

Міністерство освіти і науки України
Центральноукраїнський національний технічний університет
Механіко-технологічний факультет

Кафедра «Технологія машинобудування»



Методичні вказівки до виконання лабораторних робіт по курсу:

**«ТЕОРІЯ ПЛАНУВАННЯ ТА ВИКОНАННЯ
НАУКОВИХ ДОСЛІДЖЕНЬ»**

для студентів спеціальностей:
133 Галузеве машинобудування
131 Прикладна механіка

Кропивницький

2019

Міністерство освіти і науки України
Центральноукраїнський національний технічний університет
Механіко-технологічний факультет

Кафедра «Технологія машинобудування»



Методичні вказівки до виконання лабораторних робіт по курсу:

**«ТЕОРІЯ ПЛАНУВАННЯ ТА ВИКОНАННЯ
НАУКОВИХ ДОСЛІДЖЕНЬ»**

для студентів спеціальностей:
133 Галузеве машинобудування
131 Прикладна механіка

Затверджено на засіданні кафедри
«Технологія машинобудування»
Протокол № 8 від 26.02.2019 р.

Теорія планування та виконання наукових досліджень : метод. вказівки до викон. лаб. робіт студ. спец. 133 «Галузеве машинобудування», 131 «Прикладна механіка» / [уклад. О.І. Скібінський, В.М. Селєхова] ; М-во освіти і науки України, Центральнoукр. нац. техн. ун-т, каф. технології машинобудування. – Кропивницький : ЦНТУ, 2019. – 120 с.

Укладачі:

Скібінський О.І. – канд.техн.наук, доцент

Селєхова В.М.

Рецензент: Мажара В.А. – канд.техн.наук, доцент

Зміст

Вступ.....	4
Лабораторна робота №1 Інтерфейс програмного продукту StatGraphics. Первинна обробка результатів експерименту.....	5
Лабораторна а робота №2 Компонентний аналіз даних в програмному продукті StatGraphics.....	21
Лабораторна робота №3 Кластерний аналіз даних в програмному продукті StatGraphics.....	30
Лабораторна робота №4 Нелінійна регресія в програмному продукті StatGraphics.....	40
Лабораторна робота №5 Перевірка вірності статистичних гіпотез в програмному продукті StatGraphics.....	55
Лабораторна робота №6 Перевірка пов'язаних вибірок, початок кореляційного аналізу в програмному продукті StatGraphics.....	69
Лабораторна робота №7 Планування та перевірка експерименту в програмному продукті StatGraphics.....	82
Література.....	105
Додатки.....	106

Вступ

Курс «Теорія планування та виконання наукових досліджень» (ТПВНД) включає цикл лекцій, лабораторних та самостійних занять і спрямований на підвищення кваліфікації майбутнього фахівця.

Курс «ТПВНД» має загальну мету, яка передбачає підготовку студентів до самостійної роботи з планування, виконання та аналізу результатів наукових досліджень.

Курс «ТПВНД» складається з двох модулів.

Передбачено виконання семи лабораторних робіт. Виконана лабораторна робота захищається студентом у формі опитування.

Опитування з виконаної лабораторної роботи (ВЛР) складається з 5 запитань, які формуються з тих, що вказані у відповідній лабораторній роботі.

Критерії оцінювання відповіді в опитуванні з лабораторної роботи:

- Неправильна відповідь – 0 балів;
- Вірна відповідь – 1 бал.

Критерії оцінювання лабораторної роботи:

- Невиконана лабораторна робота – 0 балів;
- Виконана, але незахищена лабораторна робота – 3 бали;

Загальна сума балів (ЗСБ), яку студент може набрати протягом виконання лабораторних робіт з курсу – 56. Оцінки в балах по одній лабораторній роботі вказані в табл.1, по модулю та по курсу в цілому в табл.2.

Студент вважається атестованим з виконання лабораторної роботи, якщо він набрав кількість балів, що відповідає вказаній в табл.1 в стовбцях «відмінно», «добре» або «задовільно».

Студент вважається не атестованим з виконання лабораторної роботи, якщо він набрав кількість балів, що відповідає вказаній в табл.1 в стовбці «незадовільно».

Таблиця 1

Оцінювання знань при виконанні лабораторної роботи

Показники	Оцінка в балах					
	Зараховано					Не зараховано
	«5» Відмінно	«4» Добре		«3» Задовільно		«2» Незадовільно
	A	B	C	D	E	FX
ВЛР	8	7-6	5-4	3	2	1-0

Студент вважається атестованим з виконання лабораторних робіт з модуля та курсу в цілому, якщо він набрав кількість балів*, що відповідає вказаній в табл.2 в стовбцях «відмінно», «добре» або «задовільно». *Примітка. Кількість балів отриманих студентом при виконанні лабораторних робіт враховується тільки при умові виконання та захисту всіх лабораторних робіт передбачених до виконання.

Студент вважається не атестованим з виконання лабораторних робіт з модуля та курсу в цілому, якщо він набрав кількість балів, що відповідає вказаній в табл.2 в стовбці «незадовільно».

Таблиця 2

Оцінювання знань при виконанні лабораторних робіт з модуля та курсу

Показники	Оцінка в балах					
	Зараховано					Не зараховано
	«5» відмінно	«4» добре		«3» задовільно		«2» Незадовільно
	A	B	C	D	E	FX
ВЛР	56-52	51-42	41-38	37-31	30-22	21-0

До іспиту допускаються тільки ті студенти, які атестовані з виконання лабораторних робіт з курсу.

Лабораторна робота №1

Інтерфейс програмного продукту StatGraphics. Первинна обробка результатів експерименту

Мета роботи: навчитися працювати з інтерфейсом програми StatGraphics. Навчитися: будувати ряди частот, обчислювати вибіркові статистичні характеристики, будувати гістограми та полігони частот.

Теоретичні відомості

Програмний продукт StatGraphics (Statistical Graphics Systems) включає більше 250 статистичних і системних процедур, що застосовуються на виробництві, в бізнесі, економіці, маркетингу, медицині, біології, соціології, психології та в інших областях.

Інтерфейс програмного продукту StatGraphics представлено на рис. 1.1.

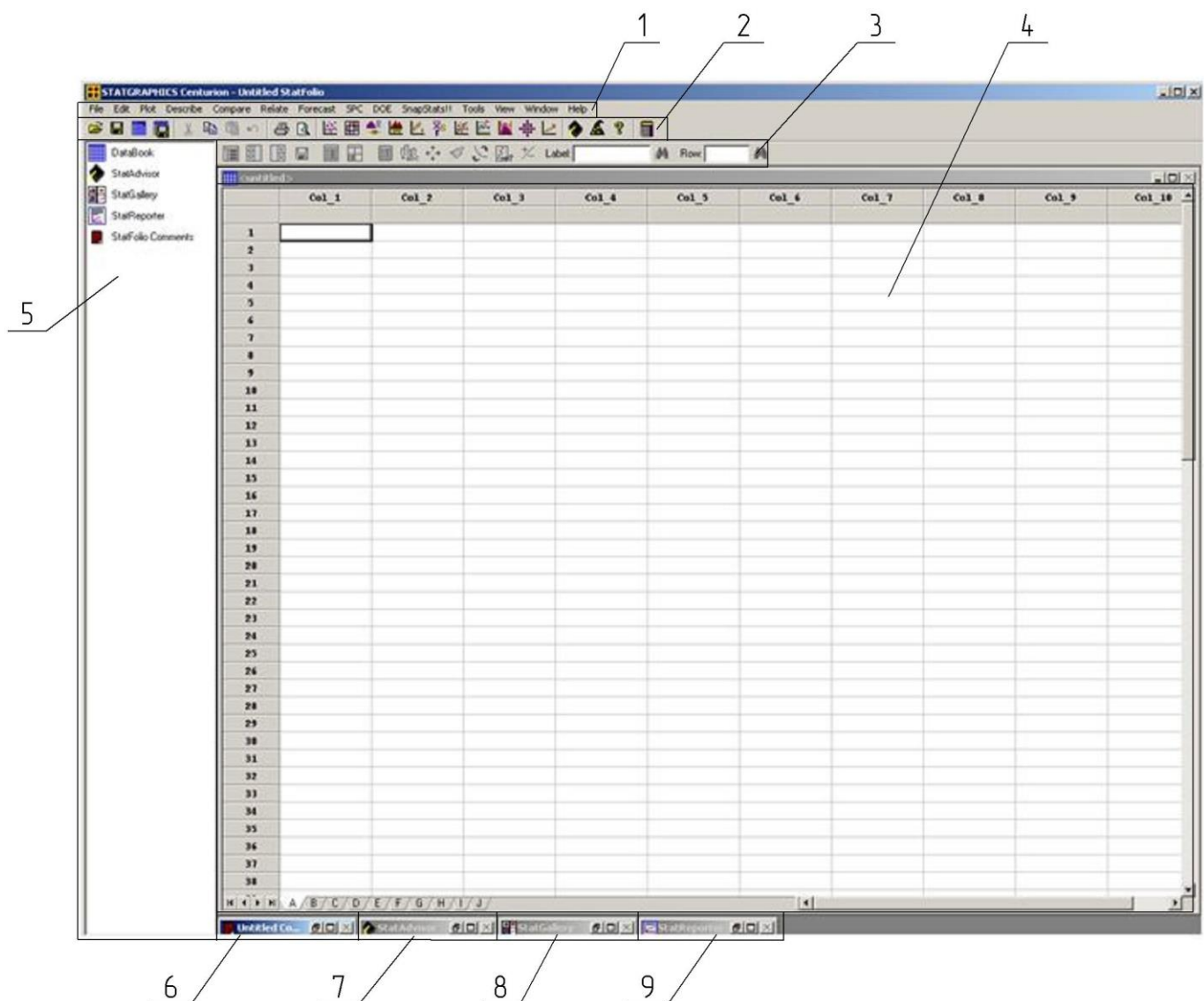


Рис. 1.1. Інтерфейс StatGraphics:

1 – строчка меню; 2 – панель інструментів; 3 – панель табличних та графічних опцій; 4 – робоче вікно документа; 5 – панель переключення вікон; 6 – вікно роботи з електронною таблицею; 7 – статистична експертна система StatAdvisor; 8 – вікно StatGallery; 9 – вікно введення коментарів

Строчка меню містить такі пункти:

- Describe (опис даних) - містить статистичні методи аналізу по одній та безлічі змінних, процедури підбору розподілів, засоби табуляції та крос-табуляції даних;
- Compare (порівняння даних) - включає методи порівняння двох і більше вибірок даних, процедури одно- і багатофакторного дисперсійного аналізу;
- Relate (відношення даних) - містить процедури простого, поліноміального та множинного регресійного аналізу.

Для розширення можливостей системи пропонуються додаткові пункти меню Special (спеціальні), до якого входять:

- контроль якості - оцінка ефективності всіх ланок виробничого процесу та формування відповідних контрольних карт;
- планування експерименту - формулювання критерію оптимальності плану експерименту, підбір найкращого плану, організація збору та обробки необхідної інформації. Пропонуються ефективні способи спрощення та інтеграції знань про досліджуваний процес. Процедура взаємодії з модулем наступна: визначення факторів; вибір плану; генерація робочої таблиці для збору та запису даних; підбір моделі; інтерпретація результатів. Дозволяє зменшити час дослідження, знизити загальні витрати та в цілому підвищити продуктивність;
- аналіз часових рядів - описові методи, процедури згладжування рядів, сезонної декомпозиції та прогнозування. Даний модуль допомагає побачити чисту картину динамічних даних. Доцільно почати роботу з описових методів, щоб отримати перше візуальне уявлення. Потім можна зробити більш точний опис динамічного ряду, враховуючи сезонні ефекти, циклічні зміни, тренди, помилки, викиди або точки зламу в даних. Результати представляються в табличній формі або на зручних для сприйняття графіках.

Панель інструментів включає такі команди: побудова X-Y графіків, побудова матриць, однофакторний аналіз, багатофакторний аналіз, регресія, автоматичний прогноз і т. і.

Панель табличних та графічних опцій містить кнопки для виклику:

- діалогових вікон;
- табличних та графічних опцій;
- збереження результатів статистичного аналізу;
- додаткових графічних опцій.

Панель задач розташована в нижній частині екрану та має такі вкладки:

- введення коментарів до проведеного статистичного аналізу (Untitled Comments);
- консультації у статистичній експертній системі (StatAdvisor);
- виклик вікна статистичної галереї (StatGallery);
- введення пояснень (StatReporter).

У вікні коментарів Untitled Comments зберігається інформація про хід статистичної обробки даних: про позначення змінних, методи обробки і т. і. При відкритті вікна коментарів

з'явиться запит на введення опису статистичного проекту (Enter a description of your StatFolio here). Для вводу коментарів українською мовою необхідно:

- вибрати Edit – Change text font;
- в діалоговому вікні Шрифт задати Набір символів «Кирилиця».

У вікні консультанта StatAdvisor відображається статистична інтерпретація отриманих результатів. Якщо результатом статистичного аналізу є таблиця даних, то її статистична інтерпретація буде наведена в цьому ж вікні безпосередньо під таблицею результатів. Якщо результатом аналізу є графік, то його статистична інтерпретація з'явиться в окремому вікні. Для виклику цього вінка потрібно натиснути вкладку StatAdvisor на панелі задач, або піктограму  з панелі табличних та графічних опцій.

Вікно статистичної галереї (StatGallery) використовується для створення текстових звітів, в яких можна легко об'єднати текст та графіку.

Генеральною сукупністю називають множину об'єктів, з яких виконується вибірка.

Вибірка – це частина генеральної сукупності, відібрана випадковим чином для проведення досліджень.

Мода – значення випадкової величини, що зустрічається найчастіше в сукупності спостережень. Іноді трапляється більше ніж одна мода. У такому випадку можна сказати, що сукупність мультимодальна. Як правило мультимодальність вказує на те, що набір даних не підпорядковується нормальному розподілу.

Медіана – це величина ознаки, що розташована посередині ранжованого ряду вибірки, тобто – це величина, що міститься в середині ряду величин, розташованих у зростаючому або спадному порядку.

Медіана ділить ряд значень ознаки на дві рівні частини, по обидві частини від неї розміщується однакова кількість одиниць сукупності.

Коефіцієнт асиметрії – безрозмірна величина, міра відхилення розподілу від симетричного (відносно математичного очікування).

Коефіцієнт ексцесу – безрозмірна характеристика гостровершинності графіка щільності ймовірності розподілу, яку застосовують в якості деякої міри відхилення розглядаємого розподілу від нормального.

Гістограма – спосіб графічного представлення табличних даних. Являє собою діаграму, що складається з прямокутників без розривів між ними. Кількісні співвідношення деякого показника представлені у вигляді прямокутників, площі яких пропорційні. Найчастіше для зручності сприйняття ширину прямокутників беруть однакою, при цьому їх висота визначає співвідношення відображуваного параметра.

Полігон частот – один із способів графічного представлення щільності вірогідності випадкових величин. Являє собою ламану, що з'єднує точки, які відповідають середнім значенням інтервалів групування та частотам цих інтервалів.

Порядок виконання роботи

1. Задання змінних та введення даних

Адміністрацію підприємства цікавить оптимальний рівень запасів комплектуючих на складі, а також середньомісячний обсяг покупок комплектуючих, які не підлягають заміні протягом тривалого часу експлуатації виробу. Для з'ясування цього питання менеджер підприємства протягом січня реєстрував частоту покупок відповідних комплектуючих і зібрав наступні дані (x_i):

4 4 9 3 3 1 7 0 4 2 3 5 7 10 6 5 7 3 2 9 8 1 4 2 1 0 8 9

1.1. В програмному продукті StatGraphics необхідно створити таблицю даних на основі прикладу (Додаток 1). Для задання назви змінної величини, подвійним клацанням ЛКМ на назві стовпчика, відкрийте вікно зміни назви ModifyColumn (модифікувати колонку). В строчку Name (Ім'я) даного вікна введіть назву стовпчика Data.

1.2. Скопіюйте дані з таблиці (Додаток 1).

	Data	Col_2	Col_3
1	4		
2	4		
3	9		
4	3		
5	3		
6	1		
7	7		
8	0		
9	4		
10	2		
11	3		
12	5		
13	7		
14	10		
15	6		
16	5		
17	7		

Рис. 1.2. Таблиця вхідних даних

1.3. Збережіть створену таблицю, вибравши File – Save – SaveDataFile.

2. Отримання статистичних оцінок

2.1. Виберіть пункт меню Describe/ Numeric Data/ One-Variable Analysis. З'явиться вікно діалогу для завдання аналізованих змінних(рис. 1.3).

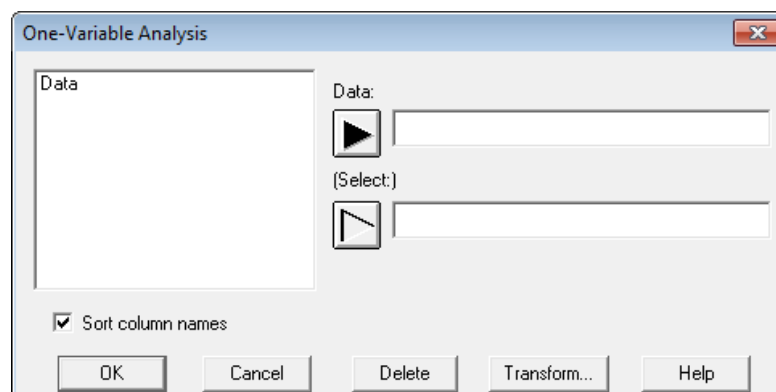


Рис. 1.3. Вікно вибору змінних

Виділіть змінну Data натисканням лівої кнопки миші, а потім клацніть по кнопці зі стрілкою, таким чином вводячи дану змінну в аналіз та натисніть ОК (рис. 1.4).

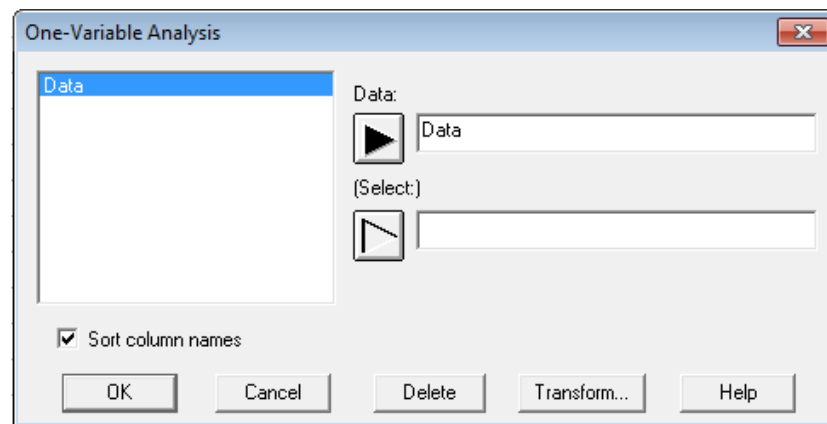


Рис. 1.4. Введення змінних в аналіз

Отримаємо вихідний звіт аналізу змінної (рис. 1.5).

One-Variable Analysis - Data

Data variable: Data
28 values ranging from 0 to 10,0

The StatAdvisor

This procedure is designed to summarize a single sample of data. It will calculate various statistics and graphs. Also included in the procedure are confidence intervals and hypothesis tests. Use the Tabular Options and Graphical Options buttons on the analysis toolbar to access these different procedures.

Рис. 1.5. Вихідний звіт аналізу множини змінних

У звіті аналізу множини змінних вказуються:

- назва змінної – Data;
- кількість значень та їх діапазон – 28 (від 0 до 10,0).

В нижній частині звіту наведена додаткова пояснювальна інформація.

2.2. Натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 1.6) встановіть прапорець таблиці загальної статистики (Summary statistics).

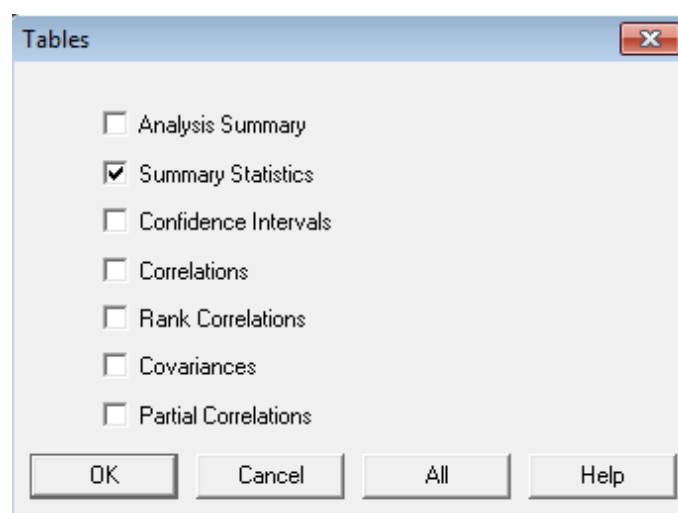


Рис. 1.6. Вікно вибору табличних опцій

За замовчуванням з'явиться наступний набір статистик (рис. 1.7).

Summary Statistics for Data	
Count	28
Average	4,53571
Standard deviation	2,99978
Coeff. of variation	66,1369%
Minimum	0,0
Maximum	10,0
Range	10,0
Std. skewness	0,56886
Std. kurtosis	-1,1974

Рис. 1.7. Звітна таблиця загальної статистики

Серед обчислених за замовчуванням числових характеристик вибірки є ті, які нас не цікавлять, і немає деяких із потрібних нам характеристик. Для вибору необхідних характеристик, клацніть ПКМ в робочому полі документа та виберіть команду Pane Options (параметри панелі). У вікно набору статистичних даних виберіть: Average (середнє), Median (медіана), Mode (мода), Variance (дисперсія), Standard Deviation (стандартне відхилення), Minimum (мінімальне значення вибірки), Maximum (максимальне значення), Range (розмах вибірки), Std.Skewness (стандартизована асиметрія), Std.Kurtosis (стандартизований експес), Sum (сума), Coeff.ofVariation (коефіцієнт варіації), Lower Quartile (нижній кuartиль), Upper Quartile (верхній кuartиль). Всі інші статистичні характеристики можна прибрати (рис. 1.8).

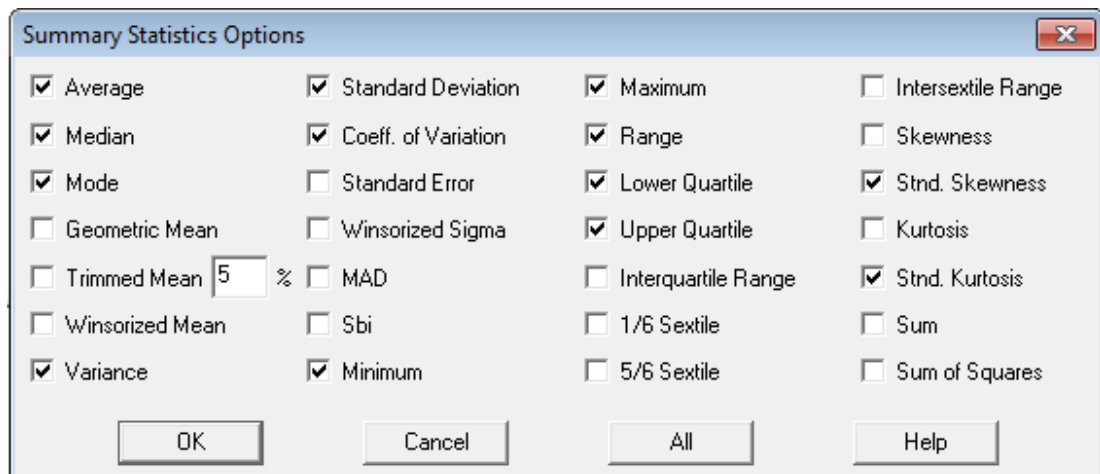


Рис. 1.8. Вікно набору статистичних даних

Summary Statistics for Data	
Count	28
Average	4,53571
Median	4,0
Mode	
Variance	8,99868
Standard deviation	2,99978
Coeff. of variation	66,1369%
Minimum	0,0
Maximum	10,0
Range	10,0
Lower quartile	2,0
Upper quartile	7,0
Std. skewness	0,56886
Std. kurtosis	-1,1974

Рис. 1.9. Звітна таблиця з додатковими статистичними даними

Розглянемо більш детально звітну таблицю. Count - кількість елементів у вибірці (28). Середнє значення рівне 4,5, медіана рівна 4,0. Оскільки, медіана не сильно відрізняється від середнього, це означає, що вибірка достатньо однорідна, немає різких викидів. Мода відсутня через те, що однаково часто зустрічаються декілька значень. В такому випадку, програмний продукт StatGraphics залишає пусту клітинку у звітній таблиці. Range – це різниця між максимальним і мінімальним значенням.

3. Побудова графіку Box-and-WiskerPlot

3.1. Натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 1.10) встановіть прапорець графіку Box-and-WiskerPlot. Отримаємо наступне зображення (рис. 1.11).

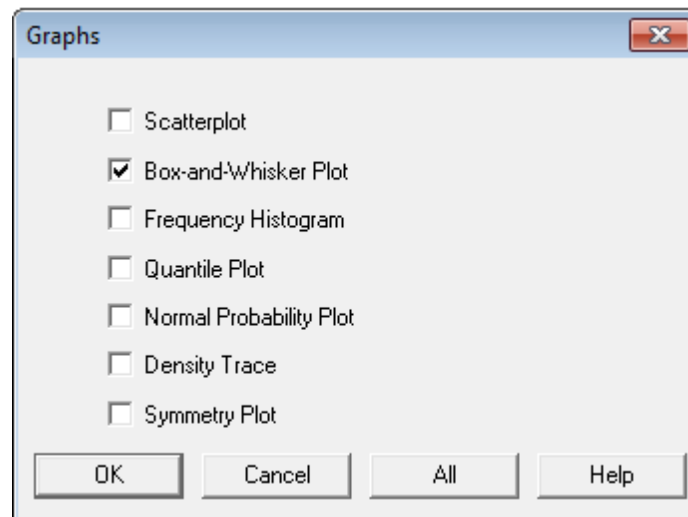


Рис. 1.10. Вікно вибору графічних опцій

Box-and-Whisker Plot

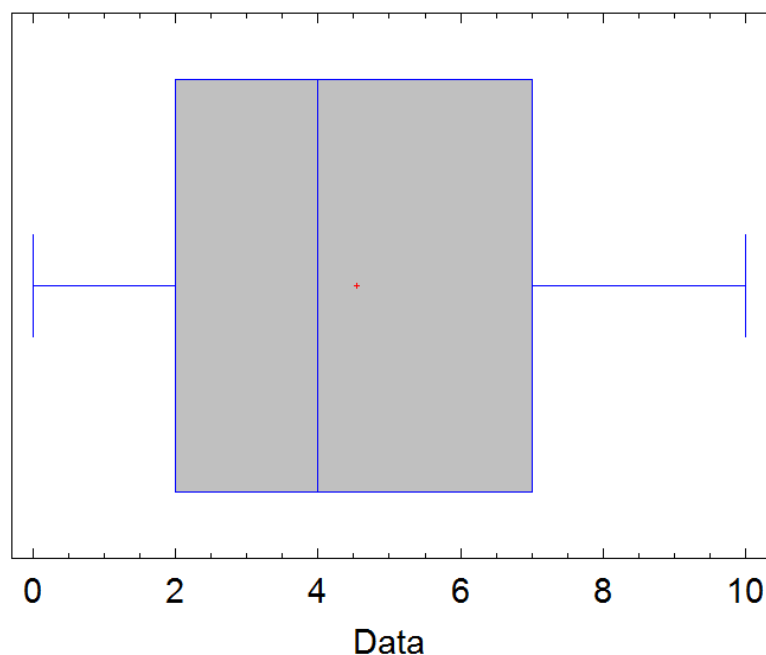


Рис. 1.11. Графік Box-and-WiskerPlot

Назва графіка Box-and-WiskerPlot в перекладі зустрічається, як скринькова діаграма, коробчата діаграма і, навіть, ящик з вусами. Цей прямокутник представляє простір між першим і третім кuartілями, тобто від 25% до 75% значень варіаційного ряду потрапляють в нього. Лінія всередині прямокутника відповідає медіані. Плюс в скриньці позначає вибіркове середнє. У нашому випадку вибіркове середнє не лежить на медіані. Відрізки («вуса») відповідають мінімальному та максимальному значенню.

4. Зміна елементів вибірки

4.1. В лівій частині вікна програми є панель переключення вікон (рис. 1.12). Переключіться на основне вікно DataBook. Збільшіть перший елемент вибірки (4) у чотири рази.



Рис. 1.12. Панель переключення вікон

4.2. Переключіться на вікно One-VariableAnalysis – Data, в якому знаходиться аналіз даних. Подивіться, як змінилися значення середнього, моди та медіани (рис. 1.13).

Summary Statistics for Data

Count	28
Average	4,96429
Median	4,0
Mode	3,0
Variance	13,6653
Standard deviation	3,69667
Coeff. of variation	74,4652%
Minimum	0,0
Maximum	16,0
Range	16,0
Lower quartile	2,0
Upper quartile	7,5
Std. skewness	2,06135
Std. kurtosis	1,38167

Рис. 1.13. Звітна таблиця після зміни елемента вибірки

4.3. Переключіться на вікно графіка Box-and-Wisker Plot. Проаналізуйте зміни на графіку (рис. 1.14).

Box-and-Whisker Plot

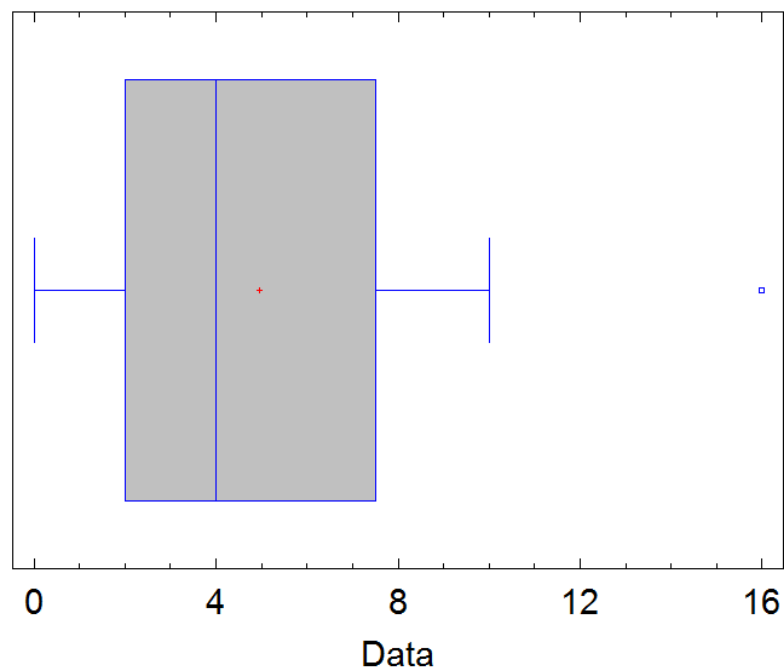


Рис. 1.14. Графік Box-and-Wisker Plot

В правій частині графіка з'явилася точка. Якщо на ній затиснути ЛКМ, то з'явиться інформація: ряд 1, X 16 (рис. 1.15). Це номер ряду, в якому був викид. Точка стає викидом, якщо вона знаходиться на відстань, яка більше, ніж в півтора рази перевищує ширину коробки.

Зверніть увагу, вибіркове середнє ще далі відійшло від медіани, коробка стала ще менш симетричною.

Row: 1
X: 16,0

Рис. 1.15. Номер ряду викиду

4.4. Знову поверніться в основне вікно DataBook. Встановіть значення першого елемента вибірки рівним 40. Переключіться на вікно графіка Box-and-WhiskerPlot (рис. 1.16).

Box-and-Whisker Plot

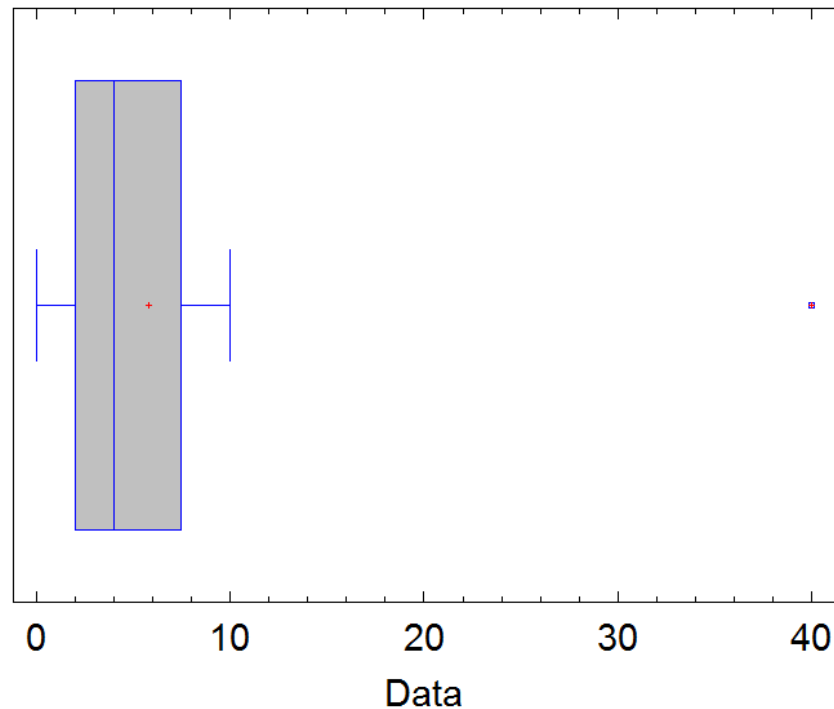


Рис. 1.16. Графік Box-and-Wisker Plot

Зверніть увагу, коробка стала ще більш несиметричною, а всередині викиду з'явився червоний хрестик. Це позначення так званого «далекого» викиду (або екстремального викиду). Так позначаються точки, якщо вони знаходяться на відстані, яка більше, ніж в 3 рази перевищує ширину коробки.

Графічне представлення Box-and-WiskerPlot використовується для початкового аналізу даних, наприклад, щоб перевірити, чи не було при введенні даних помилок: всі екстремальні викиди треба перевірити.

5. Побудова інтервального ряду частот

5.1. Якщо даних багато, зазвичай їх групують, тобто будують інтервальный ряд частот.

Натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 1.17) встановіть прапорець таблиці частот (Frequency Tabulation).

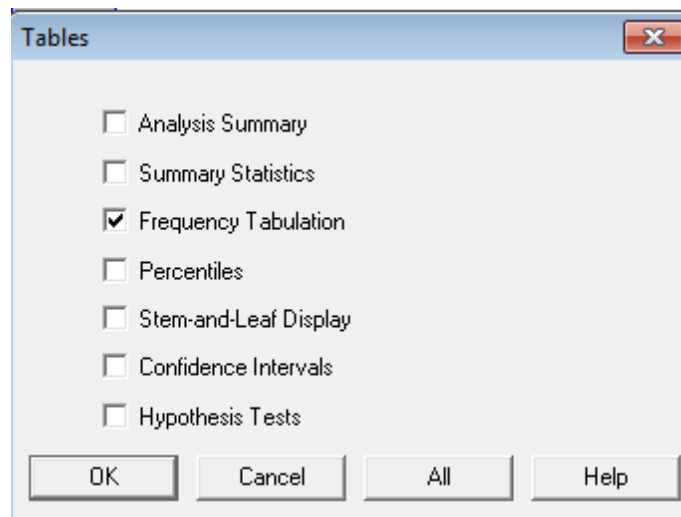


Рис. 1.17. Вікно вибору табличних опцій

На екрані з'явиться таблиця частот для даних (рис. 1.18).

Frequency Tabulation for Data

	<i>Lower</i>	<i>Upper</i>			<i>Relative</i>	<i>Cumulative</i>	<i>Cum. Rel.</i>
<i>Class</i>	<i>Limit</i>	<i>Limit</i>	<i>Midpoint</i>	<i>Frequency</i>	<i>Frequency</i>	<i>Frequency</i>	<i>Frequency</i>
	at or below	-1,0		0	0,0000	0	0,0000
1	-1,0	1,0	0,0	5	0,1786	5	0,1786
2	1,0	3,0	2,0	7	0,2500	12	0,4286
3	3,0	5,0	4,0	6	0,2143	18	0,6429
4	5,0	7,0	6,0	4	0,1429	22	0,7857
5	7,0	9,0	8,0	5	0,1786	27	0,9643
6	9,0	11,0	10,0	1	0,0357	28	1,0000
	above	11,0		0	0,0000	28	1,0000

Mean = 4,53571 Standard deviation = 2,99978

Рис. 1.18. Таблиця частот для даних

Програмний продукт StatGraphics розбиває дані на 6 груп. У таблиці наводяться дані: номер групи (Class), нижня межа (Lower Limit), верхня межа (Upper Limit), середня точка (Midpoint), частота (Frequency) - кількість елементів вибірки, що потрапили в даний інтервал), відносна частота (Relative Frequency) - частота, поділена на кількість елементів у вибір, Cumulative Frequency - накопичена частота, Cum.Rel. Frequency - накопичена відносна частота.

Під таблицею частот вказується середнє арифметичне (Mean) та стандартне відхилення (Standard deviation).

Кількість груп розбиття можна змінювати. В робочому полі документа клацніть ПКМ та у відкритшомуся меню виберіть Pane Options, відкриється вікно (рис. 1.19).

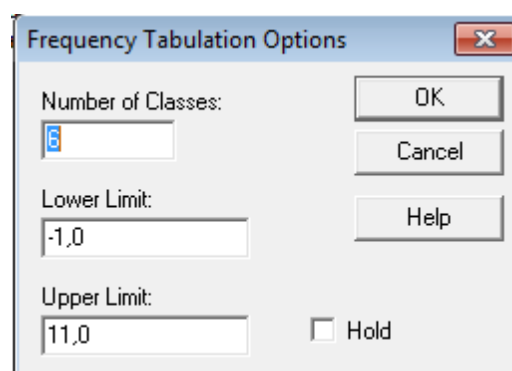


Рис. 1.19. Вікно опцій таблиці частот

В цьому вікні можна змінювати значення нижньої та верхньої меж, а також кількість груп. Спробуйте зробити це. Потім поверніть вихідні дані. Справа в тому, що кількість груп залежить від обсягу вибірки, збільшення чи зменшення їх кількості не принесе, як правило, поліпшення результатів.

6. Побудова гістограми частот

6.1. Натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 1.20) встановіть прапорець гістограми частот (Frequency Histogram). Отримаємо наступне зображення (рис. 1.21).

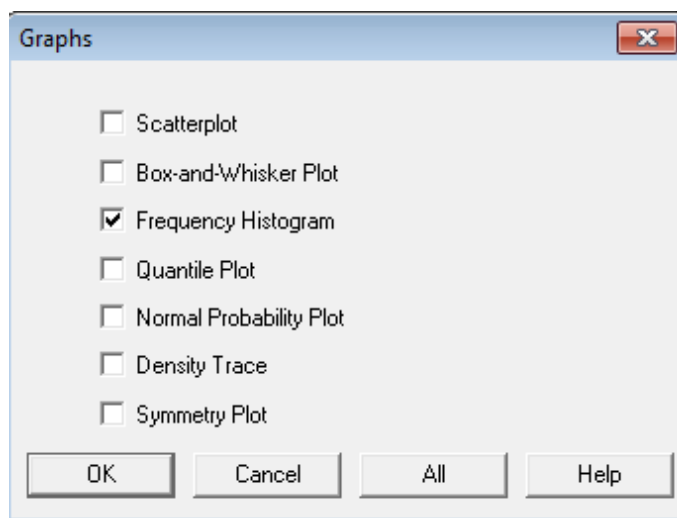


Рис. 1.20. Вікно вибору графічних опцій

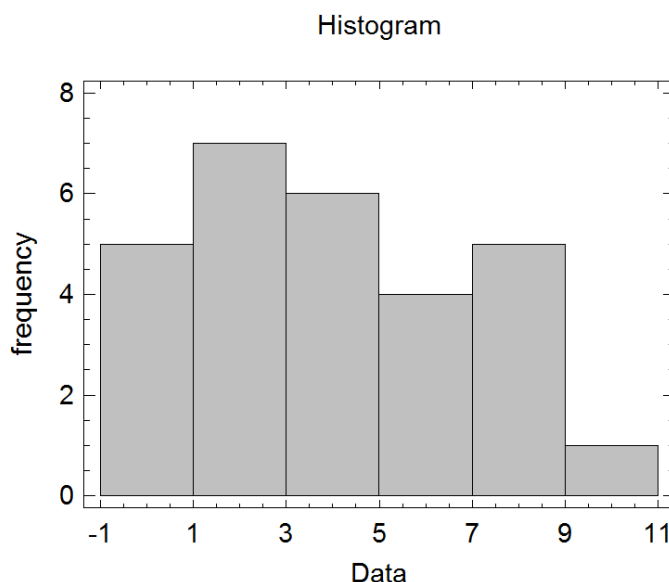


Рис. 1.21. Гістограма частот

6.2. Для внесення змін до вигляду гістограми, виберіть пункт меню Graphics Options, натиснувши ПКМ в робочому полі документа. Відкриється вікно, в якому в першій вкладці Layout (рис. 1.22) можна вибрати колір заповнення фону.

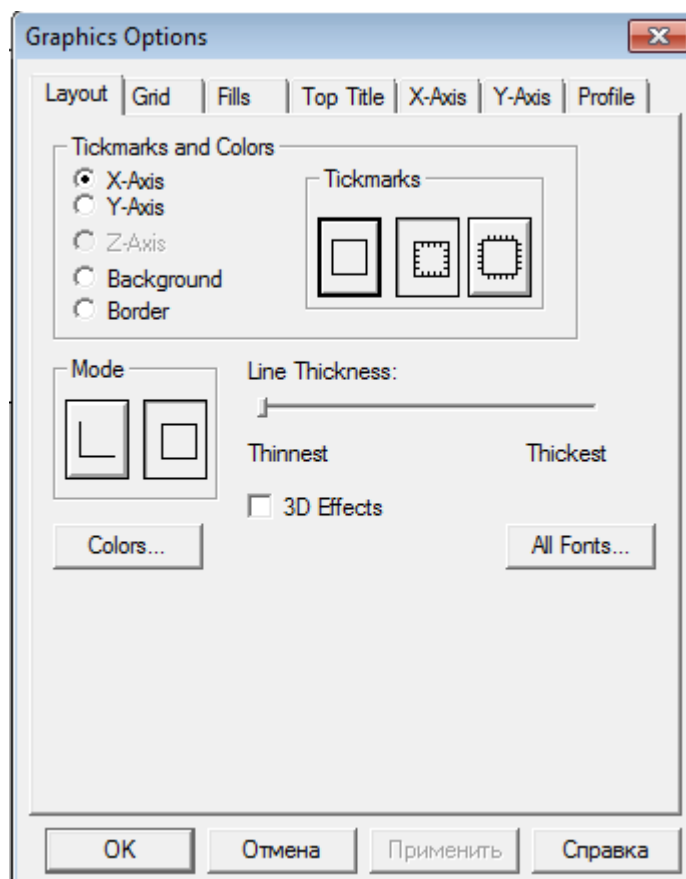


Рис. 1.22. Вікно редагування гістограми

Виберіть Background, натисніть на кнопку Colors, ввікні, що розкрилося виберіть потрібний колір, натисніть кнопку ОК. У вкладці Fills можна попрацювати з видом гістограми. На вкладці TopTitle можна перейменувати гістограму. Також на відповідних вкладках можна попрацювати з осями координат. У вкладці Profile можна налаштовувати загальний вигляд гістограми.

7. Побудова полігону частот

7.1. Необхідно перетворити гістограму в полігон. Для цього у вікні з гістограмою клацніть ПКМ, виберіть пункт меню Pane Options. Відкриється діалогове вікно (рис. 1.23).

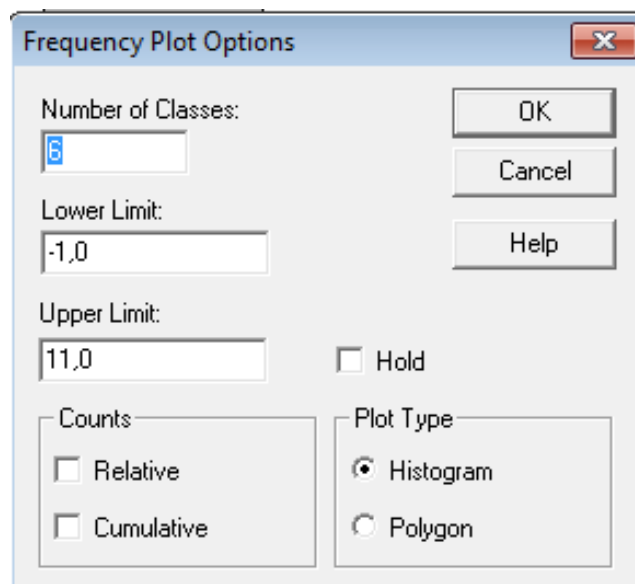


Рис. 1.23. Налаштування графіку частот

У цьому вікні виберіть Polygon і клацніть ОК. На екрані з'явиться полігон (рис. 1.24).

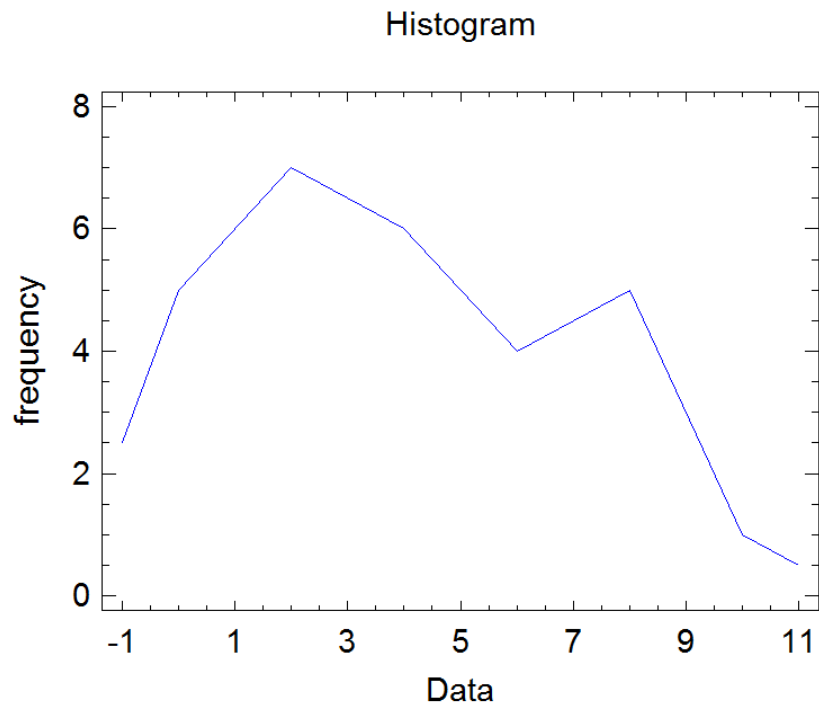


Рис. 1.24. Полігон

7.2. Для редагування полігону(зміни товщини і кольору ліній), виділіть лінію, клацнувши по ній ЛКМ, а потім клацніть по ній ПКМ і виберіть в контекстному меню Graphics Options. З'явиться наступне діалогове вікно (рис. 1.25).

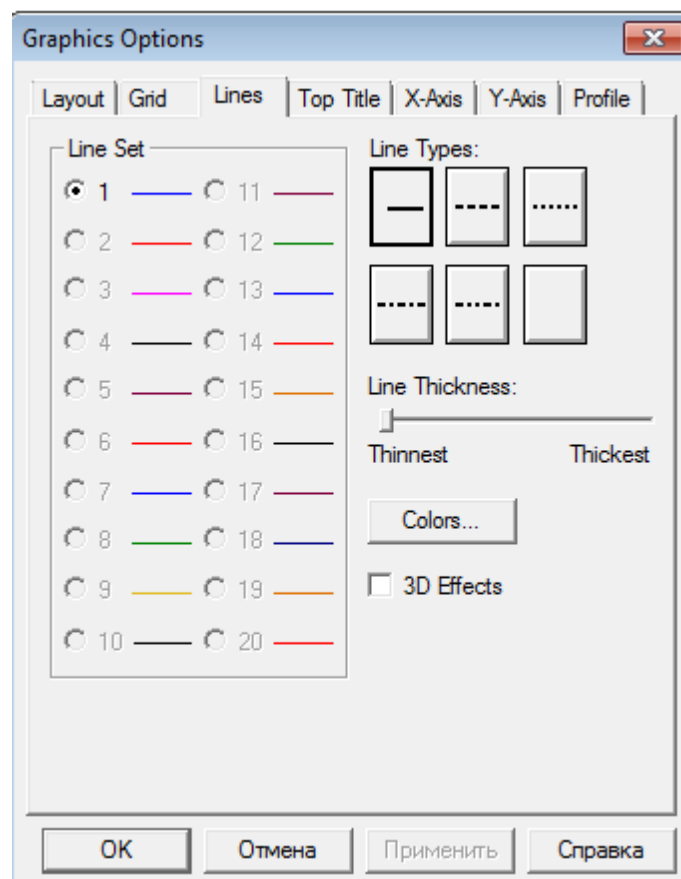


Рис. 1.25. Вікно редагування полігону

В даному діалоговому вікні можна задавати параметри: тип лінії, її колір, товщину, наявність координатної сітки та ін. Один із варіантів редагування полігону представлено на рис. 1.26.

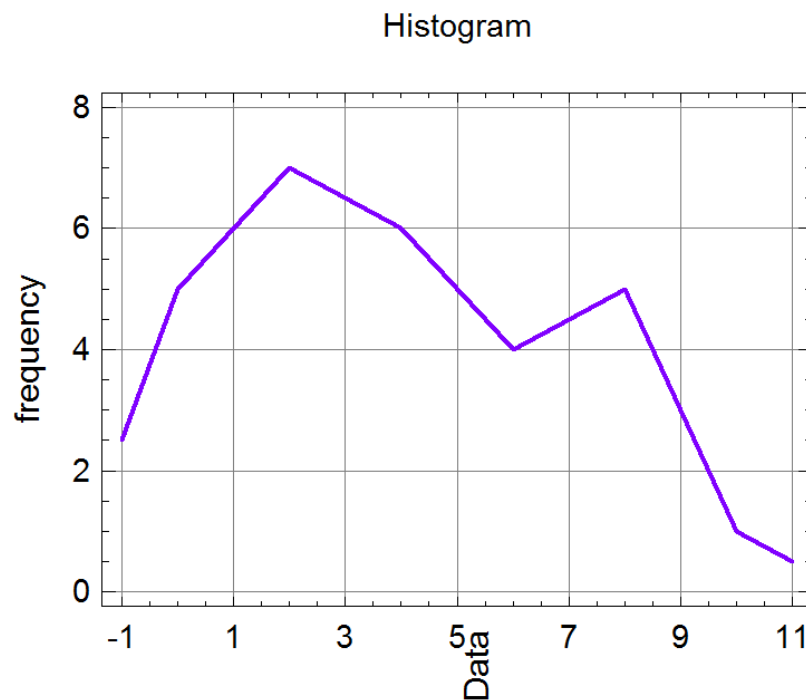


Рис. 1.26. Відредагований полігон

До речі, в тому ж вікні, де полігон і гістограма, можна вибрати кумулятивну гістограму (емпіричну функцію розподілу), якщо вибрати Cumulative Histogram, або Cumulative polygon. Якщо вибрати Relative, то на графіку по вісі Y будуть відносні частоти. Спробуйте самостійно зробити це.

7.3. Збережіть весь виконаний аналіз. Для цього виберіть File – Save As – Save StatFolio As.

Для того щоб відкрити даний аналіз на іншому комп'ютері, потрібно мати обидва файли - і StatFolio, і файл з даними Data File.

Контрольні запитання:

1. Як називаються панелі вікна програми StatGraphics?
2. Які пункти містить меню та яке їх призначення?
3. Де розташовується панель задач та які вкладники вона має?
4. Що називають генеральною сукупністю?
5. Що називають вибіркою?
6. Що називають модою?
7. Що називають медіаною?
8. Що називають коефіцієнтом асиметрії?
9. Що називають коефіцієнтом ексцесу?

10. Що називають гістограмою?
11. Що називають полігоном частот?
12. Що називають графіком «ящик з вусами?»

Лабораторна робота №2

Компонентний аналіз даних в програмному продукті StatGraphics

Мета роботи: навчитися виконувати компонентний аналіз та будувати діаграми розсіювання в середовищі StatGraphics з метою дослідження структури даних.

Теоретичні відомості

Метод головних компонентів (МГК) призначений для оцінювання моделей великого розміру, а також для оцінки параметрів моделі, якщо до неї входять мультиколінеарні змінні.

Під мультиколінеарністю розуміють наявність лінійної залежності між двома або більше факторними змінними у регресійній моделі. Під регресійною моделлю розуміють гіпотезу, яка повинна бути піддана статистичній перевірці, після чого вона приймається або відхиляється.

Регресія - форма зв'язку між випадковими величинами, закон зміни математичного очікування однієї випадкової величини залежно від значень іншої. Розрізняють прямолінійну, криволінійну, ортогональну, параболічну та інші регресії, а також лінію і площину регресії.

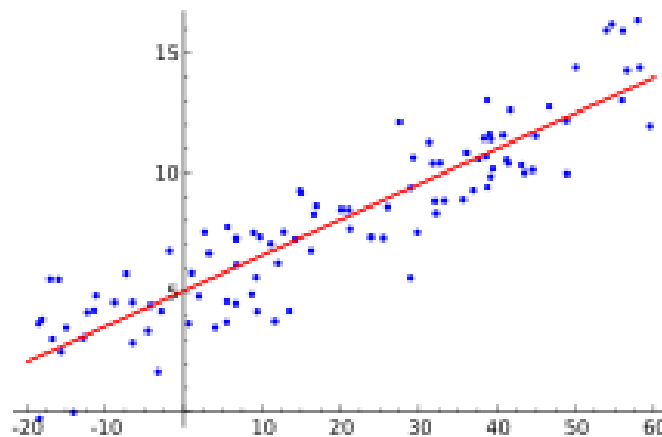


Рис. 2.1. Приклад прямолінійної регресії

Крива регресії (рис. 2.1) є залежність умовного математичного очікування величини Y від заданого значення X .

Існують різні модифікації методу головних компонентів, які різняться між собою залежно від того, що береться за основу при визначенні ортогональних змінних - коваріаційна чи кореляційна матриця незалежних змінних.

Коваріаційна матриця (або коваріаційна таблиця) – це квадратна матриця, яка складена з попарних коваріацій (спільних мінливостей) і дисперсій (відхилень) двох або більше випадкових величин.

Коваріація - це міра спільної мінливості двох випадкових змінних. Якщо більші значення однієї змінної здебільшого відповідають більшим значенням іншої, і те саме виконується для менших значень, тобто змінні схильні демонструвати подібну поведінку, то коваріація є додатною. В протилежному випадку, коли більші значення однієї змінної здебільшого відповідають меншим значенням іншої, тобто змінні схильні демонструвати протилежну

поведінку, коваріація є від'ємною. Отже, знак коваріації показує тенденцію в лінійному взаємозв'язку між цими змінними.

Дисперсія - це міра відхилення значень випадкової величини від центру розподілу. Більші значення дисперсії свідчать про більші відхилення значень випадкової величини від центру розподілу.

Нехай маємо матрицю X , яка описує незалежні змінні моделі. Оскільки спостереження, що утворюють матрицю X , як правило, корельовано (залежні) між собою, то можна поставити питання про кількість реально незалежних змінних, які входять до цієї матриці.

Точніше, ідея методу полягає в тому, щоб перетворити множину змінних X на нову множину попарно некорельованих змінних, серед яких перша відповідає максимально можливій дисперсії, а друга - максимально можливій дисперсії в підпросторі, який є ортогональним (перпендикулярним) до першого, і т.д.

Алгоритм побудови головних компонентів (з геометричної позиції) полягає в наступному.

1. Розрахунок матриці ω .

1.1. Проводиться центрування вихідних даних – визначення центру розподілу даних (центру ваги) рис. 2.2, а;

1.2. Система координат переноситься в центр розподілу даних (рис. 2.2, б);

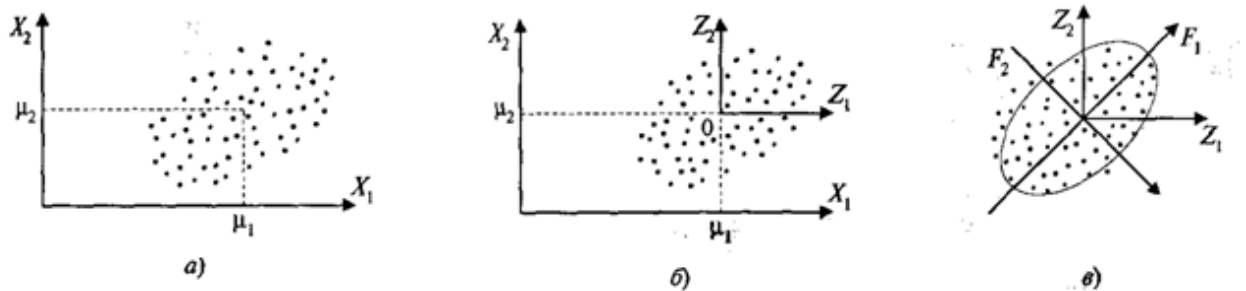


Рис. 2.2. Алгоритм побудови головних компонентів:

а) - визначення центру ваги даних; б) - перенесення системи координат в центр розподілу даних; в) - формування головних компонентів.

1.3. Формування головних компонентів $F_1, F_2, \dots, F_r$. Лінійні комбінації вибираються таким чином, що серед усіх можливих комбінацій перша головна компонента $F_1(X)$ має найбільшу дисперсію (рис. 2.2, в). Дисперсія σ прагне до максимуму. Графічно це виглядає як орієнтація нової координатної вісі F_1 вздовж напрямку найбільшої витягнутості еліпсоїда розсіювання об'єктів у вихідному просторі P ознак.

Друга головна компонента $F_2(X)$ перпендикулярна першій і будується виходячи з припущень знаходження максимальної дисперсії серед всіх, лінійних комбінацій що лишилися.

Графічно це інтерпретується напрямком найбільшої витягнутості еліпсоїда розсіювання, який перпендикулярний першій головній компоненті.

Наступні компоненти визначаються за аналогічною схемою.

На відміну від інших методів аналізу даних (регресійного, дискримінантного аналізу) розв'язання задачі виділення головних компонентів зводиться до вирішення класичних питань аналітичної геометрії: зміна масштабу простору, поворот координатної системи, координатне відображення векторів z в простір F і навпаки, F в z .

Примітки до пункту 1:

- кількість головних компонентів дорівнює кількості вихідних елементарних ознак;
- всі головні компоненти в сумі пояснюють всі 100% дисперсії елементарних ознак.

Обчислюються всі головні компоненти по спаданню дисперсії;

- відсоток пояснюваної дисперсії для будь-якої головної компоненти розраховується на підставі так званих власних значень Eigenvalue.

Власні значення - величина дисперсії вихідних ознак, які пояснюються цією головною компонентою.

2. Вибір значущих головних компонентів і визначення назви для них.

2.1. Вибір може бути заснований на різних умовах.

Наприклад, критерій «Кам'яниста осип». Цей графічний метод запропоновано Кеттелем у 1966 році. Згідно з цим критерієм будується система координат: по вісі x -головні компоненти, по вісі y -власні значення.

На графіку знаходиться точка, в якій спадання власних значень максимально сповільнюється. Згідно з графіком залишаємо 2-3 головні компоненти.

2.2. Визначення назви для головних компонент.

Для визначення назви необхідно врахувати, що $\omega_{i,j}$ визначає внесок i -ої змінної в формування j -ої головної компоненти.

Процедура. Для будь-якої головної компоненти F_j множина значень $\langle \omega_{i,j} \rangle$ умовно розбивається на чотири підмножини (рис. 2.3):

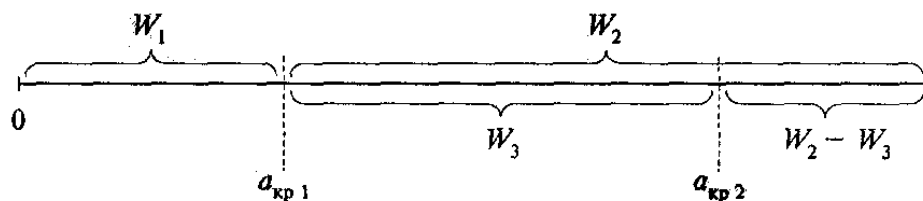


Рис. 2.3. Розбиття множини значень на підмножини:

W_1 -підмножина незначущих вагових коефіцієнтів;

W_2 -підмножина значущих вагових коефіцієнтів;

W_3 -підмножина значущих вагових коефіцієнтів, які не беруть участі в назві головної компоненти;

$W_2 - W_3$ -підмножина значущих вагових коефіцієнтів, що беруть участь у формуванні назви.

Додаткове виділення множини W_3 викликано прагненням до більш простої структури головної компоненти, що завжди легко піддається інтерпретації (знаходженню проміжних

значень). Оцінка значущості обраних коефіцієнтів проводиться за допомогою коефіцієнта інформативності:

$$K_i = \frac{\sum_{i=1}^p \{w_2 - w_3\}^2}{\sum_{i=1}^p \{w\}^2}$$

Рекомендоване значення коефіцієнта $K_i = 0,75-0,95$

Переваги МГК:

- з точки зору візуалізації багатовимірних даних метод має властивості найменшого спотворення структури вихідного простору при проектуванні в простір меншої розмірності;

- метод застосовується успішно в системі з іншими методами дослідження даних, наприклад в кореляційно-регресійному аналізі.

Недоліки МГК:

Можлива ситуація, коли вагові коефіцієнти мають близькі за величиною значення. В цьому випадку результат слабо інтерпретуємий. Ця проблема вирішується застосуванням інших видів аналізу, наприклад факторного, додаванням або вилученням змінних із аналізу.

Порядок виконання роботи

Доцільно розглянути приклад, що відноситься до порівняльного оцінювання виробів, що характеризуються одночасно кількома параметрами. це - автомобілі. У таблиці(Додаток 2) наводяться вибіркові відомості про фірму-виробника автомобіля, назву моделі, а також оціночні параметри - вага (змінна weight), число циліндрів (змінна cylinders), прискорення (змінна accel), об'єм двигуна (змінна displace) і потужність в кінських силах (змінна horsepower).

1. Задання змінних та введення даних

1.1. В програмному продукті StatGraphics необхідно створити таблицю даних на основі прикладу (Додаток 2). Для задання назв змінних величин, подвійним клацанням ЛКМ на назві стовпчика, відкрийте вікно зміни назви Modify Column (модифікувати колонку). В строчку Name (Ім'я) даного вікна введіть англomовну назву імені стовпчика, наприклад: Make. У такий спосіб дайте назви всім стовпчикам даних (Model, Weight, Cylinders, Accel, Displace, Horsepower).

1.2. Скопіюйте дані з таблиці (Додаток 2).

	Make	Model	Weight	Cylinders
1	Volkswagen	Rabbit D1	1985	4
2	Ford	Fiesta	1800	4
3	Mazda	GLC Deluxe	1985	4
4	Datsun	B210 GX	2070	4
5	Honda	Civic CVCC	1800	4
6	Oldsmobile	Cutlass	3365	8
7	Dodge	Diplomat	3735	8
8	Mercury	Monarch	3570	8
9	Pontiac	Phoenix	3535	6
10	Chevrolet	Malibu	3155	6
11	Ford	Fairmont A	2965	6
12	Ford	Fairmont M	2720	4
13	Plymouth	Volare	3430	6
14	AMC	Concord	3210	6
15	Buick	Century	3380	6
16	Mercury	Zephyr	3070	6
17	Dodge	Aspen	3620	6

Рис. 2.4. Таблиця вхідних даних

1.3. Збережіть створену таблицю під назвою Cardata, вибравши File – Save – Save DataFile.

2. Отримання статистичних оцінок

2.1. Виберіть Describe/ Multivariate Methods/ Principal Components. З'явиться вікно діалогу для завдання аналізованих змінних – вікно основних компонент (рис. 2.5).

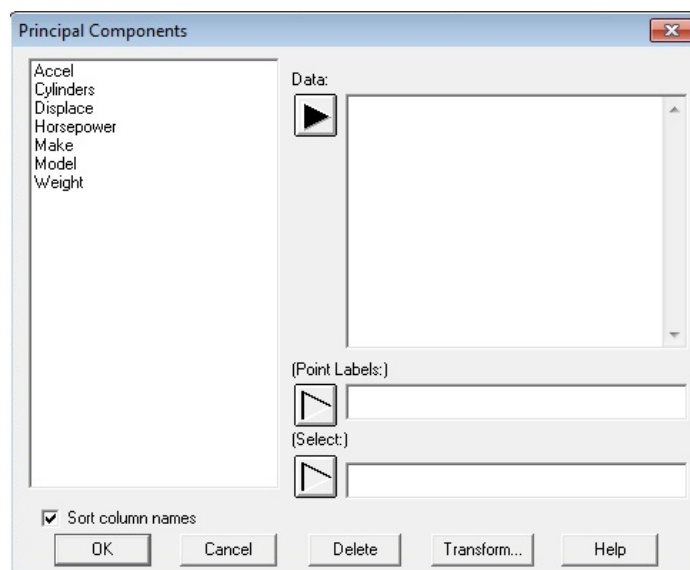


Рис. 2.5. Вікно основних компонент

Виділіть першу змінну Weight натисканням лівої кнопки миші, а потім клацніть по кнопці зі стрілкою, таким чином вводячи дану змінну в аналіз. У такий же спосіб введіть в аналіз інші змінні (Cylinders, Accel, Displace, Horsepower) та натисніть ОК (рис. 2.6).

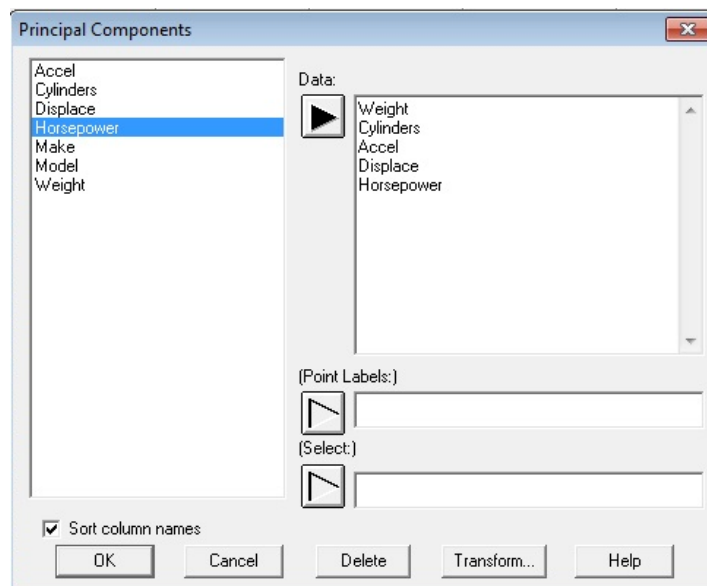


Рис. 2.6. Введення основних компонент в аналіз

Отримуємо вихідний звіт аналізу методу головних компонентів (рис. 2.7), з якого видно, що аналізу підлягають змінні ваги (weight), циліндри (cylinders), прискорення (accel), об'єм двигуна (displace), потужність (horsepower), кількість компонент 151.

Principal Components Analysis

Data variables:

Weight
Cylinders
Accel
Displace
Horsepower

Data input: observations

Number of complete cases: 151

Missing value treatment: listwise

Standardized: yes

Number of components extracted: 2

Principal Components Analysis

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	3,47615	69,523	69,523
2	1,0294	20,588	90,111
3	0,309124	6,182	96,293
4	0,145867	2,917	99,211
5	0,0394616	0,789	100,000

Рис. 2.7. Вихідний звіт аналізу методу головних компонент

Далі наведена інформація безпосередньо методу головних компонент:

- номер головного компонента (Component Number), де:

- 1 – вага;
- 2 – кількість циліндрів;
- 3 – прискорення;
- 4 – об'єм;
- 5 – потужність;

- власні значення головних компонентів, впорядковані за величиною (Eigenvalue);

- відсоток дисперсії, що припадає на кожну виділену головну компоненту (Percent of Variance);

- накопичений відсоток дисперсії (Cumulative Percentage).

Наведені в звіті цифри говорять про те, що вже перші дві головні компоненти описують 90,111% дисперсії вихідних даних. Третя головна компонента додає ще 6,182% дисперсії, так що в сумі виходить 96,293% дисперсії.

2.2. Клацанням ПКМ в робочому полі документа виберіть команду Analysis Options та встановіть кількість компонентів рівну трьом (рис. 2.8).

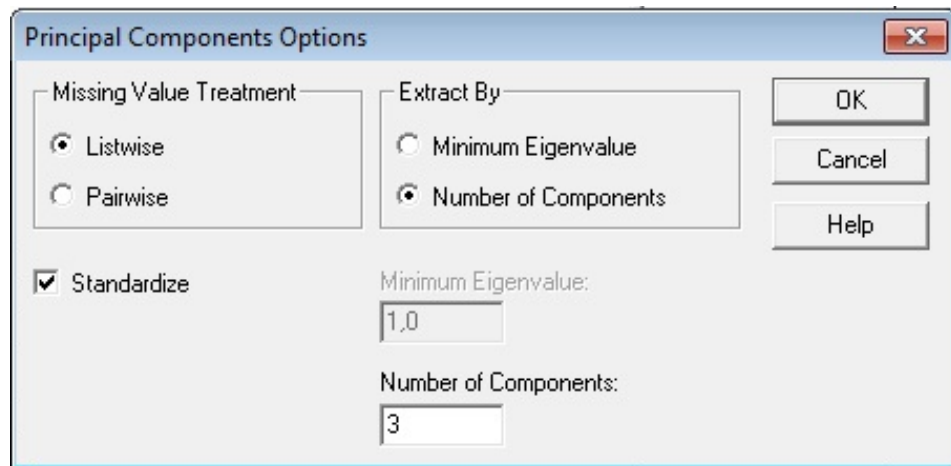


Рис. 2.8. Вікно налаштування основних компонентів

2.3. Для більш детального аналізу натисніть кнопку табличних опцій  і у відповідному вікні діалогу (рис. 2.9) встановіть прапорець компонентних ваг (Component Weights). Отримаємо наступну таблицю (рис. 2.10).

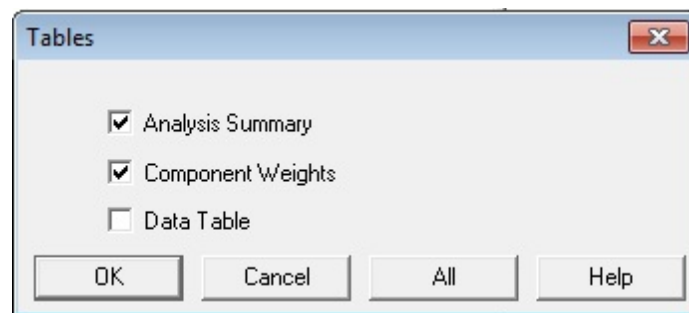


Рис. 2.9. Вікно вибору табличних опцій

Table of Component Weights			
	Component	Component	Component
	1	2	3
Weight	0,492492	0,272583	0,0518503
Cylinders	0,501315	0,0928374	-0,421138
Accel	-0,161338	0,926091	0,251486
Displace	0,517691	0,125393	-0,292074
Horsepower	0,460544	-0,209072	0,819391

Рис. 2.10. Вага ознак в головних компонентах

Аналіз даних таблиці (рис. 2.10) показав, що:

- в першій головній компоненті приблизно однакові за величиною позитивні коефіцієнти мають вага, кількість циліндрів, об'єм двигуна та потужність в кінських силах;
- у другій головній компоненті превалює тільки одна величина: прискорення;
- в третій головній компоненті спостерігається поєднання ваги машини та її потужності (з позитивним знаком), якому протиставляється кількість циліндрів (з негативним знаком).

3. Побудова діаграми розсіювання

Необхідно побудувати діаграму розсіювання всієї сукупності автомобілів в просторі виділених трьох перших головних компонент.

3.1. Для цього натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 2.11) встановіть прапорець 3Dдіаграми розсіювання (3D Scatterplot). Отримаємо наступне тривимірне зображення (рис. 2.12).

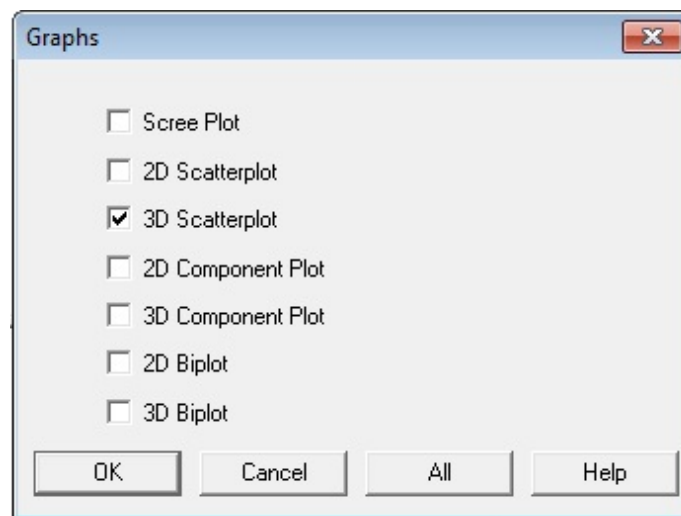


Рис. 2.11. Вікно вибору графічних опцій

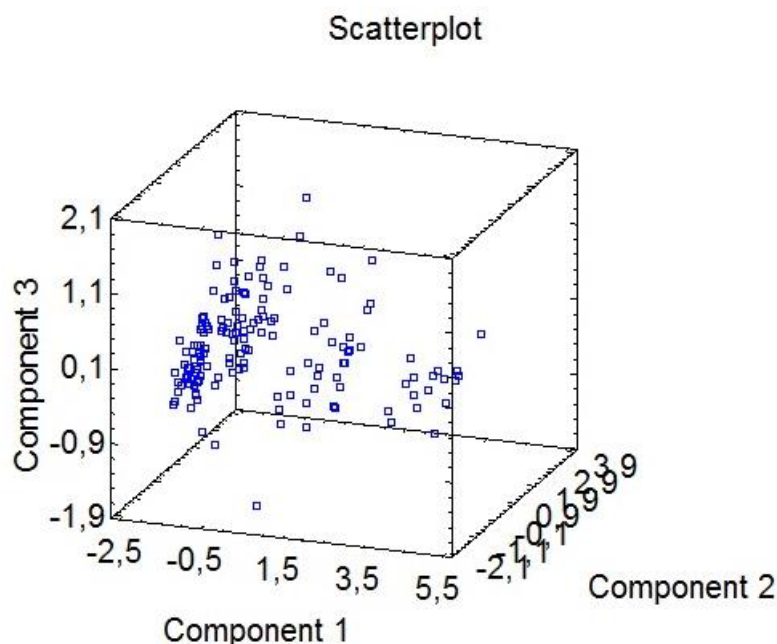


Рис. 2.12. Діаграма розсіювання

З наведеної діаграми добре видно, що вся досліджувана сукупність автомобілів розділилася на три досить чітко виражені групи.

Для першої, найбільш численної групи характерні порівняно невеликі вага, кількість циліндрів, потужність і об'єм двигуна (перша зліва група). Разом з тим, велика частка автомобілів цієї групи мають добре прискорення (більші значення 2-ої головної компоненти) і високе співвідношення ваги і потужності до кількості циліндрів (3-тя головна компонента).

Друга група не настільки численна, але їй також властиві зазначені характеристики, хоча і менш яскраво виражені.

І нарешті, третя група автомашин (порівняно нечисленна) має велику вагу, потужність, кількість, циліндрів. У той же час, показники прискорення і співвідношення ваги і потужності до кількості циліндрів тут (якщо говорити в цілому) набагато менші.

Таким чином, здійснений за допомогою методу головних компонент аналіз даних дозволяє отримати більш «об'ємне» бачення сучасного автомобільного ринку, що може сприяти кращій орієнтації як споживачів цієї продукції, так і виробників з позицій оцінки існуючих тенденцій.

4. Індивідуальне завдання

На основі даних таблиці (Додаток 3) виконати:

- отримання статистичних оцінок;
- побудову діаграми розсіювання;
- аналіз отриманих результатів.

Контрольні запитання:

1. Призначення методу головних компонентів?
2. Що називають регресією?
3. Що називають кривою регресії?
4. Що розуміють під регресійною моделлю?
5. Що називають мультиколінеарністю?
6. Що називають дисперсією?
7. Що означає поняття коваріація?
8. Яка основна ознака першої головної компоненти?
9. Назвіть переваги методу головних компонентів.
10. Назвіть недоліки методу головних компонентів.
11. Перерахуйте послідовність дій при аналізі даних методом головних компонентів в програмному продукті StatGraphics.

Лабораторна робота №3

Кластерний аналіз даних в програмному продукті StatGraphics

Мета роботи: навчитися виконувати кластерний аналіз та будувати діаграми розсіювання в середовищі StatGraphics з метою дослідження структури даних.

Теоретичні відомості

Кластерний аналіз (Data clustering) - це задача розбиття заданої вибірки (множини) об'єктів (ситуацій) на підмножини, що називаються кластерами, так, щоб кожен кластер складався із схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися.

Кластерний аналіз - це багатовимірна статистична процедура, яка виконує збір даних, що містять інформацію про вибірку об'єктів і потім упорядковує об'єкти в порівняно однорідні групи - кластери.

Мета кластерного аналізу полягає в пошуку наявних структур, що виражається в утворенні груп схожих між собою об'єктів – кластерів. Водночас його дія полягає й у привнесенні структури в досліджувані об'єкти. Це означає, що методи кластеризації необхідні для виявлення структури в даних, яку нелегко знайти при візуальному обстеженні або за допомогою експертів.

Кластер (cluster) - група однакових або подібних елементів, що характеризуються певними загальними властивостями, зібраних разом або близько один до одного.

Критерій якості кластеризації в тій чи іншій мірі повинен відображати такі неформальні вимоги:

- всередині груп об'єкти повинні бути тісно пов'язані між собою;
- об'єкти різних груп повинні бути далекі один від одного;
- розподіл об'єктів за групами має бути рівномірним.

Основними завданнями кластерного аналізу є:

- розробка типології або класифікації досліджуваних об'єктів;
- дослідження та визначення прийнятних концептуальних схем групування об'єктів;
- висунення гіпотез на підставі результатів дослідження даних;
- перевірка гіпотез.

Кластерний аналіз потребує здійснення таких послідовних кроків:

- 1) проведення вибірки об'єктів для кластеризації;
- 2) визначення множини ознак, за якими будуть оцінюватися відібрані об'єкти;
- 3) оцінка міри подібності об'єктів;
- 4) застосування кластерного аналізу для створення груп подібних об'єктів;
- 5) перевірка достовірності результатів кластерного рішення.

Кожен з цих кроків відіграє значну роль у практичному здійсненні аналізу.

Методи кластерного аналізу дозволяють вирішувати такі завдання:

- проведення класифікації об'єктів з урахуванням множини ознак з метою поглиблення знань про множину досліджуваних ознак;

- перевірка висунутих припущень про наявність певної структури в досліджуваній множині об'єктів;

- побудова нових класифікацій для слабо вивчених явищ, тобто пошук в досліджуваній множині заздалегідь невідомої структури.

StatGraphics пропонує 7 методів кластеризації:

- метод «найближчого сусіда» (Nearest neighbor) - дистанція між двома кластерами визначається, як відстань між парою значень спостережень, розташованих найближче один до одного, причому кожне спостереження береться зі свого кластера;

- метод «найвіддаленішого сусіда» (Furthest neighbor) - дистанція між двома кластерами визначається, як відстань між найвіддаленішими один від одного значеннями спостережень, причому кожне спостереження береться зі свого кластера;

- центроїдна кластеризація (Centroid) - в обох кластерах розраховуються середні значення змінних відносно до них спостережень. Потім відстань між двома кластерами розраховується, як дистанція між двома усередненими спостереженнями;

- медіанна кластеризація (Median) - цей метод схожий на центроїдну кластеризацію. Однак в попередньому методі центр ваги нового кластера знаходять, як зважене середнє центроїдів обох вихідних кластерів, причому кількості спостережень вихідних кластерів утворюють ваговий коефіцієнт. У медіанному ж методі обидва вихідних кластера беруться з однаковою вагою;

- метод «середнього групи» (Group average) – дистанція між двома різними кластерами розраховується, як середня відстань між усіма парами об'єктів в них;

- метод Варда (Ward's) - спочатку в обох кластерах для всіх наявних спостережень проводиться розрахунок середніх значень окремих змінних. Потім обчислюються квадрати евклідових відстаней від окремих спостережень кожного кластера до цього кластерного середнього значення. Ці дистанції підсумовуються. Потім в один новий кластер об'єднуються ті кластера, при об'єднанні яких виходить найменший приріст загальної суми дистанцій;

- метод «к-середніх» (K-Means) – якщо вже є гіпотези щодо числа кластерів (за спостереженнями або по змінним), то можна вказати системі створити рівно це число кластерів, щоб вони були настільки різні, наскільки це можливо.

Евклідова відстань (d_{ij}) - найкоротша відстань між двома точками:

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2},$$

Де: x_{ik} , x_{jk} - значення k-ої змінної відповідно у i-го та j-го об'єктів;

m - кількість ознак.

Всі методи кластерного аналізу діляться на дві групи:

- агломеративні (об'єднують) - побудовані на основі послідовного об'єднання об'єктів в групи;

- дивізійні (розділяють) - побудовані на основі розчленування групи на окремі об'єкти.

Узагальнений алгоритм кластерного аналізу:

1. Задається початкове (штучне або довільне) розбиття на кластери, і визначається деякий математичний критерій якості автоматичної класифікації.

2. Об'єкти переносяться з кластера в кластер до тих пір, поки значення критерію якості не перестане поліпшуватися. При цьому можливий або повний перебір варіантів, або скорочений на підставі будь-яких сукупностей прийомів та методів.

Порядок виконання роботи

1. Задання змінних та введення даних

1.1. В програмному продукті StatGraphics необхідно створити таблицю даних на основі прикладу (Додаток 2). Для задання назв змінних величин, подвійним клацанням ЛКМ на назві стовпчика, відкрийте вікно зміни назви Modify Column (модифікувати колонку). В строчку Name (Ім'я) даного вікна введіть англomовну назву імені стовпчика, наприклад: Make. У такий спосіб дайте назви всім стовпчикам даних (Model, Weight, Cylinders, Accel, Displace, Horsepower).

1.2. Скопіюйте дані з таблиці (Додаток 2).

	Make	Model	Weight	Cylinders
1	Volkswagen	Rabbit D1	1985	4
2	Ford	Fiesta	1800	4
3	Mazda	GLC Deluxe	1985	4
4	Datsun	B210 GX	2070	4
5	Honda	Civic CVCC	1800	4
6	Oldsmobile	Cutlass	3365	8
7	Dodge	Diplomat	3735	8
8	Mercury	Monarch	3570	8
9	Pontiac	Phoenix	3535	6
10	Chevrolet	Malibu	3155	6
11	Ford	Fairmont A	2965	6
12	Ford	Fairmont M	2720	4
13	Plymouth	Volare	3430	6
14	AMC	Concord	3210	6
15	Buick	Century	3380	6
16	Mercury	Zephyr	3070	6
17	Dodge	Aspen	3620	6

Рис. 3.1. Таблиця вхідних даних

1.3. Збережіть створену таблицю, вибравши File – Save – Save DataFile.

2. Виконання кластерного аналізу

2.1. Виберіть Describe/ Multivariate Methods/ Cluster Analysis. З'явиться вікно діалогу для завдання аналізованих змінних кластерного аналізу (рис. 3.2).

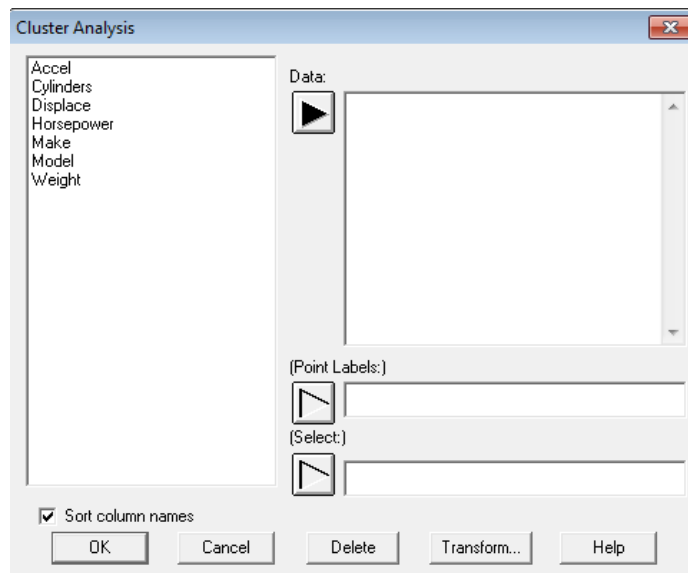


Рис. 3.2. Вікно вибору змінних

Виділіть першу змінну *Accel* натисканням лівої кнопки миші, а потім клацніть по кнопці зі стрілкою, таким чином вводячи дану змінну в аналіз. У такий же спосіб введіть в аналіз інші змінні (*Cylinders*, *Displace*, *Horsepower*, *Weight*) та натисніть **OK** (рис. 3.3).

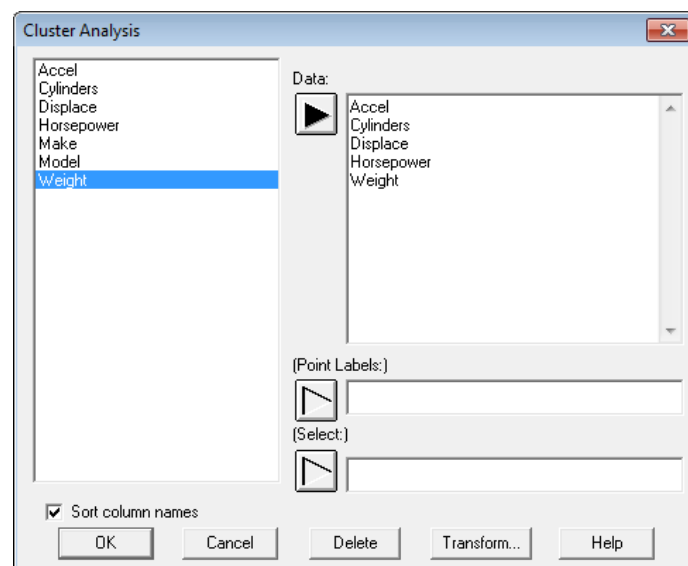


Рис. 3.3. Введення змінних в аналіз

2.2. Клацанням ПКМ в робочому полі документа виберіть команду *Analysis Options*. З’явиться діалогове вікно для вибору параметрів кластерного аналізу. Встановіть прапорець *Ward’s*, щоб виділення кластерів відбувалось за методом Варда (рис. 3.4) та натисніть **OK**.

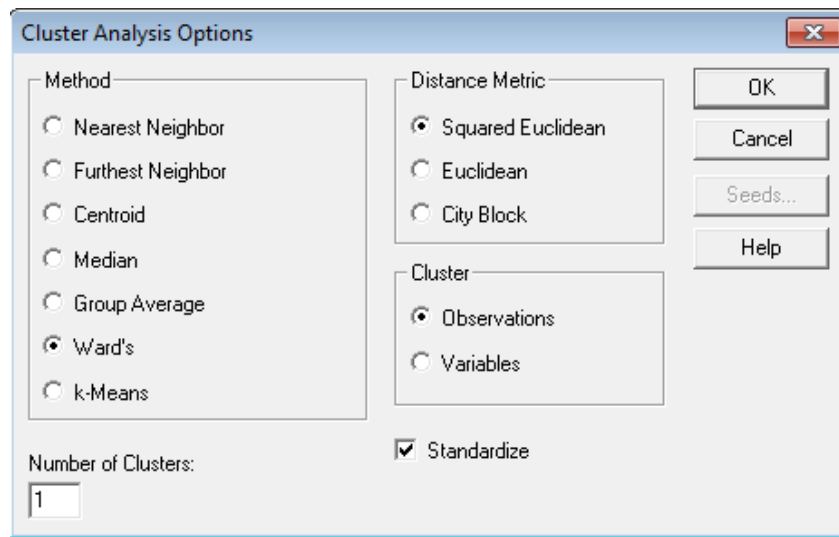


Рис. 3.4. Вибір методу кластерного аналізу

Отримуємо вихідний звіт кластерного аналізу (рис. 3.5).

Cluster Analysis

Data variables:

Accel
Cylinders
Displace
Horsepower
Weight

Number of complete cases: 151

Clustering Method: Ward's

Distance Metric: Squared Euclidean

Clustering: observations

Standardized: yes

Cluster Summary

Cluster	Members	Percent
1	151	100,00

Centroids

Cluster	Accel	Cylinders	Displace	Horsepower	Weight
1	16,2159	4,86093	154,457	89,0	2679,13

The StatAdvisor

This procedure has created 1 cluster from the 151 observations supplied. The clusters are groups of

Рис. 3.5. Вихідний звіт кластерного аналізу

У звіті кластерного аналізу перш за все вказуються: назви змінних, що беруть участь в аналізі - accel, cylinders, displace, horsepower, weight, кількість повних зразків(спостережень без пропусків) - 151, використаний метод кластерного аналізу – метод Варда, прийнята метрика - евклідова відстань.

Далі у звіті наведена інформація про:

- число кластерів - Cluster;
- кількість об'єктів в кожному кластері (населеність) - Members;
- відповідний відсоток населеності - Percent.

Нижче подана таблиця координат центроїди (Centroids) кластерів. На основі цих показників можна судити про те, які змінні грають найбільш важливу роль в кожному кластері.

Крім того, в нижній частині звіту наведено додаткову пояснювальну інформацію.

2.3. Клацанням ПКМ в робочому полі документа виберіть команду Analysis Optionста встановіть кількість груп - 3 (рис. 3.6).

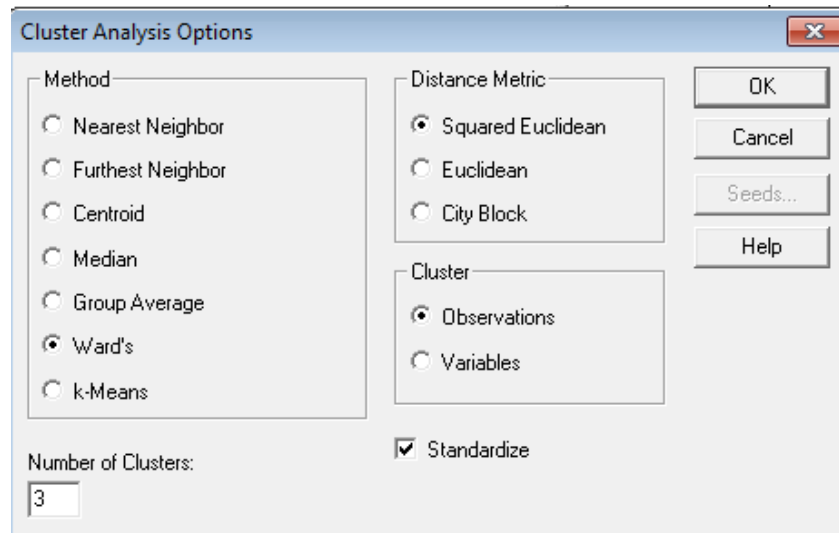


Рис. 3.6. Вікно налаштування кластерного аналізу

Вихідний звіт кластерного аналізу прийме вигляд, як на рис. 3.7.

Cluster Analysis

Data variables:

Accel
Cylinders
Displace
Horsepower
Weight

Number of complete cases: 151

Clustering Method: Ward's

Distance Metric: Squared Euclidean

Clustering: observations

Standardized: yes

Cluster Summary

Cluster	Members	Percent
1	104	68,87
2	15	9,93
3	32	21,19

Centroids

Cluster	Accel	Cylinders	Displace	Horsepower	Weight
1	16,5423	4,01923	113,51	78,7788	2366,81
2	14,0933	8,0	317,267	136,133	3772,6
3	16,15	6,125	211,219	100,125	3181,59

The StatAdvisor

This procedure has created 3 clusters from the 151 observations supplied. The clusters are groups of

Рис. 3.7. Вихідний звіт кластерного аналізу

2.4. Для побудови дендрограми натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 3.8) встановіть прапорець Дендрограма (Dendrogram). Отримаємо наступне зображення (рис. 3.9).

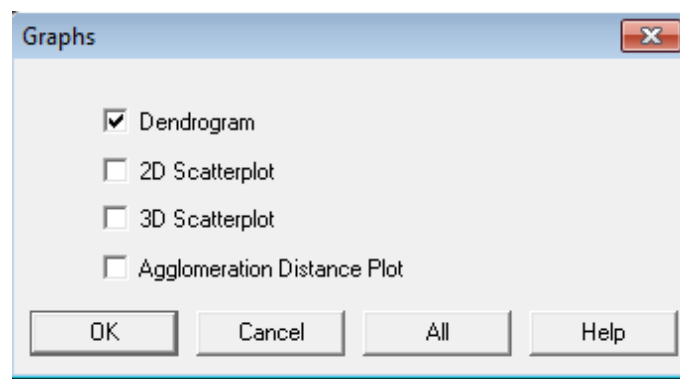


Рис. 3.8. Вікно вибору графічних опцій

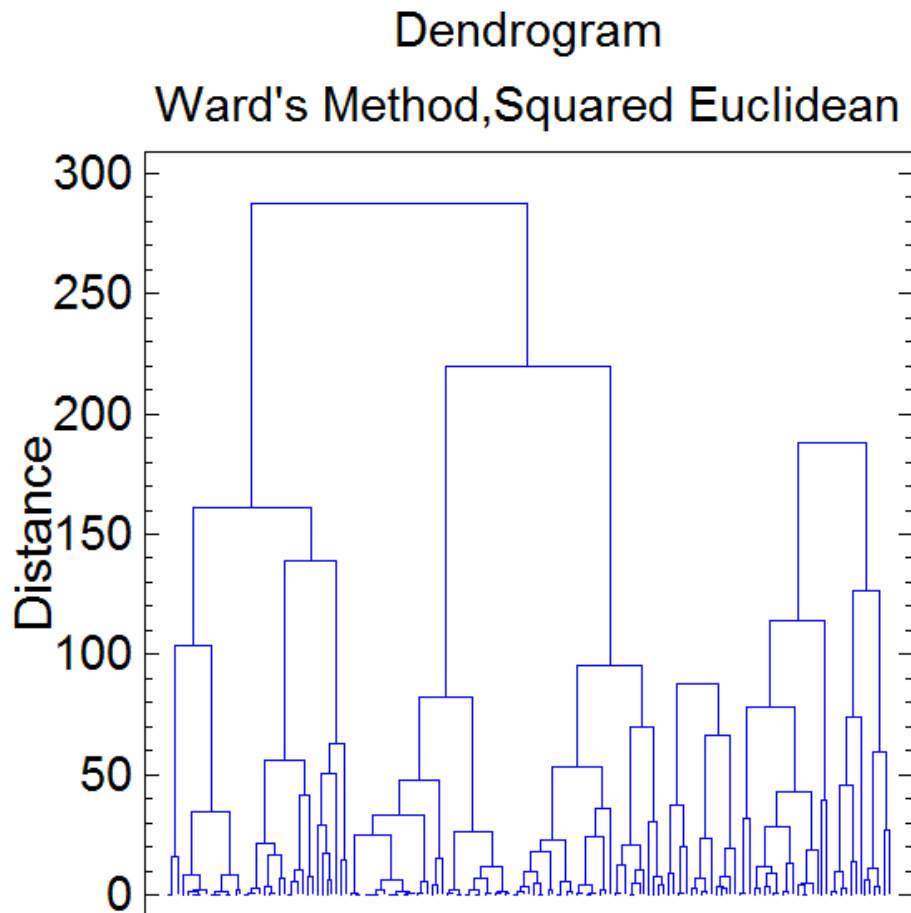


Рис. 3.9. Дендрограма, отримана методом Варда для трьох кластерів

2.5. Для більш детального аналізу натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 3.10) встановіть прапорець таблиці належності спостережень (Membership Table). В даній таблиці описуються вибрані параметри кластерного аналізу та дається повний список всіх спостережень (Row) та номери кластерів (Cluster), в які входять зазначені спостереження (рис. 3.11).

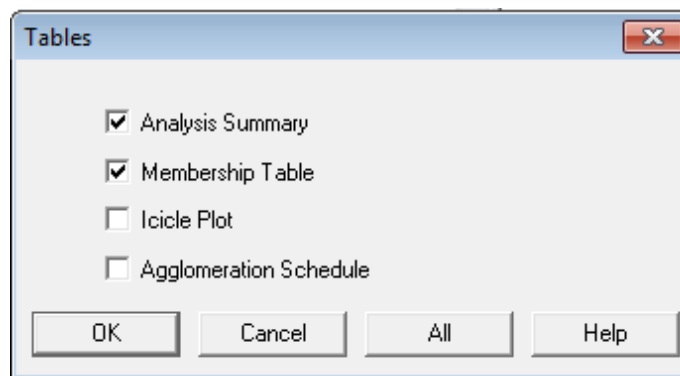


Рис. 3.10. Вікно вибору табличних опцій

Membership Table

Clustering Method: Ward's

Distance Metric: Squared Euclidean

Row	Cluster
1	1
2	1
3	1
4	1
5	1
6	2
7	2
8	2
9	3
10	3
11	3
12	1
13	3
14	3
15	3
16	3
17	3

Рис. 3.11. Таблиця належності спостережень кластерам

3. Побудова двомірної діаграми розсіювання

3.1. Натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 3.12) встановіть прапорець 2D діаграми розсіювання (2D Scatterplot).

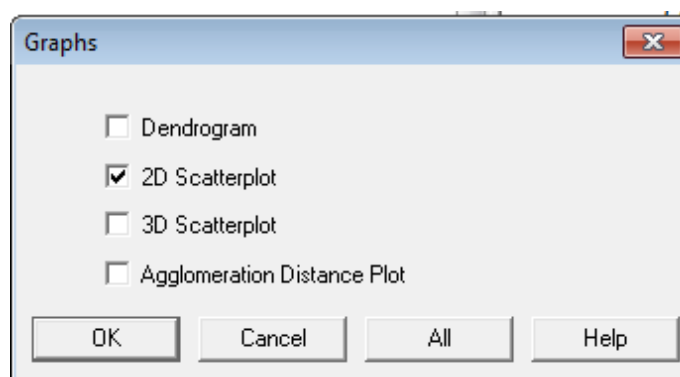


Рис. 3.12. Вікно вибору графічних опцій

3.2. Для завдання координатних осей для змінних, клацанням ПКМ в робочому полі діаграми розсіювання виберіть команду Pane Options. У вікні налаштувань 2D діаграми розсіювання (2D Scatterplot Options) необхідно задати по вісі X значення ваги (Weight), по вісі Y – потужності (Horsepower) (рис. 3.13). Отримаємо наступне зображення (рис. 3.14).

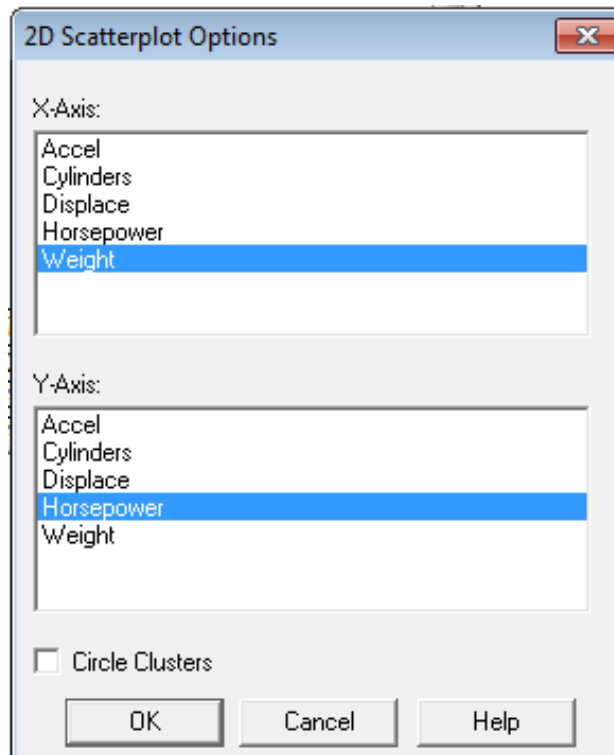


Рис. 3.13. Вікно налаштувань 2D-діаграми розсіювання

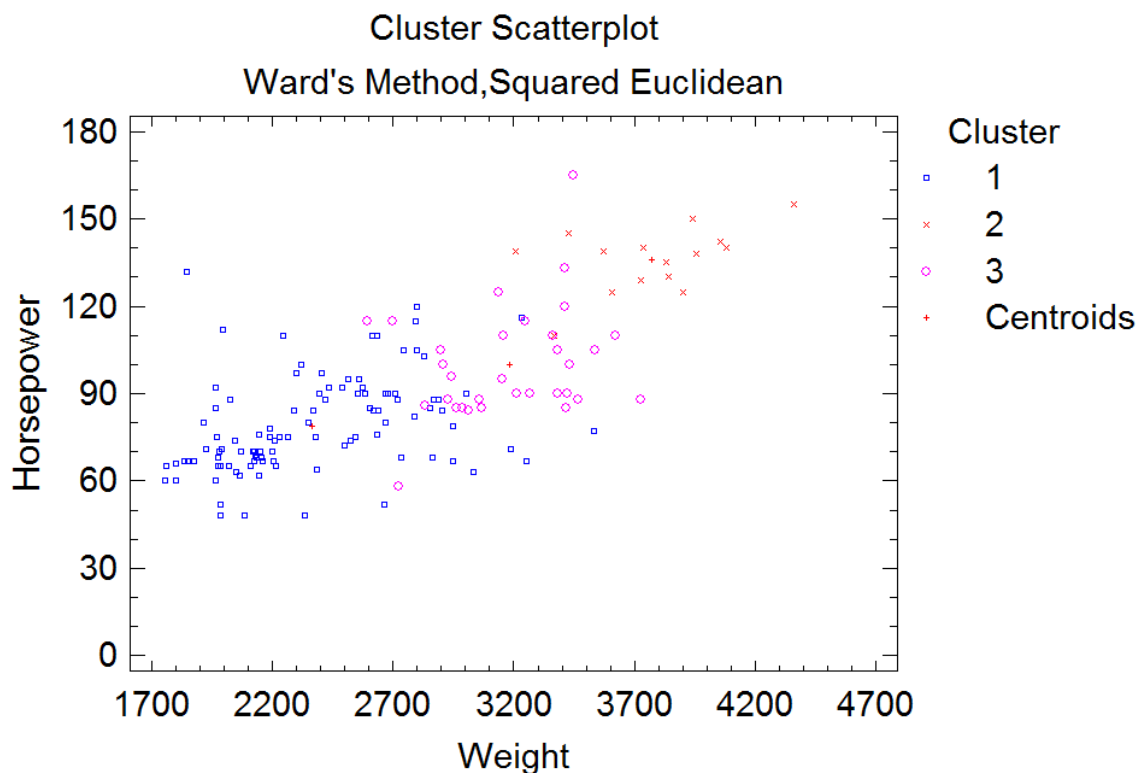


Рис. 3.14. Діаграма розсіювання

Діаграма показує, як групуються досліджувані спостереження на площині двох змінних Weight і Horsepower. Кожен кластер представлений на діаграмі власним символом і кольором. З графіка випливає, що перший кластер має низькі показники маси та потужності. У кластері 2 спостерігаються найвищі показники маси та потужності. У кластері 3 спостерігаються середні показники маси та потужностей автомобілів.

4. Індивідуальне завдання

На основі даних таблиці (Додаток 3) виконати:

- отримання статистичних оцінок;
- побудову дендрограми;
- побудову двовірної діаграми розсіювання;
- побудову тривірної діаграми розсіювання;
- аналіз отриманих результатів.

Контрольні запитання:

1. Що називають кластерним аналізом?
2. Що є метою кластерного аналізу?
3. Що називають кластером?
4. Які вимоги висуваються до критеріїв якості кластеризації?
5. Який алгоритм дій при виконанні кластерного аналізу?
6. Які завдання вирішує кластерний аналіз?
7. Які методи кластерного аналізу доступні в програмному продукті StatGraphics? Дайте характеристику цих методів?
8. Перерахуйте послідовність дій при кластерному аналізі даних в програмному продукті StatGraphics.

Лабораторна робота №4

Нелінійна регресія в програмному продукті StatGraphics

Мета роботи: навчитися виконувати регресійний аналіз, будувати матриці парних коефіцієнтів, будувати інтервальні оцінки дисперсій, перевіряти гіпотезу, виконувати прогнозний розрахунок та будувати його графік в середовищі StatGraphics з метою дослідження структури даних.

Теоретичні відомості

Регресія - форма зв'язку між випадковими величинами, закон зміни математичного очікування однієї випадкової величини залежно від значень іншої.

Однофакторною регресією Y від X називається одностороння випадкова лінійна залежність між випадковими величинами показника Y та одного фактора X , які знаходяться в причинно-наслідкових відношеннях, причому зміна фактора виключає зміну показника.

Загальний вигляд однофакторної регресії:

$$Y = \alpha + \beta X + u$$

де: Y - залежна змінна (пояснювальна змінна);

X - незалежна змінна (пояснююча змінна);

α, β - параметри моделі;

u - випадкова складова.

Багатофакторна (множинна) регресія - це зв'язок між трьома і більше пов'язаними між собою ознаками.

Загальний вигляд багатофакторної регресії:

$$\bar{Y}_{1,2,\dots,k} = f(x_1, x_2, \dots, x_k)$$

Побудова моделей багатофакторної регресії включає в себе кілька етапів:

- 1) вибір форми зв'язку (рівняння регресії);
- 2) відбір факторних ознак;
- 3) забезпечення достатнього обсягу сукупності.

Важливим етапом побудови рівняння множинної регресії є відбір і подальше включення в нього факторних ознак.

З одного боку, чим більше факторних ознак включено в рівняння, тим воно краще описує явище. Однак модель розмірністю 100 і більше факторних ознак складно реалізовується та вимагає великих витрат часу. Скорочення розмірності моделі за рахунок виключення другорядних, статистично несуттєвих факторів сприяє простоті та якості її реалізації. У той же час, побудова моделі регресії малої розмірності може призвести до того, що така модель буде недостатньо адекватна досліджуваним явищам і процесам.

Проблема відбору факторних ознак для побудови моделей взаємозв'язку може бути вирішена на основі інтуїтивно-логічних або багатовимірних математико-статистичних методів аналізу.

Найбільш прийнятним способом відбору факторних ознак є крокова регресія (кроковий регресійний аналіз). Суть методу крокової регресії полягає в реалізації алгоритмів послідовного включення-виключення факторів в рівняння регресії з наступною перевіркою їх статистичної значущості.

При використанні алгоритму включення фактори по чергові вводяться в рівняння регресії так званим прямим методом. При перевірці значущості введеного фактора визначається, наскільки зменшується сума квадратів залишків і збільшується величина множинного коефіцієнта кореляції.

Сутність алгоритму послідовного виключення полягає в тому, що відкидаються фактори, що стали незначущими за t-критерієм Стюдента або іншими статистичними критеріями. Фактор є незначущим, якщо його включення в рівняння регресії змінює тільки значення коефіцієнтів регресії, не зменшуючи її суми квадратів залишків і не збільшуючи їх значення.

Якщо при включенні в модель відповідної факторної ознаки величина множинного коефіцієнта кореляції збільшується, а значення коефіцієнта регресії не змінюється (або змінюється несуттєво), то дана ознака суттєва і її включення в рівняння регресії доцільне. В іншому випадку фактор недоцільно включати в модель регресії.

Довірчий інтервал – це інтервал, у межах якого з заданою довірчою імовірністю можна чекати значення оцінюваної (шуканої) випадкової величини.

Інтервальною оцінкою параметру θ називають числовий інтервал $(\theta_n^{(1)}, \theta_n^{(2)})$, який з заданою ймовірністю γ накриває невідоме значення параметру θ (рис. 4.1).

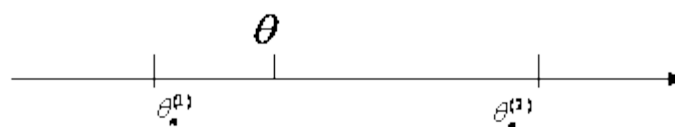


Рис. 4.1. Інтервальна оцінка параметру θ

Межі інтервалу $(\theta_n^{(1)}, \theta_n^{(2)})$ знаходять за вибірковими даними та є випадковими величинами. Інтервал $(\theta_n^{(1)}, \theta_n^{(2)})$ називають довірчим, а ймовірність γ – довірчою імовірністю або надійністю оцінки. Величина вибіркового інтервалу також залежить від об'єму вибірки (зменшується із зростанням n) та від значення довірчої ймовірності γ (збільшується при наближенні її до 1).

Вибіркою називають сукупність об'єктів, вибраних випадковим чином з генеральної сукупності.

Генеральна сукупність - сукупність всіх об'єктів, що підлягають дослідженню. Обсяг генеральної сукупності, тобто число об'єктів дослідження може бути досить великим, а інколи і нескінченним. Часто неможливо дослідити всі об'єкти генеральної сукупності.

Стандартне відхилення (standard deviation) або середнє квадратичне відхилення - найпоширеніший показник розсіювання значень випадкової величини відносно її математичного очікування. Середнє квадратичне відхилення так само, як і середнє лінійне відхилення, показує, на скільки в середньому відхиляються конкретні значення ознаки від середнього їх значення. Середнє квадратичне відхилення завжди більше середнього лінійного відхилення.

Стандартне відхилення використовують під час розрахунку стандартної похибки середнього арифметичного, для побудови довірчих інтервалів, статистичної перевірки гіпотез, виміру лінійного взаємозв'язку між випадковими величинами.

Коефіцієнт кореляції Пірсона – показник кореляції (лінійної залежності) між двома змінними X та Y , який набуває значень від -1 до $+1$ включно. Він широко використовується в науці для вимірювання ступеня лінійної залежності між двома змінними. Значення $+1$ означає, що залежність між X та Y є лінійною, і всі точки функції лежать на прямій, яка відображає зростання Y при зростанні X . Значення -1 означає, що всі точки лежать на прямій, яка відображає зменшення Y при зростанні X . Якщо коефіцієнт кореляції Пірсона $= 0$, то саме лінійної кореляції між змінними немає.

Порядок виконання роботи

1. Задання змінних та введення даних

Машинобудівним підприємством були зібрані необхідні дані для здійснення прогнозу обсягу продажів продукції певної марки (Додаток 4). Необхідно провести статистичний аналіз інформації, спрогнозувати реалізацію продукції на три подальших півріччя, припускаючи, що зв'язок між показниками лінійний.

1.1. В програмному продукті StatGraphics необхідно створити таблицю даних на основі прикладу (Додаток 6). Для задання назв змінних величин, подвійним клацанням ЛКМ на назві стовпчика, відкрийте вікно зміни назви Modify Column (модифікувати колонку). В строчку Name (Ім'я) даного вікна введіть назви стовпчиків (A , Y , X_1 , X_2 , X_3 , X_4).

1.2. Скопіюйте дані з таблиці (Додаток 4).

	A	Y	X1	X2	X3	X4
1	1	126	4,0	15,0	100,0	1
2	2	137	4,8	14,8	101,0	2
3	3	148	7,8	15,2	103,5	3
4	4	191	6,3	15,5	104,3	4
5	5	274	8,2	15,5	105,2	5
6	6	370	9,4	16,0	106,1	6
7	7	432	14,0	17,0	108,9	7
8	8	445	5,6	16,9	108,8	8
9	9	367	9,3	15,8	107,4	9
10	10	367	6,7	15,8	107,2	10
11	11	321	5,1	16,8	106,4	11
12	12	307	4,1	16,1	104,3	12
13	13	331	8,6	15,4	105,9	13
14	14	345	5,8	15,7	106,0	14
15	15	364	5,7	16,0	106,2	15

Рис. 4.2. Таблиця вхідних даних

1.3. Збережіть створену таблицю, вибравши File – Save – Save Data File.

2. Отримання статистичних оцінок

2.1. Виберіть пункт меню Describe/ Numeric Data/ Multiple–Variable Analysis. З’явиться вікно діалогу для завдання аналізованих змінних(рис. 4.3).

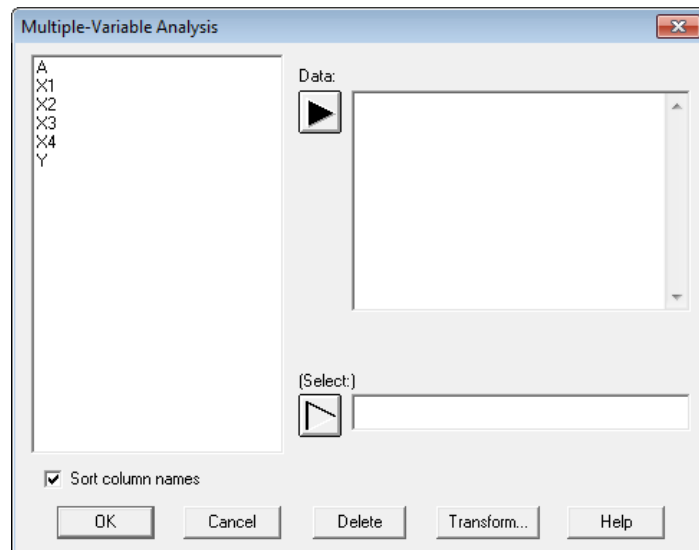


Рис. 4.3. Вікно вибору змінних

Виділіть першу змінну Y натисканням лівої кнопки миші, а потім клацніть по кнопці зі стрілкою, таким чином вводячи дану змінну в аналіз. У такий же спосіб введіть в аналіз інші змінні (X1, X2, X3) та натисніть ОК (рис. 4.4).

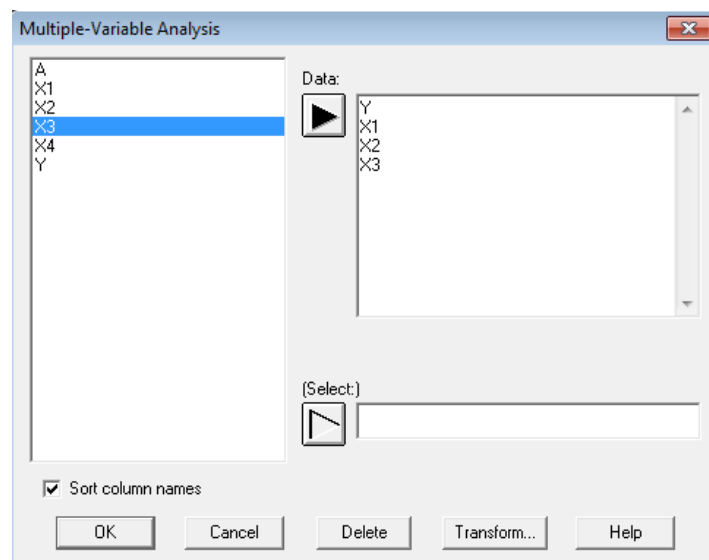


Рис. 4.4. Введення змінних в аналіз

Отримаємо вихідний звіт аналізу множини змінних (рис. 4.5).

Multiple-Variable Analysis

Data variables:

Y
X1
X2
X3

There are 15 complete cases for use in the calculations.

The StatAdvisor

This procedure is designed to summarize several columns of quantitative data. It will calculate various statistics, including correlations, covariances, and partial correlations. Also included in the procedure are a number of multivariate graphs, which give interesting views into the data. Use the Tabular Options and Graphical Options buttons on the analysis toolbar to access these different procedures.

After this procedure, you may wish to select another procedure to build a statistical model for your data. Depending on your goal, one of several procedures may be appropriate. Following is a list of goals with an indication of which procedure would be appropriate:

Рис. 4.5. Вихідний звіт аналізу множини змінних

У звіті аналізу множини змінних вказуються:

- назви змінних, що беруть участь в аналізі – Y, X1, X2, X3;
- кількість повних випадків – 15.

В нижній частині звіту наведена додаткова пояснювальна інформація.

2.2. Натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 4.6) встановіть прапорець таблиці загальної статистики (Summary statistics).

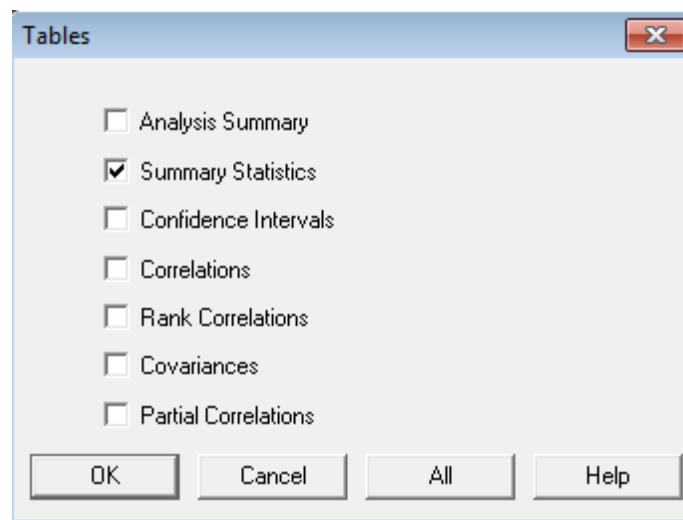


Рис. 4.6. Вікно вибору табличних опцій

За замовчуванням з'явиться наступний набір статистик (рис. 4.7).

Summary Statistics				
	Y	X1	X2	X3
Count	15	15	15	15
Average	301,667	7,02667	15,8333	105,413
Standard deviation	104,322	2,62147	0,662966	2,5125
Coeff. of variation	34,5819%	37,3074%	4,18715%	2,38348%
Minimum	126,0	4,0	14,8	100,0
Maximum	445,0	14,0	17,0	108,9
Range	319,0	10,0	2,2	8,9
Std. skewness	-0,975537	2,11001	0,733035	-1,29409
Std. kurtosis	-0,644381	1,87093	-0,339803	0,424183

Рис. 4.7. Звітна таблиця загальної статистики

Для виклику вікна повного набору статистичних даних Summary Statistics Options (рис. 4.8), клацніть ПКМ в робочому полі документа та виберіть команду параметрів панелі (Pane Options).

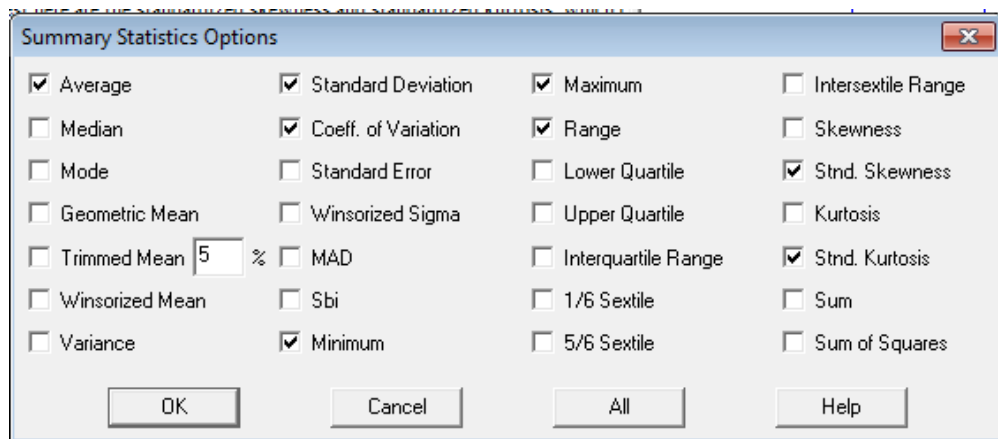


Рис. 4.8. Вікно набору статистичних даних

У вікні Summary Statistics Options можна вибрати додаткові параметри статистичних даних (табл. 4.1).

Таблиця 4.1

Додаткові параметри статистичних даних

Назва статистичної характеристики	Переклад назви	Назва статистичної характеристики	Переклад назви
Average	Середнє	Max	Максимум
Median	Медіана	Range	Розмах
Mode	Мода	Lower quartile	Нижній квартиль
Geo. Mean	Геометричне середнє	Upper quartile	Верхній квартиль
Timmed mean	Залишкове середнє	Interquartile Range	Міжквартильний розмах
Winsorized mean	Вид середнього	1/6 Sextile	1/6 секстіль
Variance	Дисперсія	5/6 Sextile	5/6 секстіль
Std. Deviation	Стандартне відхилення	Intersextile Range	Інтерсекстіальний діапазон
Coeff. of Var.	Коефіцієнт варіації	Skewness	Асиметрія
Std. Error	Стандартна помилка	Std. Skewness	Стандартна асиметрія
Winsorized Sigma	Відхилення від середнього	Kurtosis	Експес
MAD	Середнє абсолютне відхилення	Std. Kurtosis	Стандартний експес
Min	Мінімум	Sum of Squares	Сума квадратів

Summary Statistics

	Y	X1	X2	X3
Count	15	15	15	15
Average	301,667	7,02667	15,8333	105,413
Mode	367,0			104,3
Geometric mean	280,539	6,63204	15,8205	105,385
Standard deviation	104,322	2,62147	0,662966	2,5125
Coeff. of variation	34,5819%	37,3074%	4,18715%	2,38348%
Standard error	26,9358	0,67686	0,171177	0,648725
Minimum	126,0	4,0	14,8	100,0
Maximum	445,0	14,0	17,0	108,9
Range	319,0	10,0	2,2	8,9
Lower quartile	191,0	5,1	15,4	104,3
1/6 sextile	148,0	4,8	15,2	103,5
Std. skewness	-0,975537	2,11001	0,733035	-1,29409
Kurtosis	-0,815085	2,36656	-0,429821	0,536553
Std. kurtosis	-0,644381	1,87093	-0,339803	0,424183
Sum	4525,0	105,4	237,5	1581,2

Рис. 4.9. Звітна таблиця з додатковими статистичними даними

3. Побудова матриці парних коефіцієнтів

3.1. Для отримання матриці парних коефіцієнтів кореляції виберіть у вікні табличних опцій команду кореляція (Correlations) рис. 4.10. На екрані з'явиться матриця парних лінійних коефіцієнтів кореляції (рис. 4.11). Перше число є коефіцієнтом кореляції Пірсона, друге число являє собою число спостережень, а третє - рівень значущості.

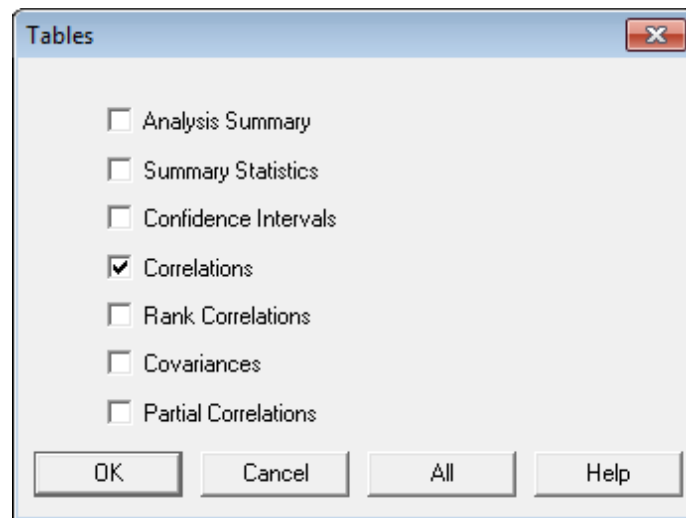


Рис. 4.10. Вікно вибору табличних опцій

Correlations				
	Y	X1	X2	X3
Y		0,4384	0,8097	0,9364
		(15)	(15)	(15)
		0,1021	0,0003	0,0000
X1	0,4384		0,3184	0,5498
	(15)		(15)	(15)
	0,1021		0,2475	0,0337
X2	0,8097	0,3184		0,8102
	(15)	(15)		(15)
	0,0003	0,2475		0,0003
X3	0,9364	0,5498	0,8102	
	(15)	(15)	(15)	
	0,0000	0,0337	0,0003	

Correlation
(Sample Size)
P-Value

Рис. 4.11. Матриця парних лінійних коефіцієнтів кореляції

4. Побудова інтервальних оцінок дисперсій

4.1. Виберіть пункт меню Describe/ Numeric Data/ One–Variable Analysis. З'явиться вікно діалогу для завдання аналізованих змінних (рис. 4.12). Введіть в аналіз змінну Y та натисніть ОК.

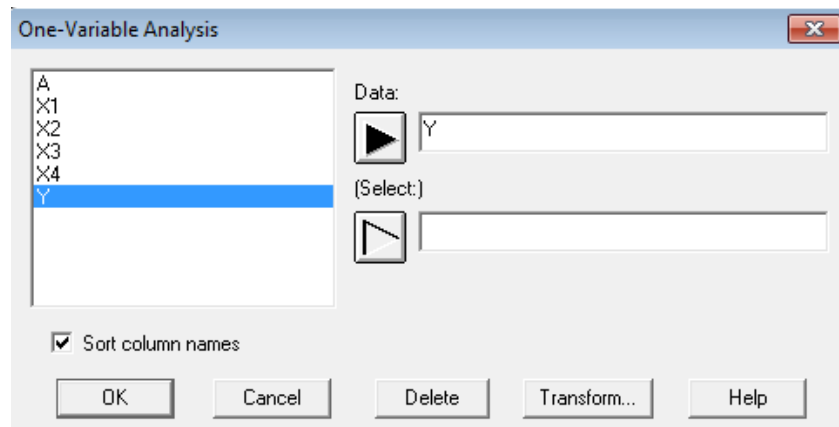


Рис. 4.12. Введення змінних в аналіз

Отримаємо вихідний звіт статистичних даних для змінної Y (рис. 4.13).

Summary Statistics for Y	
Count	15
Average	301,667
Standard deviation	104,322
Coeff. of variation	34,5819%
Minimum	126,0
Maximum	445,0
Range	319,0
Std. skewness	-0,975537
Std. kurtosis	-0,644381

Рис. 4.13. Звіт статистичних даних для змінної Y

Він включає в себе визначення середнього значення, стандартного відхилення, коефіцієнту варіації, мінімуму, максимуму, стандартної асиметрії та стандартного ексцесу. Особливо важливими статистичними показниками є стандартна асиметрія та стандартний ексцес, які можуть бути використані для визначення того, чи виходить зразок із звичайного розподілу. Значення цих статистичних показників поза діапазоном від -2 до +2 вказують на значні відхилення від норми, що призведе до недійсності будь-якої статистичної перевірки щодо стандартного відхилення. З рисунка 4.13 видно, що стандартне значення асиметрії та стандартне значення ексцесу знаходяться в межах очікуваного для даних від звичайного розподілу (-0,975537 та -0,644381 відповідно).

4.2. Натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 4.14) встановіть прапорець таблиці довірчих інтервалів (Confidence Intervals).

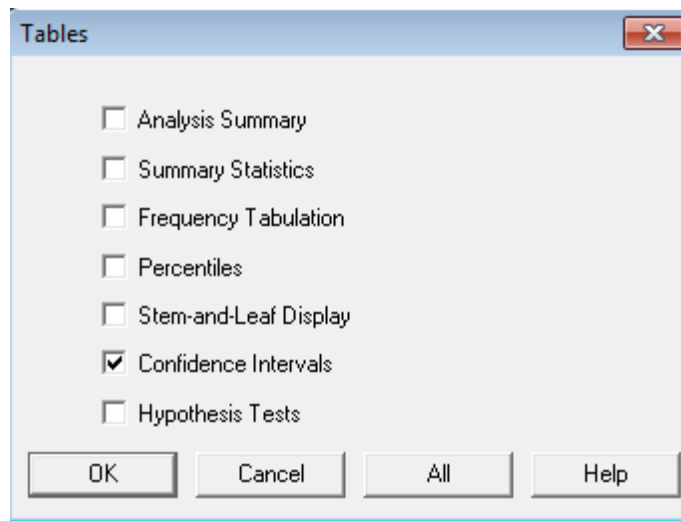


Рис. 4.14. Вікно вибору табличних опцій

Confidence Intervals for Y

95,0% confidence interval for mean: 301,667 +/- 57,7718 [243,895; 359,438]

95,0% confidence interval for standard deviation: [76,377; 164,526]

The StatAdvisor

This pane displays 95,0% confidence intervals for the mean and standard deviation of Y. The classical interpretation of these intervals is that, in repeated sampling, these intervals will contain the true mean or standard deviation of the population from which the data come 95,0% of the time. In practical terms, we can state with 95,0% confidence that the true mean Y is somewhere between 243,895 and 359,438, while the true standard deviation is somewhere between 76,377 and 164,526.

Рис. 4.15. Звіт довірчих інтервальних оцінок для змінної Y

Ця панель відображає 95,0% довірчих інтервалів для середнього та стандартного відхилення Y. Класична інтерпретація цих інтервалів полягає в тому, що при повторній вибірці ці інтервали містять справжнє середнє або стандартне відхилення населення, з якого дані належать 95,0% часу. З практичної точки зору, можна стверджувати з 95,0% упевненістю, що справжнє середнє значення Y знаходиться в межах від 243,895 до 359,438, тоді як справжнє стандартне відхилення знаходиться в межах від 76,377 до 164,526.

Вважають, що населення, з якого виходить зразок, може бути представлено нормальним розподілом. Хоча довірчий інтервал середнього є досить надійним і не дуже чутливим до порушень цього припущення, довірчий інтервал для стандартного відхилення досить чутливий. Якщо дані не походять з нормального розподілу, інтервал для стандартного відхилення може бути неправильним.

4.3. Аналогічним чином отримайте інтервальні оцінки для змінних X1, X2, X3.

5. Перевірка гіпотези

Довірчий інтервал для стандартного відхилення генеральної сукупності можна отримати, перевіряючи гіпотезу $H_0: \sigma_b = \sigma_r$, при конкуруючій гіпотезі $H_a: \sigma_b \neq \sigma_r$.

5.1. В основному вікні DataBook (рис. 4.2) виділіть стовпчик Y, клацнувши по ньому ЛКМ. Виберіть команду Describe/ Numeric Data/ Hypthesis Tests (перевірка гіпотези), на екрані з'явиться діалогове вікно перевірки гіпотези (рис. 4.16).

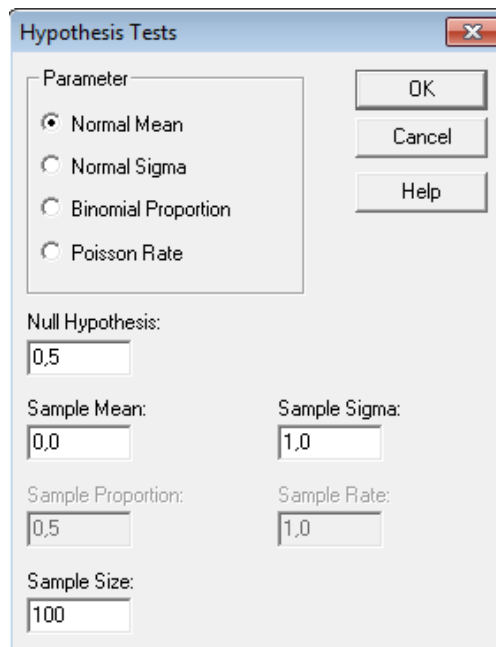


Рис. 4.16. Діалогове вікно перевірки гіпотези

Встановіть:

- параметр Normal Sigma;
- Sample Sigma- значення стандартного відхилення змінної Y зі звіту статистичних даних для Y (рис. 4.13) - 104,32;
- Sample Size- обсяг вибірки 15.

На екрані з'явиться звіт перевірки гіпотези (рис. 4.17).

Hypothesis Tests

Sample standard deviation = 104,32

Sample size = 15

95,0% confidence interval for sigma: [76,3754;164,523]

Null Hypothesis: standard deviation = 0,5

Alternative: not equal

Computed chi-squared statistic = 609429,

P-Value = 0,0

Reject the null hypothesis for alpha = 0,05.

The StatAdvisor

This analysis shows the results of performing a hypothesis test concerning the standard deviation (sigma) of a normal distribution. The two hypotheses to be tested are:

Рис. 4.17. Звіт перевірки гіпотези

Із звіту випливає, що стандартне відхилення σ_T потрапляє в довірчий інтервал (76,3754; 164,523) з надійністю 95%. Нульова гіпотеза $H_0: \sigma_T = 0,5$ відкидається, при рівні значущості $\alpha = 0,05$.

6. Визначення множинної регресії

6.1. Виберіть команду Relate/ Multiple Factors/ Multiple Regression. З'явиться вікно діалогу для завдання аналізованих змінних (рис. 4.18). Введіть спочатку змінну Y у поле Depended Variable (залежна змінна), X1, X2 та X3 в поле Independed Variables (незалежні змінні). Натисніть ОК.

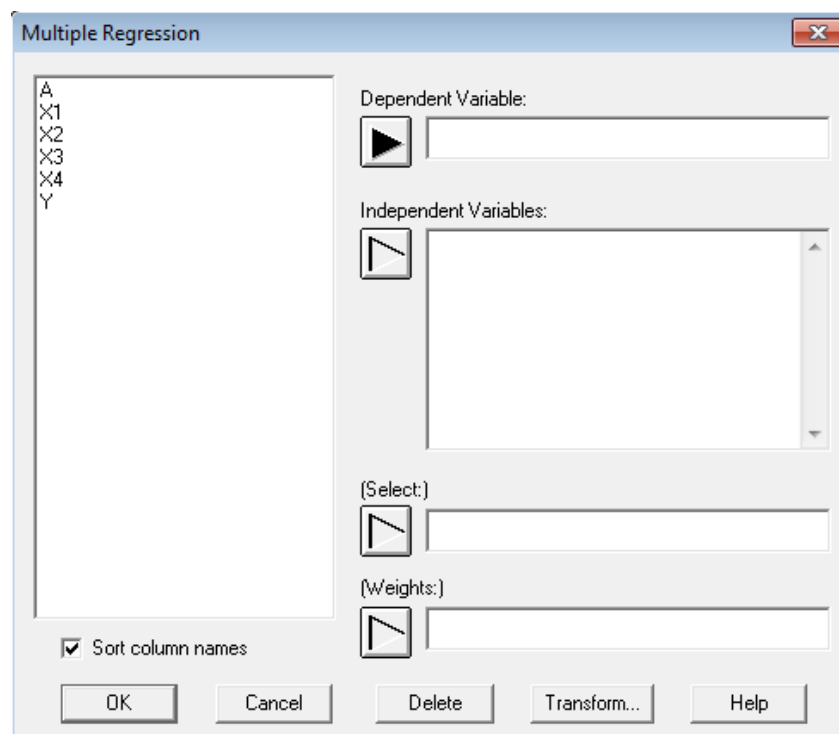


Рис. 4.18. Вікно завдання аналізованих змінних

Отримаємо журнал проведеного аналізу (рис. 4.19).

Multiple Regression - Y

Dependent variable: Y

Independent variables:

X1

X2

X3

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
CONSTANT	-3862,2	568,4	-6,79487	0,0000
X1	-3,51437	4,94623	-0,710513	0,4922
X2	18,2387	27,8708	0,654403	0,5263
X3	36,9952	8,34593	4,43272	0,0010

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	135513,	3	45171,1	29,49	0,0000
Residual	16849,9	11	1531,81		
Total (Corr.)	152363,	14			

R-squared = 88,941 percent

R-squared (adjusted for d.f.) = 85,9249 percent

Standard Error of Est. = 39,1383

Mean absolute error = 25,7918

Durbin-Watson statistic = 1,31464 (P=0,0366)

Lag 1 residual autocorrelation = 0,298105

Рис. 4.19. Звіт множинної регресії

Рівняння заданої моделі:

$$Y = -3862,2 - 3,51437 * X1 + 18,2387 * X2 + 36,9952 * X3$$

Пояснення показників до рисунка 4.19:

- Estimate- оцінка;
- St. Error - ст. помилка;

- T Statistic - Т критерій (статистика);
- P-Value - Р-значення;
- Source- джерело;
- Sumof Squares - сума квадратів;
- Df - ступені свободи;
- Mean Square - середній квадрат;
- F - Ratio - F-критерій;
- R - Squared - коефіцієнт множинної детермінації;
- R squared (adjusted ford. f.) - коефіцієнт детермінації з урахуванням ступенів свободи;
- Standard ErrorifEst - стандартна помилка залишків;
- Meanabsoluteerror - середня абсолютна помилка;
- Durbin – Watson Statistic - статистика Дарбіна-Уотсона.

Аналогічно можна отримати звіти про множинні регресії $Y(X1X2)$, $Y(X1X3)$, $Y(X2X3)$, $Y(X1)$, $Y(X2)$, $Y(X3)$, а також звіт про рівняння трендів $X1(X4)$; $X2(X4)$; $X3(X4)$.

Рівняння парної регресії можна отримати, використовуючи режим множинної регресії, так як аналіз парної регресії не містить статистику Дарбіна-Уотсона.

7. Прогнозний розрахунок

7.1. Виберіть команду Relate/ One Factors/ Simple Regression. З'явиться вікно діалогу для завдання аналізованих змінних (рис. 4.20). Введіть спочатку змінну Y у поле Depended Variable (залежна змінна), X3 в поле Independend Variables (незалежні змінні). Натисніть OK.

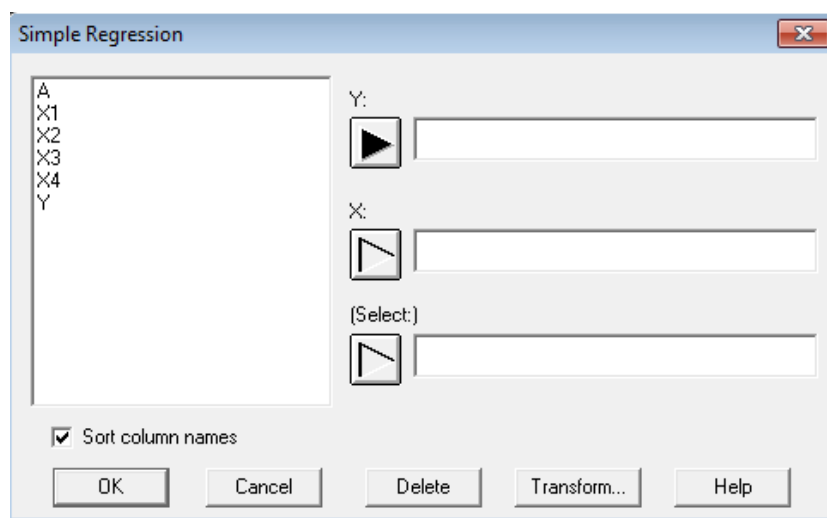


Рис. 4.20. Вікно завдання аналізованих змінних

7.2. Натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 4.21) встановіть прапорець таблиці прогнозування (Forecasts).

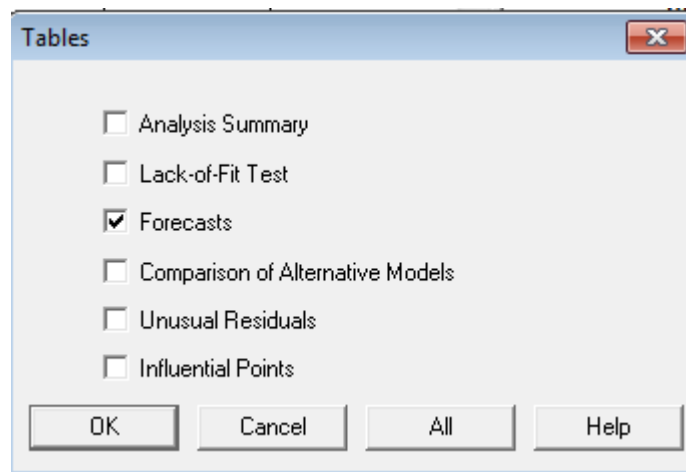


Рис. 4.21. Вікно вибору табличних опцій

На екрані з'являться інтервальні прогнозні значення для $X_3=100$, $X_3=108,9$ (рис. 4.22), за замовчуванням.

Predicted Values					
		95,00%		95,00%	
	Predicted	Prediction	Limits	Confidence	Limits
<i>X</i>	<i>Y</i>	Lower	Upper	Lower	Upper
100,0	91,2051	-5,88616	188,296	39,389	143,021
108,9	437,223	347,119	527,326	400,117	474,328

Рис. 4.22. Звіт інтервальних прогнозних значень

7.3. Для того, щоб отримати прогноз для вибраних вами значень X_3 , наприклад, $X_3=105,6$, $X_3=106$, $X_3=107$, клацніть ПКМ в робочому полі документа та виберіть команду Pane Options. З'явиться вікно для задання значень X_3 (рис. 4.23).

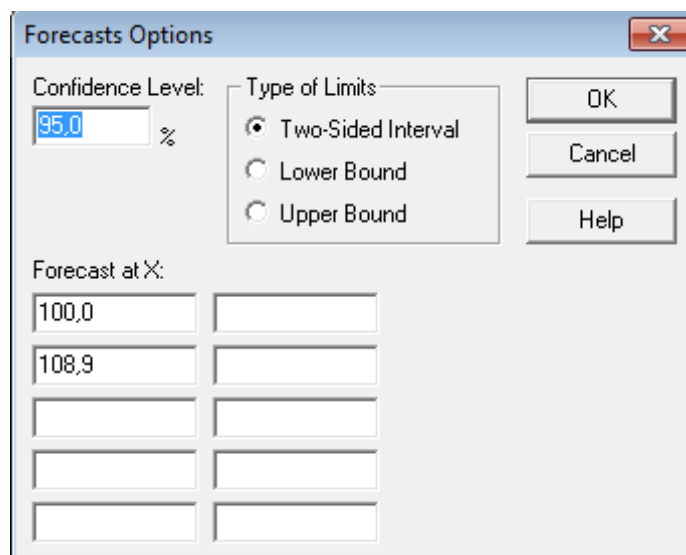


Рис. 4.23. Вікно задання значень прогнозованої змінної

У вільні поля введіть потрібні значення $X=105,6$; $X=106$; $X=107$. Натисніть ОК. З'явиться таблиця прогнозованих значень (рис. 4.24).

Predicted Values					
		95,00%		95,00%	
	<i>Predicted</i>	<i>Prediction</i>	<i>Limits</i>	<i>Confidence</i>	<i>Limits</i>
<i>X</i>	<i>Y</i>	<i>Lower</i>	<i>Upper</i>	<i>Lower</i>	<i>Upper</i>
100,0	91,2051	-5,88616	188,296	39,389	143,021
108,0	402,232	314,473	489,991	371,25	433,214
105,0	285,597	200,719	370,475	264,091	307,102
106,0	324,475	239,519	409,431	302,665	346,286
107,0	363,354	277,428	449,28	338,026	388,682

Рис. 4.24. Таблиця прогнозованих значень

8. Графік прогнозних значень по рівнянню парної регресії

8.1. Натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 4.25) встановіть прапорець Графік встановленої моделі (Plot Fitted Model). Отримаємо наступне зображення (рис. 4.26).

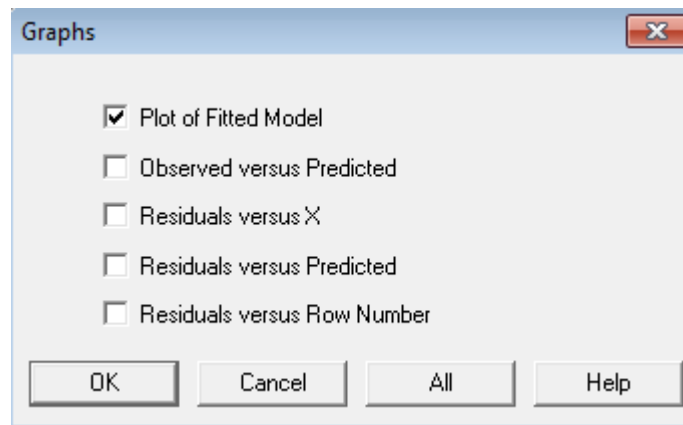


Рис. 4.25. Вікно вибору графічних опцій

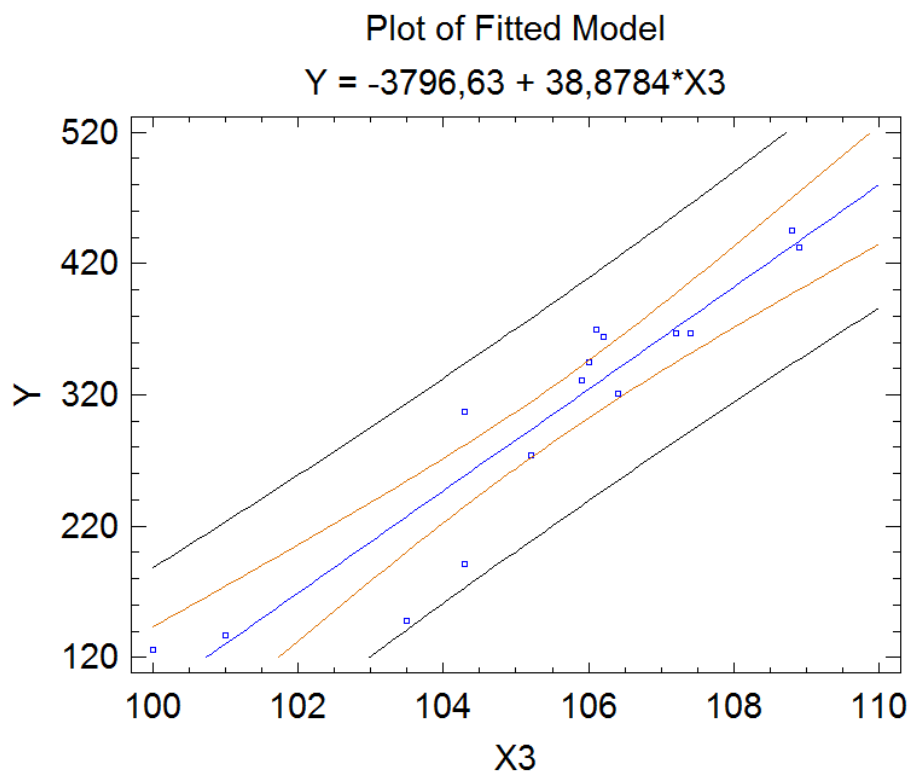


Рис. 4.26. Графік Plot Fitted Model

Контрольні запитання:

1. Що називають регресією?
2. Що є однофакторною регресією?
3. Що називають багатофакторною регресією?
4. Етапи побудови моделей багатофакторної регресії?
5. В чому полягає суть методу крокової регресії?
6. Поясніть суть алгоритму включення фактора.
7. Поясніть суть алгоритму виключення фактора.
8. Що називають довірчим інтервалом?
9. Що називають інтервальною оцінкою параметру?
10. Що називають вибіркою?
11. Що називають генеральною сукупністю?
12. Що називають стандартним відхиленням?
13. Що називають коефіцієнтом кореляції Пірсона?
14. Перерахуйте послідовність дій при регресійному аналізі даних в програмному продукті StatGraphics.

Лабораторна робота №5

Перевірка вірності статистичних гіпотез в програмному продукті StatGraphics

Мета роботи: навчитися виконувати перевірку статистичних гіпотез застосовуючи: F-критерій Фішера, t-критерій Стюдента та критерій Манна-Уїтні; побудову графіків в програмному продукті StatGraphics.

Теоретичні відомості

При проведенні дослідження часто формуються дві групи, наприклад, контрольна та експериментальна. В цьому випадку говорять про дві незалежні вибірки.

Для перевірки статистичної гіпотези про рівність дисперсій генеральних сукупностей $\sigma_x^2 = \sigma_y^2$ застосовується F-критерій Фішера.

Умови застосування F-критерію: вибірки незалежні та отримані з нормально розподілених генеральних сукупностей з параметрами $\mu_x, \sigma_x, \mu_y, \sigma_y$.

Гіпотеза: $H_0 : \sigma_x^2 = \sigma_y^2$.

Альтернативна гіпотеза: $H_1 : \sigma_x^2 \neq \sigma_y^2$.

F-критерій Фішера розраховується за формулою:

$$F = \frac{S_1^2}{S_2^2},$$

де: S_1^2 та S_2^2 є відповідно, більшою та меншою з вибірових дисперсій S_x^2 та S_y^2 .

Після розрахунку значення F порівнюється з критерієм $F_{кр}$, який отримують з довідкових таблиць. Якщо $F_{табл} \geq F_{кр}$, то гіпотеза про рівність дисперсій генеральних сукупностей відкидається на рівні значимості α .

Для перевірки статистичної гіпотези про рівність середніх арифметичних генеральних сукупностей застосовується t-критерій Стюдента.

Умови застосування t-критерію: вибірки незалежні та отримані з генеральних сукупностей X та Y, які мають нормальний розподіл з параметрами: $\mu_x, \sigma_x, \mu_y, \sigma_y$.

Нульова гіпотеза: $H_0 : \mu_x = \mu_y$.

Альтернативна гіпотеза: $H_1 : \mu_x \neq \mu_y$.

Значення t-критерію Стюдента розраховується за формулою:

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}},$$

де: S_x^2 та S_y^2 є вибіровими дисперсіями вибірок;

n_x та n_y - їх об'єми.

Після розрахунку значення $t_{табл}$ порівнюється з $t_{кр}$, отриманим з довідкових таблиць. Якщо $t_{табл} \geq t_{кр}$, то гіпотеза про рівність генеральних середніх відкидається на рівні значимості α .

Для перевірки статистичної гіпотези про рівність медіан генеральних сукупностей застосовуються критерії Манна-Уїтні та Вілкоксона. Історично склалося так, що критерій Вілкоксона (W) був одним із перших рангових критеріїв. Він був запропонований Вілкоксоном у 1945 році для порівняння вибірок однакового об'єму. У 1947 році цей критерій був модифікований Манном та Уїтні та розповсюджений на випадок вибірок довільних об'ємів.

Умови застосування: вибірки отримані із однотипних неперервних розподілів. При цьому закон розподілу генеральних сукупностей може бути будь-яким.

Гіпотеза: $H_0 : Me_x = Me_y$.

Альтернативна гіпотеза: $H_1 : Me_x \neq Me_y$.

Критерії Манна-Уїтні U та Вілкоксона W мають багато спільного. Принцип розрахунку пояснюється на нижченаведеному прикладі. Результати бігу на 100 м у контрольній (X) та експериментальній (Y) групах юнаків наведені в таблиці 5.1.

Таблиця 5.1

Результати бігу на 100 м

Групи	Результати бігу на 100 м, с									
Контр.	12,6	12,6	12,6	12,6	12,6	12,6	12,6	12,6	12,6	12,6
Експ.	11,3	11,3	11,3	11,3	11,3	11,3	11,3	11,3	11,3	11,3

Порядок розрахунку:

1. Об'єднуємо дві вибірки X та Y та ранжуємо їх за зростанням (табл. 5.2); при цьому відзначаємо дані, що відносяться до вибірки Y (стовпчик 3).
2. Присвоюємо ранги об'єднаній вибірці (стовпчик 4).
3. Після цього додаємо окремо ранги, що відносяться до кожної з вибірок. Менша із сум (83,5) приймається в якості критерію $W_{табл.}$. Його значення порівнюється з критичним $W_{кр.}$. Якщо $W_{табл.} \leq W_{кр.}$, то нульова гіпотеза відкидається – різниця вважається статистично значимою на рівні значимості α .

Таблиця 5.2

Розрахунок критеріїв Вілкоксона та Манна-Уїтні

X	Y	Об'єднана вибірка	Ранги
1	2	3	4
12,6	11,3	11,3	1
12,3	12,8	11,4	2
11,8	12,2	11,7	3
12,1	11,7	11,8	4,5
12,8	12,4	11,8	4,5
13,2	13,3	12,0	6,5
13,8	11,4	12,0	6,5
12,0	12,0	12,1	8
12,6	11,8	12,2	9
13,0	12,5	12,3	10
		12,4	11
		12,5	12

1	2	3	4
		12,6	13,5
		12,6	13,5
		12,8	15,5
		12,8	15,5
		13,0	17
		13,2	18
		13,3	19
		13,8	20
		Сума рангів	83,5
		Сума рангів	126,5

Після цього критерій Манна-Уїтні U розраховується за формулою:

$$U = W - \frac{1}{2}m(m+1),$$

де: W - значення критерію Вілкоксона;

n, m - об'єми вибірок.

В даному прикладі об'єми вибірок рівні. Тепер розраховуємо значення U критерію Манна-Уїтні:

$$U = 83,5 - \frac{1}{2} \cdot 10 \cdot (10 + 1) = 28,5$$

Отримане значення $U_{табл}$ порівнюється з критичним. Якщо $U_{табл} \leq U_{кр}$ – нульова гіпотеза відкидається на рівні значимості α . Різниця результатів вважається статистично значимою.

Порядок виконання роботи

1. Задання змінних та введення даних

Необхідно перевірити гіпотези про те, що: варіативність зросту хлопців 15 років така ж, як і дівчат того ж віку; зріст 15-річних дівчат менше зросту 15-річних хлопців.

Примітка: розподіл зросту дівчат і хлопців 15 років вважати таким, що відповідає нормальному закону.

1.1. В програмному продукті StatGraphics необхідно створити таблицю даних на основі прикладу (Додаток 5). Для задання назв змінних величин, подвійним клацанням ЛКМ на назві стовпчика, відкрийте вікно зміни назви Modify Column (модифікувати колонку). В строчку Name (Ім'я) даного вікна введіть назви стовпчиків (div10 – зріст 10-річних дівчат, hlor10 – зріст 10-річних хлопців, div12 – зріст 12-річних дівчат, hlor12 – зріст 12-річних хлопців, div15 – зріст 15-річних дівчат, hlor15 – зріст 15-річних хлопців).

1.2. Скопіюйте дані з таблиці (Додаток 5).

	div10	hlop10	div12	hlop12	div15	hlop15
1	146	142	153	142	179	167
2	138	130	153	160	171	174
3	144	143	149	155	175	163
4	142	142	155	153	163	175
5	141	136	164	132	153	176
6	144	144	151	160	163	160
7	145	145	155	154	160	158
8	142	142	153	154	167	159
9	135	147	159	158	167	160
10	136	140	156	147	170	164
11	142	148	158	145	168	161
12	143	137	158	144	159	175
13	138	136	160	147	157	164
14	143	151	156	158	165	168
15	146	129	164	163	162	160
16	150	142	155	153	170	165
17	139	140	159	144	157	166

Рис. 5.1. Таблиця вхідних даних

1.3. Збережіть створену таблицю, вибравши File – Save – Save Data File.

2. Порівняння двох вибірок

Так як розподіл зросту дівчат і хлопців 15 років відповідає нормальному закону, немає ніяких перепон для застосування F-критерію Фішера та t-критерію Стьюдента. У той же час застосування критерію Манна-Уїтні не висуває ніяких вимог до характеру розподілу результатів, тому він може бути використаний у будь-яких випадках, якщо необхідно порівняти дві вибірки.

2.1. Виберіть пункт меню Compare/ Two Samples/ Independent Samples. З'явиться вікно діалогу для завдання порівнюваних вибірок (рис. 5.2).

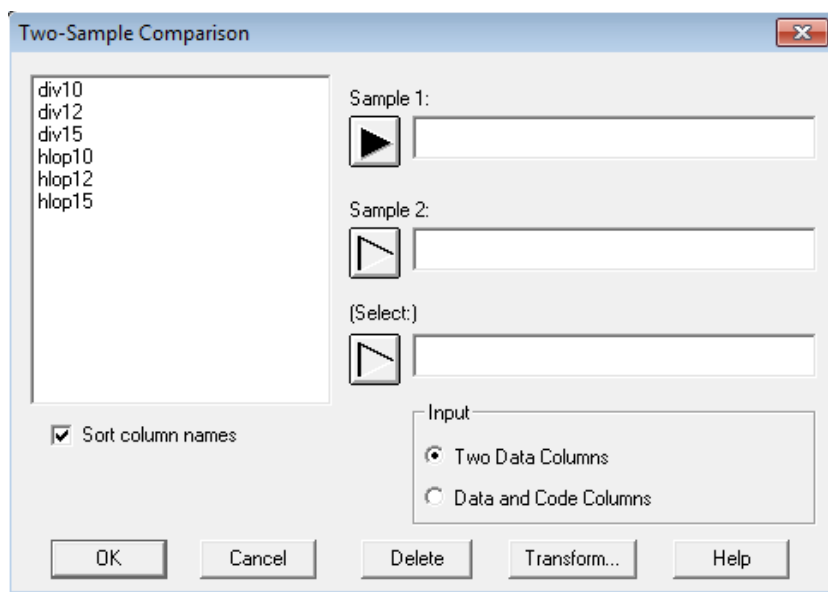


Рис. 5.2. Вікно порівнюваних вибірок

Виділіть змінну div15 клацанням по ній лівою кнопкою миші, та клацніть по кнопці зі стрілкою, таким чином вводячи дану змінну в поле Sample 1. У такий же спосіб введіть в аналіз змінну hlop15 в поле Sample 2 (рис. 5.3) та натисніть ОК.

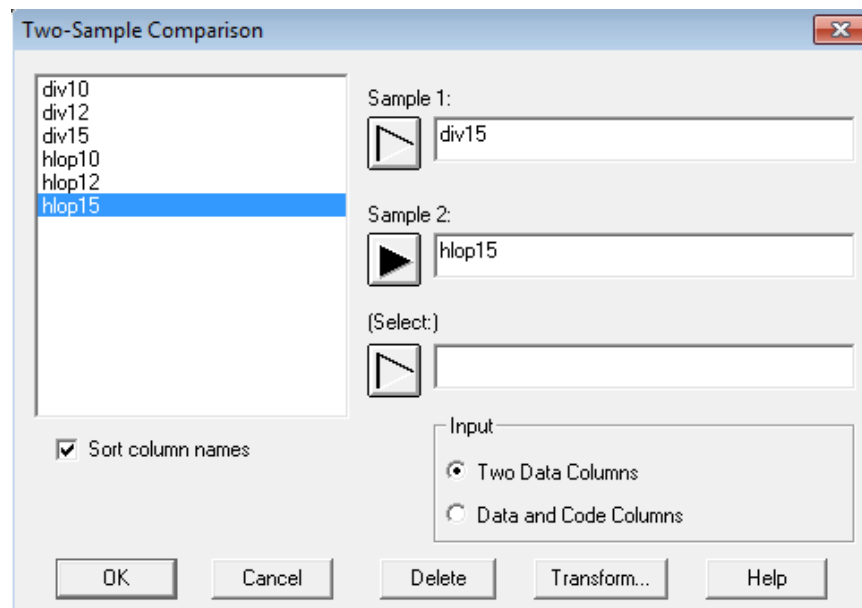


Рис. 5.3. Введення змінних в аналіз

Отримаємо вихідний звіт порівняння двох вибірок (рис. 5.4).

Two-Sample Comparison - div15 & hlop15

Sample 1: div15

Sample 2: hlop15

Sample 1: 50 values ranging from 147,0 to 179,0

Sample 2: 50 values ranging from 156,0 to 182,0

The StatAdvisor

This procedure is designed to compare two samples of data. It will calculate various statistics and graphs for each sample, and it will run several tests to determine whether there are statistically significant differences between the two samples.

Рис. 5.4. Вихідний звіт порівняння двох вибірок

У звіті порівняння двох вибірок вказуються:

- назви змінних, що порівнюються – div15, hlop15;
- кількість елементів вибірки – 50;
- діапазон значень вибірок - div15 – від 147 до 179, hlop15 – від 156 до 182.

В нижній частині звіту наведена додаткова пояснювальна інформація.

3. Розрахунок F-критерію Фішера

3.1. Натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 5.5) встановіть прапорець таблиці порівняння стандартних відхилень (Comparison of Standard Deviation).

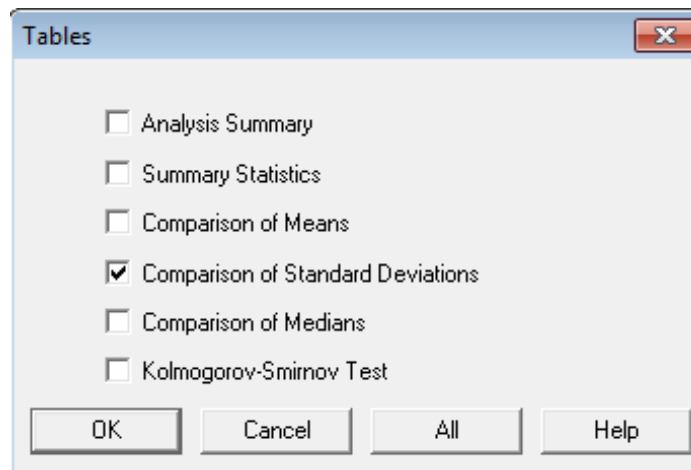


Рис. 5.5. Вікно вибору табличних опцій

З'явиться таблиця порівняння стандартних відхилень (рис. 5.6).

Comparison of Standard Deviations

	<i>div15</i>	<i>hlop15</i>
Standard deviation	5,72499	6,28649
Variance	32,7755	39,52
Df	49	49

Ratio of Variances = 0,82934

95,0% Confidence Intervals

Standard deviation of div15: [4,78228; 7,1341]

Standard deviation of hlop15: [5,25132; 7,83381]

Ratio of Variances: [0,470631; 1,46145]

F-test to Compare Standard Deviations

Null hypothesis: $\sigma_1 = \sigma_2$

Alt. hypothesis: $\sigma_1 \neq \sigma_2$

F = 0,82934 P-value = 0,51496

Do not reject the null hypothesis for $\alpha = 0,05$.

The StatAdvisor

This option runs an F-test to compare the variances of the two samples. It also constructs confidence intervals or bounds for each standard deviation and for the ratio of the variances. Of particular interest is the confidence interval for the ratio of the variances, which extends from 0,470631 to 1,46145.

Since the interval contains the value 1, there is not a statistically significant difference between the standard deviations of the two samples at the 95,0% confidence level.

Рис. 5.6. Звітна таблиця порівняння стандартних відхилень

У вікні Comparison of Standard Deviation наведено:

- значення стандартного відхилення (Standard deviation), дисперсії (Variance) та числа ступенів свободи (Df) для досліджуваних змінних;
- довірчі інтервали (Confidence Intervals) для стандартного відхилення;
- значення F-критерію Фішера, при різних альтернативних гіпотезах. У даному випадку немає підстав припускати, що значення дисперсій відрізняються, тому вибираємо першу альтернативу (1). Значення критерію Фішера $F=0,82934$ та $P\text{-value}=0,51496$ (експериментально розрахованої ймовірності помилки) свідчать про те, що між дисперсіями немає достовірних відмінностей, $P>0,05$. Відповідно, варіативність зросту 15-річних хлопців така ж, як і 15-річних дівчат.

4. Розрахунок t-критерію Стюдента

4.1. Натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 5.7) встановіть прапорець таблиці порівняння середніх арифметичних (Comparison of Means).

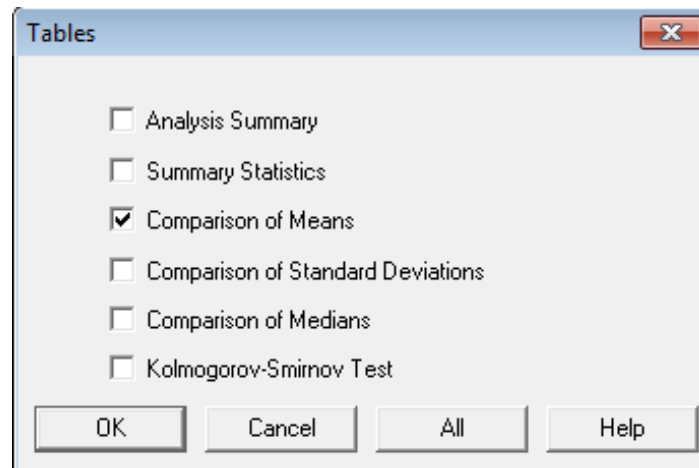


Рис. 5.7. Вікно вибору табличних опцій

З'явиться таблиця порівняння середніх арифметичних (рис. 5.8).

Comparison of Means

95,0% confidence interval for mean of div15: 164,0 +/- 1,62703 [162,373; 165,627]
95,0% confidence interval for mean of hlop15: 168,52 +/- 1,78661 [166,733; 170,307]
95,0% confidence interval for the difference between the means
assuming equal variances: -4,52 +/- 2,38625 [-6,90625; -2,13375]

t test to compare means

Null hypothesis: mean1 = mean2
Alt. hypothesis: mean1 NE mean2
assuming equal variances: t = -3,75896 P-value = 0,000290116
Reject the null hypothesis for alpha = 0,05.

The StatAdvisor

This option runs a t-test to compare the means of the two samples. It also constructs confidence intervals or bounds for each mean and for the difference between the means. Of particular interest is the confidence interval for the difference between the means, which extends from -6,90625 to -2,13375. Since the interval does not contain the value 0, there is a statistically significant difference between the means of the two samples at the 95,0% confidence level.

Рис. 5.8. Звітна таблиця порівняння середніх арифметичних

У вікні Comparison of Means наведено довірчі інтервали для математичного очікування зросту дівчат та хлопців 15 років, та їх різниці. Після цього наведені значення t-тесту ($t=3,75896$) та P-value для альтернативної гіпотези – mean1 NE mean2 (генеральні середні не рівні) - P-value=0,000290116.

4.2. Для того, щоб встановити умову іншої альтернативної гіпотези, клацніть ПКМ в робочому полі документа та виберіть команду Pane Options. З'явиться вікно опцій (рис. 5.9).

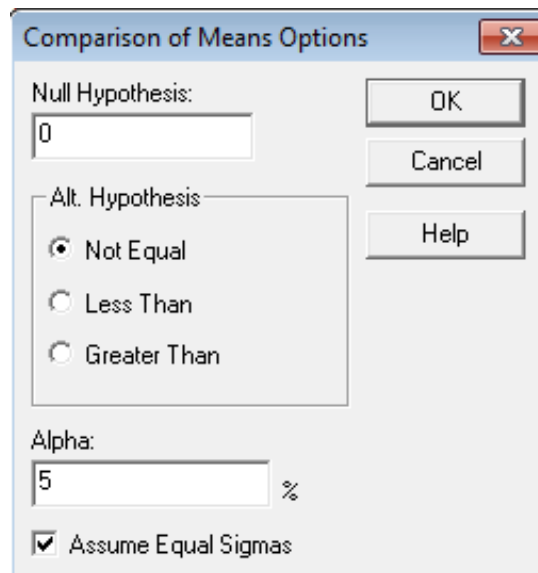


Рис. 5.9. Вікно опцій альтернативних гіпотез

Встановіть прапорець альтернативної гіпотези на значення Less Than (перша генеральна середня менше другої) та натисніть ОК.

Comparison of Means
 95,0% confidence bound for mean of div15: 164,0 + 1,3574 [165,357]
 95,0% confidence bound for mean of hlop15: 168,52 + 1,49053 [170,011]
 95,0% confidence bound for the difference between the means
 assuming equal variances: -4,52 + 1,99675 [-2,52325]

t test to compare means
 Null hypothesis: mean1 = mean2
 Alt. hypothesis: mean1 < mean2
 assuming equal variances: t = -3,75896 P-value = 0,000145058
 Reject the null hypothesis for alpha = 0,05.

Рис. 5.10. Звітна таблиця порівняння середніх арифметичних (друга альтернативна гіпотеза)

З рис. 5.10 видно, що при значенні першої генеральної середньої меншої за другу, значення P-value для альтернативної гіпотези $P\text{-value}=0,000145058$.

4.3. Клацніть ПКМ в робочому полі документа та виберіть команду Pane Options. Встановіть прапорець альтернативної гіпотези на значення Greater Than (перша генеральна середня більша за другу) та натисніть ОК. В даному випадку значення P-value буде рівним 0,999855.

Обираємо другу альтернативу (так, як припускаємо, що генеральне середнє, що характеризує зріст дівчат менше генерального середнього, що характеризує зріст хлопців). Після цього робимо висновок: так як P-value дуже малий ($P\text{-value}=0,000145$), гіпотеза про рівність середніх відкидається. Отже, зріст 15-річних дівчат достовірно менший за зріст 15-річних хлопців, $P<0,001$.

5. Розрахунок критерію Манна-Уїтні

5.1. Натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 5.11) встановіть прапорець таблиці порівняння медіан (Comparison of Medians).

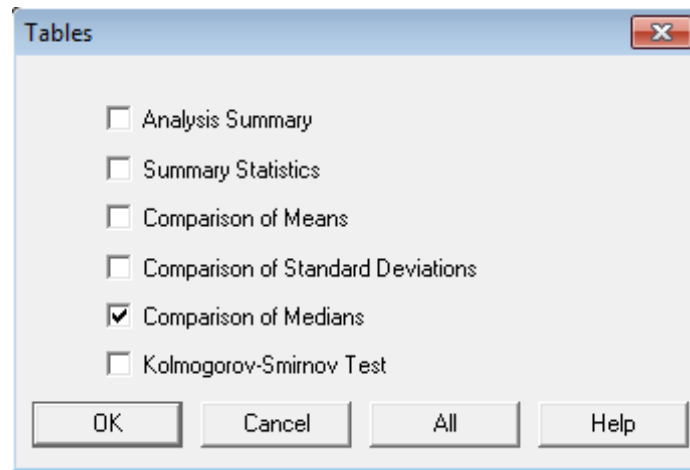


Рис. 5.11. Вікно вибору табличних опцій

З'явиться таблиця порівняння медіан (рис. 5.12).

Comparison of Medians
Median of sample 1: 163,5
Median of sample 2: 169,0

Mann-Whitney (Wilcoxon) W test to compare medians
Null hypothesis: median1 = median2
Alt. hypothesis: median1 NE median2

Average rank of sample 1: 40,69
Average rank of sample 2: 60,31

W = 490,5 P-value = 0,000715869
Reject the null hypothesis for alpha = 0,05.

Рис. 5.12. Звітна таблиця порівняння медіан

У вікні Comparison of Medians наведені значення медіани першої вибірки (div15) та другої вибірки (hlor15). Медіани відповідно рівні 163,5 см та 169,0 см. Після цього сформульована нульова гіпотеза, перевіряється рівність медіан генеральних сукупностей ($\text{median1}=\text{median2}$). Після формулювання нульової гіпотези приводяться величини середніх рангів для першої вибірки (40,69) та другої вибірки (60,31). Далі наведені значення критерію Манна-Уїтні $W=490,5$ та P-value для альтернативної гіпотези: $\text{median1 NE median2}$ (медіани не рівні).

За допомогою панелі Pane Options перевірте дві інші альтернативні гіпотези (Less Than, Greater Than). Так як при альтернативній гіпотезі ($\text{median1}<\text{median2}$) P-value дуже малий ($P\text{-value}=0,000357935$), гіпотеза про рівність медіан відкидається. Отже, зріст 15-річних дівчат достовірно менший за зріст 15-річних хлопців, $P<0,001$.

6. Побудова графіків

6.1. Натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 5.13) встановіть прапорець гістограма (Frequency Histogram).

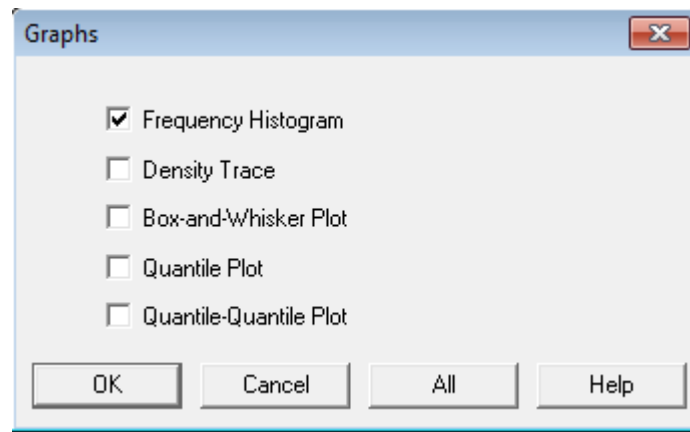


Рис. 5.13. Вікно графічних опцій

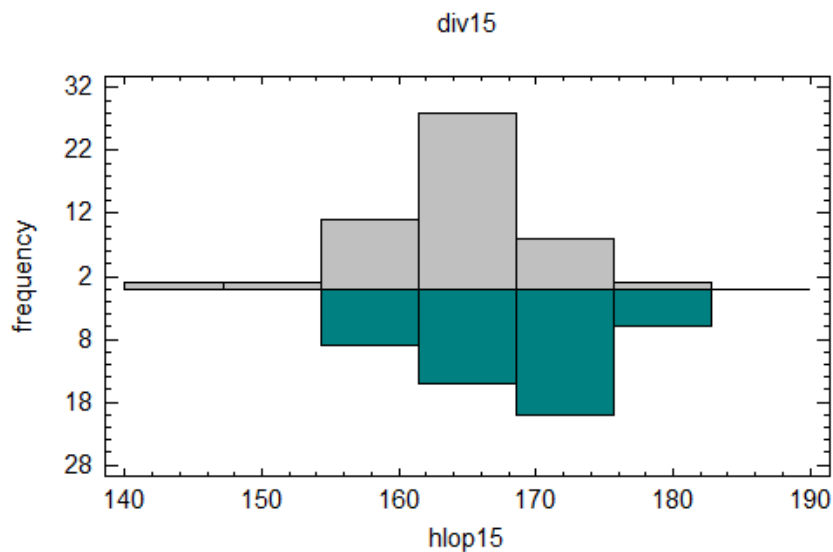


Рис. 5.14. Гістограма розподілу вибірок

На рисунку 5.14 представлена гістограма розподілу двох вибірок. При цьому одна з них представлена у інверсному вигляді. Таке представлення кінцевих результатів зручно використовувати при необхідності демонстрації форми розподілів.

6.2. Натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 5.13) встановіть прапорець графіку щільності ймовірностей розподілу (Density Trace). Цей графік (рис. 5.15) демонструє щільність ймовірності розподілу двох вибірок. Чим ближче графіки один до одного, тим менше відмінностей між вибірками і навпаки.

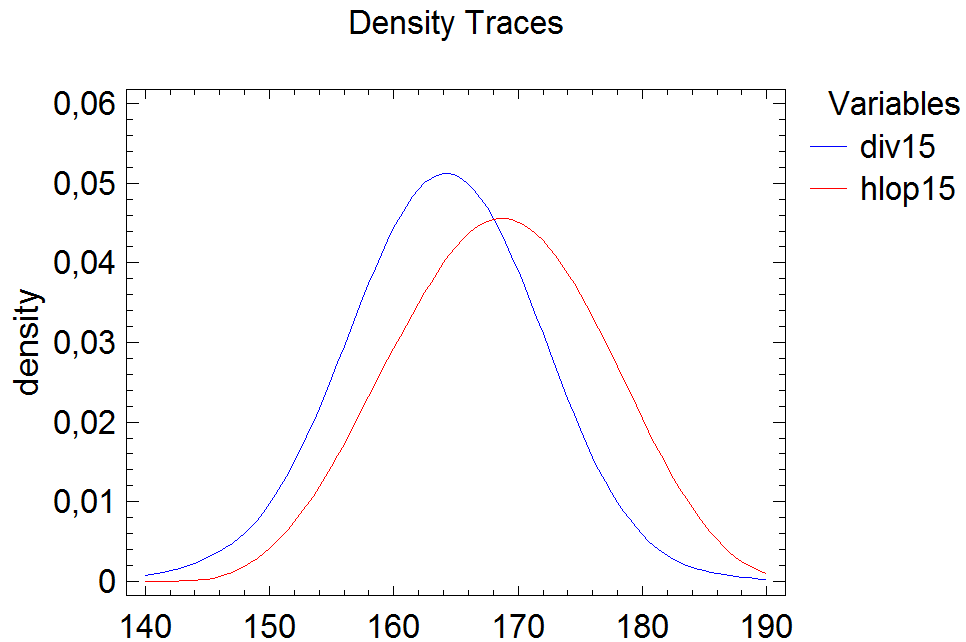


Рис. 5.15. Графік щільності ймовірностей розподілу

6.3. Натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 5.13) встановіть прапорець графіку «ящик з вусами» (Box-and-Whisker Plot). Даний графік являє собою прямокутник. Всередину цього прямокутника потрапляє 50% всіх результатів. Лінія всередині цього прямокутника відповідає медіані. Значення середнього арифметичного позначено хрестиком. «Вуса» відповідають максимальному та мінімальному значенню. Виключення становлять випадки відхилення значень від меж прямокутника на півтори довжини (позначаються на діаграмі квадратиками). Ці значення наносяться на графік окремо. На рис. 5.16 видно, що в групі дівчат екстремальні значення виходять за межі, що відповідають 1,5 довжині прямокутника. У групі хлопців таких значень немає. Цей графік наглядно демонструє відмінність вибірок: чим менший перетин проєкцій графіків на вісь абсцис, тим більше різниця між вибірками.



Рис. 5.16. Графік Box-and-Whisker Plot

6.4. Натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 5.13) встановіть прапорець графіку кумулят (Quantile Plot). На цьому графіку (рис. 5.17) представлені кумуляти, що побудовані на основі даних двох вибірок. Чим більше відмінностей між вибірками, тим далі одна від одної криві.

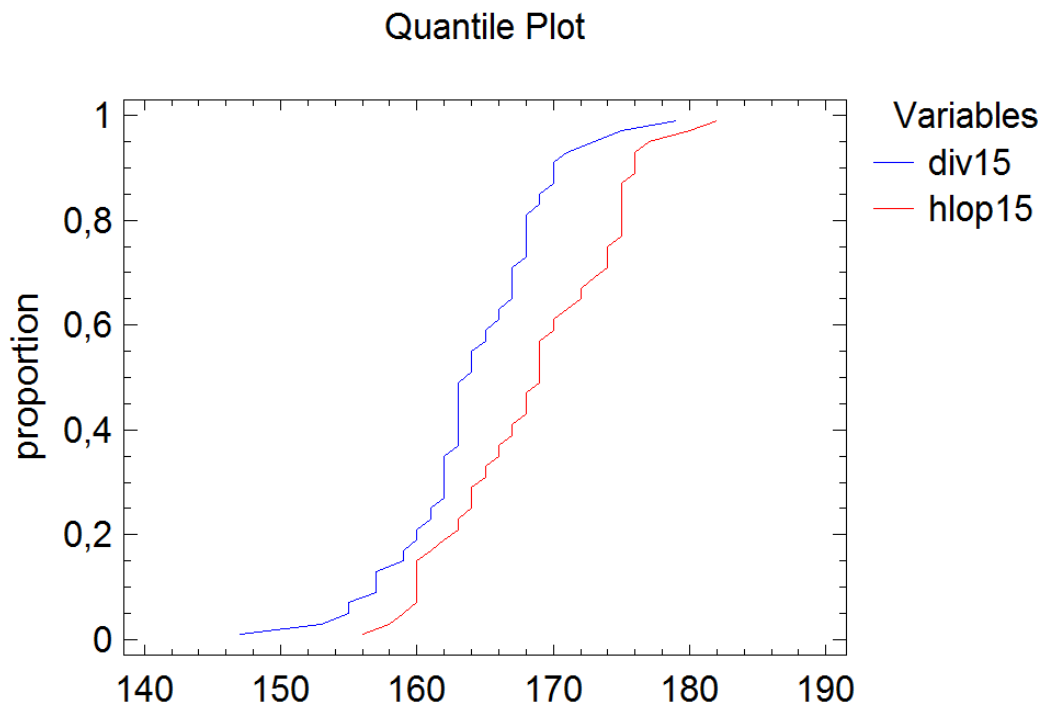


Рис. 5.17. Графік кумулят

6.5. Натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 5.13) встановіть прапорець графіку кумулянт (Quantile-Quantile Plot). Для побудови цього графіка ранжуються значення змінних (div15, hlor15), та за зображуються на координатній площині. На графіку добре видно перевищення результатів однієї вибірки над іншою. Якщо результати у вибірках дещо відрізняються один від одного, точки розташовуються на бісектрисі першого координатного кута. В даному випадку (рис. 5.18) видно, що зріст хлопців 15 років більше зросту їх ровесниць.

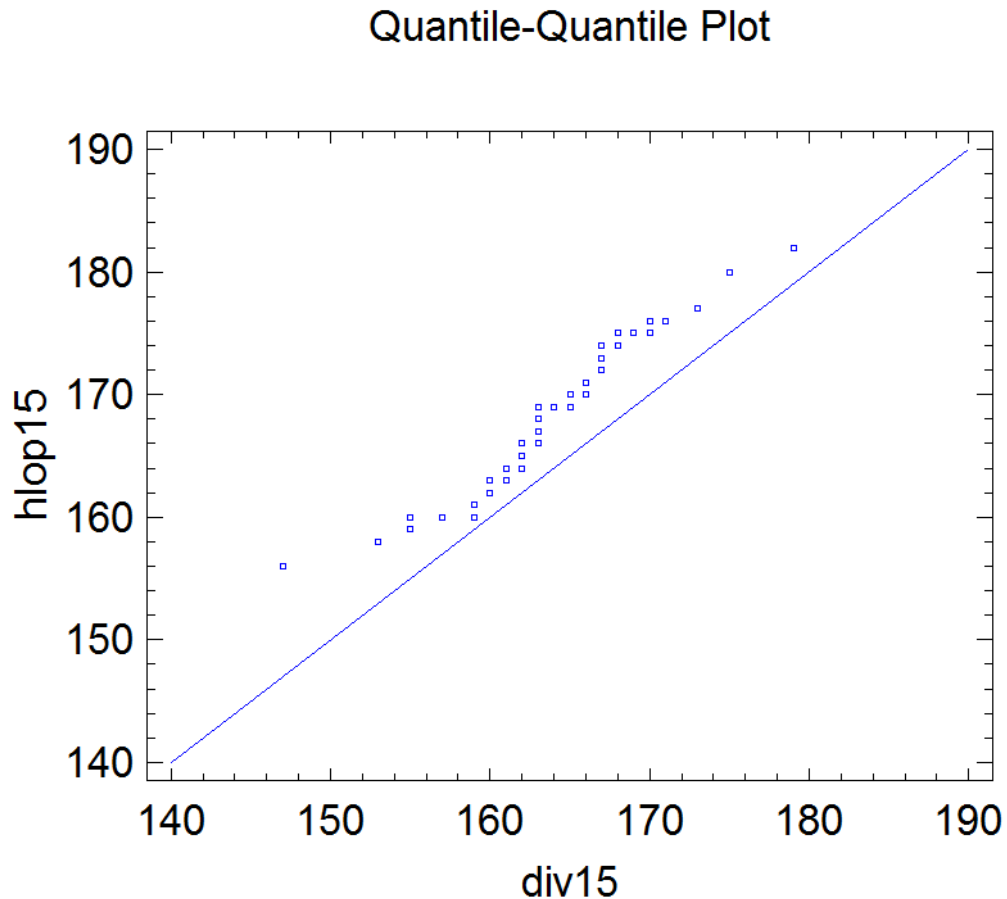


Рис. 5.18. Графік Quantile-Quantile Plot

7. Індивідуальне завдання

7.1. На основі даних таблиці (Додаток 5) перевірте гіпотези про те, що:

- варіативність зросту хлопців 10 років така ж, як і дівчат того ж віку;
- зріст 10-річних дівчат менше зросту 10-річних хлопців;
- варіативність зросту хлопців 12 років така ж, як і дівчат того ж віку;
- зріст 12-річних дівчат менше зросту 12-річних хлопців.

Вважати розподіл зросту дівчат і хлопців таким, що відповідає нормальному закону.

Виконайте:

- порівняння вибірок;
- розрахунок F-критерію Фішера;
- розрахунок t-критерію Стюдента;

- розрахунок критерію Манна-Уїтні;
- побудову графіків.

7.2. Збережіть весь виконаний аналіз. Для цього виберіть File – Save As – Save StatFolio As.

Контрольні запитання:

1. З якою метою застосовують F-критерій Фішера та які умови його застосування?
2. Як формулюються статистичні гіпотези при застосуванні F-критерію Фішера, який порядок обчислення критерію?
3. З якою метою застосовується t-критерій Стьюдента та які умови його застосування?
4. Як формулюються статистичні гіпотези при застосуванні t-критерію Стьюдента, який порядок обчислення критерію?
5. З якою метою застосовується критерій Манна-Уїтні та які умови його застосування?
6. У чому різниця критеріїв Вілкоксона та Манна-Уїтні?
7. Як формулюються статистичні гіпотези при застосуванні критерію Манна-Уїтні, який порядок обчислення критерію?
8. Як можна графічно представити результати аналізу?

Лабораторна робота №6

Перевірка пов'язаних вибірок, початок кореляційного аналізу в програмному продукті StatGraphics

Мета роботи: навчитися виконувати розрахунок t-критерію Стюдента для пов'язаних вибірок, коефіцієнту кореляції Брава-Пірсона та Спірмена, будувати графіки кореляційного поля.

Теоретичні відомості

При проведенні досліджень застосовується порівняння парних вибірок, отриманих до та після експерименту; у такому випадку вибірки вважаються пов'язаними. Для оцінки вірності різниці між такими вибірками застосовується модифікований t-критерій Стюдента, а також непараметричні критерії: знаків, знакових рангів Вілкоксона та ін.

При застосуванні t-критерію Стюдента гіпотеза про рівність нулю середнього арифметичного генеральної сукупності формулюється по відношенню до різниць пов'язаних пар спостережень.

Умови застосування: утворюють $d_i = x_i - y_i$ – різниці пов'язаних пар результатів вимірювань та передбачається нормальний розподіл цих різниць в генеральній сукупності з параметрами μ_d, σ_d

Гіпотеза: $H_0 : \mu_d = 0$.

Альтернативна гіпотеза: $H_1 : \mu_d \neq 0$.

Значення t-критерію Стюдента розраховується за формулою:

$$t = \frac{\bar{d}}{S_d} \sqrt{n},$$

де: $\bar{d}$ - середнє арифметичне різниць;

S_d - стандартне відхилення;

n - об'єм вибірки.

Після розрахунку значення $t_{\text{табл}}$ порівнюється з критичним $t_{\text{кр}}$, отриманим з довідкових таблиць. Якщо $t_{\text{табл}} \geq t_{\text{кр}}$, то гіпотеза про рівність генерального середнього нулю відкидається на рівні значимості α .

Критерій знаків Z вважається одним із найпростіших. Попарно порівнюються вибіркові значення: перевищення значень однієї вибірки позначається «+», у протилежному випадку «-» (однакові значення відкидаються). Ідея критерію знаків Z наступна: якщо попарно порівнювані вибірки не відрізняються, то кількість плюсів та мінусів виявиться приблизно однаковою. В протилежному випадку на вибіркові значення виявляє вплив якийсь невідомий фактор.

Умови застосування: всі z_i взаємно незалежні та мають рівні нулю медіани.

Гіпотеза: $H_0 : Me_z = 0$.

Альтернативна гіпотеза: $H_1 : Me_z \neq 0$.

Значенням Z критерія знаків є найбільша кількість виставлених символів «+» або «-». Після цього значення $Z_{табл}$ порівнюється з критичним $Z_{кр}$, отриманим з довідкових таблиць. Якщо $Z_{табл} \geq Z_{кр}$, то нульова гіпотеза відкидається.

Критерій знакових рангів Вілкоксона застосовується при перевірці гіпотези про різницю двох пов'язаних вибірок.

Умови застосування: цей критерій можна застосовувати для будь-якого розподілу пов'язаних вибірок.

Гіпотеза: $H_0 : Me_d = 0$.

Альтернативна гіпотеза: $H_1 : Me_d \neq 0$.

Принцип розрахунку критерію знакових рангів Вілкоксона пояснюється на нижченаведеному прикладі. Маса тіла жінок до (masado) та після (masapi) експерименту наведені в таблиці 6.1.

Таблиця 6.1

Розрахунок критерію знакових рангів Вілкоксона

№	masado	masapi	d_i	Ранги, $ d_i $
1	2	3	4	5
1	59,1	58,0	1,1	9(+)
2	62,3	62,3	0	
3	58,6	58,0	0,6	4(+)
4	60,2	59,1	1,1	9(+)
5	63,4	60,2	3,2	21(+)
6	78,6	68,3	10,3	30(+)
7	55,4	57,0	-1,6	11(-)
8	64,9	60,8	4,1	23(+)
9	65,0	62,0	3,0	18,5(+)
10	63,2	60,7	2,5	15,5(+)
11	72,9	65,4	7,5	29(+)
12	56,1	55,0	1,1	9(+)
13	69,2	62,9	6,3	28(+)
14	68,8	64,0	4,8	27(+)
15	58,8	58,0	0,8	5(+)
16	66,2	64,2	2,0	13(+)
17	62,5	62,8	-0,3	3(-)
18	68,7	65,6	3,1	20(+)
19	62,2	60,0	2,2	14(+)
20	66,5	62,0	4,5	26(+)
21	62,0	60,3	1,7	12(+)
22	60,1	60,0	0,1	1(+)
23	63,2	60,7	2,5	15,5(+)
24	61,3	60,3	1,0	7(+)
25	60,3	59,4	0,9	6(+)
26	65,1	62,4	2,7	17(+)
27	70,2	66,0	4,2	24(+)
28	65,3	62,3	3,0	18,5(+)
29	58,2	58,0	0,2	2(+)
30	64,5	60,9	3,6	22(+)
			R(-)	14
			R(+)	426

Порядок розрахунку:

1. Розраховують різниці $d_i = x_i - y_i$ (стовпчик 3); пари з однаковими значеннями відкидаються.
2. Знаходять ранги абсолютних значень різниць (стовпчик 4) та відмічають ранги, що відносяться до додатних та від'ємних значень різниць.
3. Знаходять суми рангів від'ємних $R(-)$ та додатних $R(+)$ різниць, менша з сум рангів приймається в якості значення критерію W . В даному прикладі вона рівна 14.
4. З довідкових таблиць знаходиться критичне значення критерію та порівнюється із спостерігаємим. Якщо $W_{табл} \leq W_{кр}$, то нульова гіпотеза відкидається – спостережна різниця пов'язаних вибірок є статистично значимою на рівні значимості α . В даному прикладі $W_{кр} = 69$ при $\alpha = 0,01$. Так як $W_{табл} < W_{кр}$, різниця в масі тіла жінок до та після експерименту достовірні, $P < 0,01$.

Однією із головних задач вивчення випадкових масових явищ є дослідження взаємозалежностей випадкових подій. В математичній статистиці ця задача зводиться до виявлення та вивчення залежностей випадкових величин одна від іншої. Існують методи, що дозволяють це робити. У найпростішому випадку розглядаються різні статистичні оцінки міри вказаної залежності.

Дві випадкові величини можуть бути пов'язані або функціональною залежністю (жорстко, як залежність змінних в математичному аналізі), або так званою статистичною залежністю, або бути незалежними.

Вибіркові значення (пари чисел $(x_i; y_i)$) системи випадкових величин $(X; Y)$ зображуються точками на площині з декартовою системою координат – кореляційному полі. Чим «краще» на кореляційному полі вибіркові значення вкладаються у вигляді будь-якої кривої, тим сильніша залежність випадкових величин в системі $(X; Y)$.

Статистичною називають залежність, при якій зміна однієї величини призводить до зміни розподілу іншої. Якщо при зміні однієї з величин змінюється середнє значення іншої, то говорять про кореляційну залежність.

Кореляційним аналізом називається розділ математичної статистики, який досліджує залежність між випадковими величинами на основі різних вибіркових оцінок генеральних коефіцієнтів кореляції.

Регресійні залежності – один із найбільш розповсюджених видів математичних моделей, що описує залежність однієї випадкової величини від ряду інших, випадкових або не випадкових величин. Регресійна залежність, або просто регресія, записується у вигляді функції, лінійної або нелінійної, яка описує цю залежність у «середньому», тобто наближено. Так записуються всі фізичні закони з параметрами та емпіричні залежності одних величин через інші.

В якості міри залежності випадкових величин в математичній статистиці найбільш розповсюджені дві оцінки коефіцієнта кореляції:

- Брава-Пірсона r ;

- Спірмена r_s .

Перша оцінка аналогічна визначенню коефіцієнта кореляції в теорії ймовірностей за моментами 1-го та 2-го порядків. В математичній статистиці його оцінка Брава-Пірсона розраховується за вибірковим значенням за формулою:

$$r = \frac{\frac{1}{n} \sum x_i y_i - \bar{x} \bar{y}}{\sqrt{\left(\frac{1}{n} \sum x_i^2 - \bar{x}^2 \right) \left(\frac{1}{n} \sum y_i^2 - \bar{y}^2 \right)}}.$$

Коефіцієнт кореляції Спірмена r_s розраховується за аналогічною формулою (замість вибіркових значень $(x_i; y_i)$ в неї підставляються їх ранги $(u_i; v_i)$). Тому оцінку r_s часто називають ранговим коефіцієнтом кореляції.

Введені оцінки коефіцієнта кореляції мають однакові властивості:

- обидві оцінки r та r_s приймають значення діапазону $[-1;1]$;
- $r = +1$ або $r = -1$ тоді і тільки тоді, коли випадкові величини X та Y пов'язані жорсткою лінійною залежністю, тобто $Y = aX + b$ (на кореляційному полі зображуванні точки лягають на яку-небудь пряму);
- у випадку нормального розподілу системи $(X; Y)$ рівність $r = 0$ рівносильна незалежності випадкових величин X та Y ;
- якщо ж система $(X; Y)$ розподілена не за нормальним законом, рівність $r = 0$ може виконуватись і за наявності нелінійної залежності випадкових величин X та Y ;
- якщо $r \neq 0$, то чим ближче $|r|$ до одиниці, тим сильніша лінійна залежність випадкових величин X та Y .

Термінологія, пов'язана з величиною r :

$|r| = 1$ – лінійна залежність;

$0,7 \leq |r| < 0,99$ – сильна статистична залежність;

$0,5 \leq |r| < 0,7$ – середня статистична залежність;

$0,2 \leq |r| < 0,5$ – слабка статистична залежність;

$0,09 \leq |r| < 0,2$ – дуже слабка статистична залежність;

$r > 0$ – додатня кореляція між X та Y ;

$r < 0$ – від'ємна кореляція між X та Y ;

$r = 0$ – кореляції немає; випадкові величини X та Y некорельовані.

Загальноприйнято вказувати значимість коефіцієнта кореляції. Наприклад, в експерименті взяли участь 10 чоловік. Оцінювався взаємозв'язок між результатами в бігу на 30 м та 100 м. Отримано коефіцієнт кореляції $r = 0,611$. Щоб оцінити значимість коефіцієнта кореляції, необхідно порівняти його з критичним, величина якого залежить від об'єму вибірки та рівня значимості. Якщо фактичне значення коефіцієнта кореляції більше, ніж критичне, це означає, що коефіцієнт кореляції статистично достовірний. В даному випадку критичне значення

коефіцієнту кореляції при $\alpha = 0,05$ становить $r_{кр} = 0,632$. Із цього випливає, що розрахований коефіцієнт кореляції статистично невірний. Приводити його у дослідженнях неможливо.

У науковій публікації необхідно вказати об'єм вибірки, щоб читач міг оцінити достовірність розрахованих коефіцієнтів кореляції. Іноді в публікаціях наводяться тільки значимі коефіцієнти кореляції, а замість незначимих ставляться прочерки. У таблиці 6.2 вказано, що об'єм вибірки рівний $n = 0,349$. Відповідно, всі коефіцієнти кореляції достовірні.

Таблиця 6.2

Коефіцієнти кореляції між результатами у швидкісно-силових тестах та результатом у штовханні ядра з розгону

	Вправа	1	2	3	4	5	6
1	Штовхання ядра з розгону	1	0,97	0,84	0,83	0,73	0,73
2	Штовхання ядра з місця		1	0,84	0,82	0,74	0,76
3	Кидок ядра назад			1	0,85	0,71	0,66
4	Кидок ядра вперед				1	0,66	0,62
5	Присідання зі штангою					1	0,58
6	Жим штанги лежачи						1

Порядок виконання роботи

1.1. Задання змінних та введення даних

Необхідно оцінити достовірність різниці маси тіла жінок 18-25 років, які входять до експериментальної групи до та після експерименту.

1.1.1. В програмному продукті StatGraphics необхідно створити таблицю даних на основі прикладу (Додаток 6). Для задання назв змінних величин, подвійним клацанням ЛКМ на назві стовпчика, відкрийте вікно зміни назви Modify Column (модифікувати колонку). В строчку Name (Ім'я) даного вікна введіть назви стовпчиків (masado – маси тіла жінок до експерименту, masapi – маси тіла жінок після експерименту).

1.1.2. Скопіюйте дані з таблиці (Додаток 6).

	masado	masapi	Col_3	Col_4
1	59,1	58		
2	62,3	62,3		
3	58,6	58		
4	60,2	59,1		
5	63,4	60,2		
6	78,6	68,3		
7	55,4	57		
8	64,9	60,8		
9	65	62		
10	63,2	60,7		
11	72,9	65,4		
12	56,1	55		
13	69,2	62,9		
14	68,8	64		
15	58,8	58		
16	66,2	64,2		
17	62,5	62,8		

Рис. 6.1. Таблиця вхідних даних

1.1.3.Збережіть створену таблицю, вибравши File – Save – Save Data File.

Перевірка розподілу різниць результатів показала їх близькість до нормального закону. У зв'язку із цим для перевірки гіпотези про те, що середнє арифметичне різниць рівне нулю ($H_0 : \mu_d = 0$), можна застосувати t-критерій Стьюдента.

1.2. Розрахунок t-критерію Стьюдента

1.2.1. Виберіть пункт меню Compare/ Two Samples/ Paired Samples. З'явиться вікно діалогу для завдання порівнюваних вибірок (рис. 6.2).

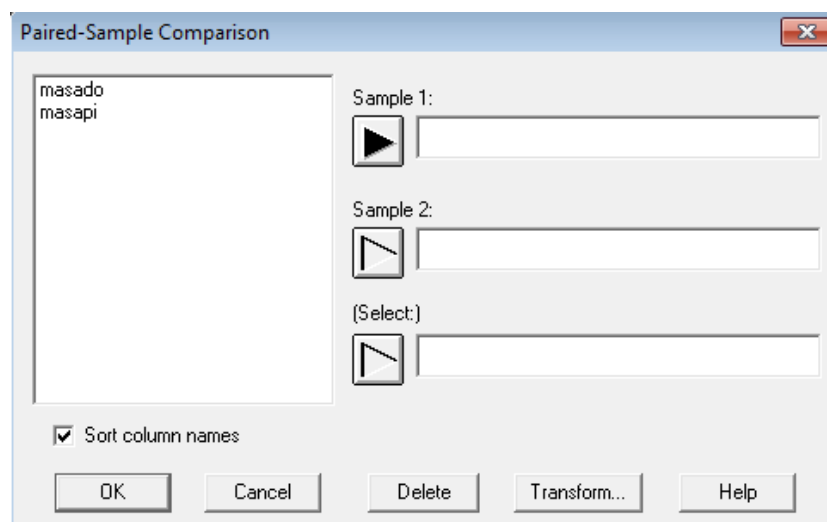


Рис. 6.2. Вікно порівнюваних вибірок

Виділіть змінну masado клацанням лівої кнопки миші, та клацніть по кнопці зі стрілкою, таким чином вводячи дану змінну в поле Sample 1. У такий же спосіб введіть в аналіз змінну masapi в поле Sample 2 (рис. 6.3) та натисніть ОК.

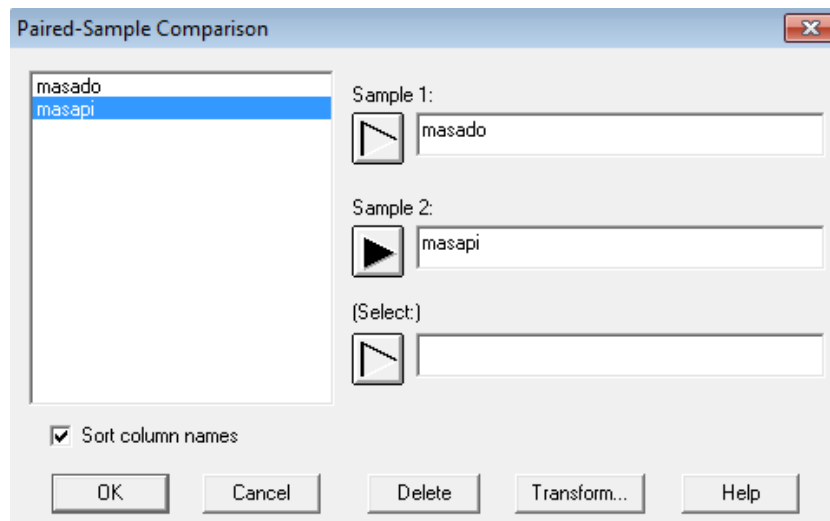


Рис. 6.3. Введення змінних в аналіз

Отримаємо вихідний звіт аналізу (рис. 6.4).

Paired Samples - masado & masapi

Data variable: masado-masapi

30 values ranging from -1,6 to 10,3

The StatAdvisor

This procedure is designed to test for significant differences between two data samples where the data were collected as pairs. It will calculate various statistics and graphs for the differences between the paired data. Also included in the procedure are tests designed to determine whether the mean difference is equal to zero. Use the Tabular Options and Graphical Options buttons on the analysis toolbar to access these different procedures.

Рис. 6.4. Вихідний звіт аналізу

У вихідному звіті вказуються:

- кількість різниць – 30;
- діапазон варіювання – від -1,6 до 10,3.

В нижній частині звіту наведена додаткова пояснювальна інформація.

1.2.2. Натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 6.5) встановіть прапорець таблиці критеріїв для перевірки гіпотез (Hypothesis Tests).

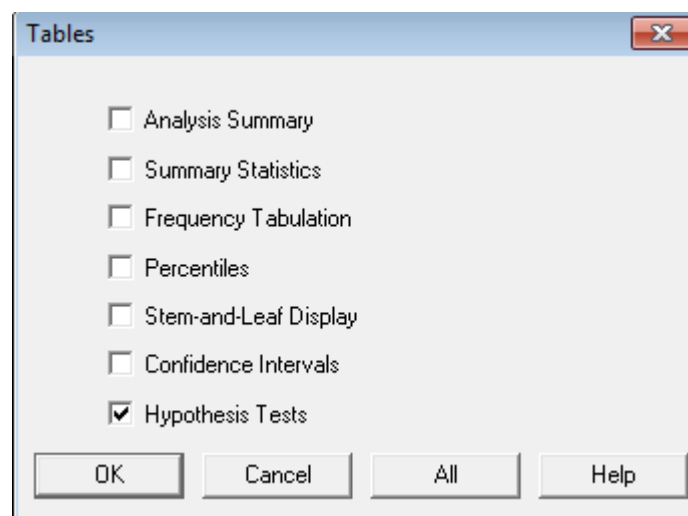


Рис. 6.5. Вікно вибору табличних опцій

З'явиться таблиця аналізу перевірки гіпотез(рис. 6.6).

Hypothesis Tests for masado-masapi

Sample mean = 2,54

Sample median = 2,35

Sample standard deviation = 2,47534

t-test

Null hypothesis: mean = 0,0

Alternative: not equal

Computed t statistic = 5,6203

P-Value = 0,00000453467

Reject the null hypothesis for alpha = 0,05.

sign test

Null hypothesis: median = 0,0

Alternative: not equal

Number of values below hypothesized median: 2

Number of values above hypothesized median: 27

Large sample test statistic = 4,45669 (continuity correction applied)

P-Value = 0,00000833082

Reject the null hypothesis for alpha = 0,05.

signed rank test

Null hypothesis: median = 0,0

Alternative: not equal

Average rank of values below hypothesized median: 7,0

Average rank of values above hypothesized median: 15,5926

Large sample test statistic = 4,39028 (continuity correction applied)

P-Value = 0,0000113296

Reject the null hypothesis for alpha = 0,05.

chi-squared test

Null hypothesis: sigma = 1,0

Alternative: not equal

Computed chi-squared statistic = 177,692

P-Value = 0,0

Reject the null hypothesis for alpha = 0,05.

Рис. 6.6. Звітна таблиця аналізу перевірки гіпотез

У вікні Hypothesis Tests наведено результати аналізу даних із застосуванням трьох тестів: одного параметричного (t-тест) та двох непараметричних: критерію знаків (signtest) та критерію знакових рангів Вілкоксона (signedranktest).

Значення t-критерію Стьюдента $t=5,6203$, $P\text{-value}=0,00000453467$. Після цих показників надається висновок СтатКонсультанта: нульова гіпотеза відкидається на рівні значимості $\alpha=0,05$.

Якщо допущення про відповідність розподілу результатів нормальному закону не підтверджується, то потрібно застосовувати два інших критерії, які мають меншу потужність, однак, і вимоги до вихідних даних у них нижче.

Значення критерію знаків (signtest) рівне 4,45669; при цьому $P\text{-value}=0,00000833082$. Після цих показників надається висновок СтатКонсультанта: нульова гіпотеза відкидається на рівні значимості $\alpha=0,05$.

Значення критерію знакових рангів (signedranktest) рівне 4,39028; при цьому P-value=0,0000113296. Після цих показників надається висновок СтатКонсультанта: нульова гіпотеза відкидається на рівні значимості $\alpha=0,05$.

Після виводу результатів аналізу даних із застосуванням усіх трьох критеріїв СтатКонсультант надає коротку методику розрахунку з використанням кожного з критеріїв.

1.2.3. Збережіть весь виконаний аналіз. Для цього виберіть File – Save As – Save StatFolio As.

2.1. Задання змінних та введення даних

Необхідно: оцінити взаємозв'язок між висотою стрибка у висоту з місця та результатом у штовханні у важкоатлетів першого розряду (вага до 60 кг); розрахувати коефіцієнти кореляції Браве-Пірсона та Спірмена, оцінити значимість розрахованих коефіцієнтів.

2.1.1. В програмному продукті StatGraphics необхідно створити таблицю даних на основі прикладу (Додаток 7). Для задання назв змінних величин, подвійним клацанням ЛКМ на назві стовпчика, відкрийте вікно зміни назви Modify Column (модифікувати колонку). В строчку Name (Ім'я) даного вікна введіть назви стовпчиків (jump – результати стрибка у висоту з місця (см), push – результати у штовханні (см)).

2.1.2. Скопіюйте дані з таблиці (Додаток 7).

	jump	push	Col_3	Col_4
1	57	107,5		
2	60	110		
3	58	110		
4	61	115		
5	63	115		
6	58	107,5		
7	55	107,5		
8	64	120		
9	65	122,5		
10	64	112,5		
11	66	120		
12	61	110		

Рис. 6.7. Таблиця вхідних даних

2.1.3. Збережіть створену таблицю, вибравши File – Save – Save Data File.

Перевірка розподілу різниць результатів показала їх близькість до нормального закону. У зв'язку із цим для перевірки гіпотези про те, що середнє арифметичне різниць рівне нулю ($H_0 : \mu_d = 0$), можна застосувати t-критерій Стьюдента.

2.2. Розрахунок коефіцієнту кореляції Браве-Пірсона та Спірмена

2.2.1. Виберіть пункт меню Describe/ Numeric Data/ Multiple Variable Analysis. З'явиться вікно діалогу для завдання кореляції (рис. 6.8).

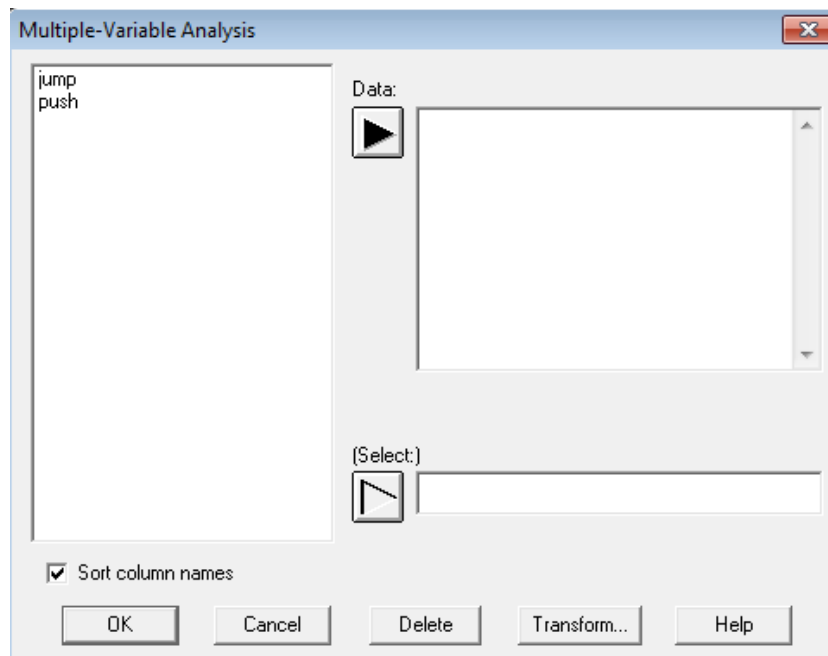


Рис. 6.8. Вікно порівнюваних вибірок

Виділіть обидві змінні та клацніть по кнопці зі стрілкою, таким чином вводячи дані змінні в поле Data та натисніть ОК. З'явиться вікно загального аналізу (рис. 6.9).

Multiple-Variable Analysis

Data variables:

jump
push

There are 12 complete cases for use in the calculations.

The StatAdvisor

This procedure is designed to summarize several columns of quantitative data. It will calculate various statistics, including correlations, covariances, and partial correlations. Also included in the procedure are a number of multivariate graphs, which give interesting views into the data. Use the Tabular Options and Graphical Options buttons on the analysis toolbar to access these different procedures.

Рис. 6.9. Вихідний звіт аналізу

У вихідному звіті вказуються:

- назви змінних;
- кількість випадків у змінних.

В нижній частині звіту наведена додаткова пояснювальна інформація.

2.2.2. Натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 6.10) встановіть прапорці таблиць коефіцієнту кореляції Браве-Пірсона (Correlations) та рангового коефіцієнту кореляції Спірмена (Rankcorrelations), натисніть кнопку ОК.

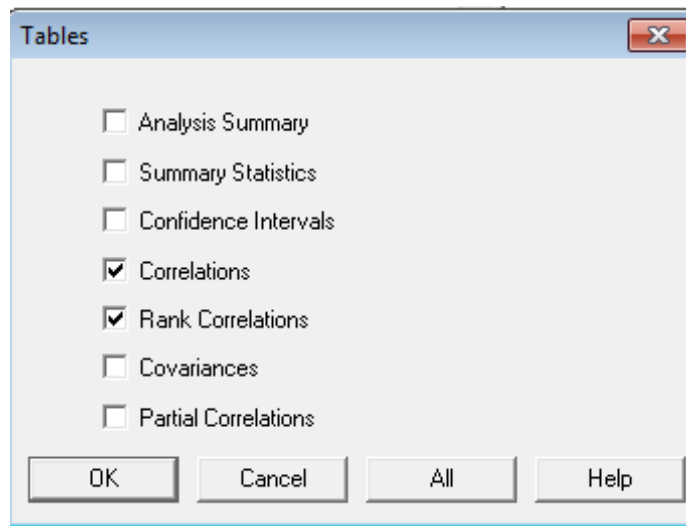


Рис. 6.10. Вікно вибору табличних опцій

З'явиться таблиця коефіцієнту кореляції Браве-Пірсона(рис. 6.11) та таблиця рангового коефіцієнту кореляції Спірмена (рис. 6.12).

Correlations

	jump	push
jump		0,8774
		(12)
		0,0002
push	0,8774	
	(12)	
	0,0002	

Correlation
(Sample Size)
P-Value

Рис. 6.11. Таблиця коефіцієнту кореляції Браве-Пірсона

Spearman Rank Correlations

	jump	push
jump		0,9196
		(12)
		0,0023
push	0,9196	
	(12)	
	0,0023	

Correlation
(Sample Size)
P-Value

Рис. 6.12. Таблиця рангового коефіцієнту кореляції Спірмена

На рисунку 6.11 представлені: коефіцієнт кореляції Браве-Пірсона, який рівний 0,8774, число парних вимірювань 12, P-value=0,0002. СтатКонсультант повідомляє, що P-values менше 0,05 мають наступні пари змінних: jump, push. Аналіз цього повідомлення СтатКонсультанта дозволяє швидко виявити всі достовірні коефіцієнти кореляції.

Таким чином, взаємозв'язок результатів стрибків у висоту з місця та результатів у штовханні у важкоатлетів першого розряду висока $r = 0,8774$, коефіцієнт кореляції достовірний, $P < 0,001$.

2.3. Побудова графіку кореляційного поля

2.3.1. Для наочного представлення взаємозв'язку між результатами потрібно побудувати графік, кореляційного поля. Виберіть Plot – Scatterplots – X-Y Plot. У діалоговому вікні (рис. 6.13) змінну jump занесіть в поле X, а змінну push – у поле Y, натисніть OK.

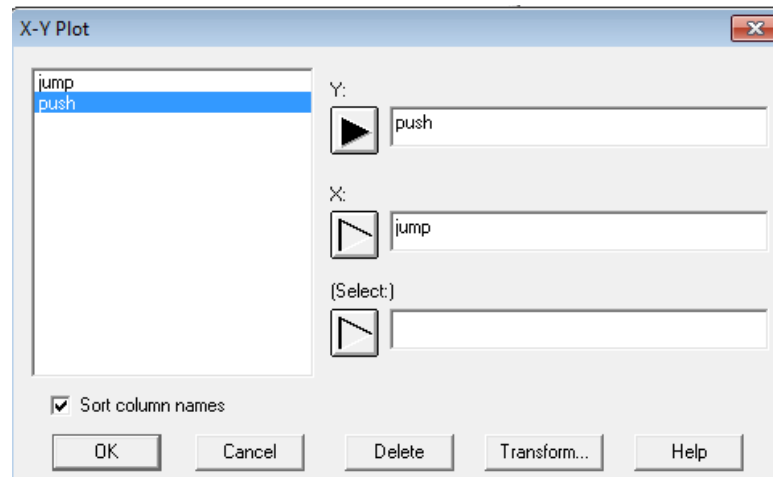


Рис. 6.13. Вікно задання осей координат

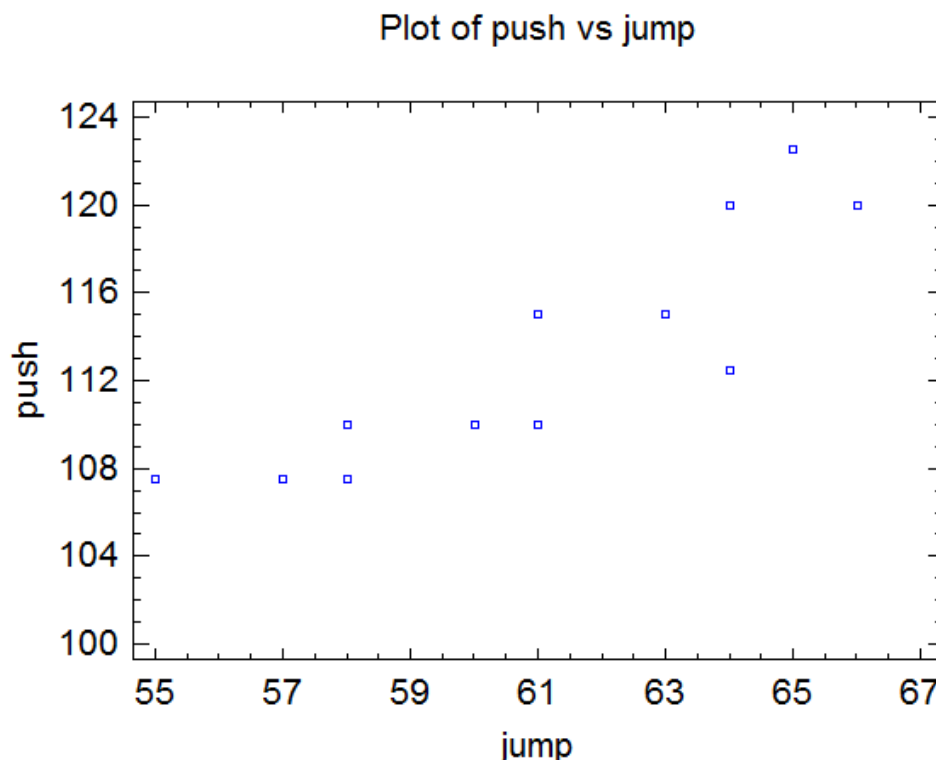


Рис. 6.14. Кореляційне поле

2.3.2. Збережіть весь виконаний аналіз. Для цього виберіть File – Save As – Save StatFolio As.

Контрольні запитання:

1. Які вибірки називаються пов'язаними?
2. Які умови застосування t-критерію Стюдента для пов'язаних вибірок?
3. Який порядок розрахунку t-критерію Стюдента для пов'язаних вибірок?
4. Який порядок розрахунку критерію знаків?
5. Який порядок розрахунку критерію знакових рангів Вілкоксона?
6. Який порядок розрахунку цих критеріїв в програмному продукті StatGraphics?
7. Що являє собою кореляційний аналіз?
8. Що використовується в якості міри зв'язку?
9. В яких випадках розраховується коефіцієнт кореляції Брава-Пірсона?
10. В яких межах знаходиться значення коефіцієнта кореляції?
11. Про що свідчить значення коефіцієнта кореляції, рівне 1?
12. Про що свідчить значення коефіцієнта кореляції, рівне 0?
13. В яких межах частіше всього заключене значення коефіцієнта кореляції?
14. В яких випадках говорять про додатну кореляцію і що це означає?
15. В яких випадках говорять про від'ємну кореляцію і що це означає?
16. Що називають значимістю коефіцієнта кореляції?
17. Як розрахувати коефіцієнт кореляції у програмному продукті StatGraphics?
18. Що називають кореляційним полем і як побудувати графік у програмному продукті StatGraphics?

Лабораторна робота №7

Планування та перевірка експерименту в програмному продукті StatGraphics

Мета роботи: навчитися виконувати: планування якісного та кількісного експериментів; створення двофакторної математичної моделі; розрахунок значень відгуку при всіх можливих комбінаціях факторів; створення карти Парето, графіків взаємодії факторів та таблиці дисперсного аналізу в програмному продукті StatGraphics.

Теоретичні відомості

Планування експерименту - процедура вибору числа та умов проведення дослідів, необхідних та достатніх для вирішення задачі досліджень із заданою точністю.

У деяких дослідженнях не вдається підібрати фактори, які можна характеризувати кількісно. Іноді дослідження проводяться на якісному рівні. В ролі незалежних змінних можуть виступати наступні фактори: тип матеріалу (сталь або алюміній), якість паперу (газетний або конструкторський), форма об'єкта (конус або циліндр), кваліфікація робітника (висока чи низька) та ін. Не зважаючи на те, що в подібних дослідженнях фактори носять якісний (описовий) характер, відгук повинен мати кількісну оцінку (наприклад, витрати палива на 100 км пробігу, строк служби тормозних колодок, похибка вимірювального засобу, кількість відмов апаратури та ін.).

Кількісні дослідження - це дослідження, націлені на строгу стандартизацію та формалізацію процесу збору та обробки інформації, які дають можливість отримати точні дані про досліджуваний об'єкт, виражені в абсолютних або відносних величинах.

Якісні дослідження являють собою неформалізований збір даних з використанням польових методів і нестандартизованою формою їх аналізу.

Карта Парето - це стовпчаста діаграма, на якій інтервали (стовпчики) впорядковані по низхідній лінії. На такій діаграмі інтервали можуть становити види дефектів, їх локалізацію, помилки тощо. А висота інтервалів (висота стовпчиків) - частоту виникнення дефектів, їх відсоткове співвідношення, вартість, час та ін.

Будується карта Парето в наступному порядку:

1. Визначається проблема, яку необхідно вирішити (наприклад, дефектні вироби, вартість втрат від браку та ін.) і вибирається часовий інтервал для вивчення проблеми.

2. Вибирається тип даних (фактор) для аналізу, який найбільш повно зможе охарактеризувати проблему (наприклад, дефекти, їх локалізація, обсяг втрат, витрати та ін.). Обраний тип даних повинен бути розбитий на підтипи. Наприклад, якщо в якості типу даних обрані дефекти, то підтипом будуть види дефектів - деформація, подряпини, тріщини та ін.

3. Визначається одиниця вимірювань, що відповідає типу даних (наприклад, кількість дефектів, їх частота, відсоток витрат та ін.).

4. Збираються статистичні дані, і виконується їх систематизація. Для збору і реєстрації даних можна застосовувати інші інструменти якості, наприклад контрольний листок. Систематизацію статистичних даних краще представити у вигляді таблиці.

5. Виконується розрахунок і впорядкування даних за спаданням.

6. При необхідності призначаються ваги для кожного з підтипів даних. Встановлення ваг може зробити істотний вплив на результат, який покаже карта Парето. Ваги перемножуються на розраховані значення по кожному з підтипів даних, що призводить до зміни співвідношення їх значимості.

7. Будується стовпчаста діаграма, на якій зазначаються підтипи даних і їх величина. У прямокутній системі координат по горизонталі відкладаються рівні відрізки, відповідні підтипам даних, а по вертикалі відкладається величина цих даних в порядку спаданням.

8. Обчислюється і відображається на діаграмі лінія сумарних значень (наприклад, накопичених відсотків).

9. Виконується аналіз отриманих результатів для розробки необхідних дій щодо вирішення проблеми.

Дисперсний аналіз (analysis of variance (ANOVA)) являє собою статистичний метод аналізу результатів, які залежать від якісних ознак. Кожен фактор може бути дискретною чи неперервною випадковою змінною, яку розділяють на декілька сталих рівнів (градацій, інтервалів). Якщо кількість вимірювань (проб, даних) на всіх рівнях кожного з факторів однакова, то дисперсний аналіз називають рівномірним, інакше - нерівномірним. В основі дисперсного аналізу є такий принцип: якщо на випадкову величину діють взаємно незалежні фактори А, В, ..., n, то загальна дисперсія дорівнює сумі дисперсій, зумовлених дією окремо кожного з факторів:

$$\sigma^2 = \sigma_A^2 + \sigma_B^2 + \dots + \sigma_n^2$$

В будь-якому експерименті середні значення досліджуваних величин змінюються у зв'язку зі зміною основних факторів (кількісних та якісних), що визначають умови досліду, а також і випадкових факторів. Дослідження впливу тих чи інших факторів на мінливість середніх є задачею дисперсного аналізу. Дисперсний аналіз використовує властивість адитивності дисперсії випадкової величини, що обумовлено дією незалежних факторів. В залежності від числа джерел дисперсії розрізняють однофакторний та багатофакторний дисперсний аналіз. Дисперсний аналіз особливо ефективний при вивченні кількох факторів. При класичному методі вивчення змінюють тільки один фактор, а решту залишають постійними. При цьому для кожного фактору проводиться своя серія спостережень, що не використовується при вивченні інших факторів. Крім того, при такому методі досліджень не вдається визначити взаємодію факторів при одночасній їх зміні. При дисперсному аналізі кожне спостереження служить для одночасної оцінки всіх факторів та їх взаємодії. Дисперсний аналіз полягає у виділенні й оцінюванні окремих факторів, що викликають зміну досліджуваної випадкової величини. При цьому проводиться розклад сумарної вибіркової дисперсії на

складові, обумовлені незалежними факторами. Кожна з цих складових є оцінкою дисперсії генеральної сукупності. Щоб дати оцінку дієвості впливу даного фактору, необхідно оцінити значимість відповідної вибіркової дисперсії у порівнянні з дисперсією відтворення, обумовленою випадковими факторами. Перевірка значимості оцінок дисперсії проводиться з допомогою критерію Фішера. Коли розрахункове значення критерію Фішера виявиться меншим за табличне, то вплив досліджуваного фактору немає підстав вважати значимим. Коли ж розрахункове значення критерію Фішера виявиться більшим за табличне, то цей фактор впливає на зміни середніх. В подальшому будемо вважати, що виконуються наступні припущення:

1. Випадкові помилки спостережень мають нормальний розподіл.
2. Фактори впливають тільки на зміну середніх значень, а дисперсія спостережень залишається постійною.

Фактори, що розглядаються в дисперсному аналізі, бувають трьох видів:

- з випадковими рівнями, коли вибір рівнів проходить з безмежної сукупності можливих рівнів та супроводжується рандомізацією і рівні вибираються випадковим чином;
- з фіксованими рівнями;
- змішаного типу - частина факторів розглядається на фіксованих рівнях, але рівні решти вибираються випадковим чином.

Дисперсний аналіз застосовується в різних формах в залежності від структури об'єкту, що досліджується; вибір відповідної форми є однією з головних труднощів у практичному застосуванні аналізу. Дисперсний аналіз використовує властивість адитивності дисперсії випадкової величини, що обумовлено дією незалежних факторів.

Адитивність - властивість величин, яка полягає в тому, що значення величини, яке відповідає цілому об'єкту, дорівнює сумі значень величин, що відповідають його частинам, незалежно від того, яким чином поділено об'єкт.

В залежності від числа джерел дисперсії розрізняють однофакторний та багатофакторний дисперсний аналіз.

Дисперсний двофакторний аналіз застосовується в тих випадках, коли досліджується одночасна дія двох факторів на різні вибірки об'єктів, тобто коли різні вибірки опиняються під впливом різних поєднань двох факторів. Може статися, що одна змінна значущо діє на досліджувану ознаку тільки при певних значеннях іншої змінної. Отже, дисперсний двофакторний аналіз дозволяє оцінити не лише вплив кожного з факторів, але й їхню взаємодію.

Суть методу залишається тією самою, як і при однофакторній моделі, але у двофакторному дисперсному аналізі можна перевірити більшу кількість гіпотез, проте розрахунки дещо складніші, ніж в однофакторних комплексах.

Дисперсний двофакторний аналіз пред'являє особливі вимоги до формування комплексів. Для кожного фактору має бути не менше двох градацій; у кожному осередку комплексу повинно бути не менше двох спостережуваних значень для виявлення взаємодії градацій;

комплекс має бути симетричною системою: кожній градації фактору А повинна відповідати однакова кількість градацій фактору В;результативна ознака повинна мати нормальний розподіл; фактори мають бути незалежними, що може бути підтверджено відсутністю кореляційного зв'язку між змінними-чинниками.

Порядок виконання роботи

1.1. Створення моделей на якісному рівні. Введення даних

На даному етапі досліджується залежність дальності польоту паперових моделей літаків від їх конструкційних особливостей. Конструкція літака характеризується чотирма якісними факторами, кожен з яких може приймати лише два значення.

Використовуючи факторний план 2^4 , провести планування експерименту та створити двофакторну математичну модель. Модель повинна описувати залежність функції відгуку від чотирьох факторів. Використовуючи отриману повну математичну модель, розрахувати по ній 16 значень відгуку при всіх можливих комбінаціях факторів. Співставити розраховані та вихідні значення відгуків.

Якісні фактори:

- конструкція літака (Design). В залежності від форми крила літака конструкція може бути або прямою (straight), або Т-подібною (tee);
- форма носа (Nose). Ніс моделі літака може бути зрізаним (cut) або не зрізаним (no);
- тип паперу (Paper). Папір, що використовується для виготовлення моделі, може бути конструкторським (construct) або газетним (writing);
- тип крила (Wing). Крило моделі може бути прямим (direct) або нахиленим (incline).

Залежною змінною (відгуком) в даному експерименті є дальність польоту моделі літака - відстань (Distance).

Обрані фактори у різному ступені впливають на величину відгуку. Для виділення факторів, які найбільш суттєво впливають на дальність польоту (Distance) моделей, застосовують карту Парето, графік взаємодії факторів та таблицю дисперсного аналізу.

1.1.1.В програмному продукті StatGraphics виконайте команду по створенню експерименту DOE – Design Creation – Create New Design. У вікні встановлення опцій проекту (Create Design Options) – рис. 7.1, встановіть перемикач на Screening (виведення на екран). В полі No. of Response Variables (число цільових змінних - відгуків) потрібно залишити 1. В полі No. of Experimental Factors (число факторів, що впливають) встановіть число 4. В полі Comment (коментар) вкажіть назву даного експерименту, наприклад Air. Натисніть кнопку ОК.

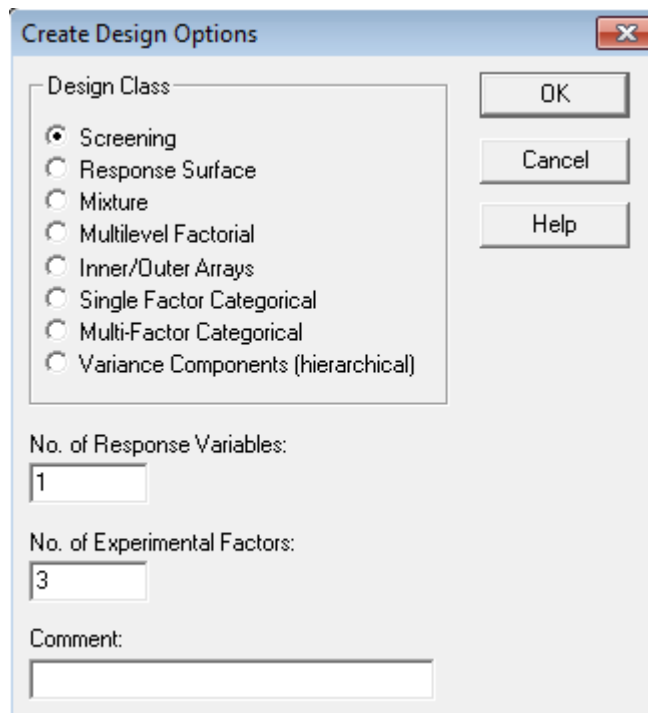


Рис. 7.1. Вікно встановлення опцій проекту

1.1.2. У наступному вікні (рис. 7.2) необхідно по чергову ввести в поле Name (ім'я) назви змінних для факторів A, B, C та D (їх вибір здійснюється за допомогою перемикача). В поля Low (нижній рівень) та High (верхній рівень) необхідно ввести назви конструкційних ознак для кожного фактору. У всіх діалогових вікнах слід прибрати прапорець у полях Continuous (безперервний).

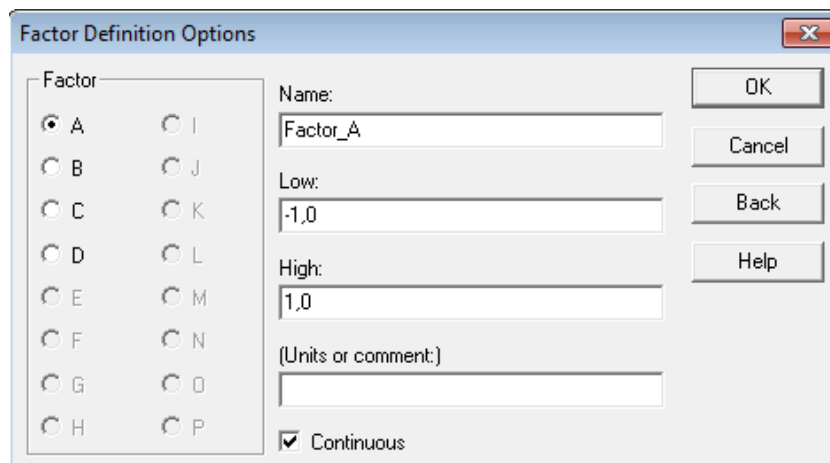
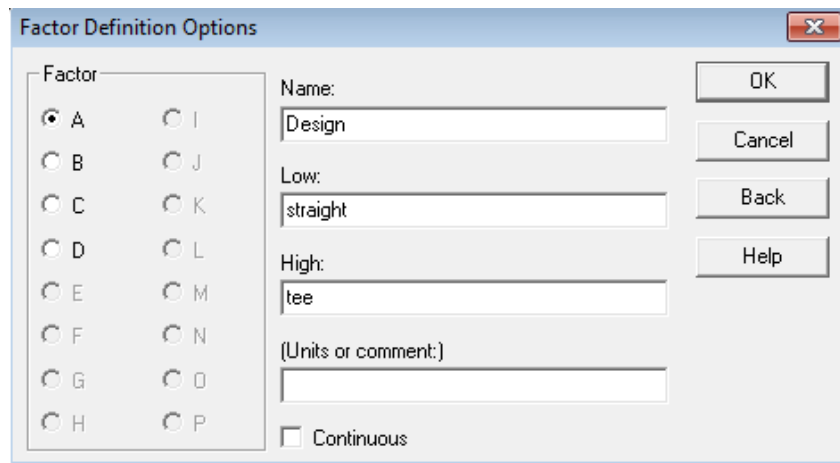


Рис. 7.2. Вікно введення факторних ознак

Для фактору A в поле Name (ім'я) введіть Design (конструкція), в поле Low (нижній рівень) – Straight (пряма), а в поле High (верхній рівень) – tee (Т-подібна). Значенню змінної Straight відповідає число «-1», а значенню змінної tee – число «+1» (рис. 7.3). Аналогічне зауваження справедливе для всіх факторів.



The dialog box is titled "Factor Definition Options". On the left, under the "Factor" heading, there are two columns of radio buttons labeled A through P. Option A is selected. To the right, there are four text input fields: "Name:" containing "Design", "Low:" containing "straight", "High:" containing "tee", and "(Units or comment:)" which is empty. Below these fields is a checkbox labeled "Continuous" which is unchecked. On the far right, there are four buttons: "OK", "Cancel", "Back", and "Help".

Рис. 7.3. Введення ознак факторну А

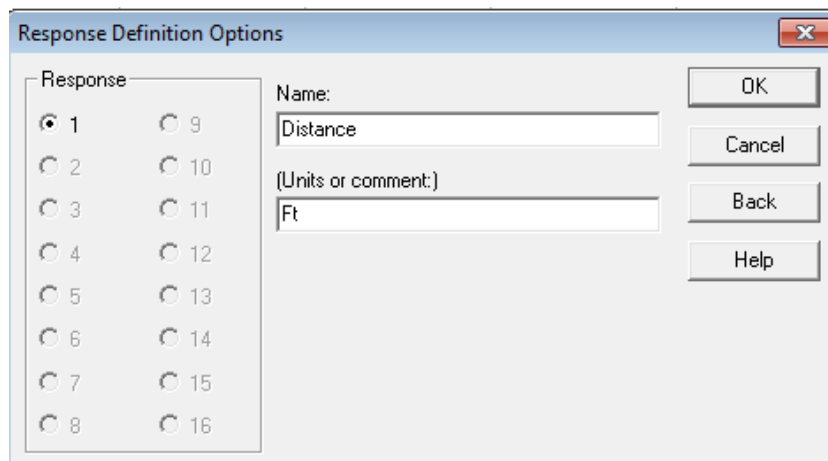
Для фактору В в поле Name (ім'я) введіть Nose (ніс), в поле Low (нижній рівень) – по (незрізаний, немає), а в поле High (верхній рівень) – cut (зрізаний).

Для фактору С в поле Name (ім'я) введіть Paper (папір), в поле Low (нижній рівень) – writting (газетний), а в поле High (верхній рівень) – construct (конструкторський).

Для фактору D в поле Name (ім'я) введіть Wing (крило), в поле Low (нижній рівень) – direct (пряме), а в поле High (верхній рівень) – incline (нахилене).

Натисніть кнопку ОК.

1.1.3. У наступному вікні (рис. 7.4) Response Definition Options (опції відгуку) в поле Name (ім'я) введіть ім'я відгуку – Distance (відстань), а в поле Units (одиниці вимірювання) – Ft (фут).



The dialog box is titled "Response Definition Options". On the left, under the "Response" heading, there are two columns of radio buttons labeled 1 through 16. Option 1 is selected. To the right, there are two text input fields: "Name:" containing "Distance" and "(Units or comment:)" containing "Ft". On the far right, there are four buttons: "OK", "Cancel", "Back", and "Help".

Рис. 7.4. Вікно введення ознак відгуку

Натисніть кнопку ОК.

1.1.4. У наступному вікні Screening Design Selection (вибір експериментального плану), використовуючи список, виберіть план Factorial 2^4 (рис. 7.5) та натисніть кнопку ОК.

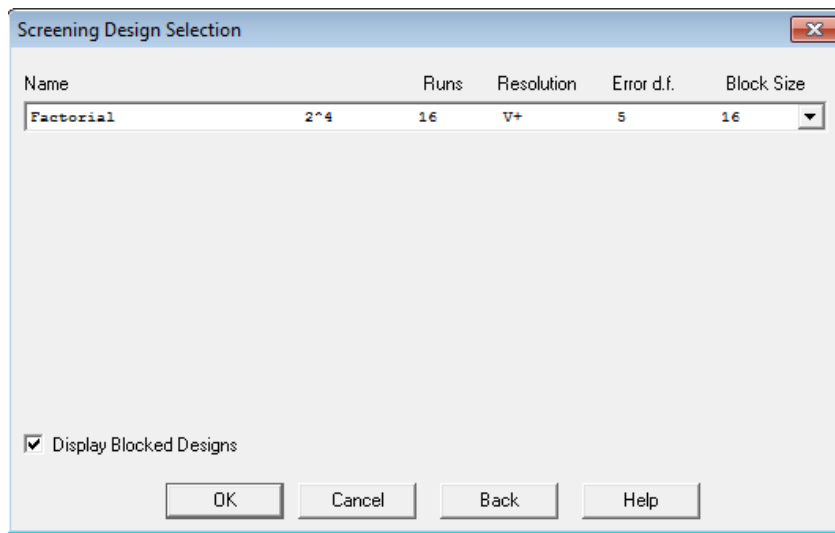


Рис. 7.5. Вікно вибору експериментального плану

1.1.5. У новому діалоговому вікні (рис. 7.6) Screening Design Options (опції проекту) приберить прапорець Randomize (довільна вибірка) та натисніть кнопку OK.

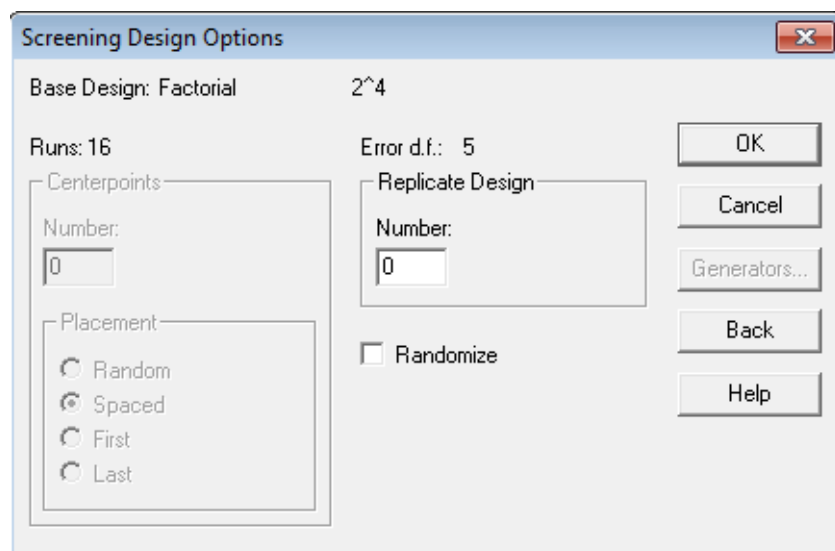


Рис. 7.6. Вікно опцій проекту

На екрані з'явиться вікно, в якому буде представлений план експерименту (рис. 7.7).

Screening Design Attributes

Design class: Screening

Design name: Factorial

2⁴

File name: <Untitled>

Comment: Air

Base Design

Number of experimental factors: 4

Number of blocks: 1

Number of responses: 1

Number of runs: 16

Error degrees of freedom: 5

Randomized: No

<i>Factors</i>	<i>Low</i>	<i>High</i>	<i>Units</i>	<i>Continuous</i>
Design	straight	tee		No
Nose	no	cut		No
Paper	writing	construct		No
Wing	direct	incline		No

<i>Responses</i>	<i>Units</i>
Distance	Ft

Рис. 7.7. План експерименту

1.1.6. Збережіть створений план. Для цього виберіть команду File – Save – Save Design File As та присвойте файлу ім'я Air.

1.1.7. Виведіть результати експериментальних досліджень. Для цього виберіть команду Window – Air.sfx. На екрані з'явиться таблиця плану (рис. 7.8).

	BLOCK	Design	Nose	Paper	Wing	Distance
						Ft
1	1	straight	no	writing	direct	
2	1	tee	no	writing	direct	
3	1	straight	cut	writing	direct	
4	1	tee	cut	writing	direct	
5	1	straight	no	construct	direct	
6	1	tee	no	construct	direct	
7	1	straight	cut	construct	direct	
8	1	tee	cut	construct	direct	
9	1	straight	no	writing	incline	
10	1	tee	no	writing	incline	
11	1	straight	cut	writing	incline	
12	1	tee	cut	writing	incline	
13	1	straight	no	construct	incline	
14	1	tee	no	construct	incline	
15	1	straight	cut	construct	incline	
16	1	tee	cut	construct	incline	

Рис. 7.8. Таблиця плану експерименту

1.1.8. Введіть у колонку числа у відповідності до таблиці 7.1. Дробова частина числа відділяється комою від цілої.

Експериментальні дані

№	Фактори				Відгук
	Design Конструкція	Nose Ніс	Peper Папір	Wing Крило	Distance Відстань
1	straight	no	writting	direct	6,25
2	tee	no	writing	direct	15,5
3	straight	cut	writing	direct	7
4	tee	cut	writing	direct	16,5
5	straight	no	construct	direct	4,75
6	tee	no	construct	direct	5,5
7	straight	cut	construct	direct	4,5
8	tee	cut	construct	direct	6
9	straight	no	writing	incline	7
10	tee	no	writing	incline	10
11	straight	cut	writing	incline	10
12	tee	cut	writing	incline	16
13	straight	no	construct	incline	4,5
14	tee	no	construct	incline	6
15	straight	cut	construct	incline	4,5
16	tee	cut	construct	incline	5,75

	BLOCK	Design	Nose	Paper	Wing	Distance
						Ft
1	1	straight	no	writting	direct	6,25
2	1	tee	no	writting	direct	15,5
3	1	straight	cut	writting	direct	7
4	1	tee	cut	writting	direct	16,5
5	1	straight	no	construct	direct	4,75
6	1	tee	no	construct	direct	5,5
7	1	straight	cut	construct	direct	4,5
8	1	tee	cut	construct	direct	6
9	1	straight	no	writing	incline	7
10	1	tee	no	writing	incline	10
11	1	straight	cut	writting	incline	10
12	1	tee	cut	writting	incline	16
13	1	straight	no	construct	incline	4,5
14	1	tee	no	construct	incline	6
15	1	straight	cut	construct	incline	4,5
16	1	tee	cut	construct	incline	5,75

Рис. 7.9. Таблиця плану експерименту із відгуками

1.1.9. Збережіть створену таблицю, вибравши команду File – Save – Save Design File.

1.2. Аналіз плану експерименту

1.2.1. Виберіть пункт меню DOE – Design Analysis – Analyze Design. З'явиться вікно діалогу. Подвійним клацанням ЛКМ введіть змінну Distance в поле Data (рис. 7.10) та натисніть кнопку ОК.

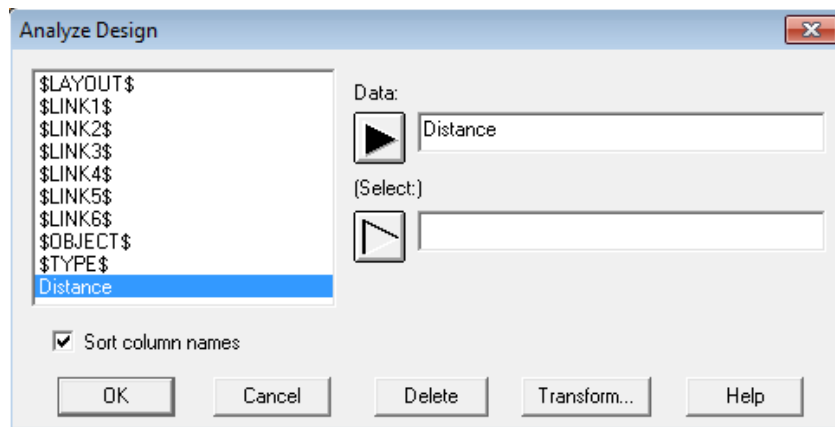


Рис. 7.10. Вікно аналізу плану

На екрані з'явиться звіт проведеного аналізу (рис.7.11).

Analyze Experiment - Distance

File name: Air.sfx

Comment: Air

Estimated effects for Distance (Ft)

<i>Effect</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>V.I.F.</i>
average	8,10938	0,385086	
A:Design	4,09375	0,770171	1,0
B:Nose	1,34375	0,770171	1,0
C:Paper	-5,84375	0,770171	1,0
D:Wing	-0,28125	0,770171	1,0
AB	0,46875	0,770171	1,0
AC	-2,84375	0,770171	1,0
AD	-1,15625	0,770171	1,0
BC	-1,34375	0,770171	1,0
BD	0,84375	0,770171	1,0
CD	0,28125	0,770171	1,0

Standard errors are based on total error with 5 d.f.

Рис. 7.11. Звіт аналізу плану

1.3. Побудова графіків

1.3.1. Визначити фактори, які сильніше за інші впливають на відгук, можна за допомогою карти Парето. Для цього натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 7.12) встановіть прапорець карти Парето (Pareto Charts).

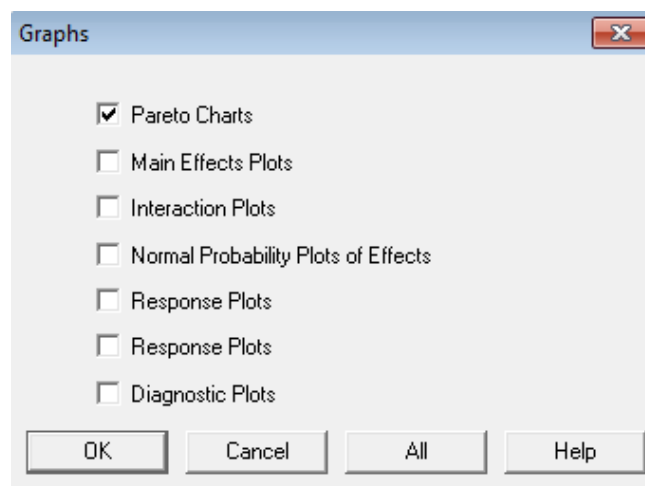


Рис. 7.12. Вікно графічних опцій

На карті Парето (рис. 7.13) добре видно, що тип паперу (C), конструкція (A) та їх взаємодія (AC) має статистично значимий вплив. Вертикальна лінія являє собою 95% тест для визначення значимості факторів.

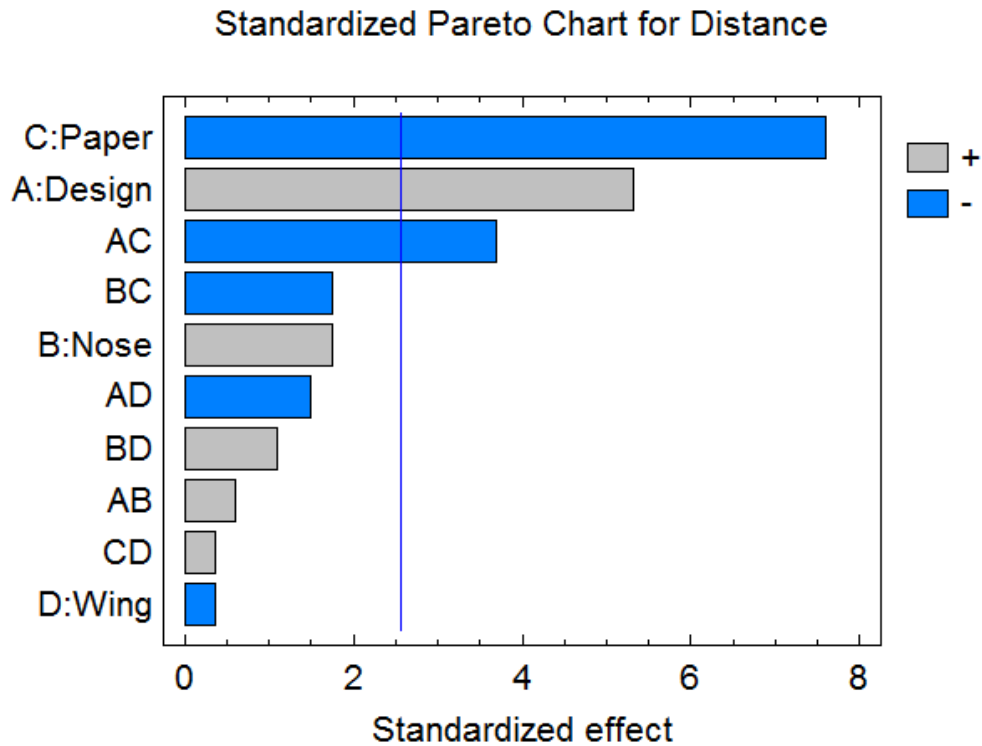


Рис. 7.13. Карта Парето

Фактори Paper (C), Design (A) та їх взаємодія (AC) набагато перевищують довжину ліній інших факторів на карті Парето. Це свідчить про те, що ці фактори діють сильніше, ніж інші фактори та взаємодії інших факторів.

1.3.2. Натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 7.14) встановіть прапорець таблиці дисперсного аналізу (ANOVA Table), за допомогою якої можна відділити фактори, що найбільш сильно впливають від шумів.

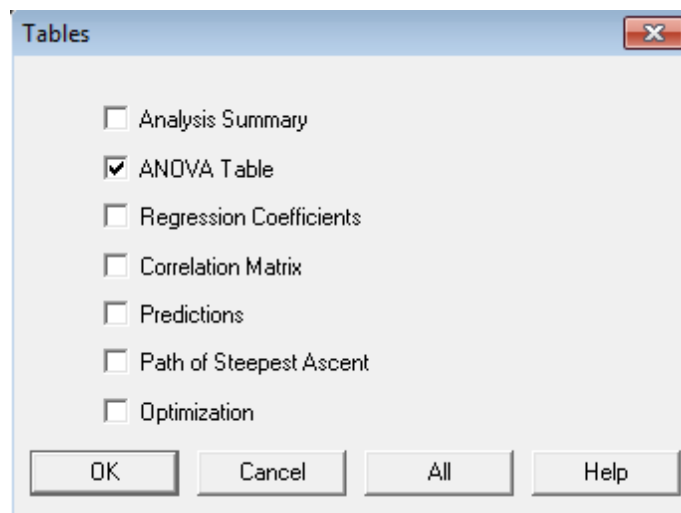


Рис. 7.14. Вікно вибору табличних опцій

З'явиться таблиця дисперсного аналізу (рис. 7.15).

Analysis of Variance for Distance - Air

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:Design	67,0352	1	67,0352	28,25	0,0032
B:Nose	7,22266	1	7,22266	3,04	0,1415
C:Paper	136,598	1	136,598	57,57	0,0006
D:Wing	0,316406	1	0,316406	0,13	0,7299
AB	0,878906	1	0,878906	0,37	0,5694
AC	32,3477	1	32,3477	13,63	0,0141
AD	5,34766	1	5,34766	2,25	0,1936
BC	7,22266	1	7,22266	3,04	0,1415
BD	2,84766	1	2,84766	1,20	0,3232
CD	0,316406	1	0,316406	0,13	0,7299
Total error	11,8633	5	2,37266		
Total (corr.)	271,996	15			

R-squared = 95,6384 percent
R-squared (adjusted for d.f.) = 86,9153 percent
Standard Error of Est. = 1,54034
Mean absolute error = 0,705078
Durbin-Watson statistic = 3,19559 (P=0,9605)
Lag 1 residual autocorrelation = -0,603165

Рис. 7.15. Таблиця дисперсного аналізу

З рисунка 7.15 видно, що фактори Design, Paper та їх взаємодія найбільш сильно впливають на дальність польоту моделі. Для цих факторів значення P-Value менше критичного значення 0,05. В той же час вплив інших факторів, B та D, а також їх взаємодій є несуттєвими, тому їх можна виключити з математичної моделі. Цей висновок співпадає з попереднім (візуальним) результатом аналізу залежності.

1.3.3. Для того, щоб виключити малозначимі фактори та графічно оцінити отриману залежність, натисніть кнопку графічних опцій  та у відповідному вікні діалогу (рис. 7.16) встановіть прапорець графіка взаємодій (Interaction Plots).

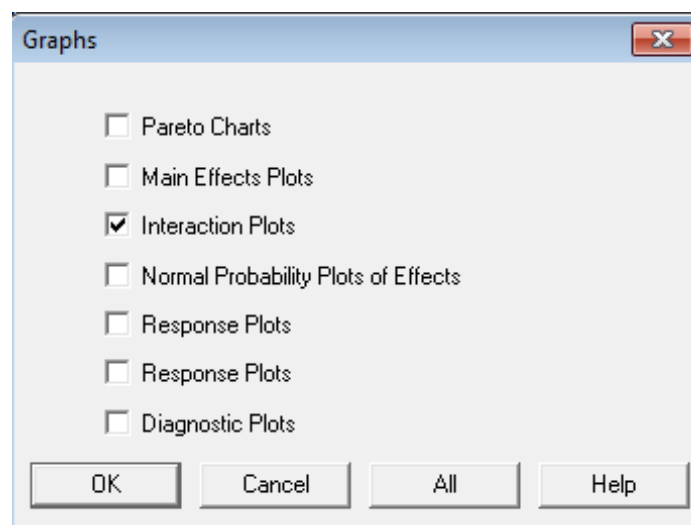


Рис. 7.16. Вікно графічних опцій

Клацніть ПКМ в робочому полі графіка взаємодій та виберіть команду Pane Options з контекстного меню. З'явиться вікно опцій (рис. 7.17) в якому потрібно залишити прапорці тільки на полях Design та Paper.

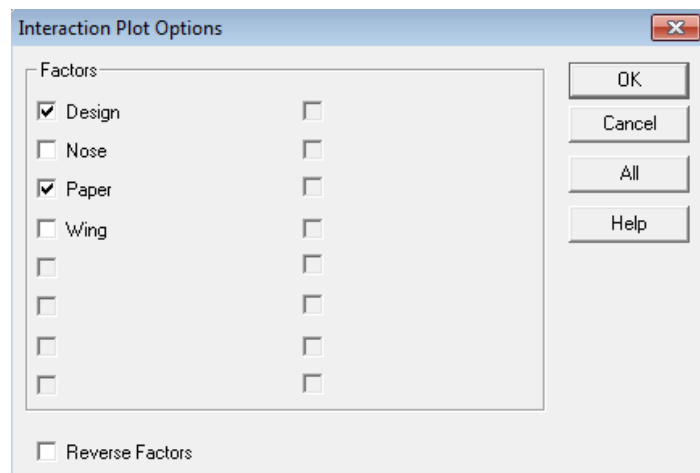


Рис. 7.17. Вікно опцій графіку взаємодій

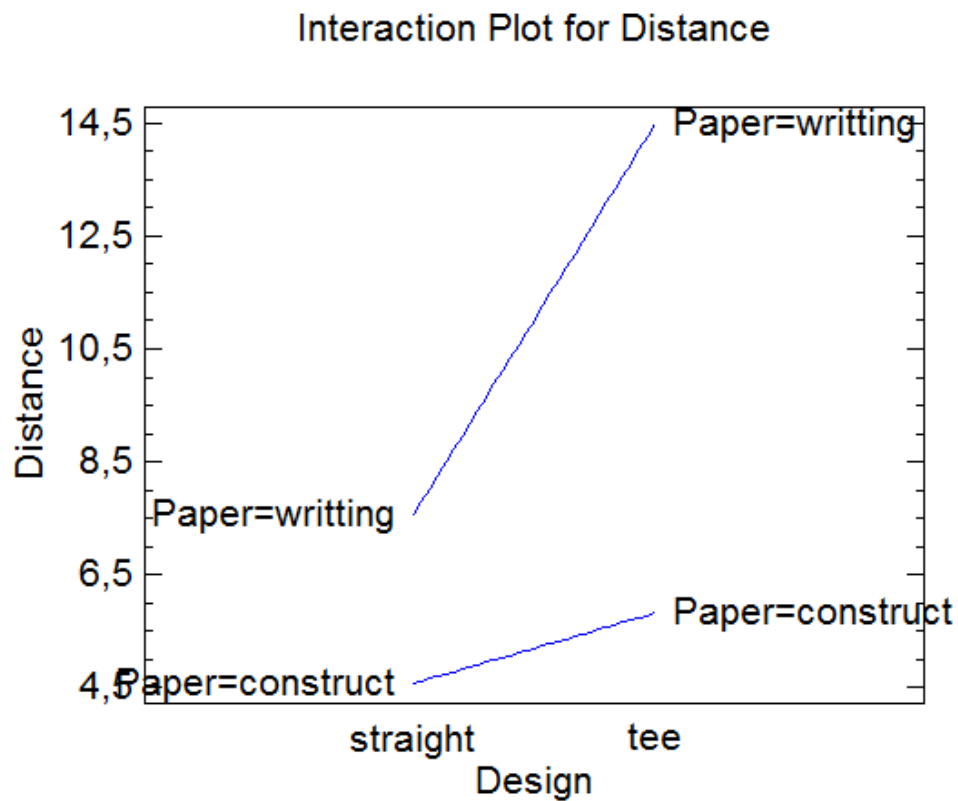


Рис. 7.18. Графік взаємодії факторів Design та Paper

З графіку взаємодії (рис. 7.18) видно, що найкращі льотні властивості мають моделі, виготовлені з газетного (writting) паперу з Т-подібною (tee) конструкцією літака. Літаки саме такої конструкції мають найбільшу дальність польоту (Distance).

1.4. Отримання математичної моделі

1.4.1. Для отримання математичної моделі (формули), що описує досліджувану залежність, натисніть кнопку табличних опцій  та у відповідному вікні діалогу (рис. 7.19) встановіть прапорець таблиці коефіцієнтів регресії (Regression Coefficients).

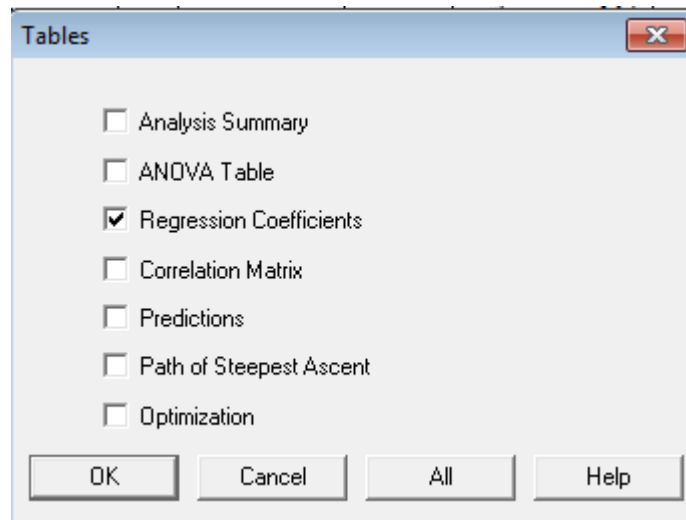


Рис. 7.19. Вікно вибору табличних опцій

З'явиться таблиця з коефіцієнтами, а СтатКонсультант (The StatAdvisor) покаже повний вигляд математичної моделі (рис. 7.20).

Regression coeffs. for Distance - Air

Coefficient	Estimate
constant	8,10938
A:Design	2,04688
B:Nose	0,671875
C:Paper	-2,92188
D:Wing	-0,140625
AB	0,234375
AC	-1,42188
AD	-0,578125
BC	-0,671875
BD	0,421875
CD	0,140625

The StatAdvisor

This pane displays the regression equation which has been fitted to the data. The equation of the fitted model is

Distance = 8,10938 + 2,04688*Design + 0,671875*Nose - 2,92188*Paper - 0,140625*Wing + 0,234375*Design*Nose - 1,42188*Design*Paper - 0,578125*Design*Wing - 0,671875*Nose*Paper + 0,421875*Nose*Wing + 0,140625*Paper*Wing

where the values of the variables are specified in their original units, except for the categorical factors which take the values -1 for the low level and +1 for the high level. To have STATGRAPHICS evaluate this function, select Predictions from the list of Tabular Options. To plot the function, select Response Plots from the list of Graphical Options.

Рис. 7.20. Аналіз коефіцієнтів регресії

Повна математична модель, отримана у результаті виконання планування експерименту:

$$\text{Distance} = 8,10938 + 2,04688 \cdot \text{Design} + 0,671875 \cdot \text{Nose} - 2,92188 \cdot \text{Paper} - 0,140625 \cdot \text{Wing} + 0,234375 \cdot \text{Design} \cdot \text{Nose} - 1,42188 \cdot \text{Design} \cdot \text{Paper} - 0,578125 \cdot \text{Design} \cdot \text{Wing} - 0,671875 \cdot \text{Nose} \cdot \text{Paper} + 0,421875 \cdot \text{Nose} \cdot \text{Wing} + 0,140625 \cdot \text{Paper} \cdot \text{Wing}$$

На основі аналізу карти Парето та таблиці дисперсного аналізу можна записати скорочену математичну модель, з якої видалені малозначимі фактори та поєднання факторів:

$$\text{Distance} = 8,10938 + 2,04688 * \text{Design} - 2,92188 * \text{Paper} - 1,42188 * \text{Design} * \text{Paper}$$

Висновок: на дальність польоту паперових літаків найбільш сильно впливають їх конструкція (Design) і тип паперу (paper), із якого вони виготовлені. Найбільша дальність польоту у Т-подібних літаків із газетного паперу.

2.1. Створення моделей на кількісному рівні

Необхідно обробити дані, які описують вольтамперну характеристику (ВАХ) польового транзистору. У даному дослідженні фактор А означає напругу стік-витік $U_{\text{св}}$, а фактор В – напругу затвор-витік $U_{\text{зв}}$. Відгуком є струм стоку $I_{\text{с}}$.

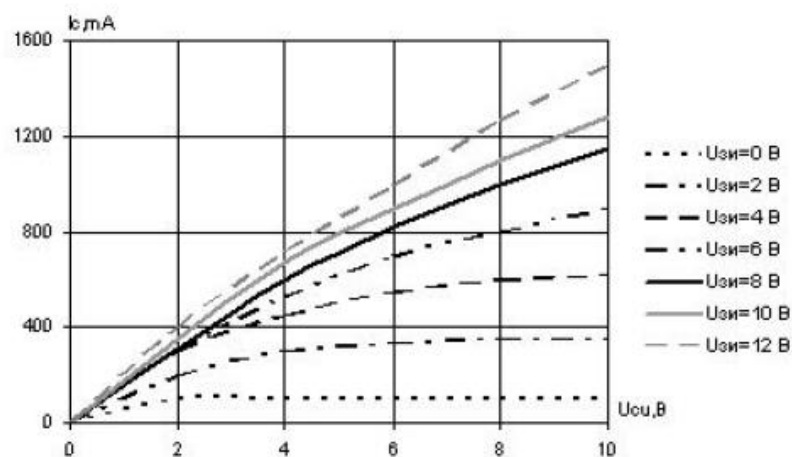


Рис. 7.21. Вольтамперна характеристика транзистора

При виконанні завдання на графіку рисунку 7.21 слід знаходити необхідні точки та приблизно визначати значення точок у відповідних точках плану.

2.1.1. У програмному продукті StatGraphics виконайте команду по створенню експерименту DOE – Design Creation – Create New Design. У вікні встановлення опцій проекту (Create Design Options) – рис. 7.22, встановіть перемикач Design Class (тип плану) на Response Surface (поверхня відгуку). В полі No. of Response Variables (число цільових змінних - відгуків) потрібно залишити 1. В полі No. of Experimental Factors (число факторів, що впливають) встановіть число 2. В полі Comment (коментар) вкажіть назву даного експерименту, наприклад ВАХ (латинськими літерами). Натисніть кнопку ОК.

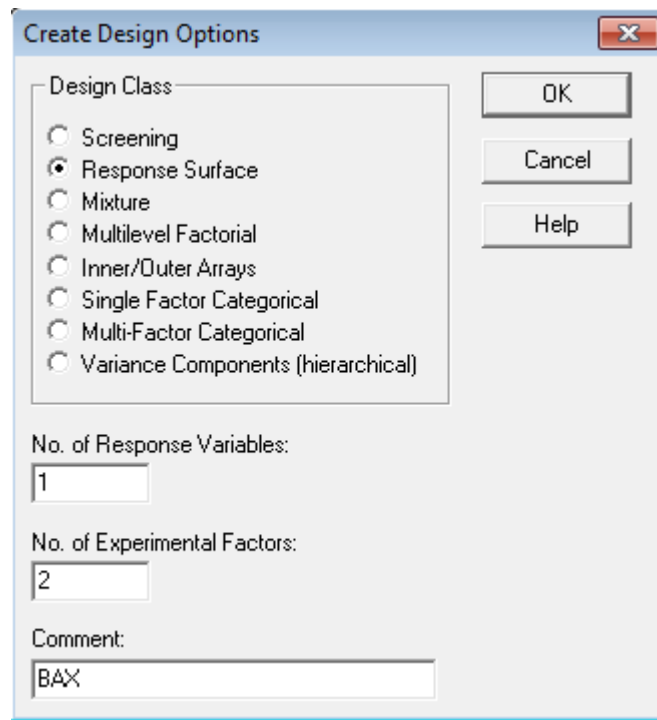


Рис. 7.22. Вікно встановлення опцій проекту

2.1.2. У наступному вікні (рис. 7.23) для фактору А необхідно ввести ім'я U_{CB} (U_{SV}), а також значення нижнього та верхнього рівнів (4 та 6В). В поле одиниць вимірювання ввести Вольти (V). Потім вибрати положення перемикача В (фактор В), задати його ім'я - U_{ZB} (U_{ZV}), а також значення нижнього та верхнього рівнів (0 та 8В). В поле одиниць вимірювання ввести Вольти (V).

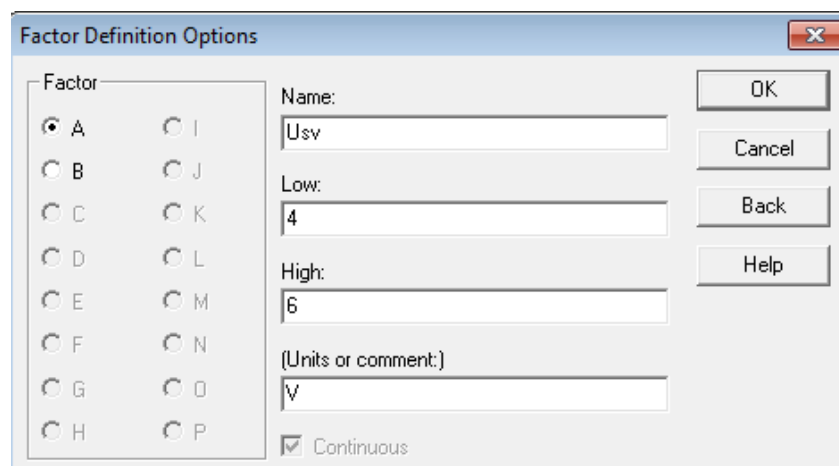


Рис. 7.23. Вікно введення факторних ознак

Натисніть кнопку OK.

2.1.3. У наступному вікні (рис. 7.24) Response Definition Options (опції відгуку) в поле Name (ім'я) введіть ім'я відгуку – I_C , а в поле Units (одиниці вимірювання) – міліампери (mA).

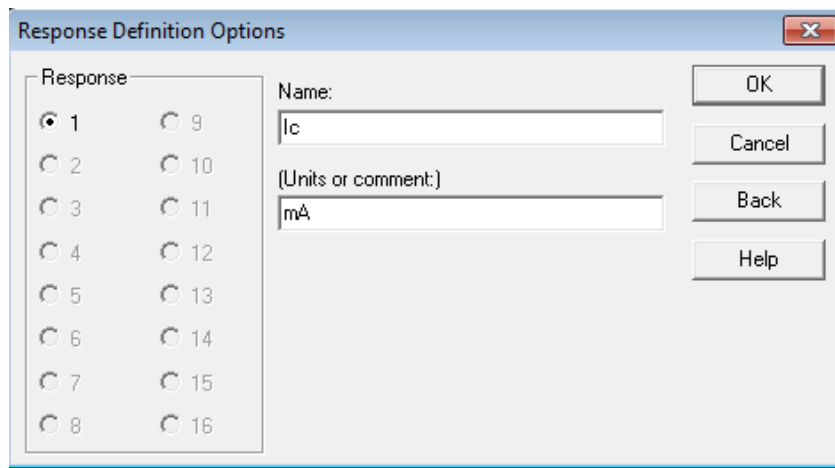


Рис. 7.24. Вікно введення ознак відгуку

Натисніть кнопку ОК.

2.1.4. У наступному вікні Response Surface Design Selection (вибір параметрів поверхні відгуку), використовуючи список, виберіть план 3-level factorial design: 3^2 (трирівневий факторний план 3^2) (рис. 7.25) та натисніть кнопку ОК.

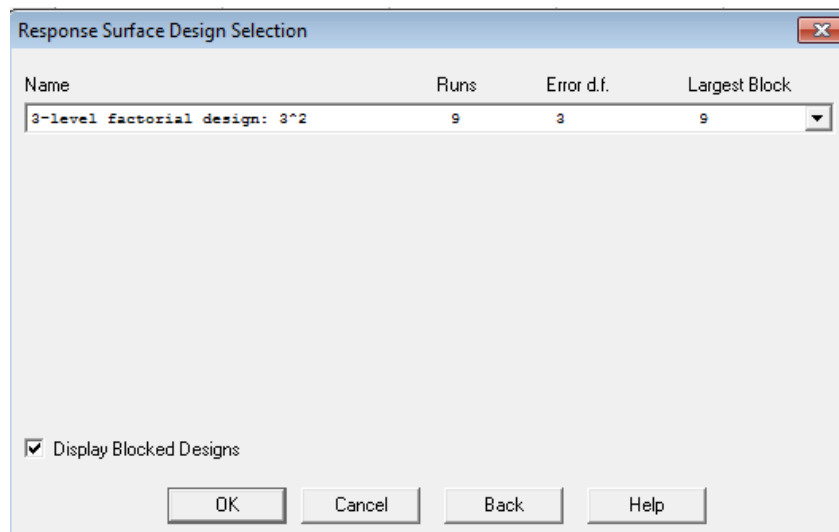


Рис. 7.25. Вікно вибору параметрів поверхні відгуку

2.1.5. У новому діалоговому вікні (рис. 7.26) Three-Level Design Options (опції проекту) приберіть прапорець Randomize (довільна вибірка), встановіть перемикач Placement (розташування) у будь-яке положення: Last (останній), First (перший), Spaced (окремий).

Опція Centerpoints (центральні точки) дає можливість виконати декілька додаткових вимірювань у центральній точці плану (в середині). У даному випадку число додаткових вимірювань потрібно залишити рівним нулю.

Опція Replicate Design (повтор експерименту) дає можливість провести декілька додаткових вимірювань в кожній точці розробленого плану. В даному випадку потрібно залишити значення числа повторів рівним нулю та натисніть кнопку ОК.

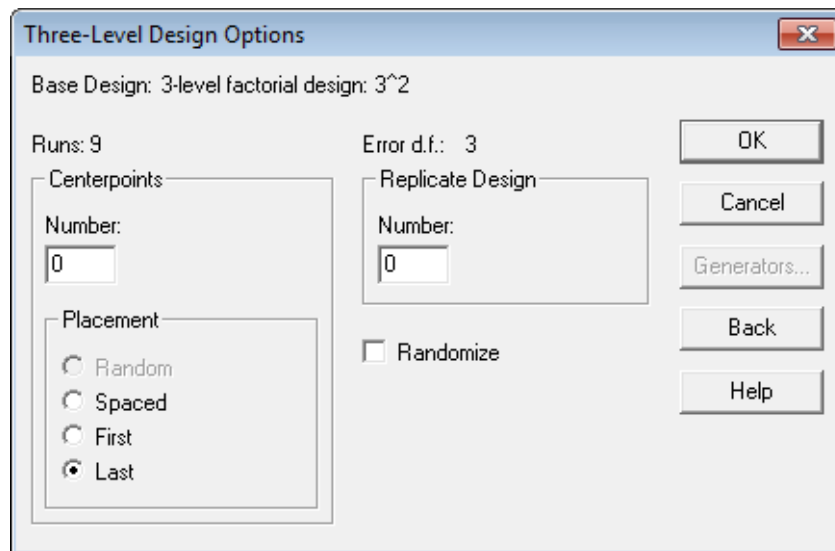


Рис. 7.26. Вікно опцій трирівневого проекту

На екрані з'явиться вікно, в якому буде представлений трирівневий факторний план (рис. 7.27).

Response Surface Design Attributes

Design class: Response Surface

Design name: 3-level factorial design: 3^2

File name: <Untitled>

Comment: BAX

Base Design

Number of experimental factors: 2

Number of blocks: 1

Number of responses: 1

Number of runs: 9

Error degrees of freedom: 3

Randomized: No

<i>Factors</i>	<i>Low</i>	<i>High</i>	<i>Units</i>	<i>Continuous</i>
Usv	4	6	V	Yes
Uzv	0	8	V	Yes

<i>Responses</i>	<i>Units</i>
Ic	mA

Рис. 7.27. План експерименту

2.1.6. Збережіть створений план. Для цього виберіть команду File – Save – Save Design File As та присвойте файлу ім'я BAX.

2.1.7. Виведіть результати експериментальних досліджень. Для цього виберіть команду Window – BAX.sfx. На екрані з'явиться таблиця плану (рис. 7.28).

	BLOCK	U_{sv}	U_{zv}	I_c
		V	V	mA
1	1	4,0	0,0	
2	1	5,0	0,0	
3	1	6,0	0,0	
4	1	4,0	4,0	
5	1	5,0	4,0	
6	1	6,0	4,0	
7	1	4,0	8,0	
8	1	5,0	8,0	
9	1	6,0	8,0	

Рис. 7.28. Таблиця плану експерименту

Щоб визначити значення відгуку I_c , які відповідають вказаним комбінаціям факторів, потрібно використати вольтамперну характеристику польового транзистора. На рис. 7.29 показані точки на графіку у відповідності до розробленого плану експерименту. Нумери точок відповідають номерам строк попередньої таблиці.

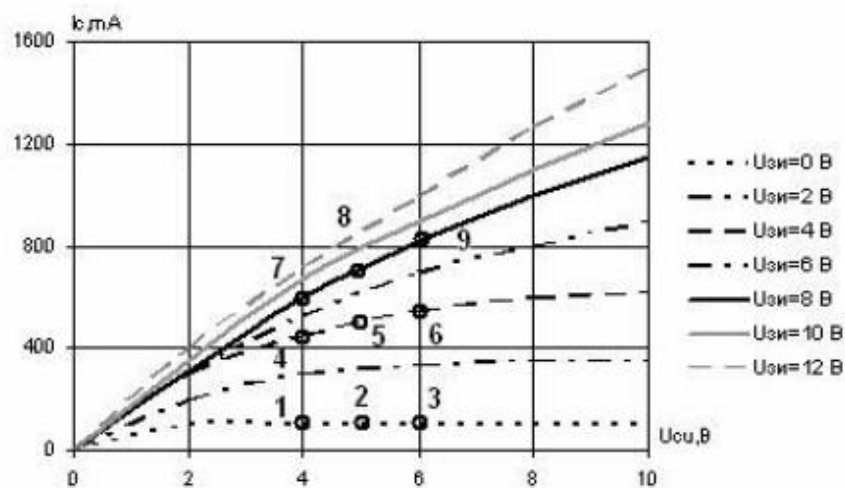


Рис. 7.29. Точки розробленого плану

За допомогою наявного графіка можна визначити значення відгуку I_c в дев'яти точках експерименту.

2.1.8. Введіть у колонку I_c числа у відповідності до рис. 7.30.

	BLOCK	U_{sv}	U_{zv}	I_c
		V	V	mA
1	1	4,0	0,0	100
2	1	5,0	0,0	102
3	1	6,0	0,0	103
4	1	4,0	4,0	420
5	1	5,0	4,0	460
6	1	6,0	4,0	500
7	1	4,0	8,0	600
8	1	5,0	8,0	700
9	1	6,0	8,0	810

Рис. 7.30. Таблиця плану експерименту із відгуками, визначеними за графіком

2.2. Аналіз експериментальних даних

2.2.1. Виберіть пункт меню DOE – Design Analysis – Analyze Design. З’явиться вікно діалогу. Подвійним клацанням ЛКМ введіть змінну I_c в поле Data (рис. 7.31) та натисніть кнопку ОК.

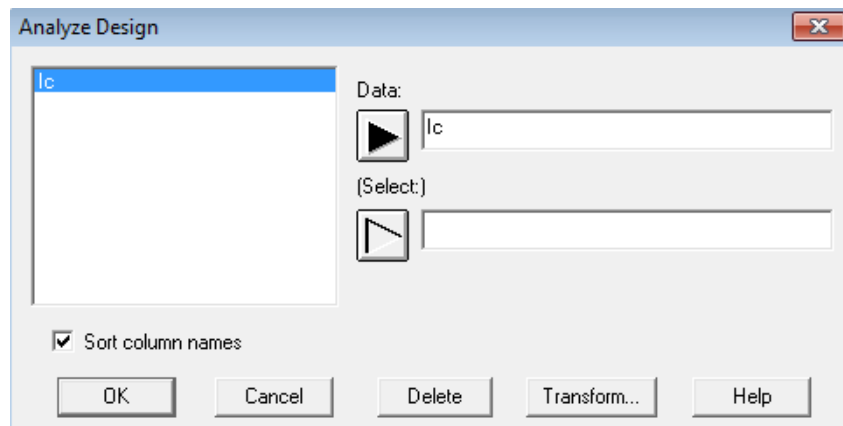


Рис. 7.31. Вікно аналізу плану

На екрані з’явиться первинний звіт проведеного аналізу (рис.7.32).

Analyze Experiment - Ic

File name: BAX.sfx

Comment: BAX

Estimated effects for Ic (mA)

Effect	Estimate	Stnd. Error	V.I.F.
average	459,0	6,7552	
A:Usv	97,6667	7,39995	1,0
B:Uzv	601,667	7,39995	1,0
AA	3,0	12,8171	1,0
AB	103,5	9,06305	1,0
BB	-115,0	12,8171	1,0

Standard errors are based on total error with 3 d.f.

Рис. 7.32. Первинний звіт проведеного аналізу

2.3. Отримання математичної моделі

2.3.1. Натисніть кнопку табличних опцій  та у відповідному вікні діалогу встановіть прапорець таблиці дисперсного аналізу (ANOVA Table) та коефіцієнту регресії (Regression Coefficients).

З’являться таблиці дисперсного аналізу та коефіцієнтів регресії. Логічно спочатку розглянути коефіцієнти регресії (рис. 7.33).

Regression coeffs. for Ic - BAX

Coefficient	Estimate
constant	152,75
A:Usv	-17,9167
B:Uzv	39,2708
AA	1,5
AB	12,9375
BB	-3,59375

Рис. 7.33. Таблиця коефіцієнтів регресії

Математична модель представлена в полі СтатКонсультанта:

$$I_c = 152,75 - 17,9167 \cdot U_{sv} + 39,2708 \cdot U_{zv} + 1,5 \cdot U_{sv}^2 + 12,9375 \cdot U_{sv} \cdot U_{zv} - 3,59375 \cdot U_{zv}^2$$

Оцінку значимості факторів і поєднань зручно виконати за допомогою таблиці дисперсного аналізу (рис. 7.34).

Analysis of Variance for Ic - BAX					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:Usv	14308,2	1	14308,2	174,19	0,0009
B:Uzv	543004,	1	543004,	6610,80	0,0000
AA	4,5	1	4,5	0,05	0,8300
AB	10712,3	1	10712,3	130,42	0,0014
BB	6612,5	1	6612,5	80,50	0,0029
Total error	246,417	3	82,1389		
Total (corr.)	574888,	8			

Рис. 7.34. Таблиця дисперсного аналізу

З результатів даного аналізу видно, що найбільше впливає на струм стоку I_c напруга затвор-витік U_{zv} . Про це каже найменше значення тесту (P-Value) та найбільше значення критерію Фішера (F-Ratio). Найменший вклад у розрахункове значення струму стоку, в даній моделі, вносить квадратичний член U_{sv} (AA). Для цього складене значення P-Value найбільше, а значення F-Ratio – найменше.

Враховуючи, що виконується співвідношення $0,83 > 0,05$, складову AA можна виключити з повної математичної моделі. В результаті видалення квадратичного члена, скорочена математична модель:

$$I_c = 152,75 - 17,9167 \cdot U_{sv} + 39,2708 \cdot U_{zv} + 12,9375 \cdot U_{sv} \cdot U_{zv} - 3,59375 \cdot U_{zv}^2$$

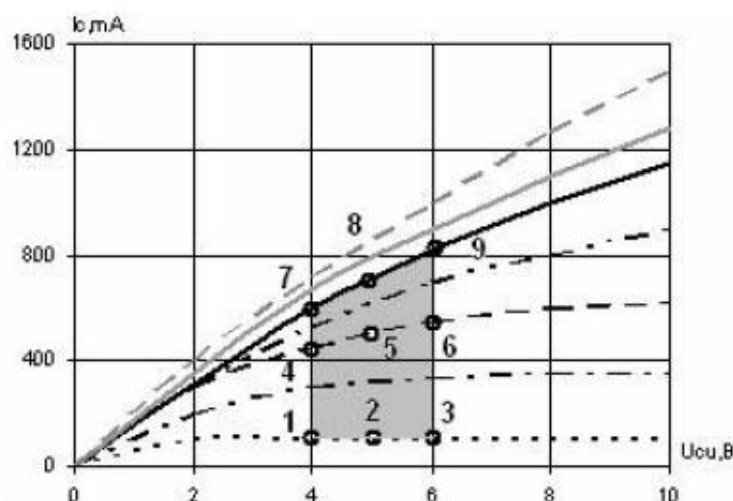


Рис. 7.35. Область дії математичної моделі

Виділять малозначимі фактори можна візуально за допомогою карти Парето та графіка відхилень від нормального розподілу.

Слід звернути увагу на те, що створена математична модель описує вольтамперну характеристику польового транзистору лише у заштрихованій на рис. 7.35 області. За межами цієї зони результати, як правило, суттєво відрізняються від істинних (вихідних) значень.

Якість моделі можна оцінити шляхом розрахунку значення відгуку в тих же точках, які були використані при формулюванні математичної моделі. Для виводу розрахункових значень потрібно у діалоговому вікні Tabular Options встановити прапорець на Predictions (передбачення) та натиснути кнопку ОК.

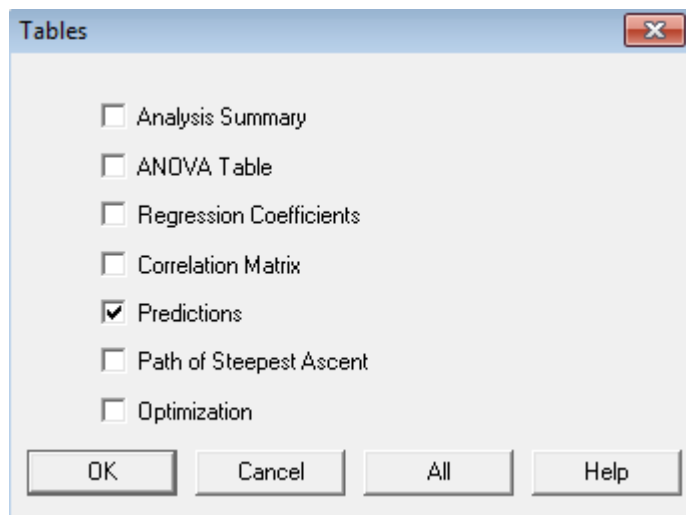


Рис. 7.36. Вікно вибору табличних опцій

З'явиться таблиця (рис. 7.37), в якій містяться вихідні дані (Observed Value), розрахункові дані (Fitted Value), а також верхні (Upper) та нижні (Lower) межі 95%-вого довірчого інтервалу для цих значень.

Estimation Results for Ic				
	<i>Observed</i>	<i>Fitted</i>	<i>Lower 95,0% CL</i>	<i>Upper 95,0% CL</i>
<i>Row</i>	<i>Value</i>	<i>Value</i>	<i>for Mean</i>	<i>for Mean</i>
1	100,0	105,083	79,1962	130,97
2	102,0	100,667	79,1686	122,165
3	103,0	99,25	73,3629	125,137
4	420,0	411,667	390,169	433,165
5	460,0	459,0	437,502	480,498
6	500,0	509,333	487,835	530,831
7	600,0	603,25	577,363	629,137
8	700,0	702,333	680,835	723,831
9	810,0	804,417	778,53	830,304

Рис. 7.37. Таблиця передбачень

2.3.2. Збережіть весь виконаний аналіз. Для цього виберіть File – Save As – Save StatFolio As.

Контрольні запитання:

1. Що називають плануванням експерименту?
2. Що називають кількісним дослідженням?

3. Що називають якісним дослідженням?
4. Що називають картою Парето?
5. Який порядок побудови карти Парето?
6. Що називають дисперсним аналізом?
7. Що називають рівномірним дисперсним аналізом?
8. Що називають нерівномірним дисперсним аналізом?
9. Чому рівна загальна дисперсія?
10. Що є задачею дисперсного аналізу?
11. Поясніть класичний метод дисперсного аналізу?
12. За допомогою якого критерію проводиться перевірка значимості оцінок дисперсії?
13. Які існують види факторів, що розглядаються в дисперсному аналізі?
14. Що називають адитивністю?
15. Що називають дисперсним аналізом?
16. Які вимоги пред'являються до формування комплексів у двофакторному дисперсному аналізі?
17. Який порядок дій при плануванні та перевірці якісного експерименту у програмному продукті StatGraphics?
18. Який порядок дій при плануванні та перевірці кількісного експерименту у програмному продукті StatGraphics?

Література.

1. Катранов А.Г. Компьютерная обработка данных экспериментальных исследований: Учебное пособие / А.Г. Катранов, А.В. Самсонова; СПб ГУФК им. Лесгафта. – СПб.: изд-во СПб ГУФК им. Лесгафта, 2005. – 132 с.
2. Новиков А.М., Новиков Д.А. Методология научного исследования. – М.: Либроком. 2010. – 280 с.
3. Реброва И.А. Планирование эксперимента: учебное пособие. – Омск: СибАДИ, 2010. – 105 с.
4. Мережа Інтернет.

ДОДАТКИ

ДОДАТОК 1

ЧАСТОТА ПОКУПОК КОМПЛЕКТУЮЧИХ НА ПІДПРИЄМСТВІ ЗА СІЧЕНЬ

Data
4
4
9
3
3
1
7
0
4
2
3
5
7
10
6
5
7
3
2
9
8
1
4
2
1
0
8
9

ВИБІРКОВІ ХАРАКТЕРИСТИКИ АВТОМОБІЛІВ

Виробник Make	Модель Model	Вага Weight	Кількість циліндрів Cylinders	Прискорення Accel	Об'єм Displace	Потужність Horsepower
1	2	3	4	5	6	7
Volkswagen	Rabbit D1	1985	4	21,5	90	48
Ford	Fiesta	1800	4	14,4	98	66
Mazda	GLC Deluxe	1985	4	19,4	78	52
Datsun	B210 GX	2070	4	18,6	85	70
Honda	Civic CVCC	1800	4	16,4	91	60
Oldsmobile	Cutlass	3365	8	15,5	260	110
Dodge	Diplomat	3735	8	13,2	318	140
Mercury	Monarch	3570	8	12,8	302	139
Pontiac	Phoenix	3535	6	19,2	231	105
Chevrolet	Malibu	3155	6	18,2	200	95
Ford	Fairmont A	2965	6	15,8	200	85
Ford	Fairmont M	2720	4	15,4	140	88
Plymouth	Volare	3430	6	17,2	225	100
AMC	Concord	3210	6	17,2	232	90
Buick	Century	3380	6	15,8	231	105
Mercury	Zephyr	3070	6	16,7	200	85
Dodge	Aspen	3620	6	18,7	225	110
AMC	Concord D1	3410	6	15,1	258	120
Chevrolet	MonteCarlo	3425	8	13,2	305	145
Buick	RegalTurbo	3445	6	13,4	231	165
Ford	Futura	3205	8	11,2	302	139
Dodge	Magnum XE	4080	8	13,7	318	140
Chevrolet	Chevette	2155	4	16,5	98	68
Toyota	Corona	2560	4	14,2	134	95
Datsun	510	2300	4	14,7	119	97
Dodge	Omni	2230	4	14,5	105	75
Toyota	Celica GT	2515	4	14,8	134	95
Plymouth	Sapporo	2745	4	16,7	156	105
Oldsmobile	Starfire	2855	4	17,6	151	85
Datsun	200-SX	2405	4	14,9	119	97
Audi	5000	2830	5	15,9	131	103
Volvo	264GL	3140	6	13,6	163	125
Saab	99GLE	2795	4	15,7	121	115
Peugeot	604SL	3410	6	15,8	163	133
Volkswagen	Scirocco	1990	4	14,9	89	71
Honda	Accord LX	2135	4	16,6	98	68
Pontiac	Lemans V6	3245	6	15,4	231	115
Mercury	Zephyr 6	2990	6	18,2	200	85
Ford	Fairmont 4	2890	4	17,3	140	88
AMC	ConcordDL6	3265	6	18,2	232	90
Dodge	Aspen 6	3360	6	16,6	225	110
Chevrolet	Caprice Cl	3840	8	15,4	305	130
Ford	LTD Landau	3725	8	13,4	302	129
Mercury	GrandMarqs	3955	8	13,2	351	138
Dodge	St. Regis	3830	8	15,2	318	135
Buick	Estate SW	4360	8	14,9	350	155
Ford	Country SW	4054	8	14,3	351	142
Chevrolet	Malibu SW	3605	8	15	267	125
Chrysler	Lebaron SW	3940	8	13	360	150

ПРОДОВЖЕННЯ ДОДАТКУ 2

1	2	3	4	5	6	7
Volkswagen	Rabbit Cus	1925	4	14	89	71
Mazda	GLC Deluxe	1975	4	15,2	86	65
Dodge	Colt Hatch	1915	4	14,4	98	80
AMC	Spirit DL	2670	4	15	121	80
Mercedes	300D	3530	5	20,1	183	77
Cadillac	Eldorado	3900	8	17,4	350	125
Peugeot	504	3190	4	24,8	141	71
Oldsmobile	Cutlass	3420	8	22,2	260	90
Plymouth	Horizon	2200	4	13,2	105	70
Plymouth	HorizonTC3	2150	4	14,9	105	70
Datsun	210	2020	4	19,2	85	65
Fiat	Strada Cus	2130	4	14,7	91	69
Buick	SkylarkLim	2670	4	16	151	90
Chevrolet	Citation	2595	6	11,3	173	115
Oldsmobile	Omega	2700	6	12,9	173	115
Pontiac	Phoenix	2556	4	13,2	151	90
Volkswagen	Rabbit	2144	4	14,7	98	76
Toyota	CorollaTer	1968	4	18,8	89	60
Chevrolet	Chevette	2120	4	15,5	98	70
Datsun	310	2019	4	16,4	86	65
Chevrolet	Citation	2678	4	16,5	151	90
Ford	Fairmont	2870	4	18,1	140	88
AMC	Concord	3003	4	20,1	151	90
Dodge	Aspen	3381	6	18,7	225	90
Audi	4000	2188	4	15,8	97	78
Toyota	Corona LB	2711	4	15,5	134	90
Mazda	626	2542	4	17,5	120	75
Datsun	510 Hatch	2434	4	15	119	92
Toyota	Corolla	2265	4	15,2	108	75
Mazda	GLC	2110	4	17,9	86	65
Dodge	Colt	2800	4	14,4	156	105
Datsun	210	2110	4	19,2	85	65
Volkswagen	Rabbit DI	2085	4	21,7	90	48
Volkswagen	Dasher DI	2335	4	23,7	90	48
Audi	5000S DI	2950	5	19,9	121	67
Mercedes	240D	3250	4	21,8	146	67
Honda	Civic1500G	1850	4	13,8	91	67
Renault	LeCar Delx	1835	4	17,3	85	67
Subaru	DL	2145	4	18	97	62
Volkswagen	Rabbit	1845	4	15,3	89	132
Datsun	280-ZX	2910	6	11,4	168	100
Mazda	RX-7 GS	2420	3	12,5	70	88
Triumph	TR7 Coupe	2500	4	15,1	122	72
Ford	Must Cobra	2905	4	14,3	140	84
Honda	Accord	2290	4	17	107	84
Plymouth	Reliant	2490	4	15,7	135	92
Buick	Skylark	2635	4	16,4	151	110
Dodge	Aries SW	2620	4	14,4	156	84
Chevrolet	Citation	2725	6	12,6	173	58
Plymouth	Reliant	2385	4	12,9	135	64
Toyota	Starlet	1755	4	16,9	79	60
Plymouth	Champ	1875	4	16,4	86	67
Honda	Civic1300	1760	4	16,1	81	65
Subaru	210	2065	4	17,8	97	62

ПРОДОВЖЕННЯ ДОДАТКУ 2

1	2	3	4	5	6	7
Datsun	Tercel	1975	4	19,4	85	68
Toyota	GLC 4	2050	4	17,3	89	63
Mazda	Horizon 4	1985	4	16	91	65
Plymouth	Escort 4W	2215	4	14,9	105	65
Ford	Escort 2H	2045	4	16,2	98	74
Ford	Jetta	2380	4	20,7	98	75
Volkswagen	18I	2190	4	14,2	105	75
Renault	Prelude	2320	4	15,8	100	100
Honda	Corolla	2210	4	14,4	107	74
Toyota	200SX	2350	4	16,8	108	80
Datsun	626	2615	4	14,8	119	110
Mazda	505S DI	2635	4	18,3	120	76
Peugeot	900S	3230	4	20,4	141	116
Saab	Diesel	2800	4	15,4	121	120
Volvo	Cressida	3160	6	19,6	145	110
Toyota	810 Maxima	2900	6	12,6	168	105
Datsun	Century	2930	6	13,8	146	88
Buick	Cutlass LS	3415	6	15,8	231	85
Oldsmobile	Granada GL	3725	8	19	350	88
Ford	Lebaron	3060	6	17,1	200	88
Chrysler	Cavalier	3465	6	16,6	225	88
Chevrolet	CavalierSW	2605	4	19,6	112	85
Chevrolet	Cavalier2D	2640	4	18,6	112	84
Chevrolet	1200 Hatch	2395	4	18	112	90
Pontiac	Aries SE	2575	4	16,2	112	92
Dodge	Phoenix	2525	4	16	135	74
Pontiac	Fairmont	2735	4	18	151	68
Ford	Concord DI	2865	4	16,4	140	68
AMC	Rabbit L	3035	4	20,5	151	63
Volkswagen	GLC Cust l	1980	4	15,3	105	70
Mazda	GLC Custom	2025	4	18,2	91	88
Mazda	Horizon	1970	4	17,6	91	75
Plymouth	Lynx l	2125	4	14,7	105	70
Mercury	Stanza XE	2125	4	17,3	98	67
Nissan	Accord	2160	4	14,5	120	67
Honda	Corolla	2205	4	14,5	107	67
Toyota	Civic M	2245	4	16,9	108	110
Honda	Civic A	1965	4	15	91	85
Honda	310 GX	1965	4	15,7	91	92
Datsun	CenturyLmt	1995	4	16,2	91	112
Buick	Cutlass DI	2945	6	16,4	181	96
Oldsmobile	Lebaron	3015	6	17	262	84
Chrysler	Granada l	2585	4	14,5	156	90
Ford	Celica GT	2835	6	14,7	232	86
Toyota	Charger2.2	2665	4	13,9	144	52
Dodge	Camaro	2370	4	13	135	84
Chevrolet	MustangGL	2950	4	17,3	151	79
Ford	Pickup	2790	4	15,6	140	82
Volkswagen	Rampage	2130	4	24,6	97	
Dodge	Ranger	2295	4	11,6	135	
Ford	S-10	2625	4	18,6	120	
Chevrolet		2720	4	19,4	119	

ВИРОБНИЧІ ХАРАКТЕРИСТИКИ ПІДПРИЄМСТВА

Y1	X5	X7	X9	X10
9,26	0,78	1,37	0,23	1,45
12,11	0,68	1,44	0,43	1,37
10,81	0,70	1,42	0,18	1,65
9,35	0,62	1,35	0,15	1,91
9,87	0,76	1,39	0,34	1,68
9,12	0,71	1,27	0,09	1,89
9,37	0,79	1,40	0,21	2,30
10,02	0,76	1,22	0,32	2,62
9,42	0,70	1,20	0,28	2,03
6,61	0,72	1,23	0,48	0,88
4,32	0,68	1,39	0,41	0,62
7,37	0,77	1,38	0,62	1,09
6,64	0,77	1,35	0,50	1,32
5,52	0,72	1,24	1,20	0,68
5,68	0,71	1,28	0,66	1,43
5,22	0,79	1,33	0,74	1,82
6,70	0,79	1,35	0,39	1,24

Y1 - продуктивність праці

X5 - питома вага працюючих у складі виробничого персоналу

X7 - коефіцієнт змінності обладнання

X9 - питома вага втрат через брак, %

X10 - фондівіддача активної частини основних виробничих фондів

СТАТИСТИЧНІ ДАНІ ОБ'ЄМУ ПРОДАЖУ ПРОДУКЦІЇ

A	Y	X1	X2	X3	X4
1	126	4,0	15,0	100,0	1
2	137	4,8	14,8	101,0	2
3	148	7,8	15,2	103,5	3
4	191	6,3	15,5	104,3	4
5	274	8,2	15,5	105,2	5
6	370	9,4	16,0	106,1	6
7	432	14,0	17,0	108,9	7
8	445	5,6	16,9	108,8	8
9	367	9,3	15,8	107,4	9
10	367	6,7	15,8	107,2	10
11	321	5,1	16,8	106,4	11
12	307	4,1	16,1	104,3	12
13	331	8,6	15,4	105,9	13
14	345	5,8	15,7	106,0	14
15	364	5,7	16,0	106,2	15

A - півріччя

Y - об'єм реалізації за 6 місяців, млн. дол.

X1 - витрати на рекламу, млн. дол.

X2 - ціна за одиницю, дол.

X3 - індекс споживчих цін, %

X4 - час

ЗРІСТ ДІВЧАТ І ХЛОПЦІВ МІСТА КИЇВ (2015 Р.)

div10	hlop10	div12	hlop12	div15	hlop15
1	2	3	4	5	6
146	142	153	142	179	167
138	130	153	160	171	174
144	143	149	155	175	163
142	142	155	153	163	175
141	136	164	132	153	176
144	144	151	160	163	160
145	145	155	154	160	158
142	142	153	154	167	159
135	147	159	158	167	160
136	140	156	147	170	164
142	148	158	145	168	161
143	137	158	144	159	175
138	136	160	147	157	164
143	151	156	158	165	168
146	129	164	163	162	160
150	142	155	153	170	165
139	140	159	144	157	166
146	132	153	147	163	169
140	135	161	155	166	168
149	149	160	160	163	171
146	150	155	145	167	169
146	141	155	156	164	162
137	146	149	137	170	167
146	144	157	143	147	166
143	129	162	150	168	169
144	142	157	144	165	170
144	138	139	145	159	170
141	137	139	141	166	175
132	144	144	141	173	165
142	146	142	151	168	175
143	133	157	146	155	173
149	144	153	130	162	169
131	137	156	145	169	172
136	138	158	147	162	177
136	134	160	142	163	175
145	140	160	143	168	176
140	139	151	139	164	182
138	132	153	150	162	174
144	139	149	146	163	160
145	142	148	149	164	180
138	139	155	145	168	176
144	143	158	145	167	160
144	148	157	147	161	169

ПРОДОВЖЕННЯ ДОДАТКУ 5

1	2	3	4	5	6
132	131	155	150	169	172
144	145	142	154	163	163
145	138	147	149	160	156
133	136	154	146	162	174
142	145	143	149	157	164
131	139	157	146	155	168
148	147	155	145	161	175

МАСА ТІЛА ЖІНОК ДО ЕКСПЕРИМЕНТУ ТА ПІСЛЯ

masado	masapi
59,1	58
62,3	62,3
58,6	58
60,2	59,1
63,4	60,2
78,6	68,3
55,4	57
64,9	60,8
65	62
63,2	60,7
72,9	65,4
56,1	55
69,2	62,9
68,8	64
58,8	58
66,2	64,2
62,5	62,8
68,7	65,6
62,2	60
66,5	62
62	60,3
60,1	60
63,2	60,7
61,3	60,3
60,3	59,4
65,1	62,4
70,2	66
65,3	62,3
58,2	58
64,5	60,9

**РЕЗУЛЬТАТИ В СТИБКАХ У ВИСОТУ З МІСЦЯ
ТА У ШТОВХАННІ У ВАЖКОАТЛЕТІВ 1 РОЗРЯДУ**

jump	push
57	107,5
60	110
58	110
61	115
63	115
58	107,5
55	107,5
64	120
65	122,5
64	112,5
66	120
61	110

Навчально-методичне видання
Методичні вказівки до виконання лабораторних робіт по курсу:
«Теорія планування та виконання наукових досліджень»
для студентів спеціальностей: 133 Галузеве машинобудування; 131 Прикладна механіка

Укладачі: Скібінський О.І., Селехова В.М.

м. Кропивницький

