

МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
„ЛЬВІВСЬКА ПОЛІТЕХНІКА”

ДМИТРІВ І. В.

**ТЕОРІЯ І ПРАКТИКА
НАУКОВИХ ДОСЛІДЖЕНЬ:
АВТОМОБІЛЬНИЙ
ТРАНСПОРТ**

ЛАБОРАТОРНИЙ ПРАКТИКУМ

ЛЬВІВ
СПОЛОМ
2022

УДК 519.242.7
ДМ 127

*Рекомендовано до друку (гриф) Науково-методичною радою
Національного університету „Львівська політехніка”
Міністерства освіти і науки України
(протокол № 61 від 21.02.2022 р.)*

Рецензенти:

- О. С. Ланець**, доктор технічних наук, професор (директор навчально-наукового інституту Механічної інженерії та транспорту, Національний університет „Львівська політехніка”);
- В. В. Братішко**, доктор технічних наук, старший науковий співробітник (декан механіко-технологічного факультету, Національного університету біоресурсів і природокористування України);
- Є. І. Калінін**, доктор технічних наук, професор (професор кафедри обчислювальної техніки та програмування, Національного технічного університету "Харківський політехнічний інститут").

Дмитрів, Ігор Васильович.

Теорія і практика наукових досліджень: Автомобільний транспорт: лабр. практ. / Національний університет „Львівська політехніка”. – Львів : СПОЛОМ, 2022. – 112 с. : рис., табл. – Бібліогр. : с. 97-101 (33 назви).

Розглянуто методи побудови математичних моделей технічного об'єкта, розв'язку дисперсійного, кореляційного та регресійного аналізу. Наведено теорію оброблення результатів багатofакторного експерименту з використанням програмного пакету STATISTICA.

Для студентів вищих навчальних закладів, які навчаються за освітнім ступенем „Магістр” спеціальностей 274 – „Автомобільний транспорт” та 133 – „Галузеве машинобудування”.

© І.В. Дмитрів, 2022
© Національний університет „Львівська політехніка”, 2022
ISBN 978-966-919-813-6 © В-во «СПОЛОМ», 2022

ЗМІСТ

ПЕРЕДМОВА.....	4
Лабораторна робота 1	
Основи моделювання в середовищі STATISTICA.....	6
Лабораторна робота 2	
Методика вибору оптимальної регресійної моделі.....	18
Лабораторна робота 3	
Визначення адекватності регресійної моделі.....	33
Лабораторна робота 4	
Визначення точності прогнозу лінійної моделі.....	45
Лабораторна робота 5	
Перевірка факторів на мультиколінеарність.....	56
Лабораторна робота 6	
Дисперсійний аналіз в середовищі STATISTICA.....	73
Лабораторна робота 7	
Кореляційно-регресійний аналіз в середовищі STATISTICA...	87
СПИСОК ЛІТЕРАТУРИ.....	97
ДОДАТКИ.....	102

ПЕРЕДМОВА

Дисципліна “Теорія і практика наукових досліджень” спрямована на забезпечення підготовки висококваліфікованих інженерів-дослідників та формування у них сучасний інструментарій планування, виконання та оброблення результатів експериментальних досліджень. Головним завданням дисципліни є вивчення сучасних методів досліджень, оптимізації технічних процесів і багатофакторного планування експериментальних досліджень щодо проблем механізації й автоматизації технологічних процесів в галузі транспорту.

Статистичний аналіз даних є невід’ємним атрибутом наукових досліджень на всіх рівнях. Оволодіння методами статистичної обробки даних експериментальних досліджень із використанням комп’ютерних технологій є важливою складовою професійної підготовки майбутнього науковця.

Програмне забезпечення статистичних досліджень достатньо розвинуте. На сьогоднішній день відомо багато статистичних пакетів прикладних програм статистичної обробки інформації. Серед них: STATISTICA, SPSS, SAS, MathCAD, MathLab, Statgraphics та інші.

Всі статистичні процедури системи розділено на окремі модулі, кожен з яких об’єднує групу логічно зв’язаних між собою статистичних методів і в рамках конкретної моделі забезпечує повний і всебічний аналіз закономірностей. У кожному модулі можна виконати певний спосіб обробки даних,

не звертаючись до інших модулів. У системі STATISTICA такими модулями є: основні статистики та таблиці, множинна регресія, дисперсійний аналіз, багатовимірний аналіз та ін.

Важливою характеристикою системи є наявність засобів всебічної графічної підтримки процесу обробки даних і візуалізації результатів аналізу. Графічні можливості й засоби системи унікальні. Вона включає сотні різних типів статистичних графіків (включаючи наукові, ділові, тривимірні та двовимірні в різних системах координат, спеціалізовані, статистичні графіки – гістограми, матричні, категоризовані графіки та ін.

У системі STATISTICA існує багато інструментів налаштування всіх компонентів графіків. Є можливість вибору різних типів ліній, форматів, кольорів, легенд, назв та інших атрибутів графіка.

Важливою характеристикою системи STATISTICA є те, що вона являється повноцінною прикладною програмою Windows.

Метою даного лабораторного практикуму є ознайомлення здобувачів вищої освіти з основами роботи в системі STATISTICA, зокрема з важливими її модулями.

ЛАБОРАТОРНА РОБОТА №1

ОСНОВИ МОДЕЛЮВАННЯ В СЕРЕДОВИЩІ STATISTICA

Мета роботи: Отримати практичні навички створення та редагування таблиць даних, автозвітів, знаходження графіків і рівнянь лінійної регресії.

Обладнання робочого місця: персональний комп'ютер з необхідним програмним забезпеченням, методичні рекомендації, довідкова література.

Загальні відомості

Генеральна сукупність. Загалом за необмеженого зростання обсягу вибірки вона переходить у генеральну сукупність – тобто вибірку, яка охоплює всі свої можливі значення. Генеральна сукупність (гранична вибірка) може мати скінченну кількість своїх значень або теоретично бути безмежною. Одиницею генеральної сукупності називається елементарна подія і значення випадкової величини, що їй відповідає.

Для досліджень властивостей випадкової величини використовуються її незалежні послідовні результати спостережень, які записують у вигляді послідовного ряду:

$$x_1, x_2, \dots, x_n. \quad (1.1)$$

Отриманий ряд називають **вибіркою** x_n , членами якої є отримані результати (x_n, y_n) (табл. 1.1), де n – обсяг вибірки; x – фактор; y – відклик.

Тобто вибірка – це результат послідовних і незалежних спостережень над випадковою величиною, що представляє генеральну сукупність.

Таблиця 1.1

Табличне зображення вибірки

n	x	y
1	x_1	y_1
2	x_2	y_2
...
n	x_n	y_n

Кореляційне поле (діаграма розсіювання) (рис. 1.1) – сукупність точок, розташованих у системі координат, що характеризує взаємозв'язок між двома ознаками: факторною x_i (на осі абсцис) та результативною (відкликом) y_i (на осі ординат). Використовується для аналізу наявності та характеру (напрямку) зв'язку між результатами експериментальних досліджень. Розміщення точок на графіку свідчить про наявність і напрям зв'язку.

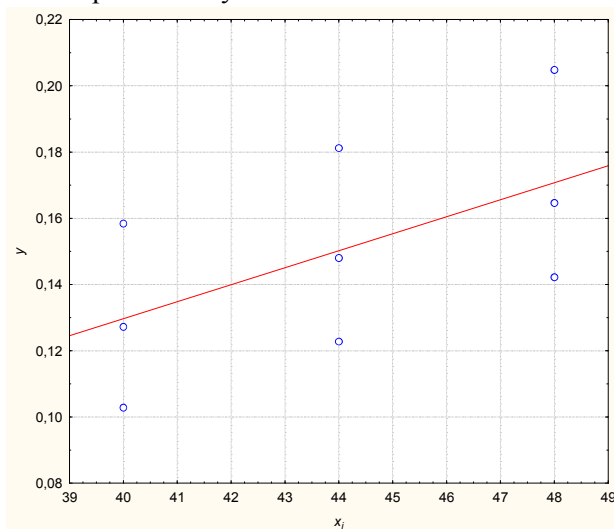


Рис. 1.1. Загальна структура наукового дослідження

Математичне моделювання – моделювання, при якому модель є системою математичних співвідношень, що описують певні технологічні, технічні чи інші процеси.

Математична модель об'єкта дослідження - це система математичних співвідношень (формул, рівнянь чи систем рівнянь), які описують основні елементи та їх взаємодію в об'єкті.

У найпростішому випадку однофакторної регресії математична модель – це формула виду $y = F(x)$. Якщо модель лінійна, то $y = b_0 + b_1x$ – рівняння лінійної регресії.

Експериментальну вибірку прийнято характеризувати параметрами, які вказують на центр групування (положення) елементів вибірки (результатів експерименту) на числовій осі, а також ступінь їхнього розсіювання навколо цього центру. Зокрема, положення вибірки на числовій осі характеризують середніми значеннями: середнім арифметичним, середнім геометричним, середнім гармонічним значеннями вибірки, а також іншими параметрами, такими як центр розмаху, медіаною, модою тощо. Розпорошення вибірки характеризують так званим розмахом, вибірковою дисперсією.

Загалом у математичній статистиці такого типу параметри називаються точковими, оскільки вони виражаються одним числом, вказують на певну точку в експериментальному розподілі та належать до широкої гами вибірових характеристик.

Середнє (арифметичне) значення фактора x (відклику y) є однією з характеристик центру, навколо якої групуються можливі значення випадкової величини, і обчислюється за формулами:

$$\bar{x}_{(n)} = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (1.2)$$

$$\bar{y}_{(n)} = \bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (1.3)$$

Графік лінійної регресії завжди проходить через центр розсіювання.

Вибіркова дисперсія. Мірою розкиду, розсіювання отриманої у результаті експерименту вибірки відносно середнього значення є **вибіркова дисперсія** – середнє значення квадратів відхилень елементів вибірки від їхнього середнього значення.

$$S_{(n)}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 . \quad (1.4)$$

Вибіркове стандартне (середньоквадратичне) відхилення. Квадратний корінь з вибіркової дисперсії називається **вибірковим стандартним відхиленням** результатів експерименту яке є випадковою величиною, тобто для різних вибірок однакового чи різного обсягу (розміру) має дещо різні значення, які проте, із збільшенням обсягу вибірки стабілізуються.

$$S_{(n)} = S = \sqrt{S_{(n)}^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} . \quad (1.5)$$

Область моделювання (прогнозів) розташована між мінімальним і максимальним значеннями фактору x . Прогноз відклику у робиться за рівнянням моделі.

Завдання до лабораторної роботи

1. Створити таблицю даних.
2. Створити автозвіт.
3. Внести в автозвіт створену таблицю даних.
4. Знайти середнє значення, середньоквадратичне відхилення, мінімальне та максимальне значення для фактора x_1 .
5. Отримати математичну модель досліджуваного процесу (згідно варіанту індивідуального завдання) та побудувати на її основі графік.
6. Здійснити моделювання критерію відгуку у для точок: x_{cp} та в будь-якій довільній точці з області прогнозів.

Приклад виконання лабораторної роботи

Розпочинаємо роботу із створення нової електронної

таблиці (spreadsheet). Для цього активуємо команду "Створити новий документ" (Create a new docement). В діалоговому вікні (рис. 1.2) вказуємо дві змінні та десять значень для кожної змінної. Натискаємо кнопку "Ok".

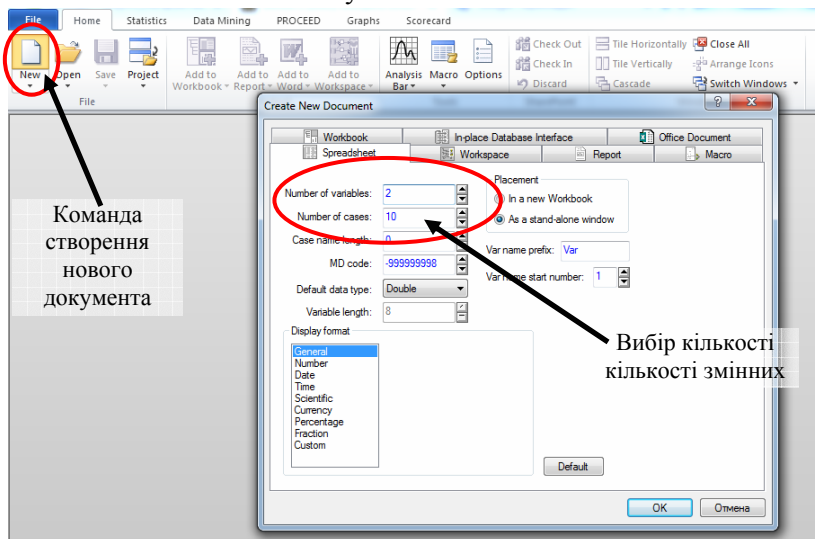


Рис. 1.2. Створення електронної таблиці

Розпочинаємо роботу із створення нової електронної таблиці (spreadsheet). Для цього активуємо команду "Створити новий документ" (Create a new docement). В діалоговому вікні (рис. 1.2) вказуємо дві змінні та десять значень для кожної змінної. Натискаємо кнопку "Ok".

Вносимо в таблицю свій варіант індивідуального завдання (дод. А) (рис. 1.3). Перейшовши на вкладку "Додому" (Home) панелі інструментів активуємо команду "Додати до звіту" (Add to Report) (рис. 1.4). В результаті виконання команди таблиця даних буде внесено в автозвіт (рис. 1.5).

Наступний етап – розрахунок середнього значення фактора (x_{cp}), середньоквадратичного відхилення та області прогнозів для фактора (x_1).

	1	2
	x_1	y
1	48	0,167
2	40	0,092
3	48	0,233
4	40	0,183
5	44	0,1
6	44	0,183
7	48	0,333
8	44	0,269
9	40	0,248
10	48	0,167

Рис. 1.3. Таблиця даних

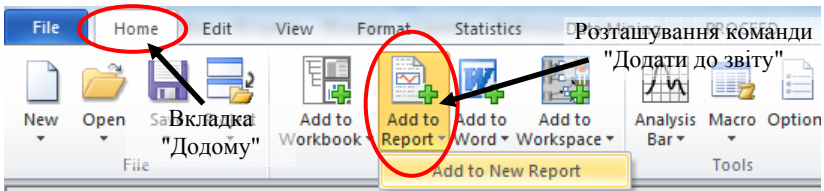


Рис. 1.4. Створення автозвіту

Report1*			
Contents			
Spreadsheet2			
	1	2	
	x_1	y	
1	48	0,167	
2	40	0,092	
3	48	0,233	
4	40	0,183	
5	44	0,1	
6	44	0,183	
7	48	0,333	
8	44	0,269	
9	40	0,248	
10	48	0,167	

Рис. 1.5. Результат створення автозвіту

Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів активуємо модуль "Основні статистичні дані/таблиці" (Basic Statistics/Tables). В діалоговому вікні, що з'явилося (рис. 1.6) обираємо команду "Описові статистики" (Descriptive statistics) та натискаємо "Ok".

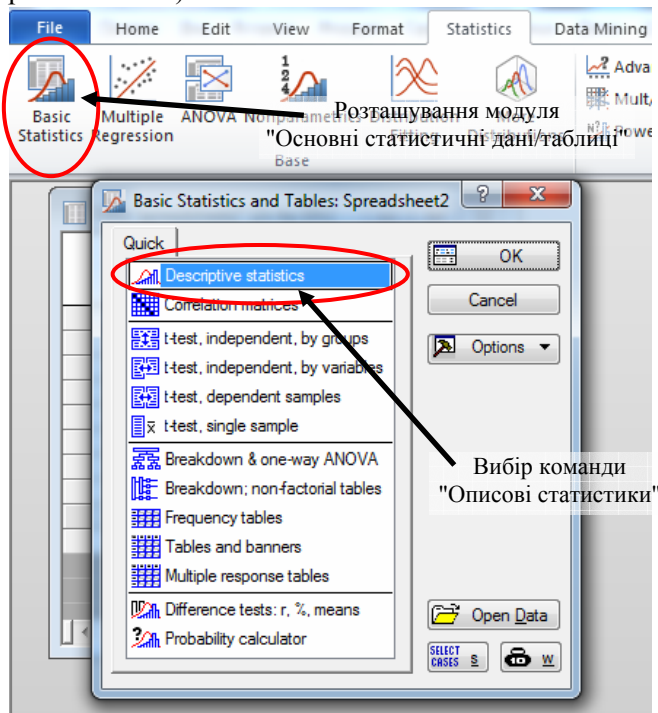


Рис. 1.6. Діалогове вікно модуля "Основні статистичні дані/таблиці"

В діалоговому вікні команди "Описові статистики" (Descriptive statistics) у вікні "змінні" (variables) вказуємо змінну " x_1 ". Переходимо на вкладку "Розширені налаштування" (Advanced) де обираємо необхідні параметри (рис. 1.7):

- середнє значення (mean);
- середньоквадратичне відхилення (standard deviation);
- мінімальне та максимальне значення (minimum &

maximum).

Після чого натискаємо кнопку "Швидкий аналіз" (Summary).

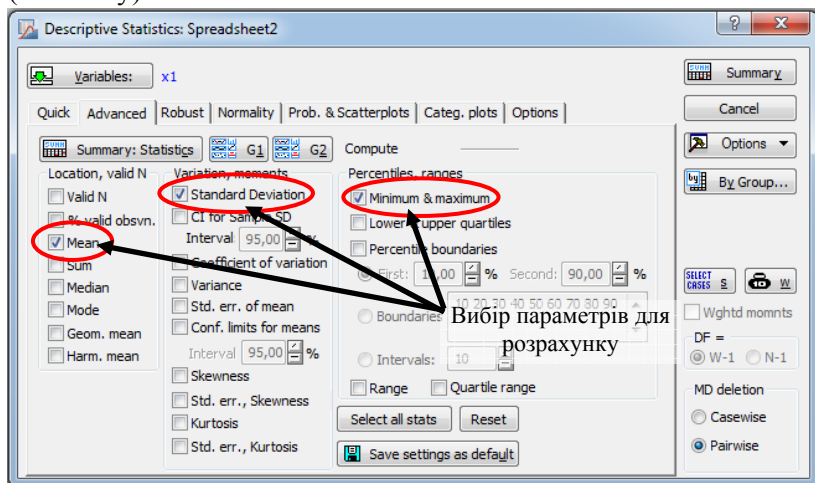


Рис. 1.7. Діалогове вікно команди "Описові статистики"

Variable	Descriptive Statistics (Spreadsheet2)			
	Mean	Minimum	Maximum	Std.Dev.
x ₁	44,40000	40,00000	48,00000	3,502380

Рис. 1.8. Результати аналізу

Отримана таблиця (рис. 1.8) додається до автозвіту.

Програмний продукт Statistica володіє потужним інструментарієм для побудови графічних залежностей. Перейшовши на вкладку "Graphs" (Графіки) панелі інструментів активуємо команду "Діаграми розсіювання" (Scatterplot).

В діалоговому вікні команди (рис. 1.9) у вікні "змінні" (variables) вказуємо незалежну змінну (фактор) та залежну змінну (критерій відгуку). Перейшовши на вкладку "Розширені налаштування" (Advanced) обираємо тип моделі: "лінійна" (linear) типу $y = b_0 + b_1x$. Натискаємо "Ok". На екрані з'явиться графік (рис. 1.10).

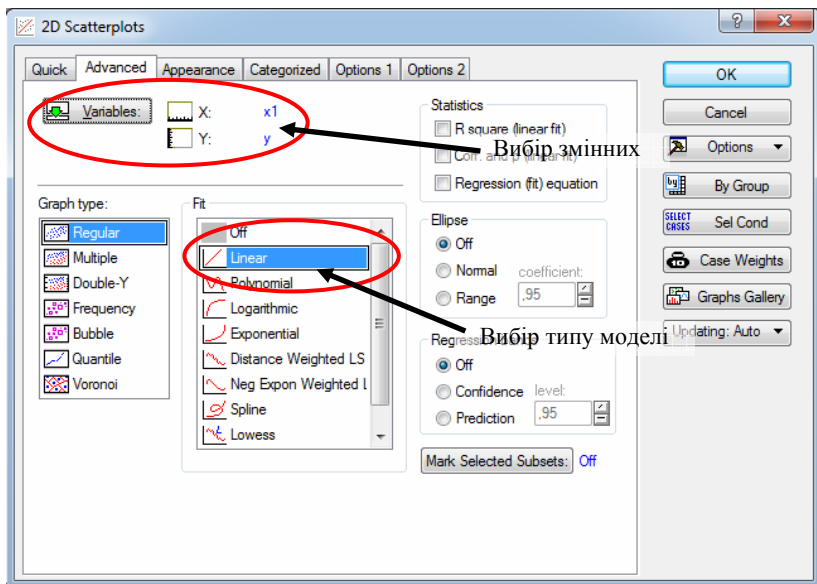


Рис. 1.9. Диалогове вікно команди "Діаграми розсіювання"

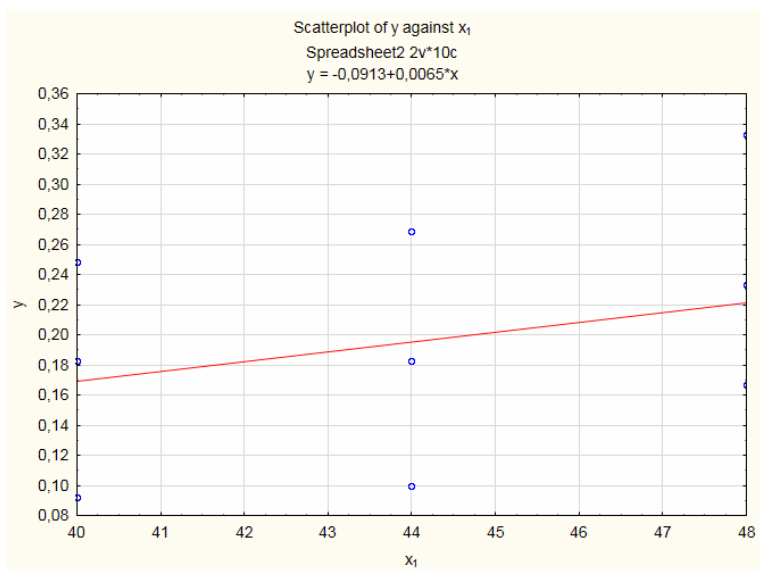


Рис. 1.10. Результат побудови графіку лінійної регресії

Моделювання критерію відгуку у проведемо для вказаних в індивідуальному завданні рівнів фактора (x_1) та двох доданих точок: $x_{cp} = 44,4$ та $x_i = 42$.

Для цього в таблиці даних (рис. 1.3) додамо два значення до змінної x_1 та нову змінну "y_reg" (рис. 1.11).

	1	2	3
	x_1	y	y_reg
1	48	0,167	
2	40	0,092	
3	48	0,233	
4	40	0,183	
5	44	0,1	
6	44	0,183	
7	48	0,333	
8	44	0,269	
9	40	0,248	
10	48	0,167	
11	42		
12	44,4		

Рис. 1.11. Загальний вигляд таблиці даних після додавання нових значень

Двічі натиснувши лівою кнопкою мишки на імені змінної " y_reg " – на екрані з'явиться діалогове вікно редагування змінної (рис. 1.12).

У полі "Довга назва" (Long name) вписуємо рівняння математичної моделі (лінійне рівняння регресії): $y = -0,0913 + 0,0065 \cdot x_1$. Варто пам'ятати імена змінних у рівнянні моделі та таблиці даних мають бути ідентичні. Після введення рівняння моделі натискаємо кнопку "Ok".

В таблиці даних в полі значень змінної "y_reg" з'являться змодельовані значення за допомогою отриманої регресійної (математичної) моделі (рис. 1.13).

Варто пам'ятати, що на усіх етапах виконання лабораторної роботи отримані результати необхідно додавати до автозвіту.

Variable 3

Name: Type: OK

Measurement Type: Length: Cancel

☐ Excluded ☐ Label ☐ Case State MD code: << >>

Display format

General
Number
Date
Time
Scientific
Currency
Percentage
Fraction
Custom

Field for entering the model equation

Long name (label or formula): Function guide

Labels: Use any text. Formulas: Must begin with an = sign.
Use variable names or V1, V2, ..., V0 is the case number.
Examples: (a) = mean(v1:v3, sqrt(v7), AGE) (b) = v1+v2: comment (after).
In case of conflict, variable names take precedence over variable text values.
Specify text values by appending \$, as in "value\$".

Рис. 1.12. Загальний вигляд вікна редагування змінної

	1 x ₁	2 y	3 y _{reg}
1	48	0,167	0,2207
2	40	0,092	0,1687
3	48	0,233	0,2207
4	40	0,183	0,1687
5	44	0,1	0,1947
6	44	0,183	0,1947
7	48	0,333	0,2207
8	44	0,269	0,1947
9	40	0,248	0,1687
10	48	0,167	0,2207
11	42		0,1817
12	44,4		0,1973

Рис. 1.13. Результат моделювання

Звіт про виконану лабораторну роботу має містити тему та завдання, отримані в процесі виконання таблиці та рисунки, а також висновок.

Зразок висновку

В результаті виконання лабораторної роботи були отримано статистичні характеристики для фактора x_1 : діапазон зміни значень фактора $40,0 \leq x \leq 48,0$, середнє значення $x_{\text{ср}}=44,4$ (визначає центр області прогнозів), а також середньоквадратичне відхилення $3,5023$ (характеризує середнє значення розсіювання значень).

Отримано математичну модель у вигляді рівняння лінійної регресії $y = -0,0913 + 0,0065 \cdot x_1$. Отримана модель дозволила провести моделювання залежності критерію відгуку (тривалості відкачування повітря) від фактору (величини тиску) x_1 . Так, для значення тиску $x_1 = 44,4$ (кПа) тривалість відкачування повітря становитиме $y = 0,1973$ (с).

ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ

1. Дайте визначення терміну "генеральна сукупність"?
2. Дайте визначення терміну "вибіркова дисперсія"?
3. Дайте визначення терміну "кореляційне поле"?
4. Як визначити середнє значення величини із використанням програмного продукту Statistica?
5. Що таке математична модель?
6. Що таке вибірка дисперсія отриманих експериментальних даних?
7. Як отримати значення прогнозу в довільній точці?
8. Що таке рівняння регресії?
9. Як отримати рівняння прямої регресії?
10. Як створити автозвіт?

ЛАБОРАТОРНА РОБОТА №2

МЕТОДИКА ВИБОРУ ОПТИМАЛЬНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ

Мета роботи: Отримати практичні навички створення та редагування графіків і рівнянь лінійної регресії та вибору оптимальної регресійної моделі.

Обладнання робочого місця: персональний комп'ютер з необхідним програмним забезпеченням, методичні рекомендації, довідкова література.

Загальні відомості

Найбільш поширеним завданням опрацювання експериментальних даних є представлення їх сукупності деякою функцією $y(x)$.

Проста лінійна регресія. Просту лінійну регресію використовують для встановлення лінійного зв'язку між двома змінними: незалежною змінною x і залежною змінною y . При цьому ознаку x називають фактором, а y – критерієм відгуку. При регресійному аналізі визначаються два параметри: b_0 – перетин і b_1 – нахил регресії. Перетин показує значення y при умові, що x дорівнює нулю, а значення нахилу b_1 показує на скільки зміниться значення критерію відгуку y при умові, що значення фактора x збільшиться на 1.

На першому етапі побудови регресійної моделі для визначення форми зв'язку між фактором x та критерієм відгуку y будують діаграму розподілу (будують точки з координатами (x, y)).

Якщо можна легко помітити, що точки на діаграмі розподілу явно концентруються (за деякими виключеннями) поблизу прямої (прямої регресії), у такому випадку говорять про лінійний зв'язок:

$$y = a + b \cdot x. \quad (2.1)$$

де b – регресійний коефіцієнт;

a – зсув по осі ординат.

Для оцінки параметрів лінійної регресії застосовуються різноманітні методи. Одним з таких є використання графіка початкових даних (рис. 2.1). Суть даного методу полягає в наступному: через дві точки проводиться пряма лінія і за графіком визначаються параметри регресії – a і b .

Зсув за віссю ординат відповідає точці на вісі y (вертикальної осі), де пряма регресії перетинає цю вісь. Коефіцієнт регресії b через співвідношення $b = \operatorname{tg}(\alpha)$ вказує на кут нахилу α прямої.

При проведенні простої лінійної регресії основним завданням є визначення параметрів b і a . Після визначення цих параметрів, наприклад, можна спрогнозувати показник y .

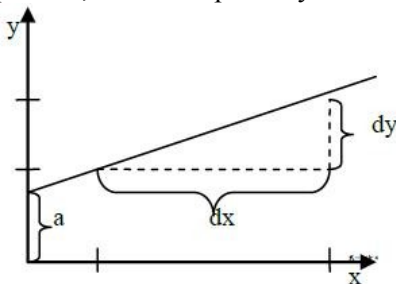


Рис. 2.1. Метод знаходження параметрів лінійної регресії за графіком первинних даних

Основа класичного методу оцінки параметрів лінійної регресії ґрунтується на використанні **методу найменших квадратів (МНК)**.

Метод найменших квадратів (МНК) уможливорює отримання оцінки параметрів a і b , при яких сума квадратів відхилень фактичних значень критерію відгуку y від

розрахункових (теоретичних) y_x мінімальна:

$$\sum_i (y_i - y_{x_i})^2 \rightarrow \min. \quad (2.2)$$

Тобто, лінія регресії обирається таким чином, щоб сума квадратів відстані по вертикалі між точками і цією лінією на графіку була б мінімальною (рис. 2.2):

$$\begin{aligned} \varepsilon_i &= y_i - y_x, \\ \sum_i \varepsilon_i^2 &\rightarrow \min. \end{aligned} \quad (2.3)$$

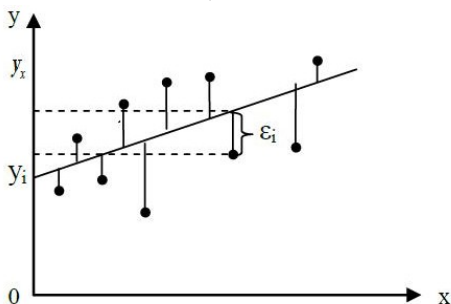


Рис. 2.2. Відхилення фактичних значень критерію відгуку від теоретичних

Експонентна регресія розраховує експонентну криву, яка описує множину даних, для яких не виконується умова лінійної зміни:

$$y = a \cdot e^{b \cdot x}. \quad (2.4)$$

де a – коефіцієнт при експоненті;
 b – регресійний коефіцієнт.

Завдання до лабораторної роботи

Використовуючи дані з лабораторної роботи № 1, за мінімумом суми квадратів залишків (параметр дисперсії) вибрати оптимальну модель із двох: лінійної ($y=b_0+b_1 \cdot x$) та експонентної ($y=b_0 \cdot \exp(b_1 \cdot x)$).

Для цього необхідно:

1. Отримати математичну модель (лінійну) та побудувати на її основі графік.
2. Здійснити моделювання критерію відгуку (y) для кожної

точки фактора (x) за лінійною моделлю.

3. Отримати математичну модель (експонентною) та побудувати на її основі графік.

4. Здійснити моделювання критерію відгуку (y) для кожної точки фактора (x) за експонентною моделлю.

5. Розрахувати похибку результатів моделювання за лінійною та експонентною моделями.

6. Розрахувати дисперсію для експонентної моделі.

7. Розрахувати дисперсію для лінійної моделі.

8. Обґрунтувати вибір оптимальної моделі.

Приклад виконання лабораторної роботи

Створення нової електронної таблиці (spreadsheet) детально описано у попередній лабораторній роботі, тому перейдемо до побудови графіка. У вкладці "Graphs" (Графіки) панелі інструментів активуємо команду "Діаграми розсіювання" (Scatterplot). В діалоговому вікні, що з'явилося на екрані (рис. 1.10) вказуємо "змінні" (variables) та обираємо тип моделі: "лінійна" (linear) типу $y = b_0 + b_1x$. Натискаємо "Ok".

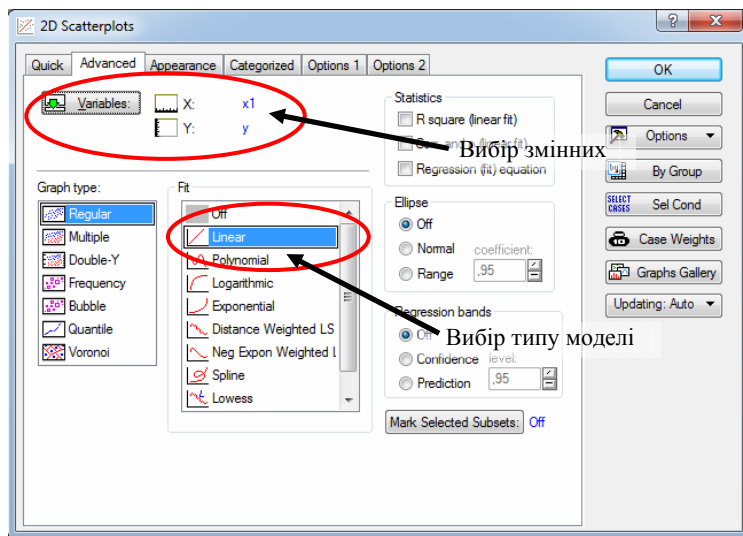


Рис. 2.3. Діалогове вікно команди "Діаграми розсіювання"

В результаті виконання команди на екрані з'явився графік та рівняння лінійної регресійної моделі (рис. 2.4).

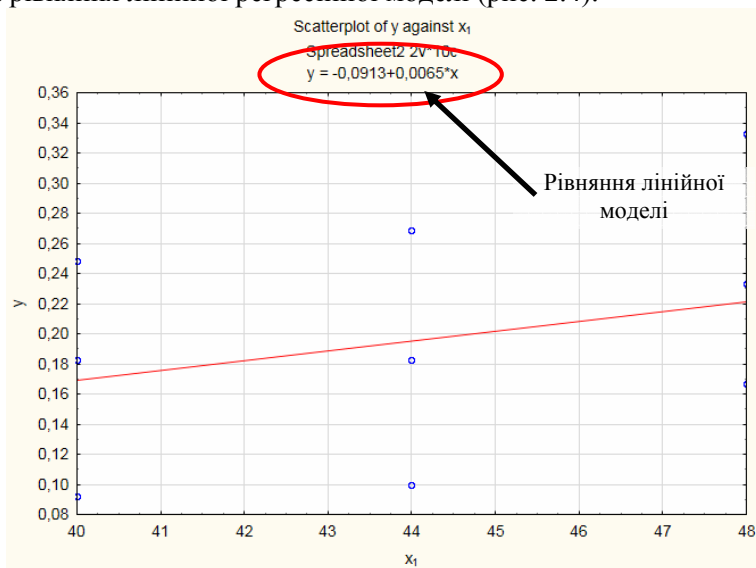


Рис. 2.4. Результат побудови графіку лінійної регресії

Наступний етап – моделювання критерію відгуку (y) для кожної точки фактора (x) за отриманим рівнянням (рис. 2.4). В таблиці даних створюємо нову змінну "y_lin". В діалоговому вікні редагування змінної (рис. 2.5) в полі "Довга назва" (Long name) вписуємо рівняння математичної моделі (лінійне рівняння регресії): $y = -0,0913 + 0,0065 \cdot x_1$.

Після введення рівняння моделі натискаємо кнопку "Ok".

В електронній таблиці в полі значень змінної "y_reg" з'являться змодельовані значення за допомогою отриманої регресійної (математичної) моделі (рис. 2.6).

Процес створення графіку експонентної моделі аналогічний. В діалоговому вікні команди "Graphs" (Графіки) вказуємо "змінні" (variables) та обираємо тип моделі (рис. 2.7): "експонентна" (exponential) типу $y = b_0 \cdot \exp(b_1 \cdot x)$. Натискаємо "Ok".

На екрані з'явився графік та рівняння (рис. 2.8).

Variable 3

Name: Type: OK

Measurement Type: Length: Cancel

☐ Excluded ☐ Label ☐ Case State MD code: << >>

Display format

☒ General
☐ Number
☐ Date
☐ Time
☐ Scientific
☐ Currency
☐ Percentage
☐ Fraction
☐ Custom

Long name (label or formula): Functions: ☒ Function guide

Labels: Use any text. Formulas: Must begin with an = sign.
 Use variable names or V1, V2, ..., V0 is the case number.
 Examples: (a) = mean(v1:v3, sort(v7), AGE) (b) = v1+v2; comment (after);
 In case of conflict, variable names take precedence over variable text values.
 Specify text values by appending \$, as in "value\$".

Рис. 2.5. Загальний вигляд вікна редагування змінної

	1	2	3
	x_1	y	y_{lin}
1	48	0,167	0,2207
2	40	0,092	0,1687
3	48	0,243	0,2207
4	40	0,183	0,1687
5	44	0,1	0,1947
6	44	0,183	0,1947
7	48	0,333	0,2207
8	44	0,263	0,1947
9	40	0,248	0,1687
10	48	0,167	0,2207

Рис. 2.6. Результат моделювання

Наступний етап – моделювання. За аналогією із лінійною моделлю створюємо нову змінну "y_log" в електронній таблиці. В діалоговому вікні редагування змінної (рис. 2.9) в полі "Довга назва" (Long name) вписуємо рівняння математичної моделі (експонентне рівняння регресії): $y = 0,0348 \cdot \exp(0,0375 \cdot x)$.

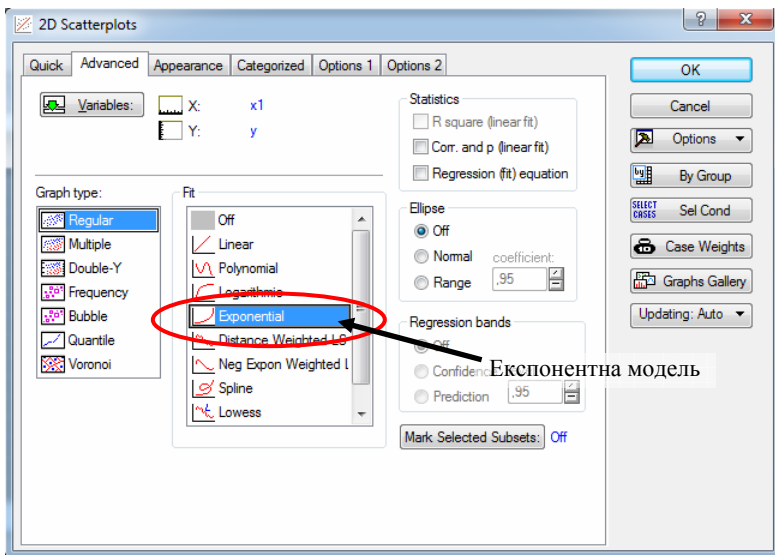


Рис. 2.7. Діалогове вікно команди "Діаграми розсіювання"

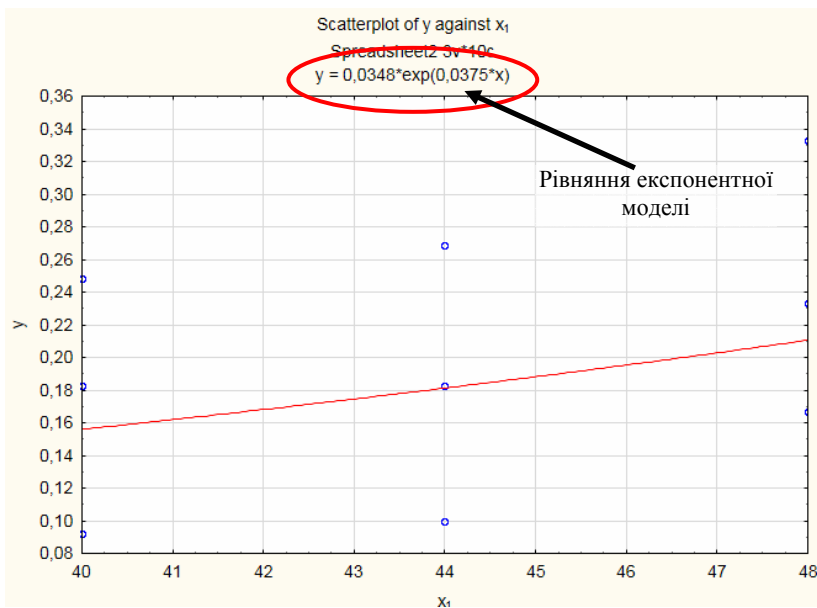


Рис. 2.8. Результат побудови графіку експонентної регресії

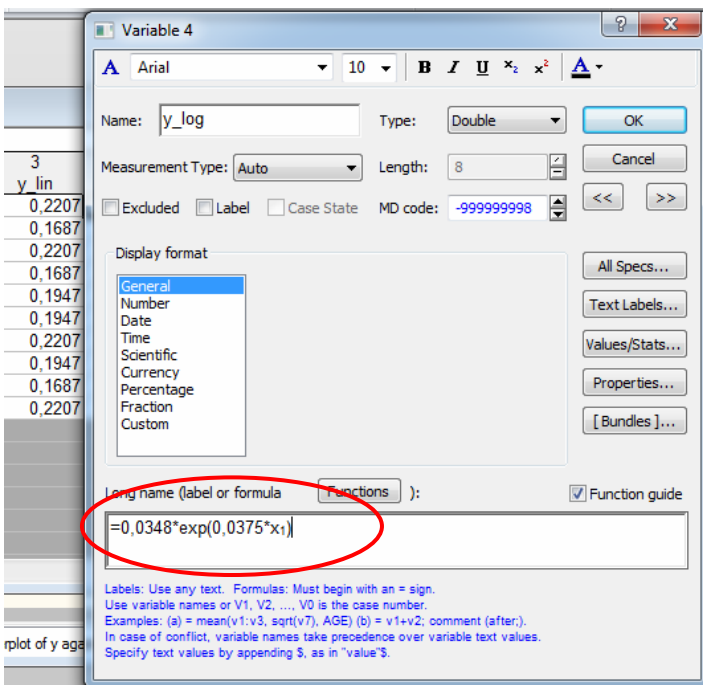


Рис. 2.9. Загальний вигляд вікна редагування змінної

В електронній таблиці (рис. 2.10) в полі значень змінної "y_log" з'являться змодельовані значення.

	1 x ₁	2 y	3 y _{lin}	4 y _{log}
1	48	0,167	0,2207	0,21052773
2	40	0,092	0,1687	0,15596278
3	48	0,233	0,2207	0,21052773
4	40	0,183	0,1687	0,15596278
5	44	0,1	0,1947	0,1812029
6	44	0,183	0,1947	0,1812029
7	48	0,333	0,2207	0,21052773
8	44	0,269	0,1947	0,1812029
9	40	0,248	0,1687	0,15596278
10	48	0,167	0,2207	0,21052773

Рис. 2.10. Результат моделювання

Розрахунок похибки результатів моделювання лінійної та експонентної моделі полягає у визначенні відсоткової розбіжності результатів. Для цього в електронній таблиці створюємо нову змінну – "RASH".

В діалоговому вікні створеної змінної в полі "Довга назва" (Long name) записуємо формулу: $(y_{lin}-y_{log})/y_{lin}*100$ (рис. 2.11). Електронна таблиця набуде вигляду (рис. 2.12).

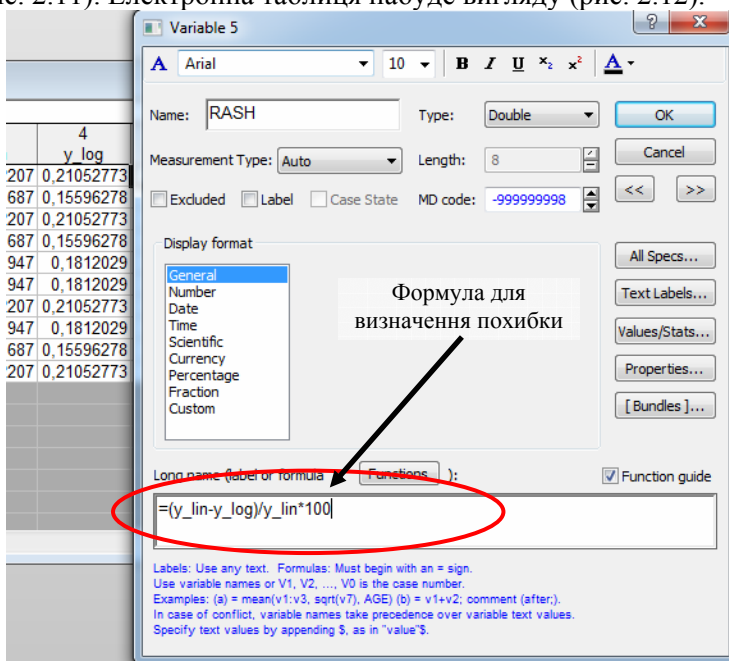


Рис. 2.11. Загальний вигляд вікна редагування змінної

Переходимо до розрахунку дисперсій двох моделей. Розглянемо два методи: перший за допомогою формул (експонентна модель) та із використанням модуля множинної регресії (лінійна модель).

Розпочнемо із експонентної моделі. В електронній таблиці додаємо нову змінну – "KV_ZAL". В діалоговому вікні створеної змінної (поле "Довга назва" (Long name)) записуємо формулу: $(y-y_{log})^2$ для порядкової дисперсії (рис. 2.13).

	1	2	3	4	5
	x_1	y	y_{lin}	y_{log}	RASH
1	48	0,167	0,2207	0,21052773	4,60909299
2	40	0,092	0,1687	0,15596278	7,55021953
3	48	0,233	0,2207	0,21052773	4,60909299
4	40	0,183	0,1687	0,15596278	7,55021953
5	44	0,1	0,1947	0,1812029	6,93225579
6	44	0,183	0,1947	0,1812029	6,93225579
7	48	0,333	0,2207	0,21052773	4,60909299
8	44	0,269	0,1947	0,1812029	6,93225579
9	40	0,248	0,1687	0,15596278	7,55021953
10	48	0,167	0,2207	0,21052773	4,60909299

Рис. 2.12. Результат розрахунку похибки

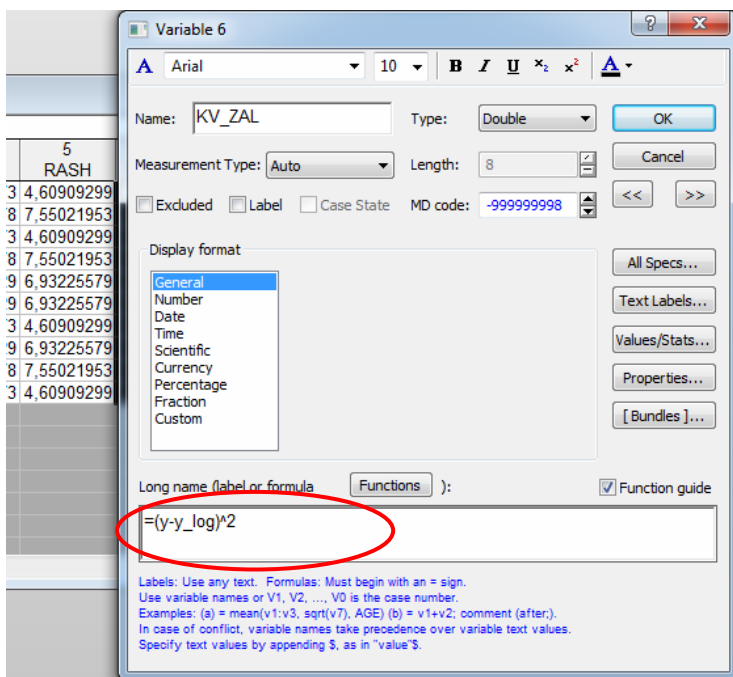


Рис. 2.13. Загальний вигляд вікна редагування змінної

Електронна таблиця набуде наступного вигляду (рис. 2.14).

	1	2	3	4	5	6
	x_1	y	y_{lin}	y_{log}	RASH	KV_ZAL
1	48	0,167	0,2207	0,21052773	4,60909299	0,00189466
2	40	0,092	0,1687	0,15596278	7,55021953	0,00409124
3	48	0,233	0,2207	0,21052773	4,60909299	0,000505
4	40	0,183	0,1687	0,15596278	7,55021953	0,00073101
5	44	0,1	0,1947	0,1812029	6,93225579	0,00659391
6	44	0,183	0,1947	0,1812029	6,93225579	3,22958E-6
7	48	0,333	0,2207	0,21052773	4,60909299	0,01499946
8	44	0,269	0,1947	0,1812029	6,93225579	0,00770833
9	40	0,248	0,1687	0,15596278	7,55021953	0,00847085
10	48	0,167	0,2207	0,21052773	4,60909299	0,00189466

Рис. 2.13. Результат розрахунку порядкової дисперсії

Дисперсія для експонентної моделі визначається, як сума порядкових дисперсій (кожного експерименту). В першій лабораторній роботі розглядався приклад використання модуля "Основні статистичні дані/таблиці" (Basic Statistics/Tables) для визначення параметрів змінних, зокрема "суми" (sum).

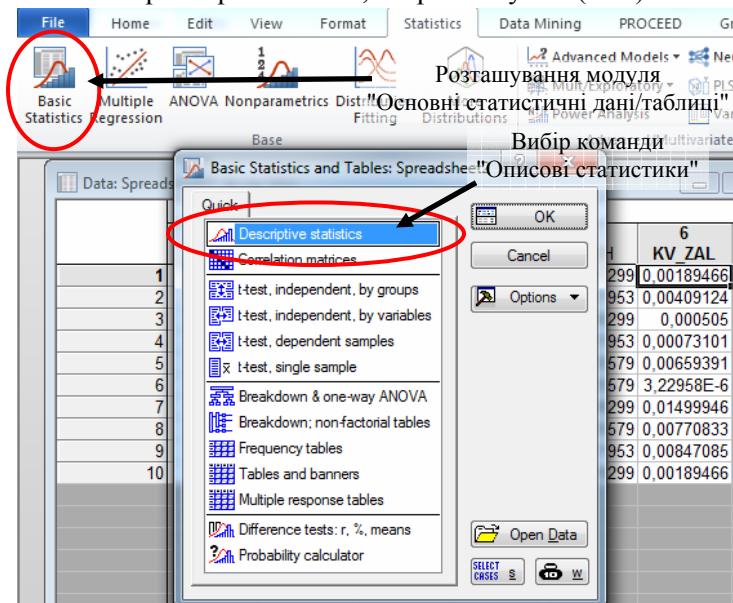


Рис. 2.14. Діалогове вікно модуля "Основні статистичні дані/таблиці"

Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів активуємо вказаний модуль. В діалоговому вікні, що з'явилось (рис. 2.14) обираємо команду "Описові статистики" (Descriptive statistics) та натискаємо "Ok".

В діалоговому вікні команди "Описові статистики" (Descriptive statistics) у вікні "змінні" (variables) вказуємо змінну "KV_ZAL". Переходимо на вкладку "Розширені налаштування" (Advanced) де обираємо необхідний параметр (рис. 2.15):

– сума (sum).

Після чого натискаємо кнопку "Швидкий аналіз" (Summary).

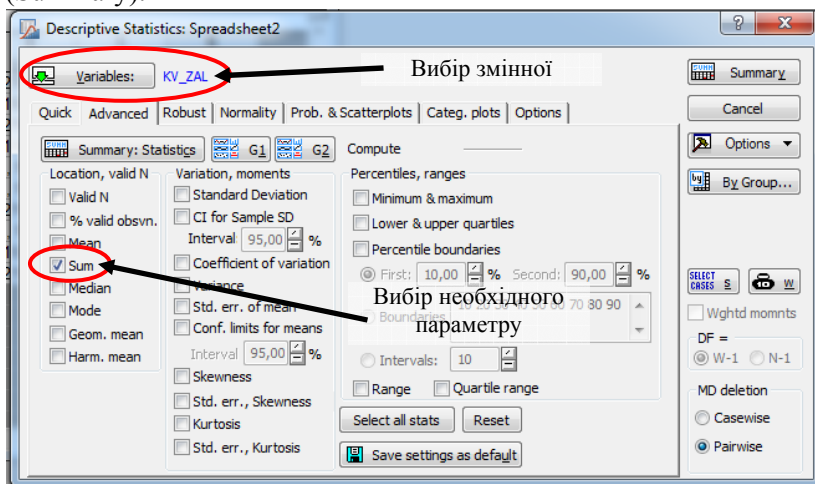


Рис. 2.15. Діалогове вікно команди "Описові статистики"

Отримана таблиця (рис. 2.16) містить параметр дисперсії для експонентної моделі.

Descriptive Statistics (Spreadsheet2)	
Variable	Sum
KV_ZAL	0.046892

Рис. 2.16. Дисперсія для експонентної моделі

Дисперсію для лінійної моделі визначимо із

використанням модуля "Множинна регресія" (Multiple regression).

Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів активуємо модуль "Множинна регресія" (Multiple regression). В діалоговому вікні, що з'явилося (рис. 2.17) у полі "змінні" (variables) вказуємо:

- залежна змінна (dependent): y;
- незалежна змінна (independent): y_lin.

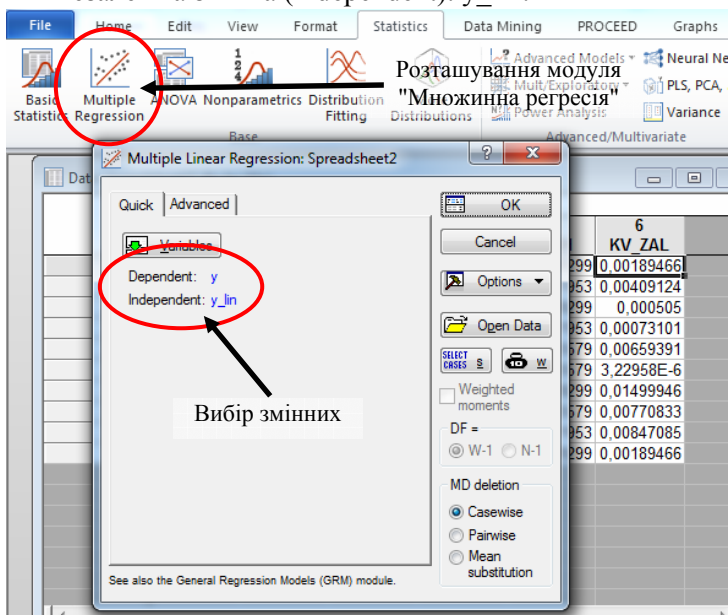


Рис. 2.17. Діалогове вікно модуля "Множинна регресія"

Після вибору змінних натискаємо "Ok", на екрані з'явиться наступне діалогове вікно: "Результати множинної регресії" (Multiple Regression Results) (рис. 2.18). Переходимо на вкладку "Розширені налаштування" (Advanced) де обираємо команду "Дисперсія (загальна підгонка)" (ANOVA (Overall goodness of fit)).

На екрані з'явиться таблиця (рис. 2.19) із розрахованою

дисперсією для лінійної моделі.

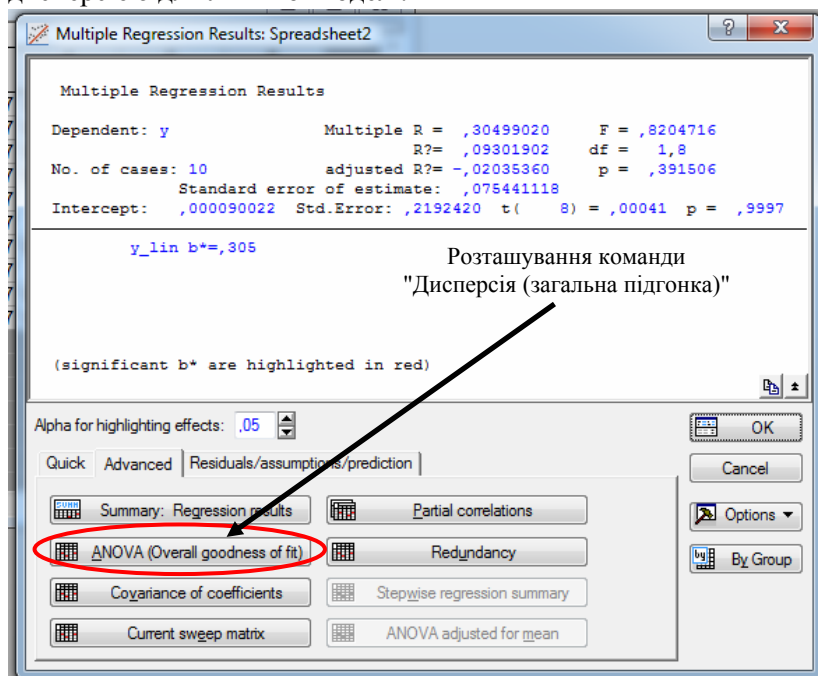


Рис. 2.18. Діалогове вікно "Результати множинної регресії"

Analysis of Variance; DV: y (Spreadsheet2)					
Effect	Sums of Squares	df	Mean Squares	F	p-value
Regress.	0,004670	1	0,004670	0,820472	0,391506
Residual	0,045531	8	0,005691		
Total	0,050201				

Рис. 2.19. Результати розрахунку дисперсії для лінійної моделі

Оскільки отриманий параметр дисперсії для лінійної моделі (0,04553) менший ніж для експонентної (0,04689) – обираємо лінійну. Дисперсія дозволяє обрати оптимальнішу модель для опису досліджуваного процесу – в якій менший квадрат залишків відносно експериментальних результатів.

Варто пам'ятати, що на усіх етапах виконання

лабораторної роботи отримані результати необхідно додавати до автозвіту.

Звіт про виконану лабораторну роботу має містити тему та завдання, отримані в процесі виконання таблиці та рисунки, а також висновок.

Зразок висновку

В результаті виконання лабораторної роботи були отримані дві математичні моделі: лінійну $y = -0,0913 + 0,0065 \cdot x_1$ та експонентну $y = 0,0348 \cdot \exp(0,0375 \cdot x)$.

За результатами розрахунків параметру дисперсії (квадрату залишків відносно експериментальних результатів) можна стверджувати, що лінійна модель (0,04553) є точнішою за експонентну модель (0,04689) для опису досліджуваного процесу.

За неправильного вибору математичної моделі (експонентної замість лінійної) максимальна похибка результатів моделювання становитиме 7,55%.

ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ

1. Яка область застосування регресійних моделей?
2. Які різновиди регресійних моделей вам відомі?
3. Як отримати рівняння експонентної регресії?
4. Як отримати рівняння лінійної регресії?
5. Як отримати прогноз на основі регресійної моделі?
6. Дайте визначення терміну "дисперсія"?
7. Дайте визначення терміну "порядкова дисперсія"?
8. Які вам відомі методи визначення параметру дисперсії?
9. На основі чого приймається рішення про вибір типу регресійної моделі?

ЛАБОРАТОРНА РОБОТА №3

ВИЗНАЧЕННЯ АДЕКВАТНОСТІ РЕГРЕСІЙНОЇ МОДЕЛІ

Мета роботи: Отримати практичні навички перевірки значимості коефіцієнтів рівняння регресії та визначення адекватності регресійної моделі.

Обладнання робочого місця: персональний комп'ютер з необхідним програмним забезпеченням, методичні рекомендації, довідкова література.

Загальні відомості

Метою будь-якого технічного дослідження є одержання математичної залежності між факторами і критеріями відгуку. За математичну модель можуть бути прийняті функції:

$$y = f(d_{\text{пер}}, P_i, V). \quad (3.1)$$

Така кореляційна залежність може бути представлена формою полінома у вигляді:

$$y = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i < j} b_{ij} x_i x_j + \sum_{i=1}^k b_{ii} x_i^2, \quad (3.2)$$

де y – функція відгуку;

b_0, b_i, b_{ij}, b_{ii} – коефіцієнти кореляційної залежності;

x_1, x_2 – кодовані значення факторів;

i, j – номери факторів.

Для апроксимації процесу, що розглядається, використовується лінійна частина кореляційної залежності (3.2):

$$y = b_0 + b_1 x_1 + b_2 x_2, \quad (3.3)$$

при її неадекватності – квадратична:

$$y = b_0 + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2. \quad (3.4)$$

Величини b_0 , b_i , b_{ij} , b_{ii} називаються **коефіцієнтами регресії** і визначаються за відомою методикою [15-17] за формулами:

$$b_0 = k_1(n) \sum_{g=1}^N \bar{y}_g - k_2(n) \sum_{i=1}^n \sum_{g=1}^N x_{ig}^2 \bar{y}_g, \quad (3.5)$$

$$b_i = k_3(n) \sum_{g=1}^N x_{ig} \bar{y}_g, \quad (3.6)$$

$$b_{ij} = k_4(n) \sum_{g=1}^N x_{ig} x_{jg} \bar{y}_g, \quad (3.7)$$

$$b_{ii} = k_5(n) \sum_{g=1}^N x_{ig}^2 \bar{y}_g + k_6(n) \sum_{i=1}^n \sum_{g=1}^N x_{ig}^2 \bar{y}_g - k_7(n) \sum_{g=1}^N \bar{y}_g, \quad (3.8)$$

де $k_i(n)$ – табличні значення коефіцієнтів, які стосуються конкретного плану факторного експерименту та області планування;

N – кількість дослідів.

Після отримання регресійного рівняння необхідно провести статистичний аналіз, тобто перевірити адекватність подання результатів дослідів поліномом, значимість коефіцієнтів регресії і взаємозв'язок коефіцієнтів регресії.

Статистична значимість коефіцієнтів регресії перевіряється за допомогою **критерію Стюдента** (t – критерій).

Для оцінки значимості коефіцієнтів спочатку визначають розрахункове значення t – критерію за формулою:

$$t_{ip} = \frac{|b_i|}{S_{b_i}}, \quad (3.9)$$

де S_{b_i} – середньоквадратична похибка коефіцієнта регресії.

Значимість коефіцієнтів регресії перевіряють за умовою:

$$t_{ip} > t_m, \quad (3.10)$$

де t_m – табличне значення критерію, знайдене для вибраної величини значимості (0,95) й ступеня свободи $fn=N \cdot (m-1)$;

N – кількість дослідів (рядків в матриці плану);

m – кількість повторностей одного дослідів.

Дисперсію коефіцієнтів регресії визначають за формулою:

$$S_{b_i}^2 = \frac{S_e^2}{N}, \quad (3.11)$$

де S_e^2 – головна дисперсія, що залежить лише від похибки дослідів та кількості дослідів.

Головну дисперсію визначають за формулою:

$$S_e^2 = \frac{\sum_{i=1}^N S_n^2}{N}. \quad (3.12)$$

Якщо виявиться, що який-небудь коефіцієнт регресії статистично незначний, то його зовсім не обов'язково видаляти із моделі, оскільки модель після цього може виявитись не адекватною.

Придатність рівняння регресії для опису реальної залежності критерію оптимізації від факторів визначається відомим методом [1-17]. Відповідно до нього адекватність кореляційної моделі визначається за допомогою **критерію Фішера** (F -критерію). Для цього порівнюється дві дисперсії – одна характеризує розсіювання середніх дослідних даних \bar{y}_{ci} відносно значень другої – змінної величини \hat{y}_{pi} , що розраховані за допомогою знайденого рівняння регресії (передбачувані значення). Цю дисперсію називають **дисперсією адекватності** й визначають з виразу [1-17]:

$$S_{ad}^2 = \frac{\sum_{i=1}^N (\hat{y}_{pi} - \bar{y}_{ci})^2}{N - d}, \quad (3.13)$$

де \bar{y}_{ci} – середнє значення змінної, отримане дослідним шляхом;

\hat{y}_{pi} – середнє значення змінної, отримане розрахунковим шляхом;

d – кількість значимих коефіцієнтів у рівнянні регресії.

Адекватність моделі оцінюється за умовами:

$$F_p \leq F_T - \text{модель адекватна};$$

$$F_p \geq F_T - \text{модель не адекватна}, \quad (3.14)$$

де F_T – табличне значення F -критерію, визначене для заданого значення значимості і ступеня вільності головної дисперсії $f_1=N-d$ та дисперсії адекватності $f_2=N-(m-1)$.

Розрахункове значення критерію Фішера $F_{роз}$ визначається зі співвідношення [15-17]:

$$F_{роз} = \frac{S_{a\partial}^2}{S_{\epsilon}^2}. \quad (3.15)$$

Коефіцієнт детермінації показує, якою мірою варіація залежної змінної (критерій відгуку) у визначається варіацією незалежної змінної (фактора) x :

$$R^2 = 1 - \frac{S_u^2}{S_y^2}, \quad (3.16)$$

де:

$$S_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (3.17)$$

Коефіцієнт варіації розраховується також з використанням наступної формули:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (3.18)$$

де \hat{y} - значення критерію відгуку, отримані розрахунковим шляхом;

\bar{y} - середнє значення критерію відгуку, отримане дослідним шляхом;

y_i – значення критерію відгуку.

Квадрат емпіричного коефіцієнта кореляції між двома рядами спостережень (дослідними значеннями критерію відгуку y_i та його розрахунковими значеннями \hat{y}).

$$R^2 = \frac{\left(\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y}) \right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{y})^2}. \quad (3.19)$$

Коефіцієнт кореляції, або **індекс кореляції**, показує вплив фактора x_i , на критерій відгуку y_i :

$$R = \sqrt{R^2} . \quad (3.20)$$

Для спрощення розрахунків щільність кореляційного зв'язку характеризують коефіцієнтом кореляції, який розраховується за формулою:

$$R_1 = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} . \quad (3.21)$$

Якщо зв'язок між критерієм відгуку і факторами лінійний, то використовується **лінійний коефіцієнт кореляції**, який характеризує не лише щільністю зв'язку, але і його напрямком:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2} \cdot \sqrt{\overline{y^2} - \bar{y}^2}} . \quad (3.22)$$

Завдання до лабораторної роботи

Використовуючи дані з лабораторної роботи № 1, виконати наступні завдання:

1) Знайти коефіцієнт кореляції, коефіцієнт детермінації, значення критерію Фішера та степені вільності критеріїв Фішера і Стюдента.

2) Знайти коефіцієнти рівняння регресії b_0 і b_1 .

3) Знайти значення критерію Стюдента для коефіцієнтів рівняння регресії b_0 і b_1 .

4) Знайти табличне значення критерію Стюдента з рівнем значимості $\alpha=0,05$.

5) Перевірити статистичну значимість коефіцієнтів рівняння регресії b_0 і b_1 .

6) Знайти табличне значення критерію Фішера з рівнем значимості $\alpha=0,05$.

7) Перевірити регресійну модель на адекватність із використанням критерію Фішера.

8) За значенням коефіцієнта кореляції зробити висновок про близькість зв'язку до лінійного.

Приклад виконання лабораторної роботи

Створюємо нову електронну таблицю (spreadsheet) із індивідуальним завданням (рис. 3.1).

	1	2
	x_1	y
1	48	0,167
2	40	0,092
3	48	0,233
4	40	0,183
5	44	0,1
6	44	0,183
7	48	0,333
8	44	0,269
9	40	0,248
10	48	0,167

Рис. 3.1. Таблиця даних

Розпочнемо виконання завдання із знаходження коефіцієнтів кореляції, детермінації та визначення критерію Фішера. Перейшовши на вкладу "Статистика" (Statistics) панелі інструментів, активуємо модуль "Множинна регресія" (Multiple regression). В діалоговому вікні "Багаторядкова регресія" (Multiple line regression) в полі "змінні" (variables) вказуємо (рис. 3.2):

- залежна змінна (dependent): y ;
- незалежна змінна (independent): x_1 .

Після вибору змінних натискаємо "Ok", на екрані з'явиться наступне діалогове вікно: "Результати множинної регресії" (Multiple Regression Results). Перейшовши вкладки "Розширені налаштування" (Advanced) – обираємо команду "Підсумки: Результати регресії" (Summary: Regression results) (рис. 3.3).

На екрані з'явиться таблиця (рис. 3.4) із результатами розрахунків:

Regression Summary for Dependent Variable: y (Spreadsheet1)						
R= ,30499020 R²= ,09301902 Adjusted R²= ----- F(1,8)=,82047 p<,39151 Std.Error of estimate: ,07544						
N=10	b*	Std.Err. of b*	b	Std.Err. of b	t(8)	p-value
Intercept			-0,091261	0,319683	-0,285473	0,782533
x ₁	0,304990	0,336709	0,006504	0,007180	0,905799	0,391506

Рис. 3.4. Результати розрахунку коефіцієнтів

- коефіцієнтом кореляції: $R = 0,305$;
- коефіцієнтом детермінації: $R^2 = 0,093$;
- розрахунковим значенням критерію Фішера: $F_{\text{роз}} = 0,8205$;
- степенями свободи критерію Фішера: $f_1 = 1, f_2 = 8$;
- степінь свободи критерію Стюдента: $f_n = 8$.

Команда "Підсумки: Результати регресії" (Summary: Regression results) також розраховують коефіцієнти регресійного рівняння b_0 та b_1 . У стовпці b таблиці (рис. 3.4) розміщені розраховані коефіцієнти: $b_0 = -0,091261$, $b_1 = 0,006504$.

Значимість коефіцієнтів регресії перевіряють за допомогою критерію Стюдента згідно умови (3.10). Розрахункові значення критерію для коефіцієнтів b_0 та b_1 містяться в стовпці $t(8)$ (рис. 3.4): $t_{\text{роз}}(b_0) = -0,285473$, $t_{\text{роз}}(b_1) = 0,905799$.

Програмний продукт Statistica передбачає можливість визначення табличних значень критерію Стюдента, Фішера та багатьох інших. Для цього на вкладці "Статистика" (Statistics) панелі інструментів, мишкою активуємо команду "Калькулятор імовірностей" (Probability Calculators) та обираємо варіант виконання "Розподіли" (Distributions) (рис. 3.5).

В діалоговому вікні "Калькулятор розподілу імовірностей" (Probability Distribution Calculator) у вікні "Розподіл" (Distribution) задаємо "t-розподіл Стюдента" (t (Student)) та вказуємо наступні параметри (рис. 3.6):

- рівень значимості: $p = 0,05$;

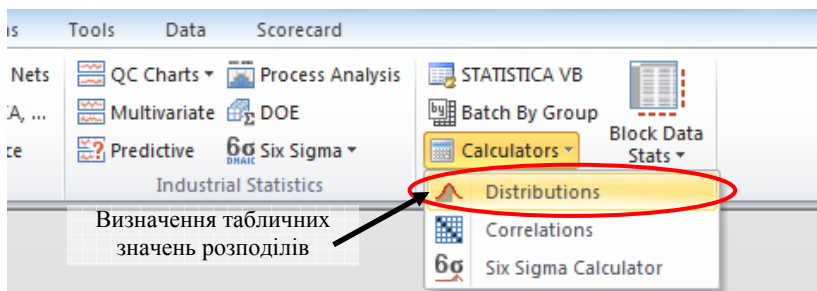


Рис. 3.5. Вибір варіанту виконання команди "Калькулятор імовірностей"

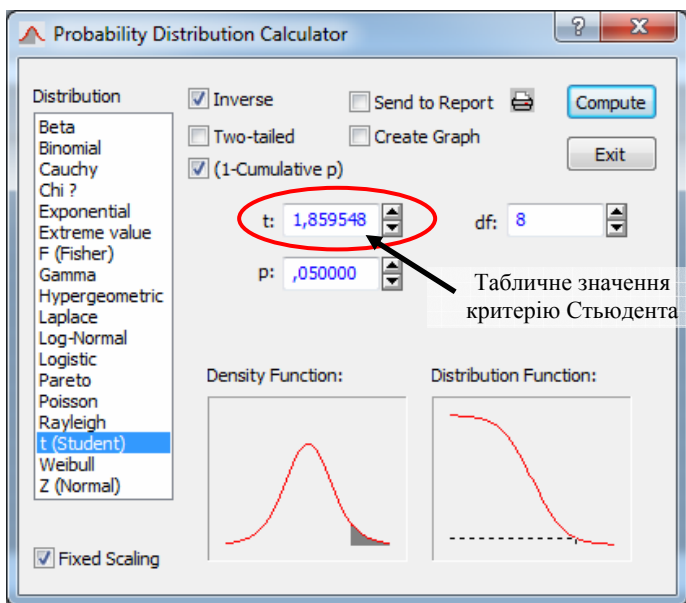


Рис. 3.6. Результат визначення табличного значення критерію Стюдента

- вказуємо степінь свободи: $df = 8$;
- натискаємо кнопку "Розрахувати" (Compute).

Маючи табличне значення критерію Стюдента ($t_{\text{табл}} = 1,859648$) визначимо статистичну значимість коефіцієнтів рівняння регресії (лінійної) ($b_0 = -0,091261$ та $b_1 = 0,006504$). Оскільки виконується умова (3.10) лише для розрахункового значення критерію Стюдента ($t_{\text{роз}}(b_1) = 0,905799$) більше

табличного значення ($t_{\text{табл}} = -1,859648$) – можна стверджувати, що лише коефіцієнт (b_1) рівняння регресії статистично значимий.

Адекватність моделі визначаємо за допомогою критерію Фішера. Для цього визначимо його табличне значення. В діалоговому вікні "Калькулятор розподілу імовірностей" (Probability Distribution Calculator) у вікні "Розподіл" (Distribution) задаємо "F-розподіл Фішера" (F (Fisher)) та вказуємо наступні параметри (рис. 3.7):

- рівень значимості: $p = 0,05$;
- вказуємо степені свободи: $df1 = 1$, $df2 = 8$;
- натискаємо кнопку "Розрахувати" (Compute).

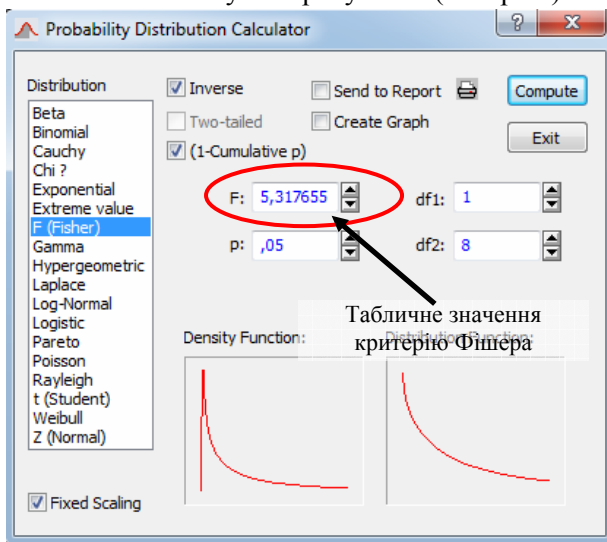


Рис. 3.7. Результат визначення табличного значення критерію Фішера

Оскільки виконується умова (3.14): $F_{\text{роз}}(0,82047)$ (рис. 3.4) є меншим $F_{\text{табл}}(5,31766)$ (рис. 3.7) отримана модель адекватна.

Оскільки коефіцієнт кореляції ($R = 0,305$) виходить за межі (від 0,6 до 0,9) – наявний недостатній лінійний зв'язок.

Варто пам'ятати, що на усіх етапах виконання лабораторної роботи отримані результати необхідно додавати до автозвіту.

Звіт про виконану лабораторну роботу має містити тему та завдання, отримані в процесі виконання таблиці, аналіз отриманих значень коефіцієнтів та висновок.

Зразок висновку

1) Статистичні характеристики досліджуваної вибірки в результаті виконання команди "Підсумки: Результати регресії" (рис. 3.4): коефіцієнт кореляції $(R) = \dots$, коефіцієнт детермінації $(R^2) = \dots$, розрахункове значення критерію Фішера $F_{\text{роз}} = (F(1,8)) = \dots$, число степенів свободи критерію Фішера $f_1 = \dots$ та $f_2 = \dots$, число степенів свободи критерію Стюдента $(t(8)) f_n = \dots$.

2) Коефіцієнти рівняння регресії b_0 та b_1 та розрахункове значення критерію Стюдента для кожного коефіцієнта (рис. 3.4) у стовпцях В і t(8): $b_0 = \dots$, $b_1 = \dots$, $t_{\text{роз}}(b_0) = \dots$, $t_{\text{роз}}(b_1) = \dots$. Табличне значення критерію Стюдента з рівнем значимості $p=0,05$: $t(8) = \dots$. Провести аналіз отриманих результатів із зазначенням чи виконується умова (3.10).

3) Табличне значення критерію Фішера з рівнем значимості $p=0,05$ для степенів свободи $f_1 = \dots$ та $f_2 = \dots$ становить: $F_{\text{табл}} = \dots$. Провести аналіз отриманих результатів із зазначенням чи виконується умова (3.14).

ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ

1. Що таке коефіцієнт кореляції?
2. Що таке коефіцієнт детермінації?
3. Що таке коефіцієнти рівняння регресії?
4. Як визначити коефіцієнти рівняння регресії?
5. У чому полягає суть статистичного аналізу результатів досліджень?
6. Як визначити значимість та взаємозв'язок коефіцієнтів рівняння регресії?
7. Опишіть методику визначення розрахункового значення критерію Стюдента?
8. Як визначити табличне значення критерію Стюдента?

9. Опишіть методику визначення розрахункового значення критерію Фішера?

10. Як визначити табличне значення критерію Фішера?

11. Для чого використовують коефіцієнт кореляції?

12. Для чого використовують коефіцієнт детермінації?

13. Як визначити головну дисперсію вибірки?

14. Як визначити дисперсію адекватності вибірки?

ЛАБОРАТОРНА РОБОТА №4

ВИЗНАЧЕННЯ ТОЧНОСТІ ПРОГНОЗУ ЛІНІЙНОЇ МОДЕЛІ

Мета роботи: Отримати практичні навички побудови довірчого інтервалу для прогнозу для різної ширини довірчих областей; розрахунку максимальної відносної помилки моделювання.

Обладнання робочого місця: персональний комп'ютер з необхідним програмним забезпеченням, методичні рекомендації, довідкова література.

Загальні відомості

Статистичне опрацювання результатів експериментальних досліджень в більшості випадків вимагає не лише знаходження для параметра y (критерію відгуку) відповідного числового значення, а і оцінки його надійності та точності. Оскільки точкова оцінка y^* за незначної кількості дослідів може містити суттєву похибку, тому наближена заміна y на y^* може спричинити значну недостовірність одержаних значень.

Для визначення точності оцінки y^* доцільно використовувати *довірчий інтервал*, а для надійності – *довірчу імовірність*.

Нехай для параметра y (критерію відгуку) отримано незміщену оцінку деяких експериментальних значень y^* . Основним завданням для нас виступає необхідність оцінки можливої похибки отриманих значень. Задавши вірогідність β (для прикладу $\beta=0,9$) знайдемо значення $\xi > 0$, для якого:

$$P = (|y^* - y| < \varepsilon) = \beta. \quad (4.1)$$

або,

$$P(y^* - \varepsilon < y < y^* + \varepsilon) = \beta, \quad (4.2)$$

де ε – неперервна випадкова величина;

P – розподіл імовірності.

Рівняння (4.2) означає, що невідому значення параметра u (критерію відгуку) з імовірністю β потрапить в інтервал $l_\beta(y^* - \varepsilon, y^* + \varepsilon)$. Критерій відгуку u є не випадковою величиною, а інтервал l_β в якому він знаходиться є величина випадкова, оскільки на положення інтервалу на осі впливає випадкова величина y^* (центр інтервалу). Дана імовірність означає, що випадковий інтервал l_β накріє точку u з певною імовірністю.

Інтервал l_β називається **довірчим інтервалом**, а імовірність β – **довірчою імовірністю** або надійністю β , що відповідає даному довірчому інтервалу l_β .

Завдання до лабораторної роботи

Використовуючи дані з лабораторної роботи № 1, виконати наступні завдання:

1. Отримати математичну модель (лінійну) та будувати на її основі графік із 80, 95 та 99% довірчими областями;
2. Розрахувати напівширину довірчого інтервалу для всіх точок вибірки, а також для двох будь-яких точок з області прогнозів (для 80, 95 та 99% коефіцієнта довіри);
3. Оцінити максимальну відносну помилку прогнозу (у відсотках) для 80, 95 та 99% коефіцієнта довіри.

Приклад виконання лабораторної роботи

Створюємо нову електронну таблицю (spreadsheet) із індивідуальним завданням (рис. 4.1).

Для створення графіку та розрахунку коефіцієнтів лінійної регресії перейдемо на вкладку "Graphs" (Графіки) панелі інструментів. Активувавши команду "Діаграми розсіювання" (Scatterplot) на екрані з'являється діалогове вікно в якому необхідно (рис. 4.2):

	1	2
	x_1	y
1	48	0,167
2	40	0,092
3	48	0,233
4	40	0,183
5	44	0,1
6	44	0,183
7	48	0,333
8	44	0,269
9	40	0,248
10	48	0,167

Рис. 4.1. Таблиця даних

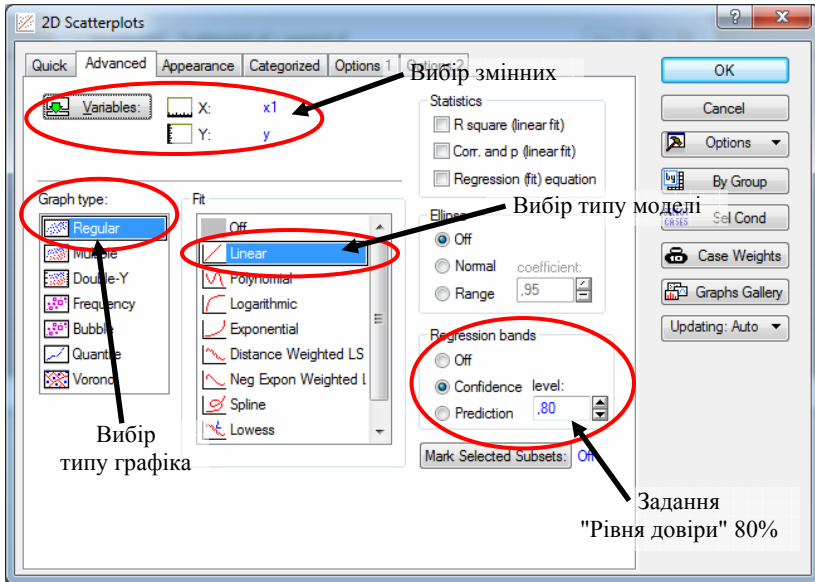


Рис. 4.2. Діалогове вікно команди "Діаграми розсіювання"

- у вікні "змінні" (variables) вказати: незалежну змінну (x_1) та залежну змінну (y);
- перейти на вкладку "Розширені налаштування"

- (Advanced) та задати тип моделі: "лінійна" (linear);
- вказати "Тип графіка" (Graph type): "звичайний" (regular);
 - перемкнути "Смуги регресії" (Regression bands) в режим: "Рівень довіри" (Confidence level);
 - задати "Рівень довіри" (Confidence level): 80%;
 - натиснути "Ok".

На екрані з'явиться графік та рівняння лінійної регресії із рівнем довіри 80% (рис. 4.3).

Аналогічно створюємо графіки із довірчими областями 95% (рис. 4.4) та 99% (рис. 4.5).

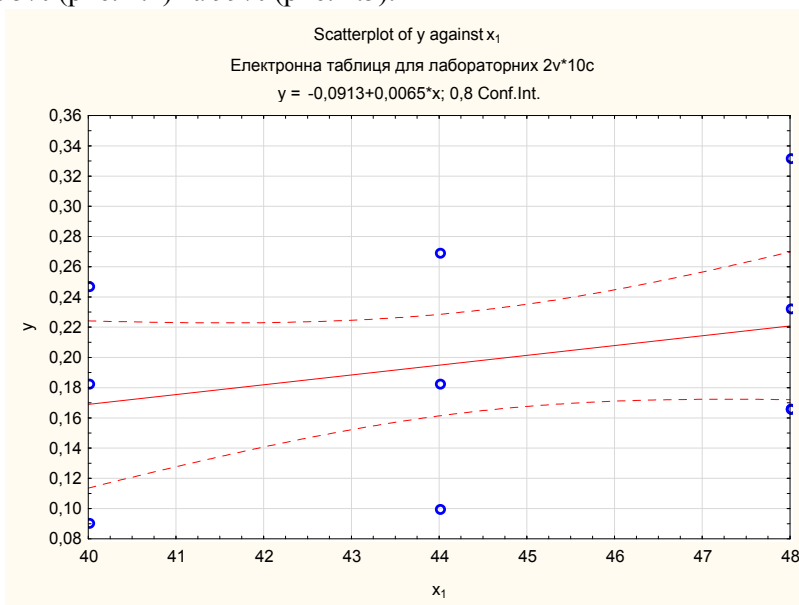


Рис. 4.2. Графік лінійної регресії із рівнем довіри 80%

Наступний крок – визначення величини параметрів необхідних для розрахунку напівширини довірчого інтервалу в точці x .

Напівширина δ довірчого інтервалу в точці x розраховується за формулою:

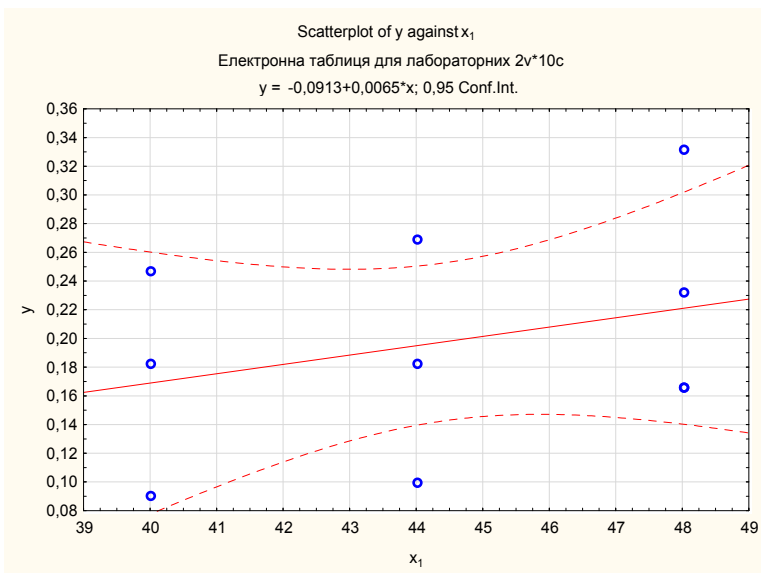


Рис. 4.3. Графік лінійної регресії із рівнем довіри 95%

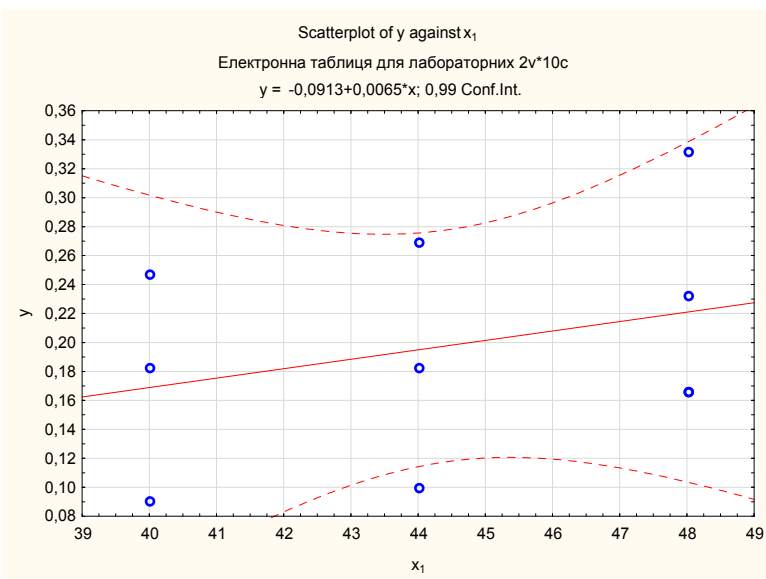


Рис. 4.4. Графік лінійної регресії із рівнем довіри 99%

$$\delta = \sigma_x t_y \sqrt{1 + \frac{1}{n} + \frac{(x - x_{cp})^2}{\sum_i (x - x_{cp})^2}}. \quad (4.3)$$

Визначаємо величину параметрів, що входять до формули для напівширини довірчого інтервалу:

1. Визначення табличного значення розподілу Стюдента (t_γ), що відповідає заданому рівню довіри (γ).

Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів, мишкою активуємо команду "Калькулятор імовірностей" (Probability Calculators) та обираємо варіант виконання "Розподіли" (Distributions).

В діалоговому вікні "Калькулятор розподілу імовірностей" (Probability Distribution Calculator) у вікні "Розподіл" (Distribution) задаємо "t-розподіл Стюдента" (t (Student)) та вказуємо наступні параметри (рис. 4.5):

- рівень значимості: $p = 0,8$;
- вказуємо степінь свободи ($n-2$): $df = 8$;
- натискаємо кнопку "Розрахувати" (Compute).

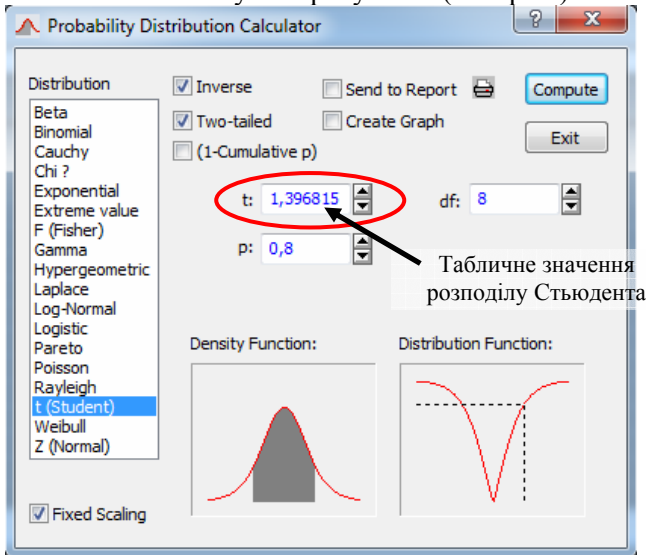


Рис. 4.5. Результат розрахунку табличного значення розподілу Стюдента

Аналогічно визначаємо табличне значення розподілу Стьюдента для рівнім значимості (довіри) 95 та 99%:

- $t(8) = 1,39682$, $p = 0,8$;
- $t(8) = 2,306$, $p = 0,95$;
- $t(8) = 3,35539$, $p = 0,99$.

2. Значення σ_e (стандартна похибка оцінки) отримаємо із таблиці "Результати множинної регресії" (Multiple Regression Results).

Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів, активуємо модуль "Множинна регресія" (Multiple regression). В діалоговому вікні "Багаторядкова регресія" (Multiple line regression) в полі "змінні" (variables) вказуємо: залежна змінна (dependent) y ; незалежна змінна (independent) x_1 .

Після вибору змінних натискаємо "Ok", на екрані з'явиться діалогове вікно "Результати множинної регресії" (Multiple Regression Results) в якому розрахований необхідний параметр (рис. 4.6): "Стандартна похибка оцінки" (Standard error of estimate) – $\sigma_e = 0,07544$.

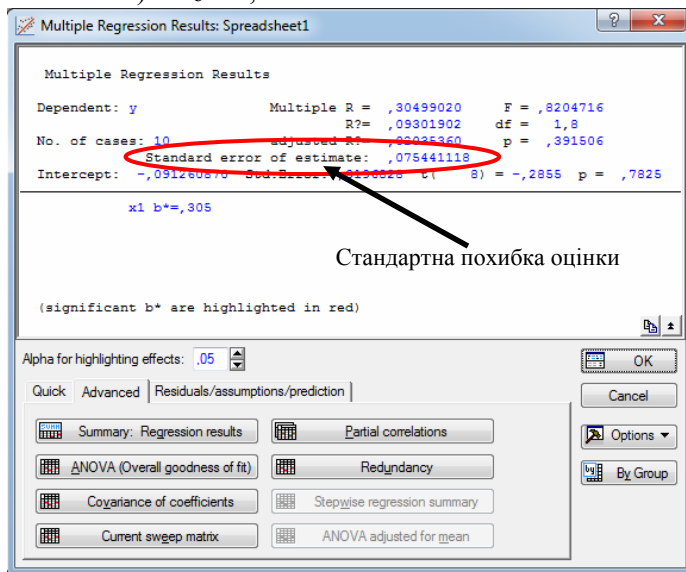


Рис. 4.6. Діалогове вікно "Результати множинної регресії"

3. Обсяг вибірки: $n=10$;

4. Середнє значення фактора: $x_{cp} = 44$.

Для кожного значення фактора (x_1) розраховуємо параметр "дисперсії" (квадрат відхилення від середнього значення): $(x_1 - x_{cp})^2$. Для цього в таблиці даних створюємо нову змінну "KV" Двічі натиснувши лівою кнопкою мишки на імені змінної – на екрані з'явиться діалогове вікно редагування змінної. У полі "Довга назва" (Long name) вписуємо рівняння: $(x_1 - 44)^2$. Після введення рівняння натискаємо кнопку "Ok".

Останній крок – визначення загальної дисперсії для фактора (x): $\Sigma((x_1 - x_{cp})^2)$.

Для цього скористаємось командою "Описові статистики" (Descriptive statistics). Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів активуємо вказаний модуль. В діалоговому вікні обираємо команду "Описові статистики" (Descriptive statistics) та натискаємо "Ok".

В діалоговому вікні вказуємо змінну "KV". Переходимо на вкладку "Розширені налаштування" (Advanced) де обираємо необхідний параметр: сума (sum). Після чого натискаємо кнопку "Швидкий аналіз" (Summary). Параметр "дисперсії" для фактора (x_1) становить 123,25 (рис. 4.7).

Descriptive Statistics (Електронна таблиця для лабо	
Variable	Sum
KV	123,2500

Рис. 4.7. Результат розрахунку сумарної дисперсії фактора x_1

Додаємо до електронної таблиці ще одну змінну: "y_reg". Та два значення до змінної " x_1 ": 41 та 42,5. В діалоговому вікні змінної "y_reg" в полі "Довга назва" (Long name) вписуємо рівняння математичної моделі (лінійне рівняння регресії): $y = -0,0913 + 0,0065 \cdot x_1$. Після введення рівняння моделі натискаємо кнопку "Ok".

Для кожного рівня довіри необхідно розрахувати напівширину довірчого інтервалу та максимальну помилку прогнозу. Розпочнемо із рівня довіри 80% (рис. 4.8):

Створюємо змінну "DELTA_80" та вводимо в діалоговому вікні в поле "Довга назва" (Long name) формулу: $=1,39682 \cdot 0,07544 \cdot (1 + 1/10 + KV/123,25)^{0,5}$.

Створюємо змінну "PROG_80" та вводимо в діалоговому вікні в поле "Довга назва" (Long name) формулу: $=DELTA_80/abs(y_reg) \cdot 100$.

	1 x1	2 y	3 KV	4 y_reg	5 DELTA_80	6 PROG_80
1	48	0,167	16	0,2207	0,11685908	52,9492869
2	40	0,092	16	0,1687	0,11685908	69,2703475
3	48	0,233	16	0,2207	0,11685908	52,9492869
4	40	0,183	16	0,1687	0,11685908	69,2703475
5	44	0,1	0	0,1947	0,11051939	56,7639378
6	44	0,183	0	0,1947	0,11051939	56,7639378
7	48	0,333	16	0,2207	0,11685908	52,9492869
8	44	0,269	0	0,1947	0,11051939	56,7639378
9	40	0,248	16	0,1687	0,11685908	69,2703475
10	48	0,167	16	0,2207	0,11685908	52,9492869
11	41		9	0,1752	0,1141288	65,1420106
12	42,5		2,25	0,18495	0,1114327	60,2501768

Рис. 4.8. Максимальна помилка прогнозу для $\gamma = 0,80$

Рівень довіри 95% (рис. 4.9):

Створюємо змінну "DELTA_95" та вводимо в діалоговому вікні в поле "Довга назва" (Long name) формулу: $=2,306 \cdot 0,07544 \cdot (1 + 1/10 + KV/123,25)^{0,5}$.

Створюємо змінну "PROG_95" та вводимо в діалоговому вікні в поле "Довга назва" (Long name) формулу: $=DELTA_95/abs(y_reg) \cdot 100$.

Рівень довіри 99% (рис. 4.10):

Створюємо змінну "DELTA_99" та вводимо в діалоговому вікні в поле "Довга назва" (Long name) формулу: $=3,35539 \cdot 0,07544 \cdot (1 + 1/10 + KV/123,25)^{0,5}$.

Створюємо змінну "PROG_99" та вводимо в діалоговому вікні в поле "Довга назва" (Long name) формулу: $=DELTA_99/abs(y_reg) \cdot 100$.

Варто пам'ятати, що на усіх етапах виконання лабораторної роботи отримані результати необхідно додавати до автозвіту.

	1	2	3	4	5	6
	x_1	y	KV	y_{reg}	DELTA 95	PROG 95
1	48	0,167	16	0,2207	0,1929218	87,4135935
2	40	0,092	16	0,1687	0,1929218	114,357914
3	48	0,233	16	0,2207	0,1929218	87,4135935
4	40	0,183	16	0,1687	0,1929218	114,357914
5	44	0,1	0	0,1947	0,18245565	93,7111729
6	44	0,183	0	0,1947	0,18245565	93,7111729
7	48	0,333	16	0,2207	0,1929218	87,4135935
8	44	0,269	0	0,1947	0,18245565	93,7111729
9	40	0,248	16	0,1687	0,1929218	114,357914
10	48	0,167	16	0,2207	0,1929218	87,4135935
11	41		9	0,1752	0,18841441	107,542473
12	42,5		2,25	0,18495	0,18396344	99,4665795

Рис. 4.9. Максимальна помилка прогнозу для $\gamma = 0,95$

	1	2	3	4	5	6
	x_1	y	KV	y_{reg}	DELTA 99	PROG 99
1	48	0,167	16	0,2207	0,28071461	127,192844
2	40	0,092	16	0,1687	0,28071461	166,398699
3	48	0,233	16	0,2207	0,28071461	127,192844
4	40	0,183	16	0,1687	0,28071461	166,398699
5	44	0,1	0	0,1947	0,26548564	136,356259
6	44	0,183	0	0,1947	0,26548564	136,356259
7	48	0,333	16	0,2207	0,28071461	127,192844
8	44	0,269	0	0,1947	0,26548564	136,356259
9	40	0,248	16	0,1687	0,28071461	166,398699
10	48	0,167	16	0,2207	0,28071461	127,192844
11	41		9	0,1752	0,27415604	156,481759
12	42,5		2,25	0,18495	0,26767957	144,730775

Рис. 4.10. Максимальна помилка прогнозу для $\gamma = 0,99$

Звіт про виконану лабораторну роботу має містити тему та завдання, отримані в процесі виконання таблиці та рисунки, а також висновок.

Зразок висновку

В результати виконання лабораторної роботи побудовано довірчі області для лінійної регресії $y = -0,0913 + 0,0065 \cdot x_1$ за трьох рівнів довіри: 80, 95 і 99%. Розраховано прогноз за лінійною регресією у всіх точках вибірки та у двох додаткових

точках з області прогнозів $x_1 = 41$ і $x_1 = 42,5$, та відносні помилки прогнозів.

Аналіз відносних похибок прогнозів показує, що підвищення рівня довіри з 80 до 99%, знижує точність прогнозу в 2,4 рази. Взявши контрольну точку, наприклад значення фактора $x_1 = 48$, порівнюємо відносні похибки: $PROG_80=52,949$; $PROG_95=87,413$; $PROG_99=127,192$.

ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ

1. Дайте визначення терміну "довірча імовірність"?
2. Дайте визначення терміну "довірчий інтервал"?
3. Як отримати графік лінійної регресії для різних значень рівня довіри?
4. Як параметри необхідні для розрахунку півширини довірчого інтервалу?
5. Як визначити табличне значення розподілу Стюдента?
6. Як оцінити максимальну відносну помилку прогнозу для різних значень рівня довіри?

ЛАБОРАТОРНА РОБОТА №5

ПЕРЕВІРКА ФАКТОРІВ НА МУЛЬТИКОЛІНЕАРНІСТЬ

Мета роботи: Отримати практичні навички використання методики перевірки факторів на колінеарність та мультиколінеарність.

Обладнання робочого місця: персональний комп'ютер з необхідним програмним забезпеченням, методичні рекомендації, довідкова література.

Загальні відомості

Однією з головних передумов застосування методу найменших квадратів для оцінки параметрів лінійних багатofакторних моделей – відсутність лінійних зв'язків між незалежними змінними моделі (факторами). Наявність таких зв'язків, тобто існування лінійної залежності, або кореляції, між двома чи більше факторами називають *мультиколінеарністю*.

Математично характеристика *мультиколінеарності* записується у вигляді співвідношення:

$$r_{x_i x_j} \rightarrow 1, i \neq j. \quad (5.1)$$

Необхідною умовою для оцінки параметрів загальної моделі МНК (методу найменших квадратів) є:

$$\sum_{i=1}^n u^2 = u'u = (Y - XA)'(Y - XA) = Y'Y - 2A'X'Y + A'X'XA. \quad (5.2)$$

яка стосується матриці вихідних даних.

Матриця має розміри $n \times m$ і ранг m , тобто серед

факторів не повинно бути лінійної залежності.

Мультиколінеарність в більшості випадків спричинює різке збільшення дисперсії оцінок факторів, кореляцію (залежність) похибок, чутливість критерію відгуку до кількості дослідів. Збільшення кількості дослідів спричинить збільшення похибки одержаних результатів. Одним з засобів усунення мультиколінеарності є *факторний аналіз*.

Ознаки мультиколінеарності. Симетрична матриця коефіцієнтів кореляції або кореляція нульового порядку між факторами надає інформацію про парну залежність коефіцієнтів кореляції:

$$r = \begin{pmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} & \dots & r_{x_1x_k} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} & \dots & r_{x_2x_k} \\ r_{x_3x_1} & r_{x_3x_2} & r_{x_3x_3} & \dots & r_{x_3x_k} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_kx_1} & r_{x_kx_2} & r_{x_kx_3} & \dots & r_{x_kx_k} \end{pmatrix}. \quad (5.3)$$

Коли рівень парних коефіцієнтів кореляції факторів дорівнює множинному коефіцієнту кореляції це вказує на наявність мультиколінеарності.

Дана ознака актуальна для двох факторних моделей. Більш загальна перевірка передбачає знаходження визначника (детермінанта) матриці r , який називається детермінантом кореляції і позначається $|r|$. Числові значення детермінанта кореляції задовольняють умову: $|r| \in [0,1]$.

Якщо $|r| = 0$, то існує повна мультиколінеарність, а коли $|r| = 1$, мультиколінеарність відсутня. Наближення значення $|r|$ до нуля збільшує імовірність мультиколінеарності між факторами. Незважаючи на те, що на числове значення $|r|$ впливає дисперсія факторів, цей показник можна вважати точковою мірою рівня мультиколінеарності.

Дослідження мультиколінеарності проводиться з використанням алгоритму **Феррара-Глобера**. Суть методу полягає у використанні трьох видів статистичних критеріїв (рис. 5.1): критерій χ^2 («хі-квадрат»); критерій Стюдента (t –

критерій); критерій Фішера (F – критерій).

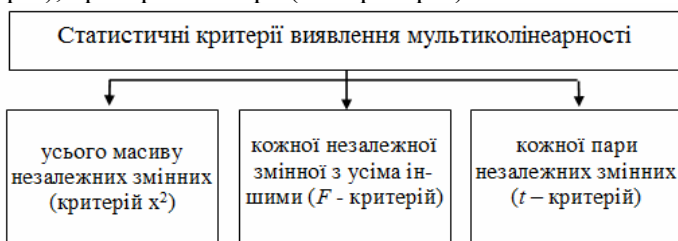


Рис. 5.1. Дослідження мультиколінеарності

Алгоритм **Феррара-Глобера** полягає в наступному:

– стандартизація змінних (факторів):

$$x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sqrt{n\sigma_{x_k}^2}}, \quad (5.4)$$

де n – число спостережень ($i = \overline{1, n}$);

m – число пояснювальних змінних, ($k = \overline{1, m}$);

\bar{x}_k – середнє арифметичне k -ї пояснювальної змінної;

$\sigma_{x_k}^2$ – дисперсія k -ї пояснювальної змінної.

– знаходження кореляційної матриці:

$$r = X^{*'} X^*, \quad (5.5)$$

де X^* – матриця стандартизованих незалежних (пояснювальних) змінних;

$X^{*'}$ – матриця, транспонована до матриці X^* .

– визначення критерію χ^2 («хі-квадрат»):

$$\chi^2 = - \left[n - 1 - \frac{1}{6}(2m + 5) \right] \ln|r|, \quad (5.6)$$

де $|r|$ м визначник кореляційної матриці r .

Розрахункове значення необхідно порівняти з табличним при рівні значимості α і ступенях свободи $1/2 \cdot m \cdot (m-1)$. Якщо $\chi_{\text{роз}}^2 > \chi_{\text{табл}}^2$, то в масиві факторів присутня мультиколінеарність.

– визначення оберненої матриці:

$$C = r^{-1} = (X^{*'} X^*)^{-1}. \quad (5.7)$$

– Визначення F -критерію:

$$F_k = (c_{kk} - 1) \frac{n - m}{m - 1}, \quad (5.8)$$

де c_{kk} – діагональні елементи матриці C .

Розрахункове значення F -критерію порівнюється з табличним при рівні значимості α і ступенях свободи $n - m$ і $m - 1$. Якщо $F_{\text{крроз}} > F_{\text{табл}}$, то відповідна k -й фактор мультиколінеарний з іншими.

– коефіцієнт детермінації для кожного фактора:

$$R_{x_k}^2 = 1 - \frac{1}{c_{kk}}. \quad (5.9)$$

– частинний коефіцієнт кореляції:

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} \cdot c_{jj}}}, \quad (5.10)$$

де c_{kj} – елемент матриці C , що міститься в k -му рядку і j -му стовпці;

c_{kk} і c_{jj} – діагональні елементи матриці C .

– визначення t -критерію:

$$t_{kj} = \frac{r_{kj} \sqrt{n - m}}{\sqrt{1 - r_{kj}^2}}. \quad (5.11)$$

Розрахункове значення критерію t_{kj} порівнюються з табличними при рівні значимості α та ступені свободи $n - m$. Якщо $t_{kj\text{роз}} > t_{\text{табл}}$ – між факторами x_k і x_j наявна мультиколінеарність.

Завдання до лабораторної роботи

Варіант індивідуального завдання для виконання лабораторної роботи (згідно порядкового номера студента в списку академічної групи) подано в дод. А.

В лабораторній роботі необхідно виконати:

1. Перевірити фактори x_1 та x_2 на мультиколінеарність;
2. Побудувати дві моделі: лінійну та степеневу (Кобба-

Дугласа) за рівні значимості $\alpha=0,05$;

3. Перевірити коефіцієнти рівнянь на значимість;
4. Перевірити адекватність отриманих моделей;
5. Вибрати оптимальну модель згідно параметру дисперсії.

Приклад виконання лабораторної роботи

Створюємо електронну таблицю (spreadsheet) для свого варіанту індивідуального завдання (дод. А). Для цього активуємо команду "Створити новий документ" (Create a new docement). В діалоговому вікні вказуємо три змінні та десять значень (рис. 5.2). Для даного прикладу (табл. 5.1) це:

- два фактори: x_1 – величина тиску, x_2 – діаметр перепускного отвору;
- критерій відгуку: y – час відкачування повітря.

Таблиця 5.1

Приклад індивідуального завдання для виконання лабораторної роботи

№	x_1	x_2	y
1	48	4	0,167
2	40	4	0,092
3	48	3	0,233
4	40	3	0,183
5	44	4	0,1
6	44	3	0,183
7	48	2	0,333
8	44	2	0,269
9	40	2	0,248
10	48	4	0,167

Проведемо перевірку факторів на мультиколінеарність. Оскільки аналіз проводиться лише для двох факторів, варто використати перевірку на мультиколінеарність за значенням парного коефіцієнта кореляції $r_{x_1x_2}$. Використаємо для цього модуль "Основні статистичні дані/таблиці" (Basic Statistics /Tables).

Перейшовши на вкладку "Статистика" (Statistics) панелі

		1	2	3
		x_1	x_2	y
	1	48	4	0,167
	2	40	4	0,092
	3	48	3	0,233
	4	40	3	0,183
	5	44	4	0,1
	6	44	3	0,183
	7	48	2	0,333
	8	44	2	0,269
	9	40	2	0,248
	10	48	4	0,167

Рис. 5.2. Таблиця даних

інструментів активуємо модуль "Основні статистичні дані/таблиці" (Basic Statistics/Tables). В діалоговому вікні, що з'явилося (рис. 5.3) обираємо команду "Кореляційні матриці" (Correlation matrices) та натискаємо "Ok".

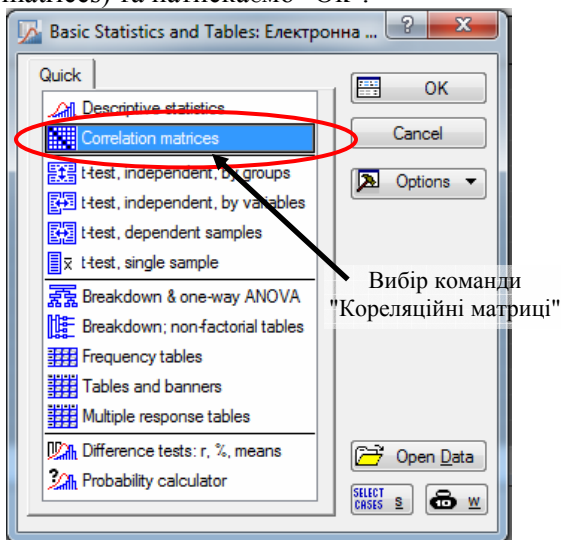


Рис. 5.3. Діалогове вікно модуля "Основні статистичні дані/таблиці"

На екрані з'явиться наступне діалогове вікно "Продукт-момент і часткові кореляції" (Product-moment and partial correlations) в якому необхідно обрати параметр "Два списки (пряма матриця)" (Two lists (rect.matrix)) та вказати два фактори (рис. 5.4) і натиснути "Швидкий аналіз" (Summary).

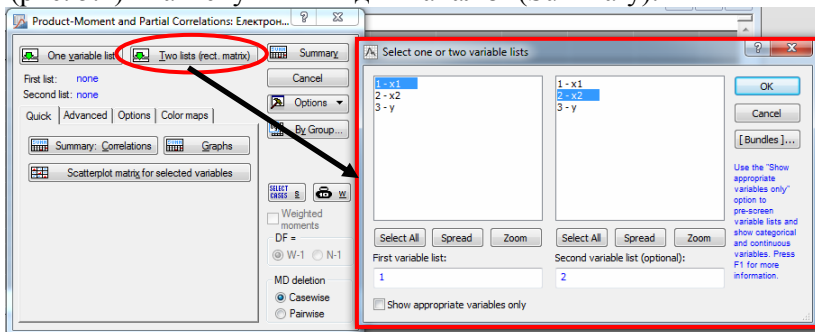


Рис. 5.4. Діалогове вікно модуля "Продукт-момент і часткові кореляції"

Парний коефіцієнт кореляції $r_{x_1x_2} = 0,13$ (рис. 5.5). Статистична значимість даного коефіцієнта перевіряється із використанням коефіцієнта Стюдента.

Correlations (Електронна таблиця для лабораторних_2)	
Marked correlations are significant at $p < ,05000$	
N=10 (Casewise deletion of missing data)	
Variable	x_2
x_1	0,130435

Рис. 5.5. Результати розрахунку коефіцієнта кореляції

В даному випадку розрахункове значення ($t_{\text{роз}}$) критерію Стюдента визначається із залежності:

$$t_{\text{роз}} = r_{x_1x_2} \cdot \left((n-2) / (1 - r_{x_1x_2}^2) \right)^{0,5}. \quad (5.12)$$

Згідно (5.12) $t_{\text{роз}} = 0,371$.

Табличне значення коефіцієнта Стюдента ($t_{\text{таб}}$) визначаємо за допомогою команди "Калькулятор імовірностей" (Probability Calculators). В діалоговому вікні "Калькулятор розподілу імовірностей" (Probability Distribution Calculator) у вікні "Розподіл" (Distribution) задаємо "t-розподіл Стюдента" (t

(Student) та вказуємо наступні параметри (рис. 5.6):

- рівень значимості: $p = 0,05$;
- вказуємо ступінь свободи ($n-2$): $df = 8$;
- натискаємо кнопку "Розрахувати" (Compute).

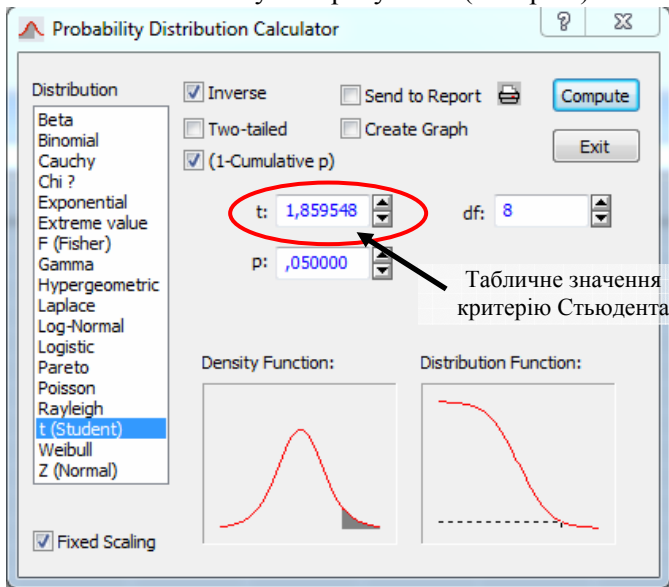


Рис. 5.6. Результат визначення табличного значення критерію Стьюдента

Оскільки не виконується умова (3.10):

$$t_{\text{роз}}(0,371) < t_{\text{таб}}(1,8595).$$

коефіцієнт кореляції статистично не значимий.

Отже, можна зробити висновок, що фактори x_1 та x_2 неколінеарні.

Наступний етап – побудова двох моделей:

– лінійної: $y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2$.

– степеневі Кобба-Дугласа: $y = Ax_1^{a_1} \cdot x_2^{a_2}$.

та перевірка коефіцієнтів регресії цих рівнянь на значимість, а моделей на адекватність.

Коефіцієнти рівняння лінійної регресії (b_0 , b_1 , b_2) розрахуємо за допомогою "Множинна регресія" (Multiple regression).

В діалоговому вікні "Багаторядкова регресія" (Multiple line regression) в полі "змінні" (variables) вказуємо (рис. 5.7):

- залежна змінна (dependent): y ;
- незалежні змінні (independent): x_1, x_2 .

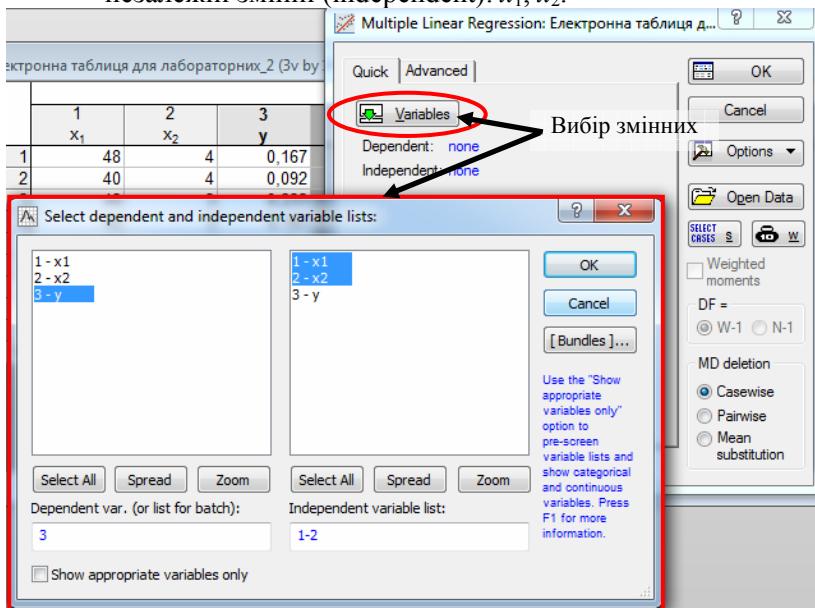


Рис. 5.7. Діалогове вікно "Багаторядкова регресія"

Після вибору змінних натискаємо "Ok", на екрані з'явиться наступне діалогове вікно: "Результати множинної регресії" (Multiple Regression Results). Перейшовши вкладку "Розширені налаштування" (Advanced) – обираємо команду "Підсумки: Результати регресії" (Summary: Regression results) (рис. 5.8).

На екрані з'явиться таблиця (рис. 5.9) із результатами розрахунків:

- розрахунковим значенням критерію Фішера: $F_{\text{роз}} = 96,745$;
- степенями свободи критерію Фішера: $f_1 = 2, f_2 = 7$;
- степінь свободи критерію Стюдента: $f_n = 7$;
- розраховані коефіцієнти: $b_0 = 0,041474, b_1 = 0,009123$,

$$b_2 = -0,080340.$$

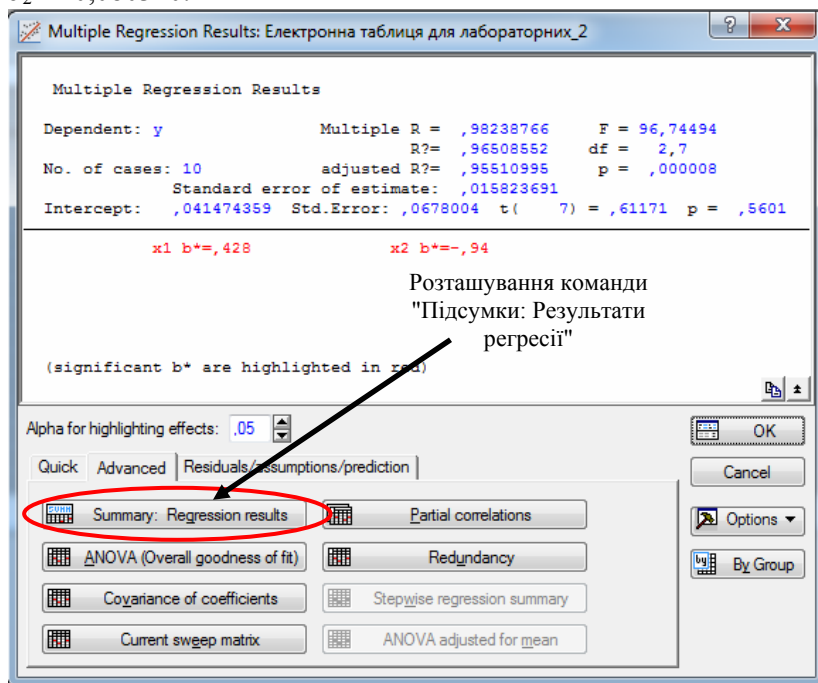


Рис. 5.8. Діалогове вікно "Результати множинної регресії"

Regression Summary for Dependent Variable: y (Електронна таблиця для лаб						
R= .98238766 R²= .96508552 Adjusted R²= .95510995						
F(2,7)=96,745 p<.00001 Std.Error of estimate: .01582						
N=10	b*	Std.Err. of b*	b	Std.Err. of b	t(7)	p-value
Intercept			0,041474	0,067800	0,6117	0,560062
x ₁	0,427846	0,071233	0,009123	0,001519	6,0063	0,000539
x ₂	-0,941892	0,071233	-0,080340	0,006076	-13,2227	0,000003

Рис. 5.9. Результати розрахунку коефіцієнтів

Значимість коефіцієнтів регресії перевіряють за допомогою критерію Стюдента згідно умови (3.10). Розрахункові значення критерію для коефіцієнтів b_0 , b_1 , b_2 містяться в стовпці $t(7)$ (рис. 5.9): $t_{\text{роз}}(b_0) = 0,6117$, $t_{\text{роз}}(b_1) = 6,0063$ та $t_{\text{роз}}(b_2) = -13,227$.

Табличне значення критерію Стюдента розраховане за допомогою команди "Калькулятор імовірностей" (Probability

Calculators) становить: $t_{\text{таб}} = 1,859548$ (для рівня значимості $p = 0,05$ та степені свободи $df = 7$).

Аналіз показав значимість лише двох коефіцієнтів рівняння лінійної регресії (b_1 та b_2). Лінійна модель регресії набуде вигляду:

$$y = 0,009123 \cdot x_1 - 0,080340 \cdot x_2. \quad (5.13)$$

Адекватність моделі визначаємо за допомогою критерію Фішера. Для цього визначимо його табличне значення. В діалоговому вікні "Калькулятор розподілу імовірностей" (Probability Distribution Calculator) у вікні "Розподіл" (Distribution) задаємо "F-розподіл Фішера" (F (Fisher)) та вказуємо наступні параметри (рис. 5.10):

- рівень значимості: $p = 0,999$;
- вказуємо степені свободи: $df1 = 2$, $df2 = 7$;
- натискаємо кнопку "Розрахувати" (Compute).

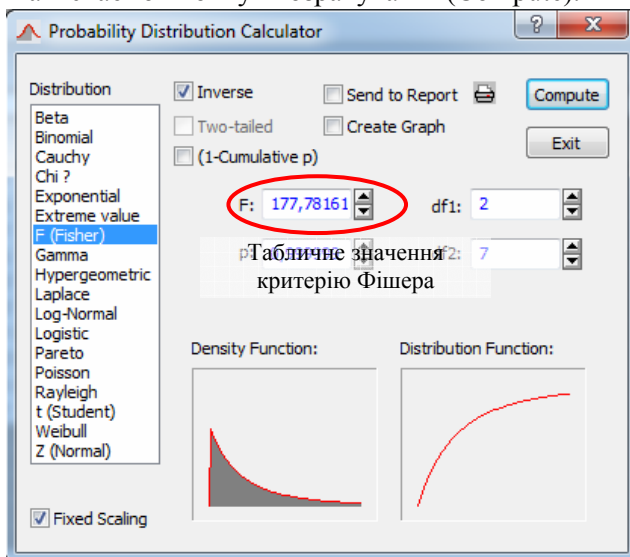


Рис. 5.10. Результат визначення табличного значення критерію Фішера

Оскільки виконується умова (3.14): $F_{\text{роз}} (96,745)$ (рис. 5.9) менше за $F_{\text{табл}} (177,78)$ (рис. 5.10) отримана модель адекватна.

Розглянемо процес створення степеневі моделі Кобба-Дугласа. Побова степеневі моделі потребує лінеаризації вибірки за формулами: $V=\ln(y)$, $U_1=\ln(x_1)$, $U_2=\ln(x_2)$. Додаємо нові змінні до електронної таблиці, розраховуємо їхні значення (згідно приведених формул) (рис. 5.11). Варто пам'ятати, що у програмному продукті Statistica натуральний логарифм $\ln(x)$ записується, як $\log(x)$.

	1	2	3	4	5	6
	x_1	x_2	y	V	U_1	U_2
1	48	4	0,167	-1,7897615	3,87120101	1,38629436
2	40	4	0,092	-2,3859667	3,68887945	1,38629436
3	48	3	0,233	-1,4567168	3,87120101	1,09861229
4	40	3	0,183	-1,6982691	3,68887945	1,09861229
5	44	4	0,1	-2,3025851	3,78418963	1,38629436
6	44	3	0,183	-1,6982691	3,78418963	1,09861229
7	48	2	0,333	-1,0996128	3,87120101	0,69314718
8	44	2	0,269	-1,3130439	3,78418963	0,69314718
9	40	2	0,248	-1,3943265	3,68887945	0,69314718
10	48	4	0,167	-1,7897615	3,87120101	1,38629436

Рис. 5.11. Результати лінеаризації вибірки

Побудуємо рівняння лінійної регресії для отриманих змінних (V , U_1 , U_2) використавши аналогічну послідовність дій, що і для змінних (y , x_1 , x_2). Активуємо модуль "Множинна регресія" (Multiple regression). В діалоговому вікні "Багаторядкова регресія" (Multiple line regression) в полі "змінні" (variables) вказуємо:

- залежна змінна (dependent): V ;
- незалежні змінні (independent): U_1 , U_2 .

Після вибору змінних натискаємо "Ok", на екрані з'явиться наступне діалогове вікно: "Результати множинної регресії" (Multiple Regression Results). Перейшовши вкладу "Розширені налаштування" (Advanced) – обираємо команду "Підсумки: Результати регресії" (Summary: Regression results).

На екрані з'явиться таблиця (рис. 5.12) із результатами розрахунків:

- розрахунковим значенням критерію Фішера: $F_{\text{роз}} = 30,998$;
- степенями свободи критерію Фішера: $f_1 = 2, f_2 = 7$;

- степінь свободи критерію Стьюдента: $f_n = 7$;
- розраховані коефіцієнти: $b_0 = -8,69358$, $b_1 = 2,19816$, $b_2 = -1,21897$.

Regression Summary for Dependent Variable: V (Електронна таблиця для ла						
R= ,94791594 R²= ,89854462 Adjusted R²= ,86955737						
F(2,7)=30,998 p<.00033 Std.Error of estimate: ,14804						
N=10	b*	Std.Err. of b*	b	Std.Err. of b	t(7)	p-value
Intercept			-8,69358	2,350189	-3,69910	0,007664
U ₁	0,427347	0,121292	2,19816	0,623894	3,52329	0,009685
U ₂	-0,899747	0,121292	-1,21897	0,164325	-7,41804	0,000147

Рис. 5.12. Результати розрахунку коефіцієнтів

Перевірка згідно критерію Стьюдента показала, що усі коефіцієнти значимі. Модель набуде вигляду:

$$V = -8,69358 + 2,19816 \cdot U_1 - 1,21897 \cdot U_2. \quad (5.14)$$

Перевіримо адекватність отриманої моделі. Оскільки виконується умова (3.14): $F_{\text{роз}}(30,998)$ (рис. 5.12) менше за $F_{\text{табл}}(177,78)$ (рис. 5.10) отримана модель адекватна.

Повернення до вихідних параметрів степеневі моделі:

$$A = e^{b_0}, \quad a_1 = b_1, \quad a_2 = b_2.$$

$$y = 0,49982 \cdot x_1^{2,19816} \cdot x_2^{-1,21897}.$$

Останній етап лабораторної роботи – вибір оптимальної моделі. Для цього використаємо вже відомий параметр "дисперсію", тобто "мінімум суми квадратів залишків".

Для початку проведемо моделювання за двома моделями. В електронній таблиці додаємо дві змінні: "y_lin" та "y_step". В діалогових вікнах створених змінних (поле "Довга назва" (Long name)) записуємо відповідні формули.

Визначимо параметр "дисперсії" для лінійної моделі. Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів активуємо модуль "Множинна регресія" (Multiple regression). В діалоговому вікні, що з'явилось (рис. 5.13) у полі "змінні" (variables) вказуємо:

- залежна змінна (dependent): y;

– незалежна змінна (independent): y_lin.

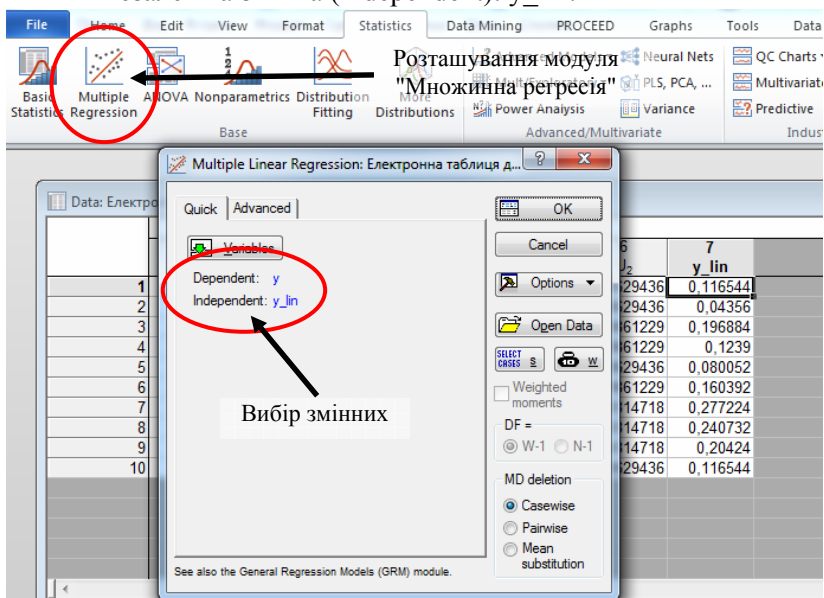


Рис. 5.13. Діалогове вікно модуля "Множинна регресія"

Після вибору змінних натискаємо "Ok", на екрані з'явиться наступне діалогове вікно: "Результати множинної регресії" (Multiple Regression Results) (рис. 5.14). Переходимо на вкладку "Розширені налаштування" (Advanced) де обираємо команду "Дисперсія (загальна підгонка)" (ANOVA (Overall goodness of fit)).

На екрані з'явиться таблиця (рис. 5.15) із розрахованою дисперсією для лінійної моделі (0,001753).

Розрахуємо "дисперсію" для степеневій функції. Для цього в електронній таблиці створимо нову змінну "KV_Zal_step". В діалоговому вікні створеної змінної (поле "Довга назва" (Long name)) записуємо формулу (рис. 5.16):

$$KV_Zal_step = (y_step - y)^2$$

Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів активуємо вказаний модуль. В діалоговому вікні,

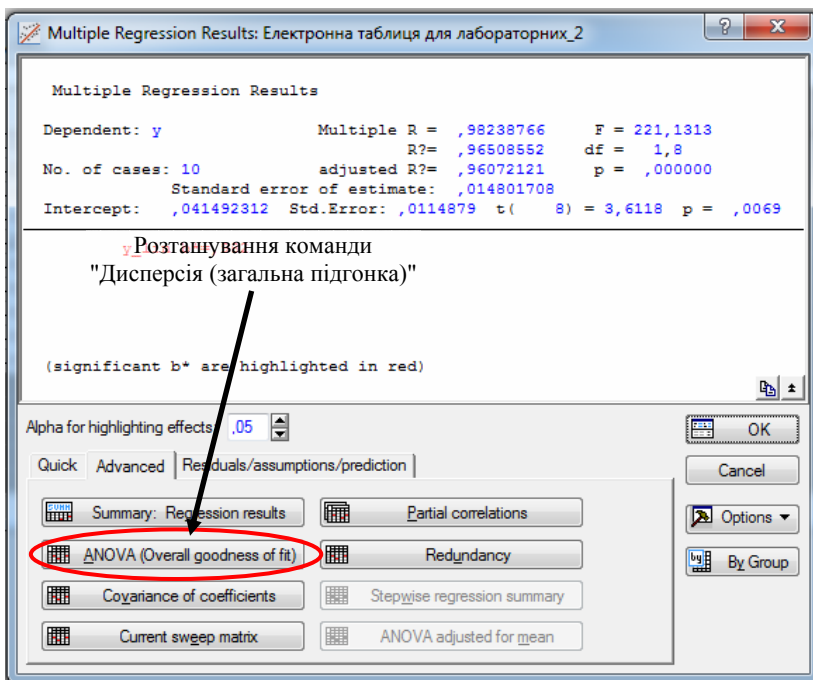


Рис. 5.14. Діалогове вікно "Результати множинної регресії"

Effect	Analysis of Variance; DV: y (Електронна таблиця для ла				
	Sums of Squares	df	Mean Squares	F	p-value
Regress.	0,048448	1	0,048448	221,1313	0,000000
Residual	0,001753	8	0,000219		
Total	0,050201				

Рис. 5.15. Результати розрахунку дисперсії для лінійної моделі

що з'явилося обираємо команду "Описові статистики" (Descriptive statistics) та натискаємо "Ok".

В діалоговому вікні команди "Описові статистики" (Descriptive statistics) у вікні "змінні" (variables) вказуємо змінну "KV_Zal_step". Переходимо на вкладку "Розширені налаштування" (Advanced) де обираємо необхідний параметр: сума (sum).

Після чого натискаємо кнопку "Швидкий аналіз" (Summary).

	1	2	3	4	5	6	7	8	9
	x_1	x_2	y	V	U_1	U_2	y_{lin}	y_{step}	KV_{Zal_step}
1	48	4	0,167	-1,7897615	3,87120101	1,38629436	0,116544	457,569351	209216,91
2	40	4	0,092	-2,3859667	3,68887945	1,38629436	0,04356	306,48458	93876,4133
3	48	3	0,233	-1,4567168	3,87120101	1,09861229	0,196884	649,760787	421886,346
4	40	3	0,183	-1,6982691	3,68887945	1,09861229	0,1239	435,216349	189254,015
5	44	4	0,1	-2,3025851	3,78418963	1,38629436	0,080052	377,914792	142744,017
6	44	3	0,183	-1,6982691	3,78418963	1,09861229	0,160392	536,649171	287795,952
7	48	2	0,333	-1,0996128	3,87120101	0,69314718	0,277224	1065,13207	1133797,06
8	44	2	0,269	-1,3130439	3,78418963	0,69314718	0,240732	879,711814	773419,664
9	40	2	0,248	-1,3943265	3,68887945	0,69314718	0,20424	713,436235	508637,459
10	48	4	0,167	-1,7897615	3,87120101	1,38629436	0,116544	457,569351	209216,91

Рис. 5.16. Результати розрахунків залишків для степеневі моделі

Отримана таблиця (рис. 5.17) містить параметр дисперсії для степеневі моделі.

Descriptive Statistics (Електронна таблиця для	
Variable	Sum
KV_Zal_step	3969845

Рис. 2.16. Дисперсія для степеневі моделі

Оскільки отриманий параметр дисперсії для лінійної моделі (0,001753) значно менший ніж для степеневі (3969845) – обираємо лінійну.

Варто пам'ятати, що на усіх етапах виконання лабораторної роботи отримані результати необхідно додавати до автозвіту.

Звіт про виконану лабораторну роботу має містити тему та завдання, отримані в процесі виконання таблиці та рисунки, а також висновок.

ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ

1. Дайте визначення терміну "мультиколінеарність" факторів?
2. Які наслідки наявності явища мультиколінеарності факторів?
3. Перерахуйте основні ознаки мультиколінеарності?

4. Які статистичні критерії використовують для визначення мультиколінеарності?

5. Дайте коротку характеристику алгоритму Фаррара - Глобера?

6. Який статистичний критерій свідчить про наявність мультиколінеарності в масиві факторів? Опишіть методику визначення.

7. Який статистичний критерій свідчить про наявність мультиколінеарності кожного фактора з усіма іншими? Наведіть відповідні формули.

8. Який статистичний критерій свідчить про наявність мультиколінеарності кожної пари факторів?

ЛАБОРАТОРНА РОБОТА №6

ДИСПЕРСІЙНИЙ АНАЛІЗ В СЕРЕДОВИЩІ STATISTICA

Мета роботи: Отримати практичні навички виявлення впливу на статистичну ознаку (дисперсійний аналіз) в середовищі Statistica із використанням спеціалізованого модуля Anova.

Обладнання робочого місця: персональний комп'ютер з необхідним програмним забезпеченням, методичні рекомендації, довідкова література.

Загальні відомості

Середні значення вимірюваних величин залежать від множини основних факторів (як якісних так і кількісних), які визначають умови проведення дослідів і випадкових факторів. Завданням дисперсійного аналізу – вивчення впливу факторів на зміни середніх значень величин. Залежно від числа джерел дисперсії (кількості факторів) розрізняють *однофакторний* і *багатофакторний дисперсійний аналіз*. Багатофакторний дисперсійний аналіз більш ефективний в порівнянні з класичним методом дослідження, при якому змінюється лише один фактор при незмінності усіх інших, що не дозволяє визначити вплив взаємодії різних факторів на результати експерименту.

При дисперсійному аналізі кожне спостереження використовується для одночасної оцінки всіх факторів і їх

взаємодії. Суть аналізу полягає у виділенні і оцінці окремих факторів, що впливають на значення середнього. При цьому сумарна вибіркова дисперсія розкладається на складові, які зумовлені дією незалежних факторів. Вплив даного фактора признається значимим, якщо відповідна йому вибіркова дисперсія значно відрізняється від дисперсії відтворюваності, яка у свою чергу зумовлена випадковими помилками. Перевірка значимості оцінки дисперсій проводиться з використанням **критерію Фішера**.

Розрізняють фактори двох типів:

- із випадковими рівнями;
- із фіксованими рівнями.

В першому випадку вибір рівнів факторів відбувається із безкінечної сукупності можливих значень. Відповідно, математична модель в такому випадку – **модель із випадковими рівнями факторів**. І відповідно **модель із фіксованими рівнями**, якщо фактор приймає лише деякі із фіксованих значень.

Однофакторний дисперсійний аналіз. Розглянемо вплив на результати досліджень одиничного фактора A , що приймає k значень (фактор A має k фіксованих рівнів a_i , $i = 1, 2, \dots, k$). Нехай y_{ij} результат j – дослідів із n_i числа вимірювань ($j = 1, 2, \dots, n_i$), виконаних на i -рівні фактора A (табл. 6.1).

Результат кожного дослідів можна представити у вигляді наступної моделі:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad (6.1)$$

де μ – сумарний ефект у всіх експериментах;

α_i – ефект, зумовлений впливом фактора A на i -рівні;

ε_{ij} – випадкова помилка дослідів на i -рівні.

Для відповіді на питання про значимість впливу фактора A , необхідно перевірити нульову гіпотезу рівності математичного сподівання сум $(\mu + \alpha_i)$ на усіх рівнях цього фактора:

$$H_0 : m_1 = m_2 = \dots = m_k = m, \quad (6.2)$$

де $m_i = M\{\mu + \alpha_i\}$.

Таблиця 6.1

Вихідні дані для однофакторного дисперсійного аналізу

Номер досліджу	Рівні фактора A			
	a_1	a_2	\dots	a_k
1	y_{11}	y_{21}	\dots	y_{k1}
2	y_{12}	y_{22}	\dots	y_{k2}
\dots	\dots	\dots	\dots	\dots
n	y_{1n}	y_{2n}	\dots	y_{kn}
Σ	B_1, C_1	B_2, C_2	\dots	B_k, C_k

Відповідно, загальна кількість експериментів:

$$N = n_1 + n_2 + \dots + n_k = kn. \quad (6.3)$$

Сума результатів всіх експериментів на i -рівні:

$$B_i = \sum_{j=1}^n y_{ij}. \quad (6.4)$$

Сума квадратів результатів на i -рівні:

$$C_i = \sum_{j=1}^n y_{ij}^2. \quad (6.5)$$

Середнє значення для вибірки із N експериментів:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^n y_{ij} = \frac{1}{k} \sum_{i=1}^k \bar{y}_i = \frac{1}{kn} \sum_{i=1}^k B_i. \quad (6.6)$$

Загальна вибіркова дисперсія експериментів:

$$S^2 = \frac{\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y})^2}{N-1} = \frac{1}{N-1} \left[\sum_{i=1}^k \sum_{j=1}^n y_{ij}^2 - \frac{1}{N} \left(\sum_{i=1}^k \sum_{j=1}^n y_{ij} \right)^2 \right], \quad (6.7)$$

$$= \frac{1}{N-1} \left[\sum_{i=1}^k C_i - \frac{1}{N} \left(\sum_{i=1}^k B_i \right)^2 \right]$$

а вибіркова дисперсія на i -рівні:

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 = \frac{1}{n-1} \left[\sum_{j=1}^n y_{ij}^2 - \frac{1}{n} \left(\sum_{j=1}^n y_{ij} \right)^2 \right] =$$

$$= \frac{1}{n-1} \left[C_i - \frac{B_i^2}{n} \right] \quad (6.8)$$

Якщо вибіркові дисперсії S_i^2 однорідні (перевірка за критерієм Кохрена), кращою оцінкою дисперсії $\sigma_{ном}^2$, що характеризує вплив випадкових факторів, буде вибіркова дисперсія

$$S_{ном}^2 = \frac{1}{k} \sum_{i=1}^k S_i^2. \quad (6.9)$$

з числом степенів свободи $f_{ном} = k(n-1) = N - k$.

Наближено оцінити дисперсію фактора A можна наступним чином:

$$\sigma_A^2 \approx S^2 - S_{ном}^2. \quad (6.10)$$

Для отримання більш точної оцінки скористаємось відхиленням середніх на фіксованих рівнях від загального середнього:

$$\frac{1}{k-1} \sum_{i=1}^k (\bar{y}_i - \bar{y})^2 \approx \sigma_A^2 + \frac{\sigma_{ном}^2}{n} \approx \sigma_A^2 + \frac{S_{ном}^2}{n}. \quad (6.11)$$

В даному випадку під дисперсією фактора A розуміють математичне сподівання середнього квадрата відхилення, зумовленого впливом цього фактора. **Вибіркова дисперсія:**

$$S_A^2 = \frac{n}{k-1} \sum_{i=1}^k (\bar{y}_i - \bar{y})^2 \approx n\sigma_A^2 + S_{ном}^2. \quad (6.12)$$

з числом степенів свободи $f_A = k-1$ використовується для перевірки нульової гіпотези (6.2) за **критерієм Фішера**.

Якщо нульова гіпотеза $(H_0 : \sigma_A^2 = \sigma_{ном}^2)$ правильна, виконується умова:

$$\left(\frac{S_A^2}{S_{ном}^2} \right) \leq F_{1-p}. \quad (6.13)$$

Різниця між дисперсіями S_A^2 і $S_{ном}^2$ незначна і відповідно вплив фактора A на результати експериментів також незначний.

Якщо:

$$\left(\frac{S_A^2}{S_{ном}^2} \right) > F_{1-p}. \quad (6.14)$$

Нульова гіпотеза про рівність математичних сподівань

сум $(\mu + \alpha_i)$ відхиляється (вплив фактора A значний). Оцінити вплив фактора можна на основі (6.12):

$$\sigma_A^2 = \frac{S_A^2 - S_{ном}^2}{n}. \quad (6.15)$$

Якщо для кожного рівня було виконано не однакову кількість експериментів, вибіркова дисперсія фактора розраховується за формулою:

$$S_A^2 = \frac{1}{k-1} \left[\sum_{i=1}^k \frac{B_i^2}{n_i} - \frac{1}{N} \left(\sum_{i=1}^k B_i \right)^2 \right]. \quad (6.16)$$

Вибіркова дисперсія, що характеризує вплив випадкових факторів:

$$S_{ном}^2 = \frac{\sum_{i=1}^k f_i s_i^2}{\sum_{i=1}^k f_i}, \quad (6.17)$$

де $f_i = n_i - 1$ – число степенів вільності $S_{ном}^2$ рівне $f_{ном} = N - k$.

Завдання до лабораторної роботи

Варіант індивідуального завдання для виконання лабораторної роботи (згідно порядкового номера студента в списку академічної групи) подано в дод. А.

Приклад виконання лабораторної роботи

Створюємо електронну таблицю (spreadsheet) для свого варіанту індивідуального завдання (дод. А). Для цього активуємо команду "Створити новий документ" (Create a new docement). В діалоговому вікні вказуємо три змінні та десять значень (рис. 6.1). Для даного прикладу (табл. 6.2) це:

– два фактори: x_1 – величина тиску, x_2 – діаметр перепускного отвору;

– критерій відгуку: y – час відкачування повітря.

Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів активуємо модуль "ANOVA" (Analysis of variance /

Таблиця 6.2

**Приклад індивідуального завдання
для виконання лабораторної роботи**

№	x_1	x_2	y
1	48	4	0,167
2	40	4	0,092
3	48	3	0,233
4	40	3	0,183
5	44	4	0,1
6	44	3	0,183
7	48	2	0,333
8	44	2	0,269
9	40	2	0,248
10	48	4	0,167

	1 x_1	2 x_2	3 y
1	48	4	0,167
2	40	4	0,092
3	48	3	0,233
4	40	3	0,183
5	44	4	0,1
6	44	3	0,183
7	48	2	0,333
8	44	2	0,269
9	40	2	0,248
10	48	4	0,167

Рис. 6.1. Таблиця даних

дисперсійний аналіз). Даний модуль призначений для проведення процедури як однофакторного, так і багатфакторного дисперсійного аналізу.

В діалоговому вікні (рис. 6.2) необхідно вказати наступні параметри:

– тип аналізу (type of analysis): "Однофакторний ANOVA" (One-way ANOVA);

– метод специфікації (specification method): "Швидке

діалогове вікон характеристик" (Quick specs dialog);
– натискаємо кнопку "Ok".

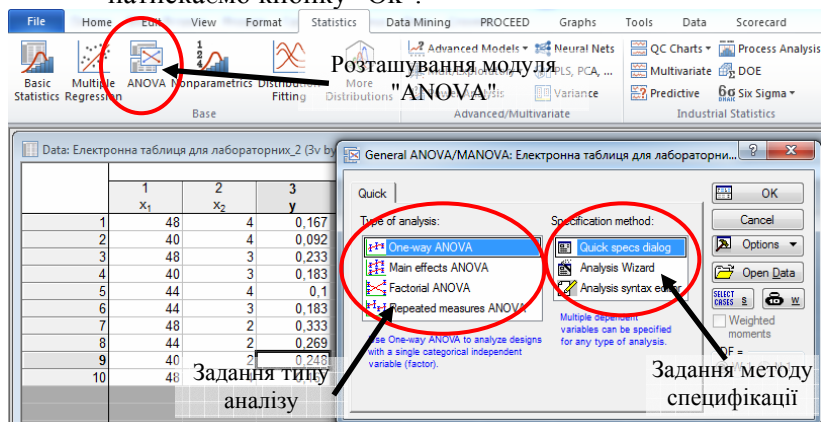


Рис. 6.2. Діалогове вікно модуля "ANOVA"

На екрані з'явиться діалогове вікно вибору змінних та опцій аналізу (рис. 6.3) в якому необхідно:

– задати змінні (variables): "залежні змінні" (dependent variables) – y , "категоріальні фактори" (categorical factors) – x_1 (рис. 6.4);

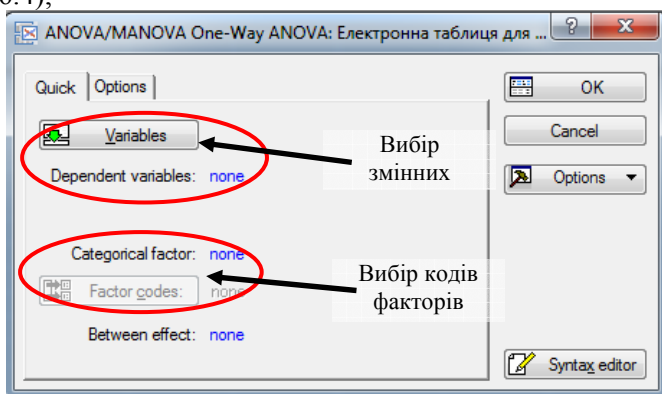


Рис. 6.3. Діалогове вікно вибору змінних та опцій аналізу

– натиснути "категоріальні фактори" (categorical factors): в діалоговому вікні (рис. 6.5) натиснути "All" для x_1 ;

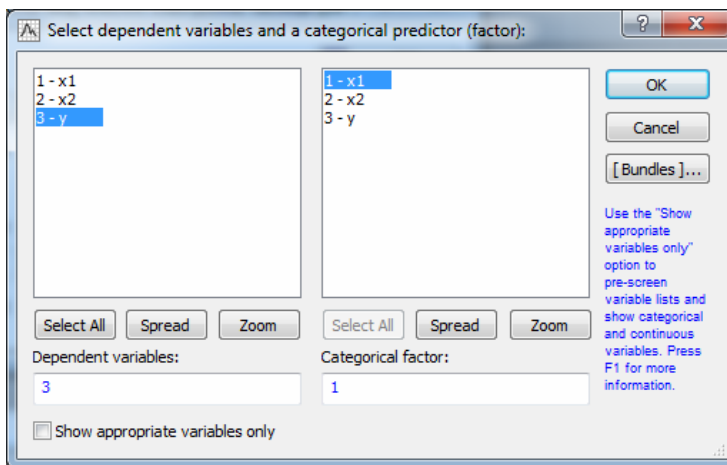


Рис. 6.4. Вибір змінних

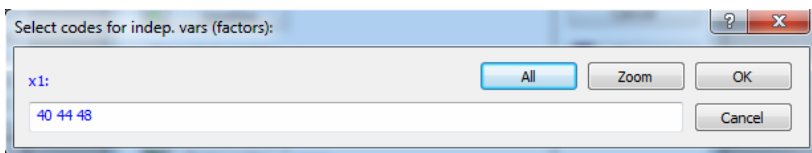


Рис. 6.5. Вибір кодів фактора

– натиснути кнопку "Ok".

На екрані з'явиться діалогове вікно "ANOVA Результати" (ANOVA Results). Рівень значимості (Significance level) залишаємо рівним 0,05 для розрахунку компонент варіації і перевірки гіпотези з використанням статистики Фішера (рис. 6.6).

Натиснувши "Усі ефекти" (All effects) – отримаємо таблицю (рис. 6.7) із наступними результатами:

– Система розрахувала компоненти "Сума квадратів" (SS) $Q_1^2=0,005182$ та $Q_2^2=0,045019$ загальної варіації Q^2 , які зумовлені фактором " x_1 (величина тиску)" та неврахованих факторів;

– Визначила число степенів свободи (Degr. of Freedom) кожної компоненти. Зокрема, для компоненти варіації, зумовленої факторною змінною, число степенів свободи

дорівнює $m-1=2$, а для компоненти, зумовленої неврахованими факторами, число степенів свободи дорівнює $m \cdot (n-1)=7$;

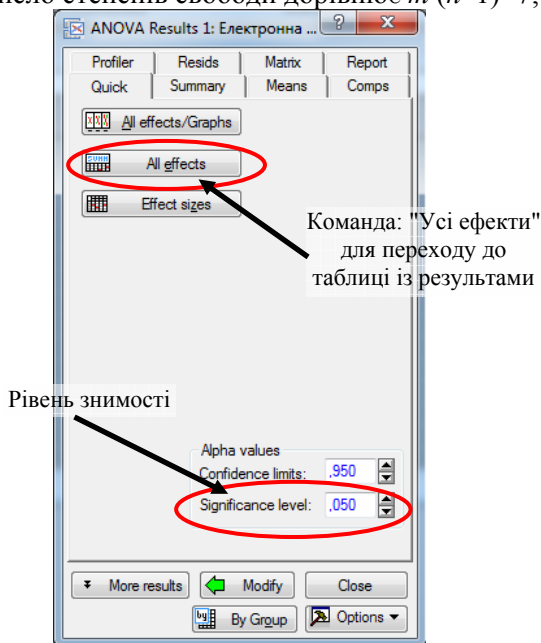


Рис. 6.6. Вікно результатів аналізу

Univariate Tests of Significance for y (Електронна таблиця для латентної Sigma-обмеженої параметризації) Effective hypothesis decomposition					
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	0,371212	1	0,371212	57,72016	0,000127
x ₁	0,005182	2	0,002591	0,40286	0,682961
Error	0,045019	7	0,006431		

Рис. 6.7. Вікно результатів однофакторного дисперсійного аналізу

– Розрахувала величини "Середній квадрат" (MS) $Q_1^2 / (m-1) = 0,002591$ та $Q_2^2 / m(n-1) = 0,006431$ відповідно.

– Розрахувала величину критерію Фішера, $F = \frac{Q_1^2 / (m-1)}{Q_2^2 / m(n-1)} = 0,40286$ та визначила імовірність $p=0,682961$, за

гіпотезі H_0 , що величина F може прийняти такі ж або більші значення. Оскільки ця імовірність істотно більша рівня значимості $\alpha=0,05$, то гіпотезу про значимість впливу фактора " x_1 (величина тиску)" на тривалість часу можна не приймати.

Наступний етап – багатофакторний дисперсійний аналіз.

В діалоговому вікні (рис. 6.8) необхідно вказати наступні параметри:

- тип аналізу (type of analysis): "Основні ефекти ANOVA" (Main effects ANOVA);
- метод специфікації (specification method): "Швидке діалогове вікно характеристик" (Quick specs dialog);
- натискаємо кнопку "Ok".

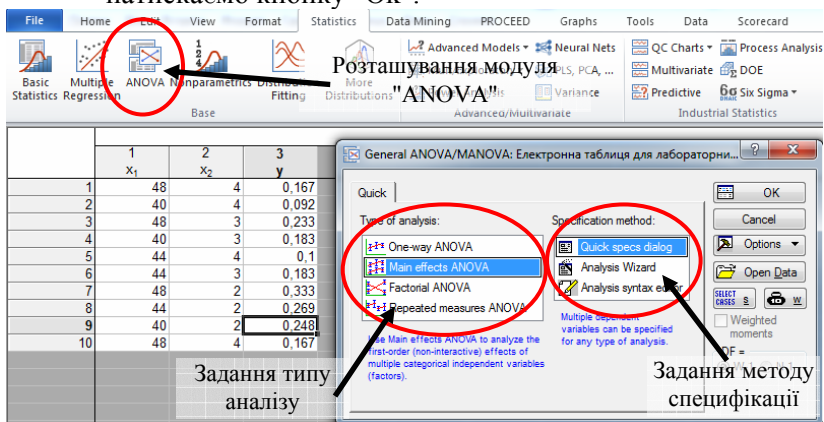


Рис. 6.8. Діалогове вікно модуля "ANOVA"

На екрані з'явиться діалогове вікно вибору змінних та опцій аналізу (рис. 6.9) в якому необхідно:

- задати змінні (variables): "залежні змінні" (dependent variables) – y , "категоріальні фактори" (categorical factors) – x_1 , x_2 (рис. 6.10);
- натиснути "категоріальні фактори" (categorical factors): в діалоговому вікні (рис. 6.11) натиснути "All" для x_1 та x_2 ;
- натиснути кнопку "Ok".

На екрані з'явиться діалогове вікно "ANOVA Результати" (ANOVA Results). Рівень значимості (Significance level) залишаємо рівним 0,05 для розрахунку компонент варіації і

перевірки гіпотези з використанням статистики Фішера (рис. 6.12).

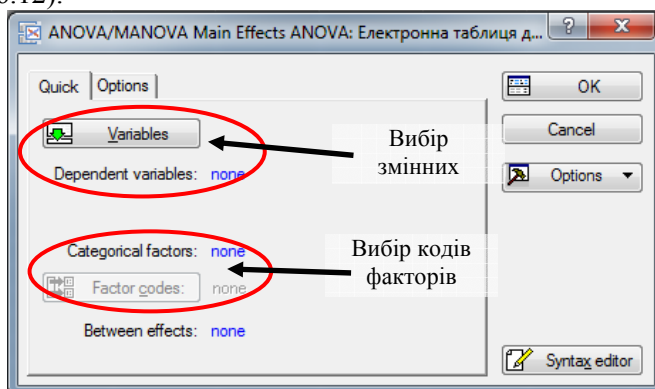


Рис. 6.9. Діалогове вікно вибору змінних та опцій аналізу

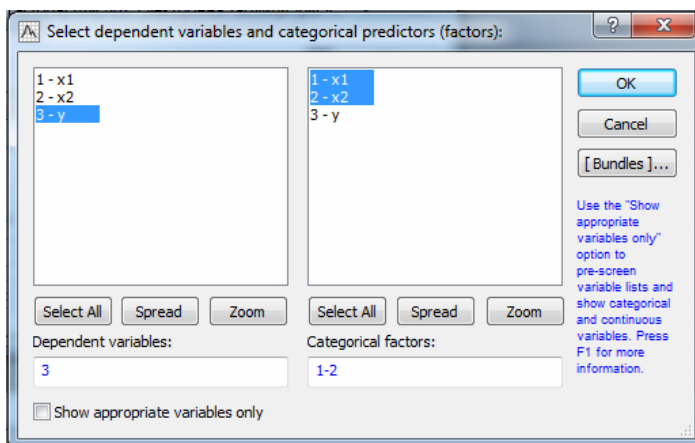


Рис. 6.10. Вибір змінних

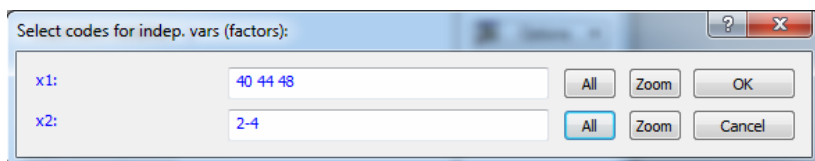


Рис. 6.11. Вибір кодів факторів

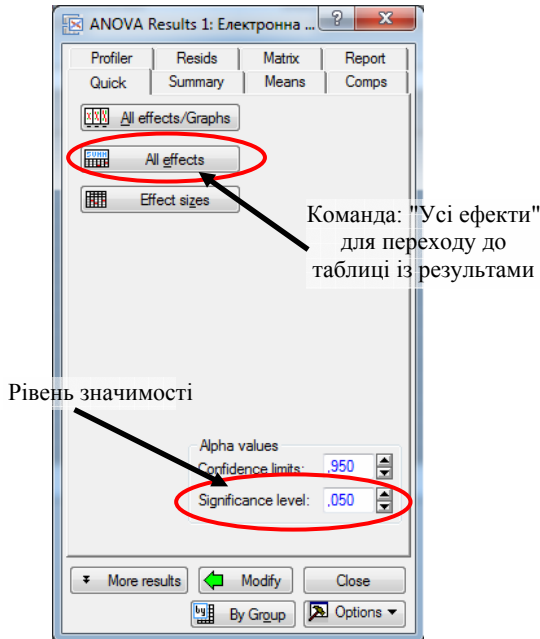


Рис. 6.12. Вікно результатів аналізу

Натиснувши "Усі ефекти" (All effects) – отримаємо таблицю (рис. 6.13) із наступними результатами:

– Система розрахувала компоненти "Сума квадратів" (SS) $Q_1^2=0,044662$ та $Q_2^2=0,000356$ загальної варіації Q^2 , які зумовлені фактором " x_1 (величина тиску)", " x_2 (діаметр перепускного отвору)" та неврахованих факторів;

Univariate Tests of Significance for (Електронна таблиця для л Sigma-restricted parameterization Effective hypothesis decomposition					
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	0,392229	1	0,392229	5506,250	0,000000
x_1	0,010304	2	0,005152	72,327	0,000204
x_2	0,044662	2	0,022331	313,494	0,000006
Error	0,000356	5	0,000071		

Рис. 6.13. Вікно результатів двофакторного дисперсійного аналізу

– Визначила число степенів свободи (Degr. of Freedom) кожної компоненти. Зокрема, для компоненти варіації, зумовленої фактором " x_1 (величина тиску)", число степенів свободи дорівнює $m-1=2$, для компоненти варіації, зумовленої фактором " x_2 (діаметр перепускного отвору)", число степенів свободи дорівнює $n-1=2$, а для компоненти, зумовленої неврахованими факторами, число степенів свободи дорівнює $= 5$;

– Розрахувала величини "Середній квадрат" (MS):

$$Q_1^2 / (m-1) = 0,005152 ,$$

$$Q_2^2 / m(n-1) = 0,022331 ,$$

$$Q_3^2 / ((m-1)(n-1)) = 0,000071$$

відповідно.

– Розрахувала величину критерію Фішера,

$$F_A = \frac{Q_1^2 / (m-1)}{Q_3^2 / ((m-1)(n-1))} = 72,327 , \quad F_B = \frac{Q_2^2 / (n-1)}{Q_3^2 / ((m-1)(n-1))} = 313,494 \quad \text{та}$$

визначила відповідні значенням F_A і F_B при імовірності $p_A=0,000204$ і $p_B=0,000006$ того, що при справедливості гіпотез H_0 величини F_A і F_B можуть прийняти такі ж або більші значення.

Оскільки імовірності p_A і p_B істотно менші рівня значимості $\alpha=0,05$, то гіпотезу про значимість впливу факторів " x_1 величина тиску" і " x_2 діаметр перепускного отвору" на "у тривалість часу" необхідно прийняти.

Варто пам'ятати, що на усіх етапах виконання лабораторної роботи отримані результати необхідно додавати до автозвіту.

Звіт про виконану лабораторну роботу має містити тему та завдання, отримані в процесі виконання таблиці та рисунки, а також висновок.

ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ

1. Які задачі розв'язує дисперсійний аналіз?
2. У чому відмінності між однофакторним та

двофакторним дисперсійним аналізом?

3. Дайте характеристику міжгрупової та внутрішньогрупової дисперсії.

4. Чим зумовлена варіація групових середніх навколо загального середнього?

5. Яка параметрична гіпотеза приймається за нульову під час дисперсійного аналізу?

6. Який порядок перевірки нульової гіпотези?

7. Що називають дисперсійним відношенням?

8. Який імовірнісний розподіл використовується для перевірки гіпотези в дисперсійному аналізі? Перерахуйте його числові характеристики.

9. Наведіть приклад однофакторного дисперсійного аналізу.

ЛАБОРАТОРНА РОБОТА №7

КОРЕЛЯЦІЙНО-РЕГРЕСІЙНИЙ АНАЛІЗ В СЕРЕДОВИЩІ STATISTICA

Мета роботи: Отримати практичні навички проведення кореляційно-регресійного аналізу отриманих результатів експериментальних досліджень.

Обладнання робочого місця: персональний комп'ютер з необхідним програмним забезпеченням, методичні рекомендації, довідкова література.

Загальні відомості

Методи виявлення і вимірювання зв'язків між фізичними величинами займають важливу частину методології наукових досліджень. В загальному розрізняють два типи зв'язків: **функціональний** та **статистичний**.

Функціональний зв'язок — коли із зміною значення фактора досліджувана величина змінюється чітко визначеним шляхом, тобто значенню одної змінної обов'язково відповідає одне або декілька точно заданих значень іншої змінної.

Статистичний зв'язок — із зміною значення фактора досліджувана величина може у певних межах приймати будь-які значення з деякою імовірністю, але її середнє значення чи інші статистичні характеристики змінюються за певним законом.

Кореляційним зв'язком називають частковий випадок **статистичного зв'язку**, тобто різним значенням однієї змінної відповідають різні середні значення іншої.

Із зміною значення фактора x закономірним чином

змінюється середнє значення досліджуваної величини \bar{y} , але в кожному окремому випадку значення досліджуваної величини y (з різними імовірностями) може приймати множину значень.

Оскільки кореляційний зв'язок є статистичним, першою умовою можливості його вивчення є наявність великої кількості інформації про досліджувану величину. Так число дослідів, достатнє для аналізу кореляційного зв'язку залежить від мети аналізу, необхідної точності і надійності параметрів зв'язку, від числа факторів.

Основним методом знаходження параметрів рівняння зв'язку є **метод найменших квадратів** (МНК). Він полягає в мінімізації суми квадратів відхилень фактично виміряних значень залежної змінної y від її значень вирахованих за рівнянням зв'язку (рівнянням регресії) з факторами x .

Завдання до лабораторної роботи

Варіант індивідуального завдання для виконання лабораторної роботи (згідно порядкового номера студента в списку академічної групи) подано в дод. А.

В даній лабораторній роботі необхідно:

1. Побудувати кореляційну матрицю та діаграми розсіювання.
2. Провести багатофакторний регресійний аналіз.

Приклад виконання лабораторної роботи

Створюємо електронну таблицю (spreadsheet) для свого варіанту індивідуального завдання (дод. А). Для цього активуємо команду "Створити новий документ" (Create a new docement). В діалоговому вікні вказуємо три змінні та десять значень (рис. 7.1). Для даного прикладу (табл. 7.1) це:

– два фактори: x_1 – величина тиску, x_2 – діаметр перепускного отвору;

– критерій відгуку: y – час відкачування повітря.

Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів активуємо модуль "Основні статистичні дані/таблиці" (Basic Statistics/Tables). В діалоговому вікні, що

з'явилося (рис. 7.2) обираємо команду "Кореляційні матриці" (Correlation matrices) та натискаємо "Ok".

Таблиця 7.1

**Приклад індивідуального завдання
для виконання лабораторної роботи**

№	x_1	x_2	y
1	48	4	0,167
2	40	4	0,092
3	48	3	0,233
4	40	3	0,183
5	44	4	0,1
6	44	3	0,183
7	48	2	0,333
8	44	2	0,269
9	40	2	0,248
10	48	4	0,167

	1	2	3
	x_1	x_2	y
1	48	4	0,167
2	40	4	0,092
3	48	3	0,233
4	40	3	0,183
5	44	4	0,1
6	44	3	0,183
7	48	2	0,333
8	44	2	0,269
9	40	2	0,248
10	48	4	0,167

Рис. 7.1. Таблиця даних

На екрані з'явиться наступне діалогове вікно "Продукт-момент і часткові кореляції" (Product-moment and partial correlations) в якому необхідно обрати параметр "Один список змінних" (One variable list) та вказати усі змінні (рис. 7.3) після чого натиснути "Швидкий аналіз" (Summary).

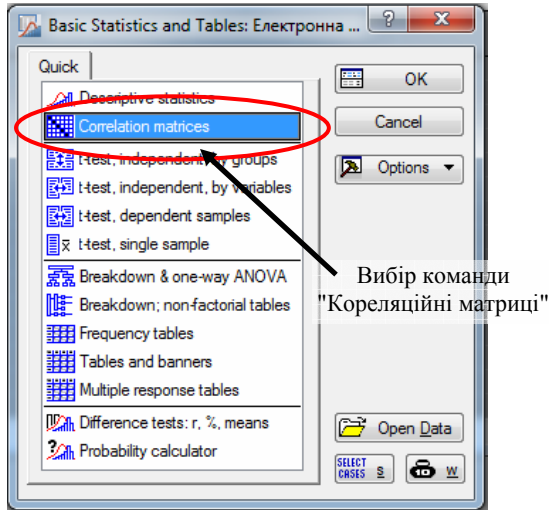


Рис. 7.2. Діалогове вікно модуля "Основні статистичні дані/таблиці"

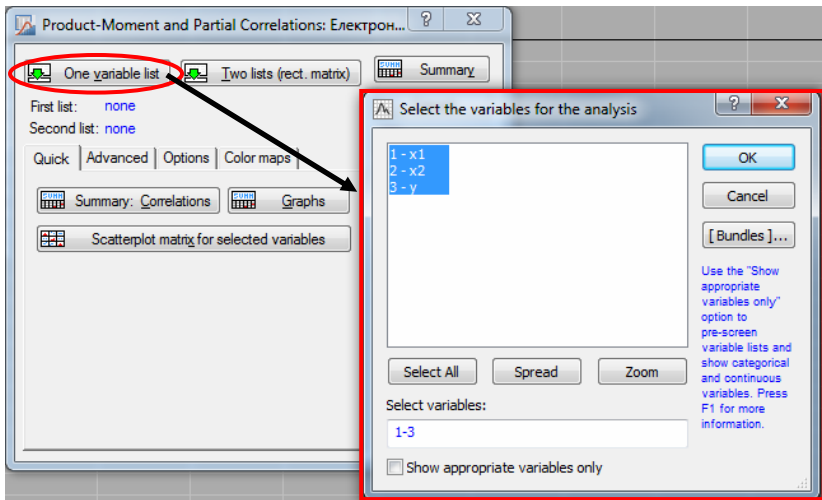


Рис. 7.3. Діалогове вікно модуля "Продукт-момент і часткові кореляції"

На екрані з'явиться кореляційна матриця (рис. 7.4). Для візуалізації отриманого значення кореляцій між змінними необхідно побудувати графік кореляції.

Correlations (Електронна таблиця для лабораторних_2) Marked correlations are significant at $p < .05000$ N=10 (Casewise deletion of missing data)					
Variable	Means	Std.Dev.	x ₁	x ₂	y
x ₁	44,40000	3,502380	1,000000	0,130435	0,304990
x ₂	3,10000	0,875595	0,130435	1,000000	-0,886086
y	0,19750	0,074685	0,304990	-0,886086	1,000000

Рис. 7.4. Кореляційна матриця

Для цього виконаємо наступну послідовність дій: натискаємо правою кнопкою мишки на відповідний коефіцієнт кореляції, на екрані з'явиться контекстне меню (рис. 7.5). В цьому меню послідовно вибираємо наступні пункти: "Графіки вихідних даних" (Graphs of Input Data) → "Діаграми розсіювання" (Scatterplot) → "Регресія, 95% підтв." (Regression, 95% conf.).

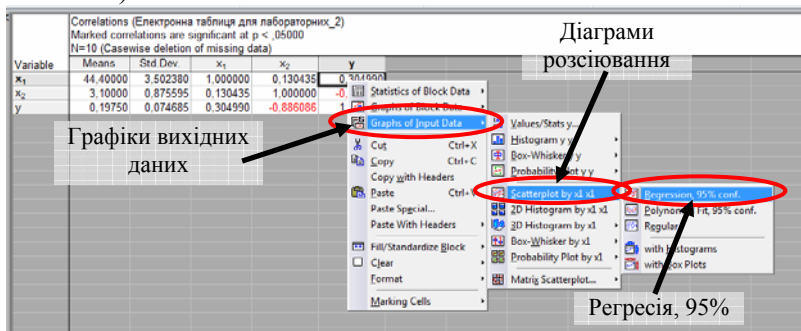


Рис. 7.5. Контекстне меню побудови діаграми розсіювання для коефіцієнта кореляції

В результаті ми отримаємо графік із параметрами, заданими за згодою (діаграма розсіювання для вибраного коефіцієнта з прямою регресією, 95%-ва довірна смуга та рівняння регресії в заголовку графіка (рис. 7.6).

Скористаємось модулем "Множинна регресія" (Multiple regression). Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів активуємо модуль "Множинна регресія" (Multiple regression). В діалоговому вікні, що з'явилося (рис. 7.7) у полі "змінні" (variables) вказуємо:

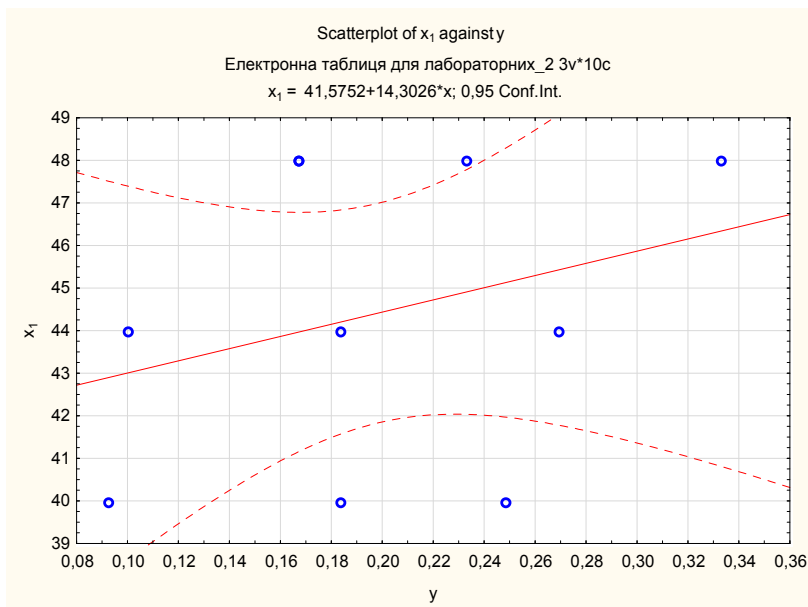


Рис. 7.6. Діаграма розсіювання коефіцієнта кореляції

- залежна змінна (dependent): y ;
- незалежна змінна (independent): x_1, x_2 .

Після вибору змінних натискаємо "Ok", на екрані з'явиться наступне діалогове вікно: "Результати множинної регресії" (Multiple Regression Results) (рис. 7.8). Обираємо команду "Підсумки: Результати регресії" (Summary: Regression results).

На екрані з'явиться таблиця (рис. 7.9) із результатами розрахунків:

- розрахунковим значенням критерію Фішера: $F_{\text{роз}} = 96,745$;
- степенями свободи критерію Фішера: $f_1 = 2, f_2 = 7$;
- степінь свободи критерію Стюдента: $f_n = 7$;
- розраховані коефіцієнти: $b_0 = 0,041474, b_1 = 0,009123, b_2 = -0,080340$.

Regression Summary for Dependent Variable: y (Електронна таблиця для лаб R= ,98238766 R²= ,96508552 Adjusted R²= ,95510995 F(2,7)=96,745 p<,00001 Std.Error of estimate: ,01582						
N=10	b*	Std.Err. of b*	b	Std.Err. of b	t(7)	p-value
Intercept			0,041474	0,067800	0,6117	0,560062
x ₁	0,427846	0,071233	0,009123	0,001519	6,0063	0,000539
x ₂	-0,941892	0,071233	-0,080340	0,006076	-13,2227	0,000003

Рис. 7.9. Результати розрахунку коефіцієнтів

Переходимо до останнього етапу роботи – оцінка адекватності отриманої моделі. Аналіз адекватності ґрунтується на основі аналізу залишків (різниця між дослідними значеннями та отриманими за допомогою моделі). Для цього знову використаємо модуль "Множинна регресія" (Multiple regression).

Перейшовши на вкладку "Статистика" (Statistics) панелі інструментів активуємо модуль "Множинна регресія" (Multiple regression). В діалоговому вікні, що з'явилося (рис. 7.7) у полі "змінні" (variables) вказуємо:

- залежна змінна (dependent): y;
- незалежна змінна (independent): x₁, x₂.

Після вибору змінних натискаємо "Ok", на екрані з'явиться наступне діалогове вікно: "Результати множинної регресії" (Multiple Regression Results).

Переходимо на вкладку "Залишки/припущення/прогноз" (Residuals/assumptions/prediction) де обираємо необхідну команду "Виконати аналіз залишків" (Perform residual analysis) (рис. 7.10).

В діалоговому вікні "Аналіз залишків" (Residual analysis) натискаємо на вкладку "Залишки" (Residuals) та активуємо команду "Залишки у порівняння із незалежною змінною" (Residuals vs. independent var) (рис. 7.11).

На екрані з'явиться графік (рис. 7.12) аналіз якого дозволяє стверджувати про достатню адекватність отриманої регресійної моделі.

Варто пам'ятати, що на усіх етапах виконання лабораторної роботи отримані результати (графіки, таблиці)

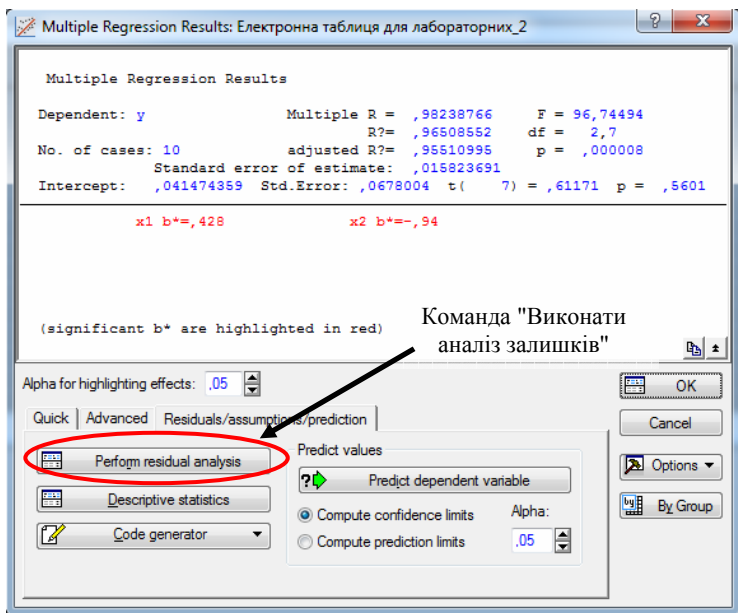


Рис. 7.10. Діалогове вікно "Результати множинної регресії"

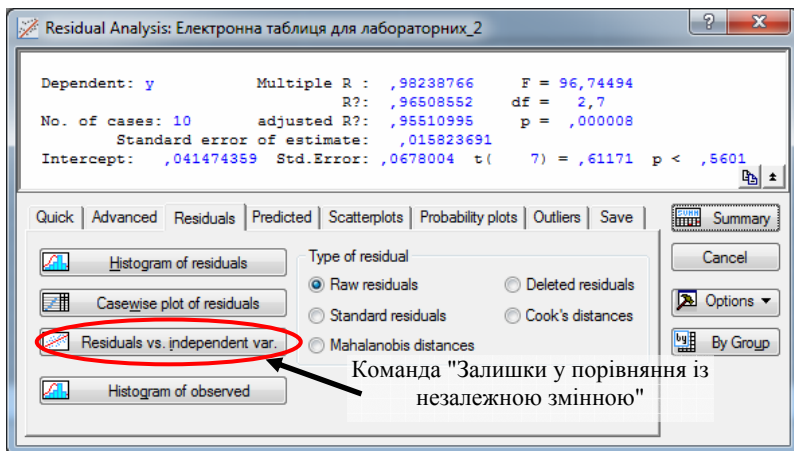


Рис. 7.11. Діалогове вікно "Аналіз залишків"

необхідно додавати до автозвіту.

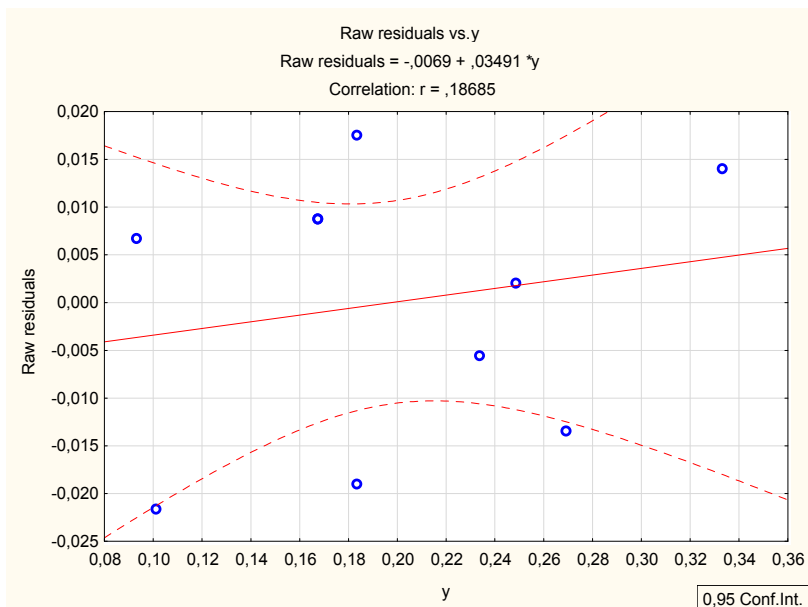


Рис. 7.12. Графік "дослідні змінні-залишки"

Звіт про виконану лабораторну роботу має містити тему та завдання, отримані в процесі виконання таблиці та рисунки, а також висновок.

ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ

1. Дайте визначення та поясніть відмінності між статистичним та функціональним зв'язком?
2. Що називається кореляційним зв'язком?
3. Перерахуйте причини виникнення кореляційного зв'язку між факторами?
4. Призначене модуля „Множинна регресія”?
5. Опишіть послідовність дій для побудови кореляційної матриці дослідних даних?
6. Як отримати діаграму розсіювання коефіцієнта кореляції? Про що вона говорить?
7. Як перевірити адекватність моделі в STATISTICA.

БІБЛІОГРАФІЧНИЙ СПИСОК

1. Дмитрів І.В. Автомобільний транспорт. Теорія і практика наукових досліджень. Навчальний посібник. Національний університет „Львівська політехніка”. – Львів : СПОЛОМ, 2019. 316 с.

2. Дмитрів І.В. Теорія та технологія наукових досліджень: механічна інженерія : навч. Посібн. Львівський національний аграрний університет. Львів : СПОЛОМ, 2017. 212 с.

3. Дмитрів В.Т. Механіко-технологічні основи систем доїльних установок. Теорія та експеримент : монографія. Львів : СПОЛОМ, 2017. 350 с.

4. Адамчук В.В., Дмитрів В.Т., Дмитрів І.В., Лаврик Ю.М. Адаптивний мікропульсатор автоматизованого доїльного апарата. Теорія та експеримент : монографія. Львів. Нац. Аграрн. Ун-т. Львів : СПОЛОМ, 2016. 152 с.

5. Дмитрів В.Т., Адамчук В.В., Лаврик Ю.М., Дмитрів І.В. Енергоощадний пневмоелектромагнітний пульсатор автоматизованого доїльного апарата. Теорія та експеримент : монографія. Львів. Нац. Аграрн. Ун-т. Львів : СПОЛОМ, 2016. 178 с.

6. Дорожовець М. Опрацювання результатів вимірювань: Навч. Посібник. Львів : Вид-во Національного університету „Львівська політехніка”, 2007. 624 с.

7. Дмитрів І. В. Багатофакторне моделювання відкачування повітря в системі “доїльний стакан-пульсатора”. *Вісник Львівського національного аграрного університету : агроінженерні дослідження*. 2014. № 18. С. 99-105.

8. Adamchuk V., Dmytriv V., Dmytriv I. Experimental studies of duration of air pumping out from the „TEAT CUP – PULSATOR” system. *Econtechmod: an International quarterly journal on economics in technology new technologies and modeling processes*. Lublin : Rzeszow, 2015. Vol. 4, № 4. P. 3-6.

9. Дмитрів В. Т., Дмитрів І. В. Результати експеримен-

тальних досліджень тривалості наповнення повітрям системи „доїльний стакан – пульсатор”. *Науковий вісник Національного університету біоресурсів і природокористування України: техніка та енергетика АПК*. К., 2015. Вип. 212/2. С. 142-148.

10. Адамчук В.В., Дмитрів І.В., Дмитрів В.Т. Результати експериментальних досліджень доїльного апарата з адаптивним пневмоелектромагнітним пульсатором. *Технічний сервіс агропромислового, лісового та транспортного комплексів [Науковий журнал]* . Харків, 2016. № 5. С. 238–245. URL : <http://www.khntusg.com.ua/node/1323>

11. Семенова І. Ю. Математичні моделі МСС : навч. посібн. К. : Київський нац. ун-тет імені Т. Шевченка, 2014. 82 с.

12. Дмитрів В. Т., Жінчин Я. С., Банга В. І. Експериментальні дослідження охолодника молока пластинчатого типу. *Вісник Харківського нац. техніч. університету с. г. ім. Петра Василенка*. Вип. 107 “Механізація сільськогосподарського виробництва”. Харків : ХНТУСГ, 2011. – Т. 2. С. 222-225.

13. Дмитрів В., Жінчин Я., Дмитрів Г., Гончаров В. Методологічні основи дослідження якості процесу змішування кормів. *Вісник Львівського національного аграрного університету : агроінженерні дослідження*. Львів : Львів. національний агроуніверситет, 2012. № 16. С. 55-62.

14. Дмитрів В. Т., Дмитрів І. В., Максимів І. І. Термоанемометричний вимірювач пульсуючих потоків двофазних середовищ. *Вісник Львівського національного аграрного університету: Агроінженерні дослідження*. Львів : Львів. національний агроуніверситет, 2013. № 17. С. 71-79.

15. Дмитрів В. Т., Дмитрів І. В. Імітатор інтенсивності молоковіддачі при експериментальних дослідженнях доїльних апаратів. *Вісник Львівського національного аграрного університету: Агроінженерні дослідження*. Львів : Львів. національний агроуніверситет, 2013. № 17. С. 124– 129.

16. Дмитрів В. Т., Банга В. І., Банга Ю. В. Методика калібрування вимірювача крутного моменту приводу обертових елементів. *Вісник Львівського національного аграрного університету: Агроінженерні дослідження*. Львів : Львів. національний агроуніверситет, 2013. № 17. С. 137-143.

17. Дмитрів В., Городняк Р. Контроль якості змішування на основі теорії розмірностей. *Управління якістю в освіті та промисловості: досвід, проблеми та перспективи: тези доповідей II Міжнародної науково-практичної конференції, 28-30 травня 2015 р.* Львів : В-во Львівської політехніки, 2015. С. 138-139.

18. Дмитрів В. Т., Дмитрів Г. М. Обґрунтування повторюваності вимірювань при експериментальних дослідженнях. *Праці Таврійської державної агротехнічної академії*. Вип. 26. Мелітополь : ТДАТА, 2005. С. 27-32.

19. Дмитрів В. Т. Тензометричні первинні перетворювачі для дослідження параметрів доїльного обладнання. *Агромех-2004: Матеріали Міжнародної науково-практичної конференції, 22-24 вересня 2004 р.* Львів : ЛДАУ, 2004. С. 110-118.

20. Дмитрів В., Жінчин Я., Дмитрів Г., Михайленко П., Остапенко М., Дріго В. Динамічні характеристики витратомірів стосовно АСК ТП машинного доїння. *Вісник Львівського державного аграрного університету: Агроінженерні дослідження*. 2005. № 9. С. 255-263.

21. Дмитрив И., Дмитрив В. Алгоритмы диагностики двигателей внутреннего сгорания. *MOTROL. Commission of Motorization and Energetics in Agriculture* . Lublin-Rzeszow, 2016. Vol.18, № 5. Р. 27-33. URL: http://www.pan-ol.lublin.pl/wydawnictwa/Motrol18_5/Motrol18_5_2016.pdf

22. Дмитрів В. Т., Ванько В. М., Лаврик Ю. М., Гонсьор О. Й., Банга В. І. Мікроконтролери і мікроконтролерні системи : лабораторний практикум. Львів : СПОЛОМ, 2017. 176 с.

23. Dmytriv V. T., Dmytriv I. V., Bubela T. Z., Yatsuk V. O.,

Lavryk Y. M., Krasnytsia B. Study of the pressure regulator work with a spring-damper system applied to milking machine / Дослідження роботи регулятора тиску з пружинно-демпферною системою стосовно до доїльних установок. *INMATEH - Agricultural Engineering. National institute of research-development for machines and installations designed to agriculture and food industry - INMA Bucharest*. vol. 52, no.2 / 2017. P. 61-68 eISSN: 2068-2239. URL : <https://www.scopus.com/record/display.uri?eid=2-s2.0-85028618607&origin=resultslist>.

24. Vasyl Dmytriv, Dmytriv Ihor, Yuriy Lavryk, Ivan Horodeckyy. Models of adaptation of the milking machines systems. *BIO Web Conf. Contemporary Research Trends in Agricultural Engineering. Volume 10, 2018. Kraków, Poland, September 25-27, 2017*. P. 1-7. eISSN: 2117-4458. DOI: <https://doi.org/10.1051/bioconf/20181002004>.

25. Dmytriv V., Dmytriv I., Dmytriv T. Research in thermoanemometric measuring device of pulse flow of two-phase medium. *17th International Scientific Conference Engineering for Rural Development. Proceedings. Volume 17, 2018. Jelgava, Latvia, May 23-25, 2018*. P. 898-904. ISSN 1691-5976. DOI: 10.22616/ERDev2018.17.N200.

26. Dmytriv V. T., Dmytriv I. V., Horodetskyi I.M., Yatsunskyi P. P. Adaptive cyber-physical system of the milk production process / Адаптивна кібер-фізична система процесу виробництва молока. *INMATEH - Agricultural Engineering. National institute of research-development for machines and installations designed to agriculture and food industry - INMA Bucharest*. vol. 61, no.2 / 2020. P. 199-208 eISSN: 2068-2239. DOI: 10.35633/inmateh-61-22/

27. Dmytriv V., Dmytriv I., Krasnytsia B. Influence of the mode of work of milking rubber on the dynamics of the stream of two-phase mixture. *VI International Scientific Congress / Agricultural machines. Research and testing. New machine designs. Volume 1, Burgas, Bulgaria, 25.06-28.06.2018*. P. 33-36. ISSN 2535-0269.

28. Дмитрів В. Т., Дмитрів І. В., Красниця Б. С. Коливання тиску в системах повітропроводу технологічного обладнання. *Вібрації в техніці та технологіях: збірник тез XVII-ої Міжнародної науково-технічної конференції*. Львів:

Національний університет „Львівська політехніка”, 2018. С. 47-48

29. Дрогомирецька Х. Т., Рибицька О. М., Слюсарчук О. З., Пабирівська Н. В., Гошко Л. В., Веселовська О. В., Білонова Д. В. Теорія ймовірностей та математична статистика. Львів : В-во Львівська політехніка, 2012. 396 с.

30. Багатофакторний дисперсійний аналіз. Методичні вказівки до виконання контрольної роботи з дисципліни “Теорія і практика наукових досліджень” для студентів спеціальності 274 «Автомобільний транспорт». Укл. Дмитрів І. В. Львів: 2018. 27 с.

31. Теорія і практика наукових досліджень. Частина 1. Методичні вказівки до виконання практичних робіт з дисципліни “Теорія і практика наукових досліджень” для студентів спеціальності 274 «Автомобільний транспорт». Укл. Дмитрів І. В. Львів: 2018. 70 с.

32. Теорія і практика наукових досліджень. Частина 2. Методичні вказівки до виконання практичних робіт з дисципліни “Теорія і практика наукових досліджень” для студентів спеціальності 274 «Автомобільний транспорт» другого (магістерського) рівня вищої освіти. Укл. Дмитрів І. В. Львів: 2018. 81 с.

33. Теорія і практика наукових досліджень. Електронний навчально-методичний комплекс «Теорія і практика наукових досліджень». Адреса розміщення: <http://vns.lpnu.ua/course/view.php?id=329>. 2019 р. 486 с.

ДОДАТКИ

ДОДАТОК А

Таблиця А.1

Варіанти індивідуальних завдань для виконання лабораторних робіт

Варіант 1									
$P, \text{кПа} (x_1)$	48	40	44	48	40	44	48	40	44
$d, \text{мм} (x_2)$	3	3	3	4	4	4	2	2	2
$t, \text{мс} (y_1)$	0,167	0,133	0,15	0,142	0,112	0,122	0,201	0,163	0,187
Варіант 2									
$P, \text{кПа} (x_1)$	48	40	44	48	40	44	48	40	44
$d, \text{мм} (x_2)$	3	3	3	4	4	4	2	2	2
$t, \text{мс} (y_1)$	0,15	0,127	0,156	0,126	0,102	0,131	0,192	0,159	0,177
Варіант 3									
$P, \text{кПа} (x_1)$	44	48	40	40	44	40	44	48	48
$d, \text{мм} (x_2)$	3	3	3	4	4	2	2	4	2
$t, \text{мс} (y_1)$	0,217	0,167	0,1	0,1	0,2	0,268	0,247	0,183	0,333
Варіант 4									
$P, \text{кПа} (x_1)$	44	40	40	44	48	40	44	48	48
$d, \text{мм} (x_2)$	2	3	2	3	4	4	4	2	3
$t, \text{мс} (y_1)$	0,254	0,092	0,264	0,233	0,183	0,117	0,183	0,333	0,15
Варіант 5									
$P, \text{кПа} (x_1)$	48	40	44	40	48	44	48	44	40
$d, \text{мм} (x_2)$	4	4	4	3	2	3	3	2	2
$t, \text{мс} (y_1)$	0,167	0,113	0,203	0,097	0,342	0,226	0,166	0,252	0,263
Варіант 6									
$P, \text{кПа} (x_1)$	46	42	38	42	46	38	46	38	42
$d, \text{мм} (x_2)$	5	5	5	4	2	4	4	2	2
$t, \text{мс} (y_1)$	0,177	0,123	0,213	0,107	0,352	0,236	0,176	0,262	0,273
Варіант 7									
$P, \text{кПа} (x_1)$	45	49	41	41	45	41	45	49	49
$d, \text{мм} (x_2)$	2	2	2	3	3	1	1	3	1
$t, \text{мс} (y_1)$	0,207	0,157	0,095	0,093	0,197	0,258	0,237	0,173	0,323
Варіант 8									
$P, \text{кПа} (x_1)$	48	40	44	48	40	44	48	40	44
$d, \text{мм} (x_2)$	3,5	3,5	3,5	4,5	4,5	4,5	2,5	2,5	2,5
$t, \text{мс} (y_1)$	0,155	0,132	0,161	0,131	0,107	0,136	0,197	0,164	0,182

продовження таблиці А.1

Варіант 9									
$P, \text{кПа} (x_1)$	47	39	43	47	39	43	47	39	43
$d, \text{мм} (x_2)$	3,4	3,4	3,4	4,4	4,4	4,4	2,4	2,4	2,4
$t, \text{мс} (y_1)$	0,164	0,130	0,147	0,139	0,109	0,119	0,198	0,160	0,184
Варіант 10									
$P, \text{кПа} (x_1)$	48	40	48	40	44	44	48	44	40
$d, \text{мм} (x_2)$	3,8	3,8	3,2	3,2	3,8	3,2	2,5	2,5	2,5
$t, \text{мс} (y_1)$	0,167	0,092	2,333	0,183	0,1	0,183	0,333	0,269	0,248
Варіант 11									
$P, \text{кПа} (x_1)$	48	40	48	40	44	44	48	44	40
$d, \text{мм} (x_2)$	3,8	3,8	3,2	3,2	3,8	3,2	2,5	2,5	2,5
$t, \text{мс} (y_1)$	0,165	0,1	0,217	0,181	0,103	0,201	0,328	0,267	0,247
Варіант 12									
$P, \text{кПа} (x_1)$	47	39	47	39	43	43	47	43	39
$d, \text{мм} (x_2)$	3,8	3,8	3,2	3,2	3,8	3,2	2,5	2,5	2,5
$t, \text{мс} (y_1)$	0,15	0,092	0,231	0,182	0,117	0,189	0,338	0,264	0,254
Варіант 13									
$P, \text{кПа} (x_1)$	48	40	48	40	44	44	48	44	40
$d, \text{мм} (x_2)$	3,9	3,9	3,1	3,1	3,9	3,1	2,7	2,7	2,7
$t, \text{мс} (y_1)$	0,166	0,091	0,225	0,161	0,112	0,198	0,351	0,261	0,252
Варіант 14									
$P, \text{кПа} (x_1)$	48	40	48	40	44	44	48	44	40
$d, \text{мм} (x_2)$	3,7	3,7	3,3	3,3	3,7	3,3	2,4	2,4	2,4
$t, \text{мс} (y_1)$	0,163	0,095	0,229	0,176	0,110	0,193	0,339	0,265	0,243
Варіант 15									
$\omega, \text{хв.}^{-1} (x_1)$	75,36	25,12	75,36	25,12	25,12	50,24	50,24	50,24	75,36
$\alpha, \text{град} (x_2)$	20	1	1	20	10	10	20	1	10
$Q, \text{кг/с} (y_1)$	0,056	0,051	0,095	0,037	0,046	0,067	0,041	0,066	0,061
Варіант 16									
$\omega, \text{хв.}^{-1} (x_1)$	75,36	25,12	75,36	25,12	25,12	50,24	50,24	50,24	75,36
$\alpha, \text{град} (x_2)$	20	1	1	20	10	10	20	1	10
$Q, \text{кг/с} (y_1)$	0,057	0,052	0,096	0,036	0,045	0,064	0,038	0,063	0,064
Варіант 17									
$\omega, \text{хв.}^{-1} (x_1)$	25,12	75,36	25,12	50,24	75,36	25,12	50,24	75,36	50,24
$\alpha, \text{град} (x_2)$	10	10	20	1	20	1	20	1	10
$Q, \text{кг/с} (y_1)$	0,031	0,077	0,047	0,056	0,050	0,041	0,039	0,053	0,026

Варіант 18									
$\omega, \text{хв.}^{-1}(x_1)$	25,12	75,36	25,12	50,24	75,36	25,12	50,24	75,36	50,24
$\alpha, \text{град}(x_2)$	10	10	20	1	20	1	20	1	10
$Q, \text{кг/с}(y_1)$	0,032	0,073	0,042	0,059	0,054	0,045	0,034	0,056	0,023
Варіант 19									
$\omega, \text{хв.}^{-1}(x_1)$	70,36	25,12	70,36	25,12	50,24	50,24	50,24	25,12	70,36
$\alpha, \text{град}(x_2)$	24	5	5	24	15	5	24	15	15
$Q, \text{кг/с}(y_1)$	0,059	0,051	0,095	0,034	0,047	0,062	0,037	0,065	0,062
Варіант 20									
$\omega, \text{хв.}^{-1}(x_1)$	70,36	25,12	70,36	25,12	50,24	50,24	50,24	25,12	70,36
$\alpha, \text{град}(x_2)$	24	5	5	24	15	5	24	15	15
$\gamma, \text{\%}(y_1)$	98,06	99,81	99,74	98,12	98,05	99,81	98,13	99,73	98,91
Варіант 21									
$\omega, \text{хв.}^{-1}(x_1)$	75,36	25,12	75,36	25,12	25,12	50,24	50,24	50,24	75,36
$\alpha, \text{град}(x_2)$	20	1	1	20	10	10	20	1	10
$\gamma, \text{\%}(y_1)$	98,16	99,61	99,54	98,32	98,25	99,61	98,33	99,53	98,71
Варіант 22									
$\omega, \text{хв.}^{-1}(x_1)$	25,12	50,24	75,36	25,12	50,24	25,12	75,36	50,24	75,36
$\alpha, \text{град}(x_2)$	1	20	20	1	10	1	10	20	1
$\gamma, \text{\%}(y_1)$	99,45	99,12	98,76	99,38	98,56	99,39	98,96	99,06	99,71
Варіант 23									
$\omega, \text{хв.}^{-1}(x_1)$	50,24	25,12	75,36	50,24	25,12	75,36	25,12	75,36	75,36
$\alpha, \text{град}(x_2)$	20	1	20	10	1	20	10	1	20
$\gamma, \text{кг/с}(y_1)$	99,89	99,61	98,56	98,72	99,46	99,75	98,32	99,62	98,39
Варіант 24									
$\omega, \text{хв.}^{-1}(x_1)$	75,36	50,24	25,12	75,36	50,24	25,12	75,36	50,24	75,36
$\alpha, \text{град}(x_2)$	20	1	10	20	1	20	10	1	10
$\gamma, \text{кг/с}(y_1)$	98,76	99,36	99,32	99,61	99,39	99,65	99,14	99,83	98,18
Варіант 25									
$\omega, \text{хв.}^{-1}(x_1)$	50,24	25,12	50,24	75,36	50,24	25,12	25,12	50,24	75,36
$\alpha, \text{град}(x_2)$	20	10	1	20	10	20	1	10	1
$\gamma, \text{кг/с}(y_1)$	99,56	98,23	99,87	98,45	99,93	98,76	99,06	99,61	99,78

ДОДАТОК Б

Таблиця Б.1

Табличні значення критерію Фішера (F - критерію)

f_2	f_1									
	1	2	3	4	5	6	8	12	24	∞
1	2	3	4	5	6	7	8	9	10	11
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,557	9,277	9,117	9,013	8,941	8,845	8,744	8,638	8,526
4	7,709	6,944	6,591	6,388	6,256	6,163	6,041	5,912	5,774	5,628
5	6,608	5,786	5,409	5,192	5,050	4,950	4,818	4,678	4,527	4,365
6	5,987	5,143	4,757	4,534	4,387	4,284	4,147	4,000	3,841	3,669
7	5,591	4,737	4,347	4,120	3,972	3,866	3,726	3,575	3,410	3,230
8	5,318	4,459	4,066	3,838	3,688	3,581	3,438	3,284	3,115	2,928
9	5,117	4,256	3,863	3,633	3,482	3,374	3,230	3,073	2,900	2,707
10	4,965	4,103	3,708	3,478	3,326	3,217	3,072	2,913	2,737	2,538
11	4,844	3,982	3,587	3,357	3,204	3,095	2,948	2,788	2,609	2,404
12	4,747	3,885	3,490	3,259	3,106	2,996	2,849	2,687	2,505	2,296
13	4,667	3,806	3,411	3,197	3,025	2,915	2,767	2,604	2,420	2,206
14	4,600	3,739	3,344	3,112	2,958	2,848	2,699	2,534	2,349	2,131
15	4,543	3,682	3,287	3,056	2,901	2,790	2,641	2,475	2,288	2,066
16	4,494	3,634	3,239	3,007	2,852	2,741	2,591	2,425	2,235	2,010
17	4,451	3,592	3,197	2,965	2,810	2,699	2,548	2,381	2,190	1,960
18	4,414	3,555	3,160	2,928	2,773	2,661	2,510	2,342	2,150	1,917
19	4,381	3,522	3,127	2,895	2,740	2,628	2,477	2,308	2,114	1,878
20	4,351	3,493	3,098	2,866	2,711	2,599	2,447	2,278	2,082	1,843
21	4,325	3,467	3,077	2,840	2,685	2,573	2,420	2,250	2,054	1,811
22	4,301	3,443	3,049	2,817	2,661	2,549	2,397	2,226	2,028	1,783
23	4,279	3,422	3,028	2,796	2,640	2,528	2,375	2,204	2,005	1,757
24	4,260	3,403	3,009	2,777	2,621	2,508	2,355	2,183	1,984	1,733
25	4,242	3,385	2,991	2,759	2,603	2,490	2,337	2,165	1,964	1,711
26	4,225	3,369	2,975	2,743	2,587	2,474	2,321	2,148	1,946	1,691
27	4,210	3,354	2,960	2,728	2,572	2,459	2,305	2,132	1,930	1,672
28	4,196	3,340	2,947	2,714	2,558	2,445	2,291	2,118	1,915	1,654
29	4,183	3,328	2,934	2,701	2,545	2,432	2,278	2,104	1,901	1,638
30	4,171	3,316	2,922	2,690	2,534	2,421	2,266	2,092	1,887	1,622
40	4,085	3,232	2,839	2,606	2,449	2,336	2,180	2,003	1,793	1,509

закінчення таблиці Б.1

60	4,001	3,150	2,758	2,525	2,368	2,254	2,097	1,917	1,700	1,389
120	3,920	3,072	2,680	2,447	2,290	2,175	2,016	1,834	1,608	1,254
∞	3,841	2,996	2,605	2,372	2,214	2,098	1,938	1,752	1,517	1,000

Таблиця Б.2

Табличні значення критерію Стьюдента (t - критерію)

Чис. ст. св. f	P				Чис. ст. св. f	P			
	0,9	0,95	0,99	0,999		0,9	0,95	0,99	0,999
1	6,314	12,760	63,657	636,619	18	1,734	2,101	2,878	3,922
2	2,920	4,303	9,925	31,598	19	1,729	2,093	2,861	3,883
3	2,353	3,182	5,841	12,994	20	1,725	2,086	2,845	3,850
4	2,132	2,776	4,604	8,610	21	1,721	2,080	2,831	3,819
5	2,015	2,571	4,032	6,859	22	1,717	2,074	2,819	3,792
6	1,943	2,447	3,707	5,959	23	1,714	2,069	2,807	3,767
7	1,895	2,365	3,499	5,405	24	1,711	2,064	2,797	3,745
8	1,860	2,306	3,355	5,041	25	1,708	2,060	2,787	3,725
9	1,833	2,262	3,250	4,781	26	1,706	2,056	2,779	3,707
10	1,812	2,228	3,169	4,587	27	1,703	2,052	2,761	3,690
11	1,796	2,201	3,106	4,437	28	1,701	2,048	2,763	3,674
12	1,782	2,179	3,055	4,318	29	1,699	2,042	2,756	3,659
13	1,771	2,160	3,012	4,221	30	1,697	2,042	2,750	3,646
14	1,716	2,145	2,977	4,140	40	1,684	2,021	2,704	3,551
15	1,753	2,131	2,947	4,073	60	1,671	2,000	2,660	3,460
16	1,746	2,120	2,921	4,015	120	1,658	1,980	2,617	3,373
17	1,740	2,110	2,898	3,965	∞	1,645	1,960	2,576	3,291

Таблиця Б.3

Табличні значення критерію Кохрена (G - критерію)

n	Число ст. св. f							
	1	2	3	4	5	6	7	8
2	0,999	0,975	0,939	0,906	0,858	0,853	0,833	0,816
3	0,967	0,871	0,798	0,746	0,707	0,677	0,653	0,633
4	0,907	0,768	0,684	0,629	0,590	0,560	0,537	0,518
5	0,841	0,684	0,598	0,544	0,506	0,478	0,456	0,439
6	0,781	0,616	0,532	0,480	0,445	0,418	0,398	0,382
7	0,727	0,561	0,480	0,431	0,391	0,373	0,356	0,338
8	0,680	0,516	0,438	0,391	0,360	0,336	0,319	0,304
9	0,640	0,478	0,403	0,358	0,329	0,307	0,290	0,277
10	0,602	0,445	0,373	0,331	0,303	0,282	0,267	0,254
12	0,541	0,392	0,326	0,288	0,262	0,244	0,230	0,219
15	0,471	0,335	0,276	0,242	0,220	0,203	0,191	0,182