,X25	14	X4,X5, X8, X9, X10	23	X64,X65,X68, X69,X70
6	X26,X27,X28, X29,X30	15	X11,X12,X13, X16, X17	24	X71,X72,X73, X1,X2
7	X31,X32,X33, X34,X35	16	X14,X15,X18, X19,X20	25	X74,У1,X5, X8, X9
8	X36,X37,X38, X39,X40	17	X21,X22,X23, X26,X27	26	X11,X12,X13, X1,X2
9	X41,X42,X43, X44,X45	18	X24,X25,X28, X29,X30	27	X21,X22,X23, X28,X29

Лабораторна робота № 8.

Методи класифікації. Дискримінантний аналіз.

Мета.

Освоїти використання дискримінантного аналізу при обробці даних спостережень чи експерименту в наукових дослідженнях.

Завдання.

1. По заданих навчальних вибірках (табл.. А13) визначити приналежність об’єктів (табл..А14) до певної категорії.
2. Визначити мінімальну кількість факторів, достатніх для визначення приналежності до категорії.

Склад звіту.

Завдання, вихідні дані, результати розрахунків, висновки.

Контрольні питання.

- Геометрична суть дискримінантного аналізу?
- Які задачі розв’язуються за допомогою дискримінантного аналізу?
- Передумови дискримінантного аналізу.
- Для чого використовується покроковий дискримінантний аналіз.
- Передумови дискримінантного аналізу.
- На що впливає мультиколінеарність вибірки.

Варіанти завдань.

В табл.. А13 і А14. Варіанти – в першому стовпчику таблиць. Нумерація варіантів наскрізна через обидві таблиці.

Теоретичні відомості.

Дискримінантний аналіз дозволяє вивчати відмінності між двома або більшою кількістю групами об`єктів по кільком ознакам одночасно. Кожна група об`єктів називається класом. Класи розглядаються як значення деякої класифікуючої змінної, яка вимірюється в шкалі найменувань.

Якщо класифікуюча змінна залежить від дискримінантних, то дискримінантний аналіз є аналогом багатовимірного регресійного аналізу, коли відгук вимірюється в шкалі найменувань. Наприклад, дискримінантні змінні описують стан хворого та методику лікування, а класифікуюча – показник типу: вилікуваний – не вилікуваний.

У випадку, коли дискримінантні змінні залежать від класифікуючої, тобто належність до певного класу викликає зміну значень одночасно кількох параметрів дискримінантний аналіз є аналогом багатовимірного дисперсійного аналізу. Наприклад, класифікуюча змінна – захворювання, а дискримінантні – показники, на основі яких ставиться діагноз.

В дискримінантному аналізі розрізняють дві задачі: інтерпретації та класифікації. Задача ***інтерпретації*** — визначення кількості та значимості дискримінантних функцій, а також їх значень для пояснення відмінностей між класами. Задача ***класифікації*** — визначення класу, до якого належить новий об`єкт.

В дискримінантному аналізі, на відміну від кластерного, є навчальна вибірка, в якій відомо, до яких класів відноситься об`єкт. По цій виборці формуються правила, які в наступному дозволяють визначити, до яких класів належать нові об`єкти.

Призначення

Вивчення відмінностей між двома чи більшою кількістю груп об`єктів одночасно по кільком змінним, що їх описують.

Передумови

- Об`єкти (спостереження) належать двом чи більшій кількості класів.
- В кожному класі є щонайменше два об`єкти.
- Число дискримінантних змінних не повинно бути більше числа об`єктів мінус два.
- Дискримінантні змінні повинні вимірюватися в шкалі інтервалів або відношень.
- Дискримінантні змінні повинні бути лінійно незалежні.
- Дискримінантні змінні повинні бути розподілені по багатовимірному нормальному закону розподілу (тобто кожна змінна має нормальний закон розподілу при зафіксованих значеннях інших).
- Коваріаційні матриці класів приблизно рівні між собою.

Розв'язання задачі для випадку двох класів

Маємо два класи, об`єкти яких характеризуються p змінними. Для першого класу сформована вибірка X**,** розміром n_1, для другого — Y розміром n_2.

1. Розраховуються середні значення для кожної змінної по кожному класу (вибірці) $\bar{x}_j = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{ij}$ и $\bar{y}_j = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{ij}$.

2. Визначаються оцінки коваріаційних матриць для кожного класу S_x та S_y. $s_{kj}(x) = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$ и $s_{kj}(y) = \frac{1}{n_2}\sum_{i=1}^{n_2}(y_{ij} - \bar{y}_j)(y_{ik} - \bar{y}_k)$.

3. Розраховується незміщена оцінка об`єднаної коваріаційної матриці. $S = \frac{1}{n_1 + n_2 + 1}(n_1 S_x + n_2 S_y)$. (5.19)

4. Знаходиться матриця S^{-1}, обернена до S.
5. Розраховується вектор оцінок коефіцієнтів дискримінантної функції $A = S^{-1}(\overline{X} - \overline{Y})$.
6. Визначаються оцінки векторів дискримінантних функцій для вихідних класів U_x=XA и U_y=YA.
7. Обчислюються середні значення оцінок дискримінантних функцій $\overline{u}_x = \frac{1}{n_1}\sum_{i=1}^{n_1} u_{xi}$ и

$\overline{u}_y = \frac{1}{n_2}\sum_{i=1}^{n_{12}} u_{yi}$.

8. Визначається дискримінантна константа $C = \frac{1}{2}(\overline{u}_x + \overline{u}_y)$.

Для того, щоб визначити, до якого класу належить нове спостереження (об`єкт) Z, необхідно спочатку обчислити для нього оцінку дискримінантної функції $u_z = \sum_{i=1}^{p} a_i z_i$ . Якщо це значення більше, або дорівнює константі C, то новий об`єкт належить до класу X, якщо менше — до класу Y (при $\overline{u}_x > \overline{u}_y$).

Розширення алгоритму для випадку числа класів більше двох

Існує k класів, об’єкти яких характеризуються p змінними. Для кожного класу сформована вибірка $X^{(l)}$**,** розміром n_l.

- Розраховуються середні значення по кожній змінній для кожної вибірки (класу)

$$\overline{x}_j^{(l)} = \frac{1}{n_l}\sum_{i=1}^{n_l} x_{ij}^{(l)} .$$

- Визначаються оцінки коваріаційних матриць для кожного класу S_l.

$$s_{mj}^{(l)}(x) = \frac{1}{n_l}\sum_{i=1}^{n_l}(x_{ij} - \overline{x}_j)(x_{im} - \overline{x}_m)$$

- Розраховується незміщена оцінка об’єднаної коваріаційної матриці $S = \frac{1}{\sum_{i=1}^{k} n_i - k - 1}\sum_{i=1}^{k} n_i S_i$.
- Знаходиться матриця S^{-1}, обернена до S.
- Розраховуються вектори оцінок коефіцієнтів дискримінантної функції $A^{(l)} = S^{-1}\overline{X}^{(l)}$.
- Визначаються дискримінантні константи $\lambda_l = \frac{1}{2}\overline{X}^{(l)}(S^{-1}\overline{X}^{(l)})$.
- Визначається приналежність нового об’єкта до класу. Об’єкт, описаний вибіркою Z,

належить тому класу, для котрого значення виразу $Z^T A^{(l)} - \lambda_l$ максимальне.

Покроковий дискримінантний аналіз

Теоретичні відомості

Вихідними даними для покрокового дискримінантного аналізу є вибірка X (M змінних, N спостережень), яка складається з K груп розміром N_k кожна ($N = \sum_{k=1}^{K} N_k$). Розраховується внутрішньо групове (W) і загальне розсіяння (T).

$$W_{ij} = \sum_{k=1}^{K} \sum_{l=1}^{N_k} (X_{kli} - \overline{X}_{ki})(X_{klj} - \overline{X}_{kj}) \quad (1)$$

$$T_{ij} = \sum_{k=1}^{K} \sum_{l=1}^{N_k} (X_{kli} - \overline{X}_{i})(X_{klj} - \overline{X}_{j}) \quad (2)$$

де X_{kli} – значення l-го спостереження в k-й групі для i-ї змінної; $\overline{X}_{ki}$ – середнє для i-ї змінної в k-й групі; $\overline{X}_i$ – загальне середнє для i-ї змінної. Даля розраховується $V_i = \frac{W_{ii}}{T_{ii}}$. Значущими вважаються ті змінні, для яких значення F-включення більше критичного. $F_{вкл} = \frac{(N-p-K)}{K-1} \frac{1-V_i}{V_i}$, тут p – кількість змінних, які вже включені в модель.

Зверніть увагу, що використання операції знаходження оберненої матриці вимагає доброї обумовленості, а застосування критерію Фішера – незалежності параметрів. Тобто, ми маємо справу з тими ж вимогами, що і до плану регресійного чи дисперсійного аналізу.

Додаток А.

Вихідні дані для розрахунків

Таблиця А1.

X1	x2	x3	x4	x5	x6	x7	x8	x9
49	0.15	3	5	3	5.7	5.4	5.9	17
39	-0.05	2	3	1	5.7	5.1	5.9	34
44	-0.02	1	3	1	5.1	3.9	5.7	23
41	-0.8	4	4	2	5.3	4.8	5.8	27
41	0.33	4	3	3	5.5	4.8	5.3	23
44	0.77	1	2	1	5.5	4.8	5.7	21
37	0.74	4	4	3	5.4	5.5	6.2	32
47	0.47	5	4	3	6.1	6	5.3	34
41	-0.04	3	4	4	5.3	5.4	5.3	25
36	-0.08	2	2	3	6.7	6.2	7	26
46	-0.2	2	4	2	6.6	5.9	6.4	30
54	0.12	6	3	2	5.5	5.2	5.2	24
40	0.34	2	2	2	6.4	5.7	6.7	23
42	0.36	2	2	2	4.9	4.6	5.1	26
31	0.45	2	3	4	5.6	6.1	6.4	29
42	0.4	3	2	2	5.8	4.1	6.4	24

Таблиця А2.

x10	x11	x12	x13	x14	x15	x16	x17	x18
23	19	20	31	28	19	22	23	33
31	30	27	31	28	19	22	23	33
27	23	24	17	14	21	21	22	17
28	23	24	28	28	23	24	21	29
23	19	23	20	25	20	15	20	25
23	17	14	22	17	22	21	18	21
32	23	21	31	29	19	18	24	24
32	36	19	29	34	16	13	30	34
27	15	20	16	20	27	19	20	21
32	22	21	24	28	22	23	27	25
20	25	27	27	33	13	17	28	27
15	18	24	21	22	17	22	23	16
29	15	21	19	23	18	19	16	21
25	22	26	27	27	23	19	25	23
33	31	27	30	35	24	23	27	36
27	22	29	25	25	19	20	20	23

Таблиця А3.

x19	x20	x21	x22	x23	x24	x25	x26	x27
12.33	10.92	2.8	18.31	20.1	58.11	6	48.33	17
19.67	20.43	1.4	21.82	21.8	53.2	4	66.08	14
12.33	8.36	0.4	2.53	31.53	75	8	75.5	17
14.67	10.77	1.2	21.35	28.6	63.25	4	68.69	7
9	5.9	0.2	17.72	19.7	53.8	0	66.58	2
9	4.38	1	20.45	23.07	58.3	3	65.9	7
14.33	12.38	1.2	23.7	23.07	58.3	3	65.9	7
14	11.18	1.4	39.99	23.7	55.78	6	67.27	5
9.67	8.23	0.4	14.94	24.27	55.7	1	65.19	6
13	24.92	1.6	33.32	19.16	55.44	3	65.1	3
20.33	25.59	1.6	14.04	23.9	52.5	3	61.29	3
9.67	5.76	0.2	18.27	22.63	71.78	1	70.82	3

9	7.6	1.4	10.84	22	56.78	2	65.64	3
9.67	6.73	1.2	38.69	21.36	57.7	0	68.75	4
11.67	24.07	1.4	22.84	25.13	59.1	1	72	6
9.67	5.44	0.4	30.75	19.03	48.1	4	61.36	7

Таблиця А4.

x28	x29	x30	x31	x32	x33	x34	x35	x36
100.2	417	35	36	30	28	27	26	44
86.84	448	36	29	29	25	26	24	55
171.96	197	32	26	26	26	27	24	47
93.48	386	30	25	21	31	23	25	52
103.24	407	32	24	26	25	25	23	61
101.86	378	36	29	28	25	26	24	64
101.86	378	43	32	34	21	25	26	66
90.52	384	34	28	33	19	25	22	69
108.32	374	29	27	32	26	25	26	49
102.48	285	32	32	28	28	23	21	57
94.68	444	39	25	32	23	28	27	48
97.4	386	41	33	34	27	27	26	71
83.96	411	43	33	34	33	36	23	53
110.72	347	36	26	21	21	25	24	47
93.8	412	31	26	21	19	21	20	71
88.48	391	41	32	28	23	28	25	50

Таблиця А5.

x37	x38	x39	x40	x41	x42	x43	x44	x45
43	41	1.02	7	5	3	1	6	7
61	51	0.9	5	1	6	2	8	4
56	61	0.84	2	4	4	4	4	4
67	59	0.78	6	4	6	4	4	3
58	57	1.05	5	5	4	3	6	3
53	64	1.21	6	4	4	4	6	3
70	72	0.94	5	6	5	4	2	4
70	78	0.99	7	4	3	2	8	3
44	57	1.11	5	2	3	3	5	3
50	48	1.14	5	5	4	4	7	3
52	48	0.92	6	3	4	2	6	3
71	61	1	3	0	5	4	3	4
46	51	1.15	5	5	5	5	3	6
51	47	0.92	3	4	3	4	5	4
68	60	1.04	6	3	5	2	7	3
59	59	0.85	3	1	1	7	2	3

Таблиця А6.

x46	x47	x48	x49	x50	x51	x52	x53	x54
560	515	0.919	36	22	0.587	17	249	0.519
860	767	0.892	40	46	0.574	63	253	0.303
700	682	0.974	136	9	1.21	27	175	0.472
560	496	0.887	25	31	1.275	17	199	0.519
720	669	0.929	52	25	1.145	10	207	0.551
1020	1000	0.98	260	10	0.675	8	336	0.561
800	749	0.937	64	25	0.915	20	254	0.504
720	564	0.784	17	76	0.892	37	171	0.425
760	717	0.943	69	21	0.818	17	232	0.519
880	815	0.927	61	32	0.755	28	270	0.467
580	545	0.94	50	17	0.652	12	208	0.542
740	701	0.947	72	19	0.944	8	201	0.561

820	778	0.948	80	21	0.782	25	271	0.481
700	647	0.924	47	26	0.848	50	241	0.364
840	766	0.912	48	37	1.028	31	325	0.452
1080	1038	0.961	139	21	0.982	19	271	0.509

Таблиця А7.

x55	x56	x57	x58	x59	x60	x61	x62	x63
2	7	3	0.25	46.6	33.3	46.6	36.6	0
6	3	3	0.4	46.6	46.6	46.6	60	50
6	6	0	0.3	26.6	6.6	40	13.3	0
0	8	4	0.35	46.6	66.6	53.3	66.6	100
10	2	0	0.86	33.3	53.3	40	56.6	90
5	6	1	0.44	53.3	53.3	53.3	53.3	70
8	4	0	0.3	46.6	46.6	50	46.6	100
8	3	1	0.35	40	33.3	46.6	33.3	80
6	3	3	0.44	20	26.6	20	36.6	60
6	5	1	0.32	13.3	33.3	13.3	30	0
4	8	0	0.3	33.3	40	33.3	40	0
3	4	5	0.4	26.6	53.3	30	60	0
7	4	1	0.34	46.6	46.6	46.6	46.6	0
3	4	5	0.51	53.3	60	53.3	63.3	20
1	8	3	0.34	53.3	66.6	66.6	66.6	0
9	3	0	0.61	53.3	53.3	13.3	13.3	60

Таблиця А8.

x64	x65	x66	x67	x68	x69	x70	x71	x72
0	0	0	0	0	0	-1	-1	-1
70	83.33333	67.77778	53.33333	93.33333	73.33333	-1	1	-1
0	0	0	0	0	0	-1	-1	-1
100	75	91.66667	66.66667	66.66667	66.66667	-1	-1	1
50	0	46.66667	100	100	100	-1	-1	-1
76.66667	83.33333	76.66667	66.66667	66.66667	66.66667	-1	-1	1
70	91.66667	87.22222	13.33333	0	6.666667	0	-1	0
50	98.33333	76.11111	0	50	25	-1	-1	-1
13.33333	0	24.44444	53.33333	66.66667	60	-1	-1	-1
0	16.66667	5.555556	66.66667	66.66667	66.66667	-1	-1	-1
33.33333	100	44.44444	66.66667	100	83.33333	-1	-1	0
46.66667	85	43.88889	40	36.66667	38.33333	-1	-1	1
0	0	0	66.66667	66.66667	66.66667	-1	-1	-1
73.33333	83.33333	58.88889	66.66667	93.33333	80	-1	1	1
0	83.33333	27.77778	33.33333	100	66.66667	-1	-1	-1
96.66667	80	78.88889	66.66667	96.66667	81.66667	-1	-1	1

Таблиця А9.

x73	x74	y1	y2	y3	y4
-1	-1	5	10	56.95	19.44
1	1	10	13	55.7	19.5
-1	-1	9	11	57.53	19.5
1	1	10	10	56	19.1
0	0	9	13	57.2	19.7
0	1	12	8	48.68	19.15
-1	-1	2	7	56.65	20.25
-1	-1	5	9	52	18.28
-1	-1	9	11	57	20.2
-1	-1	2	3	58.15	21.5
-1	0	4	4	59.53	20.16
1	1	4	7	60.59	21.75

1	1	10	12	60.85	16.75
-1	-1	4	9	57.37	19.78
-1	0	4	7	56.44	19.32
-1	1	5	7	55.94	21.5

Таблиця А10

Варіант	Назва групи	Наявність дії	
		Дія була	Дії не було
1	А	87	16
	Б	430	198
2	А	87	26
	Б	430	188
3	А	15	23
	Б	67	79
4	А	179	236
	Б	223	164
5	А	137	281
	Б	151	185
6	А	265	56
	Б	17	274
7	А	82	188
	Б	175	38
8	А	226	27
	Б	219	98

Таблиця А11

№	Категорія	З позитивною реакцією	З негативною реакцією
9	a	19	42
	b	23	71
	c	23	227
10	a	20	6
	b	18	6
	c	17	2
11	a	13	8
	b	6	1
	c	5	1
12	a	19	17
	b	9	11
	c	6	3
13	a	13	18
	b	10	1
	c	10	11
14	a	9	17
	b	12	17
	c	5	14
15	a	19	6
	b	29	1
	c	14	3

16	a	7	9
	b	20	10
	c	9	2

Таблиця А12

№	Значення	Виробник							
		1	2	3	4	5	6	7	
17	Високе	15,4	16,8	31,9	32,4	20,6	37,9	29,6	184,6
	Номінал	80,1	78,4	65,1	64,3	77	58,4	67,8	491,1
	Низьке	4,5	4,8	3	3,3	2,4	3,7	2,6	24,3
18	Високе	17	16,8	31,9	32,4	20,6	37,9	29,6	186,2
	Номінал	60	78,4	65,1	64,3	77	58,4	67,8	471
	Низьке	22,5	4,8	3	3,3	2,4	3,7	2,6	42,3
19	Високе	82	12,3	62,3	39,2	23,5	43,7	26	289
	Номінал	14,8	47,8	6	10,5	61,5	21,9	11,2	173,7
	Низьке	3,2	39,9	31,7	50,3	15	34,4	62,8	237,3
20	Високе	46,8	71,3	33,9	28	21,8	37,7	26,7	266,2
	Номінал	20	0,9	22,4	23,1	41,1	32,3	23,1	162,9
	Низьке	33,2	27,8	43,7	48,9	37,1	30	40,2	260,9
21	Високе	19	13	79,4	20,9	69,2	68,9	47,4	317,8
	Номінал	20,4	37,5	15,8	68,2	17,8	6,6	46,9	213,2
	Низьке	60,6	49,5	4,8	10,9	13	24,5	5,7	169
22	Високе	43,3	43,9	13,6	76,1	21,8	58,9	29,2	286,8
	Номінал	34	30,1	25,6	16,2	46,9	28,1	40,5	221,4
	Низьке	22,7	26	60,8	7,7	31,3	13	30,3	191,8
23	Високе	27,5	40,3	30,6	41,2	64,9	39,8	6,4	250,7
	Номінал	38,7	32,1	46,8	42,8	27,8	30,3	2,7	221,2
	Низьке	33,8	27,6	22,6	16	7,3	29,9	90,9	228,1
24	Високе	27,3	24	23,3	58,3	73,8	26	11,4	244,1
	Номінал	31,9	31,4	24,9	29,4	13,6	15,2	55	201,4
	Низьке	40,8	44,6	51,8	12,3	12,6	58,8	33,6	254,5

Таблиця А13. Навчальна вибірка для дискримінантного аналізу

Категорія	№ точки	Координати точок		
		X_1	X_2	X_3
1	1	5	6	5,6
	2	4	7	6,5
	3	6	7	1,5
	4	3	8	5,1
	5	7	8	7,4
	6	4	9	7
	7	6	9	7,3
	8	5	10	4,8
2	9	1	2	7,1
	10	2	4	7,7
	11	2	3	8,6
	12	2	3	0,1
	13	3	3,5	7,8
	14	3	3	5,6

	15	3	2	3,4
	16	3,5	4	0,6
3	17	6	4	7,6
	18	6,5	2	6,3
	19	7	4	4,1
	20	8	3	5,8
	21	8	1	8,8
	22	8	4	5
	23	9	3	4,3
	24	9	1	6,3

Таблиця А14. Варіанти точок, приналежність яких необхідно визначити

№ варіанту	№ точки	Координати		
		X_1	X_2	X_3
1	1	6,1	1,6	3,5
	2	5	0,9	8,3
	3	0,6	4,8	0,9
2	4	3,7	6,3	3,3
	5	2,8	4,5	1,4
	6	1,4	5,7	1,6
3	7	9,3	6,5	3,1
	8	7,8	4,8	6,3
	9	4,3	9,3	2,5
4	10	5,8	5,9	0,7
	11	6,6	8,3	2,9
	12	6	4,8	2
5	13	7,7	2,9	5,7
	14	9,9	1,2	9,5
	15	3,6	9,2	4,5
6	16	2,5	9,8	8,2
	17	3,2	0	1,2
	18	4,8	0,1	8,9
7	19	3,4	3	4,4
	20	5,3	10	4,7
	21	3,6	9,7	0,1
8	22	1,3	7,3	9,8
	23	6,3	2	9,7
	24	3,4	7,9	8,3
9	25	1,3	3,5	6
	26	7,5	8,3	5,2
	27	2,7	2,5	4
10	28	0,6	5	6,5
	29	7,9	0,7	8,2
	30	9,4	4,9	1,3
11	31	9,5	1,9	1,9
	32	7	5,9	7,6
	33	4,7	0,9	8,6
12	34	8,5	4,5	6,9
	35	1,7	3,1	1,1
	36	3,5	9,6	2,3
13	37	5,8	1,1	0,9
	38	3,9	5,1	7,5
	39	1,6	7,5	9,6
14	40	5,5	0,9	4,3

	41	3,9	6,8	3,3
	42	2,9	0,4	5,5
15	43	2,8	4	8,2
	44	5,6	0,2	4,4
	45	1,7	5,6	3,5
16	46	8	9,2	1,6
	47	7	2,5	0,2
	48	0,9	3,6	3,1
17	49	6,7	9,1	9,2
	50	2,7	7,8	5,1
	51	1,1	7,7	3
18	52	7,7	1	6
	53	6,3	7,6	8
	54	5	3,3	7,4
19	55	7,4	4,7	1
	56	2,2	6,8	3,3
	57	4,8	3,9	4,2
20	58	6,8	5,1	0,5
	59	3,5	5,1	2,9
	60	2	3,3	4,7
21	61	8,1	1,6	1,7
	62	9	3,1	6,5
	63	7,7	7,7	0
22	64	9,1	6,8	9,3
	65	0,9	6,2	8,8
	66	9,9	6,8	6
23	67	9	3,1	5,4
	68	6,3	1,9	5,7
	69	7,9	6,2	7,9
24	70	9,2	5,1	6,3
	71	6,2	9,1	6,9
	72	6,2	2,7	8,2

Додаток Б. Приклади виконання лабораторних робіт

Б.1. Лабораторна робота №1

Варіаційний ряд числа циклів до руйнування при консольному вигині з обертанням зразків, виготовлених з алюмінієвого сплаву АВ приведено в табл.Б.1.1

Таблиця Б.1.1. Варіаційний ряд для розрахунку вибіркових характеристик

№	$n*10^{-6}$	№	$n*10^{-6}$	№	$n*10^{-6}$	№	$n*10^{-6}$
1	0,736	26	2,143	51	3,330	76	5,339
2	0,825	27	2,338	52	3,492	77	5,339
3	1,051	28	2,369	53	3,519	78	5,485
4	1,093	29	2,405	54	3,612	79	5,536
5	1,103	30	2,415	55	3,633	80	5,673
6	1,108	31	2,415	56	3,637	81	6,036
7	1,160	32	2,465	57	3,676	82	6,273
8	1,202	33	2,493	58	3,677	83	6,274
9	1,208	34	2,556	59	3,796	84	6,292
10	1,277	35	2,564	60	3,822	85	6,365
11	1,283	36	2,567	61	3,905	86	6,520
12	1,313	37	2,640	62	3,942	87	6,560
13	1,314	38	2,709	63	4,070	88	7,327
14	1,330	39	2,718	64	4,437	89	7,444
15	1,447	40	2,897	65	4,440	90	7,901
16	1,474	41	2,904	66	4,720	91	8,037
17	1,495	42	2,915	67	4,720	92	8,146
18	1,514	43	2,954	68	4,764	93	8,297
19	1,546	44	2,960	69	4,893	94	8,511
20	1,567	45	3,049	70	4,916	95	9,005
21	1,611	46	3,144	71	5,012	96	12,372
22	1,622	47	3,151	72	5,122	97	14,736
23	1,697	48	3,295	73	5,167	98	18,210
24	1,823	49	3,312	74	5,221	99	18,944
25	2,049	50	3,322	75	5,221	100	28,744

На рис.Б.1.1 показано фрагмент вихідних даних і результати розрахунків основних вибіркових характеристик. Напівширина довірчого інтервалу розрахована для середнього значення.

В табл..Б.1.2 показано використання функцій для розрахунку. В функції =ДОВЕРИТ(0,05;G7;СЧЁТЗ(C5:C104)) 0,05 – рівень значущості, G7 – посилання на значення середньоквадратичного відхилення, а СЧЁТЗ(C5:C104) – розрахунок розміру вибірки за діапазоном виділених комірок.

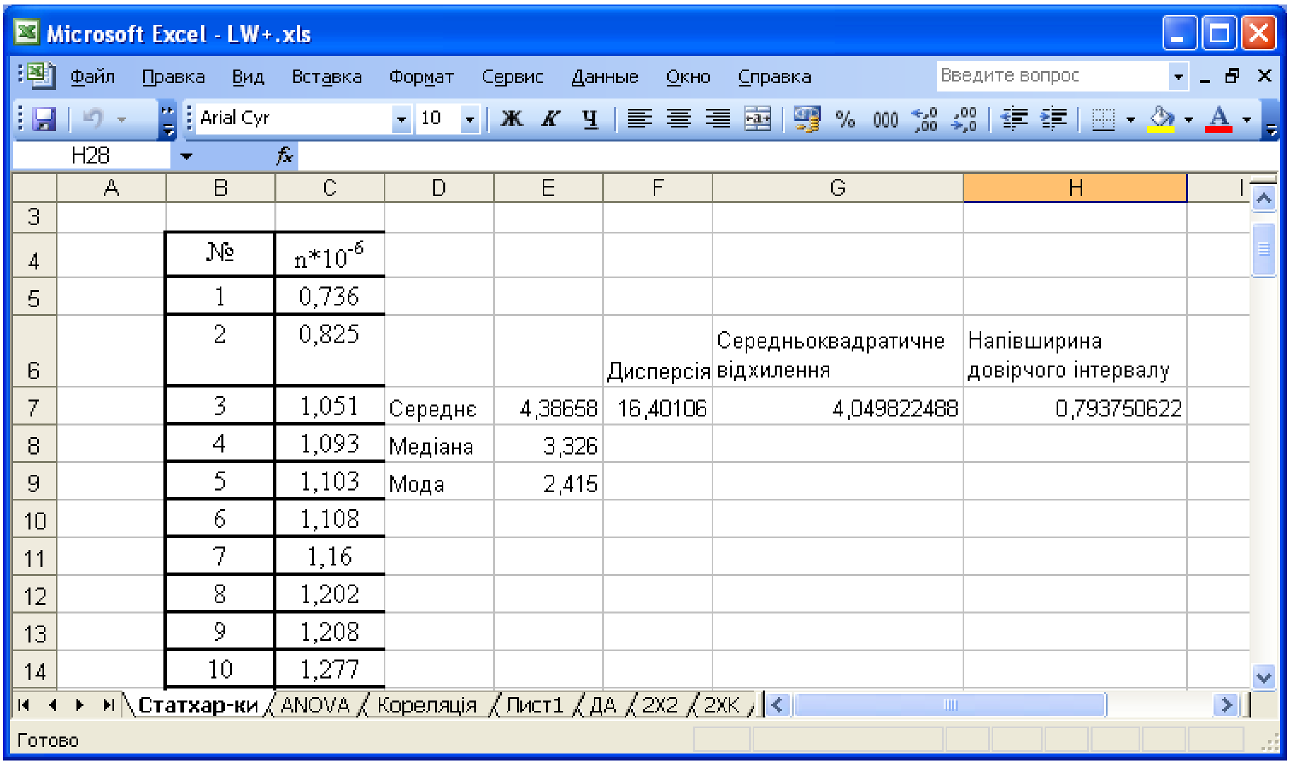


Рис. Б.1.1. Розрахунок основних статистичних параметрів вибірки

Таблиця Б.1.2. Розшифрування функцій

	E	F	G	H
6	=СРЗНАЧ(C5:C104)	=ДИСП(C5:C104)	=СТАНДОТКЛОН(C5:C104)	=ДОВЕРИТ(0,05;G7;СЧЁТЗ(C5:C104))
7	=МЕДИАНА(C5:C104)			
8	=МОДА(C5:C104)			

Гістограму можна побудувати за допомогою функції, яка є в надбудові “Анализ данных”. Для цього в меню послідовно вибираємо “Сервис”, “Анализ данных”. В вікні, яке з’явилось (рис. Б.1.2.) вибирається “Гистограмма”.

Після цього з’являється вікно (рис.Б.1.3), в якому необхідно задати параметри для побудови гістограми.

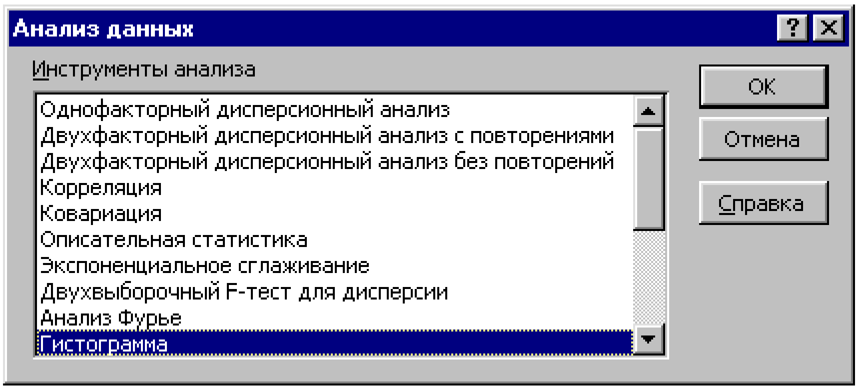


Рис. Б.1.2. Вікно вибору функції

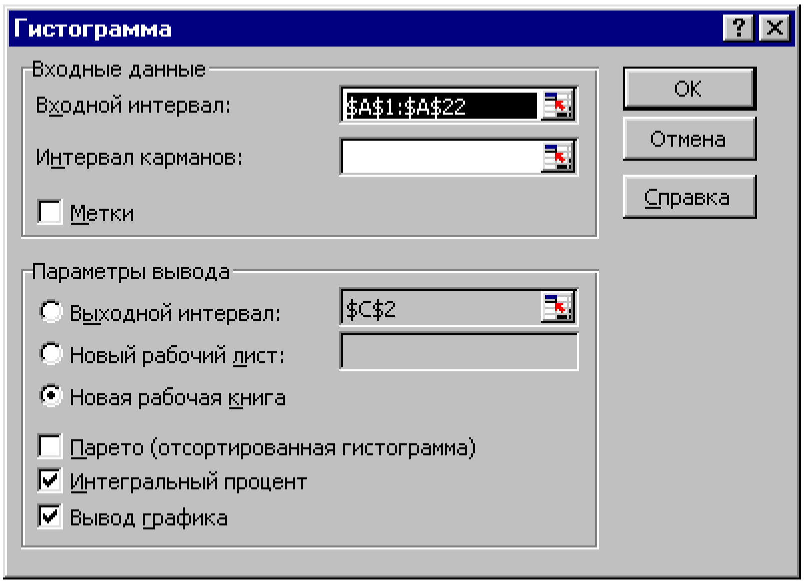


Рис. Б.1.3. Вікно встановлення параметрів для побудови гістограми

Вікна завдання параметрів мають наступне значення.

Входной диапазон.

В цьому вікні необхідно задати посилання на діапазон комірок, в якому знаходяться вихідні дані. Ці дані мають бути переліком значень, а не частотами.

Интервал карманов (необов'язковий). Вводиться діапазон комірок, в якому знаходиться набір граничних значень (розміщений в порядку збільшення) визначаючих відрізки (карманы). Якщо вони не задані, то програма вибирає їх самостійно. Обчислюється кількість значень між поточним початком відрізка і сусіднім наступним, якщо він є. При цьому включаються значення на нижній границі відрізку, але не включаються на верхній.

Метки

Відмічається, якщо перший рядок чи стовпець вхідного інтервалу містить заголовки. В противному разі назви створюються автоматично.

Выходной диапазон. Вводиться посилання на комірку, з якої почнеться виведення даних.

Новый лист

Відмічається, якщо виведення планується з нового листа в книзі. Виведення почнеться з комірки A1.

Новая книга

Перемикач встановлюється щоб результат було виведено в перший лист нової книги. При виведенні графіка нова книга обов'язкова.

Парето (отсортированная диаграмма). В математичній статистиці ця форма не використовується.

Интегральный процент

Розраховуються значення і виводиться графік для накопичених частот.

Вывод графика. Автоматично будується графік.

Результати приведені на рис.Б.1.4 і Б.1.5.

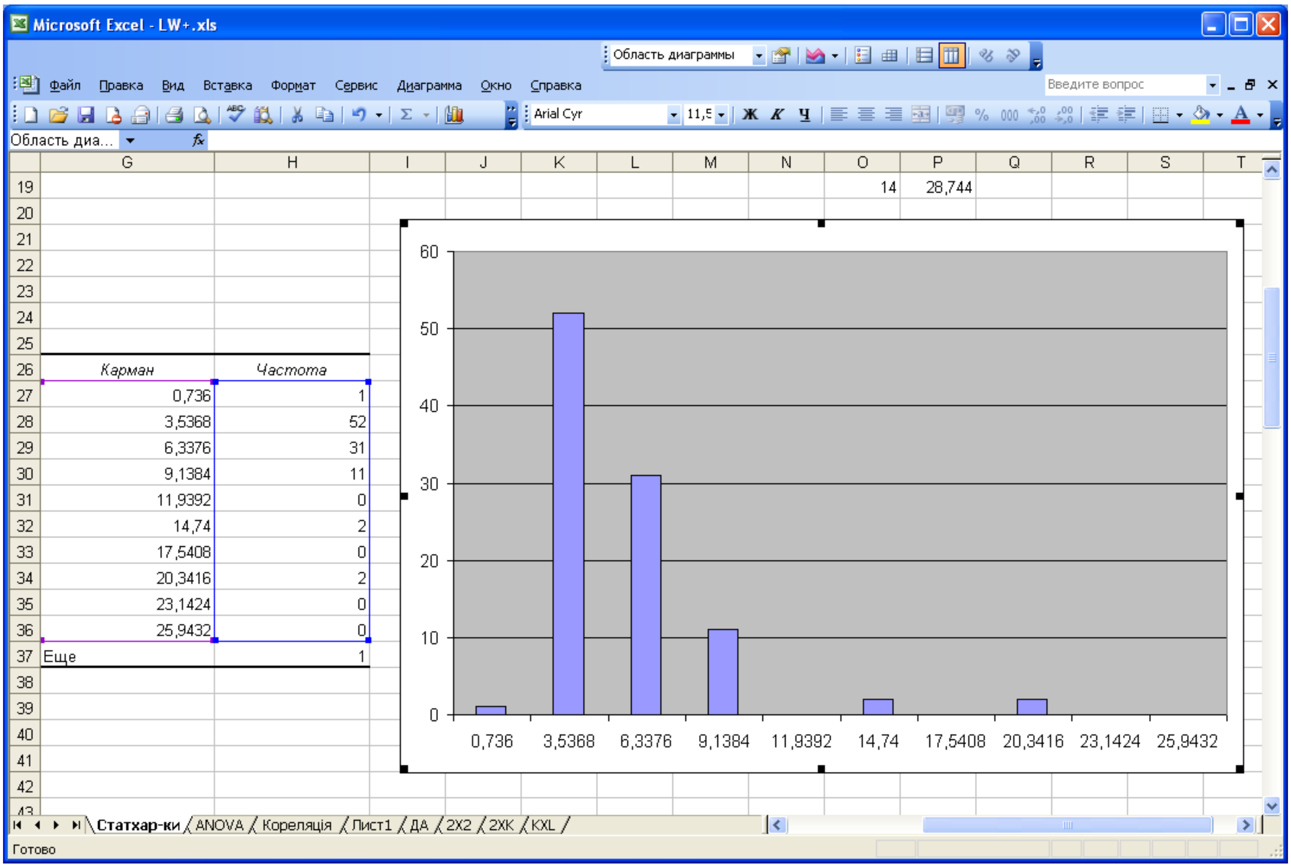


Рис. Б.1.4. Гістограма при автоматичній побудові “карманів”

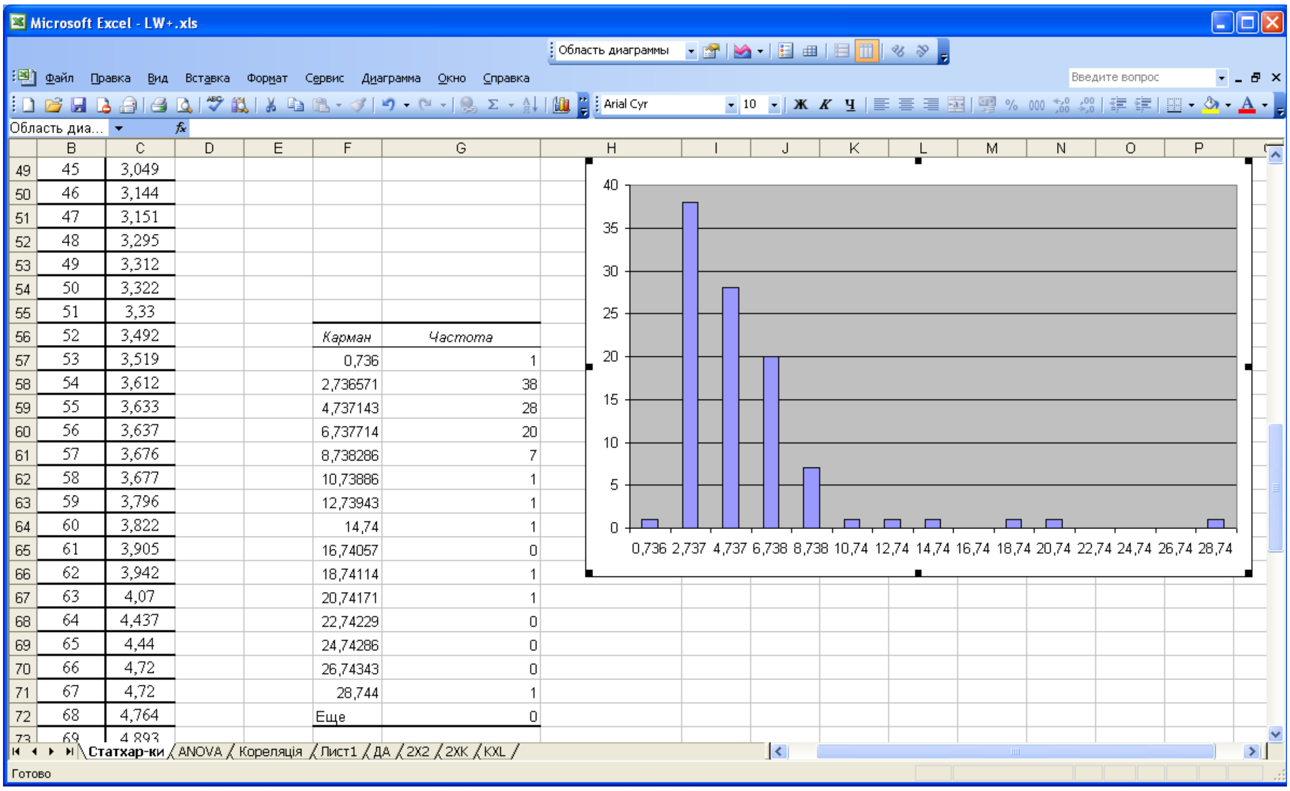


Рис. Б.1.5. Гістограма при заданих “карманах”

Б.2. Лабораторна робота №2

Порівняння дисперсій і середніх

Б.2.1. Параметричні критерії

Проведені випробування стійкості свердел діаметром 6 мм (стійкість в хв.). Одна група свердел звичайні, друга — з потовщеною серцевиною. Необхідно визначити, чи можемо ми вважати, що стійкість цих груп статистично значимо розрізняється.

Таблиця Б.2.1. Результати випробувань двох видів свердел.

Свердла з стовщеною серцевиною	Свердла з нормальною серцевиною
15,68	5,04
18,88	6,48
19,2	7,12
22,56	7,2
24,4	9,44
24,64	11,36
26,56	12,16
27,2	14,24
30,24	15,68
32,16	16,32
33,6	17,84
36,8	18
39,2	21,28
	23,04
	24,6
	30,4

Вихідні дані розміщені в комірках B4:C20 (з заголовками). Спочатку необхідно визначити, чи можемо ми вважати, що закон розподілу обох вибірок нормальний (Гауса). В табл..Б.2.2. показані виклики функцій, які використовуються для визначення закону розподілу і параметрів положення і розсіяння.

Таблиця Б.2.2. Попередні розрахунки

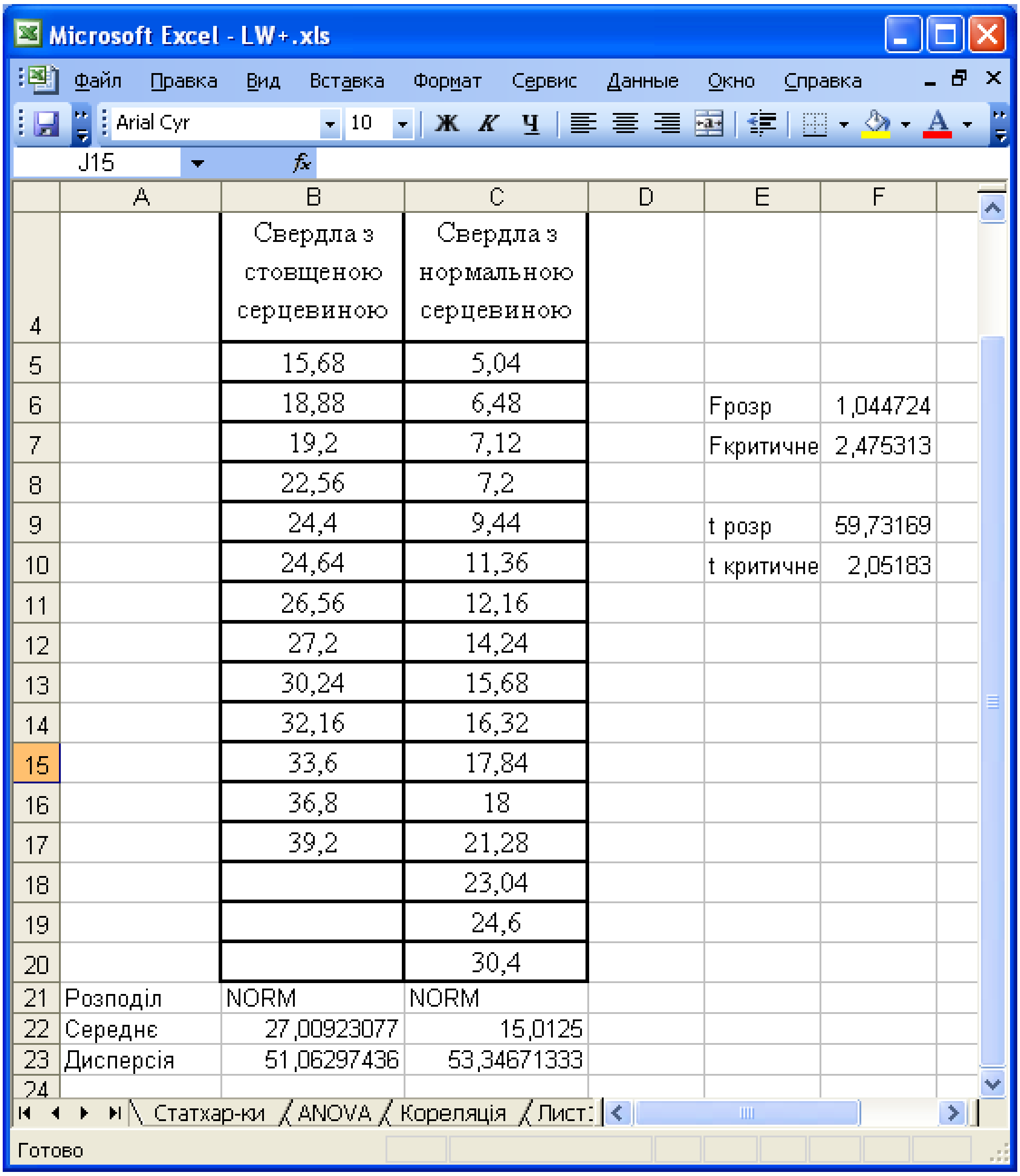
	A	B	C
21	Розподіл	=NORMSAMP_1(B3:B12)	=NORMSAMP_1(C5:C20)
22	Середнє	=СРЗНАЧ(B5:B17)	=СРЗНАЧ(C5:C20)
23	Дисперсія	=ДИСП(B5:B17)	=ДИСП(C5:C20)

Перевірка показала, що ми можемо вважати закон розподілу обох вибірок нормальним. Тому для перевірки гіпотези про середні ми будемо використовувати параметричні критерії. Так як вибірки мають різний розмір, то ми не можемо використати критерій, в якому не враховується співвідношення дисперсій вибірок. Тому перевіряємо гіпотезу про дисперсії (табл..Б.2.3, комірки E6:F7). Так як розрахункове значення критерію Фішера менше критичного, то дисперсії статистично значимо не відрізняються. Тоді використовуємо формули для однакових дисперсій (табл..Б.2.3, комірки E9:F10).

Таблиця Б.2.3. Перевірка гіпотез

	E	F
6	Fрозр	=C23/B23
7	Fкритичне	=FРАСПОБР(0,05;13-1;16-1)
8		
9	t розр	=(B22-C22)/КОРЕНЬ((1/13+1/16)*(B23/12+C23/15)/(13+16-2))
10	t критичне	=СТЬЮДРАСПОБР(0,05;13+16-2)

Вихідні дані і результати представлені на рис.Б.2.1.



	A	B	C	D	E	F
4		Свердла з стовщеною серцевиною	Свердла з нормальною серцевиною			
5		15,68	5,04			
6		18,88	6,48		Fрозр	1,044724
7		19,2	7,12		Fкритичне	2,475313
8		22,56	7,2			
9		24,4	9,44		t розр	59,73169
10		24,64	11,36		t критичне	2,05183
11		26,56	12,16			
12		27,2	14,24			
13		30,24	15,68			
14		32,16	16,32			
15		33,6	17,84			
16		36,8	18			
17		39,2	21,28			
18			23,04			
19			24,6			
20			30,4			
21	Розподіл	NORM	NORM			
22	Середнє	27,00923077	15,0125			
23	Дисперсія	51,06297436	53,34671333			
24						

Рис. Б.2.1. Порівняння середніх. Нормальний закон розподілу

Б.2.2. Непараметричні критерії

Розглянемо випадок використання непараметричного критерію Уілкоксона. В табл..Б.2.4 приведені вихідні дані і необхідні формули і функції.

Таблиця Б.2.4. Виконання дій по перевірки гіпотези з використанням критерію Уілкоксона

Рядок/ Стовпець	A	B	C	D
3	Результати досліджень			
4	Контрольна вибірка	Експериментальна вибірка		
5	27	55	N1=	12
6	41	24	N2=	13
7	34	40	Q=	0,05
8	44	60	Wкрит.н.=	=Wperculc1(D7/2;D5;D6)
9	69	39	Wнабл.=	=W_crit1(A5:A16;B5:B17)
10	38	28	Wкрит.в.=	=D5*(D5+D6+1)-D8
11	24	22		
12	30	37		
13	37	72		
14	59	42		
15	43	33		
16	39	41		
17		25		
18	=NORMSAMP_1(A5:A16)	=NORMSAMP_1(B5:B17)		

Необхідність використання непараметричного критерію викликана тим, що перевірка (комірки A18 і B18) показала, що закон розподілу обох вибірок відмінний від нормального.

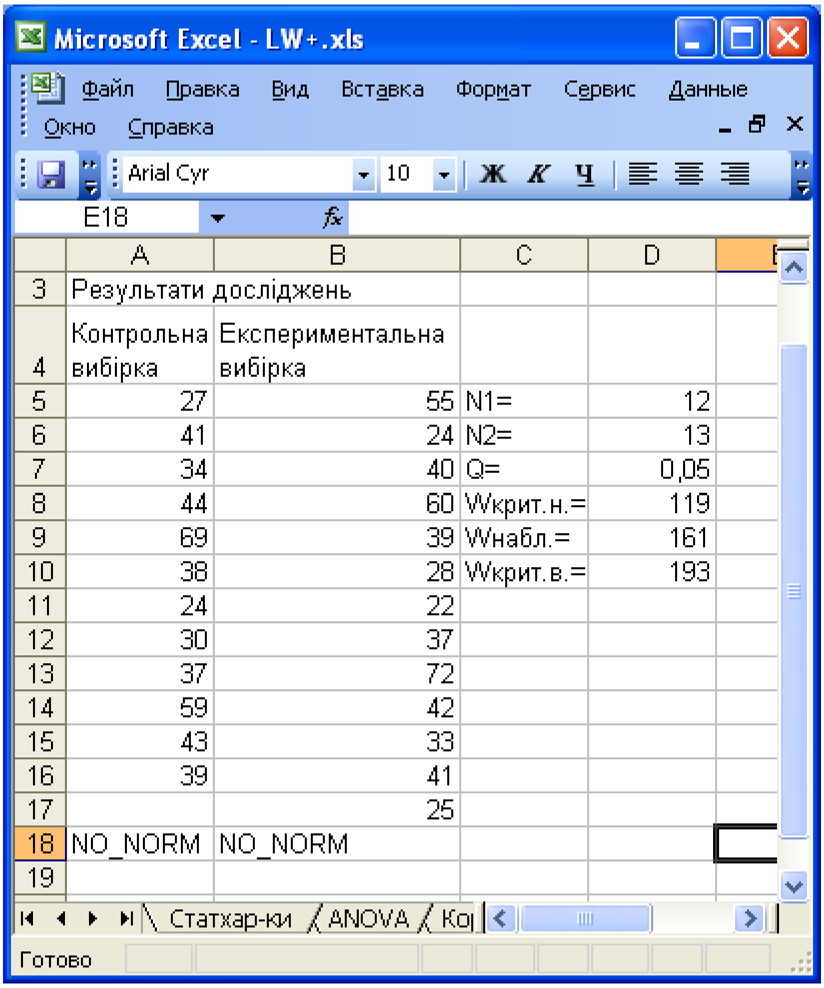
В комірках D5 – D7 розміщені допоміжні дані: розміри обох вибірок і рівень значущості.

В D9 розміщено виклик функції, яка розраховує емпіричне значення критерію Уілкоксона. В D8 – виклик функції для обчислення нижнього критичного значення процентної точки Уілкоксона, а в D10 формула розрахунку верхньої критичної точки через нижню.

На рис. Б.2.2. Показані вихідні дані і результати перевірки гіпотези.

Оскільки емпіричне значення критерію (161) знаходиться між нижнім (119) і верхнім (193) критичними, то нульова гіпотеза про рівність середніх приймається.

Ці середні приймають значення 40,42 і 39,85 відповідно.



	A	B	C	D
3	Результати досліджень			
4	Контрольна вибірка	Експериментальна вибірка		
5	27	55	N1=	12
6	41	24	N2=	13
7	34	40	Q=	0,05
8	44	60	Wкрит.н.=	119
9	69	39	Wнабл.=	161
10	38	28	Wкрит.в.=	193
11	24	22		
12	30	37		
13	37	72		
14	59	42		
15	43	33		
16	39	41		
17		25		
18	NO_NORM	NO_NORM		
19				

Рис.Б.2.2.

Б.3. Лабораторна робота №3 Множинні порівняння середніх

LSD-критерій

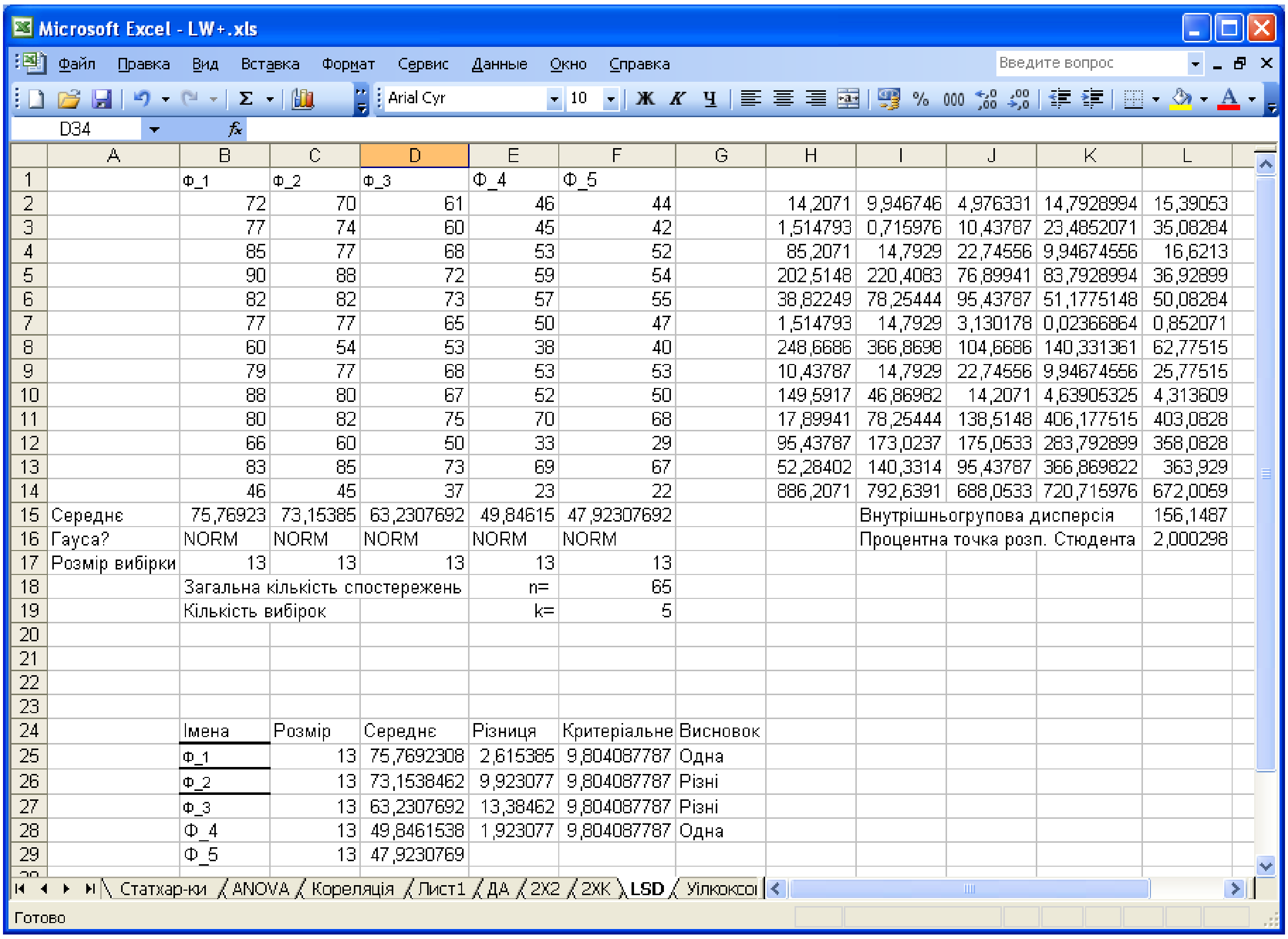
Вихідні дані розміщені в комірках B1:F13. Це п'ять вибірок, які потрібно розбити на групи. Оскільки всі вибірки розподілені за нормальним законом (B16:F16) ми можемо використати LSD-критерій. В комірках B15:F19 розміщені попередні обчислення, відповідно назвам. В комірках H2:L14 знаходяться допоміжні обчислення, необхідні для розрахунку внутрішньо групової дисперсії. Вони отримані введенням формули =(B2-B$15)*(B2-B$15) в комірку H2 і наступним розтягненням її на всі інші комірки. В L15 обчислюється дисперсія =СУММ(H2:L14)/(F18-F19), а в L16 – процентна точка розподілу Стюдента =СТЬЮДРАСПОБР(0,05;F18-F19).

В C24:G29 розміщена таблиця аналізу. Вона приведена для загального випадку вибірок різних розмірів. В третьому стовпці розміщуються значення середніх вибірок, впорядковані по їх зменшенню; в другому – відповідні розміри вибірок, а в першому їх імена. Формули для четвертого – шостого стовпців приведені в табл..Б.6.5 (перший рядок, на інші поширюється розтягненням).

Таблиця Б.2.5. Фрагмент таблиці аналізу з формулами

Різниця	Критеріальне	Висновок
=D26-D27	=L$16*КОРЕНЬ(L$15*(C25+C26)/(C25*C26))	=ЕСЛИ(E25>F25;"Різні";"Одна")

З таблиці видно, що середні розбиваються на наступні групи: (Ф1, Ф2); (Ф3), (Ф4, Ф5).



	A	B	C	D	E	F	G	H	I	J	K	L
1		Ф_1	Ф_2	Ф_3	Ф_4	Ф_5						
2		72	70	61	46	44		14,2071	9,946746	4,976331	14,7928994	15,39053
3		77	74	60	45	42		1,514793	0,715976	10,43787	23,4852071	35,08284
4		85	77	68	53	52		85,2071	14,7929	22,74556	9,94674556	16,6213
5		90	88	72	59	54		202,5148	220,4083	76,89941	83,7928994	36,92899
6		82	82	73	57	55		38,82249	78,25444	95,43787	51,1775148	50,08284
7		77	77	65	50	47		1,514793	14,7929	3,130178	0,02366864	0,852071
8		60	54	53	38	40		248,6686	366,8698	104,6686	140,331361	62,77515
9		79	77	68	53	53		10,43787	14,7929	22,74556	9,94674556	25,77515
10		88	80	67	52	50		149,5917	46,86982	14,2071	4,63905325	4,313609
11		80	82	75	70	68		17,89941	78,25444	138,5148	406,177515	403,0828
12		66	60	50	33	29		95,43787	173,0237	175,0533	283,792899	358,0828
13		83	85	73	69	67		52,28402	140,3314	95,43787	366,869822	363,929
14		46	45	37	23	22		886,2071	792,6391	688,0533	720,715976	672,0059
15	Середнє	75,76923	73,15385	63,2307692	49,84615	47,92307692			Внутрішньогрупова дисперсія			156,1487
16	Гауса?	NORM	NORM	NORM	NORM	NORM			Процентна точка розп. Стюдента			2,000298
17	Розмір вибірки	13	13	13	13	13						
18		Загальна кількість спостережень			n=	65						
19		Кількість вибірок			k=	5						
20												
21												
22												
23												
24		Імена	Розмір	Середнє	Різниця	Критеріальне	Висновок					
25		Ф_1	13	75,7692308	2,615385	9,804087787	Одна					
26		Ф_2	13	73,1538462	9,923077	9,804087787	Різні					
27		Ф_3	13	63,2307692	13,38462	9,804087787	Різні					
28		Ф_4	13	49,8461538	1,923077	9,804087787	Одна					
29		Ф_5	13	47,9230769								

Рис. Б.2.3. Множинні порівняння з використанням LSD-критерію

Б.3.2. Критерій Шеффе

На рис. Б.2.4 приведена вся інформація (вихідні дані, допоміжні розрахунки і результати) перевірки гіпотез про середні за критерієм Шеффе. Комірки A21: F31 місця допоміжні дані для перевірки: 1 і -1 для тих вибірок, які ми порівнюємо і 0 для всіх інших. В M16 – процентна точка розподілу Фішера =FРАСПОБР(0,05;F18-1;F18-F17).

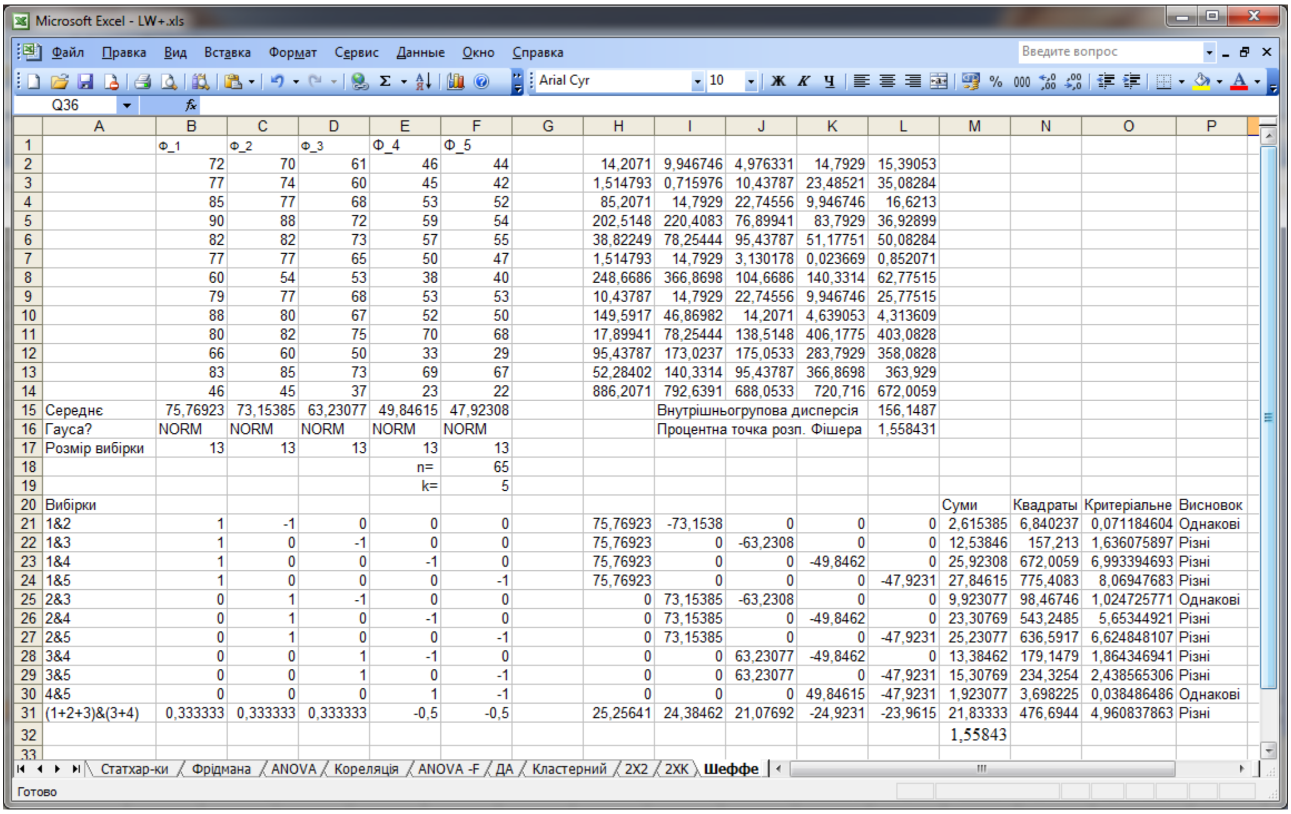
В H21 : L31 розміщені попередні розрахунки. Для їх виконання в комірку H21 потрібно помістити формулу =B21*B$15 і розтягнути її на всі інші комірки вказаного діапазону.

В M21:P31 розміщена таблиця аналізу. Вона приведена для загального випадку вибірок різних розмірів. Формули приведені в табл..Б.2.6 (перший рядок, на інші поширюється розтягненням).

Таблиця Б.2.6. Фрагмент таблиці аналізу з формулами для критерію Шеффе

Суми	Квадрати	Критеріальне	Висновок
=СУММ(H21:L21)	=M21*M21	=N21/((F19-1)*L15*2/B$17)	=ЕСЛИ(O21<M32;"Однакові";"Різні")

Тут приведені всі варіанти можливих перевірок. Використовуватись одночасно вони не можуть – тільки один варіант.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		Ф_1	Ф_2	Ф_3	Ф_4	Ф_5										
2		72	70	61	46	44		14,2071	9,946746	4,976331	14,7929	15,39053				
3		77	74	60	45	42		1,514793	0,715976	10,43787	23,48521	35,08284				
4		85	77	68	53	52		85,2071	14,7929	22,74556	9,946746	16,6213				
5		90	88	72	59	54		202,5148	220,4083	76,89941	83,7929	36,92899				
6		82	82	73	57	55		38,82249	78,25444	95,43787	51,17751	50,08284				
7		77	77	65	50	47		1,514793	14,7929	3,130178	0,023669	0,852071				
8		60	54	53	38	40		248,6686	366,8698	104,6686	140,3314	62,77515				
9		79	77	68	53	53		10,43787	14,7929	22,74556	9,946746	25,77515				
10		88	80	67	52	50		149,5917	46,86982	14,2071	4,639053	4,313609				
11		80	82	75	70	68		17,89941	78,25444	138,5148	406,1775	403,0828				
12		66	60	50	33	29		95,43787	173,0237	175,0533	283,7929	358,0828				
13		83	85	73	69	67		52,28402	140,3314	95,43787	366,8698	363,929				
14		46	45	37	23	22		886,2071	792,6391	688,0533	720,716	672,0059				
15	Середнє	75,76923	73,15385	63,23077	49,84615	47,92308			Внутрішньогрупова дисперсія			156,1487				
16	Гауса?	NORM	NORM	NORM	NORM	NORM			Процентна точка розп. Фішера			1,558431				
17	Розмір вибірки	13	13	13	13	13										
18					n=	65										
19					k=	5										
20	Вибірки												Суми	Квадраты	Критеріальне	Висновок
21	1&2	1	-1	0	0	0		75,76923	-73,1538	0	0	0	2,615385	6,840237	0,071184604	Однакові
22	1&3	1	0	-1	0	0		75,76923	0	-63,2308	0	0	12,53846	157,213	1,636075897	Різні
23	1&4	1	0	0	-1	0		75,76923	0	0	-49,8462	0	25,92308	672,0059	6,993394693	Різні
24	1&5	1	0	0	0	-1		75,76923	0	0	0	-47,9231	27,84615	775,4083	8,06947683	Різні
25	2&3	0	1	-1	0	0		0	73,15385	-63,2308	0	0	9,923077	98,46746	1,024725771	Однакові
26	2&4	0	1	0	-1	0		0	73,15385	0	-49,8462	0	23,30769	543,2485	5,65344921	Різні
27	2&5	0	1	0	0	-1		0	73,15385	0	0	-47,9231	25,23077	636,5917	6,624848107	Різні
28	3&4	0	0	1	-1	0		0	0	63,23077	-49,8462	0	13,38462	179,1479	1,864346941	Різні
29	3&5	0	0	1	0	-1		0	0	63,23077	0	-47,9231	15,30769	234,3254	2,438565306	Різні
30	4&5	0	0	0	1	-1		0	0	0	49,84615	-47,9231	1,923077	3,698225	0,038486486	Однакові
31	(1+2+3)&(3+4)	0,333333	0,333333	0,333333	-0,5	-0,5		25,25641	24,38462	21,07692	-24,9231	-23,9615	21,83333	476,6944	4,960837863	Різні
32													1,55843			
33																

Рис. Б.2.4. Множинні порівняння за критерієм Шеффе

Останній варіант (31-й рядок), коли ми порівнюємо всі вибірки одночасно. При цьому вважаємо, що 1, 2 і 3 належать до однієї генеральної сукупності, а 4 і 5 до іншої. Тоді коефіцієнти мають значення 1/3, 1/3, 1/3, -1/2, -1/2.

Б.4. Лабораторна робота №4. Кореляційний аналіз

Б.4.1. Коефіцієнт парної кореляції Пірсона

Нам потрібно визначити залежність між границею обмеженої витривалості при вигині з обертанням зразків на базі 10^7 циклів і границею міцності для алюмінієвих сплавів (див. табл..Б.4.1).

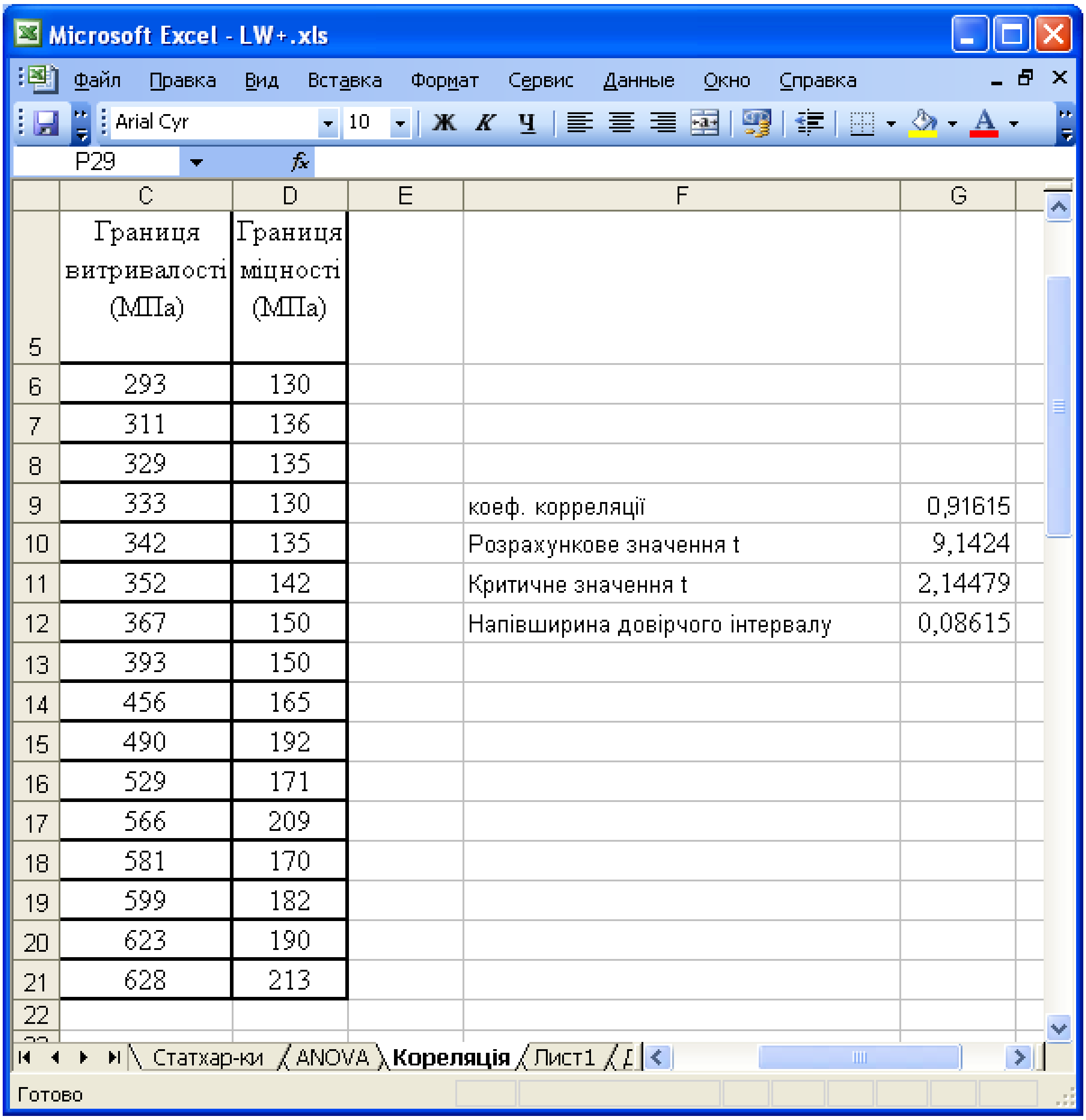
Таблиця Б.4.1. Результати випробувань границь витривалості та міцності

Границя витривалості (МПа)	Границя міцності (МПа)
293	130
311	136
329	135
333	130
342	135
352	142
367	150
393	150
456	165
490	192
529	171
566	209
581	170
599	182
623	190
628	213

Набираємо вихідні дані (див. рис.Б4.1). Для отримання коефіцієнта кореляції набираємо функцію =КОРРЕЛ(C6:C21;D6:D21) в комірці G9. Щоб проаналізувати значимість коефіцієнта кореляції розрахуємо емпіричне значення t-критерію. Для цього в комірці G10 наберемо функцію =G9*КОРЕНЬ(ЧСТРОК(C6:C21))/КОРЕНЬ(1-G9^2). Критичне значення t-критерію визначимо за формулою =СТЬЮДРАСПОБР(0,05;ЧСТРОК(C6:C21)-2), поміщеною в комірці G11. Оскільки розрахункове значення більше критичного, то коефіцієнт кореляції статистично значимий.

В деяких випадках необхідно знайти довірчий інтервал для коефіцієнта кореляції. Для визначення його на півширини скористаємось формулою =G11*(1-G9^2)/КОРЕНЬ(ЧСТРОК(C6:C21)).

Результати розрахунків приведені на рис. Б.4.1.



	C	D	E	F	G
5	Границя витривалості (МПа)	Границя міцності (МПа)			
6	293	130			
7	311	136			
8	329	135			
9	333	130		коеф. корреляції	0,91615
10	342	135		Розрахункове значення t	9,1424
11	352	142		Критичне значення t	2,14479
12	367	150		Напівширина довірчого інтервалу	0,08615
13	393	150			
14	456	165			
15	490	192			
16	529	171			
17	566	209			
18	581	170			
19	599	182			
20	623	190			
21	628	213			

Рис.Б.4.1. Вихідні дані і розрахунки для аналізу коефіцієнта парної кореляції.

В тому випадку, коли необхідно знайти коефіцієнти кореляції одночасно для групи змінних (з однаковою кількістю експериментів), можливо скористатись спеціальною функцією.

Після підготовки вихідних даних вибираємо в меню выбираем ***Сервис***, а потім ***Анализ данных***. З'явиться вікно вибору методу обробки даних (рис. Б.4.2).

В цьому вікні вибираємо ***Корреляция*** — з'явиться вікно вводу даних для кореляції (рис. Б.4.3).

В ньому необхідно задати дані для розрахунку коефіцієнтів кореляції.

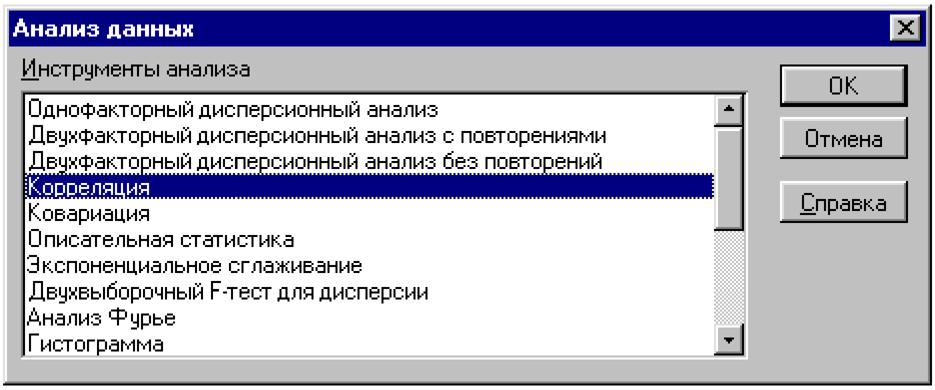


Рис. Б.4.2. Вікно вибору методу обробки даних

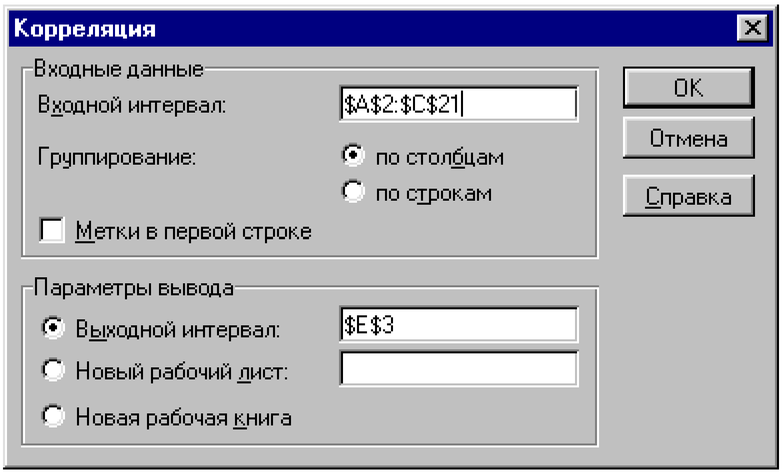
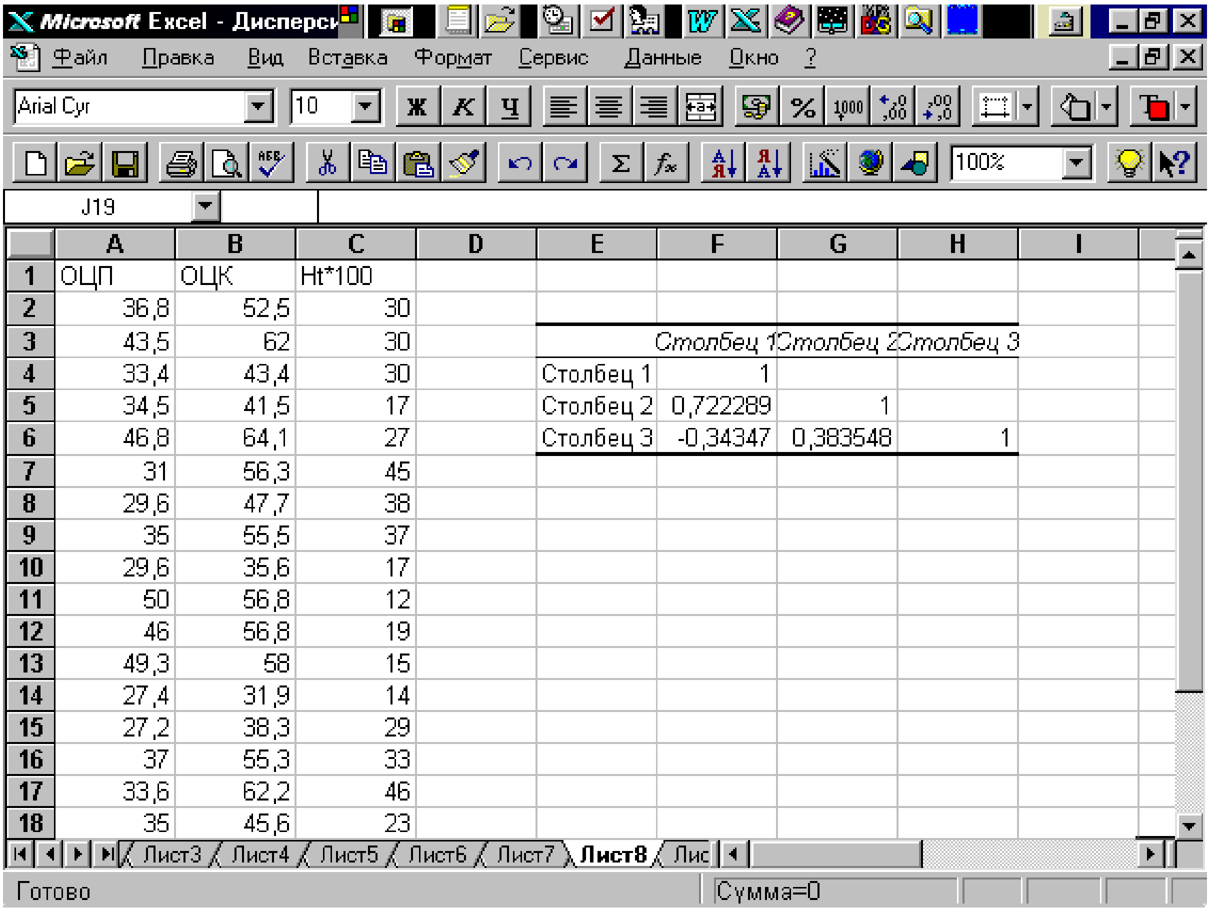


Рис. Б.4.3. Вікно вводу даних для кореляції

Входной интервал — необхідно відмітити таблицю, в якій розміщені вихідні дані (ліва верхня і права нижня комірки). *Группирование* — необхідно вказати в рядках чи стовбцях розміщені дані. *Выходной интервал* — виводиться посилання на комірку, яка знаходиться в лівому верхньому кутку вихідного діапазону (місце виведення результату). Розміри вихідної області будуть розраховані автоматично. *Новый рабочий лист* — вибирається в тому випадку, коли ви хочете помістити результат на інший робочий лист. *Новая рабочая книга* — вибирається у випадку розміщення результату в новій книзі.

Після встановлення параметрів і натиснення на ОК отримуємо матрицю коефіцієнтів парної кореляції (приклад див. на рис. Б.4.4).



	A	B	C	D	E	F	G	H
1	ОЦП	ОЦК	Ht*100					
2	36,8	52,5	30					
3	43,5	62	30			Столбец 1	Столбец 2	Столбец 3
4	33,4	43,4	30		Столбец 1	1		
5	34,5	41,5	17		Столбец 2	0,722289	1	
6	46,8	64,1	27		Столбец 3	-0,34347	0,383548	1
7	31	56,3	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					
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					

Рис. Б.4.4. Матриця кореляції, розрахована функцією ***Корреляция (Анализ данных)***

Б.4.2. Рангова кореляція

Розглянемо визначення тісноти і значимості між змінними, інформація про які приведена в табл.Б.4.2.

Оскільки карбідна неоднорідність визначається за шкалою рангів, то в даному випадку використовуються коефіцієнти рангової кореляції.

У тому випадку, коли немає такої інформації про вихідні дані, використовується функція перевірки належності розподілу даних до нормального NORMSAMP_1(R_1). Ця функція виводить результатом NO_NORM, або NORM.

Таблиця Б.4.2.Залежність між діаметром прокату швидкорізальної сталі Р18 і її карбідною неоднорідністю.

Діаметр в мм	Карбідна неоднорідність в балах
20	3
21	3
25	2
40	4
40	3
40	3
42	4

42	4
42	5
42	5
42	5
45	4
45	4
45	4
45	4
45	4
48	5
50	4
50	4
50	4
52	5
52	5
52	5
52	5
52	4

Обчислення рангових коефіцієнтів кореляції зручніше виконувати за допомогою спеціальних функцій.

Вибираємо послідовно пункти меню «Сервис», «Макрос», «Макросы». В вікні, які з'явиться (рис. Б.4.5) выберем пункт Candall_cor або Spirmen_cor, в залежності від вибраного коефіцієнта кореляції.

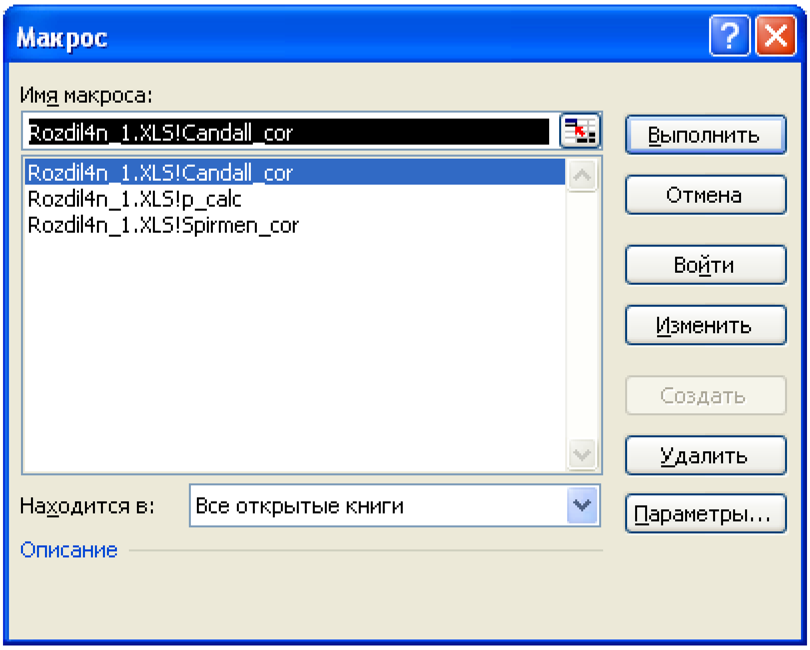


Рис. Б.4.5. Вікно вибору макросу

Після цього послідовно відповідаємо на питання в вікнах запиту. Спочатку вводимо діапазон комірок, в яких знаходяться вихідні дані (рис. Б.4.6).

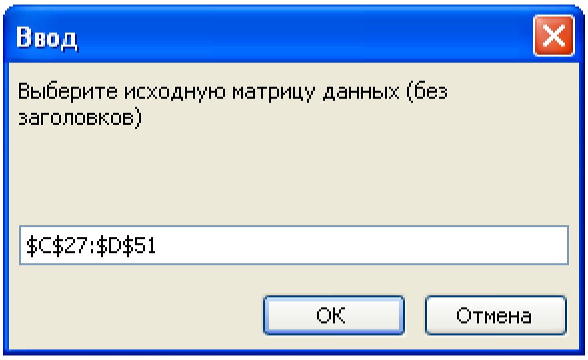


Рис. Б.4.6 Запит на діапазон вхідних даних

Після цього вводяться посилання на комірки, в яких знаходяться імена змінних (рис. Б.4.7).

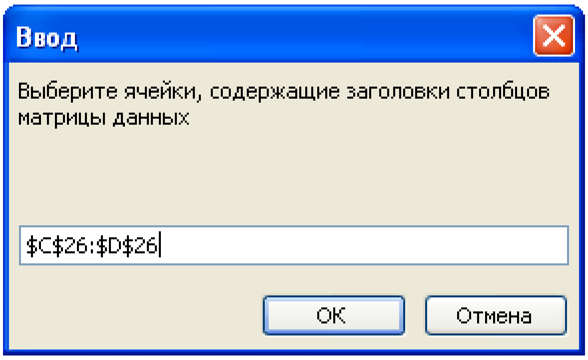


Рис. Б.4.7. Введення імен змінних

Потів вказується місце виведення результату (рис. Б.4.8).

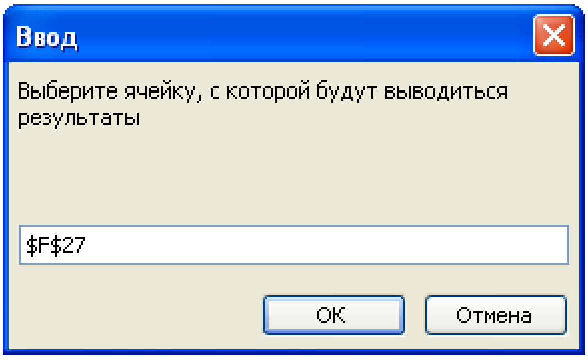
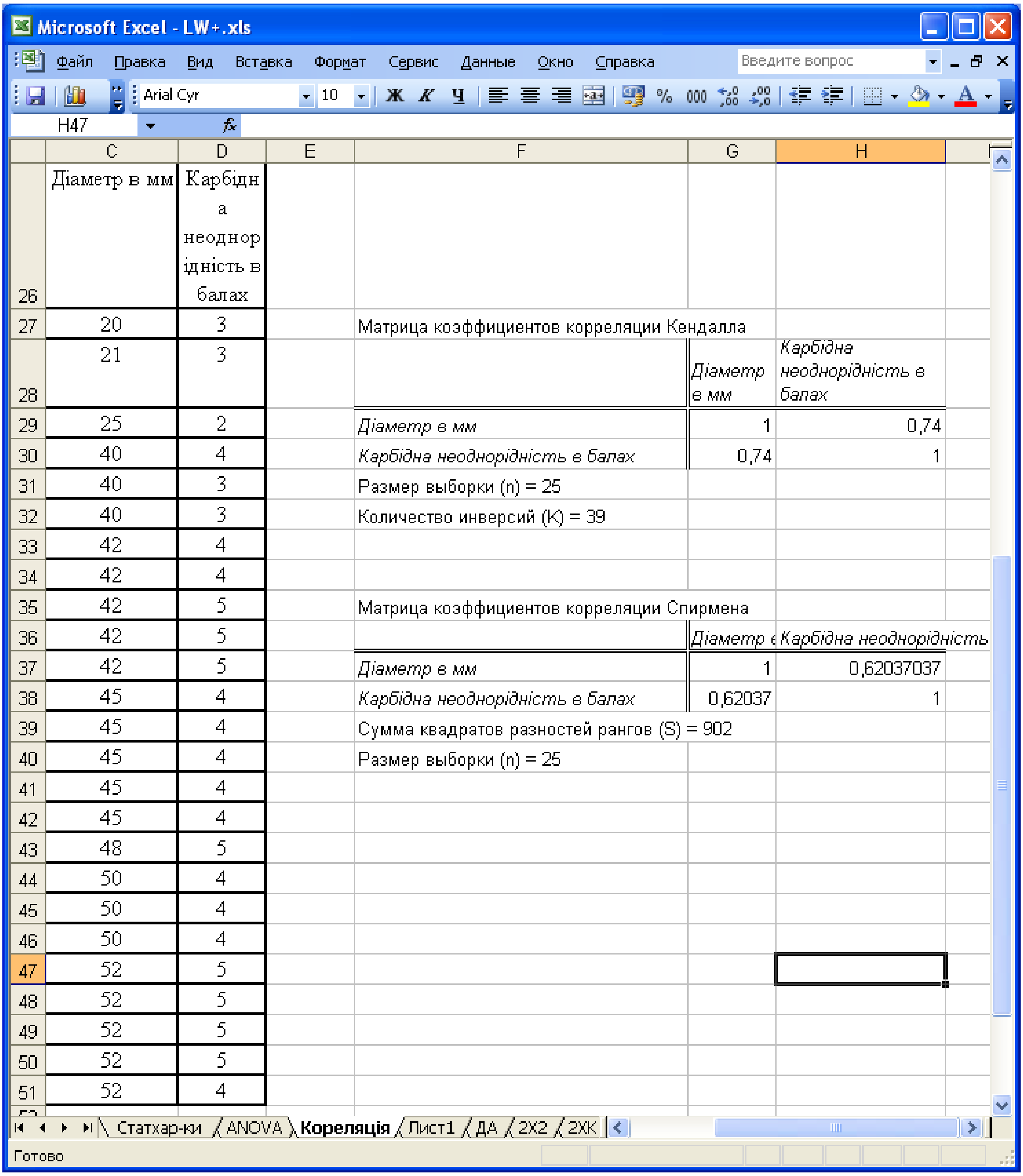


Рис. Б.4.8

Результати роботи приведені на рис. Б.4.9.

Перевірка значимості коефіцієнта кореляції Спірмена може виконуватись аналогічно коефіцієнту кореляції Пірсона. Для перевірки значимості коефіцієнта кореляції Кендала необхідно скористатись спеціальними таблицями (наприклад Додаток Д в [2]). Для рівня значущості α = 0,05 і існуючого розміру вибірки критичне значення 0,24. Це означає, що коефіцієнт кореляції статистично значущий.



Microsoft Excel - LW+.xls

	C	D
26	Діаметр в мм	Карбідна неоднорідність в балах
27	20	3
28	21	3
29	25	2
30	40	4
31	40	3
32	40	3
33	42	4
34	42	4
35	42	5
36	42	5
37	42	5
38	45	4
39	45	4
40	45	4
41	45	4
42	45	4
43	48	5
44	50	4
45	50	4
46	50	4
47	52	5
48	52	5
49	52	5
50	52	5
51	52	4

Матрица коэффициентов корреляции Кендалла

	Діаметр в мм	*Карбідна неоднорідність в балах*
Діаметр в мм	1	0,74
Карбідна неоднорідність в балах	0,74	1

Размер выборки (n) = 25

Количество инверсий (K) = 39

Матрица коэффициентов корреляции Спирмена

	Діаметр в	*Карбідна неоднорідність*
Діаметр в мм	1	0,62037037
Карбідна неоднорідність в балах	0,62037	1

Сумма квадратов разностей рангов (S) = 902

Размер выборки (n) = 25

Статхар-ки / ANOVA / Кореляція / Лист1 / ДА / 2Х2 / 2ХК

Готово

Рис. Б.4.9. Вихідні дані і результати розрахунків рангових коефіцієнтів кореляції

Б.4.3. Коефіцієнт конкордації.

Часто використовується для визначення питання про узгодженість експертних висновків.

Вихідні дані готуються в вигляді представленому в табл..Б.4.3. В ній знаходиться інформація про оцінки, які поставили за десятибальною шкалою вісім експертів в 32 спостереженнях.

Таблиця Б.4.3. Вихідні дані для конкордації

№ спостереження	Експерти							
	1	2	3	4	5	6	7	8
1	5	4	4	5	5	5	5	4
2	4	3	3	3	4	4	4	4
3	4	4	3	4	4	4	4	4
4	3	4	3	3	2	3	3	2
5	2	3	3	3	3	2	3	2
6	3	3	2	3	3	3	3	3
7	5	4	4	4	3	4	4	4
8	4	3	3	2	3	3	3	2
9	4	3	4	4	3	4	4	3
10	3	3	3	2	3	3	3	2
11	4	4	4	3	3	4	3	4
12	3	3	3	2	3	3	3	2
13	4	4	3	4	3	4	3	4
14	4	3	4	4	3	4	4	4
15	4	4	3	3	4	4	3	4
16	4	3	3	3	3	3	4	3
17	4	3	4	3	4	4	3	4
18	5	5	4	4	4	5	5	5
19	4	3	4	4	3	4	4	4
20	4	3	4	4	4	4	4	4
21	5	4	5	5	5	4	5	4
22	4	3	3	4	4	4	3	4
23	4	3	3	4	4	4	3	4
24	4	3	3	4	3	4	4	4
25	4	4	5	4	5	5	4	5
26	4	4	3	3	5	3	3	4
27	3	2	2	2	3	3	3	3
28	3	2	3	2	3	3	2	3
29	4	3	3	4	4	4	3	4
30	3	3	2	2	3	3	3	2
31	4	3	3	4	4	5	3	3
32	4	5	5	5	5	5	5	4

Після цього необхідно на основі цієї таблиці побудувати таблицю рангів. В макросі обчислення конкордації в якості вихідних даних використовується матриця рангів. Для запуску макросу послідовно вибираються пункти меню «Сервис»–«Макрос»–«Макроси». В вікні, що з’явилось (рис.Б.4.10) вибирається макрос з іменем rang. Потім вводиться вихідна матриця (рис.Б.4.11) і місце виведення результатів (рис.Б.4.12).

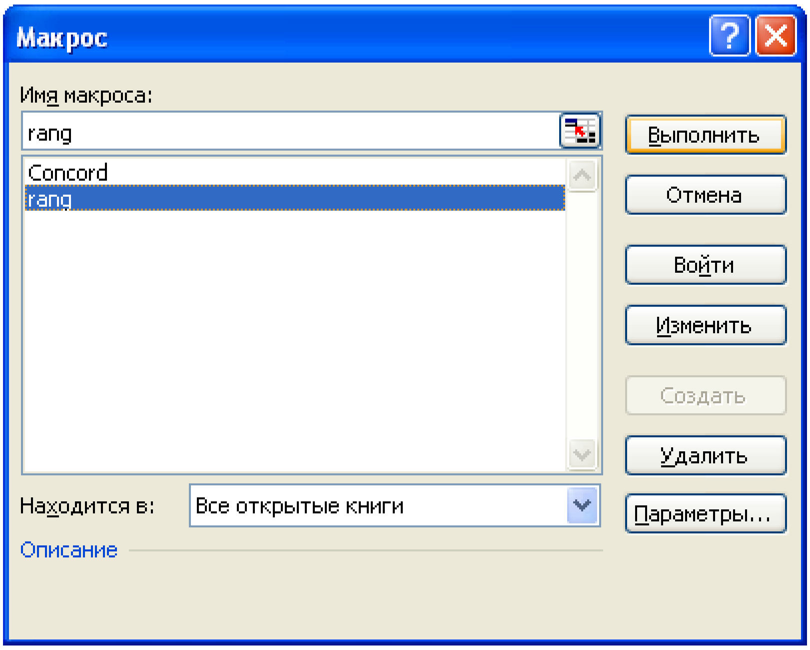


Рис. Б.4.10. Вікно вибору макросу

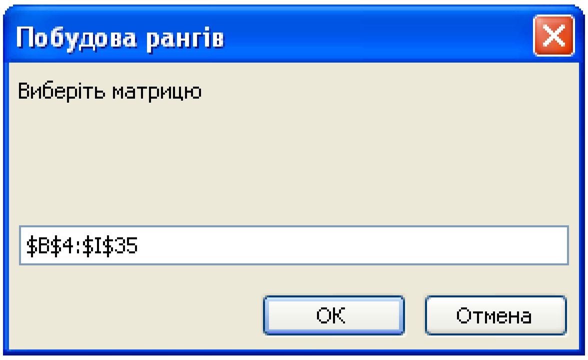


Рис. Б.4.11. Запит на введення вихідної матриці

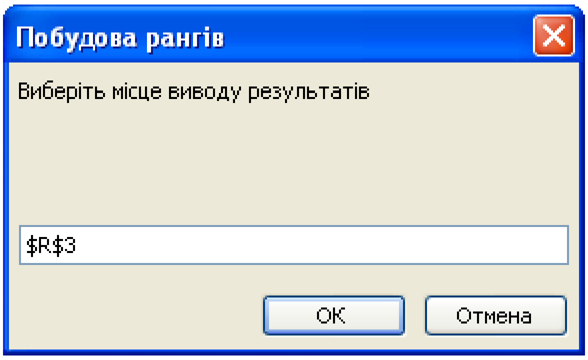


Рис. Б.4.12. Запит місця виведення результатів

Результатом роботи макросу буде матриця рангів (див. табл..Б.4.4).

Для запуску макросу послідовно вибираються пункти меню «Сервис»–«Макрос»–«Макроси». В вікні, що з'явилось (рис.Б.4.13) вибирається макрос з іменем Concord. На запит макросу (Б.4.14) необхідно ввести матрицю рангів (див. рис. Б.4.14.) без заголовків.

Якщо ви хочете змінити стандартне значення рівня значущості, то це можна зробити в наступному запиті макросу (Б.4.15). В противному разі вибираєте ОК, або натискаєте Enter.

В випадку введення помилкового значення після повідомлення про помилку (рис. Б.4.16) необхідно повторити введення (рис. Б.4.15). Наостанок необхідно відмітити комірку, з якої почнеться виведення результатів (рис. Б.4.17).

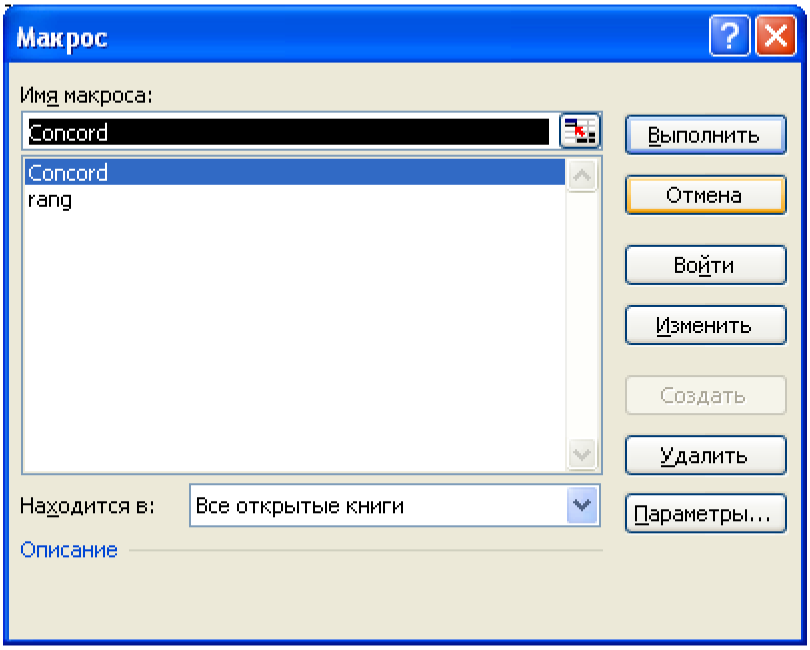


Рис. Б.4.13.. Вікно запуску макросу

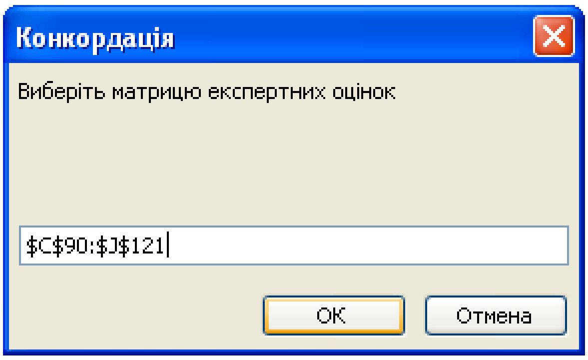


Рис. Б.4.14. Запит на введення матриці вихідних даних

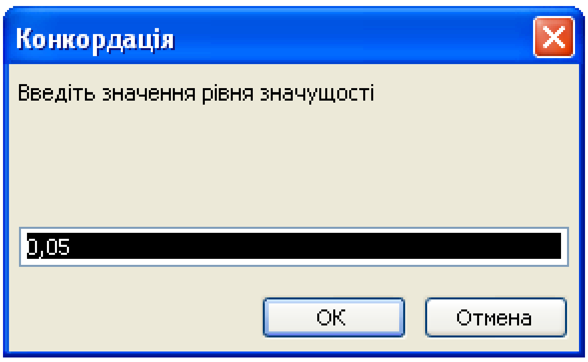


Рис. Б.4.15. Запит на зміну рівня значущості

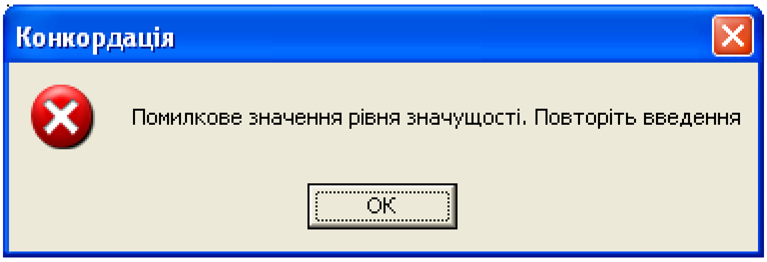


Рис. Б.4.16. Повідомлення про помилку при введенні рівня значущості

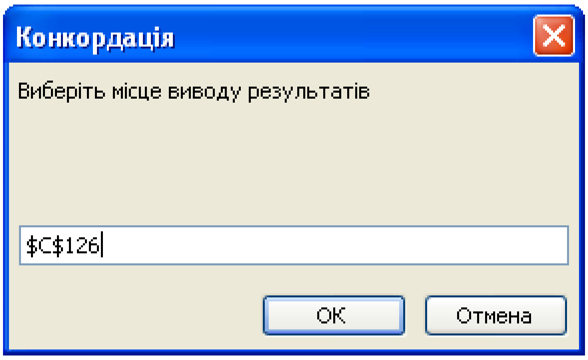


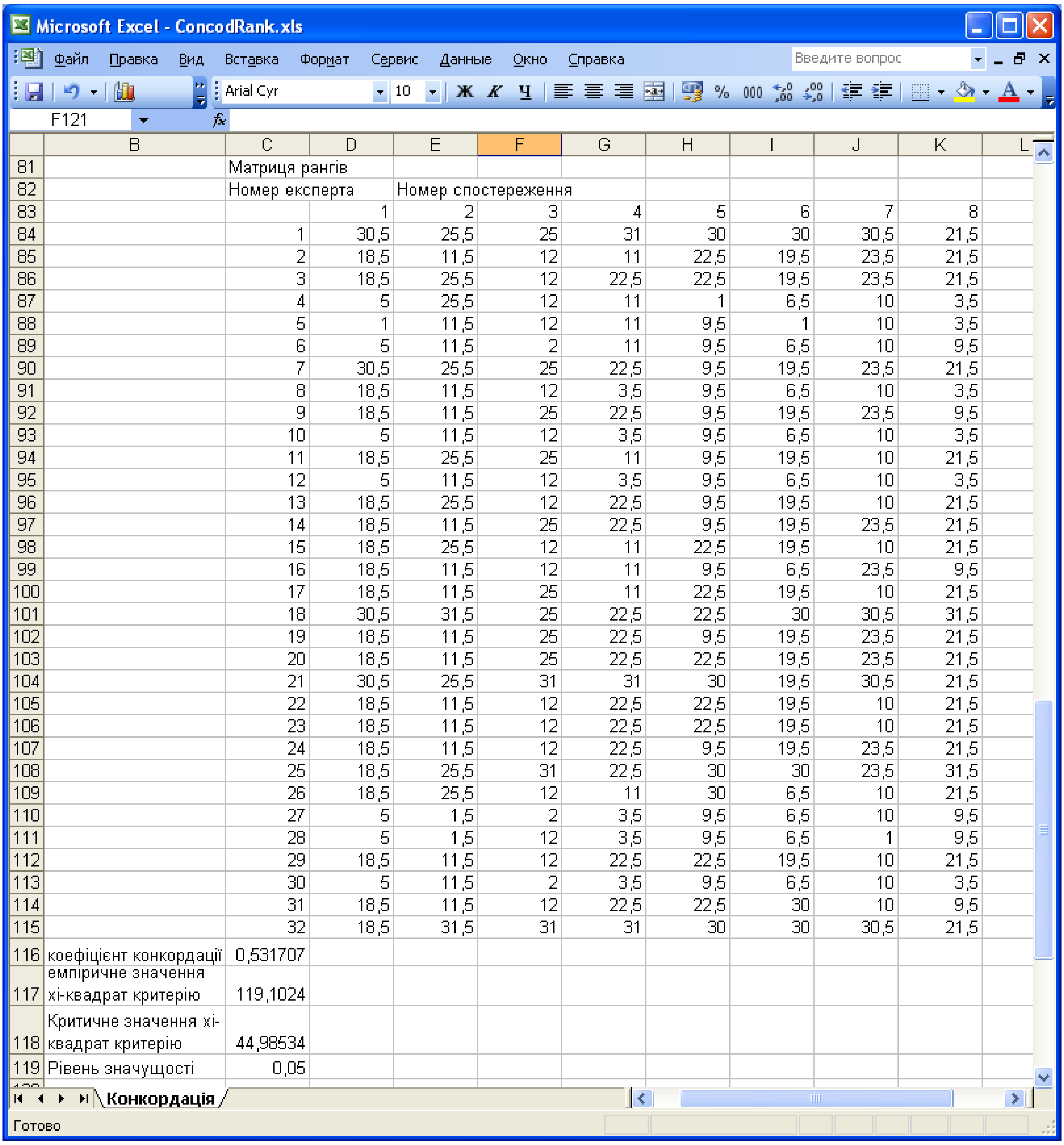
Рис. Б.4.17. Запит місця введення результатів

Таблиця Б.4.4. Розрахована матриця рангів

Матриця рангів								
Номер експерта	Номер спостереження							
	1	2	3	4	5	6	7	8
1	30,5	25,5	25	31	30	30	30,5	21,5
2	18,5	11,5	12	11	22,5	19,5	23,5	21,5
3	18,5	25,5	12	22,5	22,5	19,5	23,5	21,5
4	5	25,5	12	11	1	6,5	10	3,5
5	1	11,5	12	11	9,5	1	10	3,5
6	5	11,5	2	11	9,5	6,5	10	9,5
7	30,5	25,5	25	22,5	9,5	19,5	23,5	21,5
8	18,5	11,5	12	3,5	9,5	6,5	10	3,5
9	18,5	11,5	25	22,5	9,5	19,5	23,5	9,5
10	5	11,5	12	3,5	9,5	6,5	10	3,5
11	18,5	25,5	25	11	9,5	19,5	10	21,5
12	5	11,5	12	3,5	9,5	6,5	10	3,5
13	18,5	25,5	12	22,5	9,5	19,5	10	21,5
14	18,5	11,5	25	22,5	9,5	19,5	23,5	21,5
15	18,5	25,5	12	11	22,5	19,5	10	21,5
16	18,5	11,5	12	11	9,5	6,5	23,5	9,5
17	18,5	11,5	25	11	22,5	19,5	10	21,5
18	30,5	31,5	25	22,5	22,5	30	30,5	31,5
19	18,5	11,5	25	22,5	9,5	19,5	23,5	21,5
20	18,5	11,5	25	22,5	22,5	19,5	23,5	21,5
21	30,5	25,5	31	31	30	19,5	30,5	21,5
22	18,5	11,5	12	22,5	22,5	19,5	10	21,5
23	18,5	11,5	12	22,5	22,5	19,5	10	21,5
24	18,5	11,5	12	22,5	9,5	19,5	23,5	21,5
25	18,5	25,5	31	22,5	30	30	23,5	31,5
26	18,5	25,5	12	11	30	6,5	10	21,5
27	5	1,5	2	3,5	9,5	6,5	10	9,5
28	5	1,5	12	3,5	9,5	6,5	1	9,5
29	18,5	11,5	12	22,5	22,5	19,5	10	21,5

30	5	11,5	2	3,5	9,5	6,5	10	3,5
31	18,5	11,5	12	22,5	22,5	30	10	9,5
32	18,5	31,5	31	31	30	30	30,5	21,5

Результат роботи має вигляд, представлений на рис. Б.4.18. Якщо коефіцієнт конкордації статистично незначущий, то в останньому рядку виводиться «Не узгоджені».



Microsoft Excel - ConcodRank.xls

	B	C	D	E	F	G	H	I	J	K
81		Матриця рангів								
82		Номер експерта		Номер спостереження						
83			1	2	3	4	5	6	7	8
84		1	30,5	25,5	25	31	30	30	30,5	21,5
85		2	18,5	11,5	12	11	22,5	19,5	23,5	21,5
86		3	18,5	25,5	12	22,5	22,5	19,5	23,5	21,5
87		4	5	25,5	12	11	1	6,5	10	3,5
88		5	1	11,5	12	11	9,5	1	10	3,5
89		6	5	11,5	2	11	9,5	6,5	10	9,5
90		7	30,5	25,5	25	22,5	9,5	19,5	23,5	21,5
91		8	18,5	11,5	12	3,5	9,5	6,5	10	3,5
92		9	18,5	11,5	25	22,5	9,5	19,5	23,5	9,5
93		10	5	11,5	12	3,5	9,5	6,5	10	3,5
94		11	18,5	25,5	25	11	9,5	19,5	10	21,5
95		12	5	11,5	12	3,5	9,5	6,5	10	3,5
96		13	18,5	25,5	12	22,5	9,5	19,5	10	21,5
97		14	18,5	11,5	25	22,5	9,5	19,5	23,5	21,5
98		15	18,5	25,5	12	11	22,5	19,5	10	21,5
99		16	18,5	11,5	12	11	9,5	6,5	23,5	9,5
100		17	18,5	11,5	25	11	22,5	19,5	10	21,5
101		18	30,5	31,5	25	22,5	22,5	30	30,5	31,5
102		19	18,5	11,5	25	22,5	9,5	19,5	23,5	21,5
103		20	18,5	11,5	25	22,5	22,5	19,5	23,5	21,5
104		21	30,5	25,5	31	31	30	19,5	30,5	21,5
105		22	18,5	11,5	12	22,5	22,5	19,5	10	21,5
106		23	18,5	11,5	12	22,5	22,5	19,5	10	21,5
107		24	18,5	11,5	12	22,5	9,5	19,5	23,5	21,5
108		25	18,5	25,5	31	22,5	30	30	23,5	31,5
109		26	18,5	25,5	12	11	30	6,5	10	21,5
110		27	5	1,5	2	3,5	9,5	6,5	10	9,5
111		28	5	1,5	12	3,5	9,5	6,5	1	9,5
112		29	18,5	11,5	12	22,5	22,5	19,5	10	21,5
113		30	5	11,5	2	3,5	9,5	6,5	10	3,5
114		31	18,5	11,5	12	22,5	22,5	30	10	9,5
115		32	18,5	31,5	31	31	30	30	30,5	21,5
116	коефіцієнт конкордації	0,531707								
117	емпіричне значення хі-квадрат критерію	119,1024								
118	Критичне значення хі-квадрат критерію	44,98534								
119	Рівень значущості	0,05								

Б.4.18.Результат роботи макросу розрахунку коефіцієнта конкордації

Значимість коефіцієнтів конкордації при малій кількості експертів виконується за допомогою спеціальної таблиці, фрагмент якої приведено в табл..Б.4.5.

Таблиця Б.4.5. Дані для розрахунку значимості коефіцієнта конкордації

N=3; m=10		N=5; m=3	
50	0,092	56	0,096

62	0,046	62	0,056
104	0,0034	78	0,053
126	0,0008	86	0,0009

Для отримання критичного значення коефіцієнта конкордації необхідно взяте з таблиця значення підставити в формулу $12*S/m^2(n^3–n)$. Якщо ж число експертів більше 7, то вираз n(m-1)W порівнюється з табличним значенням, розподіленим за χ^2 с N–1 степенями свободи.

Б.4.4. Розрахунок бісеріального коефіцієнта кореляції

Використовується для визначення наявності зв'язку між змінними, одна з яких виміряна в дихотомічній шкалі (так-ні; є-немає тощо). Вид коефіцієнту залежить від шкали вимірювання другої (не дихотомічної) змінної.

Якщо друга виміряна в шкалі відношень або інтервалів, то використовується точково-бісеріальний коефіцієнт кореляції

$$r_{pb} = \frac{m_1 - m_0}{\sigma_x}\sqrt{\frac{n_1 n_0}{n(n-1)}}$$

де: m_1 и m_0 – середні значення X для для 1 або 0 по Y; σ_x – середнє квадратичне відхилення всіх значень по X; n_1 ,n_0 – кількість значень X з 1 або 0 по Y; n – загальна кількість пар значень;

Значимість перевіряється за критерієм хі-квадрат.

В тому випадку, якщо друга змінна виміряна в порядковій шкалі, то використовується рангово-бісеріальний коефіцієнт кореляції.

$$r_{rb}=2 * (m_1 - m_0)/ n$$

де: n – загальна кількість пар значень; m_1 и m_0 - середні ранги об'єктів з 1 або 0 по дихотомічній змінній.

Значимість перевіряється за критерієм Стюдента.

Коли вважають, що дихотомічна змінна розподілена за нормальним законом, то використовується бісеріальний коефіцієнт кореляції.

$$r_{bis} = \frac{x_1 - x_0}{s_x}\frac{n_1 n_0}{U_n\sqrt{n^2 - n}}$$

Тут всі описані параметри аналогічні точковому коефіцієнту кореляції, за винятком U_n. Це ординати нормального розподілу, які відповідає n_1 /n % площі під кривою. Якщо значення коефіцієнту за абсолютною величиною більше 1, то це означає, що гіпотеза про нормальний розподіл не підтверджується.

Значимість перевіряється за критерієм Стюдента.

Дані для розрахунку готуються в наступному вигляді (табл..Б.4.6).

Таблиця Б.4.6. Вихідні дані

№	Назви	
	Значення параметру 1	Наявність параметру 2
1	74,80	0
2	58,50	0
3	92,60	0
4	24,20	0
5	42,90	1
6	76,40	0
7	84,70	0

8	228,50	1
9	142,70	0
10	93,80	0
11	49,90	1
12	35,80	1
13	83,30	0
14	190,00	0
15	1323,10	1
16	71,40	0
17	92,90	0
18	88,80	0
19	98,70	0
20	256,10	1
21	82,60	0
22	88,90	0
23	35,50	1
24	36,80	0
25	60,70	0
26	37,30	0
27	16,02	0
28	23,03	0
29	27,10	0
30	18,20	0
31	51,80	0
32	26,09	0
33	54,40	0
34	24,09	0
35	22,03	0
36	25,01	0
37	18,02	1
38	22,03	0
39	38,40	0
40	93,90	0
41	83,70	0
42	20,02	0
43	91,80	0

Для запуску макросу послідовно вибираються пункти меню «Сервис»–«Макрос»–«Макроси». В вікні, що з’явилось (рис.Б.4.19) вибирається макрос з іменем Byser_M.

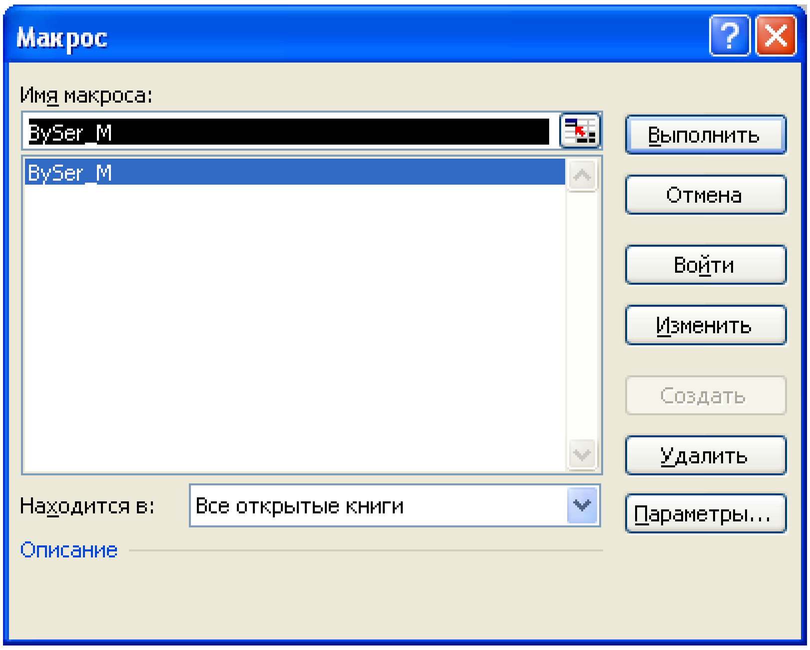


Рис. Б.4.19. Вікно запуску макросу

Після запуску макросу необхідно на запит (рис. Б.4.20) ввести табл.. Б.4.6 без заголовків.

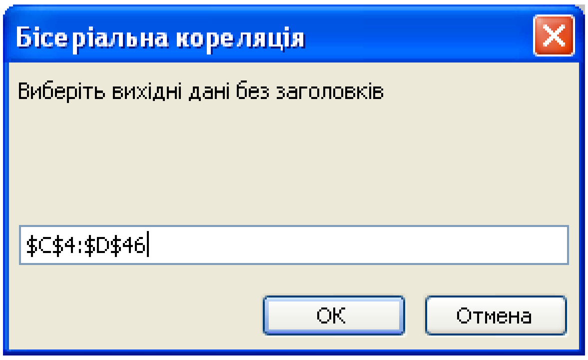


Рис. Б.4.20. Запит на введення вихідних даних

У випадку, якщо таблиця занадто мала, то після повідомлення про помилку (рис. Б.4.21) необхідно повторити введення таблиці (рис. Б.4.20).

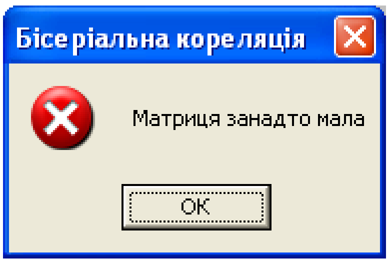


Рис. Б.4.21. Повідомлення про помилку при введенні матриці

Потім можливо змінити або прийняти стандартний рівень значущості (рис. Б.4.22).

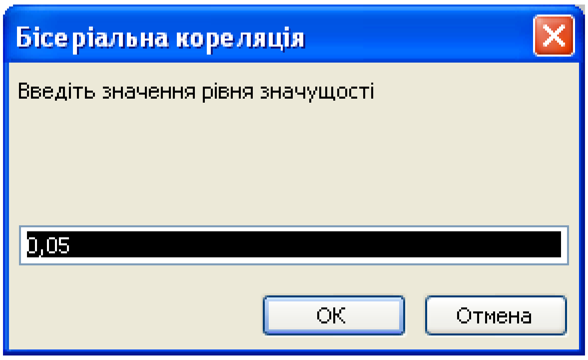


Рис. Б.4.22 Запит на введення рівня значущості

Якщо значення рівня значущості неправильне, то після повідомлення про помилку (рис. Б.4.23) необхідно повторити його введення (рис. Б.4.22).

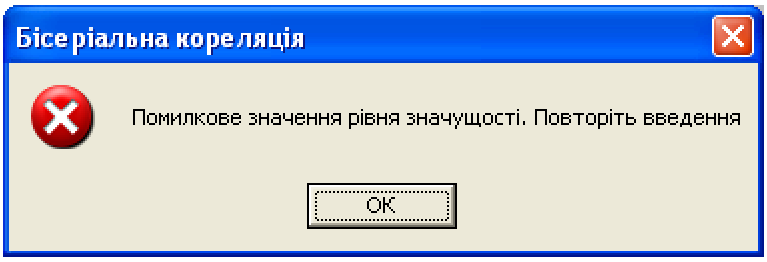


Рис. Б.4.23. Повідомлення про неправильний рівень значущості

Потім необхідно вказати комірку, з якої почнеться виведення результатів (рис. Б.4.24).

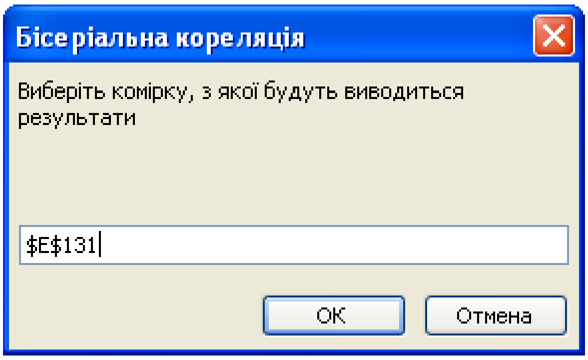


Рис. Б.4.24. Введення місця виведення результату

Наостанок пропонується вибрати вид коефіцієнта бісеріальної кореляції (рис. Б.4.25)

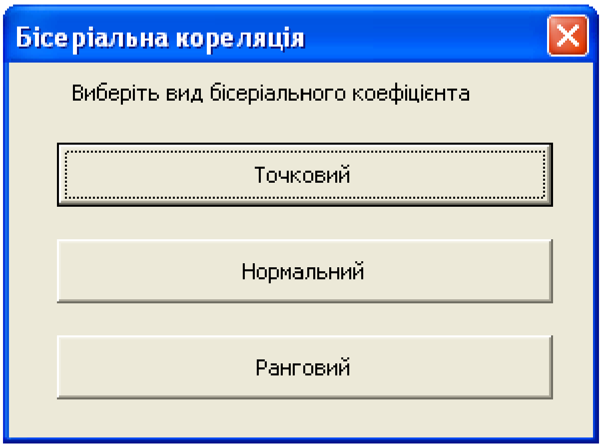


Рис. Б.4.25. Запит виду бісеріального коефіцієнта кореляції

Результати роботи представлені в табл.. Б.4.7 – табл.. Б.4.9 (форма залежить від виду коефіцієнта.

Таблиця Б.4.7. Коефіцієнт кореляції точково-бісеріальний

Коефіцієнт кореляції точково-бісеріальний=	0,367576
Рівень значущості	0,05
Число степенів свободи =	40
Хі-квадрат розрахункове =	36,39643
Критичне значення хі-квадрат критерію	55,75848

Таблиця Б.4.8. Коефіцієнт кореляції рангово-бісеріальний

Коефіцієнт кореляції рангово-бісеріальний=	0,114286
t розрахункове =	0,736612
Критичне значення критерію Стюдента	2,019541
Рівень значущості	0,05

Таблиця Б.4.9. Коефіцієнт кореляції бісеріальний

Коефіцієнт кореляції бісеріальний=	33,06013	
t розрахункове =		
Критичне значення критерію Стюдента	2,019541	
Рівень значущості	0,05	
Закон розподілу не Гауса		

Б.4.5.Часткові коефіцієнти кореляції

Для того, щоб «очистити» значення кореляційного зв’язку між двома змінними від можливого впливу третьої, введено поняття часткової кореляції. За ним коефіцієнт кореляції між змінними X і Z визначається за формулою

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

Тут r_{12}, r_{13} і r_{23} — коефіцієнти парної кореляції між змінними X і Y, X і Z, Y і Z відповідно.

Для випадку, коли закорельованих між собою змінних більше трьох, використовується наступна формула

$$r_{ij,X^{i,j}} = \frac{-R_{ij}}{\sqrt{Rii * R_{jj}}}$$

Довірчий інтервал для часткового коефіцієнта кореляції розраховується за формулою $th(z_1) < r < th(z_2)$**,** де th() – гіперболічний тангенс, а Z розраховується за формулою

$$z_{1,2} = \frac{1}{2}\ln\frac{1+\hat{r}}{1-\hat{r}} \mp \frac{u_{\alpha/2}}{\sqrt{(n-1)-3}} - \frac{\hat{r}}{2((n-1)-1)},$$

В якій $u_{\alpha/2}$ – квантиль стандартного нормального розподілу.

Вихідні дані мають бути розміщені в таблиці, де кожен стовпчик є окремою змінною. Послідовно вибираються пункти меню «Сервис»–«Макрос»–«Макроси». В вікні, що з'явилось (рис. Б.4.26) вибирається макрос з іменем Par_Corr.

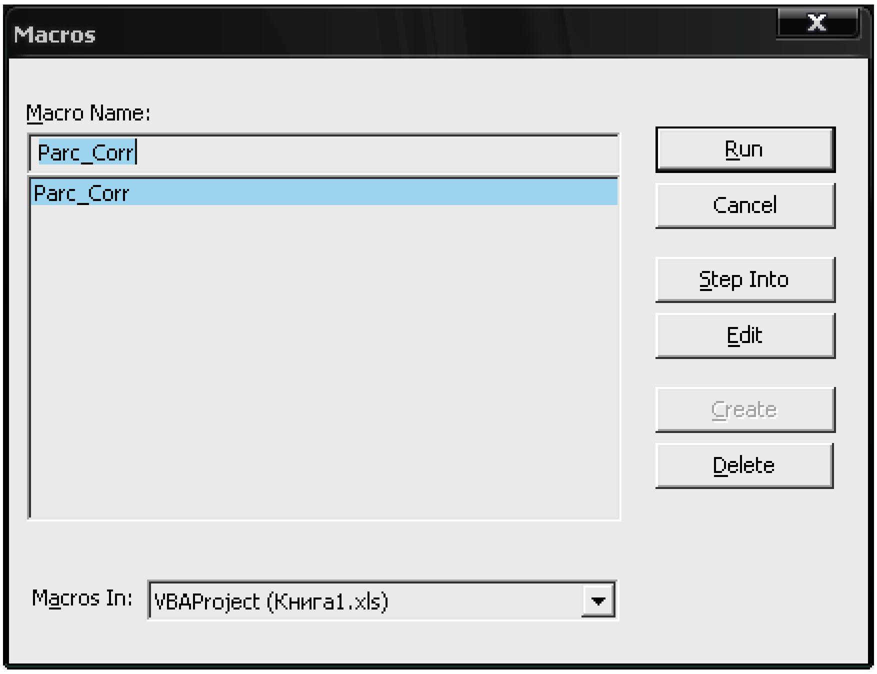


Рис. Б.4.26. Форма вибору макросу

Після запуску програми з'являється запит, на який необхідно ввести матрицю незалежних змінних (рис. Б.4.27)

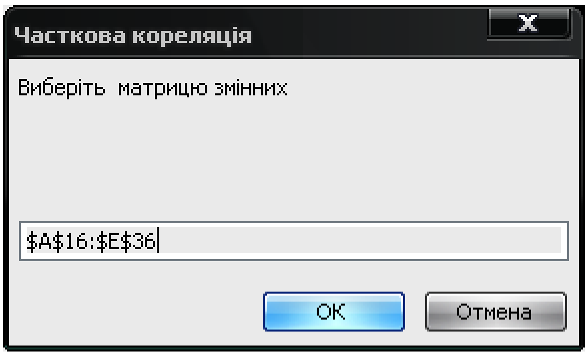


Рис. Б.4.27 Запит на введення вихідної матриці даних

Якщо введена матриця має менше 4 дослідів, або менше двох факторів, відається повідомлення про помилку (рис. Б.4.28)

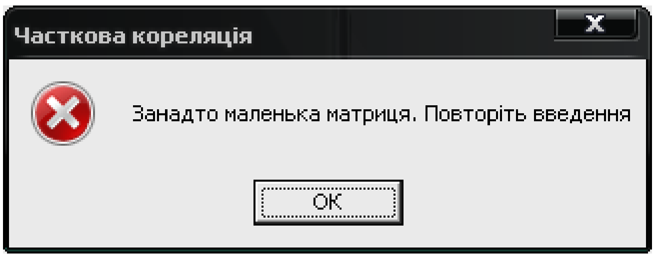


Рис. Б.4.28 Повідомлення про помилку

Після цього вводиться номер опорної змінної (рис. Б.4.29, тобто тієї змінної, для якої обчислюються коефіцієнти часткових кореляцій з усіма іншими.

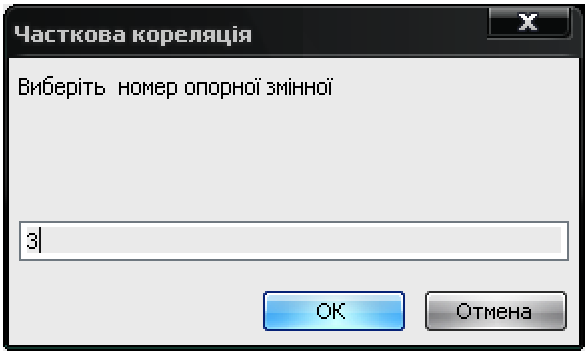


Рис. Б.4.29 Введення номеру опорної змінної

При введенні неіснуючого номера опорної змінної з'являється повідомлення про помилку (рис. Б.4.30).

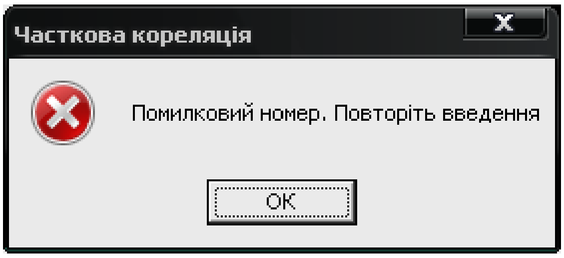


Рис. Б.4.30. Помилка при введенні номера опорної змінної

Далі вводиться значення рівня значущості для побудови довірчих інтервалів часткових коефіцієнтів кореляції (рис. Б.4.31). Якщо немає необхідності змінювати стандартний рівень значущості, то достатньо натиснути «ОК» або Enter. При введенні помилкового значення рівня значущості (за межами інтервалу [0,1]), макрос видає повідомлення про помилку (рис. Б.4.32). Після чого необхідно повторити введення.

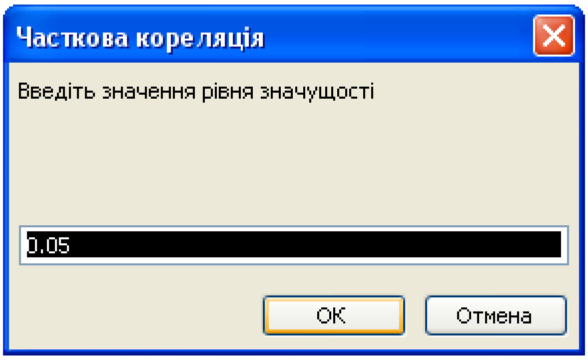


Рис. Б.4.31. Запит рівня значущості

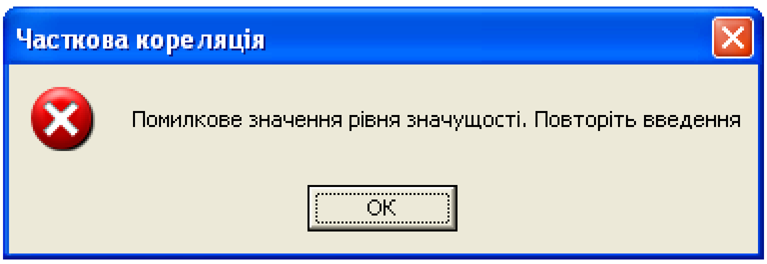


Рис. Б.4.32. Помилка при введенні рівня значущості

Наостанок необхідно ввести місце виводу результату (рис.8).

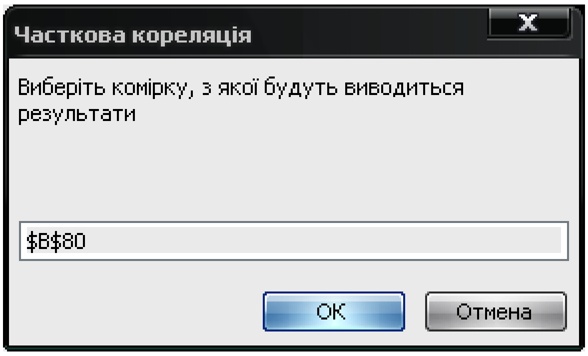


Рис. Б.4.33. Встановлення місця виведення результату

При обчисленні можливі грубі помилки (рис. Б.4.34, Б.4.35).

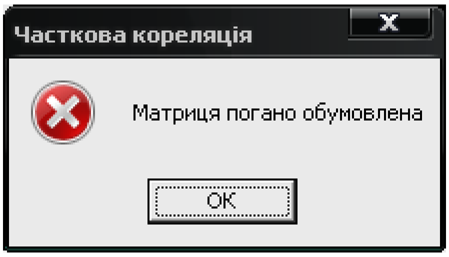


Рис. Б.4.34. Погано підібрані дані

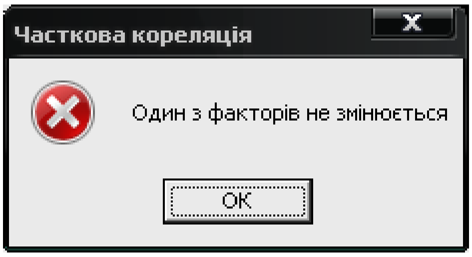


Рис. Б.4.35. Помилка в даних

Після таких помилок макрос припиняє роботу. Для отримання успішного результату необхідно виправити змінити вихідну матрицю.

Форма виведення результатів представлена в табл.. Б.4.10.

Таблиця Б.4.10. Форма вихідних даних для розрахунку часткової кореляції

	Кореляційна матриця						
		X 1	X 2	X 3	X 4	X 5	
	X 1	1	0,022466	0,050662	-0,02008	0,220808	
	X 2	0,022466	1	0,720876	-0,34169	-0,10531	
	X 3	0,050662	0,720876	1	0,387248	-0,0666	
	X 4	-0,02008	-0,34169	0,387248	1	0,089659	
	X 5	0,220808	-0,10531	-0,0666	0,089659	1	
Опорна змінна		2					
	Часткова кореляція						
	X 1	X 2	X 3	X 4	X 5		
	-0,31472	1	0,986473	-0,975	0,24263		
Довірчий інтервал	-0,66002		0,963558	-0,98975	-0,22999		
	0,156571		0,994472	-0,93334	0,614763		

Рівень значущості	0,05					

Б.5. Лабораторна робота №5.

Дисперсійний аналіз

Б.5.1. Однофакторний

Розглядається вплив погодних умов на частку браку. Вихідні дані приведені в табл.. Б.5.1. На рис. Б.5.3. представлені вихідні дані і результати роботи.

Таблиця Б.5.1. Вихідні дані для одно факторного дисперсійного аналізу

Погода	Робітники			
	1	2	3	4
Тиха погода	13,8	11	13,7	12,1
Вітер і хуртовина	16	12,2	15,8	14,3

Після того як наші дані набрані в електронній таблиці (рядки 2–4 і стовпці A — E) і перевірені на нормальний закон розподілу (рядки 3–4 і стовпець F), входимо в меню, вибираючи послідовно ***Сервис***, потім ***Анализ данных***, в результаті чого з'явиться вікно вибору функції (див. рис.Б.5.1). І цьому вікні необхідно вибрати "Однофакторный дисперсионный анализ". В результаті з'явиться вікно введення даних для дисперсійного аналізу (рис. Б.5.2)

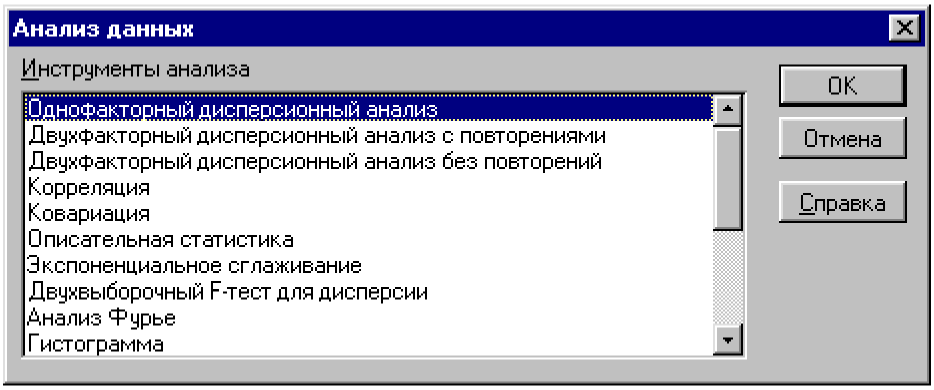


Рис. Б.5.1. Вікно вибору методу обробки даних

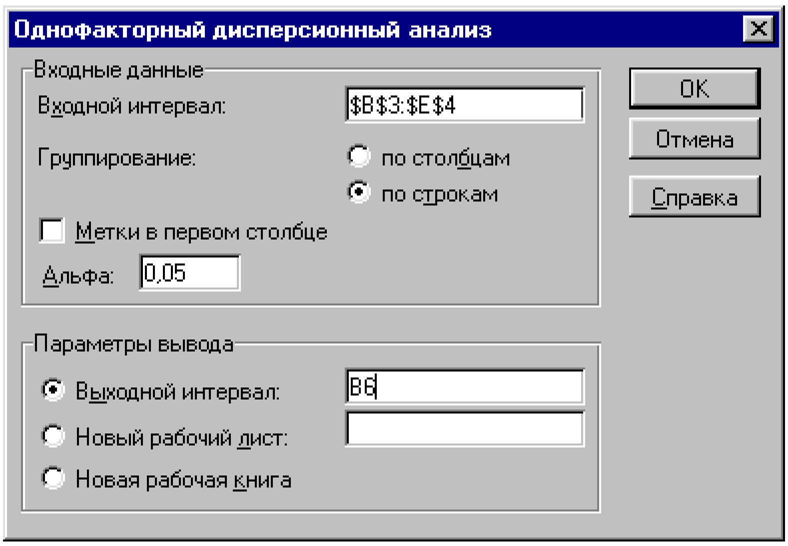


Рис.Б.5.2. Вікно встановлення параметрів для одно факторного дисперсійного аналізу

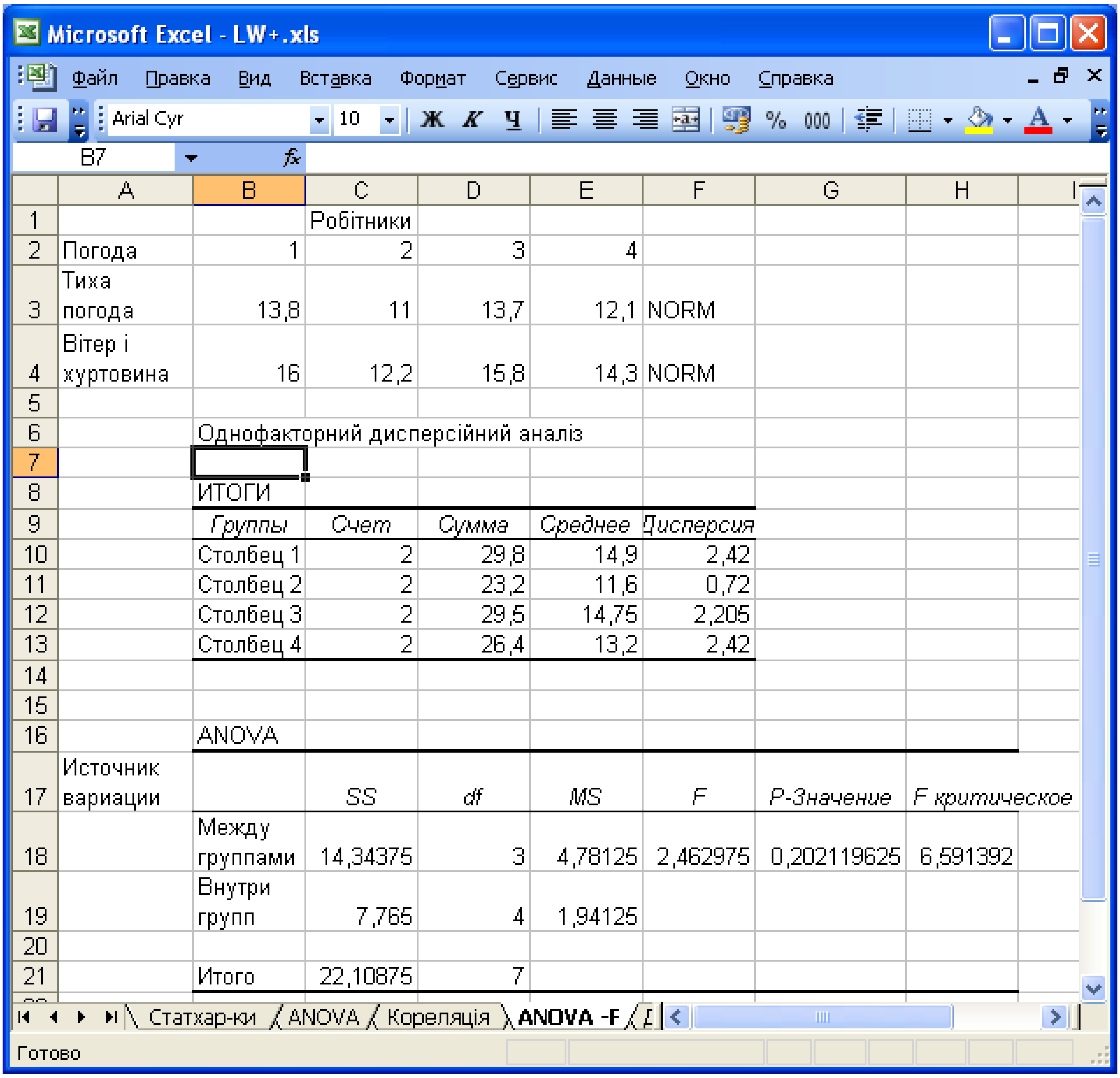
В цьому вікні встановлюються необхідні параметри для проведення дисперсійного аналізу.*Входной интервал* — необхідно встановити посилання на таблицю, в якій розміщені вихідні дані.

Группирование — вказується в рядках чи стовпцях знаходяться дані, які відносяться до одного рівня фактору (в даному випадку – в рядках).

Метки в первом столбце – відмічається при наявності імен в матриці вихідних даних.

Альфа — рівень значущості при перевірці гіпотез.

Выходной интервал — водиться посилання на першу комірку, з якої почнеться виводитись результат. *Новый рабочий лист* — встановлюється при необхідності вивести дані в новий робочий лист. *Новая рабочая книга* — встановлюється при необхідності вивести дані в нову робочу книгу. Результати будуть на першому листі, починаючи з комірки A1.



	A	B	C	D	E	F	G	H
1			Робітники					
2	Погода	1	2	3	4			
3	Тиха погода	13,8	11	13,7	12,1	NORM		
4	Вітер і хуртовина	16	12,2	15,8	14,3	NORM		
5								
6		Однофакторний дисперсійний аналіз						
7								
8		ИТОГИ						
9		*Группы*	*Счет*	*Сумма*	*Среднее*	*Дисперсия*		
10		Столбец 1	2	29,8	14,9	2,42		
11		Столбец 2	2	23,2	11,6	0,72		
12		Столбец 3	2	29,5	14,75	2,205		
13		Столбец 4	2	26,4	13,2	2,42		
14								
15								
16		ANOVA						
17	Источник вариации		*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
18		Между группами	14,34375	3	4,78125	2,462975	0,202119625	6,591392
19		Внутри групп	7,765	4	1,94125			
20								
21		Итого	22,10875	7				

Рис. Б.5.3. Результати обробки даних. Однофакторний дисперсійний аналіз

Тут

SS — сума квадратів;

Между группами — міжгрупова сума квадратів;

Внутри групп — внутрішньогрупова сума квадратів;

Итого —загальна сума квадратів;

df — число степенів свободи;

MS —дисперсія;

F — розрахункове значення критерію Фішера;

P-Значение — розрахункове значення мінімальної значущості;

F критическое — критичне значення розподілу Фішера.

В нашому прикладі розрахункове значення критерію Фішера (3,026) меньше критичного (5,99). Тому ми приймаємо нульову гіпотезу про відсутність впливу погоди на кількість браку при рівні значущості 0,05.

P-Значение в нашому випадку дорівнює 0,133. Оскільки воно достатньо велике, то немає підстав відкидати нульову гіпотезу про рівність дисперсій.

Зауваження до дисперсійного аналізу в Excel

В функції дисперсійного аналізу в Excel перевіряється наступна альтернативна гіпотеза: дисперсія чисельника більше дисперсії знаменника, тобто одностороння гіпотеза. При цьому не враховується можливість ситуації, коли F-розрахункове 1. В такому випадку одностороння перевірка, яка виконується в Excel неправомочна. Для коректної перевірки необхідно або перерахувати розрахункове і критичне значення, або ж обчислити дві критичні границі.

Припустимо, що така ситуація (розрахункове значення критерію Фішера менше 1) склалася в нашому прикладі. Тоді нове розрахункове значення обчислюється за формулою =1/F16, а відповідне критичне викликом функції = FРАСПОБР(0,05;D17;D16). Нові значення порівнюються аналогічно. Для другого варіанту перевірки критичне значення не змінюється, але розраховуються нові значення для критеріальних границь: нижньої =FРАСПОБР(0,975; D16;D17) і верхньої = FРАСПОБР(0,025; D16;D17). Якщо розрахункове значення знаходиться всередині цих границь, то нульова гіпотеза приймається.

Б.5.2. Двофакторний дисперсійний аналіз з повторними дослідами (рівномірне дублювання)

В Excel є два варіанти двофакторного дисперсійного аналізу: з повторами і без.. Двофакторний дисперсійний аналіз без повторів – це такий його різновид, в кожній комірці якого є тільки одне спостереження. Таблиця вихідних даних в цьому випадку подібна до таблиці одно факторного аналізу, тільки замість номера досліду — номер рівня другого фактору. В таблиці аналізу в цьому випадку відсутня взаємодія.

На жаль функція двофакторного дисперсійного аналізу з повторами в Excel працює неправильно і нею не варто користуватись.

В прикладі досліджувалась стійкість кінцевих фрез діаметром 16мм з двох марок швидкорізальної сталі від двох виробників (табл.. Б.5.2.).

Таблиця Б.5.2. Вихідні дані для двофакторного дисперсійного аналізу

Виробник	№ випробування	Марка сталі	
		Р9К5	EMo5Co5
A	1	25,7	19
	2	19	21,3
	3	19	21,3

	4	20,2	15,7
	5	20,2	23,5
Б	1	37,2	21,4
	2	35,2	12,1
	3	38,6	20,1
	4	35,7	28,3
	5	40	10

Вихідні дані розміщені в комірках B4:E15 (див. рис.Б.5.4).

Виконуємо наступні дії.

- Обчислюємо $\overline{X}_{ij*}$,.
- Обчислюємо середнє по рядкам $\overline{X}_{i**}$.
- Обчислюємо середнє по стовпцям $\overline{X}_{j*}$.
- Обчислюємо загальне середнє $\overline{X}$.

Результати розміщуються в комірках від D18:F20. Використані формули представлені в табл..Б.5.3.

Таблиця Б.5.3. Формули розрахунку середніх

Комірка	D	E	F
18	=СРЗНАЧ(D6:D10)	=СРЗНАЧ(E6:E10)	=СРЗНАЧ(D18:E18)
19	=СРЗНАЧ(D11:D15)	=СРЗНАЧ(E11:E15)	=СРЗНАЧ(D19:E19)
20	=СРЗНАЧ(D18:D19)	=СРЗНАЧ(E18:E19)	=СРЗНАЧ(F18:F19)

- Знаходимо квадрати різниць $(X_{ijk} - \overline{X}_{ij*})^2$ для кожної комірки. Масив розміщений в комірках H6 по I15.
- Знаходимо квадрати різниць по рядкам $(\overline{X}_{i**} - \overline{X})^2$. Результати розміщені в H18–H19.
- Знаходимо квадрати різниць по стовпцям $(\overline{X}_{j*} - \overline{X})^2$. Результати розміщені в D22 – E22.
- Знаходимо квадрати різниць, необхідних для розрахунку дисперсії взаємодії $(\overline{X}_{j*} - \overline{X}_{i**} - \overline{X}_{j*} + \overline{X})^2$. Результати розміщені в D24 – E25.

  По п.п.5–6 набираємо одно формулу в першій комірці діапазону (приведена в табл.. Б.5.4), а потім розмножаємо розтягуванням на весь діапазон.
- Формуємо таблицю дисперсійного аналізу (комірки від J22 до O26).

Таблиця Б.5.4. Формули розрахунку квадратів різниць

Комірка	Зміст
B22	=(D20-$F20)^2

H18	=(F18-F$20)^2
H6	=(D6-D$18)^2
D24	=(D18-D$20-$F18+F20)^2

Розрахункові формули представлені в таблицях Б.5.5 і Б.5.6.

Таблиця Б.5.5. Дисперсії

Комірка	J	K	L
23	=СУММ(H18:H19)*ЧИСЛСТОЛБ(D6:E6)*ЧСТРОК(D6:D10)	=ЧИСЛСТОЛБ(D6:E6)-1	=J23/K23
24	=СУММ(D22:E22)*2*ЧСТРОК(D6:D10)	2-1	=J24/K24
25	=ЧСТРОК(D6:D10)*СУММ(D24:E25)	=(2-1)*(ЧИСЛСТОЛБ(D6:E6)-1)	=J25/K25
26	=СУММ(H6:I15)	=2*ЧИСЛСТОЛБ(D6:E6)*(ЧСТРОК(D6:D10)-1)	=J26/K26

Таблиця Б.5.6. Критерій Фішера

Комірка	M	N	
23	=L23/L$26	=FРАСПОБР(0,975;$K23;$K$26)	=FРАСПОБР(0,025;$K23;$K$26)
24	=L24/L$26	=FРАСПОБР(0,975;$K24;$K$26)	=FРАСПОБР(0,025;$K24;$K$26)
25	=L25/L$26	=FРАСПОБР(0,975;$K25;$K$26)	=FРАСПОБР(0,025;$K25;$K$26)

Оскільки всі три розрахункових значення критерію Фішера знаходяться за межами критичної області, то нульова гіпотеза відкидається. Це означає, що як обидва фактори, так і взаємодія статистично значимо впливають на відгук. Наявність статистично значущої взаємодії означає різний рівень якості різних сталей в одного і того ж виробника.

Як вихідні дані так і результати приведені на рис. Б.5.4.

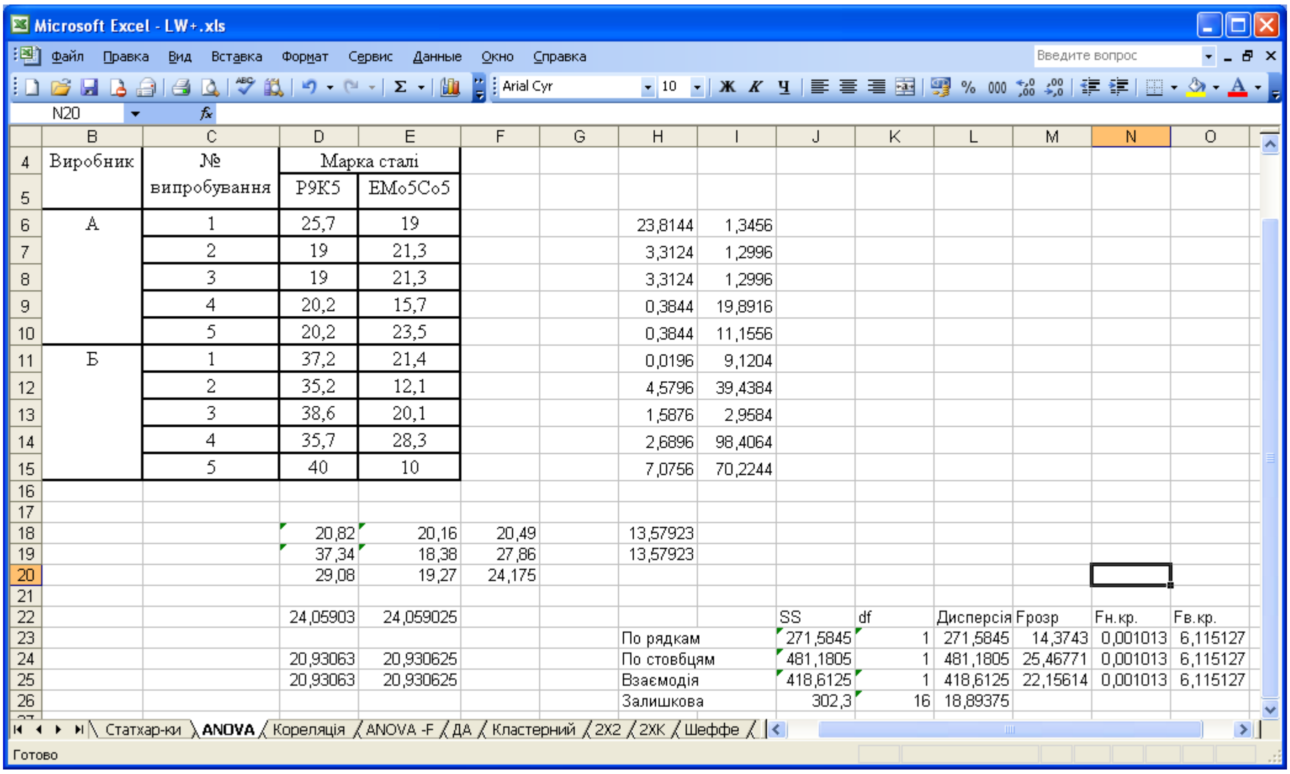


Рис. Б.5.4. Результати обробки даних. Двофакторний дисперсійний аналіз

Зауваження

- Існують і більш складні схеми, які враховують значну кількість факторів, нерівномірність дублювання експериментів тощо.
- З ростом кількості компонент зростає складність і громіздкість розрахунків. Якщо факторів більше двох, то доцільніше використання регресійного аналізу з деякими обмеженнями в висновках.

Б.5.3. Непараметричний дисперсійний аналіз Фрідмана

В тому випадку, коли закон розподілу залежної величини не є нормальним використовується непараметричний дисперсійний аналіз Фрідмана.

Нульова гіпотеза. Середні значення всіх вибірок рівні.

Передумови

- Всі випадкові величини взаємно незалежні.
- Дані кожної вибірки розподілені за одним законом розподілу. В різних вибірках можуть бути різні закони розподілу.

Вихідні дані представляються в наступному вигляді (табл. Б.5.7).

Таблиця Б.5.7 Загальний вигляд вихідних даних

Номери елементів сукупностей	1	2	…	j	…	n
Номери сукупностей						
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_{jn}
…			…		…	
m	X_{m1}	X_{m2}		X_{mj}		X_{mn}

Потім в кожному стовпчику значення X замінюють на їх ранги. Далі розраховується значення критерію

$$\chi^2 = \frac{12\sum_{j=1}^{n}(\sum_{i=1}^{m} R_{ij})^2}{mn(n+1)} - 3m(n+1),$$

де R_{ij} — відповідні значення рангів.

Якщо розрахункове значення χ^2 буде більше критичного, взятого з заданим рівнем значущості і (n-1) степенями свободи, гіпотеза про наявність зв’язку приймається.

При розрахунках можливо перевірити правильність визначення рангів, знаючи, що має місце співвідношення

$$\sum_{i=1}^{m}\sum_{j=1}^{n} R_{ij} = \frac{nm(m+1)}{2}$$

Зауваження:

При малих значеннях m і n критерій χ^2 дає занадто грубе наближення, що може привести до прийняття помилкових рішень. Тому критерій χ^2 використовується в разі виконання наступних умов: $m = 3$ і $n > 9$ або $m = 4$ і $n > 4$ або $m > 4$ і $n \geq 9$ ([4]).

Якщо ці умови не виконуються, то перевірка виконується за критерієм Фрідмана.

Виконання роботи

Вихідні дані і результати приведені на рис.Б.5.5/

Спочатку будуємо ранги по стовпчикам. Для цього в комірці B8 набираємо виклик функції побудови рангів =Rank1(B3;B$3:B$6;1), а потім розмножуємо її розтягування на комірки від B8 до F11.

Знаходимо суми рангів по рядкам і розміщуємо їх в стовпцю G. Для цього в G8 набираємо формулу =СУММ(B8:F8), а потім розмножуємо її перетягуванням на інші комірки стовпця. Тут бажано виконати перевірку правильності визначення рангів. Для цього в комірці G12 формуємо суму рангів (=СУММ(G8:G11)), а в комірці G13 — контрольне значення, яке визначається за формулою

=(ЧСТРОК(B8:B11)*ЧИСЛСТОЛБ(B8:F9)*(ЧСТРОК(B9:B11)+1))/2.

Якщо ці значення співпадають, то це означає, що розрахунки рангів виконані правильно.

Знаходимо суму квадратів =СУММКВ(G8:G11) і поміщаємо її в комірку H14.

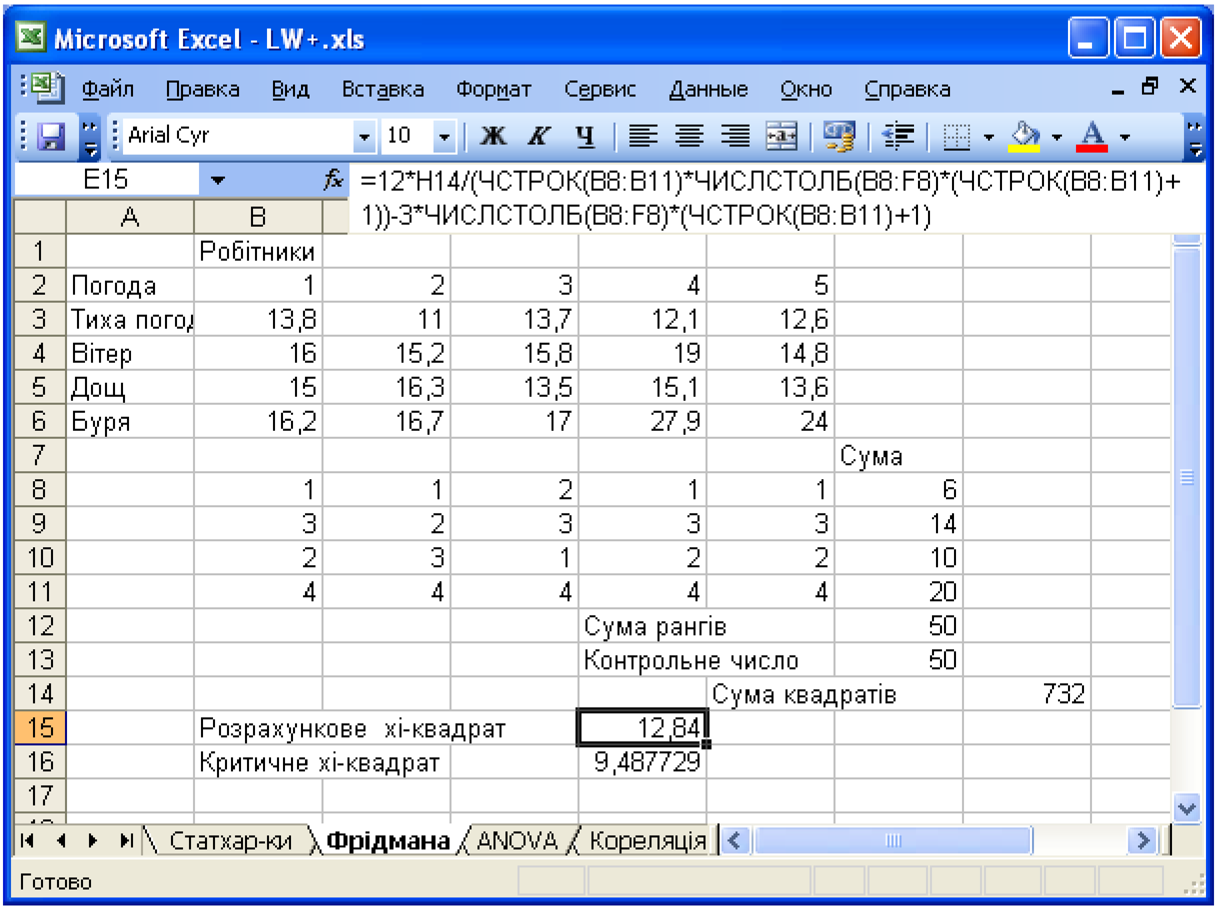
Тепер можна розрахувати критеріальне значення за формулою

=12*H14/(ЧСТРОК(B8:B11)*ЧИСЛСТОЛБ(B8:F8)*(ЧСТРОК(B8:B11)+1))-3*ЧИСЛСТОЛБ(B8:F8)*(ЧСТРОК(B8:B11)+1)

І визначити критичне значення χ^2 викликом функції

=ХИ2ОБР(0,05;ЧСТРОК(B8:B11)-1)).

Так як розрахункове значення (12,84) більше критичного (9,4877), то приймається гіпотеза про наявність впливу погодних умов на кількість браку.



	A	B	C	D	E	F	G	H
1		Робітники						
2	Погода	1	2	3	4	5		
3	Тиха погод	13,8	11	13,7	12,1	12,6		
4	Вітер	16	15,2	15,8	19	14,8		
5	Дощ	15	16,3	13,5	15,1	13,6		
6	Буря	16,2	16,7	17	27,9	24		
7							Сума	
8		1	1	2	1	1	6	
9		3	2	3	3	3	14	
10		2	3	1	2	2	10	
11		4	4	4	4	4	20	
12					Сума рангів		50	
13					Контрольне число		50	
14						Сума квадратів		732
15		Розрахункове хі-квадрат			12,84			
16		Критичне хі-квадрат			9,487729			
17								

Рис. Б.5.5. Результати обробки даних методом дисперсійного аналізу Фрідмана

Б.6. Лабораторна робота №6. Таблиці часток та пропорцій

Б.6.1. 4-х кліткова таблиця

При виготовленні заготовок інструмента з швидко ріжучої сталі методами пластичної деформації можлива поява браку в вигляді тріщин. Виникла необхідність в створенні методу оперативного контролю швидкорізальної сталі по пластичності. Було запропоновано випробування методом гарячого кручення і як показник використовувати кількість оборотів до руйнування.

Було проведено 45 плавок з відповідними випробуваннями. Середнє значення кількості обертів, розраховане по всіх випробуваннях 7,57.

Таблиця Б.6.1. Інформація про випробування в вигляді 4-х кліткової таблиці

Число обертів до руйнування зразка	Кількість плавок	
	З браком	Без браку
<=7,57	19	7
>7,57	1	18

Для розв'язання задачі по перевірці наявності зв'язку між вибраним критерієм і якістю плавка виконаємо наступні дії.

1. Знайдемо суми по стовпцям і рядкам, для чого впишемо функцію знаходження суми у відповідні крмірки (табл..6.2).

Таблиця Б.6.2. Інформація про випробування в вигляді 4-х кліткової таблиці

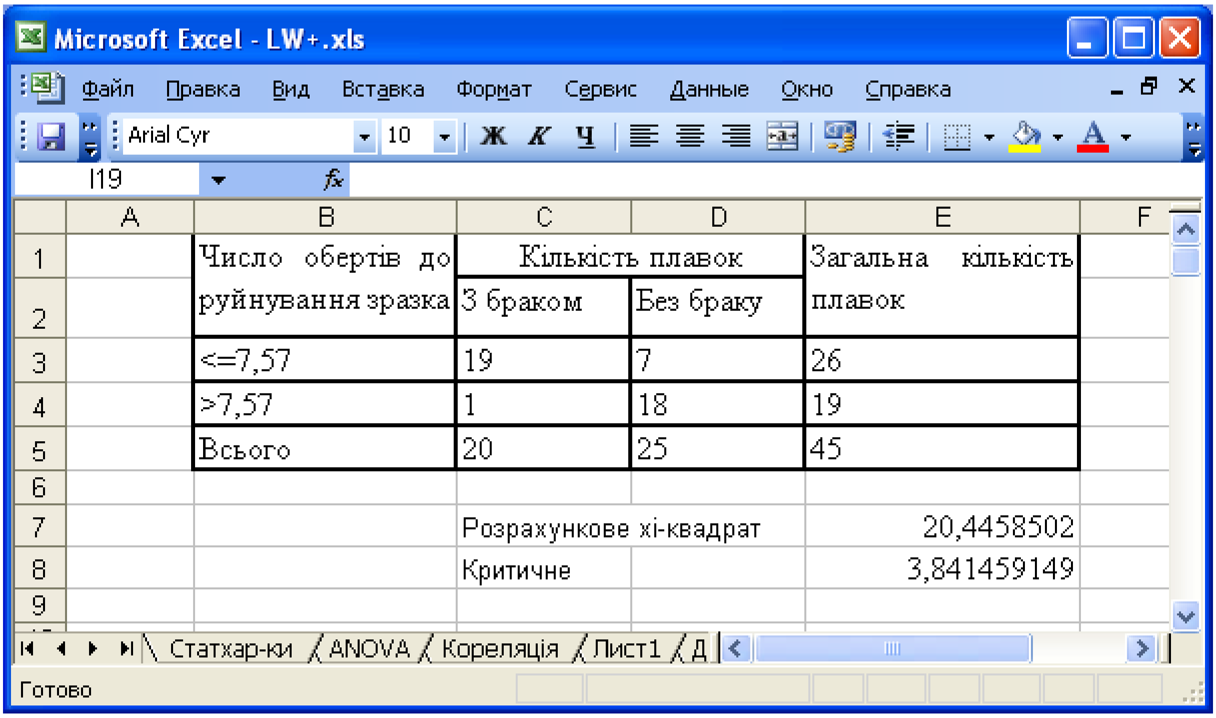
Рядок/Стовпець	C	D	E
3	19	7	=СУММ(C3:D3)
4	1	18	=СУММ(C4:D4)
5	=СУММ(C3:C4)	=СУММ(D3:D4)	=СУММ(E3:E4)

2. Визначимо розрахункове значення критерію хі-квадрат. Для цього введемо формулу =(E5*(C3*D4-D3*C4)*(C3*D4-D3*C4))/(E3*E4*C5*D5) в комірку E7.

3. Розрахуємо критичне значення для критерію хі-квадрат, введенням в комірку E8 виклику функції =ХИ2ОБР(0,05;1).

4. Виконаємо порівняння розрахункового і критичного значення. Оскільки розрахункове (20,45) більше критичного (3,84), то з рівнем значущості 0,05 нульова гіпотеза відкидається. Це означає, що існує статистично значущий зв'язок між вибраною характеристикою і наявністю браку в плавці.

Результати роботи приведені на рис. Б.6.1.



	A	B	C	D	E
1		Число обертів до руйнування зразка	Кількість плавок		Загальна кількість плавок
2			З браком	Без браку	
3		<=7,57	19	7	26
4		>7,57	1	18	19
5		Всього	20	25	45
6					
7			Розрахункове хі-квадрат		20,4458502
8			Критичне		3,841459149
9					

Рис. Б.6.1. Аналіз чотирьох кліткових таблиць часток і пропорцій

Б.6.2. Таблиці виду 2×К

Як приклад розглянемо дані з табл. Б.6.3.

Таблиця Б.6.3 Залежність перемог на конкурсах і олімпіадах в залежності від занять науковою роботою

Категорія студентів	Кількість		Всього
	Стали лауреатами конкурсів	Не мають конкурсних відзнак	
Постійно займаються науковою роботою	19	42	61
Епізодично займаються науковою роботою	23	71	94
Не займаються науковою роботою	23	227	250
Разом	65	340	405

Необхідно встановити чи є зв'язок між заняттями науковою студентською роботою і перемогами на конкурсах і олімпіадах. В табл. Б.6.4 показано, як за допомогою операцій та функцій Excel розрахувати попередні дані, необхідні для обчислення значення критерію χ^2.

Таблиця Б.6.4. Підготовчі розрахунки

Рядок / стовпець	C	D	E	F

3	19	42	=СУММ(C3:D3)	=C3*C3/E3
4	23	71	=СУММ(C4:D4)	=C4*C4/E4
5	23	227	=СУММ(C5:D5)	=C5*C5/E5
6	=СУММ(C3:C5)	=СУММ(D3:D5)	=СУММ(C6:D6)	=СУММ(F3:F5)

Для обчислення критеріального значення вставляємо в комірку E8 формулу =E5*E5*(F6-C6*C6/E6)/(C6*(E6-C6)). Для отримання критичного значення (процентної точки) розподілу χ^2 для порівняння в комірку E9 розміщуємо виклик функції =ХИ2ОБР(0,05;3-1). Тут 0,05 — рівень значущості, а 3 — кількість різних вибірок. Оскільки 9,133466 > 5,991476, то нульова гіпотеза про відсутність зв'язку відхиляється. Таким чином між інтенсивністю заняттями науковою роботою студентів і кількістю перемог на конкурсах існує статистично значущий зв'язок. На рис. Б.6.2 приведені результати обчислень.

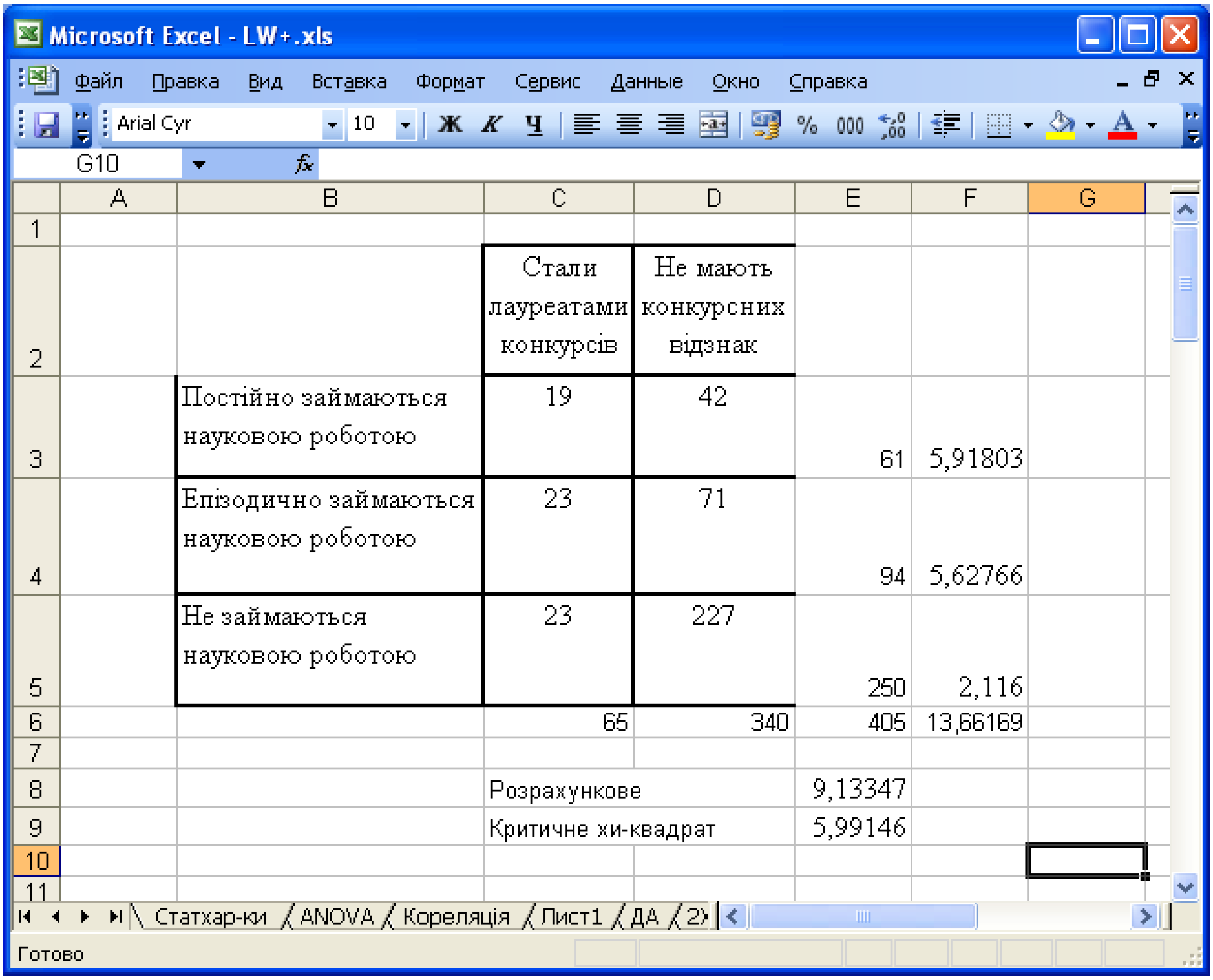


Рис. Б.6.2. Приклад аналізу таблиць суміжності виду 2ХК

Б.6.3. Таблиці виду K×L

Розглянемо задачу про розподіл фракцій різної дисперсності в інструменті різних виробників. Вихідні дані приведені в табл.Б.6.5.

Таблиця Б.6.5 Розподіл фракцій різної дисперсності в інструменті різних виробників

Виробник інструменту	Частка фракцій		
	високодисперсні	середньодисперсні	низькодисперсні
А	15,4	80,1	4,5
Б	16,7	78,5	4,8
В	31,9	65,6	2,5
Г	32,4	64,3	3,3
Д	20,7	77	2,3
Ж	37,9	58,4	3,7
К	29,6	67,8	2,6

Набираємо приведені вище дані в Excel (см. рис. Б.6.3). Для зручності подальшого використання таблиця транспонована. Спочатку обчислюємо суми по рядкам і стовпцям (табл.Б.6.6).

Таблиця Б.6.6 Підготовчі розрахунки

Адреса комірки	Зміст
L5	=СУММ(L2:L4)
M5	=СУММ(M2:M4)
N5	=СУММ(N2:N4)
O5	=СУММ(O2:O4)
P5	=СУММ(P2:P4)
Q5	=СУММ(Q2:Q4)
R5	=СУММ(R2:R4)
S2	=СУММ(L2:R2)
S3	=СУММ(L3:R3)
S4	=СУММ(L4:R4)
S5	=СУММ(L5:R5)

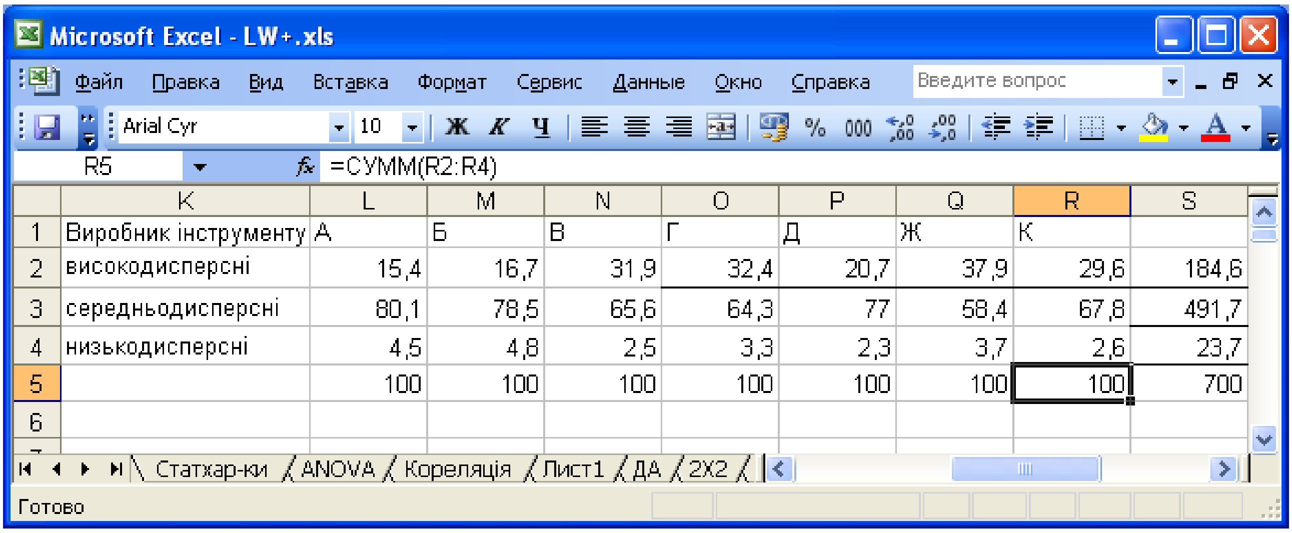


Рис. Б.6.3. Вихідні дані і попередні розрахунки

Після цього формуємо допоміжну таблиця такого ж розміру як вихідна (див. Табл.Б.6.7.

Таблиця Б.6.7. Формування допоміжної таблиці

Стовпець / Рядок	2	3	4
T	=L2*L2/(L5*$S2)	=L3*L3/(L5*$S3)	=L4*L4/(L5*$S4)
U	=M2*M2/(M5*$S2)	=M3*M3/(M5*$S3)	=M4*M4/(M5*$S4)
V	=N2*N2/(N5*$S2)	=N3*N3/(N5*$S3)	=N4*N4/(N5*$S4)
W	=O2*O2/(O5*$S2)	=O3*O3/(O5*$S3)	=O4*O4/(O5*$S4)
X	=P2*P2/(P5*$S2)	=P3*P3/(P5*$S3)	=P4*P4/(P5*$S4)
Y	=Q2*Q2/(Q5*$S2)	=Q3*Q3/(Q5*$S3)	=Q4*Q4/(Q5*$S4)
Z	=R2*R2/(R5*$S2)	=R3*R3/(R5*$S3)	=R4*R4/(R5*$S4)

Насправді набирається тільки формула =L2*L2/(L5*$S2). У всіх інших комірках фона створюється перетягуванням.

Тепер ми можемо отримати розрахункове значення для критерію χ^2. Для цього в комірку W10 вставляємо формулу =S5*(СУММ(T2:Z4)-1). Потім визначаємо критичне значення помістивши в комірку W12 виклик функції =ХИ2ОБР(0,05;(3-1)*(7-1)). Тут 0,05 — рівень значущості, 3 і 7 — кількість рівнів варіювання (різних значень) першого і другого факторів. Оскільки розрахункове значення (24,9496) більше критичного (21,02606), то нульова гіпотеза

відхиляється. Це означає, що якість інструменту різних виробників відрізняється статистично значуще. Результати представлені на рис. Б.5.4.

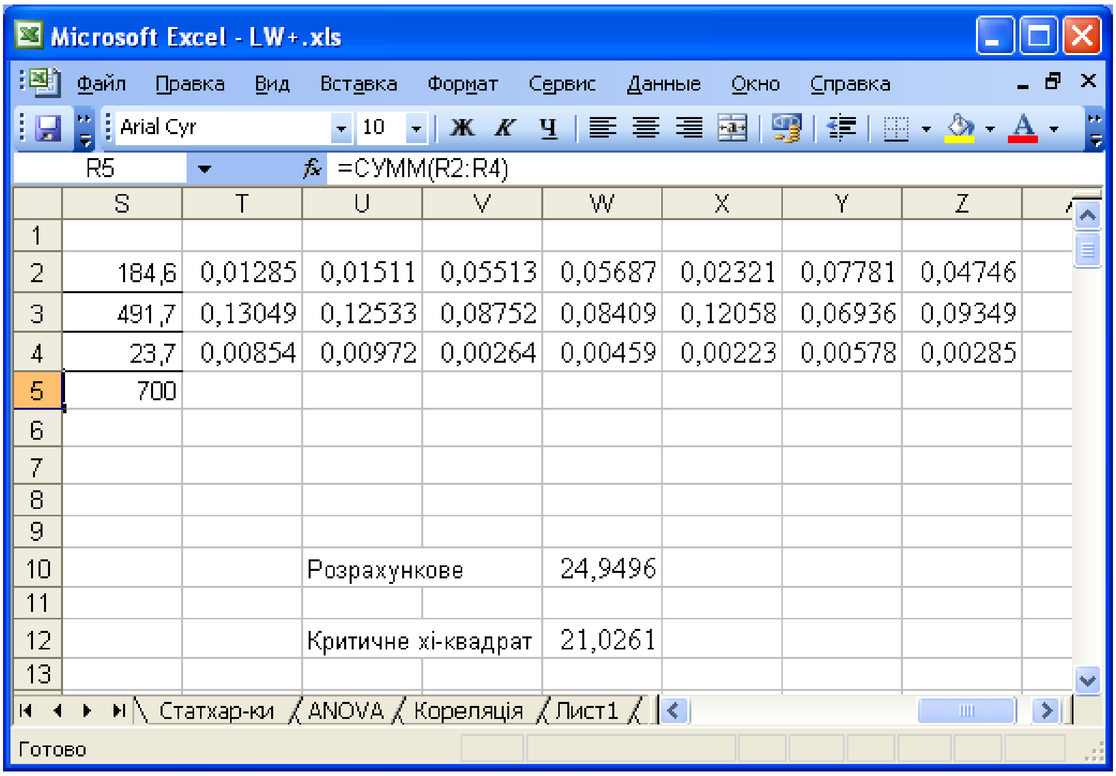


Рис. Б.6.4. Результати розрахунку по таблиці L×K

Б.7. Лабораторна робота №7. Кластерний аналіз

Б.7.1. Класичний кластер ний аналіз

Приклад вихідної інформації представлено в табл..Б.7.1.

Таблиця Б.7.1. Матриця вихідних даних для кластерного аналізу

Номер об'єкта	Змінні		
	X_1	X_2	X_3
1	5	6	5,6
2	4	7	6,5
3	6	7	1,5
4	3	8	5,1
5	7	8	7,4
6	4	9	7
7	6	9	7,3
8	5	10	4,8
9	1	2	7,1
10	2	4	7,7
11	2	3	8,6
12	2	3	0,1
13	3	3,5	7,8
14	3	3	5,6
15	3	2	3,4
16	3,5	4	0,6
17	6	4	7,6
18	6,5	2	6,3
19	7	4	4,1
20	8	3	5,8
21	8	1	8,8
22	8	4	5
23	9	3	4,3
24	9	1	6,3

Для запуску макросу вибираємо послідовно пункти меню «Сервис», «Макрос», «Макросы» (або «Вид», «Макросы», «Макросы» в більш нових версіях)і в діалоговому вікні, що з'явилося (рис.Б.7.1), запускаємо макрос Claster.

Після запуску макросу з'явиться запит на введення матриці вихідних даних (рис.Б.7.2). Матриця відмічається без заголовків.

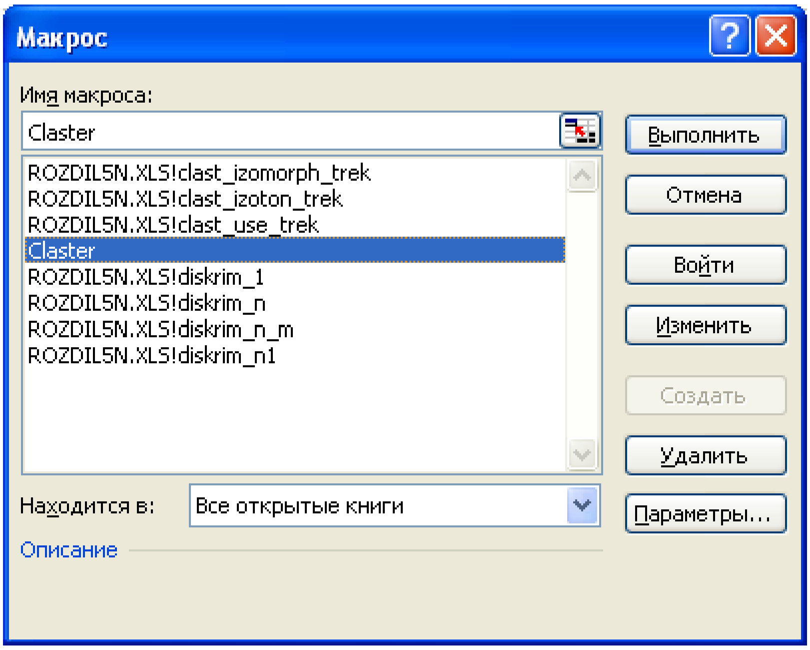


Рис. Б.7.1. Форма запуску макросу

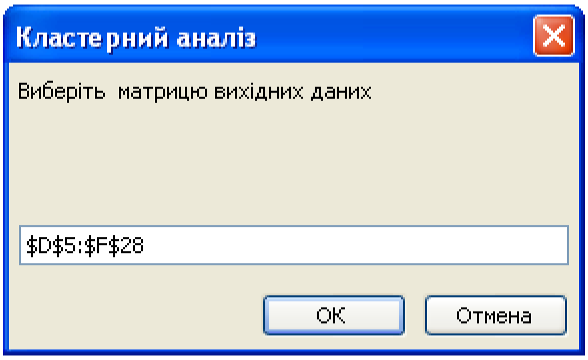


Рис. Б.7.2. Запит на матрицю вихідних даних

Якщо кількість стовпців менше 2 або число рядків менше 4, то з'являється повідомлення про помилку (рис. Б.7.3), і виконується повернення на введення матриці вихідних даних..

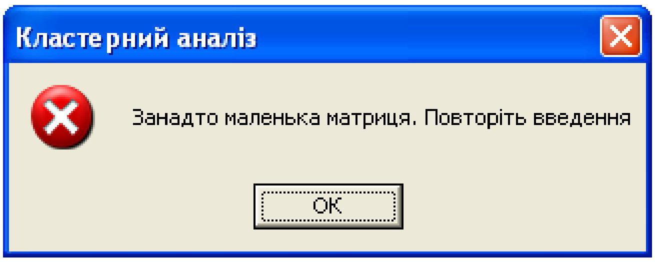


Рис. Б.7.3. Повідомлення про помилку при введенні матриці вихідних даних

Після цього необхідно ввести матрицю імен об'єктів (рис. Б.7.4).

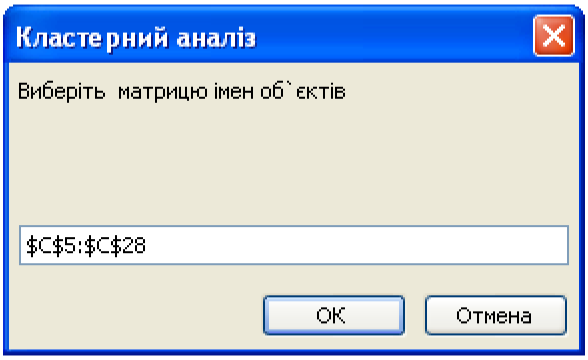


Рис. Б.7.4. Запит на список імен об’єктів

Якщо довжина введеної матриці імен не відповідає розмірам вихідної матриці, то після повідомлення про помилку (рис. Б.7.5) необхідно ввести імена знову.

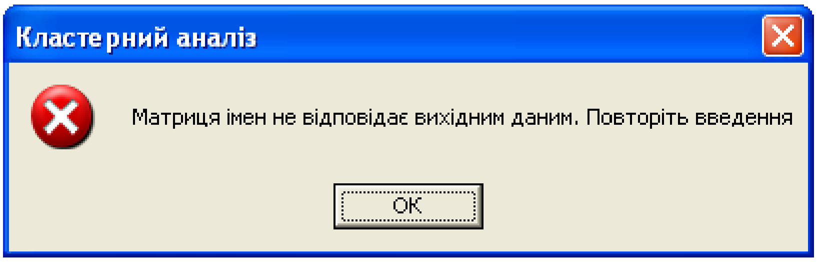


Рис. Б.7.5. Повідомлення про невідповідність матриці вихідних даних і списку імен

Після цього програма видає запит на місце, з якого почнеться виводитись результат (рис. Б.7.6).

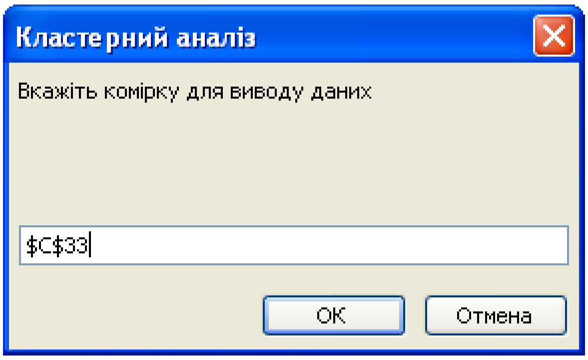


Рис. Б.7.6. Запит на місце виведення результатів

На наступному кроці необхідно вибрати вид нормування вихідних даних (рис. Б.7.7). Формули можливих видів нормування приведені в табл.. Б.7.2. Після цього вибирається спосіб визначення відстані між об’єктами (рис. Б.7.8). Формули розрахунків відстані для різних способів приведені в табл.. Б.7.3. Потім макрос запитує про необхідність виведення матриці відстані між усіма об’єктами. Слід взяти до уваги, що вона має досить великий розмір при значній кількості об’єктів. Вигляд матриці відстані (фрагмент) приведено в табл.. Б.7.2.

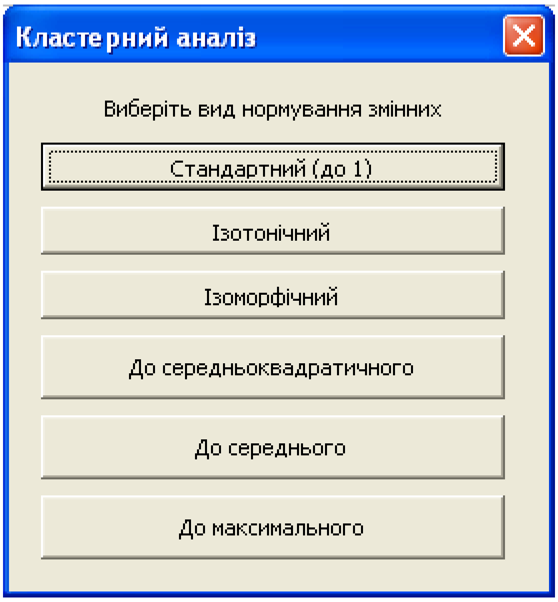


Рис. Б.7.7. Запит способу нормування вихідних даних

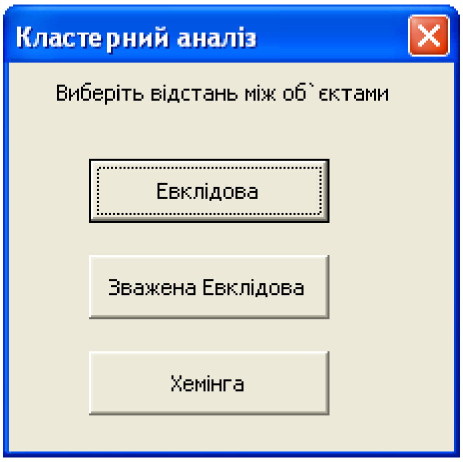


Рис. Б.7.8. Запит на спосіб визначення відстані між об'єктами

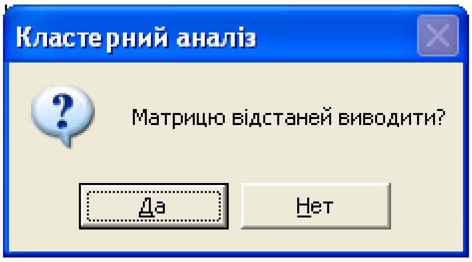


Рис. Б.7.9. Запит на необхідність виведення матриці відстаней між об'єктами

Останнім макрос запрошує зробити вибір міри зв'язку між кластерами (рис. Б.7.10). Приклад розрахованої таблиці відстаней між кластерами приведено в табл..Б.7.2. Інформація про первинні кластери (ланцюжки) має вигляд представлений в табл.. Б.7.3, а відстані між ними табл.. Б.7.4.

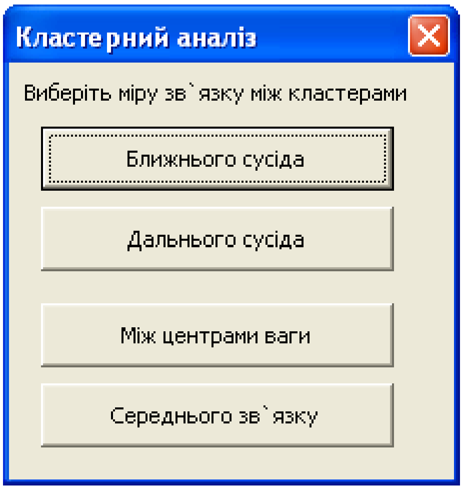


Рис. Б.7.10. Запит на спосіб визначення відстані між кластерами

Крім уже вказаного раніше, програма виводить інформацію про сформовані кластери (див. табл.. Б.7.5) як список ланцюжків і про критичні відстані між ними – табл.. Б.7.6, яка дозволяє прийняти рішення про об’єднання чи розрив відповідних кластерів.

Критична відстань визначається за формулою.

$$d_{кр} = \bar{d} + q * s \quad ,$$

де $d_{кр}$ – критична відстань, q – коефіцієнт запасу (може змінюватись від 1 і вище), s – середнє квадратичне відхилення відстані., $\bar{d}$ – середня відстань. Критичні відстані для різних значень коефіцієнта приводяться в табл.. Б.7.7.

Таблиця Б.7.2. Матриця відстаней (фрагмент)

Нормування простору стандартне (до 1); Відстань Евклідова							
Матриця відстаней між об`єктами							
	1	2	3	4	5	6	7
1	0	0,013968	0,032897	0,024792	0,027936	0,030233	0,031074
2	0,013968	0	0,040734	0,016089	0,027231	0,018478	0,025225
3	0,032897	0,040734	0	0,037619	0,045609	0,047728	0,046826
4	0,024792	0,016089	0,037619	0	0,037231	0,018718	0,031064
5	0,027936	0,027231	0,045609	0,037231	0	0,026561	0,012278
6	0,030233	0,018478	0,047728	0,018718	0,026561	0	0,016679
7	0,031074	0,025225	0,046826	0,031064	0,012278	0,016679	0
8	0,036686	0,031074	0,037539	0,024613	0,031235	0,020458	0,022287
9	0,050279	0,051789	0,074119	0,05868	0,073568	0,068032	0,075649
10	0,03445	0,033017	0,062938	0,041874	0,054981	0,048454	0,056117
11	0,043021	0,042756	0,072098	0,052867	0,061926	0,057996	0,064303
12	0,055036	0,062085	0,050119	0,059176	0,081912	0,076558	0,083159
13	0,032456	0,034136	0,061794	0,045416	0,052537	0,050806	0,055731
14	0,031785	0,03773	0,053452	0,045402	0,057619	0,055904	0,061019
15	0,043034	0,051464	0,053501	0,055755	0,070202	0,069281	0,073966
16	0,043213	0,051809	0,034769	0,049499	0,068635	0,065844	0,07045
17	0,024853	0,032823	0,052916	0,047661	0,03716	0,04838	0,045304
18	0,038616	0,049765	0,05781	0,062168	0,055068	0,066836	0,063918
19	0,026936	0,04088	0,034354	0,049585	0,043751	0,055933	0,051803
20	0,036797	0,049299	0,051076	0,061499	0,047515	0,064195	0,057847

21	0,056832	0,065837	0,078588	0,080496	0,06473	0,080709	0,075096
22	0,03102	0,044211	0,041103	0,054941	0,041207	0,057983	0,051127
23	0,043859	0,057326	0,048577	0,067393	0,053418	0,071134	0,063734
24	0,05628	0,06825	0,069573	0,080943	0,06598	0,083524	0,076887

Таблиця Б.7.3. Первинні ланцюжки (кластери)

Ланцюжок 1 (4):	1	0,013968	2	0,040734	3	0,037619	4
Ланцюжок 2 (4):	5	0,012278	7	0,016679	6	0,020458	8
Ланцюжок 3 (2):	9	0,016582	11				
Ланцюжок 4 (4):	10	0,009451	13	0,016995	14	0,018715	15
Ланцюжок 5 (2):	12	0,015794	16				
Ланцюжок 6 (2):	17	0,020937	18				
Ланцюжок 7 (4):	19	0,01064	22	0,010834	20	0,013894	23
Ланцюжок 8 (2):	21	0,020367	24				

Таблиця Б.7.4. Матриця між ланцюжкових відстаней (фрагмент)

Матриця міжланцюжкових відстаней Середнього звязку

	Ланцюжок1	*Ланцюжок2*	*Ланцюжок3*	*Ланцюжок4*	*Ланцюжок5*	*Ланцюжок6*
Ланцюжок1	0 (0; 0)	0,032329 (0; 0)	0,055701 (0; 0)	0,044888 (0; 0)	0,050713 (0; 0)	0,045827 (0; 0)
Ланцюжок2	0,032329 (0; 0)	0 (0; 0)	0,06957 (0; 0)	0,060993 (0; 0)	0,073367 (0; 0)	0,056214 (0; 0)
Ланцюжок3	0,055701 (0; 0)	0,06957 (0; 0)	0 (0; 0)	0,02291 (0; 0)	0,0585 (0; 0)	0,042029 (0; 0)
Ланцюжок4	0,044888 (0; 0)	0,060993 (0; 0)	0,02291 (0; 0)	0 (0; 0)	0,04492 (0; 0)	0,034474 (0; 0)
Ланцюжок5	0,050713 (0; 0)	0,073367 (0; 0)	0,0585 (0; 0)	0,04492 (0; 0)	0 (0; 0)	0,058485 (0; 0)
Ланцюжок6	0,045827 (0; 0)	0,056214 (0; 0)	0,042029 (0; 0)	0,034474 (0; 0)	0,058485 (0; 0)	0 (0; 0)
Ланцюжок7	0,046178 (0; 0)	0,057271 (0; 0)	0,060019 (0; 0)	0,047561 (0; 0)	0,055231 (0; 0)	0,025443 (0; 0)
Ланцюжок8	0,0696 (0; 0)	0,078216 (0; 0)	0,060675 (0; 0)	0,055308 (0; 0)	0,076022 (0; 0)	0,02943 (0; 0)
Мін. відстань Між ланцюжками	0,032329 (1; 2)	0,032329 (2; 1)	0,02291 (3; 4)	0,02291 (4; 3)	0,04492 (5; 4)	0,025443 (6; 7)

Таблиця Б.7.5. Інформація про кластери

Сформовано кластерів:	3		
Номер кластера	Склад кластера в ланцюжках		
1 –й	1	2	
2 –й	3	4	5
3 –й	6	7	8

Таблиця Б.7.6. Інформація про дендрит

Об`єднання кластерів в один	
Ланцюжки 1 і 6	
Відстань	0,024853
Ланцюжки 6 і 4	
Відстань	0,025247

Таблиця Б.7.7. Інформація про критичні відстані

Середня відстань по ланцюжку		0,019001	Середнє квадратичне по ланцюжку		0,00541
Значення коефіцієнта	0	1	1,25	1,5	2
Критична відстань	0,019001	0,024411	0,025763	0,027116	0,029821

Б.7.2. Нечіткий кластерний аналіз

Спочатку готуємо вихідні дані. Це матриця опису об'єктів (табл. Б.7.8).

Таблиця Б.7.8. Вихідні дані

Назви об'єктів	Назви факторів	
	X1	X2
1	1	1
2	1	4
3	1	7
4	3	2
5	3	4
6	3	6
7	5	4
8	7	4
9	9	4
10	11	2
11	11	4
12	11	6
13	13	1
14	13	4
15	13	7

Вибираємо послідовно пункти меню «Макрос», «Макросы» і в діалоговому вікні, яке з'явилося, запускаємо макрос FuzzyClasterU. Після запуску макросу з'явиться запит на введення матриці вихідних даних (рис. Б.7.11). Форму матриці показано в табл..1. Матриця відмічається без заголовків.

Якщо кількість стовпців менше 2 або число рядків менше 4, то з'являється повідомлення про помилку (рис. Б.7.12), і виконується повернення на введення матриці вихідних даних..

Після цього необхідно ввести матрицю імен об'єктів (рис. Б.7.13).

Якщо довжина введеної матриці імен не відповідає розмірам вихідної матриці, то після повідомлення про помилку (рис. Б.7.14) необхідно ввести імена знову.

Після цього програма видає запит на місце, з якого почнеться виводитись результат (рис. Б.7.15).

Потім послідовно вводяться параметри алгоритму:

1. Кількість кластерів, на які необхідно розбити вибірку (рис. Б.7.16). Якщо введена кількість кластерів менше двох або більше кількості об'єктів, то з'являється повідомлення про помилку (рис. Б.7.17) і повторний запит на введення.
2. Значення параметру зупинки процесу (рис. Б.7.18). При помилковому значенні (менше чи рівне нулю, або більше 3) повідомляється про помилку (рис. Б.7.19) і значення необхідно ввести заново. Якщо натиснути ОК, то залишиться значення параметру за умовчуванням.

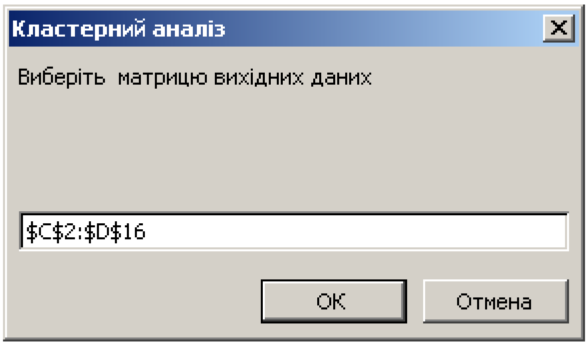


Рис. Б.7.11. Запит вибірки

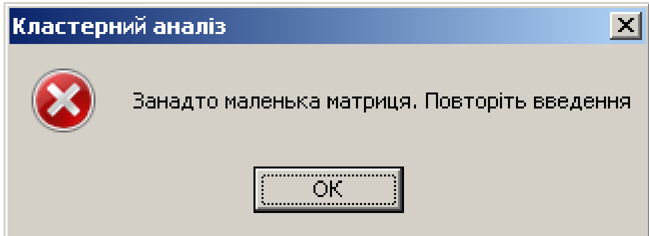


Рис. Б.7.12 Помилкова матриця

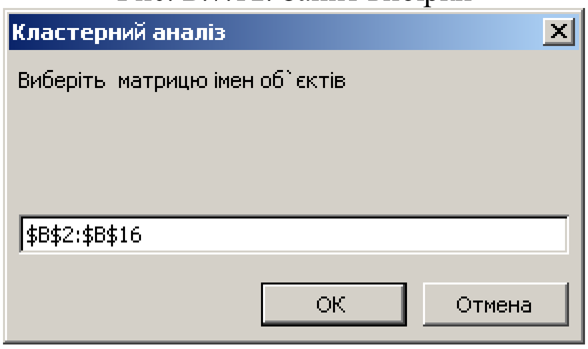


Рис. Б.7.13. Запит імен об'єктів

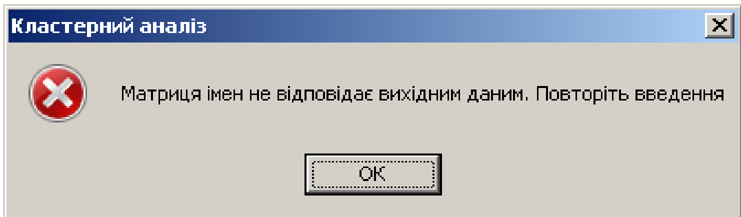


Рис. Б.7.14. Конфлікт розмірів вибірки і списку імен

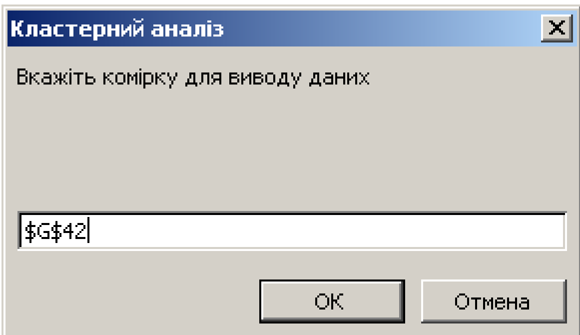


Рис. Б.7.15. Запит місця виводу

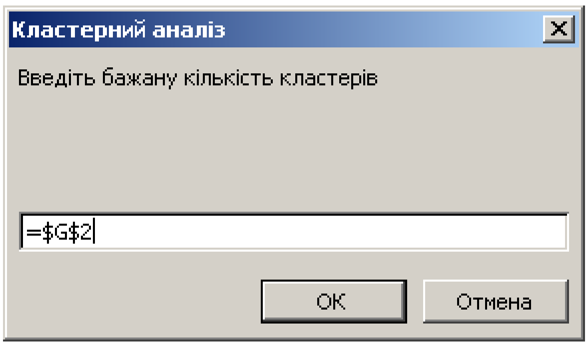


Рис. Б.7.16. Запит кількості кластерів

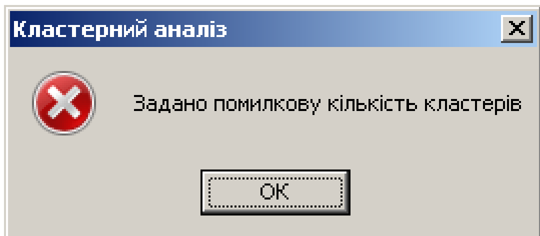


Рис. Б.7.17. Задано мало або багато кластерів

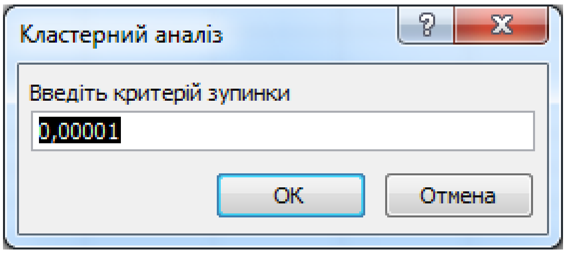


Рис. Б.7.18. Введення точності

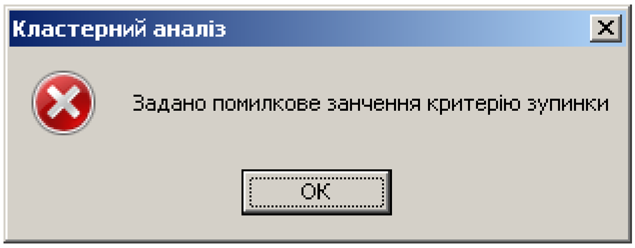


Рис. Б.7.19. Помилкове значення точності

Результати роботи мають вигляд таблиці (табл.. Б.7.9)

Таблиця Б.7.9. Результати нечіткого кластерного аналізу

Приналежність до кластерів			
		Номер кластера	
Назва об`єкту		1	2
	1	0,091363	0,908637
	2	0,024438	0,975562
	3	0,09132	0,90868
	4	0,053	0,947
	5	0,001861	0,998139
	6	0,052952	0,947048
	7	0,121215	0,878785
	8	0,50048	0,49952
	9	0,879302	0,120698
	10	0,947131	0,052869
	11	0,998184	0,001816
	12	0,947085	0,052915
	13	0,908644	0,091356
	14	0,97546	0,02454
	15	0,908602	0,091398
Wj/n		2,765395	2,763596
S/n		15,66161	15,66161
d2jmax		9,639401	9,636905
Кластер?		Так	Так
Сгущення?		Так	Так
Параметр експоненціальної ваги		2	
Критерій зупинки		0,00001	
Внутрігрупове розсіяння		82,93487	
Міжгрупове розсіяння		152,4181	
Якість розділення		0,646972	

Б.8. Лабораторна робота №8. Дискримінантний аналіз

Б.8.1. Дискримінація

Розглянемо учбову задачу, яка дозволяє наочну геометричну інтерпретацію. В табл. Б.8.1 представлені вихідні дані. Ми маємо три класи X1(B5:D10), X2(B13:D20) і X3(B22:D28), а також об'єкти P1–P3 (B29:D31) , приналежність якого до одного з цих класів необхідно визначити.

Таблиця Б.8.1. Вихідні дані для дискримінантного аналізу

		X1	X2
		1	2
	1	5	6
	2	4	7
	3	6	7
	4	3	8
X1	5	7	8
	6	4	9
	7	6	9
	8	5	10
	9	1	2
	10	2	4
	11	2	3
	12	2	3
X2	13	3	3,5
	14	3	3
	15	3	2
	16	3,5	4
	17	6	4
	18	6,5	2
X3	19	7	4
	20	8	3
	21	8	1
	22	8	4
	23	9	3
	24	9	1
	P1	5	8
	P2	4	2
	P3	7	2

Для виконання задачі класифікації за допомогою дискримінантного аналізу необхідно виконати наступні дії. В меню послідовно вибираємо «Сервис», «Макрос», «Макросы» (або «Вид», «Макросы», «Макросы» в більш нових версіях). З'являється вікно (рис.Б.8.1), в якому необхідно вибрати discrim_n_m. Після запуску макросу у відповідь на запит (рис.Б.8.2) необхідно ввести інформацію про кількість класів. Якщо кількість класів менше двох, або більше 16, то після повідомлення про помилку (рис.Б.8.3) необхідно повторити введення.

Після цього потрібно ввести навчальні вибірки. Вони вводяться на запити виду (рис.Б.8.4), в яких змінюється тільки номер вибірки (1-й, 2-й тощо).

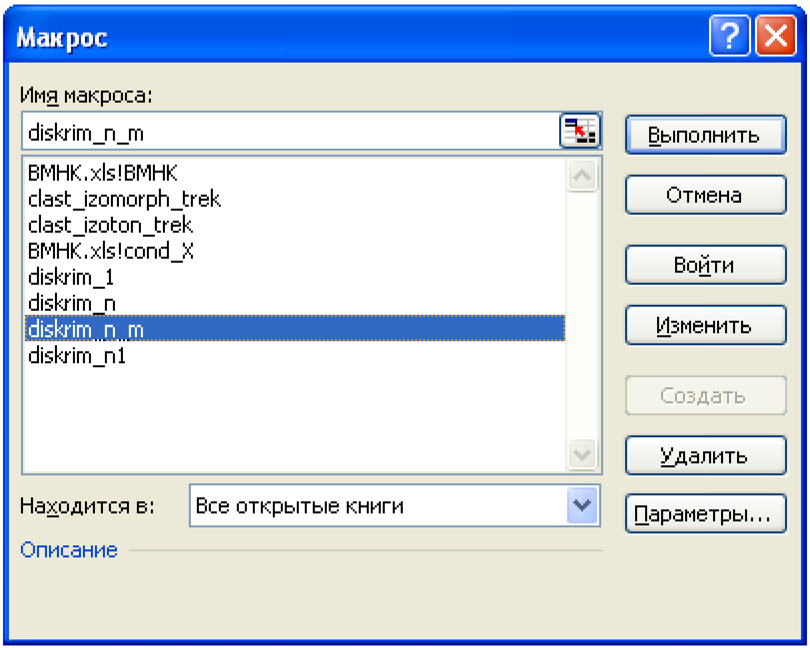


Рис.Б.8.1 Форма вибору макросу

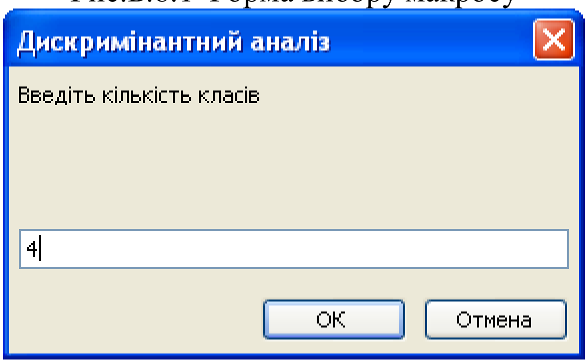


Рис. Б.8.2. Запит інформації про кількість класів

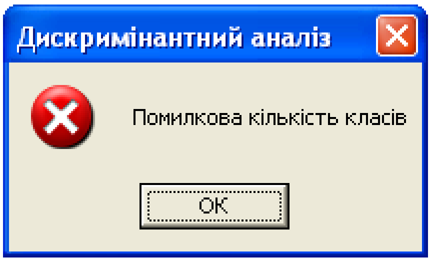


Рис. Б.8.3. Повідомлення про помилкову кількість класів

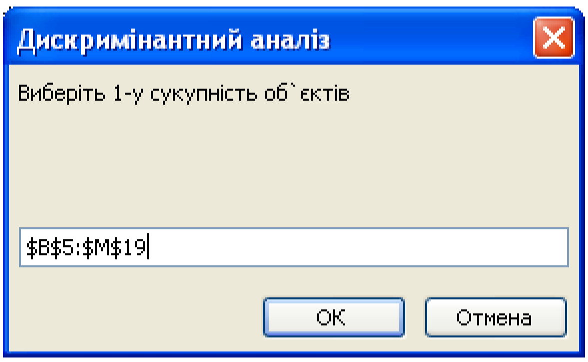


Рис. Б.8.4. Запит на введення інформації по навчальній вибірці

Якщо в вибірках буде не однакова кількість факторів, то після повідомлення про помилку (Рис.Б.8.5) необхідно буде вводити всі вибірки спочатку.

Далі потрібно ввести опис об'єкту, належність якого до якогось із класів необхідно визначити (рис. Б.8.9).

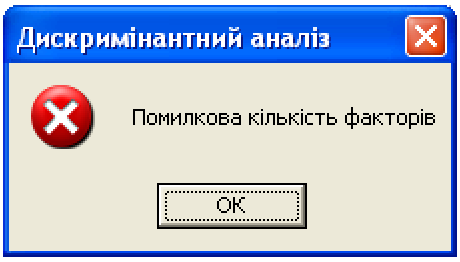


Рис.Б.8.5. Різна кількість факторів в вибірках

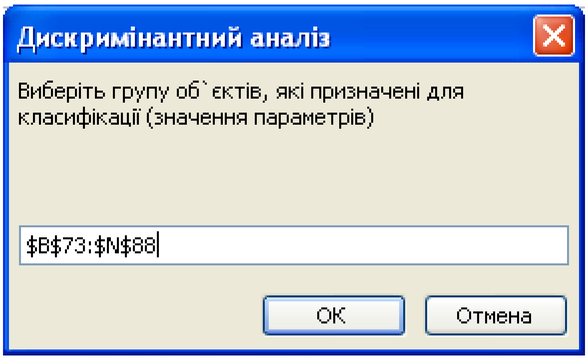


Рис. Б.8.6. Запит інформації про об'єкти, який необхідно класифікувати

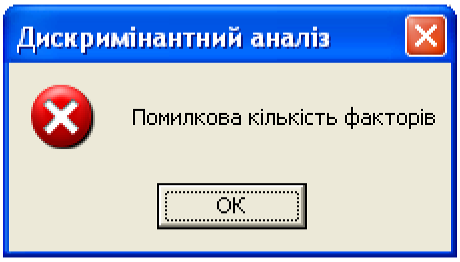


Рис.Б.8.7. Помилка: кількість факторів в вибірці для класифікації не відповідає навчальним

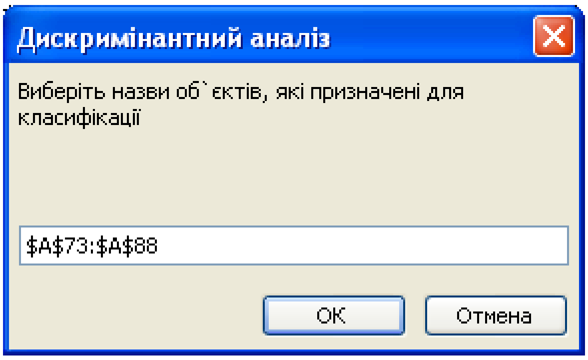


Рис.Б.8.8. Запит на введення імен об'єктів для класифікації

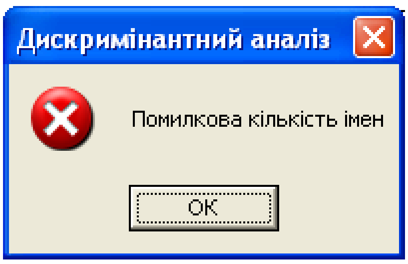


Рис.Б.8.9. Помилка: кількість імен не відповідає попередньо введеним об'єктам

Наостанок вводиться інформація про місце виведення результату (рис.Б.8.10). Результати роботи макросу дискримінантного аналізу представлені на рис.Б.8.11.

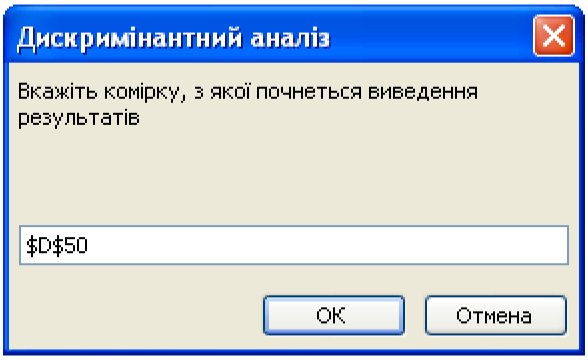


Рис. Б.8.10. Запит місця виведення результату

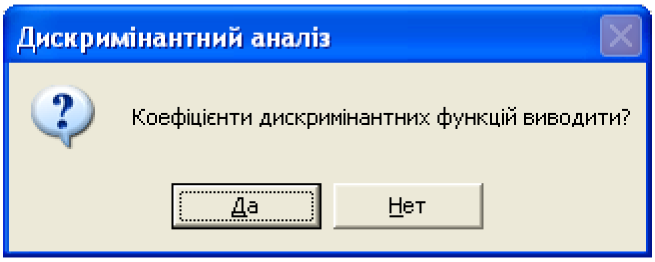
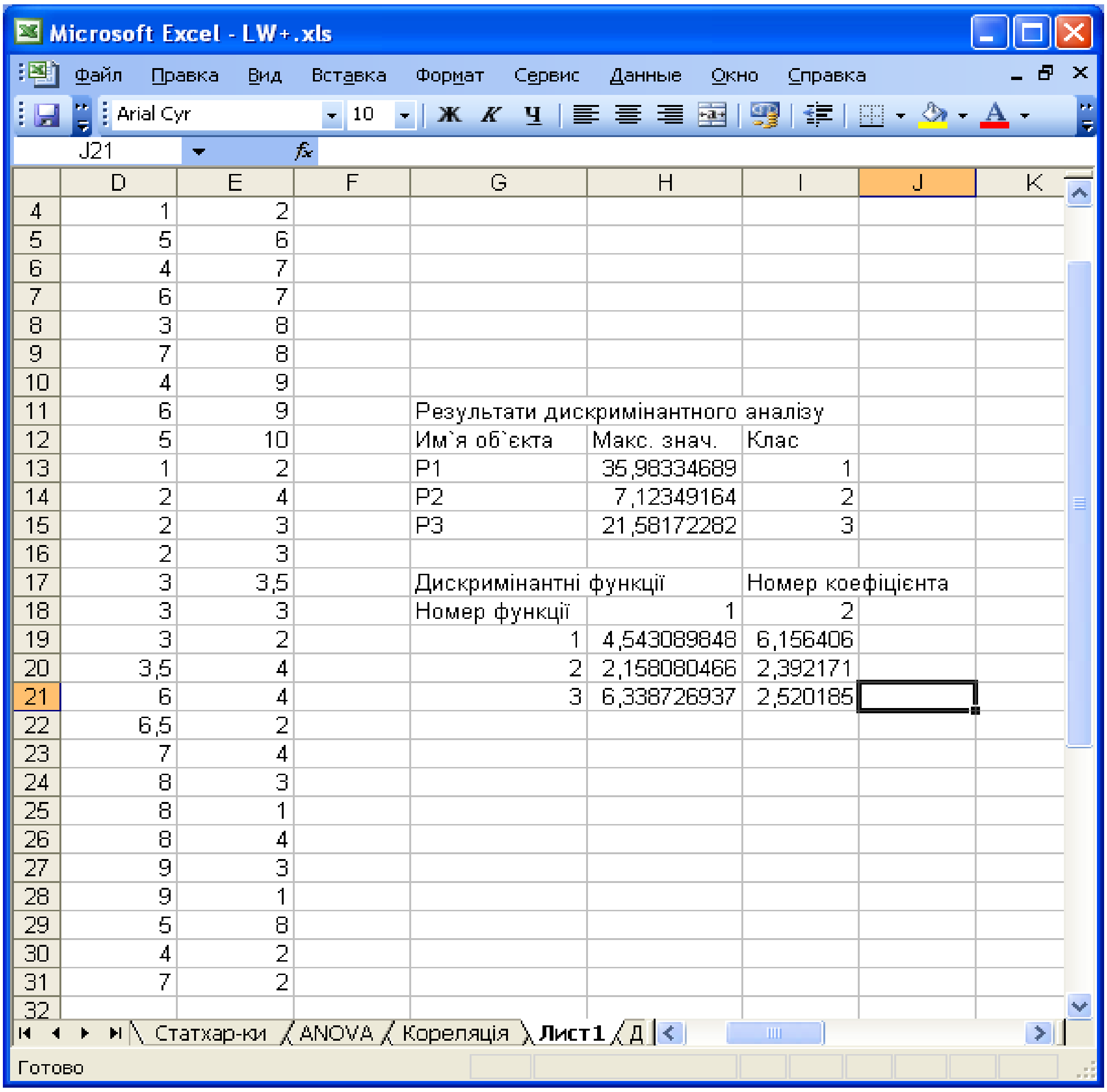


Рис. Б.8.11. Запит на виведення коефіцієнтів дискримінант них функцій



Microsoft Excel - LW+.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

J21

	D	E	F	G	H	I	J	K
4	1	2						
5	5	6						
6	4	7						
7	6	7						
8	3	8						
9	7	8						
10	4	9						
11	6	9		Результати дискримінантного аналізу				
12	5	10		Им`я об`єкта	Макс. знач.	Клас		
13	1	2		P1	35,98334689	1		
14	2	4		P2	7,12349164	2		
15	2	3		P3	21,58172282	3		
16	2	3						
17	3	3,5		Дискримінантні функції		Номер коефіцієнта		
18	3	3		Номер функції	1	2		
19	3	2		1	4,543089848	6,156406		
20	3,5	4		2	2,158080466	2,392171		
21	6	4		3	6,338726937	2,520185		
22	6,5	2						
23	7	4						
24	8	3						
25	8	1						
26	8	4						
27	9	3						
28	9	1						
29	5	8						
30	4	2						
31	7	2						
32								

Статхар-ки / ANOVA / Кореляція / Лист1 / Д

Готово

Рис. Б.8.12. Результати роботи

На рис.Б.8.13 можна побачити геометричну інтерпретацію учбового прикладу. Зрозуміло, що в реальних задачах число факторів перевищує не тільки два, але 3 три і проста інтерпретація неможлива.

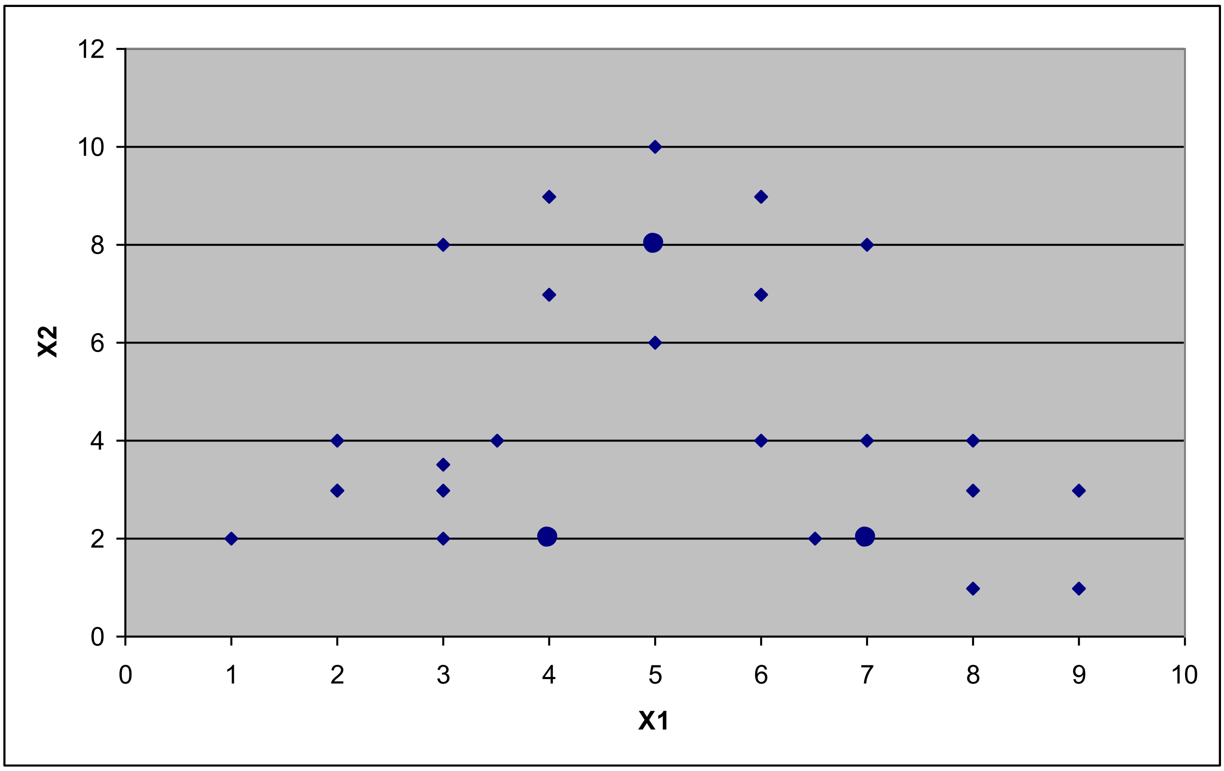


Рис. Б.8.13. Форма навчальних вибірок прикладу

Б.8.2. Покроковий дискримінантний аналіз

Спочатку готується вихідні дані (див. приклад табл..1). В даному випадку маємо дві змінні (M=2) , три групи (K=3) і всього 22 (N=22) спостереження.

Таблиця 1. Вихідні дані

Група	Об'єкт	X1	X2
1	1	2	5
1	2	2,2	4
1	3	2,4	7,5
1	4	2,5	6
1	5	2,6	4
1	6	2,8	9
1	7	3	7
2	8	4,5	6,5
2	9	4,5	7,5
2	10	5	10
2	11	5,3	4,7
2	12	5,3	5,1
2	13	5,5	11
2	14	6	8
3	15	7,5	5
3	16	7,6	4
3	17	8,5	7
3	18	8,5	6
3	19	8,5	3
3	20	8	4
3	21	9	6,5
3	22	9	3

Після цього набравши «Сервис», «Макрос», «Макросы» (або «Разработчик», «Макросы» для 2007) отримуємо вікно виклику макросів (рис.1). В Вікні вибираємо макрос StepWiseDiskrim. Після запуску макрос запитує необхідну вхідну інформацію. Спочатку потрібно ввести кількість груп (рис. 2). Якщо введена кількість менше двох чи більше 6, видається повідомлення про помилку (рис. 3), після якого потрібно повторити введення числа груп.

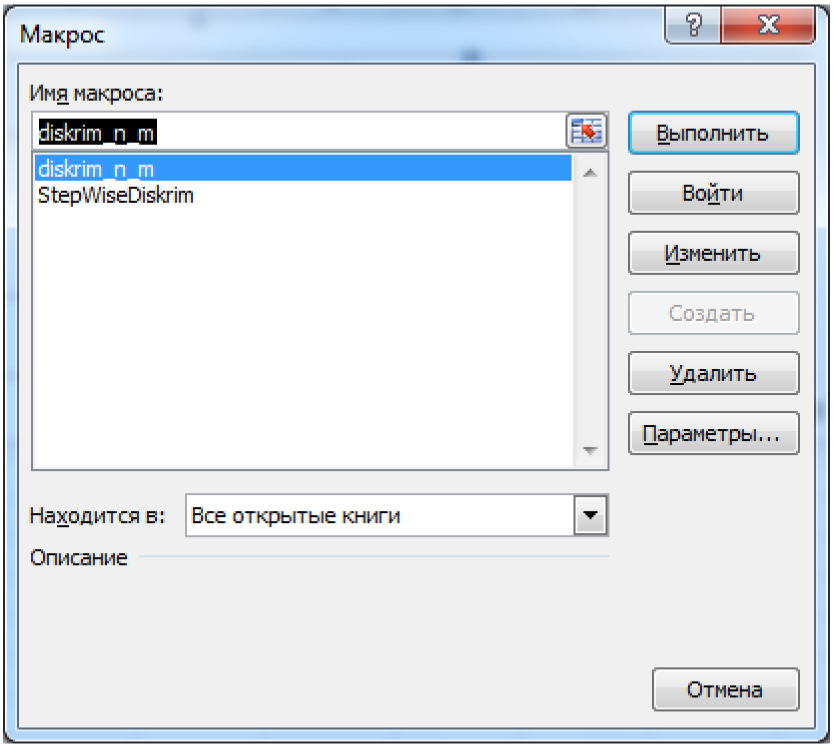


Рис. 1. Вікно запуску макросу

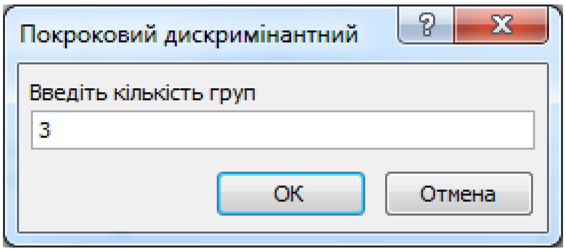


Рис. 2. Запит кількості навчальних груп

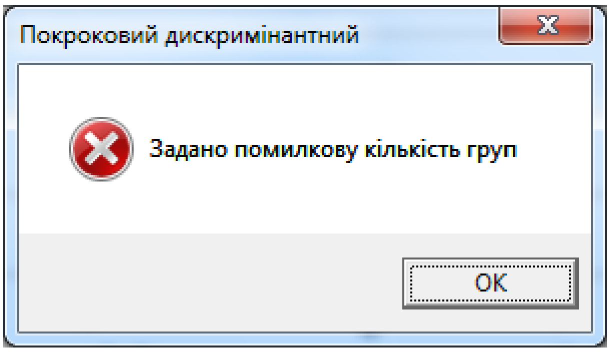


Рис. 3. Повідомлення про помилку при введенні кількості груп

Після цього послідовно вводяться дані всіх груп по одній (рис.4). Якщо користувач в групах ввів групи з різною кількістю змінних, то після повідомлення про помилку (рис.5) потрібно повторити введення.

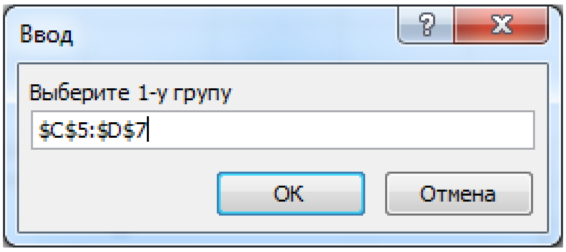


Рис. 4. Внедення даних поточної групи

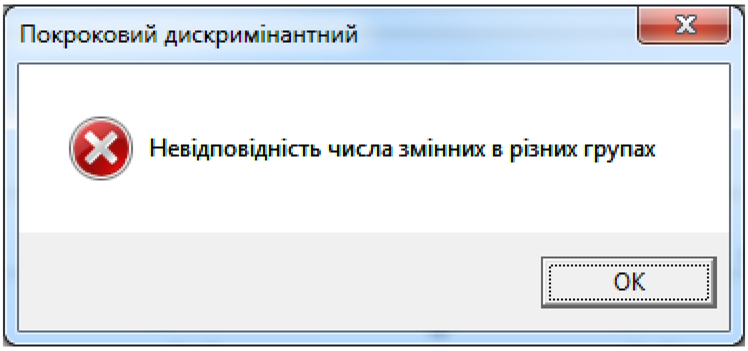


Рис. 5. Помилка при введенні поточної групи

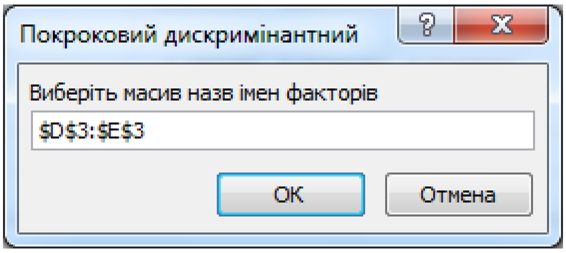


Рис.6. Запит назв імен факторів

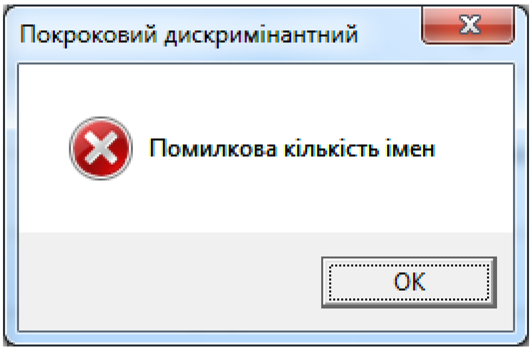


Рис. 7. Невідповідність кількості імен розмірам введених груп

Потім макрос видає запит на введення рівня значущості (рис. 8). Якщо вас влаштовує стандартний рівень, то достатньо вибрати «ОК» або Enter. При введенні неправильного значення – за межами інтервалу (0, 1) видається повідомлення про помилку (рис.9), після якого необхідно повторити введення.

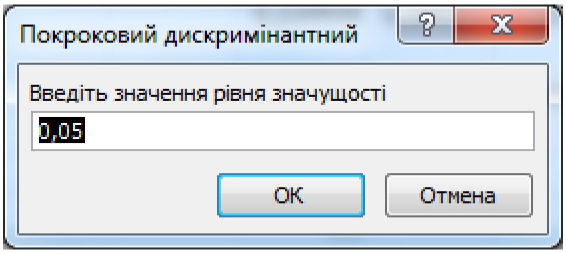


Рис. 8. Запит рівня значущості для перевірки критерію Фішера

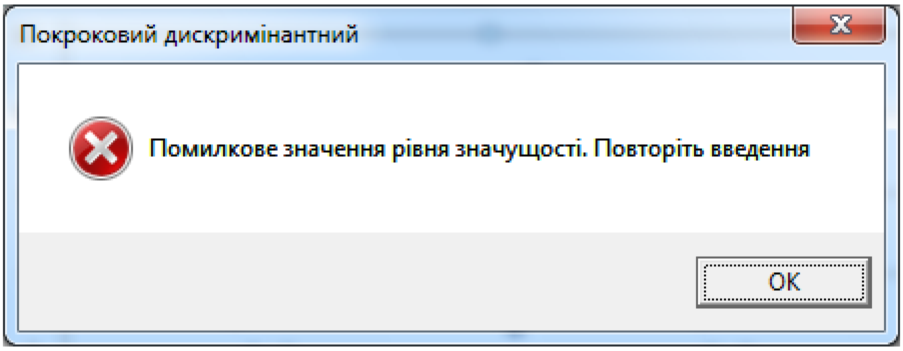


Рис. 9. Помилка при введенні рівня значущості

Після цього макрос видає запит на введення рівня мультиколінеарності (рис. 10). Якщо вас влаштовує стандартний рівень, то достатньо вибрати «ОК» або Enter. При введенні неправильного значення – за межами інтервалу (0,3; 1] видається повідомлення про помилку (рис.11), після якого необхідно повторити введення.

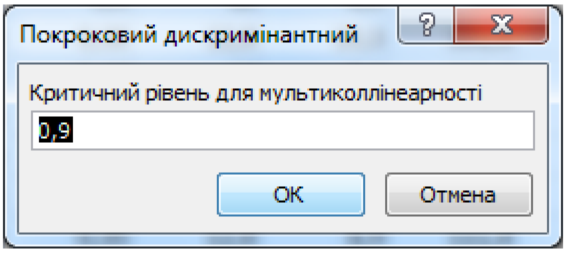


Рис. 10. Запит на рівень мультиколінеарності

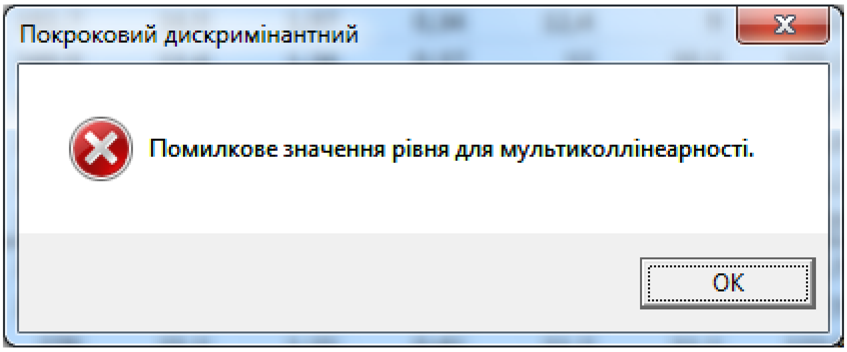


Рис. 11. Помилка при введенні рівня мультиколінеарності

Останнім запитується місце виведення результату (рис. 12).

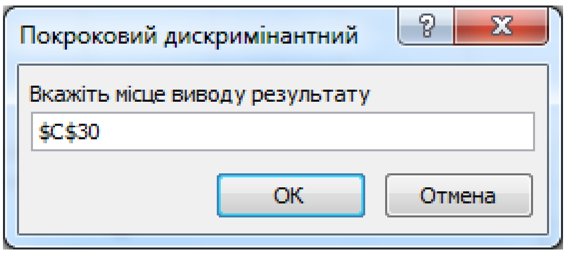


Рис. 12. Запит на місце виведення результатів

Приклади розрахунків.

Приклад 1.

Результатом роботи макросу є таблиця (табл. 2, 3), яка включає тільки ті змінні, які необхідно включати в дискримінаційну модель. Для кожної змінної виводиться також розрахункове і критичне значення F-включення. Якщо по результатах перевірки на контрольних дослідах цих змінних виявиться недостатньо, то необхідно виконати новий розрахунок, збільшивши значення рівня значущості.

Таблиця 2. Результати роботи

Ім’я фактору	F-розрахункове	F-критичне
1	240,7197	19,44314
Рівень значущості	0,05	
Критичний рівень для мультиколлінеарності	0,9	

В даному прикладі в таблиці одна змінна. Як добре видно з рис. 13 саме цієї однієї змінної достатньо для класифікації груп в тестовій задачі.

Імя фактору		F-розрахункове включення	F-критичне	F-розрахункове виключення	F-критичне
X1		240,7197	19,44022	9,157616	19,44314
Рівень значущості		0,05			
Критичний рівень для мультиколлінеарності		0,9			
Класс		Власне значення	Відносний % склад	Канонічна кореляція	
	1	13,78531	0,07009	0,965591	
	2	51,37267	0,261199	0,990407	
	3	131,5224	0,668711	0,99622	
Статистика Уілкса					
Номер функції		Лямбда статистика	Статистика хі-квадрат розр.	Статистика хі-квадрат кр.	
	0	9,74E-06	213,4673	9,487729	
	1	0,000144	163,635	3,841459	

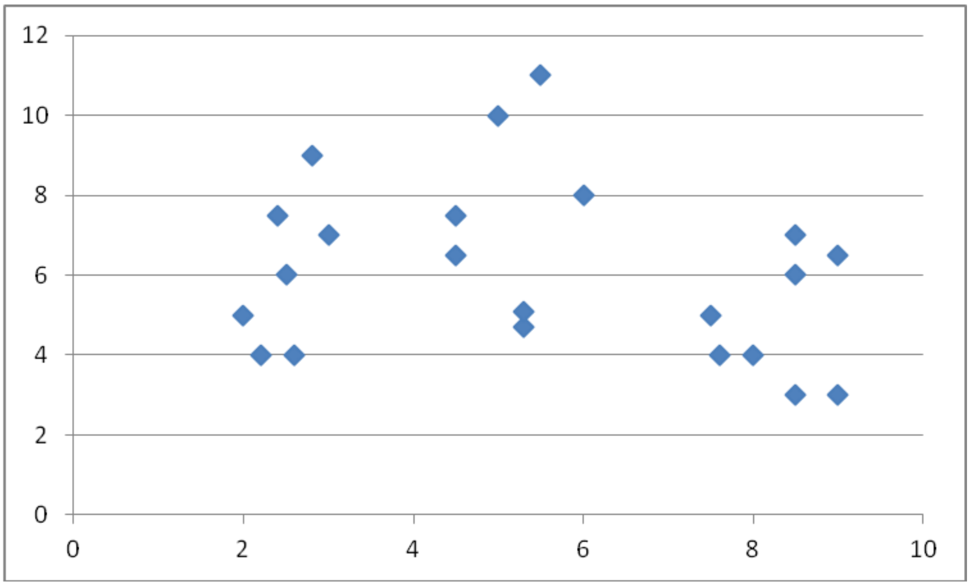
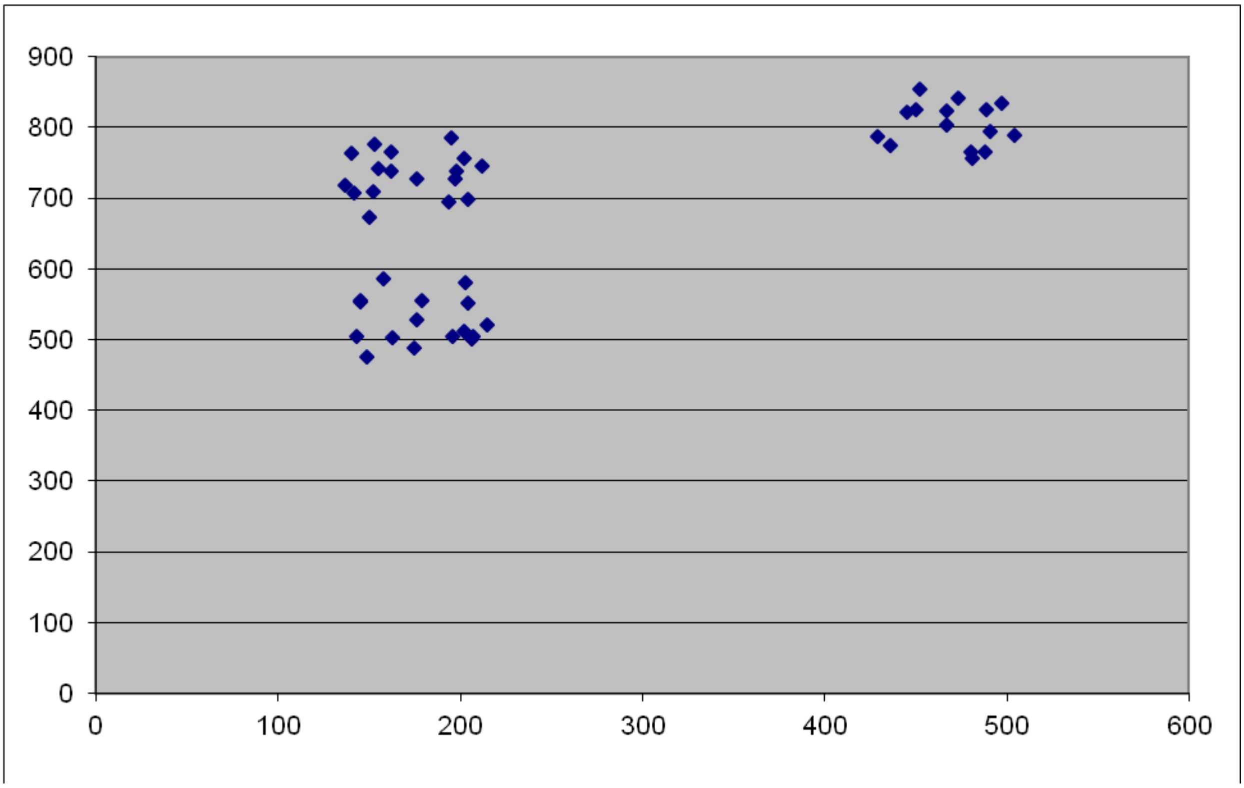


Рис. 13. Діаграма розсіяння для тестової задачі

Приклад 2.

Таблиця 3. Вихідні

	X	Y		X	Y		X	Y
1	728	176	17	803	467	33	528	176
2	739	198	18	825	450	34	512	202
3	695	194	19	775	436	35	555	179
4	707	142	20	765	480	36	552	204
5	742	155	21	794	491	37	503	163
6	766	162	22	825	489	38	488	175
7	727	197	23	842	473	39	501	206
8	710	152	24	821	445	40	521	215
9	718	137	25	787	429	41	505	143
10	757	202	26	757	481	42	556	145
11	786	195	27	789	504	43	553	145
12	776	153	28	765	488	44	505	196
13	673	150	29	854	452	45	580	203
14	698	204	30	834	497	46	475	149
15	746	212	31	823	467	47	504	207
16	739	162	32	764	140	48	586	158



Імя фактору	F-розрахункове включення	F-критичне	F-розрахункове виключення	F-критичне
X	316,7186	19,47301	21,03861	19,47351
Y	131,8937	19,47248	18,91643	19,47301

Рівень значущості	0,05
Критичний рівень для мультиколлінеарності	0,9

Класс		Власне значення	Відносний % склад	Канонічна кореляція
	1	259,6041	0,363633	0,99808
	2	319,8436	0,448011	0,99844
	3	134,471	0,188356	0,996302

Статистика Уілкса

Номер функції		Лямбда статистика	Статистика хі-квадрат розр.	Статистика хі-квадрат кр.
	0	8,83E-08	722,8008	9,487729
	1	2,3E-05	475,2472	3,841459

Приклад 3.

В табл.3 Приведено вигляд результатів у разі числа змінних більше двох і наявності мультиколінеарності.

Таблиця 3. Повний варіант виведення вихідних результатів.

Мультиколінеарність в групі 1

Фактор	Коефіцієнт кореляції	Фактор
X3	0,962255	X4

Мультиколінеарність в групі 2

Фактор	Коефіцієнт кореляції	Фактор
X2	0,966139	X4

Мультиколінеарність в групі 3

Фактор	Коефіцієнт кореляції	Фактор
X1	0,955717	X5
X2	0,966403	X4
X3	0,962192	X4
X3	0,955603	X8

Мультиколінеарність в групі 4

Фактор	Коефіцієнт кореляції	Фактор
X3	0,963085	X4

Імя фактору	F-розрахункове включення	F-критичне	F-розрахункове виключення	F-критичне
X12	36,94565	5,152749	12,47585	5,152415
X1	8,462647	5,153096	5,862925	5,152749
X5	5,590334	5,153455	4,326698	5,153096

Рівень значущості	0,1
Критичний рівень для мультиколлінеарності	0,95

Класс		Власне значення	Відносний % склад	Канонічна кореляція
	1	12035,37	0,251325	0,999958
	2	11891,17	0,248314	0,999958
	3	11904,72	0,248597	0,999958
	4	12056,34	0,251763	0,999959

Статистика Уілкса

Номер функції		Лямбда статистика	Статистика хі-квадрат розр.	Статистика хі-квадрат кр.
	0	4,87E-17	1915,638	47,21217
	1	5,86E-13	1436,458	30,81328
	2	6,97E-09	957,8924	15,98718

Література

- С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин Прикладная статистика: Исследование зависимостей –М.: Финансы и статистика, 1985. –487с.
- С.Н. Лапач, А.В. Чубенко, П.Н. Бабич Статистика в науке и бизнесе –К.: 2002, Морион. – 640с.
- П.Г. Кацев Статистические методы исследования режущего инструмента. Изд. 2-е, перераб. И доп. М.: Машиностроение, 1974. - 231с.
- М.Н.Степнов Статистические методы обработки результатов механических испытаний: Справочник. М. Машиностроение, 1985. 232с.