

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНА МЕТАЛУРГІЙНА АКАДЕМІЯ УКРАЇНИ**

**РОБОЧА ПРОГРАМА,  
методичні вказівки та індивідуальні завдання  
до вивчення дисципліни «Математичні методи  
та спеціалізовані програми обробки даних»  
для студентів спеціальності 136 – металургія, (Комп'ютерне проектування в  
металургії МЕ-909)**

**Затверджено  
на засіданні Вченої ради  
академії  
Протокол №1 від 27.01.2017**

**Дніпро НМетАУ 2017**

УДК 658.516:658.562:006.063(075.8)

Робоча програма, методичні вказівки та індивідуальні завдання до вивчення дисципліни «Математичні методи та спеціалізовані програми обробки даних» для студентів спеціальності 136 – металургія, Комп'ютерне проектування в металургії (МЕ-909) / Укл.: І.А. Соловйова, Ю.М. Николаєнко. – Дніпро: НМетАУ, 2017. – 54 с.

Викладені робоча програма, методичні вказівки та контрольні питання до вивчення дисципліни «Математичні методи та спеціалізовані програми обробки даних», наведені основні теоретичні відомості, завдання для самостійної роботи, а також питання для контролю знань.

Призначена для студентів спеціальності **136 – металургія**, Комп'ютерне проектування в металургії (МЕ-909) заочної форми навчання.

Укладачі: І.А. Соловйова, канд. техн. наук, доц.  
Ю.М. Николаєнко, ст. викл.

Відповідальний за випуск В.Ф.Балакін, д-р техн. наук, проф.

Рецензент Н.М. Мосьпан, канд. техн. наук, доц. (НМетАУ)

Підписано до друку .02.2017. Формат 60x84 1/16. Папір друк. Друк  
плоский.

Облік.–вид. арк. . Умов. друк. арк. . Тираж 100 пр. Замовлення №

Національна металургійна академія України  
49600, м. Дніпропетровськ–5, пр. Гагаріна, 4

---

Редакційно–видавничий відділ НМетАУ

## ХАРАКТЕРИСТИКА ДИСЦИПЛІНИ

Навчальна дисципліна «Математичні методи та спеціалізовані програми обробки даних» є базовою і входить до циклу дисциплін самостійного вибору студента.

**Мета вивчення дисципліни** – вивчення основ методології та розкриття логічних передумов, на яких базуються сучасні комп'ютерні методи постановки та обробки результатів інженерного експерименту в металургії. Знайомство з сучасними комп'ютерними спеціалізованими програмними засобами обробки статистичних або експериментальних даних.

У результаті вивчення дисципліни студент повинен:

**знати** – логічні основи постановки і планування інженерного експерименту; методологічні підходи і принципи обробки результатів інженерного експерименту; перелік сучасних комп'ютерних спеціалізованих програмних засобів, їх характеристики; основні можливості вводу–виводу даних для розглянутих програм; основні методики обробки даних та їх реалізація в програмних засобах.

**вміти** – виконувати за допомогою сучасних математичних методів і комп'ютерних функцій систем MS Excel і MathCad обробку результатів експериментів; виконувати розрахунки коефіцієнтів регресійних моделей, аналізувати їх значимість та адекватність моделей; представляти результати експериментів; систематизувати дані для вводу в програмі Statistica, SPSS; ввести дані в програми; виконувати візуалізацію чисельної інформації; зробити вивід результатів статистичного аналізу в програмах у вигляді таблиць або графіків.

**Критерії успішності** – отримання позитивної оцінки при складанні контрольних заходів.

**Засоби діагностики успішності навчання** – захист індивідуальних завдань та екзамен.

**Зв'язок з іншими дисциплінами** – дисципліна є продовженням наступних дисциплін – «Теорія систем та прийняття рішень», «Інформаційні технології в управлінні проектами». Набуті знання і вміння використовуються при опануванні програми підготовки магістрів за фахом та виконанні кваліфікаційної роботи магістра.

## 1. ЗАГАЛЬНІ МЕТОДИЧНІ ВКАЗІВКИ

Структуру вивчення дисципліни «Математичні методи та спеціалізовані програми обробки даних» наведено в таблиці 1.1.

Таблиця 1.1

Загалом годин	Курс/семестри					Примітки
	2/1					
	Лекцій, годин	Лабор. роб., годин/ кількість	Практ. зан., годин/ кількість	Самост. роб., годин	Вид контролю	
120	12	—	20	88	індивідуальне завдання, екзамен	

Робоча програма передбачає самостійну роботу, контрольовану викладачем, що включає:

- вивчення лекційного матеріалу;
- самостійне вивчення розділів дисципліни, які не викладаються на лекціях;
- підготовку до контрольного заходу (індивідуальне завдання, екзамен).

## 2. РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

### Основна література

1. Соловйова І.А., Николаєнко Ю.М. Математична та статистична обробка даних в MS Excel, Statistica. Частина 1, 2: Навч. посібник. – Дніпропетровськ: НМетАУ, 2013. – 52 с., 50с.

2. Соловйова І.А., Николаєнко Ю.М. Застосування MathCad в системному аналізі та інженерних розрахунках. Частина 1–3: Навч. посібник. – Дніпропетровськ: НМетАУ, 2011. – по 52 с.

### Додаткова література

1. Гайдышев И.П. Решение научных и инженерных задач средствами Excel, VBA, C/C++. – СПб.: БХВ–Петербург, 2004. – 512 с.: ил.

2. Методы планирования и обработки результатов инженерного эксперимента: Конспект лекций / Н.А. Спирин, В.В. Лавров. – Екатеринбург: ГОУ ВПО УГТЦ–УПИ, 2004. – 257 с.

3. Боровиков В.П. Statistica. Искусство анализа данных на компьютере: для профессионалов / В.П. Боровиков. – СПб.: Питер, 2003. – 688 с.: ил.

4. Макарова Н.В. Статистика в Excel: учебное пособие / Н.В. Макарова, В.Я. Трофимец. – М.: Финансы и статистика, 2002. – 368 с.: ил.

### 3. ПРОГРАМА І ПИТАННЯ ДЛЯ ЗАКЛЮЧНОГО КОНТРОЛЮ

Розподіл навчальних годин за темами і видами занять наведено в таблиці 3.1.

Таблица 3.1

№№ тем	Назва розділу / теми та її зміст	Види занять			
		ауд.	лекц.	практ.	самост.
1.	<b>Математичні методи обробки результатів експерименту</b> Методи первинної обробки результатів експериментів. Математичні прийоми аналізу та обробки результатів експерименту. Розрахунки кількісних характеристик. Функції програмного забезпечення обробки даних – середніх, математичного сподівання, дисперсії.	12	4	8	18
2.	<b>Системи обробки даних</b> Сучасні комп'ютерні програми обробки даних. Коротка характеристика системи Statistica. Інтерфейс системи. Статистичні процедури. Модулі системи. Приклади використання системи Statistica при обробці експериментальних даних. Типи графіків системи. Побудова графіків статистичних показників.	8	4	4	22
3.	<b>Промислова статистика.</b> Види статистичного аналізу в Statistica. Регресія в системі Statistica.	8	2	6	22
4.	<b>Контрольні карти якості.</b> Огляд можливостей системи SPSS.	4	2	2	26
<b>Загалом:</b>		32	12	20	88

## 4. ЛЕКЦІЙНИЙ МАТЕРІАЛ

### **Тема 1. Методи первинної обробки результатів експериментів.**

Математичні прийоми аналізу та обробки результатів експерименту. Розрахунки кількісних характеристик. Функції програмного забезпечення для обробки даних – середніх, математичного сподівання, дисперсії (8 годин).

Обчислення, що виконуються вченими та інженерами, можуть бути різними за складністю. Для вирішення складних задач може знадобитися значна кількість часу. Будь-який інструмент, здатний спростити чисельні розрахунки, значно підвищує продуктивність наукового працівника або інженера. Такими інструментами є універсальні та спеціалізовані програми обробки даних. Програмне забезпечення статистичних досліджень достатньо розвинуте. На сьогоднішній день відомо багато статистичних пакетів прикладних програм статистичної обробки інформації. Серед них є Statistica, SPSS, SAS, MathCAD, MathLab, Statgraphics та інші.

Прикладами найбільш поширених програм є MS Excel та Statistica, що забезпечують засоби для виконання обчислень, побудови графіків і діаграм, а також можливості для програмування в єдиному пакеті, користуватися яким просто і зручно.

Всі статистичні процедури систем розділено на окремі модулі, кожен з яких об'єднує групу логічно зв'язаних між собою статистичних методів і в рамках конкретної моделі забезпечує повний і всебічний аналіз закономірностей. У кожному модулі можна виконати певний спосіб обробки даних, не звертаючись до інших модулів. У системі Statistica такими модулями є: основні статистики та таблиці, множинна регресія, дисперсійний аналіз, багатовимірний аналіз та ін.

MS Excel та Statistica містять велику кількість інструментів, що дозволяють виконати обробку та аналіз даних, які відносяться до самих різних сфер діяльності людини. При статистичній обробці даних спостережень або експериментів, як правило, увага дослідників звертається не тільки на окремі значення, скільки на поведінку всього масиву даних. Під час аналізу отриманих результатів типовими питаннями є:

- Як розподіляються дані?
- В якій області вони знаходяться?

- Яка величина середнього значення і дисперсії?
- Як пов'язані між собою вхідні і вихідні величини?

Аналіз статистичних даних у системі Statistica проводиться за певною схемою.

1. Робота розпочинається з запуску системи Statistica (аналогічно запуску інших прикладних програм – через меню Пуск або використовуючи ярлик).

2. Формування файлу з даними.

3. Перетворення та візуалізація первинних даних (якщо потрібно).

4. Виклик Стартової панелі модуля, в якому проводитиметься аналіз. Стартові панелі різних модулів деякою мірою схожі між собою. В них обов'язково міститься кнопка для вибору змінних для аналізу, перелік опцій для визначення початкових умов аналізу і, звичайно, кнопка виконання команд.

5. Наступним вікном є вікно Результатів аналізу. Воно складається з двох частин – інформаційної та функціональної. В інформаційній частині міститься інформація про загальні результати аналізу, а у функціональній – ряд опцій, що дозволяють змінити зовнішній вигляд отриманих результатів.

6. Усі результати аналізу можна зберегти окремим файлом, що дозволить у будь-який момент продивитися результати аналізу, не проводячи його. Після проведення аналізу формується звіт. Після чого дається математична та економічна інтерпретація отриманих результатів. Обов'язково робляться висновки.

### **Первинна обробка даних у ППП Statistica за допомогою модуля «Основная статистика / Таблицы»**

STATISTICA – пакет для всебічного статистичного аналізу, розроблений компанією StatSoft. У пакеті STATISTICA реалізовані процедури для аналізу даних (data analysis), керування даними (data management), видобутку даних (data mining), візуалізації даних (data visualization).

Меню займає другий рядок основного вікна програми й містить у собі систему меню, що випадають. Ряд пунктів меню, таких як **File (Файл), Edit (Виправлення), View (Вид), Windows (Вікно), Help (Довідка)**, стандартний для Windows, пункт Analysis (Аналіз) специфічний для Statistica (рис. 4.1).

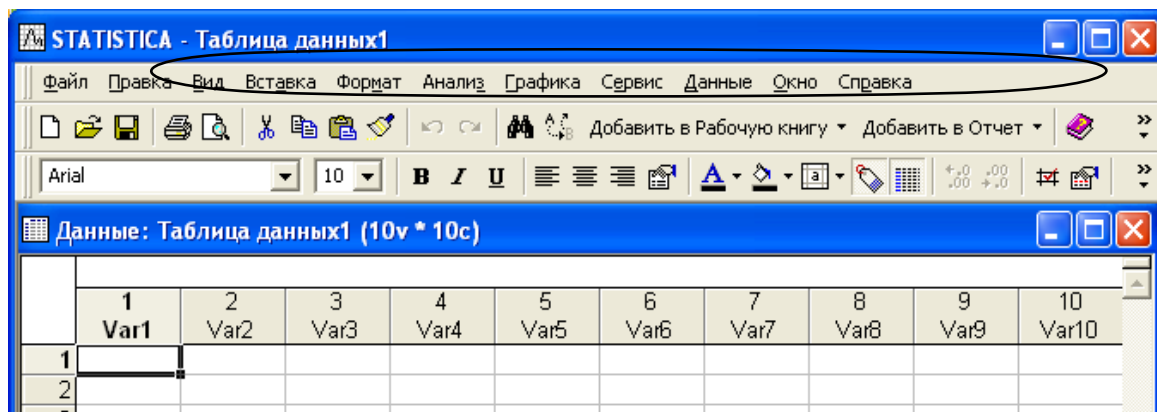


Рис. 4.1. Меню программы Statistica

Панель інструментів займає третій і четвертий рядок робочого вікна й розташовується нижче рядка меню.

Панель інструментів містить кнопки для швидкого доступу до найбільше часто використовуваних команд меню. Кожному типу документа Statistica відповідає своя панель інструментів. Зовнішній вигляд панелі інструментів і її розташування у вікні системи можна налаштувати за допомогою команди Toolbar (Панель інструментів) з меню View (Вид) (рис. 4.2). Ці установки діють тільки для поточного сеансу роботи.

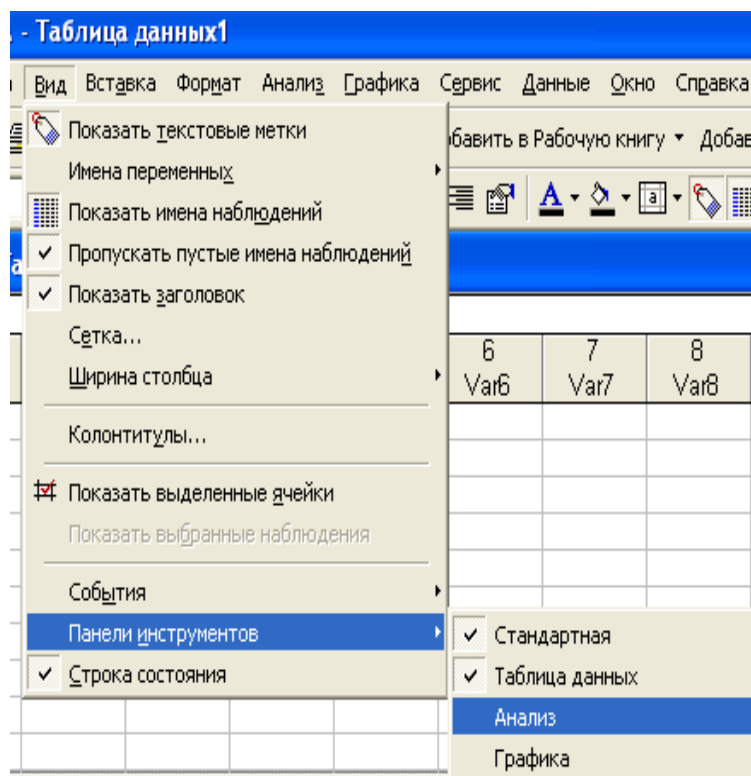


Рис. 4.2. Панель інструментів з меню Вид



Вхідні дані в системі Statistica організовані у вигляді електронної таблиці. Таблиці з вхідними даними в Statistica носять особливу назву (розширення \*.sta), щоб не плутати їх з іншими таблицями системи.

Statistica використовує стандартний інтерфейс електронних таблиць. Поточний файл даних завжди відображається у вигляді електронної таблиці. Дані організовані у вигляді спостережень і змінних. Спостереження можна розглядати як еквівалент стовпців електронної таблиці. Кожне спостереження складається з набору значень змінною.

Електронна таблиця системи SPREADSHEETS складається з рядків і стовпців. На відміну від звичайних електронних таблиць, де рядка й стовпці рівноправні, в Statistica вони мають різні значення.

Стовпці електронної таблиці з вхідними даними мають назву Variables (Змінні), а рядки – Cases (Випадки). У якості змінних звичайно виступають досліджувані величини, а випадки – це значення, які приймають змінні в окремих вимірах.

Система може працювати як із чисельними, так і з текстовими даними, що, звичайно, важливо в практичних статистичних дослідженнях. Зокрема, електронні таблиці можуть містити й чисельну, і текстову інформацію. Аналогічно MS Excel вони підтримують різні типи операцій з даними, такі як операції з використанням буфера обміну Windows; операції з виділеними блоками значень, у тому числі й з використанням методу drag-and-drop, автозаповнення блоків і т. д.

### ***Модульна структура системи Statistica***

STATISTICA організована за модульним принципом. Це означає, що всі методи статистичної обробки, реалізовані в системі, розбиті на кілька груп-модулів відповідно до розділів статистичного аналізу.

Наприклад, модуль Basic Statistica and Tables (Основні статистики й таблиці) містить основні описові статистики, методи статистичного аналізу різних таблиць, різнобічний інструментарій для проведення розвідницького аналізу даних.

У модулі Cluster Analysis (Кластерний аналіз) реалізовані методи кластерного аналізу.

Модуль Time Series Analysis/Forecasting (Аналіз тимчасових рядів і прогнозування) містить методи аналізу тимчасових рядів і прогнозування.

Кожний модуль є окремим Windows–додатком і може працювати незалежно від інших модулів системи.

Модулі STATISTICA відкриваються простим клацанням миші на вкладці Аналіз (Analysis) (рис. 3.3).

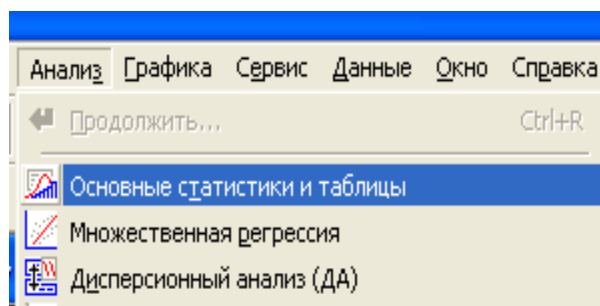


Рис. 4.3. Вкладка Аналіз

Модуль Основні статистики/таблиці запуститься, і робоче вікно системи з'явиться на екрані.

Після вибору в Перемикачі модулів відкривається стартова панель обраного модуля. Далі ви відкриваєте файл даних, задаючи, при необхідності, умови вибору випадків – кнопка Select cases (Вибрати випадки) – і ваги змінних – кнопка Weight.

Потім необхідно:

- обрати змінні для аналізу з відкритого файлу даних;
- зробити запуск обчислювальної процедури;
- використовуючи графічні можливості й спеціальні таблиці висновку з обчисленими різноманітними статистиками, здійснити всебічний перегляд і аналіз результатів у вікні Results (Результати);
- вибрати наступний крок аналізу.

**Приклад 4.1.** На підприємстві проводяться вимірювання діаметра партії труб після обробки (в міліметрах). Довільним чином відібрали результати 55 представників з партії (табл. 3.1). Виконати первинну обробку даних.

#### *Рішення*

1. У клітинку A1 введемо заголовок «Наблюдения», а в діапазон A2:A56 – емпіричні дані (див. табл. 3.1).
2. Виконати первинну обробку даних (описову статистику) в програмі Statistica на прикладі даних табл. 3.1.

Таблиця 4.1

## Емпіричні дані спостережень діаметрів труб

Спостереження				
64	62	58	63	61
57	62	63	58	58
63	60	61	60	60
62	64	59	59	64
58	61	62	60	60
61	59	60	59	59
63	59	60	61	61
60	63	58	62	64
60	61	61	62	62
61	62	60	63	59
65	58	63	57	65

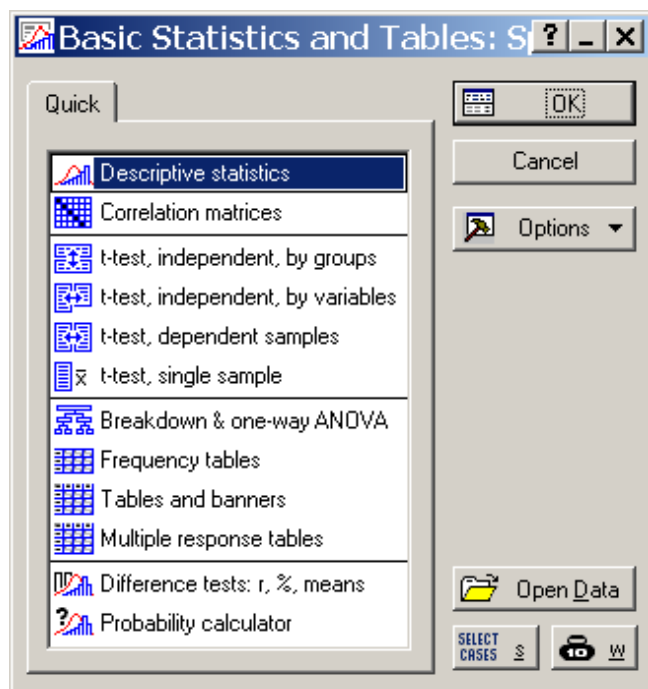


Рис. 4.4. Стартове вікно модуля  
з переліком статистичних процедур

Розглянемо первинну обробку даних (описову статистику) в програмі Statistica.

Стартове вікно модуля «Основная статистика/ Таблицы» представлено на рис. 4.4.

За даними прикладу 4.1 є вибірка об'ємом 55 вимірювань. Файл даних містить одну змінну. Створення файлу Statistica аналогічно створенню таблиці MS Excel, але при створенні нового документа необхідно вказати число створюваних стовпців (число

змінних) та рядків (число реєстрів) (рис. 4.5).

Дані скопіюємо з таблиці MS Excel або введемо в стовпчик Statistica. Назву змінної редагуємо, активуючи вікно «Опису змінних» подвійним клацанням по стовпцю. Після цього вікно файлу даних має вигляд, показаний на рис. 4.6.

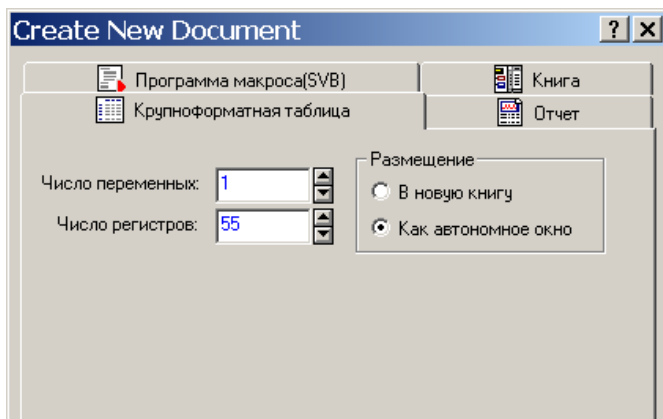


Рис. 4.5. Діалогове вікно створення нового документа – таблиці

	1				
	Наблюдения				
1	64				
2	57				
3	63				
4	62				
5	58				
6	61				
7	63				
8	60				
9	60				
10	61				
11	65				
12	62				

Рис. 4.6. Вікно файлу даних

Визначення змінних для процедур статистичної обробки здійснюється у вікні Variables (рис. 4.7).

Одна із найпростіших можливостей програми – висновок певного списку статистичних показників за допомогою команди «Статистика даних блоку», яка застосовується до виділеного стовпця або рядку, результати поміщаються безпосередньо під блоком (рис. 4.8).

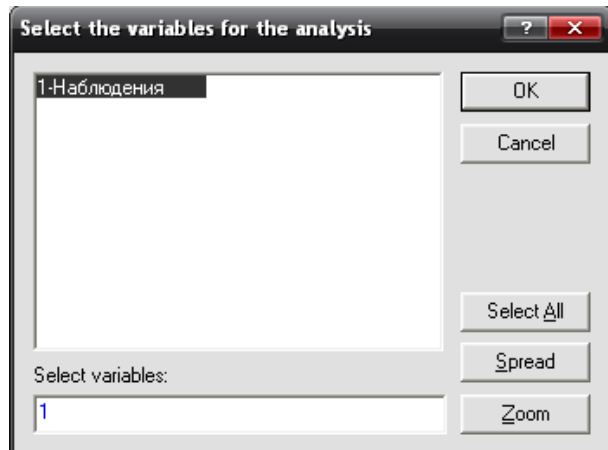


Рис. 4.7. Вибір змінних для обробки

	1				
	Наблюдения				
52	64				
53	62				
54	59				
55	65				
56					
MEAN case 1-56	60,8545455				
MEDIAN case 1-56	61				
SD case 1-56	2,04955444				
VALID_N case 1-56	55				
SUM case 1-56	3347				
MIN case 1-56	57				
MAX case 1-56	65				
25th% case 1-56	59				
75th% case 1-56	62				

Рис. 4.8. Статистика даних блоку

Щоб вибрати статистики, що підлягають обчисленню, зручніше скористатися вкладкою Advanced вікна Descriptive Statistics (рис. 4.9), в якій прапорцями визначаються статистичні параметри.

Після натискання на ОК вікна Descriptive Statistics на екрані з'являється таблиця з результатами описових статистик (рис. 4.10):

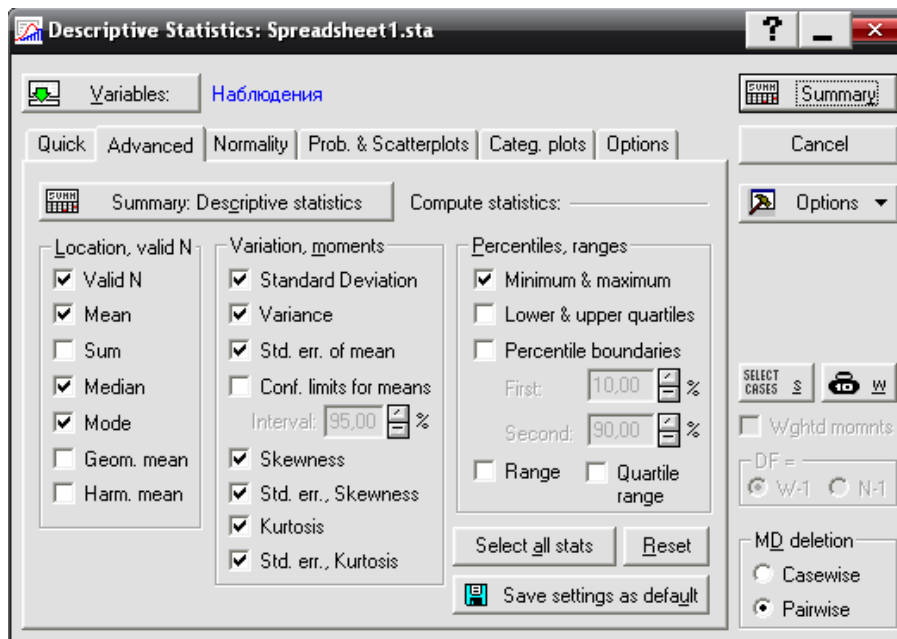


Рис. 4.9. Вибір розрахункових статистик

Descriptive Statistics (Spreadsheet1.sta)													
Valid N	Mean	Median	Mode	Frequency of Mode	Minimum	Maximum	Variance	Std.Dev.	Standard Error	Skewness	Std.Err. Skewness	Kurtosis	Std.Err. Kurtosis
55	60,85455	61,00000	60,00000	10	57,00000	65,00000	4,200673	2,049554	0,276362	0,096321	0,321742	-0,743698	0,633507

Рис. 4.10. Обчислення статистики

Valid N – об’єм вибірки;

Mean – середня арифметична;

Sum – сума;

Median – медіана;

Mode – мода;

Standard Deviation (Std.Dev. або SD) – стандартне відхилення;

Variance – дисперсія;

Standard error of mean – стандартна помилка середнього;

95% confidence limits of mean – довірчий інтервал для середнього;

Minimum, maximum – мінімальне і максимальне значення;

Lower, upper quartiles – нижній і верхній квартили (верхній квартиль – це таке значення випадкової величини, більше якого за величиною 25% вибірки. Нижній квартиль – це таке значення випадкової величини, менше якого за величиною 25% вибірки);

Range – розмах;

Quartile range – інтерквартільна широта (відстань між верхнім і нижнім квантилем);

Skewness – асиметрія;

Standard error of Skewness – стандартна помилка асиметрії;

Kurtosis – ексцес;

Standard error of Kurtosis – стандартна помилка ексцесу.

На першому етапі обробки даних часто виникає необхідність їх угруповання. Угруповання дозволяє представити первинні дані в компактному вигляді, виявити закономірності досліджуваної величини. За групованими даними будуються гістограми і таблиці. Для побудови гістограм і таблиць використовується група кнопок Distribution вкладки Normality (рис. 4.11). Під кнопками знаходяться дві опції Categorization (Угруповання), що дозволяють задати число інтервалів угруповання або встановити величину інтервалу, який дорівнює цілому числу (перемикач Integer intervals (categories)). Результати угруповання діаметрів (змінна «Наблюдения») по цілим числам наведені на рис. 4.12, а по інтервалах – на рис. 4.13.

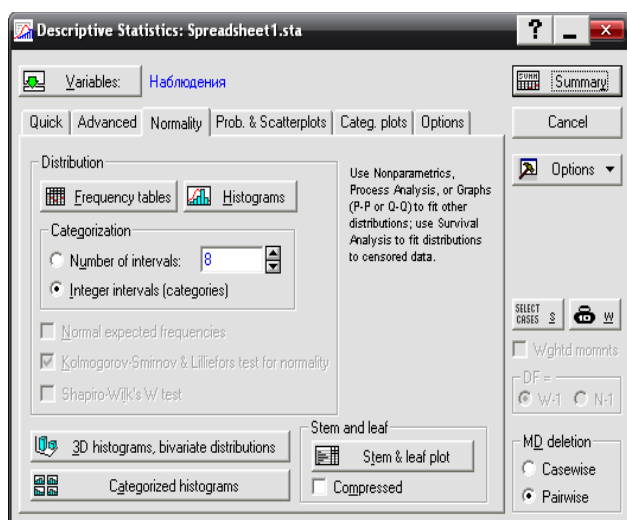


Рис. 4.11. Група кнопок Distribution для групованих даних

Frequency table: Наблюдения (Spreadsheet1.sta)						
K-S d=.11619, p>.20; Lilliefors p<.10						
Category	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
56,00000<x<=57,00000	2	2	3,63636	3,6364	3,07692	3,0769
57,00000<x<=58,00000	6	8	10,90909	14,5455	9,23077	12,3077
58,00000<x<=59,00000	7	15	12,72727	27,2727	10,76923	23,0769
59,00000<x<=60,00000	10	25	18,18182	45,4545	15,38462	38,4615
60,00000<x<=61,00000	9	34	16,36364	61,8182	13,84615	52,3077
61,00000<x<=62,00000	8	42	14,54545	76,3636	12,30769	64,6154
62,00000<x<=63,00000	7	49	12,72727	89,0909	10,76923	75,3846
63,00000<x<=64,00000	4	53	7,27273	96,3636	6,15385	81,5385
64,00000<x<=65,00000	2	55	3,63636	100,0000	3,07692	84,6154
Missing	10	65	18,18182		15,38462	100,0000

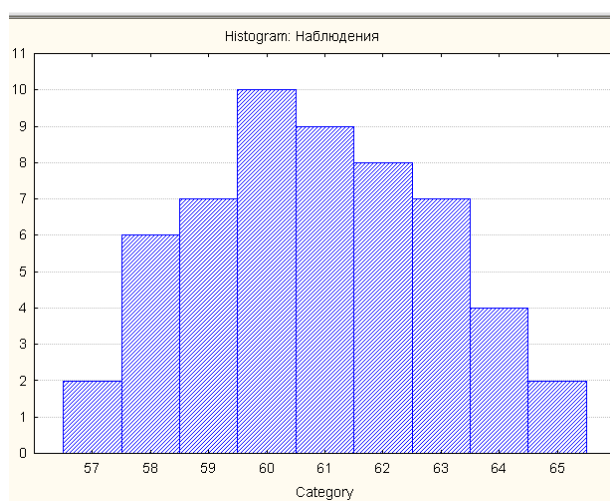
Рис. 4.12. Таблиця згрупованих даних по цілим значенням

У таблицях (див. рис. 4.12, 4.13) перший стовпець – кількість, другий – кількість з накопиченням, третій – %, четвертий – % з накопиченням, п'ятий – % від загальної кількості.

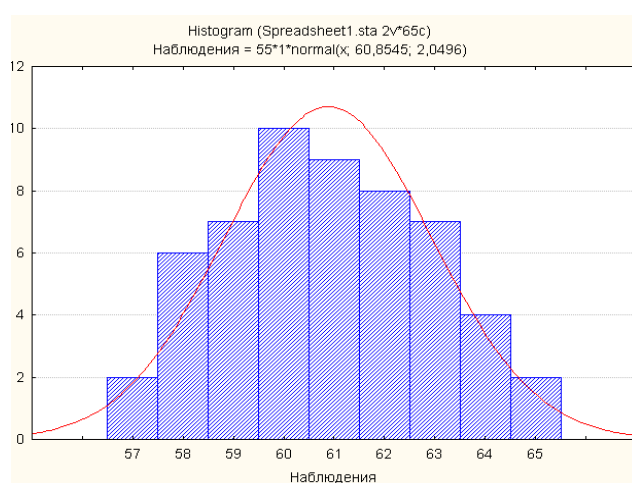
Frequency table: Наблюдения (Spreadsheet1.sta) K-S d=.11619, p> .20; Lilliefors p<.10 Shapiro-Wilk W=.96563, p=.11667										
Category	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All	Expected Count	Cumulative Expected	Percent Expected	Cumulative % Expected
56,00000<x<=58,00000	8	8	14,54545	14,5455	12,30769	12,3077	4,50151	4,50151	8,18456	8,18456
58,00000<x<=60,00000	17	25	30,90909	45,4545	26,15385	38,4615	14,10831	18,60982	25,65148	33,83604
60,00000<x<=62,00000	17	42	30,90909	76,3636	26,15385	64,6154	20,54347	39,15330	37,35177	71,18781
62,00000<x<=64,00000	11	53	20,00000	96,3636	16,92308	81,5385	12,41313	51,56643	22,56933	93,75714
64,00000<x<=66,00000	2	55	3,63636	100,0000	3,07692	84,6154	3,10205	54,66848	5,64010	99,39724
Missing	10	65	18,18182		15,38462	100,0000				

Рис. 4.13. Таблиця згрупованих даних по інтервалах (обрано 5 інтервалів)

Уявімо розподілені дані на Гістограмах (рис. 4.14). Для цього призначена кнопка Histograms вікна Descriptive Statistics.



а



б

Рис. 4.14. Гістограма: а – згрупованих даних по цілим числам; б – з накладанням щільності нормального розподілу

На гістограму можна накласти щільність нормального розподілу, перевірити близькість розподілу до нормального вигляду за допомогою критеріїв Колмогорова–Смирнова, Лілієфорса; обчислити статистику Шапіто–Уїлкса (рис. 4.15). Для цього в групі опцій Distribution необхідно включити відповідні прапорці (див. рис. 4.11).

Про нормальність розподілу можна судити за графіком на нормальному ймовірному папері (опції Normal probability plots вікна Descriptive Statistics, рис. 4.16). Чим ближче розподіл до нормального вигляду, тим краще значення лягають на пряму (рис. 4.17).

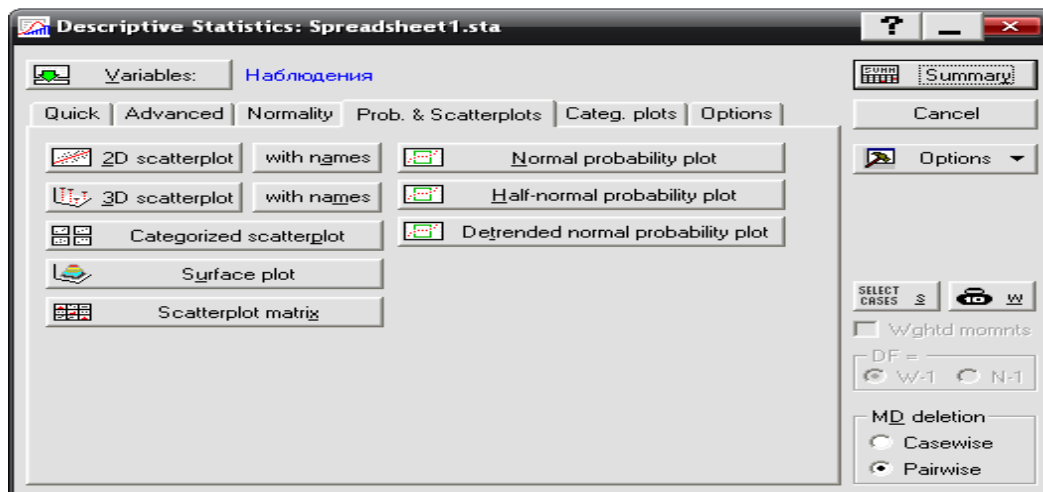


Рис. 4.15. Вибір опцій побудови гістограм і графіків

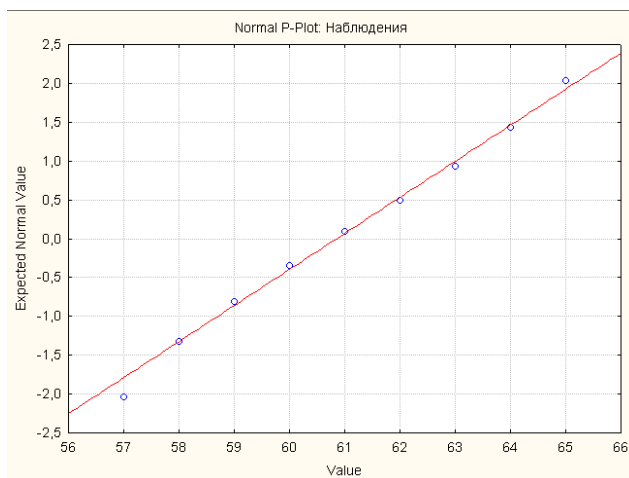


Рис. 4.16. Графік на нормальному ймовірному папері для вибірки спостережень діаметрів

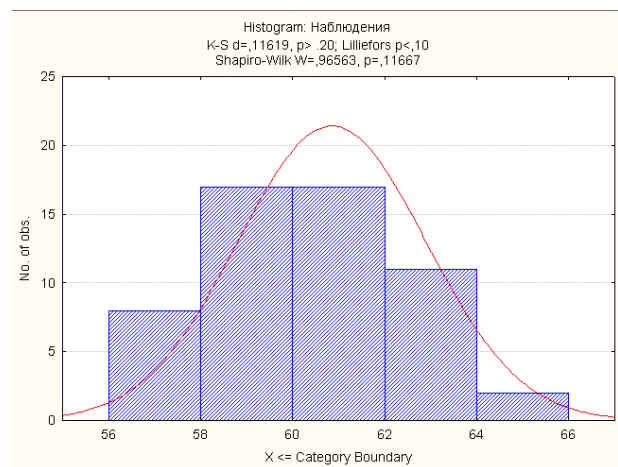


Рис. 4.17. Гістограма згрупованих даних по 5-х інтервалах з висновком статистик

На гістограмі (див. рис. 4.17) статистика Колмогорова–Смирнова дорівнює 0,116. Чим менше величина цієї статистики, тим ближче розподіл випадкової величини до нормального. Ймовірність нульової гіпотези (p) більше 0,2. Однак детальна перевірка гіпотези про нормальність вибірки вимагає досить значних об'ємів вибірки, не менше 100.

### Самостійне вивчення (18 годин)

Етапи робіт, що передують обробці експериментальних даних. Побудова активного та пасивного експеримента. Перевірка статистичних гіпотез.



### **Контрольні питання**

1. Етапи робіт, що передують обробці експериментальних даних.
2. Побудова пасивного експерименту.
3. Побудова активного експерименту.
4. Мета експерименту. Похибка вимірювань.
5. Перелічить мінімальний набір статистичних методів аналізу.
6. Назвіть основні програмні засоби, які використовуються для аналізу даних.
7. Типи розподілу даних експеримента, побудова гістограм розподілу.
8. Перелічить групи основних статистичних функцій.
9. Первинна обробка даних у MS Excel за допомогою модуля «Описова статистика».
10. Функції MS Excel для обробки даних.
11. Первинна обробка даних у ППП Statistica за допомогою модуля «Основная статистика».
12. Побудова гістограм розподілу даних в різних ПП.

**Тема 2. Системи обробки даних.** Сучасні комп'ютерні програми обробки даних. Коротка характеристика системи Statistica. Інтерфейс системи. Статистичні процедури. Модулі системи (4 години).

Статистика дозволяє адекватно оцінювати ситуацію, що складається, і виявляти тенденції, приймати оперативні і стратегічні рішення. В умовах сучасної ринкової економіки статистична інформація стала важливим інструментом боротьби за виживання на ринку. Тому пакети статистичного аналізу даних є настільним робочим інструментом фахівців будь-якого рівня. А для фахівця в області управління і економіки знання статистичних методів обробки інформації і сучасних комп'ютерних технологій, які дозволяють автоматизувати громіздкі розрахунки, абсолютно потрібні.

Сучасний технолог повинен володіти декількома основними програмними засобами, в яких закладені методи статистичного аналізу. MS Excel пропонує широкий діапазон засобів для аналізу статистичних і експериментальних даних. До групи статистичних функцій входять функції кореляційного аналізу. Окрім вбудованих засобів можна використати надбудову Пакет аналізу (Сервис/Надстройки/Анализ данных). MathCad також

має розвинений апарат роботи із завданнями математичної статистики і обробки експерименту. По–перше, є велика кількість вбудованих спеціальних функцій, що дозволяють розраховувати щільність вірогідності і інші основні характеристики основних законів розподілу випадкових величин. Разом з цим, в MathCad запрограмована відповідна кількість генераторів псевдовипадкових чисел для кожного закону розподілу, що дозволяє ефективно проводити моделювання методами Монте–Карло. По–друге, передбачена можливість побудови гістограм і розрахунку статистичних характеристик вибірок випадкових чисел і випадкових процесів, таких як середні, дисперсії, кореляції і т. п. При цьому випадкові послідовності можуть як створюватися генераторами випадкових чисел, так і вводитися з файлів. По–третє, є цілий арсенал засобів, спрямованих на інтерполяцію–екстраполяцію даних, побудову регресії по методу найменших квадратів, фільтрацію сигналів. різних інтегральних перетворень, що здійснюють розрахунок, що дозволяє організувати спектральний аналіз різного типу.

Потреба в засобах статистичного аналізу даних у різних областях діяльності, особливо в науці, дуже велика, що і послужило причиною розвитку ринку комп'ютерних програм для статистичної обробки даних. За останні 20 років активний розвиток отримали комп'ютерні програми, що дозволяють проводити статистичний аналіз великих об'ємів даних. Мета аналізу – виявлення закономірностей, порівняння можливих альтернатив вибору, побудова прогнозів розвитку подій, виявлення зв'язків між явищами і процесами. Існуючі програми постійно удосконалюються в частині прискорення роботи з даними, поліпшення представлення результатів аналізу даних, підвищення зручності інтерфейсу, вдосконалення довідкової системи, збільшення числа вбудованих в програму статистичних процедур, засобів обробки даних та інше.

Галузь розвивається стрімкими темпами. На сьогоднішній день на ринку представлено близько тисячі комп'ютерних програм для статистичної обробки даних (далі – статистичні пакети). Різноманітність статистичних пакетів обумовлена багатоплановістю задач обробки даних із застосуванням різних типів статистичних процедур аналізу для пошуку відповідей на запитання з різних галузей людської діяльності.

Перед користувачами різних категорій постає питання вибору оптимального статистичного пакета для пошуку вірних відповідей на існуючі

питання. Очевидно, що оптимальним є варіант, що поєднує в собі необхідні функціональні можливості, високу якість роботи і помірну ціну. При виборі пакета враховуються такі параметри:

- відповідність характеру вирішуваних задач;
- об'єм оброблюваних даних;
- вимоги, що ставляться до кваліфікації користувача (рівень знань в області статистики);
- наявне комп'ютерне обладнання.

Статистичні пакети за ознакою функціональності можуть бути розділені на три основні групи.

1. Універсальні пакети, або пакети загального призначення (наприклад, SPSS, STATA, Statistica, S –PLUS, Stadia, STATGRAPHICS, SYSTAT, Minitab). Ці пакети не орієнтовані на специфічну предметну область і можуть застосовуватися для аналізу даних з різних галузей діяльності. Як правило, вони пропонують широкий діапазон статистичних методів і мають відносно простий інтерфейс. Багатопрофільність універсального пакету дозволяє провести пробний аналіз різних типів даних з використанням широкого діапазону статистичних методів. Більшість існуючих універсальних пакетів мають багато перетинів за складом вбудованих статистичних процедур.

Для того щоб статистичний пакет вважався універсальним, він повинен задовольняти ряду вимог:

- містити досить широкий набір стандартних статистичних методів;
- бути досить простим для швидкого освоєння;
- працювати з досить великими базами даних і відповідати високим вимогам до введення, перетворенню і організації зберігання даних;
- здійснювати обмін даними з широко поширеними пакетами і базами даних;
- мати великий набір засобів графічного представлення даних і результатів їх аналізу;
- мати докладний документаційний супровід та довідкову систему.

2. Професійні пакети (наприклад, SAS, BMDP). Вони відрізняються від універсальних тим, що дозволяють працювати із надвеликими об'ємами даних, застосовувати вузькоспеціалізовані методи аналізу, створювати власну систему обробки даних і, відповідно, вимагають професійних знань.

3. Спеціалізовані пакети (наприклад, BioStat, MESOSAUR, DATASCOPE). У деяких областях діяльності аналізовані дані настільки специфічні, що до них слід застосовувати особливі методи статистичного аналізу, як правило, не представлені в універсальних пакетах. Спеціалізовані пакети дозволяють проводити аналіз з використанням обмеженого числа спеціалізованих статистичних методів або застосовні до використання для вирішення питань, що відносяться до окремо взятої предметної області. Як правило, з подібними статистичними пакетами працюють фахівці, добре знайомі з методами аналізу даних в тій області, на яку орієнтований пакет. Так, статистичний пакет BioStat створений для аналізу даних у галузі біології. Російський статистичний пакет MESOSAUR спеціалізується на аналізі одновимірних і багатовимірних числових рядів і побудові регресійних моделей. Ще один російський статистичний пакет DATASCOPE спеціалізується на проведенні аналізу багатовимірних даних.

Враховуючи різноманіття пропозиції, часом буває складно зробити правильний вибір. На думку М. Мітчелла, який має 20-річний досвід роботи зі статистичними пакетами та 11-річний досвід роботи в якості консультанта за статистикою в Каліфорнійському університеті в Лос-Анджелесі, статистичний пакет – всього лише інструмент в руках майстра. Якщо фахівець не володіє достатніми знаннями, то навіть найдосконаліший програмний продукт не дозволить провести якісний аналіз даних. У той же час неправильно підібраний пакет, який не володіє необхідними для аналізу технічними характеристиками, здатний уповільнити роботу навіть видатного вченого, ускладнивши виявлення необхідних закономірностей і отримання вірних результатів аналізу даних.

Необхідно відзначити, що існує мінімальний набір статистичних методів аналізу, який включений в усі розглянуті пакети:

- описова статистика (базові статистичні методи, перевірка нормальності розподілу даних);
- дисперсійний аналіз;
- непараметрична статистика (аналіз таблиць спряженості, непараметричні порівняння, дисперсійний аналіз);
- контроль якості;
- аналіз виживаємості;
- кластерний аналіз;

- факторний аналіз;
- дискримінантний аналіз;
- регресійний аналіз;
- обробка даних (сортування, відбір, трансформація даних).

Проте провідним пакетом статистичного аналізу є система Statistica, яка ґрунтована на найсучасніших технологіях, повністю відповідає останнім досягненням в області ІТ, дозволяє вирішувати будь-які завдання в області аналізу і обробки даних, ідеально підходить для застосування у будь-якій області: маркетингу, фінансах, страхуванні, економіці, бізнесі, промисловості, медицині та ін. Наприклад, рішення поставлених на початку статті завдань, може бути реалізовано на базі промислових модулів Statistica (карти контролю якості, планування експериментів, аналіз процесів).

Statistica (американської компанії StatSoft, <http://www.statsoft.com>) – система, що реалізовує відомі методи статистичної обробки і візуалізації даних, управління базами даних і розробки призначених для користувача застосувань за допомогою вбудованої мови програмування Statistica Basic.

Завдяки широкому набору процедур аналізу Statistica застосовується в наукових дослідженнях, техніці, бізнесі. Statistica широко використовується в учбовому процесі. Окрім загальних статистичних і графічних засобів в системі є спеціалізовані модулі, наприклад, для проведення соціологічних або біомедичних досліджень, рішення технічних і, що дуже важливо, промислових завдань: Карти контролю якості, Аналіз процесів і Планування експерименту. Модуль Карти контролю дозволяє автоматизувати процес контролю за якістю вироблюваної продукції, аналізувати причини появи відхилень від планових специфікацій. Statistica здійснює аналіз придатності (придатності процесів/механізмів), як одній з найважливіших характеристик виробничого процесу. Обчислення показників (чи індексів) придатності дозволяє дати відповідь на важливе питання: яка кількість виробів потрапляє в задані межі інженерного допуску?

Таким чином, Statistica є однією з найбільш простих для непідготовленого користувача систем, з найменшим періодом оволодіння її можливостями і вдалим набір графічних можливостей.

Набори файлів даних системи Statistica (розширення \*.sta) можна розглядати як «робочі книги» файлів, оскільки вони містять і автоматично

зберігають інформацію про усі додаткові файли (наприклад, графіках, звітах і програмах), які використовуються з поточним набором даних.

Система складається з ряду модулів, працюючих незалежно. Кожен модуль включає певний клас процедур. Майже усі процедури є інтерактивними, тобто для запуску обробки необхідно вибрати з меню змінні і відповісти на низку запитань системи.

Модулі і процедури: Описові статистики, Регресійний аналіз (у тому числі і багатовимірний, нелінійний), Дисперсійний аналіз, Непараметрична статистика, Аналіз багатовимірних таблиць, Підгонка розподілів, Промислова статистика, Кореляційний аналіз, Кластерний аналіз, Прогнозування тимчасових рядів, Аналіз Монте–Карло та ін. Види аналізу наведені на рисунках 4.17– 4.21.

Statistica дозволяє будувати різні типи графіків (рис. 4.22): Матричні графіки, Піктографики, Діаграми розсіювання, Гістограми, Тернарні графіки, Карті ліній рівня, Кругові діаграми, Категоризовані, Імовірнісні, Графіки поверхонь, та ін.

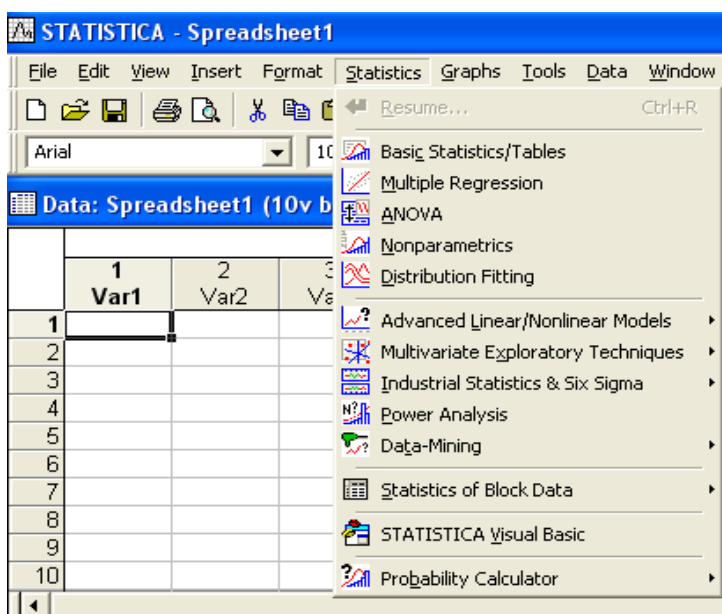


Рис.4.17. Основне меню модуля  
Статистики/таблиці системи Statistica

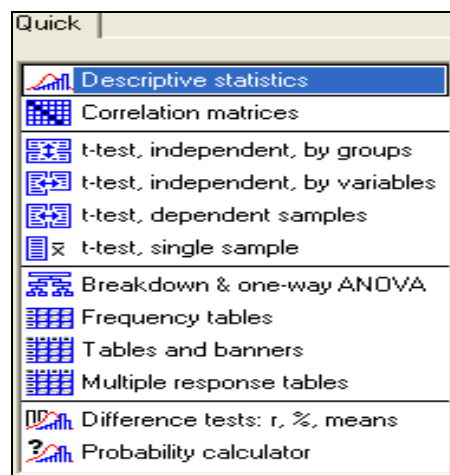


Рис. 4.18. Процедури Basic  
Statistics/Tables  
(Статистики/таблиці системи)

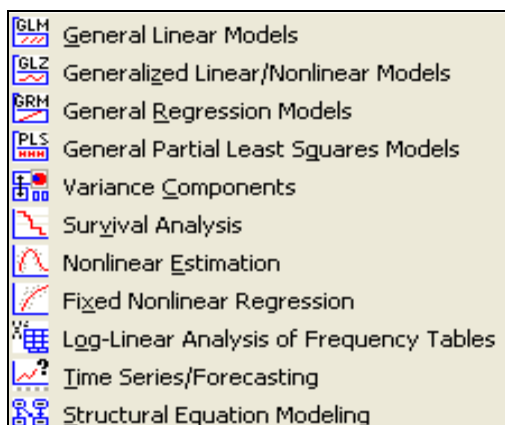


Рис. 4.19. Типи аналізу даних (моделей) модуля Advanced Linear/Nonlinear Models (прогресивні лінійні/нелінійні моделі)

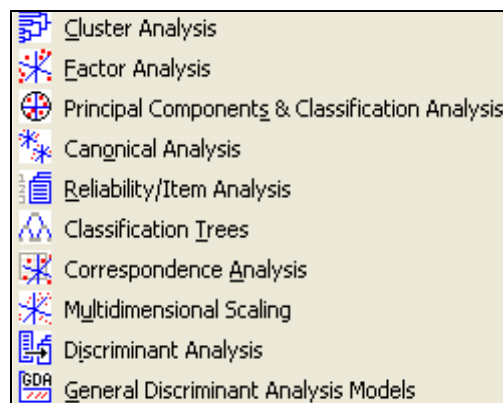


Рис. 4.20. Типи аналізу модуля Multivariate Exploratory Techniques (Багатомірні Дослідницькі методи)

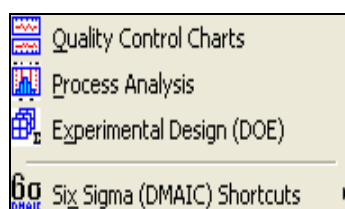


Рис. 4.21. Процедури модуля Industrial Statistic and Six Sigma (промислова статистика і статистика 6–ти сигм)

Statistica включає в себе велику кількість різноманітних категорій і типів графіків. Це всілякі графіки на площині (див. рис. 4.22 – 2D лінії) й у просторі (див. рис. 4.22 – 3D лінії), включаючи аналітичні графіки, ділові графіки та діаграми, спеціалізовані статистичні графіки (гістограми, матричні, категоризовані графіки, діаграми розсіювання та ін), піктографіки (див. рис. 4.22 – типи графіків). Є два основних типи графіків, що відрізняються підмножиною даних, на яких вони будуються:

- статистичні графіки – графіки для візуалізації всіх значень змінних (стовпців) електронної таблиці. Вони включають в себе Stats Graphs – Статистичні графіки і Quick Stats Graphs – Швидкі статистичні графіки;

- користувацькі графіки. Графіки для візуалізації значень з попередньо виділеного блоку електронної таблиці. Ця група графіків включає в себе Custom Graphs – Користувацькі графіки і Block Stats Graphs – Блокові статистичні графіки.

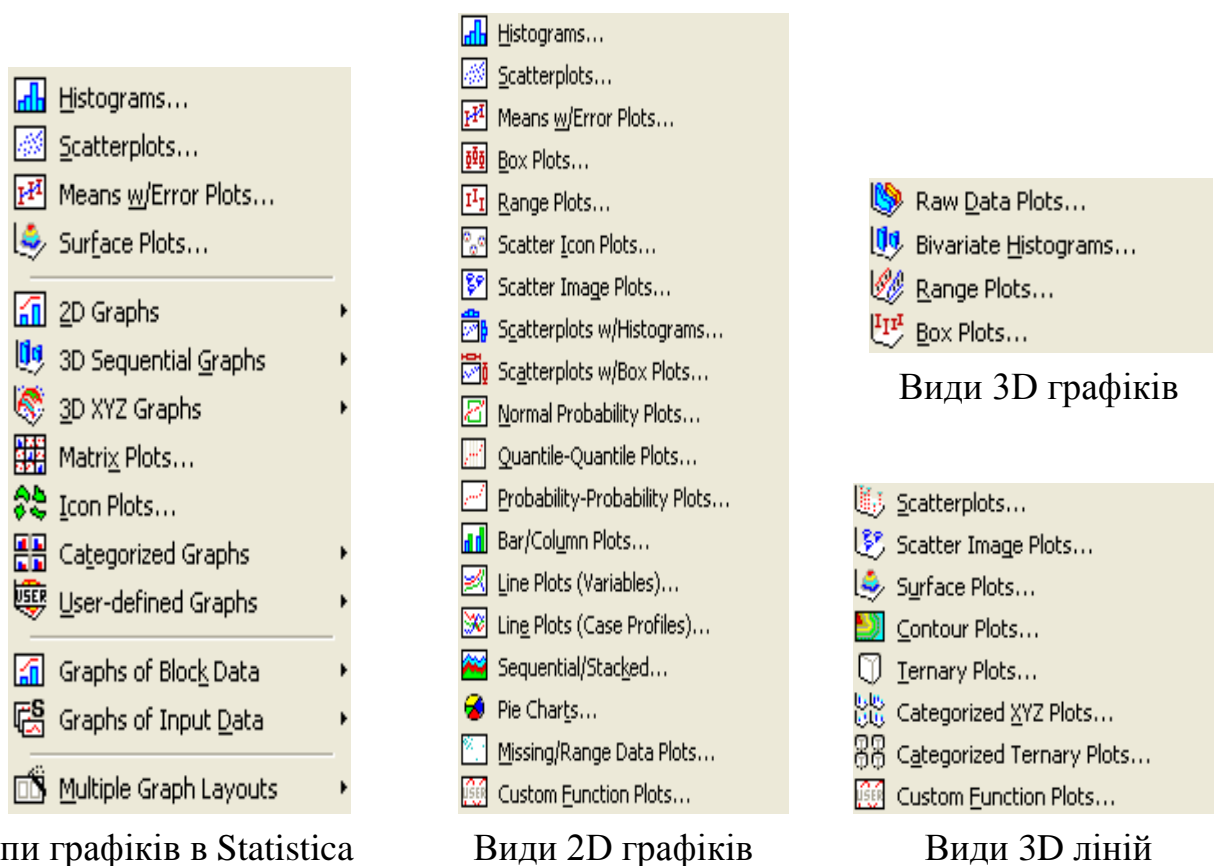


Рис. 4.22. Графічне представлення даних в Statistica

Statistica повністю підтримує технологію OLE (в режимах і сервера, і клієнта), за допомогою якої ви можете впроваджувати документи (наприклад, зображення, таблиці тощо) з різних додатків у графічний документ Statistica і навпаки.

У систему Statistica включені зручні інструментальні засоби для проведення інтерактивного графічного аналізу даних. За допомогою так званої Щітки, включеної в Statistica, ви можете виділити якісь точки на графіку (наприклад, викиди значень змінної) та провести подальший статистичний аналіз без значень, які відповідають цим точкам. Ці точки можуть бути позначені спеціальним маркером або міткою. Чисельні значення, відповідні виділеним точкам на графіці, можуть бути відзначені в спеціальній електронній таблиці, яка пов'язана з графіком, а також переглянуті в спеціальному Редакторі даних графіка.

**Приклад 4.2.** Розглянемо побудову графіків на прикладі візуалізації даних з табл. 4.2.



Таблиця 4.2

Наявний дохід на душу населення, грн. на рік

Область	Роки		Область	Роки	
	2000	2001		2000	2001
АР Крим	12 265,70	14 718,84	Львівська	13 657,00	16 388,40
Вінницька	12 380,20	14 856,24	Миколаївська	13 297,60	15 957,12
Волинська	11 130,80	13 356,96	Одеська	12 621,70	15 146,04
Дніпропетровська	16 646,60	19 975,92	Полтавська	14 747,30	17 696,76
Донецька	17 380,90	20 857,08	Рівненська	11 531,30	13 837,56
Житомирська	12 385,30	14 862,36	Сумська	13 655,60	16 386,72
Закарпатська	10 028,10	12 033,72	Тернопільська	10 733,20	12 879,84
Запорізька	16 174,00	19 408,80	Харківська	14 902,20	17 882,64
Івано–Франківська	12 015,40	14 418,48	Херсонська	11 606,30	13 927,56
Київська	15 085,80	18 102,96	Хмельницька	12 332,40	14 798,88
Кіровоградська	11 757,60	14 109,12	Черкаська	12 404,40	14 885,28
Луганська	14 368,40	17 242,08	Чернівецька	10 275,40	12 330,48
			Чернігівська	12 996,00	15 595,20

Для побудови графіка необхідно виконати наступні дії:

1. Виконати команду «2D Graphs /Line Plots (Variables)» з меню «Graphs» (рис. 4.23).

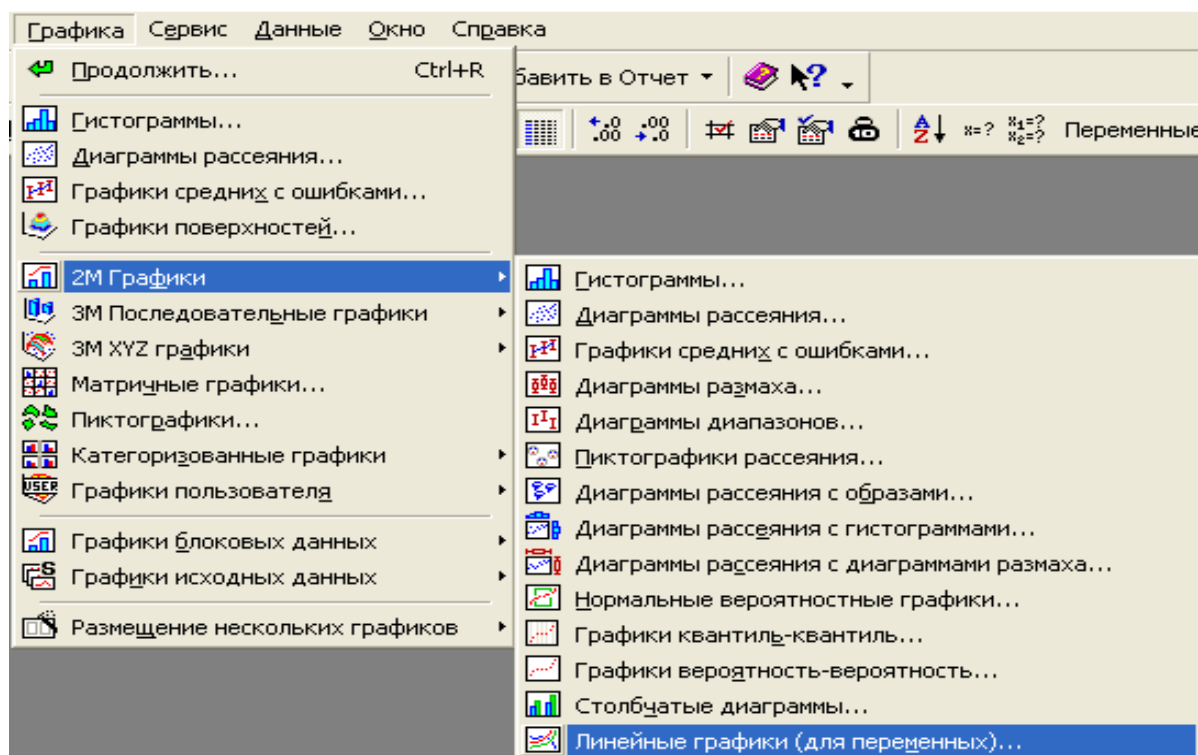


Рис. 4.23. Виклик діалогового вікна побудови графіка

2. Натиснути кнопку «Variables» на вкладці «Quick». Вибрати всі змінні. Вибрати тип графіка «Multiple» (рис. 4.24).

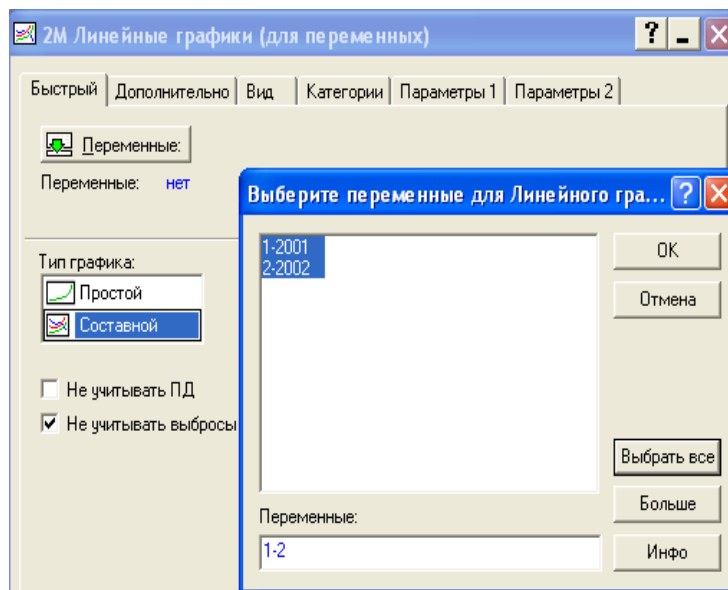


Рис. 4.24. Вибір змінних та типу графіка

3. На вкладці «Options 1» у списку «Case labels» вибрати значення «Case names». У рядок «Custom title» ввести заголовок «Наявний дохід на душу населення» натиснути «OK» (рис. 4.25).

4. Зберегти графік у файлі з розширенням .STW. Результат побудови графіка представлений на рис. 4.26.

Програма також дозволяє генерувати інший вид графіка. Для цього необхідно зробити наступні дії:

1. Натиснути кнопку «2D Line Plots – Variables» у лівому нижньому куті вікна програми.

2. У лівому нижньому куті вікна, що відкрилося, натиснути першу кнопку «Graphs Gallery».

3. У вікні «Graphs Gallery» вибрати вид графіка «2D Graphs / Line Plots (Cases Profiles)», натиснути «OK», знову вибрати всі змінні, «OK» (рис. 4.27). Зберегти графік під новим ім'ям.

Для того, щоб додати для кривої лінії «2000» стовпчикову діаграму, необхідно:

1. Кладнути правою клавішею миші по кривій «2000».
2. У контекстному меню вибрати «Bar Options».

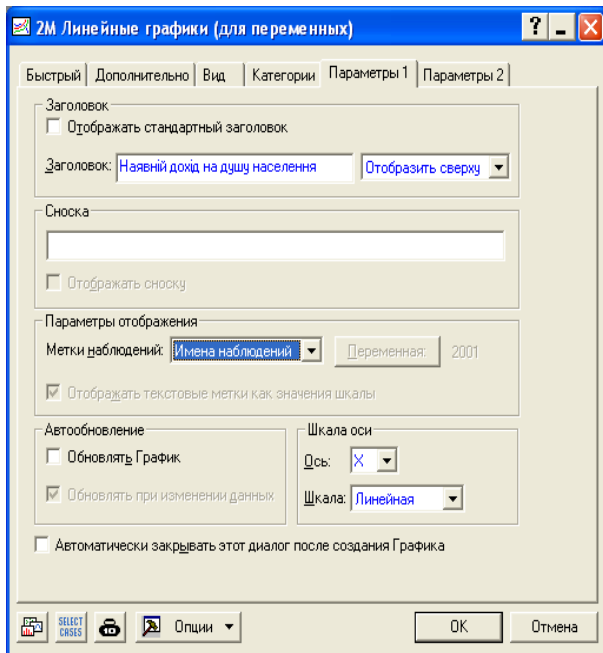


Рис. 4.25. Вибір підпису значень та заголовка діаграми

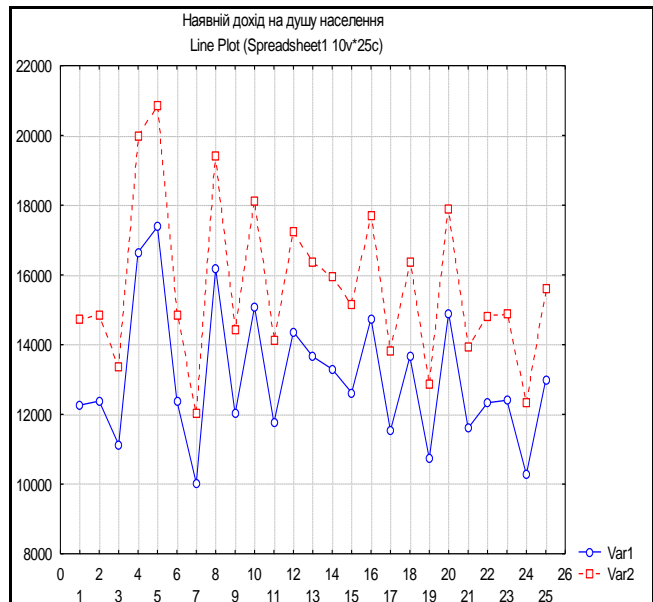


Рис. 4.26. Розподіл наявного доходу населення серед регіонів України за 2 роки

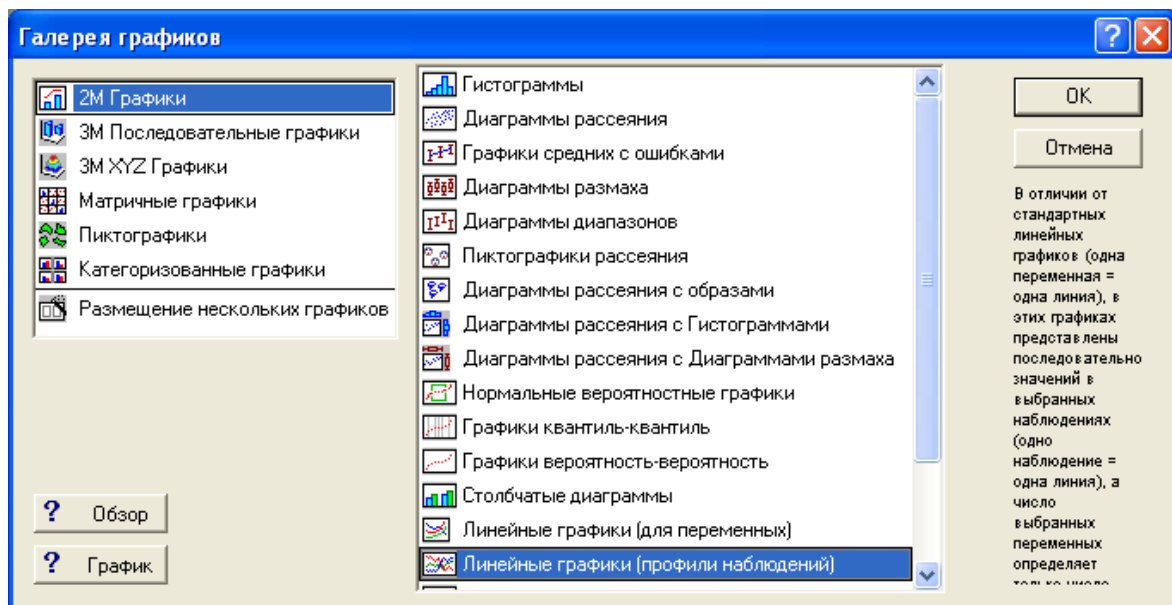


Рис. 4.27. Зміна типу графіку

3. У вікні «Bars» включити перемикач «Display bars» і встановити товщину стовпця діаграми «Width», «OK». Зберегти графік під новим ім'ям (рис. 4.28).

Для того, щоб навчитись змінювати різні формати графіків, необхідно виконати наступні дії:

1. Змінити тип і товщину кривій «2001», клацнувши по ній правою клавішею миші й вибравши команду «Properties».

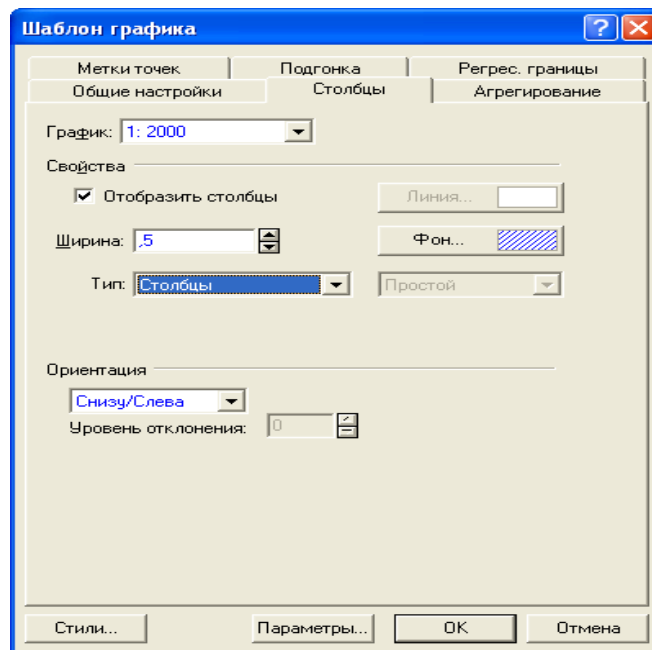


Рис. 4.28. Додавання стовпчикової діаграми

2. У вікні «General» клацнути по кнопці «Line».

3. Вибрати безперервну лінію товщиною 2 зеленого кольору. Натиснути «Close» (виберіть свій вид лінії й колір) (рис. 4.29).

4. У вікні «General» клацнути по кнопці «Markers».

5. Вибрати форму маркера коло, розмір 8 (виберіть свій вид маркера й колір). Натиснути «Close», «OK» (рис. 4.30).

Для того, щоб змінити параметри графіка, необхідно кликати контекстне меню «Graph Properties (All Options)» на фоні підкладки:

1. На вкладці «Graph Window» натисканням кнопки «Outside Background Color» змінити колір підкладки на власний (рис.4.31).

Також ця вкладка дозволяє змінити розмір графіка, масштаб, поля відступу та формат границі.

2. На вкладці «Graph Titles / Text» змінити заголовок на «Наявний дохід на душу населення». Шрифт «Arial» 14 пт, жирний (рис. 4.32).

У наведеній вкладці можливо змінити положення заголовка, зняти відображення його на графіку, вирівнювання та додати підзаголовок, виноску та ін. (вікно Статус).

3. На вкладці «Plot: Point Labels» встановити параметри, як на рис. 3.11. Натисканням кнопки «Font» змінити шрифт на «Times New Roman» 11 пт. При цьому установку зробити для двох змінних, тобто змінювати по черзі значення в списку, що розкривається, «Plot» (рис. 4.33).

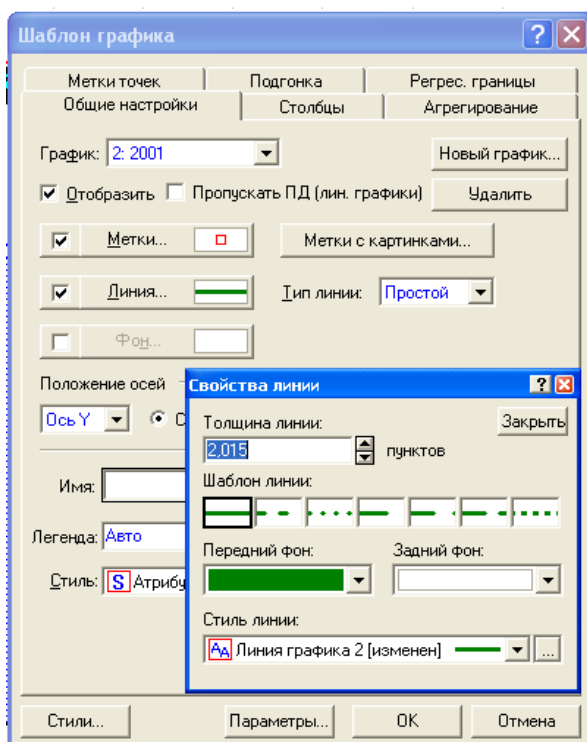


Рис. 4.29. Зміна кольору та товщини лінії графіка

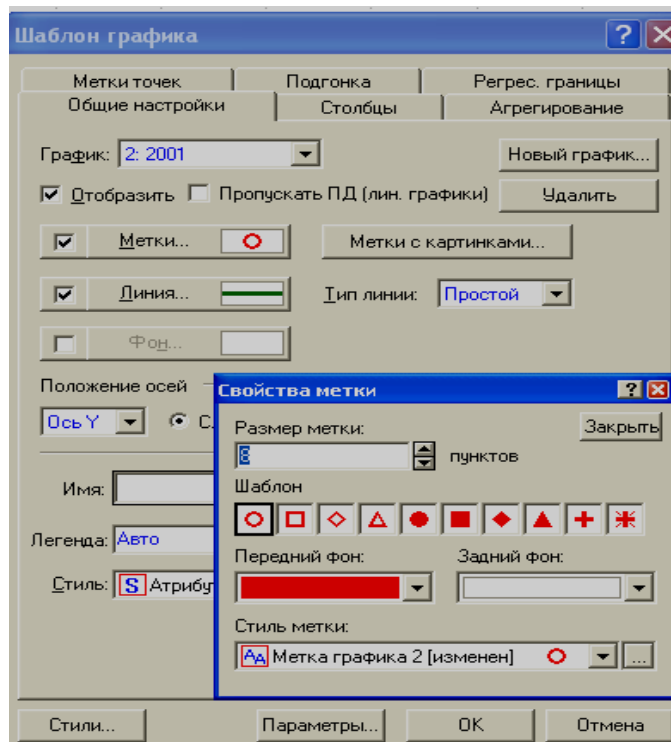


Рис. 4.30. Зміна властивостей маркера

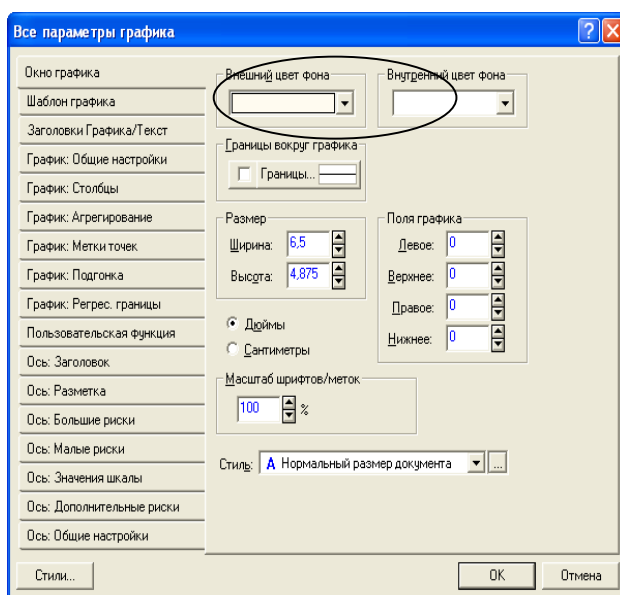


Рис. 4.31. Зміна кольору вікна графіка

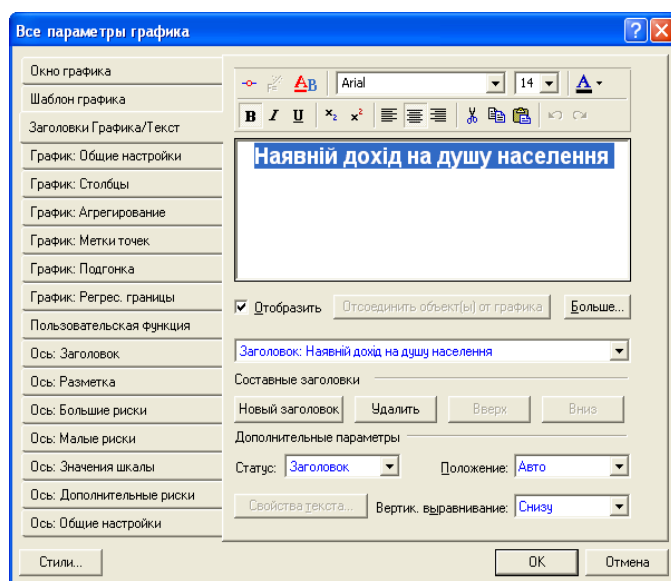


Рис.4.32. Вікно зміни назви та формату заголовка

4. На вкладці «Axis Titles» ввести назву осі X, колір шрифту зелений, змінити шрифт, його розмір. Аналогічно ввести назви осі Y (вікно Axis) (рис. 4.34).

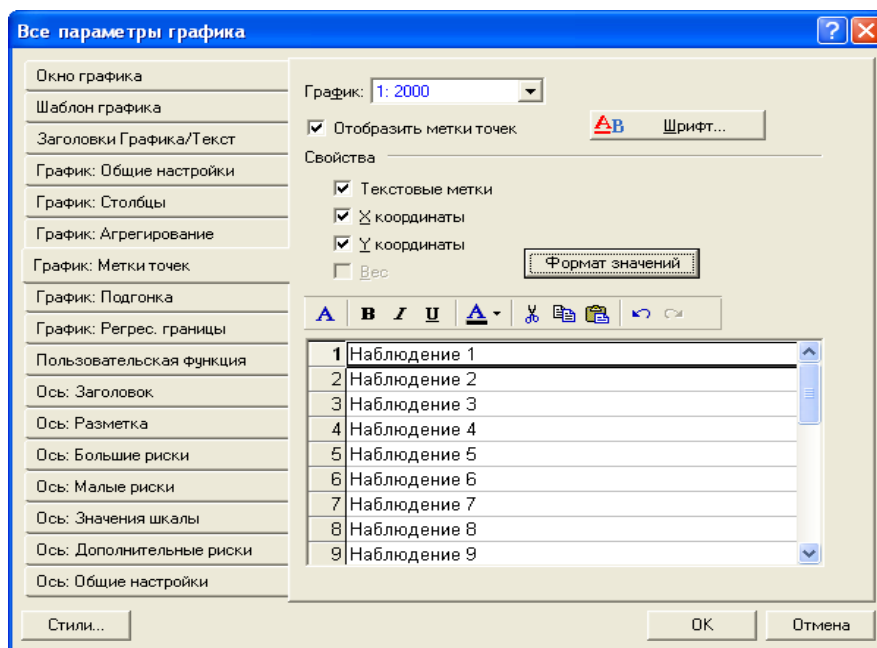


Рис. 4.33. Зміна властивостей графіка змінної

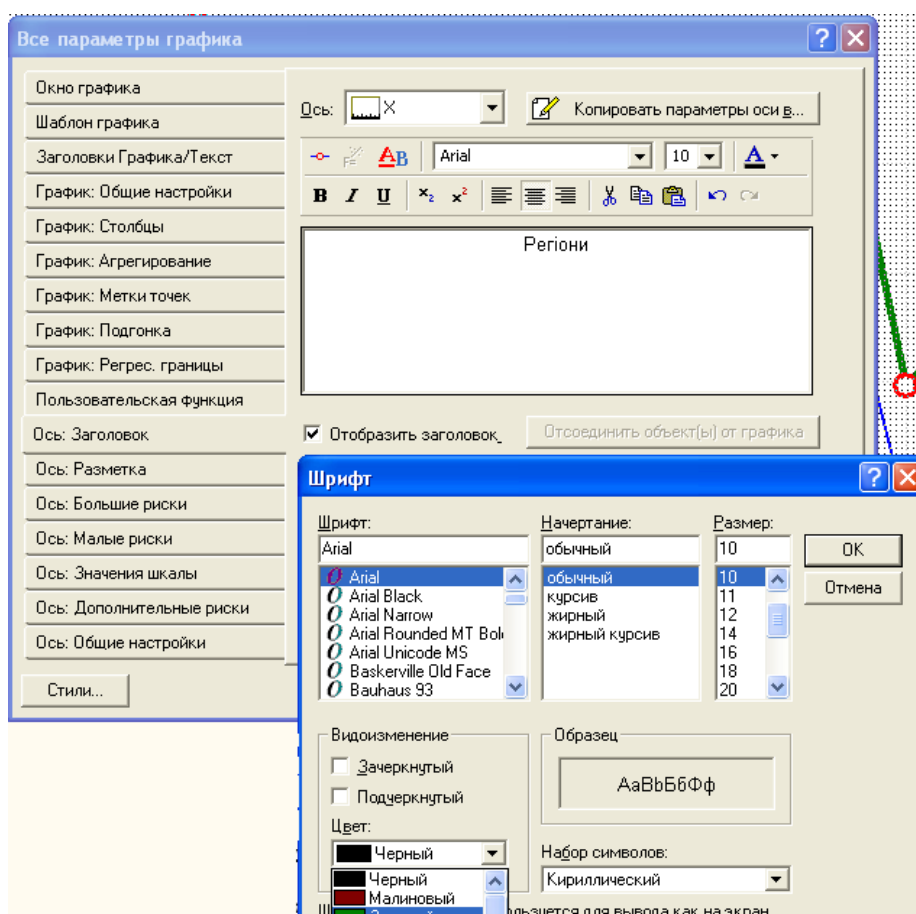


Рис. 4.34. Додавання назви осей

5. На вкладці «Axis Scaling» змінити діапазон осі значень наявного доходу (вісь Y), потім натисканням кнопки «Edit step» змінити ціну розподілу цієї осі (рис. 4.35).

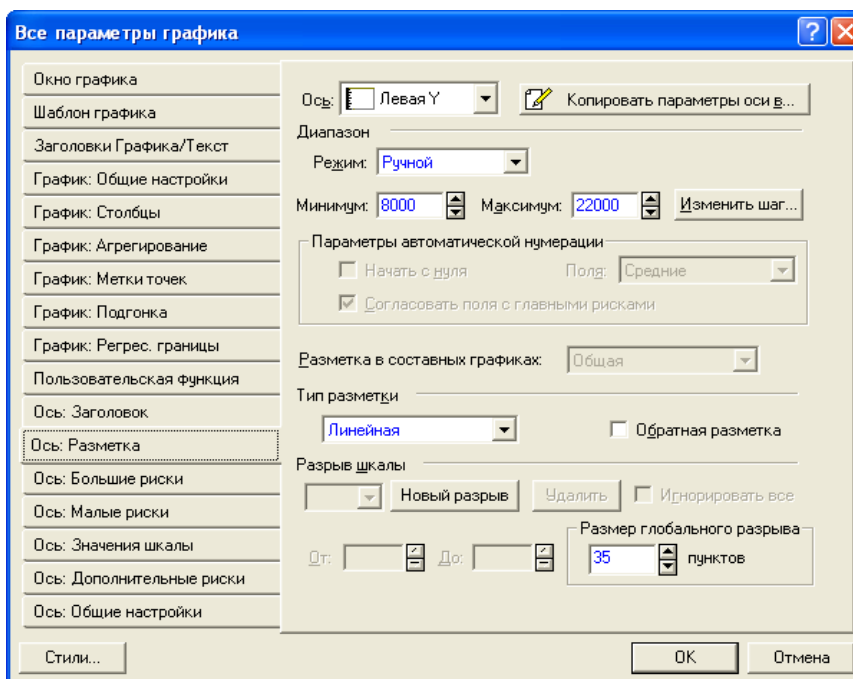


Рис. 4.35. Зміна діапазону та шагу осей,

6. На вкладці «Axis Major Units» натисканням кнопки «Gridlines» змінити тип лінії сітки осі Y і вісі X (змінити тип і колір) (рис. 4.36).

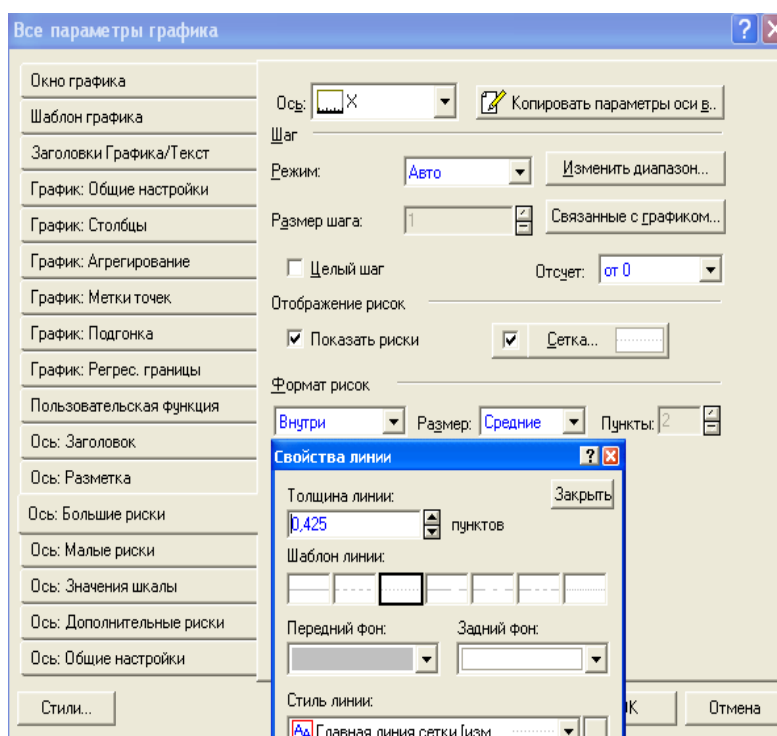


Рис. 4.36. Зміна формату основної сітки графіка

Вийти з вікна форматування графіка.

7. Контекстною командою «Rename», викликаній на фоні значка поточного графіка в деревоподібній ієрархічній структурі, розташованій ліворуч від графіка, змінити ім'я робочого аркуша на «Дослідження доходу населення» (рис. 4.37).

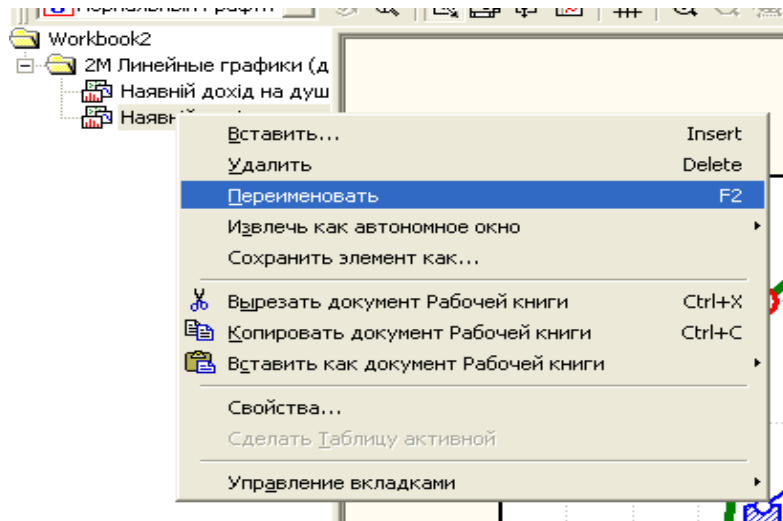


Рис. 4.37. Перейменування графіка

Зберегти графік під новим ім'ям. Закрити всі файли.

### Самостійне вивчення (22 години)

Приклади побудови графіків описової статистики – гістограми, кореляційне поле в MS Excel та Statistica.

### Контрольні питання

1. Перелік програмних засобів обробки даних. Характеристика та основні функції систем обробки даних.
2. Призначення системи Statistica.
3. Інтерфейс системи Statistica.
4. Ввод даних в систему Statistica, приклади.
5. Модулі системи Statistica.
6. Види статистичного аналізу в системі Statistica.
7. Навести приклад описової статистики.
8. Побудова 2D та 3D графіків та гістограм в системі Statistica.



### **Тема 3. Промислова статистика. Види статистичного аналізу в Statistica. Регресія в системі Statistica. (2 год)**

Статистика – наука про збір, вимір і аналіз масових кількісних даних. Статистичні дані завжди є наближеними, усередненими. Тому вони носять оцінний характер і для достовірності результатів потрібне велике число початкових даних.

Існує декілька видів статистичного аналізу даних: кореляційного, регресійного, дисперсійного, факторного, кластерного та ін. Розглянемо деякі з них.

#### **Кореляційний аналіз**

Іноді кореляцію і регресію розглядають як сукупний процес статистичного дослідження. Кореляційно–регресійний аналіз є одним зі значимих методів побудови математичних моделей в економіці і вважається одним з головних методів в маркетингу.

Кореляція в широкому значенні слова означає зв'язок між об'єктивно існуючими явищами.

Кореляційний аналіз – вид статистичного аналізу, який полягає в кількісній оцінці сили і напрямку зв'язку між двома(парна кореляція) Для кількісної оцінки сили зв'язку використовуються коефіцієнти парної кореляції  $r$  і множинній кореляції  $R$ .

Коефіцієнт кореляції (безрозмірна величина) – кількісний показник лінійного зв'язку між двома або більше наборами даних, значення якого лежить в інтервалі від  $-1$  до  $1$ . Якщо коефіцієнт дорівнює  $1$ , то зв'язок функціональний, якщо рівний  $0$ , то зв'язок відсутній.

Для якісної оцінки сили зв'язку використовуються спеціальні табличні співвідношення(наприклад, шкала Чеддока, табл. 4.3).

Напрямок зв'язку визначається знаками  $\pm 1$ : близькість до  $+1$  означає, що зростанню одного набору значень відповідає зростання іншого набору, близькість до  $-1$  означає зворотне.

Для наочності виміру усіх зв'язків у разі множинної кореляції доцільно використати кореляційну матрицю – матрицю з попарних коефіцієнтів кореляції.

Шкала Чеддока

Значення коефіцієнта кореляції	Характер зв'язку
$0 <  r  < 0,3$	Дуже слабка
$0,3 \leq  r  < 0,5$	Слабка
$0,5 \leq  r  < 0,7$	Помітна
$0,7 \leq  r  < 0,9$	Сильна
$0,9 \leq  r  < 1$	Дуже сильна

### Регресійний аналіз

Регресійний аналіз – вид статистичного аналізу, який полягає в уявленні залежності одних чинників від інших у вигляді деякої функції (рівняння регресії) за допомогою якої здійснюється прогнозування і пошук відповіді на питання «Що буде через якийсь час»? чи «Що буде, якщо».

У разі парної регресії (табл. 4.4) рівняння визначається по двох наборах даних, один з яких представляє значення залежної змінної  $y$ , а інший – незалежній змінній  $x$ . У разі множинної регресії (див. табл. 4.4) рівняння визначається по декількох наборах даних, один з яких представляє значення залежної змінної  $y$ , а інші незалежними змінними  $x_1, x_2, \dots, x_m$ .

Отримання рівняння регресії відбувається в два етапи: підбір виду функції і обчислення параметрів функції.

Таблиця 4.4

Видя функцій, які використовуються в регресійних моделях

Парна (проста) регресія	Множинна регресія
Лінійна регресія $y = ax + b$	$y = a_0 + a_1 x_1 + \dots + a_m x_m$
Квадратична (параболічна) $y = ax^2 + bx + c$	$y = a_0 + a_1 x_1^2 + \dots + a_m x_m^2$
Степенева $y = ax^b$	$y = a_0 x_1^{a_1} x_2^{a_2} \dots x_m^{a_m}$
Логарифмічна $y = a \ln x + b$ ,	Гіперболічна $y = a_0 + a_1 (1/x_1) + \dots + a_m (1/x_m)$
Експоненціальна $y = ae^{bx}$	
де $a, b, c$ – коефіцієнти парної регресії.	де $a_0, a_1, a_2, \dots, a_m$ – коефіцієнти множинної регресії, $n$ – об'єм сукупності, $m$ – кількість факторних ознак.

Для кількісної оцінки точності побудови рівняння регресії призначений коефіцієнт детерміації  $R^2$ , рівний квадрату коефіцієнта кореляції і що вказує, який відсоток зміни функції  $y$  пояснюється дією чинників  $x_k$ . Чим його значення ближче до 1, тим рівняння точніше описує досліджувану залежність.

Значиме рівняння (з  $R^2$  близьким до 1) використовується, як правило, для прогнозування явища, що вивчається.

### ***Виконання багатовимірної регресійного аналізу в пакеті Statistica***

Розглянемо приклад побудови регресійної моделі в пакеті Statistica 6.0 на прикладі 4.3.

**Приклад 4.3.** Побудувати регресійну модель для передбачення змін рівня захворюваності органів дихання ( $Y$ ) залежно від вмісту в повітрі двоокису вуглецю ( $X_1$ ) і ступеня запилення ( $X_2$ ). На рис. 4.38 наведені дані спостережень протягом 24 місяців. Передбачити рівень захворюваності при вмісті двоокису вуглецю, що дорівнює 0,7, та запилення – 1,5.

	A	B	C
1	Вміст CO2 X1	Міра запилення X2	Рівень захворюваності Y
2	1	1,3	1160
3	1	1,3	1155
4	1,1	1,4	1210
5	1,1	1,4	1210
6	1,1	1,5	1215
7	1,1	1,5	1210
8	1	1,4	1157
9	1	1,5	1159
10	1,2	1,6	1256
11	1,2	1,7	1260
12	0,6	1	1040
13	0,6	1	1035
14	0,7	1,1	1040
15	0,7	1,15	1038
16	0,75	1,2	1040
17	0,7	1,2	1039
18	0,7	1,3	1040
19	0,7	1,3	1039
20	0,8	1,4	1233
21	0,8	1,4	1235
22	0,78	1,5	1233
23	0,78	1,5	1245
24	0,8	1,5	1248

Рис. 4.38. Вихідні дані для регресійної моделі

	Вміст CO2 X1	Міра запилення X2	Рівень захворюваності Y
Вміст CO2 X1	1		
Міра запилення X2	0,69398113	1	
Рівень захворюва ності Y	0,65410545	0,874651073	1

Рис. 4.39. Розрахунок коефіцієнтів кореляції в MS Excel

Використовуємо дані прикладу 4.3 і скопіюємо з таблиці MS Excel в таблицю пакета Statistica. Вкажемо число рядків – 24, змінних –3, дамо назву змінним (рис. 4.41). Для виконання регресійного аналізу використовується модуль **Multiple Regressions** (Множинна регресія), який дозволяє передбачити залежну змінну за декількома незалежними змінними. У стартовому діалоговому вікні цього модуля (рис. 4.42) за допомогою кнопки **Variables** вказуються залежна (dependent) і незалежна (independent) змінні (див. рис. 4.41). У полі **Input file** вказується тип файлу з даними:

- **Raw Data** – дані у вигляді рядкової таблиці;
- **Correlation Matrix** – дані у вигляді кореляційної матриці.

У полі MD deletion вказується спосіб виключення з обробки відсутніх даних:

- **Casewise** – ігнорується весь рядок, в якому є хоча б одне пропущене значення;
- **Mean Substitution** – замість пропущених даних підставляються середні значення змінної;
- **Pairwise** – попарне виключення даних з пропусками із тих змінних, кореляція яких обчислюється.

	1 CO2	2 PUL	3 Y
1	1	1,3	1160
2	1	1,3	1155
3	1,1	1,4	1210
4	1,1	1,4	1210
5	1,1	1,5	1215
6	1,1	1,5	1210
7	1	1,4	1157
8	1	1,5	1159
9	1,2	1,6	1256
10	1,2	1,7	1260
11	0,6	1	1040
12	0,6	1	1035
13	0,7	1,1	1040
14	0,7	1,15	1038
15	0,75	1,2	1040
16	0,7	1,2	1039
17	0,7	1,3	1040
18	0,7	1,3	1039
19	0,8	1,4	1233
20	0,8	1,4	1235
21	0,78	1,5	1233
22	0,78	1,5	1245
23	0,8	1,5	1248

Рис. 4.40. Вікно файлу даних

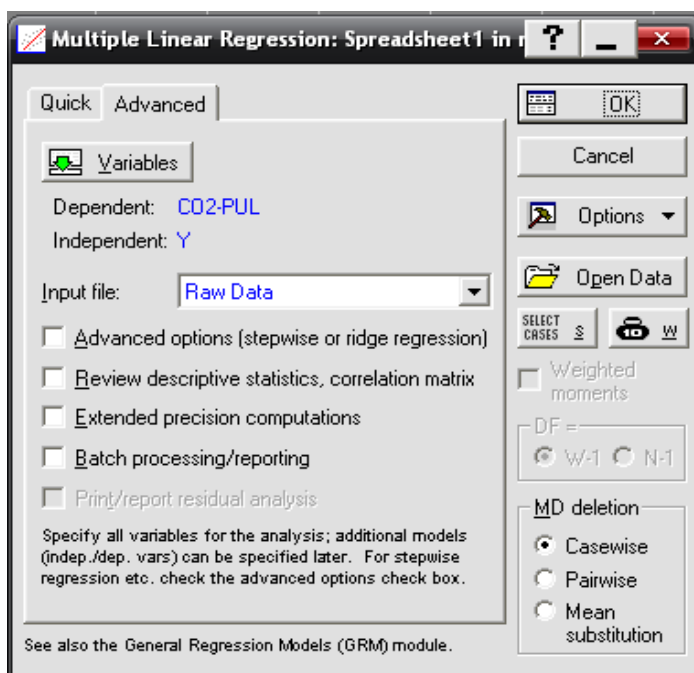


Рис. 4.41. Стартове меню модуля  
Multiple Regression

На першому етапі виведемо кореляційну матрицю (рис. 4.42, 4.43) і порівняємо її результати з відповідною таблицею MS Excel (рис. 3.39). Як бачимо, результати повністю ідентичні. Лінійна залежність  $Y$  від  $CO_2$  і  $PUL$  простежується.

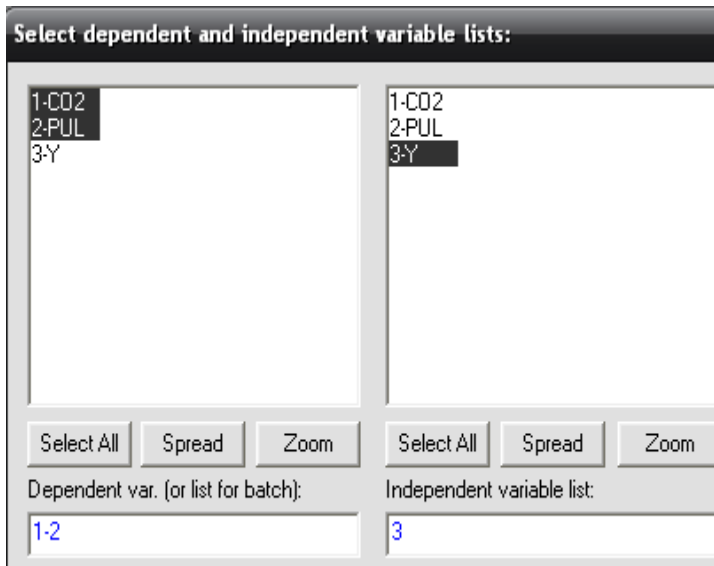


Рис. 4.42. Вибір залежної і незалежної змінних

Correlations (Spreadsheet1 in re			
Variable	Y	CO2	PUL
Y	1,000000	0,654105	0,874651
CO2	0,654105	1,000000	0,693981
PUL	0,874651	0,693981	1,000000

Рис. 4.43. Кореляційна матриця

Врахуємо, що за наявності однієї залежної змінної ( $Y$ ) і двох незалежних змінних ( $CO_2$  і  $PUL$ ) можна запропонувати різні моделі лінійної регресії. виберемо багатовимірну  $Y = b_0 + b_1 * CO_2 + b_2 * PUL$ .

Тепер більш докладно розглянемо послідовність дій створення моделі та аналіз отриманих результатів.

Після вибору всіх опцій стартового діалогового вікна регресійного аналізу і натиснення кнопки ОК з'являється вікно результатів регресійного аналізу Multiple Regressions Results (рис. 4.44). Детально проаналізуємо отримані результати регресійної моделі.

Про якість запропонованої моделі регресії будемо судити за величиною коефіцієнта детермінації.

У верхній частині вікна наведені найбільш важливі параметри отриманої регресійної моделі:

- **Multiple R** (0,877) – коефіцієнт множинної кореляції, який характеризує тісноту лінійного зв'язку між залежною і всіма незалежними змінними. Може приймати значення від 0 до 1;

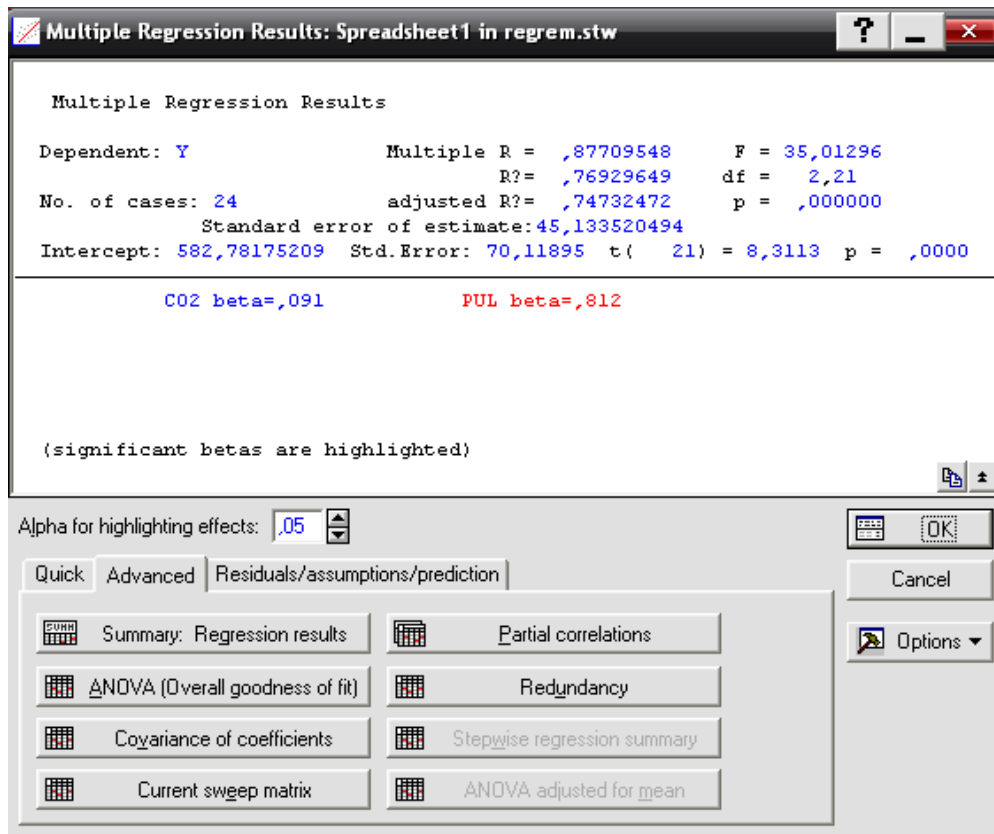

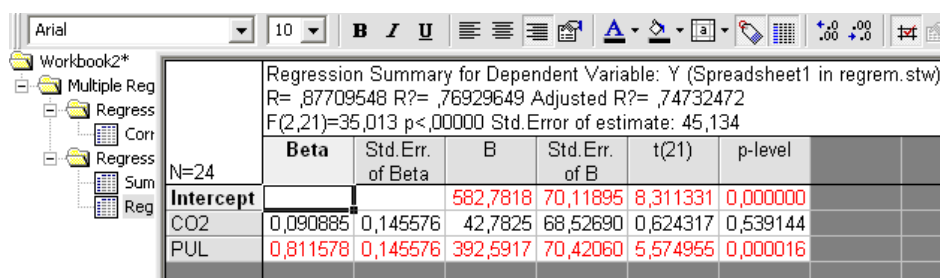


Рис. 4.44. Результати регресійного аналізу

- $R^2$  – коефіцієнт детермінації (0,769). Чисельно виражає частку варіації залежної змінної, пояснену за допомогою регресійного рівняння. Чим більше, тим більшу частку варіації пояснюють змінні, включені в модель;
- **adjusted R** (0,747) – скоригований коефіцієнт множинної кореляції. Включення нової змінної в регресійне рівняння збільшує не завжди, а тільки в тому випадку, коли окремий F–критерій при перевірці гіпотези про значущість «включаємий» змінної більше або дорівнює 1. В іншому випадку включення нової змінної зменшує значення і adjusted R;
- **F** (35,01) – F-критерій використовується для перевірки значущості регресії. У даному випадку в якості нульової гіпотези перевіряється гіпотеза: між залежною і незалежними змінними немає лінійної залежності;
- **df** (2,21) – числа ступенів свободи для F-критерію;
- **p** – ймовірність нульової гіпотези для F-критерію;
- **Standard error of estimate** – стандартна помилка оцінки (рівняння). Ця оцінка є мірою розсіювання спостережених значень щодо регресійної прямої;
- **Intercept** (582,78) – оцінка вільного члена рівняння;
- **Std.Error** – стандартна помилка оцінки вільного члена рівняння;

- $t(8,31)$  – t-критерій для оцінки вільного члена рівняння;
- $p$  – ймовірність нульової гіпотези для вільного члена рівняння;
- **Beta** –  $\beta$  – коефіцієнти рівняння. Це стандартизовані регресійні коефіцієнти, розраховані за стандартизованим значенням змінних. За їх величиною можна оцінити значущість залежних змінних. Коефіцієнт показує, на скільки одиниць стандартного відхилення зміниться залежна змінна при зміні на одне стандартне відхилення незалежної змінної, за умови сталості інших незалежних змінних. Вільний член в такому рівнянні дорівнює 0.

Натискання кнопки  – у вікні результатів (див. рис. 4.44) дозволяє отримати основні результати регресійної моделі (рис. 4.45), частина з яких вже була описана: **B** – коефіцієнти рівняння регресії; **St. Err. of B** – стандартні помилки коефіцієнтів рівняння регресії; **t (11)** – t-критерій для коефіцієнтів рівняння регресії; **p-level** – ймовірність нульової гіпотези для коефіцієнтів рівняння регресії.



	Beta	Std. Err. of Beta	B	Std. Err. of B	t(21)	p-level
Intercept			582,7818	70,11895	8,311331	0,000000
CO2	0,090885	0,145576	42,7825	68,52690	0,624317	0,539144
PUL	0,811578	0,145576	392,5917	70,42060	5,574955	0,000016

Рис. 4.45. Параметри рівняння регресії

У результаті проведеного аналізу було отримано наступне рівняння:

$$Y = 582,78 + 42,78 \cdot CO_2 + 392,59 \cdot PUL$$

Це рівняння пояснює 76,9% ( $R^2 = 0,769$ ) варіації залежної змінної. Отримані результати свідчать про те, що коефіцієнт  $b_1$  при змінній  $CO_2$  незначимо відрізняється від нуля. Перевірка якості рівняння регресії здійснювалася за допомогою  $F$  статистики  $F=35,01$ . За статистичними таблицями Фішера-Снедекора з даними ступенями свободи  $df=2,21$  гіпотезу  $H_0$  (лінійна залежність відсутня) можна прийняти з ймовірністю  $p=0,0000$ ; при рівні значущості  $\alpha=0,05$  приймаємо альтернативну гіпотезу – лінійна залежність значима.

Одночасно перевірялася статистична значущість коефіцієнтів множинної регресії (критерій Стюдента). Видно (див. рис. 4.45), що  $t$  для коефіцієнтів  $b_0$  і

$b_2$  значимо відрізняються від нуля, а для коефіцієнта  $b_1$  незначимо відрізняється від нуля.

Для розрахунку за отриманим регресійним рівнянням значень залежної змінної за значеннями незалежних змінних скористаємося кнопкою

 Predict dependent variable (розділ Residuals/assumptions/prediction) (рис. 4.46).

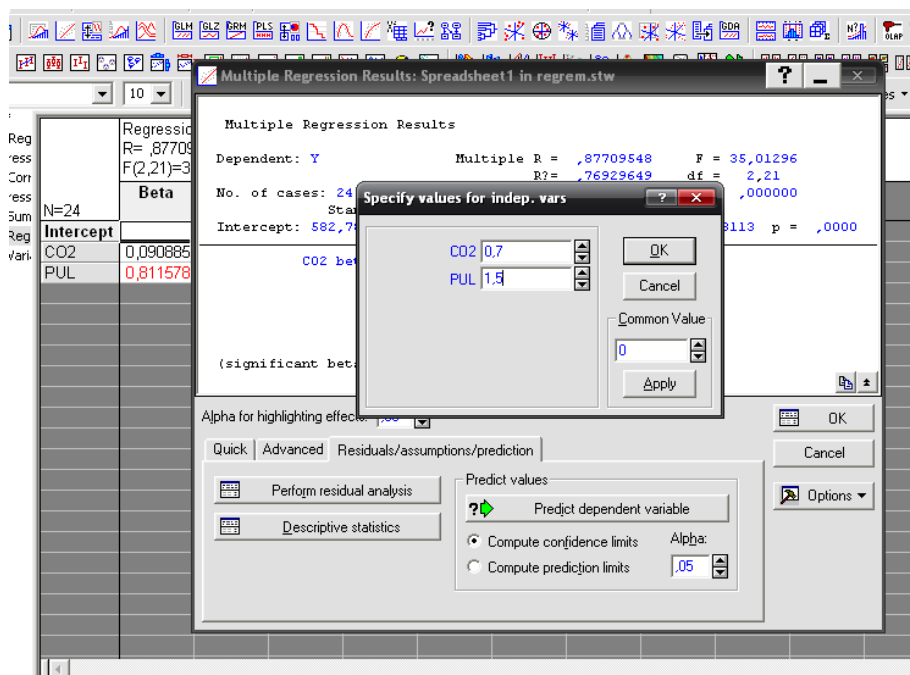


Рис. 4.46. Вікно завдання значень незалежних змінних

Задаємо значення двоокису вуглецю ( $\text{CO}_2=0,7$ ) і рівня запиленості ( $\text{PUL}=1,5$ ). Врахуємо, що в пакеті Statistica наводиться як точкова, так і інтервальна оцінка (рис. 4.47).

Predicting Values for (Spreadsheet1 in re variable: Y			
Variable	B-Weight	Value	B-Weight * Value
CO2	42,7825	0,700000	29,948
PUL	392,5917	1,500000	588,888
Intercept			582,782
Predicted			1201,617
-95,0%CL			1156,083
+95,0%CL			1247,151

Рис. 4.47. Передбачені точкові та інтервальні значення

Про отримані результати можна сказати наступне:  $Y=1201,617$  – це точкова оцінка. 95% – довірчий інтервал дорівнює (1156.083; 1247,151).



Порівнюючи отримані результати з результатами, отриманими в MS Excel (рис. 4.48), переконуємося в їх ідентичності та повноті аналізу, зробленого в пакеті Statistica.

Вывод итогов									
Регрессионная статистика									
Множественный R	0,877095483								
R-квадрат	0,769296486								
Нормированный R-квадрат	0,747324722								
Стандартная ошибка	45,13352049								
Наблюдения	24								
Дисперсионный анализ									
	df	SS	MS	F	Значимость F				
Регрессия	2	142645,2302	71322,61511	35,01296079	2,05149E-07				
Остаток	21	42777,72812	2037,034672						
Итого	23	185422,9583							
Коэффициенты		Стандартная ошибка	t-статистика	P-значение	Нижние 95%	Верхние 95%	Нижние 95%	Верхние 95%	Средние 95,0%
У-пересечение	582,7817521	70,11894764	8,311330556	4,44139E-08	436,9614183	728,6021	436,9614	728,6021	
Содержание CO2 X1	42,78250622	68,52689749	0,624316988	0,53914349	-99,72697801	185,292	-99,727	185,292	
Степень запыленности X	392,5916792	70,4206019	5,574954894	1,56316E-05	246,1440211	539,0393	246,144	539,0393	
	t-критическое	2,073873058	F-критическое	3,466800112					

Рис. 4.48. Результати регресійного аналізу в MS Excel (Пакет аналізу–Регресія)

### **Контрольні питання**

1. Навести приклади виконання кореляційного аналізу в програмі Statistica.
2. Модулі системи Statistica для проведення регресійного аналізу.
3. Навести приклади виконання регресійного аналізу в програмі Statistica.
4. Виконання багатовимірною регресійного аналізу в пакеті Statistica.
5. Оцінка коефіцієнтів регресії.
6. Оцінка адекватності моделі регресії.

### **Тема 4. Контрольні карти якості.**

Одним з основних інструментів у великому арсеналі статистичних методів контролю якості є контрольні карти (КК). Прийнято вважати, що ідея контрольної карти належить відомому американському статистику Уолтеру Л. Шухарта.

Контрольна карта складається з центральної лінії, двох контрольних меж (над і під центральною лінією) і значень характеристики (показника якості), нанесених на карту для представлення стану процесу.

У певні періоди часу відбирають (всі; вибірково; періодично з безперервного потоку і т.п.)  $n$  виготовлених виробів і вимірюють контрольований параметр. Результати вимірювань наносять на контрольну карту, і залежно від цього значення приймають рішення про коригування процесу або про продовження процесу без коректувань. Сигналом про можливу розладнання технологічного процесу можуть служити:

- вихід точки за контрольні межі;
- сильне розсіювання точок на контрольній карті щодо середньої лінії, що свідчить про зниження точності технологічного процесу.

При наявності сигналу про порушення виробничого процесу повинна бути виявлена і усунена причина порушення. Таким чином, контрольні картки використовуються для виявлення певної причини, але не випадкової.

### ***Типи контрольних карток***

Існують два основних типи контрольних карт: для якісних ознак (придатний – непридатний) і для кількісних ознак. Для контролю будуються такі контрольні картки:

**X–карта.** На цю контрольну карту наносяться значення вибірових середніх для того, щоб контролювати відхилення від середнього значення безперервної змінної (наприклад, діаметрів труб, міцності матеріалу тощо).

**R–карта.** Для контролю за ступенем мінливості безперервної величини у контрольній карті цього типу будуються значення розмахів вибірок.

**S–карта.** Для контролю за ступенем мінливості безперервної змінної в контрольній карті даного типу розглядаються значення вибірових стандартних відхилень.

**S<sup>2</sup>–карта.** У контрольній карті даного типу для контролю мінливості будується графік вибіркової дисперсії.

Для контролю якості продукції за альтернативною ознакою зазвичай застосовуються наступні типи контрольних карток:

**C–карта.** У таких контрольних картках будується графік числа дефектів (у партії, в день, на один стан, в розрахунку на 100 м труби тощо). При використанні карти цього типу робиться припущення, що дефекти

контрольованої характеристики продукції зустрічаються порівняно рідко, при цьому контрольні межі для даного типу карток розраховуються на основі властивостей розподілу Пуассона (розподілу рідкісних подій).

**U–карта.** У карті даного типу будується графік відносної частоти дефектів, тобто відношення числа виявлених дефектів до  $n$  – числу перевірених одиниць продукції (тут  $n$  позначає, наприклад, число см довжини труби, обсяг партії виробів).

**Np–карта.** У контрольних картах цього типу будується графік для числа дефектів (у партії, в день, на стан), як і у випадку C–карти. Однак, контрольні межі цієї карти розраховуються на основі біноміального розподілу, а не розподілу рідкісних подій Пуассона. Тому цей тип карт повинен використовуватися в тому випадку, коли виявлення дефекту не є рідкісною подією (наприклад, коли виявлення дефекту відбувається більш ніж у 5% перевірених одиниць продукції). Цією картою можна скористатися, наприклад, при контролі числа одиниць продукції, які мають невеликий брак.

**P–карта.** У картах даного типу будується графік відсотка виявлених дефектних виробів (у розрахунку на партію, в день, на верстат і т.п.). Графік будується так само, як і у випадку U–карти. Однак контрольні межі для даної карти знаходяться на основі біноміального розподілу (для часток), а не розподілу рідкісних подій. Тому P–карта найбільш часто використовується, коли поява дефекту не можна вважати рідкісною подією (якщо, наприклад, очікується, що дефекти будуть присутні в більш ніж 5% загального числа вироблених одиниць продукції).

КК складається зазвичай з трьох ліній. Центральна лінія (ЦЛ) є необхідним середнім значенням характеристики контрольованого параметра якості. Так, у разі ('x – R) –карти це будуть номінальні значення 'x і R, нанесені на відповідні карти.

Дві інші лінії, одна з яких знаходиться над центральною – верхньою контрольною межею (ВКП), а інша під нею – нижня контрольна межа (НКП), є максимально допустимі межі зміни значень контрольованої характеристики (показника якості).

При організації будь-якого виробничого процесу виникає завдання встановлення меж характеристик виробу, в рамках яких вироблена продукція задовольняє своєму призначенню. Існує два «ворога» якості продукції: відхилення від планових специфікацій і занадто великий розкид реальних

характеристик виробів (щодо планових специфікацій). На ранніх стадіях налагодження виробничого процесу для оптимізації цих двох показників якості часто використовуються методи планування експерименту. Методи, які містить модуль «Контроль якості», призначені для побудови процедур контролю якості продукції в процесі її виробництва, тобто поточного контролю якості.

**Приклад 4.4.** Припустимо, вам необхідно контролювати концентрацію деякої речовини на виході хімічного процесу. Ви спостерігаєте процес у реальному часі протягом 20 годин і знімаєте з датчиків потрібну характеристику щогодини. Дані представлені в таблиці (табл. 4.5).

Таблиця 4.5

Вхідні дані

1	102	11	101
2	95	12	99
3	98	13	101
4	98	14	98
5	102	15	97
6	99	16	97
7	99	17	100
8	98	18	101
9	100	19	97
10	98	20	101

Особливістю процесів, що протікають в реальному часі, є те, що тут не природно групувати вимірювання. Тому потрібно використовувати карти для індивідуальних спостережень.

Вважається, що процес виходить з-під контролю, якщо значення концентрації перевищують допустимий рівень і виходять за верхній контрольний кордон.

Крок 1. Введіть вихідні дані в таблицю даних в програмі (рис. 4.49).

Крок 2. Оберіть з меню «Аналіз» пункт «Промислова статистика й Шість Сігма». Далі оберіть «Карти контролю якості». Повинно з'явитись наступне вікно (рис. 4.50).

Крок 3. На стартовій панелі оберіть опцію «Окремі спостереження і ковзаючи розмахи», натисніть кнопку ОК (див. рис. 4.50). У вікні, що з'явилося, оберіть Var2 в якості змінної з вимірами (рис. 4.51).

	1 Var1	2 Var2
1	1	102
2	2	95
3	3	98
4	4	98
5	5	102
6	6	99
7	7	99
8	8	98
9	9	100
10	10	98
11	11	101
12	12	99
13	13	101
14	14	98
15	15	97
16	16	97
17	17	100
18	18	101
19	19	97
20	20	101

Рис. 4.49. Вихідні дані

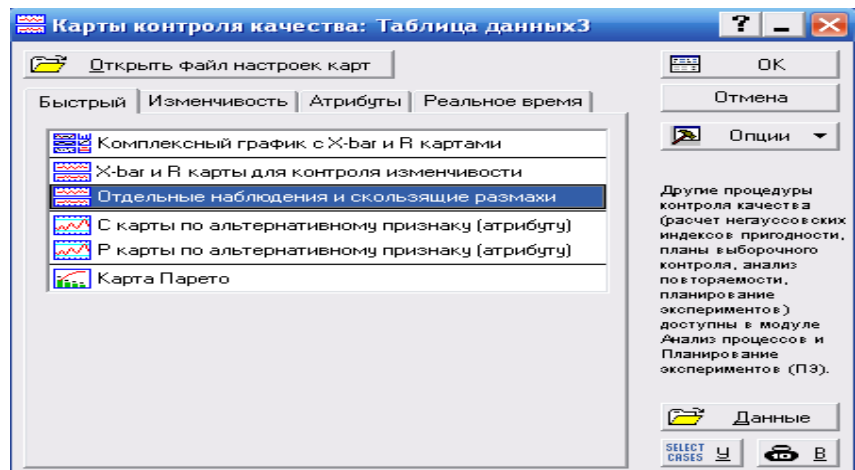


Рис. 4.50. Стартовая панель модулю контролю якості

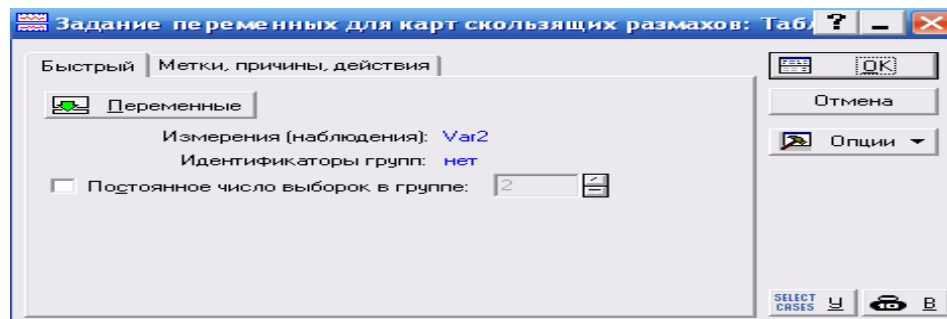


Рис. 4.51 Вікно завдання змінних

Після натискання кнопки «ОК» з'явиться наступне вікно (рис. 4.52).

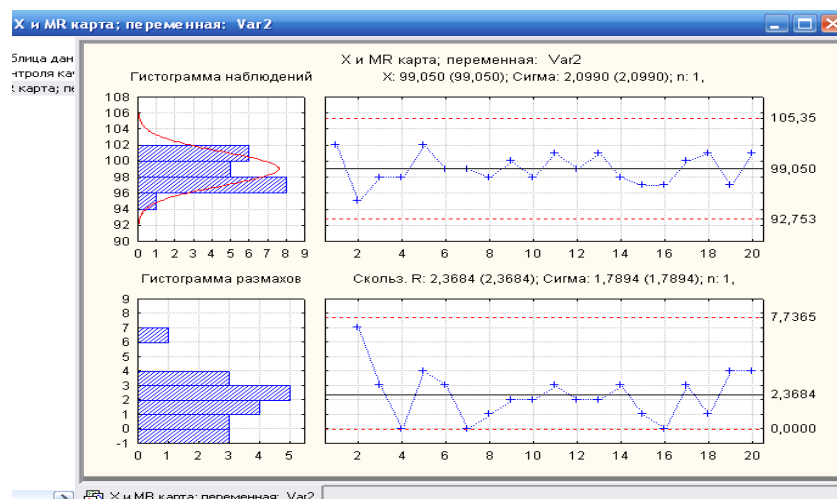


Рис. 4.52. X– и MR карты для змінної Var2

Крок 4. Аналіз карти. На X-карті всі точки потрапляють в область всередині контрольних меж.

На контрольній карті ковзних розмахів (MR– карті) видно, що всі крапки знаходяться нижче контрольного кордону (див. рис. 4.52). Це хороший результат.

Можна сказати, що концентрація речовини в хімічному процесі підкоряється вимогам статистичного контролю.

Слід мати на увазі, що карти для окремих спостережень не здатні відображати малі зміни середнього рівня концентрації, які, проте, можуть відігравати суттєву роль у хімічному процесі. Тому для аналізу цих даних потрібно далі використовувати контрольні карти накопичених сум (рис. 4.53).

Виявлення малих змін середніх значень

Крок 1. Запустіть модуль Карти контролю якості:

У вікні, що з'явилось, оберіть вкладку «Мінливість» і у списку оберіть «CuSum карту для окремих спостережень» (рис. 4.54). Натисніть кнопку «ОК».

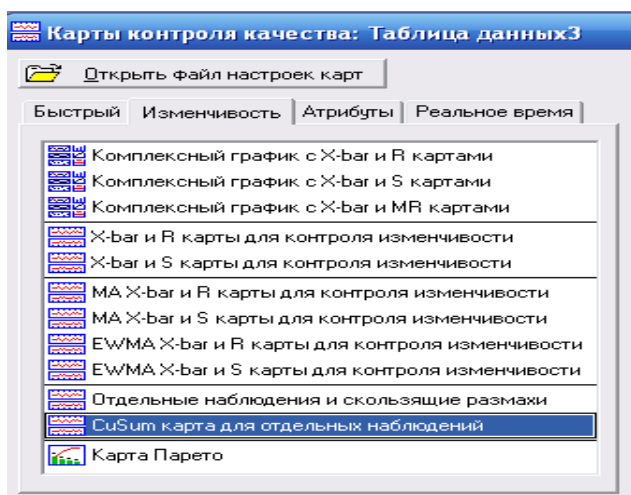


Рис. 4.53. Стартова панель вказує доступні контрольні карти

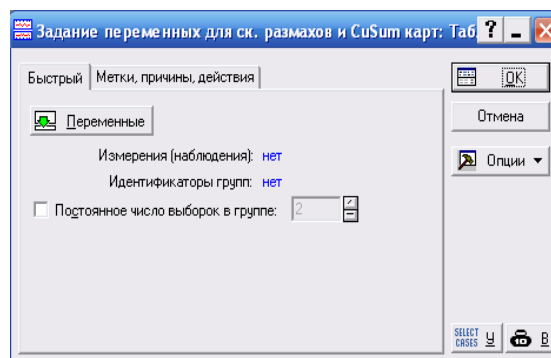


Рис. 4.54. Вікно завдання змінних

У вікні завдання змінних натисніть на кнопку «Змінні» і оберіть змінну Var2 в якості змінної з вимірами. Натисніть кнопку «ОК» (2 рази). В результаті виконання попередніх дій повинно з'явитись вікно представлене на рис. 4.55.

З наведеного графіка випливає, що всі точки даних потрапляють всередину контрольного інтервалу.

На карті зображена також так звана V–маска, що має наступний сенс.

Якщо в процесі є значне зміщення середнього значення, то точки виходять за межі V–маски.

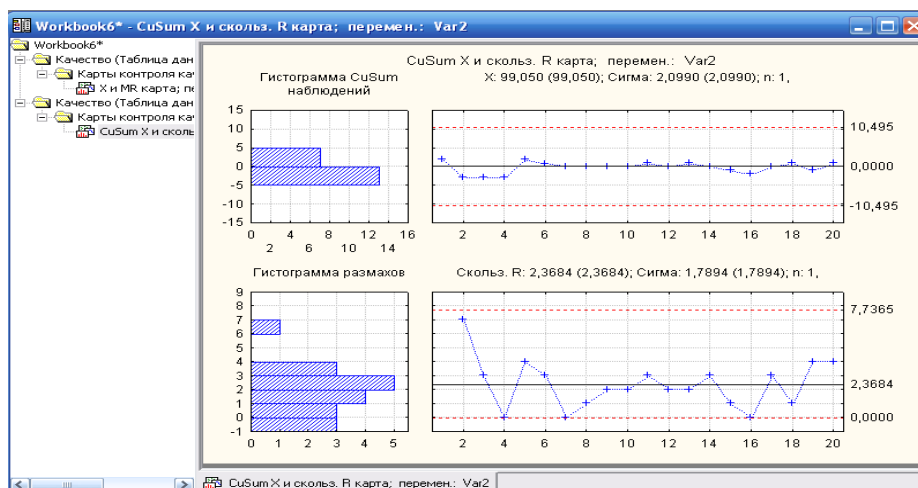


Рис.4.55. Карта накопиченої суми для змінної Var2

У нашому випадку точки не виходять за межі маски, тому можна зробити остаточний висновок про те, що досліджений хімічний процес задовольняє вимогам статистичного контролю.

Крок 2. Використовуючи кнопки діалогового вікна, яке з'явилося на екрані, можна отримати додаткові результати (рис. 4.56).

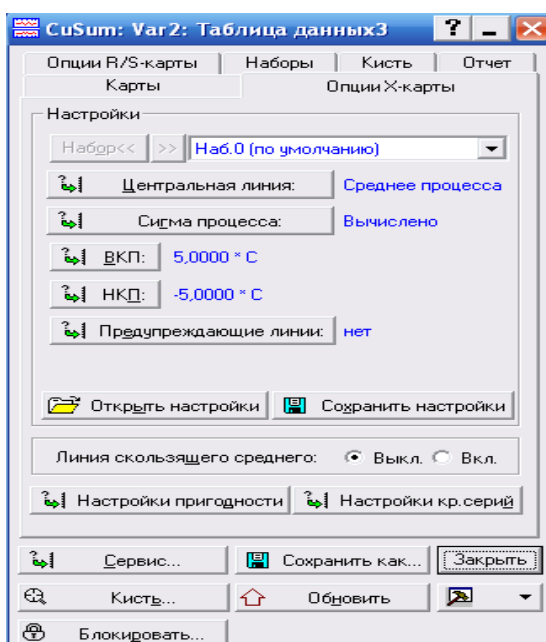


Рис. 4.56. Карта накопиченої суми для змінної Var2

### Контрольні питання

1. Назвіть найпоширеніші типи контрольних карт.
2. Назвіть переваги контрольних карт за альтернативною ознакою.
3. Назвіть переваги контрольних карт для безперервних змінних.

4. Алгоритм побудови контрольних карт.
5. Обчислення вибірових статистик, центральної лінії, контрольних меж.
7. Оцінка керованості процесу.
8. Вдосконалення системи якості процесу.

## **5. ЗАВДАННЯ ДЛЯ ІНДИВІДУАЛЬНИХ РОБІТ**

Вимоги до індивідуальних завдань (тематику завдань див. нижче).

1. В роботі 1 повинна бути вирішена задача, побудовані графіки та гістограми та наведені теоретичні визначення наведених функцій вибірових характеристик з тематики індивідуального завдання.
2. Номер варіанта завдання вибирається згідно з номером залікової книжки: – остання цифра номера залікової книжки.
3. Індивідуальне завдання повинне мати титульний аркуш, зміст або план, основний зміст, перелік посилань.

### **Варіанти завдання до тем 1, 2**

0. Служба контролю якості підприємства справила заміри довжин (у сантиметрах) випадково відібраних заготовок: 39, 41, 40, 43, 41, 44, 42, 41, 41, 43, 42, 39, 40, 42, 43, 42, 41, 39, 42, 42, 41, 42, 40, 41, 43, 41, 39, 40, 41, 40.

Побудуйте варіаційний ряд, полігон частот, діаграму відносних і накопичених частот в системі Statistica. Визначте вибірові характеристики.

1. Побудуйте варіаційний ряд, полігон частот і гістограму в системі Statistica за наступними даними про масу заготовок з 40 штук (кг): 181, 169, 178, 178, 171, 179, 172, 181, 179, 168, 174, 167, 169, 171, 179, 181, 181, 183, 172, 176, 165, 178, 176, 176, 190, 183, 169, 192, 185, 173, 179, 187, 181, 168, 179, 184, 177, 184, 185, 186.

Визначте вибірові характеристики цього розподілу.

2. Отримано такі дані про успішність у групі з 20 студентів: 4, 4, 5, 3, 4, 5, 4, 5, 3, 5, 3, 3, 5, 4, 5, 4, 3, 5, 3, 5.

Побудуйте варіаційний ряд, полігон частот, діаграму відносних і накопичених частот в системі Statistica. Визначте вибірові характеристики цього розподілу.





9. У робочій зоні проводилися виміри концентрації шкідливої речовини (у  $\text{мг/м}^3$ ). Одержаний наступний ряд значень: 14, 16, 15, 14, 11, 20, 16, 14, 18, 14, 15, 17, 23, 16, 11, 12, 21, 17, 16, 15, 10, 13, 14, 19, 18, 16, 17, 20, 14, 18, 14, 11, 20, 16, 14, 18, 17.

Побудуйте варіаційний ряд і гістограму в системі Statistica. Визначте вибіркові характеристики.

### Завдання до теми 3

Вимоги до індивідуальних завдань (тематику завдань див. нижче).

1. В роботі повинні бути надані відповіді на питання першого та другого блоку з тематики індивідуального завдання.

2. Номер варіанта завдання першого блоку вибирається згідно з номером залікової книжки: – остання цифра номера залікової книжки, якщо вона менш 6 і остання цифра номера залікової книжки мінус 5, якщо ця цифра більш 5, завдання другого блоку – єдине, незалежно від номера залікової книжки.

3. Індивідуальне завдання повинне мати титульний аркуш, зміст або план, основний зміст, перелік посилань.

#### Перший блок:

1. За даними дослідження (табл. 5.1 (згідно варіанту)) визначити коефіцієнт кореляції між твердістю  $y$  і ступенем деформації  $x$ . За наявності взаємозв'язку визначити рівняння регресії і його адекватність експериментальним результатам.

Таблиця 5.1

№	$x$	$y$	№	$x$	$y$	№	$x$	$y$
Варіант 1								
1	0,20	64	8	0,26	67	15	0,17	62
2	0,19	65	9	0,28	70	16	0,30	70
3	0,28	69	10	0,25	68	17	0,19	64
4	0,26	69	11	0,25	67	18	0,25	68
5	0,23	66	12	0,22	66	19	0,29	69
6	0,21	65	13	0,18	63	20	0,27	68
7	0,24	67	14	0,26	68	$\Sigma$		

№	x	y	№	x	y	№	x	y
Варіант 2								
1	0,21	64	8	0,25	67	15	0,17	62
2	0,19	65	9	0,29	70	16	0,30	70
3	0,27	69	10	0,24	68	17	0,19	64
4	0,28	69	11	0,23	67	18	0,25	68
5	0,22	66	12	0,21	66	19	0,29	69
6	0,20	65	13	0,19	63	20	0,27	68
7	0,23	67	14	0,26	68	Σ		
Варіант 3								
1	0,20	63	8	0,26	67	15	0,17	62
2	0,19	64	9	0,28	70	16	0,30	70
3	0,28	70	10	0,25	68	17	0,19	64
4	0,26	68	11	0,25	67	18	0,25	68
5	0,23	66	12	0,22	66	19	0,29	69
6	0,21	63	13	0,18	63	20	0,27	68
7	0,24	64	14	0,26	68	Σ		
Варіант 4								
1	0,21	65	8	0,26	67	15	0,17	62
2	0,19	63	9	0,28	70	16	0,30	70
3	0,28	70	10	0,25	68	17	0,19	64
4	0,26	68	11	0,25	67	18	0,25	68
5	0,23	64	12	0,22	66	19	0,29	69
6	0,21	63	13	0,18	63	20	0,27	68
7	0,24	68	14	0,26	68	Σ		
Варіант 5								
1	0,21	64	8	0,26	67	15	0,19	61
2	0,19	65	9	0,27	70	16	0,28	70
3	0,27	70	10	0,25	68	17	0,22	64
4	0,26	69	11	0,24	69	18	0,25	68
5	0,23	66	12	0,22	64	19	0,29	69
6	0,22	64	13	0,18	62	20	0,27	68
7	0,24	67	14	0,26	68	Σ		

### Другий блок:

1. Для маркетингового дослідження в компанії А була створена вибірка, що складається з 11 магазинів з приблизно однаковими об'ємами продажів подарункових наборів цукерок. Розглянемо дві незалежні змінні – ціна набору в грн. (X1) і місячний бюджет рекламної кампанії, що проводиться в магазині, виражений в доларах (X2). До цього бюджету входять витрати на оформлення

вивісок і вітрин, а також на роздачу купонів і безкоштовних зразків. Залежна змінна  $Y$  є кількістю наборів, проданих за місяць (табл. 5.2).

Таблиця 5.2

Дані маркетингового дослідження

Магазин	Об'єм продаж $Y$	Ціна $X_1$	Витрати на рекламу $X_2$
1	3825	59	200
2	1096	59	200
3	761	59	200
4	2088	69	300
5	820	69	100
6	2114	69	300
7	1882	69	400
8	2159	79	400
9	1602	79	200
10	3354	79	500
11	2927	79	500

Скласти рівняння множинної регресії, оцінити коефіцієнти рівняння і зробити висновок про адекватність рівняння. Порівняти результати в програмах MS Excel и Statistica.

**Завдання до теми 4:**

1. В роботі повинні бути надані теоретичні визначення типів контрольних карт та їх призначення, надане обґрунтування вибору контрольних карт для завдання, виконана побудова карт, зроблені висновки з тематики індивідуального завдання.

2. Номер варіанта завдання вибирається згідно з останньою цифрою номера залікової книжки: остання цифра 1– 1–20 вимірювання з табл. 3.8, 2 – 11–30 вимірювання, 3 – 21–40, 4 – 31–50, 5 – 41–60, 6 – 15–35, 7 – 26–46, 8 – 36–56, 9 – 5–26 и 0 – 11–40 вимірювання з таблиці 5.3.

3. Індивідуальне завдання повинне мати титульний аркуш, зміст або план, основний зміст, перелік посилань.

**Завдання до варіантів 0-9:**

Припустимо, що ви здійснюєте контроль товщини стінки труб при холодній прокатці. Кожні 15 хвилин проводиться автоматично замір товщини, дані з датчиків заносяться в таблицю (табл. 5.3) в режимі реального часу.

Вважається, що процес виходить з-під контролю, якщо значення товщини стінки виходить за контрольні межі. Побудувати потрібні контрольні карти, проаналізувати процес. Зробити висновки до якості процесу.

Таблиця 5.3

Вхідні дані вимірювань товщини стінки труб

№	п	№	п	№	п	№	п	№	п
1	2,89	13	3,23	25	3,20	37	3,49	49	3,20
2	3,19	14	3,20	26	3,22	38	3,23	50	3,00
3	3,20	15	3,19	27	3,15	39	3,04	51	3,36
4	3,22	16	3,45	28	3,29	40	3,05	52	3,16
5	3,22	17	3,10	29	3,08	41	3,33	53	3,20
6	3,20	18	3,00	30	3,19	42	3,24	54	3,51
7	3,18	19	3,22	31	3,20	43	3,26	55	3,16
8	2,98	20	3,31	32	3,20	44	3,21	56	3,01
9	3,20	21	3,14	33	3,20	45	3,19	57	2,88
10	3,23	22	3,19	34	3,22	46	2,91	58	3,20
11	3,18	23	2,97	35	3,21	47	3,22	59	3,20
12	3,27	24	3,36	36	3,21	48	3,20	60	3,29

При оцінці якості виконання індивідуального завдання враховуються: відповіді на всі питання; глибина розкриття теми; ясність викладу, стиль. Обсяг роботи: 15...20 сторінок формату А4; шрифт 14; міжрядковий інтервал 1,5.

**Захист роботи** здійснюється у ході співбесіди студента з керівником. Наявність заліку враховується при формуванні заключної оцінки з навчальної дисципліни.

## ЗМІСТ

ХАРАКТЕРИСТИКА ДИСЦИПЛІНИ.....	3
1. ЗАГАЛЬНІ МЕТОДИЧНІ ВКАЗІВКИ.....	4
2. РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ.....	4
3. ПРОГРАМА І ПИТАННЯ ДЛЯ ЗАКЛЮЧНОГО КОНТРОЛЮ .....	5
4. ЛЕКЦІЙНИЙ МАТЕРІАЛ.....	6
5. ЗАВДАННЯ ДЛЯ ІНДИВІДУАЛЬНИХ РОБІТ .....	48