

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ,
МОЛОДІ ТА СПОРТУ УКРАЇНИ**

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

**Лабораторний практикум
з модуля 1 "Методи описової статистики
та статистичний аналіз рядів розподілу"
навчальної дисципліни**

"СТАТИСТИКА"

**для студентів напряму підготовки
6.030506 "Прикладна статистика"
денної форми навчання**

Харків. Вид. ХНЕУ, 2012

Затверджено на засіданні кафедри статистики та економічного прогнозування.

Протокол № 2 від 23.09.2011 р.

Укладачі: Раєвнєва О. В.

Горохова О. І.

Чанкіна І. В.

Шаповалова В. О.

Дмитрусенко К. О.

Бобкова О. Ю.

Л-12 Лабораторний практикум з модуля 1 "Методи описової статистики та статистичний аналіз рядів розподілу" навчальної дисципліни "Статистика" для студентів напряму підготовки 6.030506 "Прикладна статистика" денної форми навчання / укл. Раєвнєва О. В., Горохова О. І., Чанкіна І. В. та ін. – Х. : Вид. ХНЕУ, 2012. – 60 с. (Укр. мов.)

Подано лабораторні роботи, метою яких є закріплення й поглиблення знань теоретичного та практичного матеріалу, придбання навичок аналізу різних типів даних за допомогою системи STATISTICA.

Рекомендовано для студентів напряму підготовки 6.030506 "Прикладна статистика" денної форми навчання.

Вступ

Статистичний аналіз даних є невід'ємним атрибутом системи управління на всіх її рівнях – від окремого економічного об'єкта до національної економіки в цілому. Оволодіння методами статистичної обробки даних із використанням комп'ютерних технологій є важливою складовою професійної підготовки статистика.

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.

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

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

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

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

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

Загальні методичні рекомендації

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.

STATISTICA являється комплексною системою аналізу даних NF може використовувати дані майже з усіх типів файлів і генерувати табличні звіти, графіки, малюнки розподілів і трендів, описову статистику і проводити складні види статистичного аналізу.

Кваліфікаційні вимоги в галузі статистика до студентів

Необхідна навчальна база перед початком вивчення дисципліни

З метою кращого засвоєння навчального матеріалу дисципліни студенти повинні до його початку опанувати знаннями та навичками в галузі загальної теорії статистики.

У свою чергу знання з даної дисципліни забезпечують успішне засвоєння виконання курсових проектів і дипломних проектів.

У результаті вивчення дисципліни студенту необхідно набути наступних компетенцій.

**Вимоги до компетенцій, які повинен здобути студент, вивчаючи
дисципліну "Статистика"**

Назва компетенцій	Зміст компетенцій	Вміння студента щодо даної компетенції
1	2	3
1. Обліково-статистична	1.1. Здатність до здійснення статистичних спостережень	1.1.1. Володіти сучасними інформаційними технологіями збору, обробки та поширення даних
	1.2. Здатність проводити статистичні розрахунки	1.2.1. Проводити групування та вторинне групування.
		1.2.2. Проводити розрахунок абсолютних та відносних показників. 1.2.3. Проводити розрахунок середніх показників. 1.2.4. Проводити розрахунок показників центра, форми та варіації розподілу 1.2.5. Проводити розрахунок показників оцінки параметрів генеральної сукупності при вибірковому обстеженні
2. Аналітична	2.1. Здатність щодо проведення статистичного аналізу	2.1.1. Збирати та обробляти інформацію відповідно до наукових методів її первинної обробки. 2.1.2. Аналізувати інформацію, наведену у вигляді просторових та динамічних рядів. 2.1.3. Аналізувати систему показників, які характеризують явище в цілому та його структуру 2.1.4. Проводити дисперсійний аналіз рядів розподілу. 2.1.5. На основі аналізу інформації виробляти пропозиції щодо підвищення якості управлінських рішень. 2.1.6. Використовувати засоби візуалізації отриманих результатів
	2.2. Здатність аналізувати та використовувати зведену інформацію	2.2.1. Узагальнювати систему статистичних показників для складання статистичних бюлетенів, щорічників з метою висвітлення економічних та соціальних процесів. 2.2.2. Складати ґрунтовні аналітичні огляди, інформаційні записки відповідно до потреб зацікавлених установ

1	2	3
3. Організаційно-методологічна	3.1. Здатність щодо засвоєння загальних положень статистики	3.1.1. Знати сучасні інформаційні системи та ППП обробки статистичної інформації. 3.1.2. Знати основні категорії та поняття, закономірності дослідження процесів та явищ. 3.1.3. Володіти методами оцінювання явищ та процесів соціальної та економічної сфери. 3.1.5. Знати методи аналітичної та комбінаційної статистики
	3.2. Здатність впроваджувати статистичну методологію	3.2.1. Володіти методами аналізу рядів розподілу. 3.2.2. Володіти вибірковим методом. 3.2.4. Володіти методами побудови та аналізу структури
4. Контрольна	4.1. Здатність перевіряти достовірність статистичної інформації	4.1.1. Досліджувати, прогнозувати та перевіряти достовірність отримання результатів шляхом логічного та арифметичного аналізів. 4.1.2. Контролювати достовірність інформації шляхом перевірки взаємозв'язку показників
5. Інформаційна	5.1. Здатність використовувати при вирішенні економічних завдань сучасні засоби обробки інформації	5.1.1. Використовувати різноманітні можливості пакета прикладних програм STATISTICA та Excel для аналізу соціально-економічних процесів чи явищ з метою підвищення якості управлінських рішень

Лабораторна робота № 1 "Знайомство з програмою STATISTICA"

Мета – придбання знань щодо основних модулів програми STATISTICA.

Завдання – знайомство зі змістом і розрахунковими можливостями модулів програми STATISTICA.

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. 1.1).

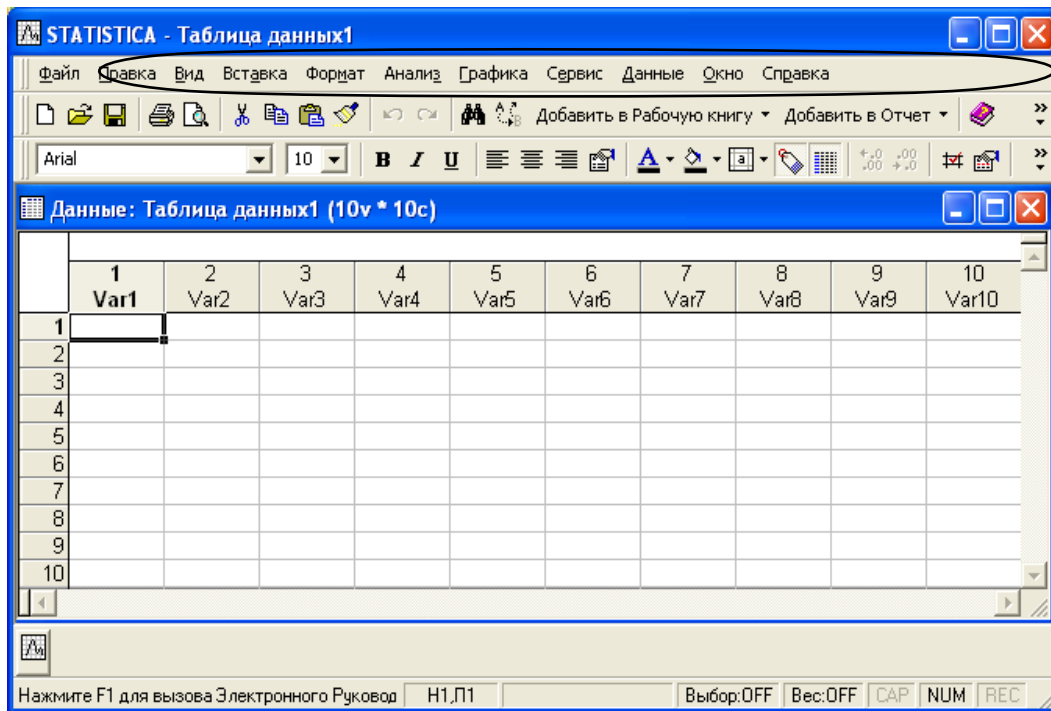


Рис.1.1. Меню програми STATISTICA

Вибір команд із меню, що випадають, можна здійснити декількома різними способами (за допомогою миші або клавіатури). Зазвичай використовують мишу, але іноді корисна й клавіатура.

Для відкриття якого-небудь пункту меню зробіть одну з наступних дій:

- підведіть курсор і клацніть лівою кнопкою миші на пункті меню;
- натисніть комбінацію клавіш **ALT+<підкреслена буква в назві меню>**. Наприклад, для відкриття меню File (Файл) натисніть **ALT+<F>**. Цей спосіб відкриття меню часто використовується для написання макрокоманд, які становлять послідовність команд, що вводяться за допомогою клавіатури.

За допомогою стрілок клавіатури: {вліво}, {вправо} виберіть необхідний пункт меню. Далі натисніть стрілку {вниз}.

Після того, як пункт меню обраний, у ньому можна ініціювати необхідну вам команду. Тут існують різні способи дій, що приводять до одного результату:

- клацніть на імені команди мишею;
- натисніть на клавішу з буквою, що підкреслена в назві команди. Наприклад, для вибору команди Open Data (Відкрити дані) з меню File (Файл) натисніть на клавішу з буквою {O} (рис. 1.2);

• скористайтеся стрілками {верх} і {вниз} для вибору необхідної команди й натисніть на клавішу ENTER. Зазначимо, що багато команд із меню, що випадає, викликаються за допомогою певної комбінації

"гарячих" клавіш. Наприклад, для вибору команди Open Data (Відкрити дані) з меню File (Файл) використовуйте комбінацію CTRL+O.

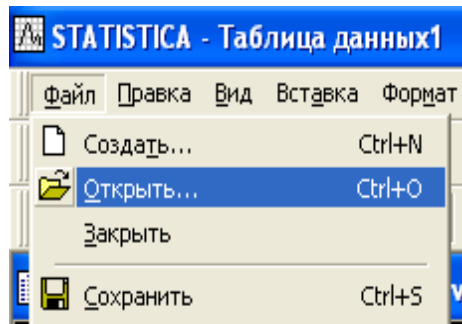


Рис. 1.2. Меню Файл, що викликане за допомогою клавіатури

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. За допомогою клацання мишею на якій-небудь кнопці можна одержати швидкий доступ до відповідної команди. 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. 1.3). Ці установки діють тільки для поточного сеансу роботи. Панель інструментів може бути виведена в одну й два рядки й може бути розташована в різних частинах основного вікна системи.

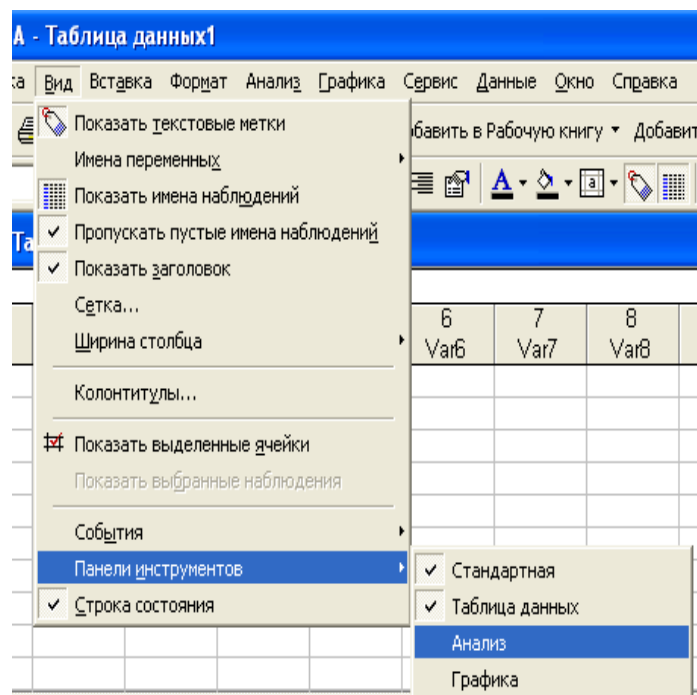


Рис. 1.3. Панель інструментів з меню Вид

Вихідні дані в системі STATISTICA організовані у вигляді електронної таблиці. Якщо ви маєте досвід роботи з електронними таблицями (наприклад, з MS Excel), то вам буде нескладно освоїтися з електронними таблицями в STATISTICA. Таблиці з вихідними даними в STATISTICA носять особливу назву, щоб не плутати їх з іншими таблицями системи.

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 вихідними даними мають назву Variables (Змінні), а рядка 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.

1. 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 (Кластерный 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.

1.1. Запуск модулів

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

Чисельна реалізація статистичних методів в STATISTICA (більша частина статистичних процедур написана на асемблері) така, що вони дозволяють на вашому комп'ютері вирішувати завдання граничної складності як за обсягом й розмірністю оброблюваних даних, так і за точністю й швидкістю обчислень.

Приклад: запуск модуля Основні статистики/Таблиці із вкладки Аналіз.

Крок 1. Клацніть на вкладці Аналіз.

Крок 2. У меню, що випадає, виділіть слова: Basic Statistics/Tables (Основні статистики/таблиці) (рис. 1.4).

Крок 3. Клацніть на даному модулі.

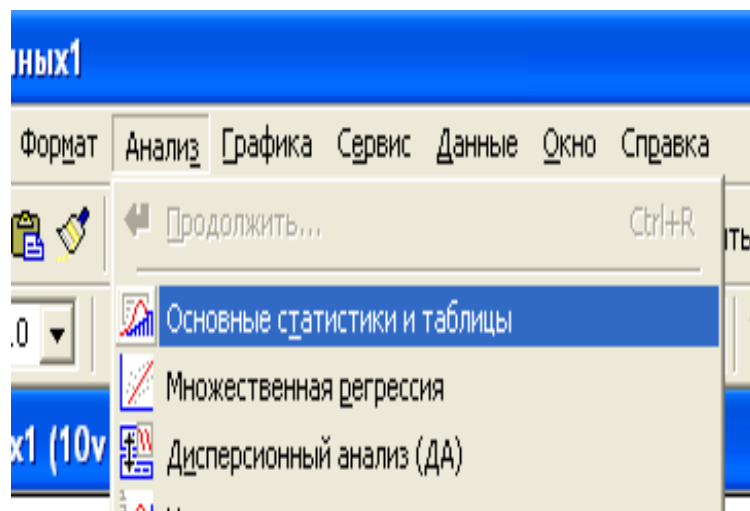


Рис. 1.4. Вкладка Аналіз

Модуль Основні статистики/таблиці запуститься, і робоче вікно системи з'явиться на екрані. У цьому модулі ми навчимося робити перші кроки з обробки даних у системі.

Після вибору в Перемикачі модулів відкривається стартова панель обраного модуля. Далі ви відкриваєте файл даних, задаючи, при необхідності, умови вибору випадків – кнопка S Select cases (Вибрати випадки) – і ваги змінних – кнопка W Weight (Вага) (рис. 1.5).

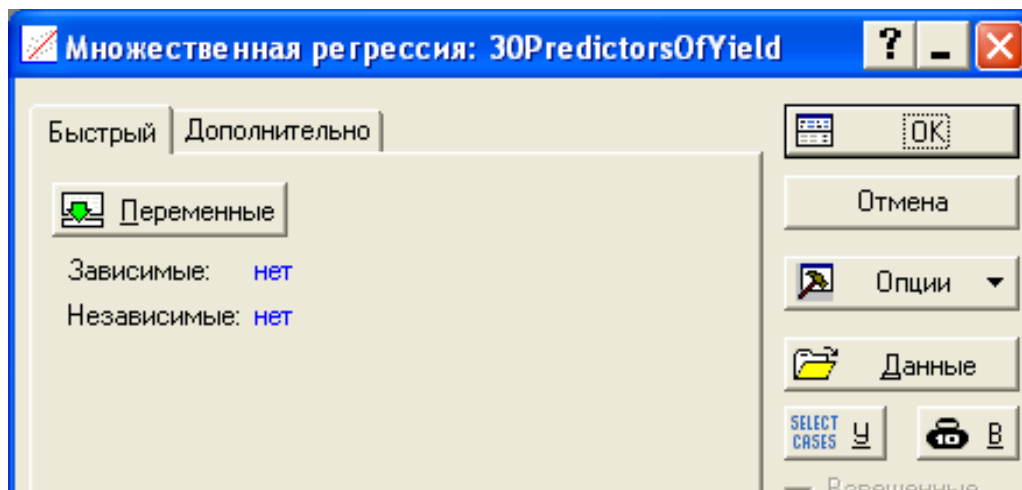


Рис. 1.5. Модуль Множинна регресія

Потім необхідно:

- обрати змінні для аналізу з відкритого файла 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.

Як вже говорилося, у складному проекті потрібно працювати з різними модулями, послідовно перемикаючись між ними. На першому етапі знайомства із системою краще попрацювати в кожному модулі окремо, всебічно експериментуючи в ньому з даними, і відчути його можливості.

Відзначимо ще раз, що без експериментування з даними вам не вдасться освоїти систему й ви ніколи не зрозумієте, що таке обробка даних. STATISTICA – це, насамперед, середовище обробки.

Лабораторна робота № 2 "Подання статистичних даних у програмі STATISTICA"

Мета – придбання студентами навичок роботи з подання й візуалізації статистичних даних за допомогою програми STATISTICA.

Завдання – вивчити основні прийоми по формуванню бази даних в STATISTICA, таблиць і різноманітних графіків; придбання навичок з їхнього редагування.

Методичні рекомендації

1. Створення робочої книги

Рішення будь-якого завдання статистичного аналізу з використанням пакетів прикладних програм починається із процедури введення даних. У середовищі ППП STATISTICA процедура введення реалізується у такий спвісіб. У першу чергу необхідно створити файл, для чого в головному меню системи вибирається File і далі опція New (рис. 2.1).

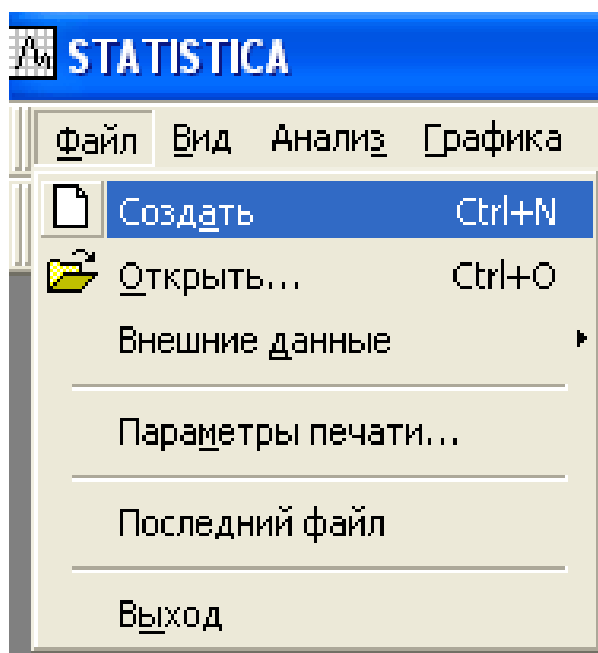


Рис. 2.1. Вікно ППП STATISTICA з відкритим меню Файл

Програма надає можливість створення файлів декількох типів:

1. Spreadsheet – робочий аркуш (електронна таблиця).
2. Report – звіт.
3. Macro (SVB) Program – макрос.
4. Workbook – робоча книга.
5. Browser Window – посилання в мережі.

Робоча книга (Workbook) є найбільш загальним поняттям.

У робочу книгу можуть бути включені будь-які елементи, доступні в STATISTICA, наприклад, робочі аркуші, звіти, документи Microsoft Office.

Для уведення даних буде потрібна електронна таблиця. Для виводу її на екран необхідно вибрати закладку Spreadsheet (рис. 2.2).

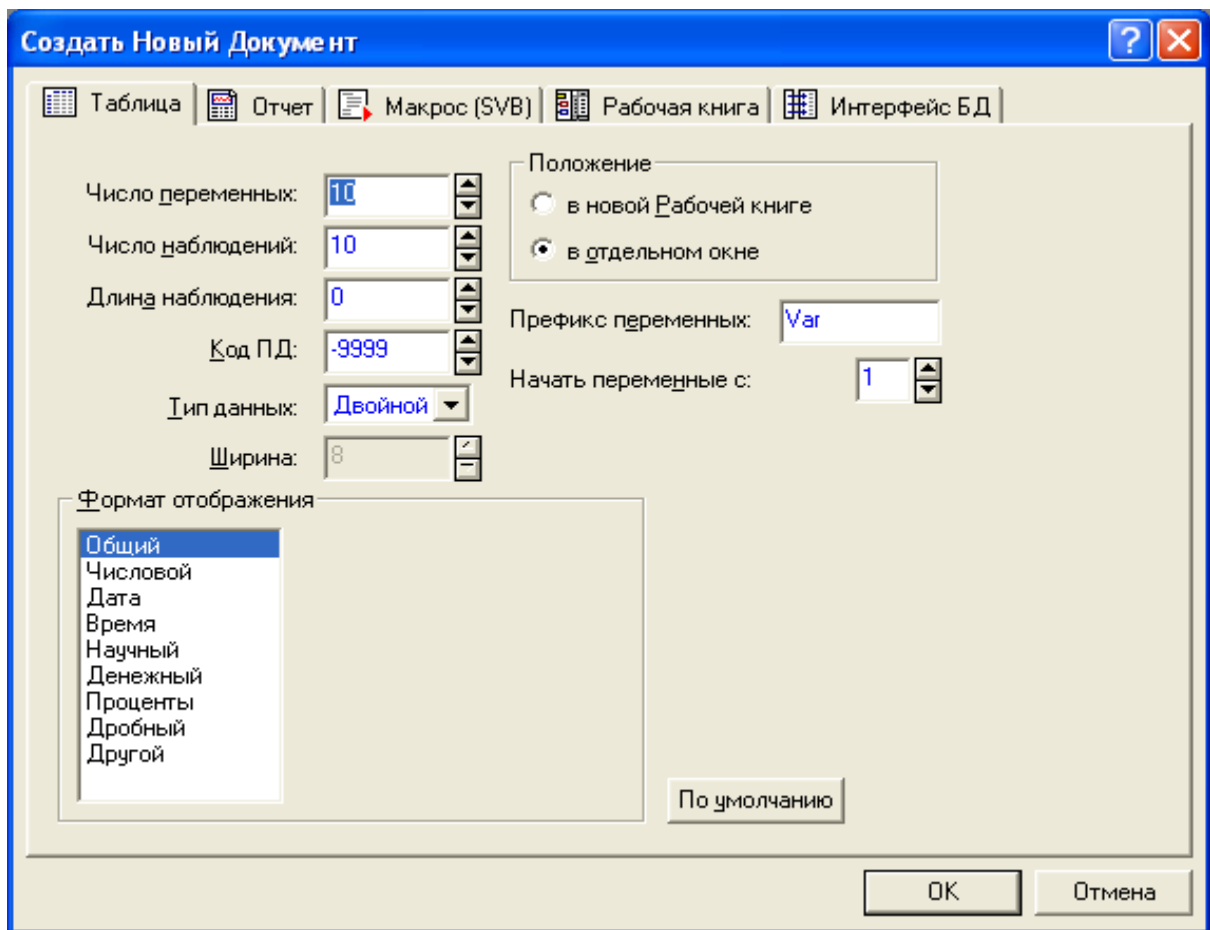


Рис. 2.2. Вікно вибору параметрів створення нового документа STATISTICA

Користувачеві надається можливість визначити розмірність таблиці, тобто задати число змінних (number of variables) і число спостережень (number of cases). Число змінних – це число ознак, що варіюють, значення яких зареєстровані в кожній одиниці досліджуваної статистичної сукупності. Число спостережень – число одиниць сукупності (обсяг сукупності). Потім вибирається місце, де буде зберігатися створювана електронна таблиця (Placement – Положення). Доцільно вибрати "In a new Workbook – У новій Робочій книзі", що означає створення нового робочого аркуша (електронної таблиці) усередині нової робочої книги. Процедура "As a stand-alone window – В окремому вікні" означає створення окремого робочого аркуша.

Далі можна вибрати різні варіанти висновку даних. Поле Case name length (Довжина спостереження) задає довжину ім'я рядка, поле MD Code (Код ПД) задає код висновку пропущених осередків (за замовчуванням вони не враховуються при розрахунку показників), поле

Default data type (Тип даних) дозволяє вибрати формат даних (числовий, текстовий, бінарний; за замовчуванням обраний тип Double – це означає, що система буде сприймати як текстові, так і числові змінні), поле Variable length призначено для того, щоб задавати довжину даних у випадку вибору текстового формату даних. Поле Display format (Формат відображення) задає параметри висновку що вводяться або розрахованих даних (звичайний, числовий, відсотковий і т. д.), поле Var name prefix (Префікс змінних) пропонує задати назва створюваних змінних, а поле Var name start number (почати змінні с) задає стартовий порядковий номер першої змінної. Кнопка Default (За замовчуванням) дозволяє скинути всі налаштування за замовчуванням.

Після того, як всі необхідні умови обрані, натискаємо кнопку ОК.

З'являється робочий аркуш, що відповідає заданим умовам, а ліворуч від нього дерево робочої книги. За замовчуванням робоча книга називається Workbook1. У середині цієї робочої книги будуть утримуватися всі елементи, створювані на основі вихідних даних. З ними можна робити всі елементарні перетворення (збереження, видалення, перейменування). Для цього досить клацнути по них правою кнопкою миші в дереві робочої книги й вибрати відповідний пункт (рис. 2.3).

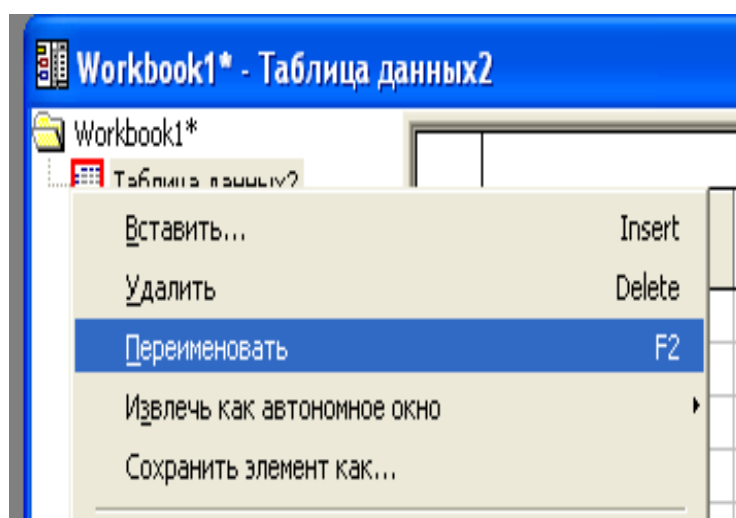


Рис. 2.3. Вікно STATISTICA з обраною функцією перейменування документа

Робочі книги, що створені, повинні мати стандартні й легко ідентифіковані з конкретним користувачем імена. Для задання ім'я робочої книги вибирається розділ меню File і в ньому опція Save as (рис. 2.4).

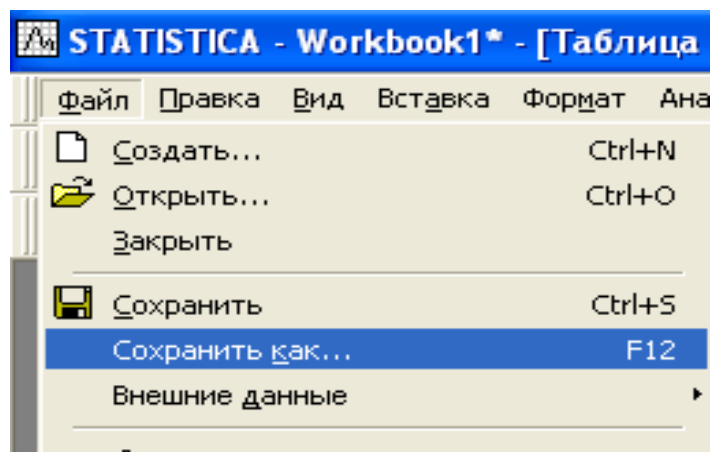


Рис. 2.4. Вибір меню збереження документа STATISTICA

Звертаємо увагу, що для робочої книги характерне розширення *.stw. При запуску STATISTICA відкриває файл, що використовувався при останньому сеансі. Щоб відкрити потрібний файл, необхідно скористатися меню File/Open і вибрати в провіднику Windows папку, що цікавить нас, і файл. При цьому система пропонує всі можливі підтримувані формати (*.stw, *.sta, *.stg).

2. Введення даних

Далі можна приступати до безпосереднього введення вихідних даних. Спочатку задамо назви спостережень, у нашому прикладі – це назва регіону (табл. 2.1). Для цього необхідно двічі клацнути лівою кнопкою миші по заголовку рядка і надрукувати текст.

Переважно введення даних виробляється стандартним чином і додаткових коментарів не вимагає (після запису числа на екрані натискається клавіша "Enter", десяткові дробі вводяться через кому).

Таблица 2.1

Наявний дохід та середньомісячна номінальна заробітна плата по регіонах України

Область	Наявний дохід на душу населення, грн	Середньомісячна номінальна заробітна плата найманих працівників (на одного штатного працівника), грн
1	2	3
АР Крим	12 265,7	1 707
Вінницька	12 380,2	1 511
Волинська	11 130,8	1 427

1	2	3
Дніпропетровська	16 646,6	1 963
Донецька	17 380,9	2 116
Житомирська	12 385,3	1 493
Закарпатська	10 028,1	1 562
Запорізька	16 174	1 843
Івано-Франківська	12 015,4	1 627
Київська	15 085,8	1 987
Кіровоградська	11 757,6	1 537
Луганська	14 368,4	1 873
Львівська	13 657	1 667
Миколаївська	13 297,6	1 806
Одеська	12 621,7	1 787
Полтавська	14 747,3	1 733
Рівненська	11 531,3	1 614
Сумська	13 655,6	1 593
Тернопільська	10 733,2	1 412
Харківська	14 902,2	1 804
Херсонська	11 606,3	1 482
Хмельницька	12 332,4	1 521
Черкаська	12 404,4	1 532
Чернівецька	10 275,4	1 523
Чернігівська	12 996	1 465
м. Київ	27 792,3	3 161
м. Севастополь	13 023,9	1 882

Оскільки це масив даних, число змінних (ознак) дорівнює двом, число спостережень – 27. Вихідні дані поміщені в робочу книгу з ім'ям Workbook1.stw , під ім'ям Var1 і Var2.

3. Редагування вихідних даних

Якщо в дослідженні беруть участь тільки регіони без міст, то файл необхідно перетворити. Для цього необхідно з вихідного ряду забрати два останніх рівні ряду. Щоб здійснити цю операцію, потрібно створити ще одну змінну, бажано зробити для цього новий робочий аркуш.

Для цього клацаємо правою кнопкою миші по робочому аркуші Spreadsheet1 (Таблиця даних1) у дереві робочої книги й вибираємо там опцію Copy Workbook document (З документ Робочої книги). Після цього клацаємо правою кнопкою миші в дереві робочої книги й вибираємо

опцію Paste as Workbook document/As next sibling (Вставити як документ Робочої книги/Наступної на рівні). Це означає додавання копії робочого аркуша в дерево робочої книги після поточних елементів, але не в дерево старого робочого аркуша (опція As a next child). Тому що обоє отриманих робітників аркуша будуть мати однакові назви, то доцільно перейменувати новий робочий аркуш, наприклад в Spreadsheet2. Після цього необхідно видалити в новій змінній дані по містах.

Тепер у робочій книзі втримується два робочих аркуша, при цьому Spreadsheet1 є активним (підсвічується червоним кольором у дереві робочої книги).

Варто звернути увагу на загальне правило роботи в ППП STATISTICA: всі процедури, що виконуються з даними активного в даний момент робочого аркуша (виділеного червоним кольором).

Тому необхідно зробити активним новий робочий аркуш Spreadsheet2. Для цього клацаємо по ньому правою кнопкою миші в дереві робочої книги й вибираємо функцію Use as Active Input (Зробити таблицю активною).

4. Основні операції з рядками й стовпцями

У цьому підрозділі буде представлена інформація, корисна для використання можливостей програми при рішенні будь-яких завдань. Так само, як і елементи в дереві робочої книги, рядки й стовпці на робочому аркуші системи можна додавати, переміщати, видаляти, копіювати.

Стандартна процедура створення нової змінної полягає в наступному. Необхідно двічі клацнути лівою кнопкою миші по сірому полю праворуч від заголовка існуючої змінної. З'явиться діалогове вікно (рис. 2.5), що пропонує додати змінні (Add Variables) або спостереження (Add Cases).

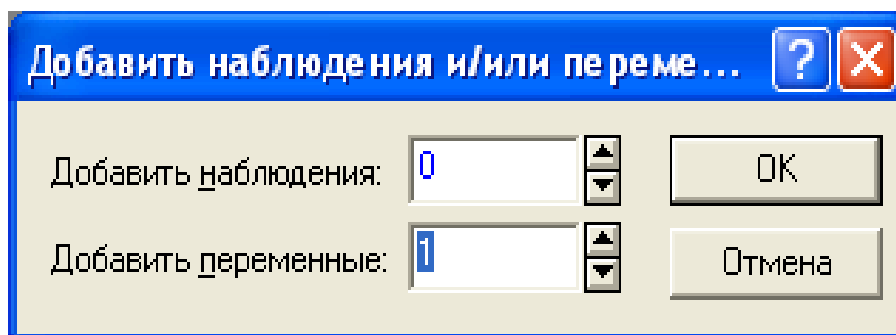


Рис. 2.5. Вікно вибору параметрів створення змінної

Додамо, наприклад, одну змінну, тому у відповідному полі ставиться цифра 1. Клацаємо ОК. На екрані з'являється панель для вибору параметрів нової змінної (рис. 2.6).

У вікні, що з'явилося, можна задати ім'я нової змінної, наприклад Var2. Кнопки ">>" і "<<", що знаходяться під кнопкою Cancel у правому верхньому куті вікна, дозволяють переміщатися по змінним і надбудовам.

Надалі, при необхідності виклику даного вікна, потрібно двічі клацнути лівою кнопкою миші на заголовку кожної з існуючих змінних.

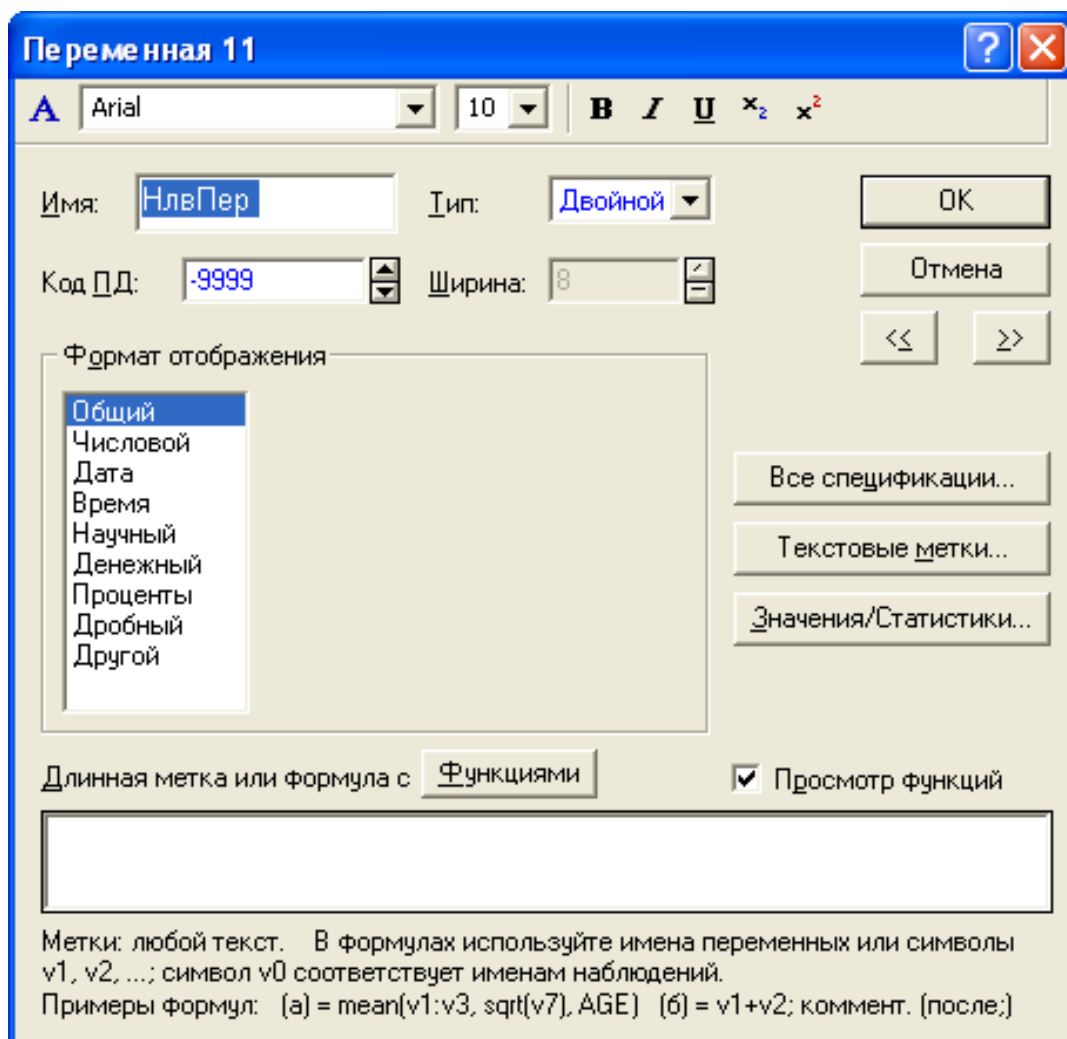


Рис. 2.6. Вікно вибору параметрів змінної

У поле Display Format можна також задати формат змінної, за замовчуванням він звичайний (General), але нас цікавить формат числової (Number). Використання цього формату дозволяє встановлювати необхідну точність обчислень, тобто кількість десяткових знаків після коми (поле Decimal places) (рис. 2.7).

Коли на робочому аркуші перебуває трохи змінних, то можна додавати нові зі вказівкою місця їхнього додавання. Для цього необхідно клацнути правою кнопкою миші по заголовку будь-якого стовпця й вибрати функцію Add Variables (у цьому ж меню можна вибрати функції переміщення, видалення й копіювання змінних, виконувани аналогічним чином).

У вікні, що з'явилося, крім вже відомих нам процедур існує поле After, у якому вказується ім'я змінної, після якої ми хочемо помістити знову створювану.

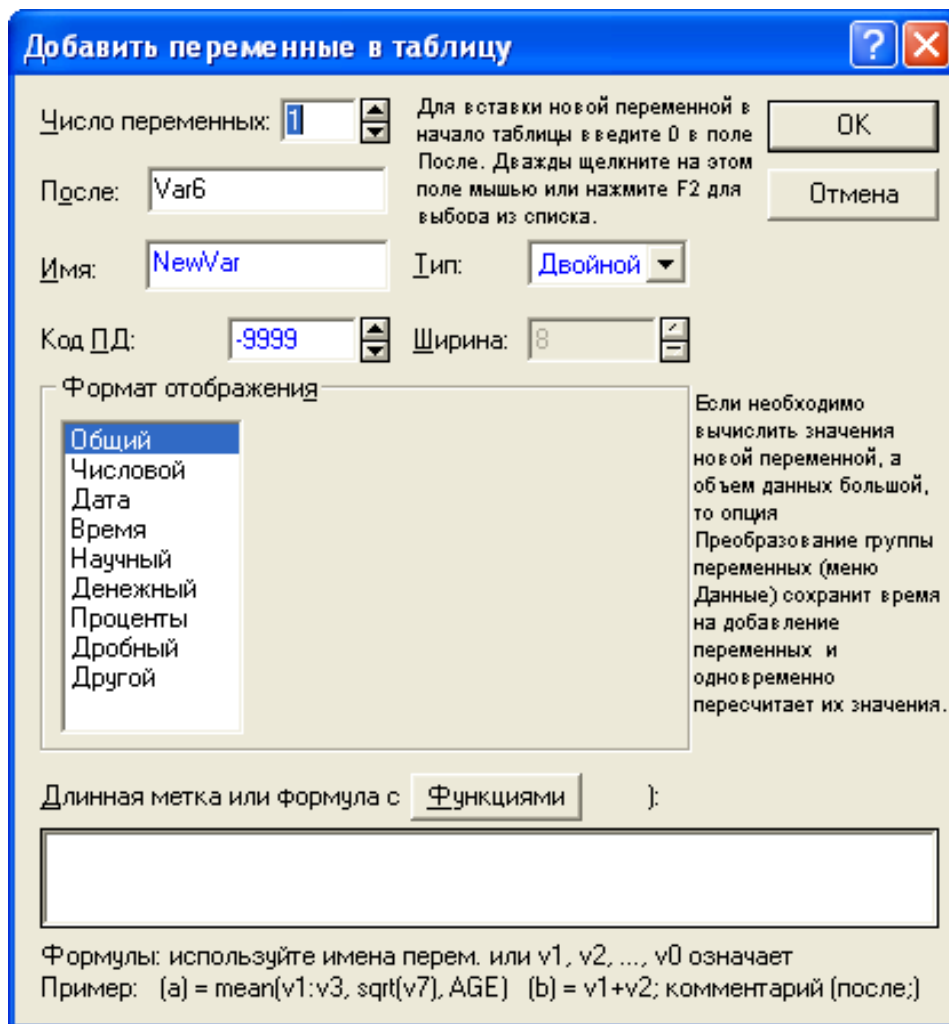


Рис. 2.7. Вікно функції додавання нової змінної

При виборі інших функцій у цьому меню також потрібна вказівка імен змінних, наприклад, від якої й до якої змінної варто видалити стовпці.

Дані зі стовпців поточного робочого аркуша, інших робітників аркушів, документів Excel можна копіювати однаковим чином.

Для цього виділяємо область, що цікавить нас, потім, клацнувши правою кнопкою миші, у підменю, що з'явився, вибираємо функцію Copy.

Далі клацаємо правою кнопкою миші на заголовку нової змінної й у меню, що з'явився, вибираємо функцію Paste (Вставка).

Всі ці функції справедливі й для рядків. Наприклад, ми хочемо видалити рядка.

Для цього потрібно клацнути правою кнопкою миші по будь-якому номері рядка й вибрати функцію Delete cases.

Далі у вікні, що з'явилося, у поле From case ставимо номер рядка, починаючи з якого потрібно видалити дані; у поле To case ставимо номер рядка, по якому потрібно видалити дані включно, і натискаємо ОК (рис. 2.8).

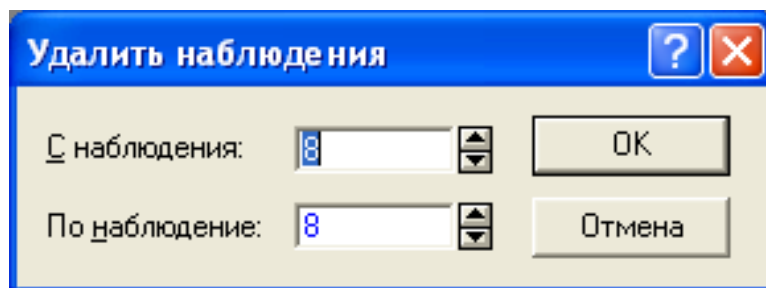


Рис. 2.8. Функція видалення рядків в STATISTICA

Всі функції зі змінними й спостереженнями можна запусити за допомогою кнопок: Vars і Cases відповідно в панелі інструментів.

Лабораторна робота № 3 "Графічний аналіз даних у середовищі "STATISTICA"

Мета – придбання навичок щодо побудови графіків в системі STATISTICA.

Завдання – вивчити основні прийоми створення та форматування графіків в системі STATISTICA.

Методичні вказівки

STATISTICA включає в себе велику кількість різноманітних категорій і типів графіків. Це всілякі графіки на площині й у просторі, включаючи аналітичні графіки, ділові графіки та діаграми, спеціалізовані

статистичні графіки (гістограми, матричні, категоризовані графіки, діаграми розсіювання та ін), піктографіки. 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.

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.

Розглянемо побудову графіків на прикладі даних з табл. 3.1.

Таблиця 3.1

Наявний дохід на душу населення, грн

Область	Роки		Область	Роки	
	2000	2001		2000	2001
1	2	3	4	5	6
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

1	2	3	4	5	6
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. 3.1).

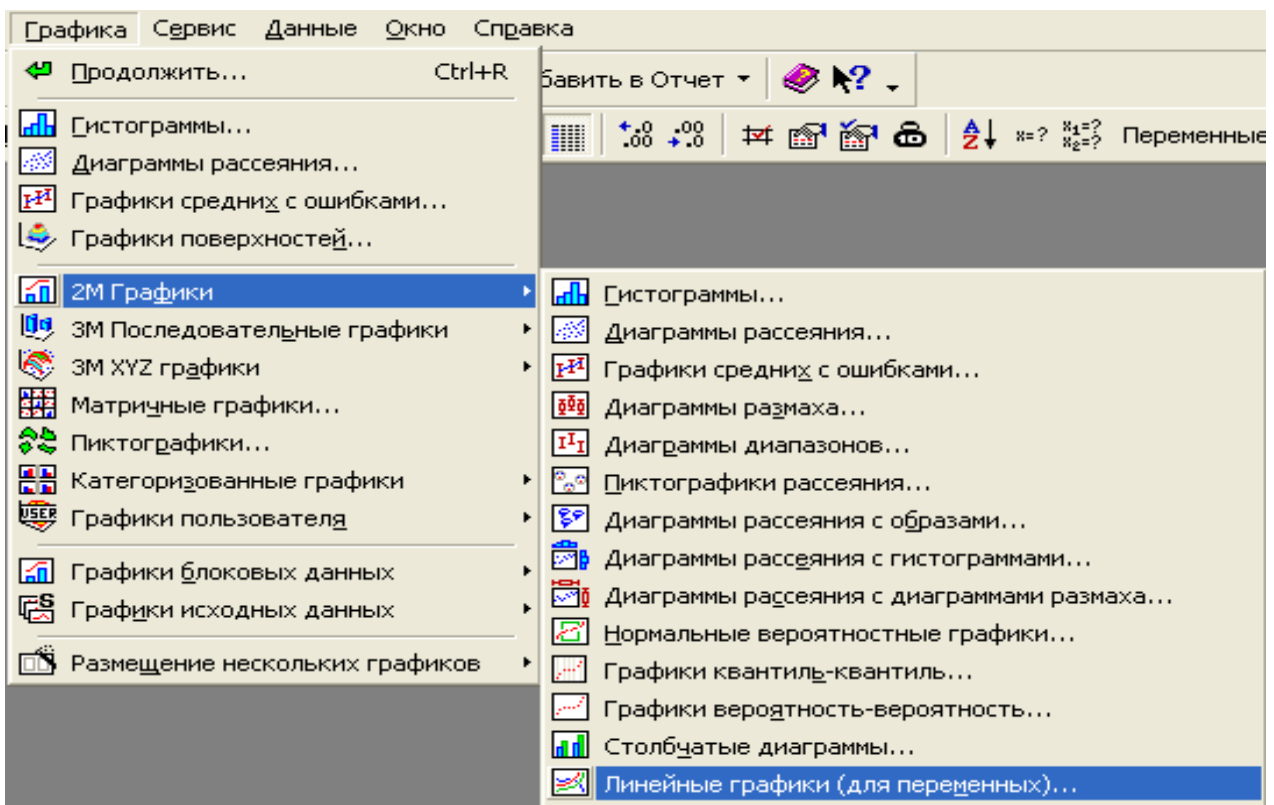


Рис. 3.1. Вибір діалогового вікна побудови графіка

2. Натиснути кнопку "Variables" на вкладці "Quick". Вибрати всі змінні. Вибрати тип графіка "Multiple" (рис. 3.2).

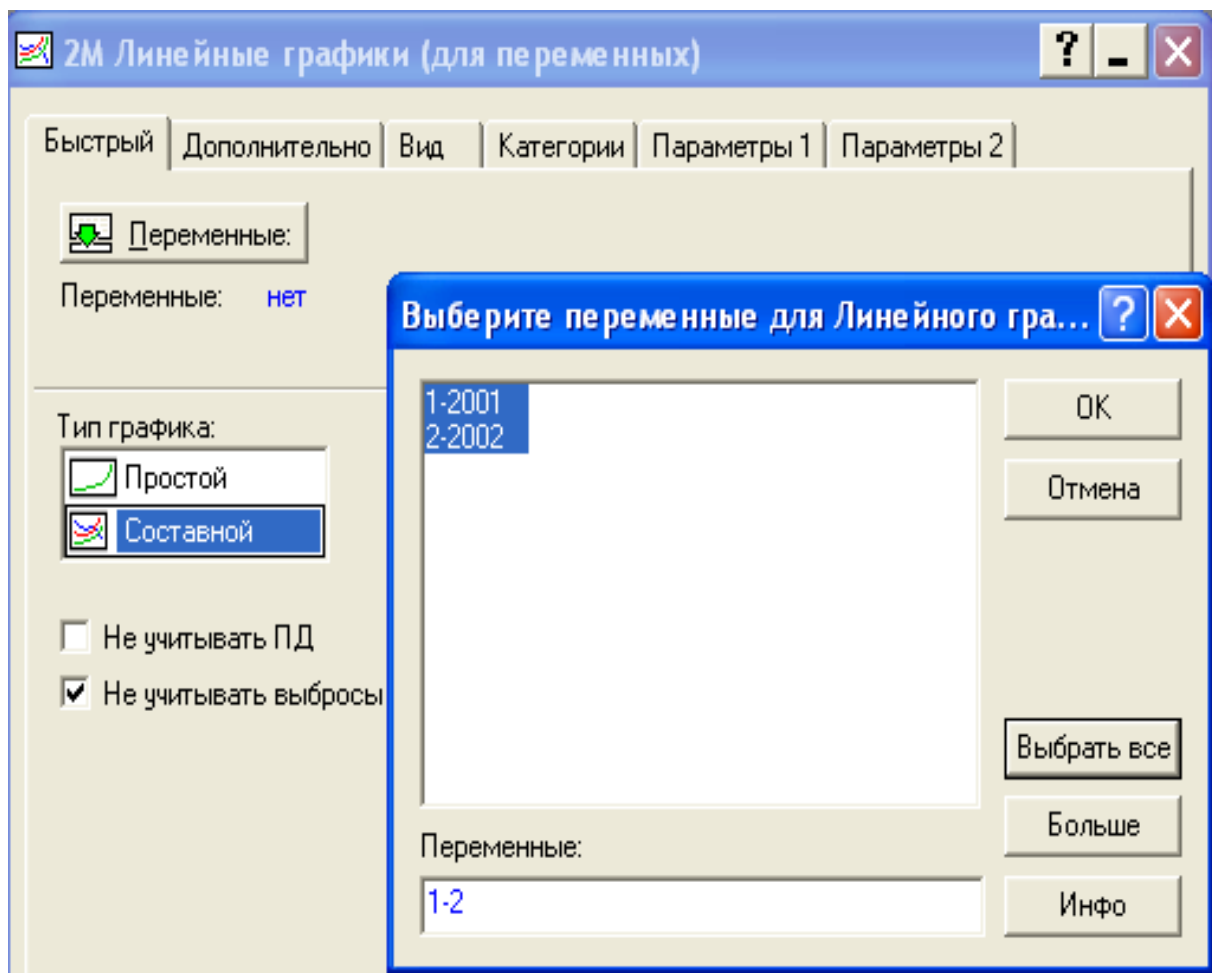


Рис. 3.2. Вибір змінних та типу графіка

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 "ОК" (рис. 3.3).

4. Зберегти графік у файлі з розширенням .STW. Результат побудови графіка представлений на рис. 3.4.

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 "ОК", знову вибрати всі змінні, "ОК" (рис. 3.5). Зберегти графік під новим ім'ям.

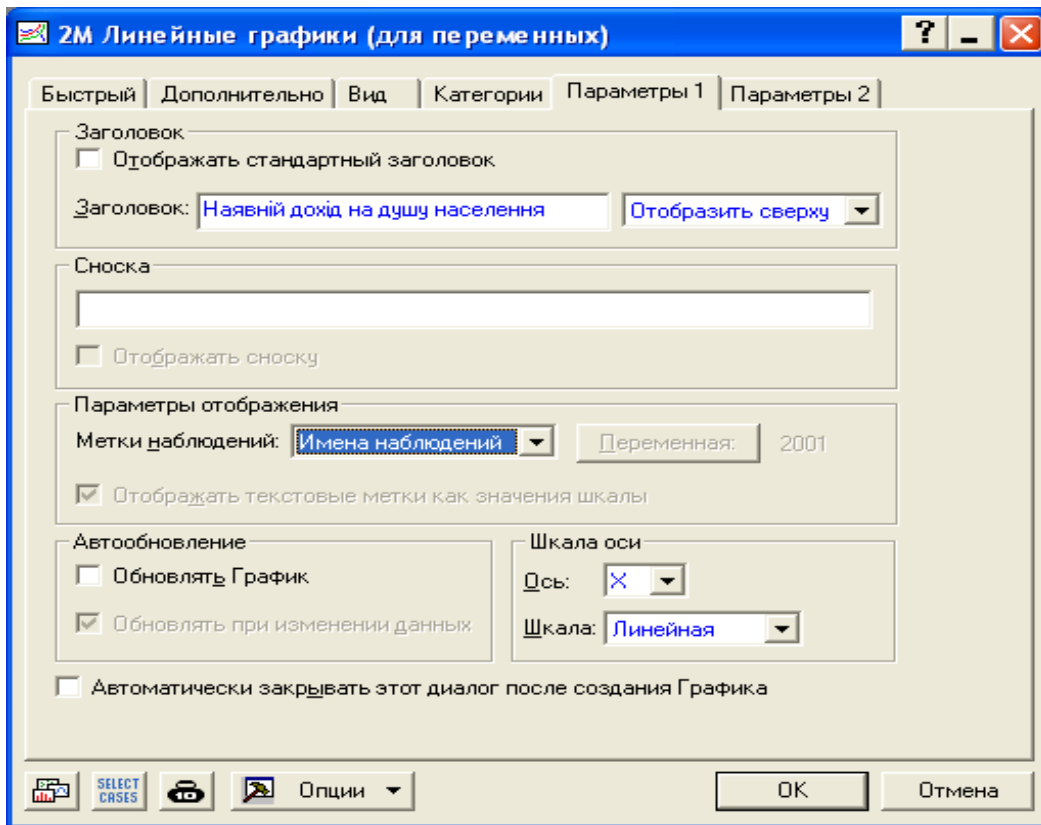


Рис. 3.3. Вибір підпису значень та заголовка діаграми

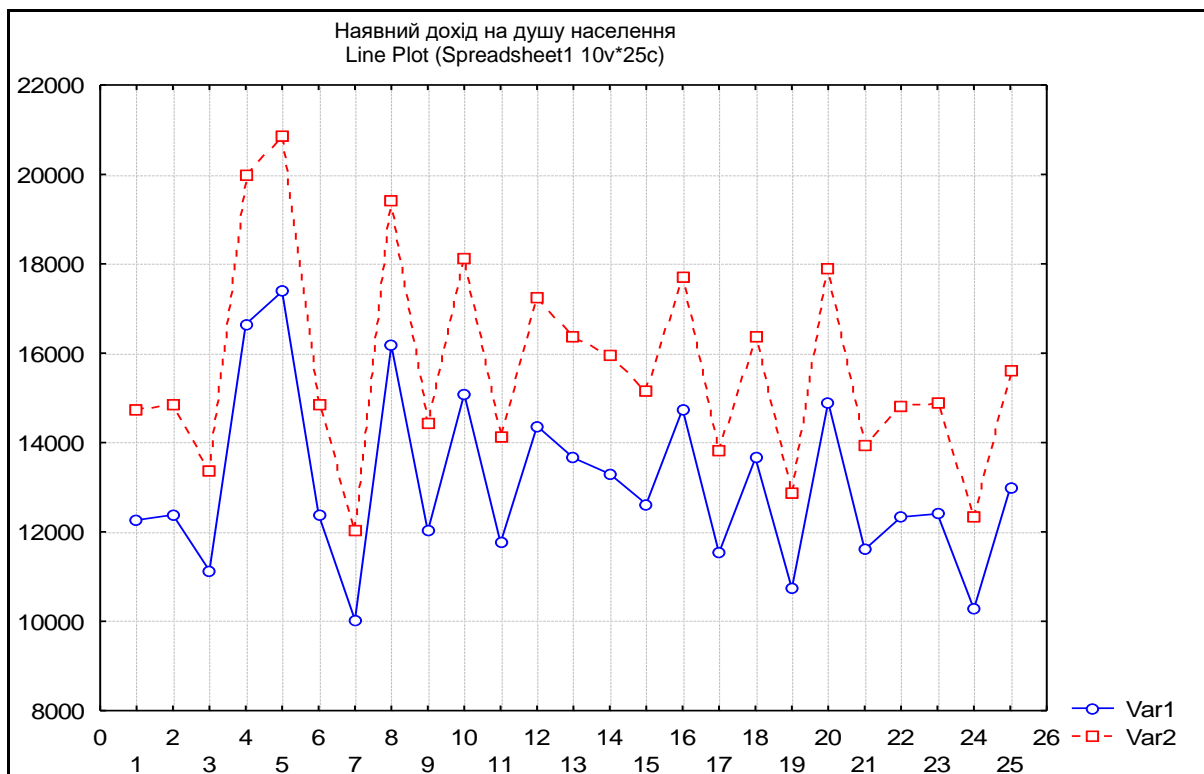


Рис. 3.4. Розподіл наявного доходу населення серед регіонів України за 2 роки

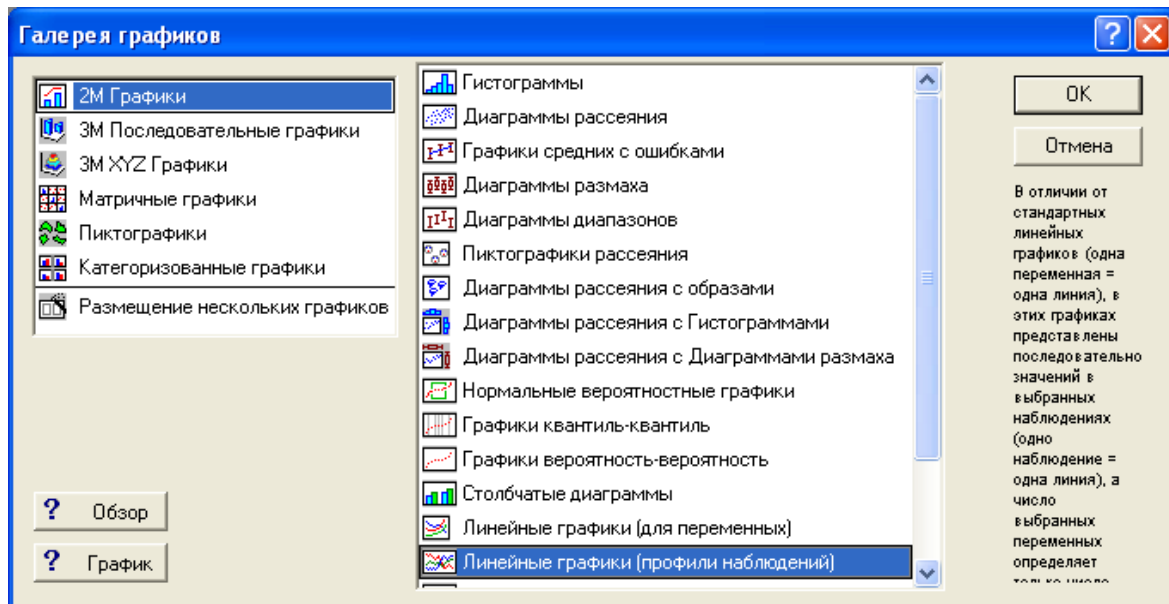


Рис. 3.5. Зміна типу графіку

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

1. Клацнути правою клавішею миші по кривій "2000".
2. У контекстному меню вибрати "Bar Options".
3. У вікні "Bars" включити перемикач "Display bars" і встановити товщину стовпця діаграми "Width", "OK". Зберегти графік під новим ім'ям (рис. 3.6).

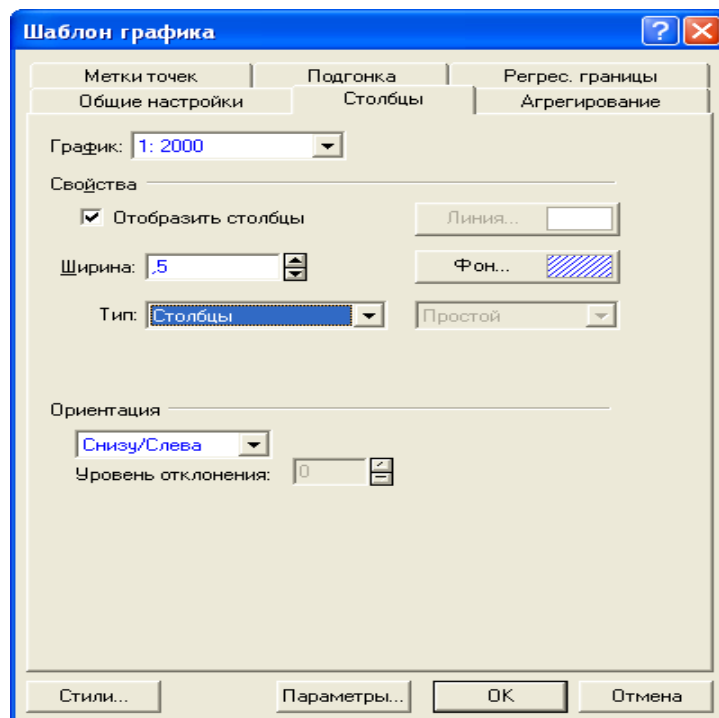


Рис. 3.6. Додавання стовпчикової діаграми

Форматування графіків

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".
- 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. 3.7).

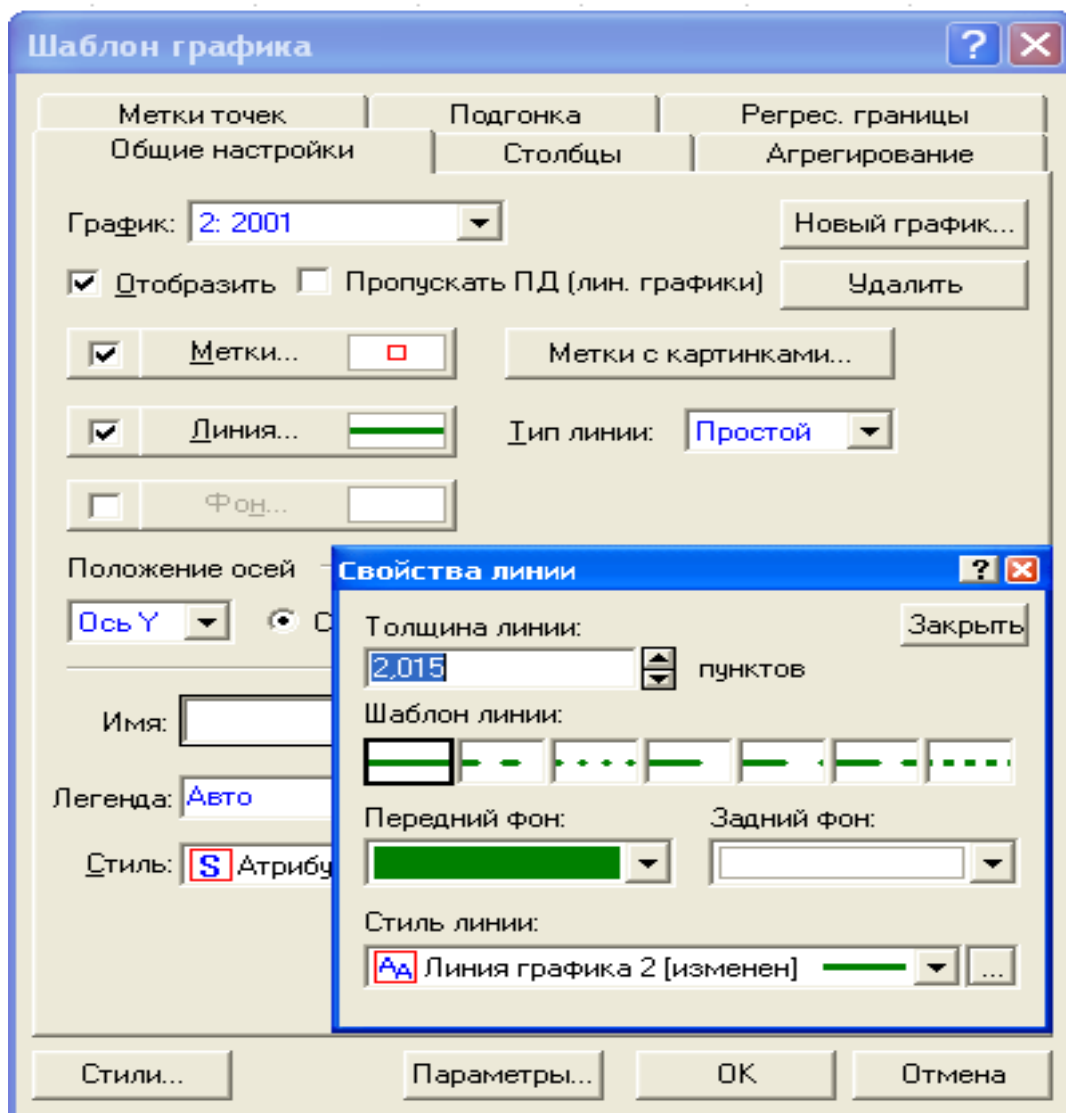


Рис. 3.7. Зміна кольору та товщини лінії графіка

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. 3.8).

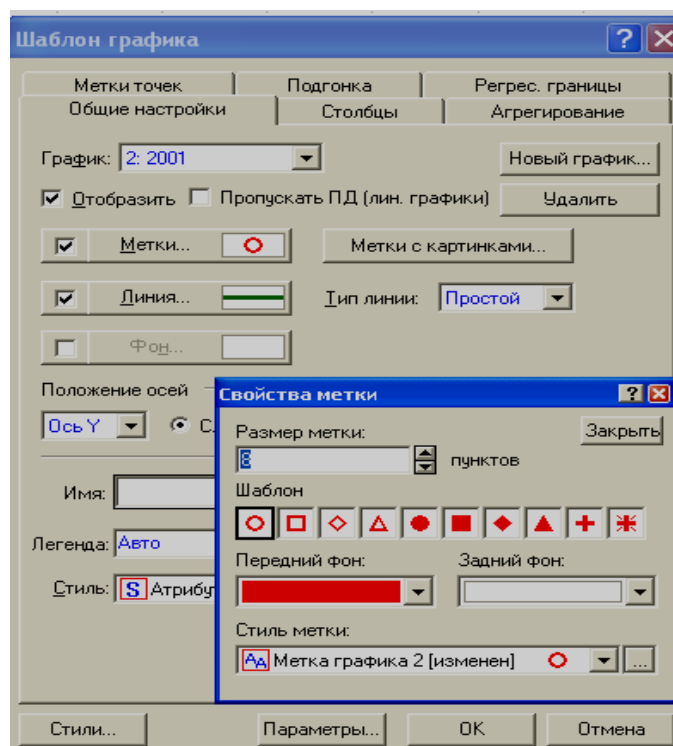


Рис. 3.8. Зміна властивостей маркера

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. 3.9).

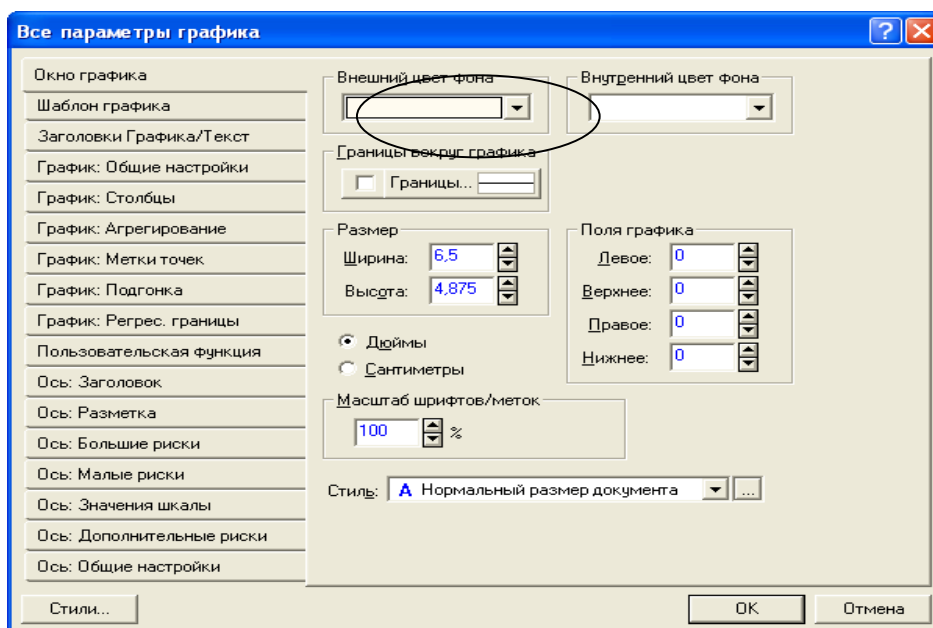


Рис. 3.9. Зміна кольору вікна графіка

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. 3.10).

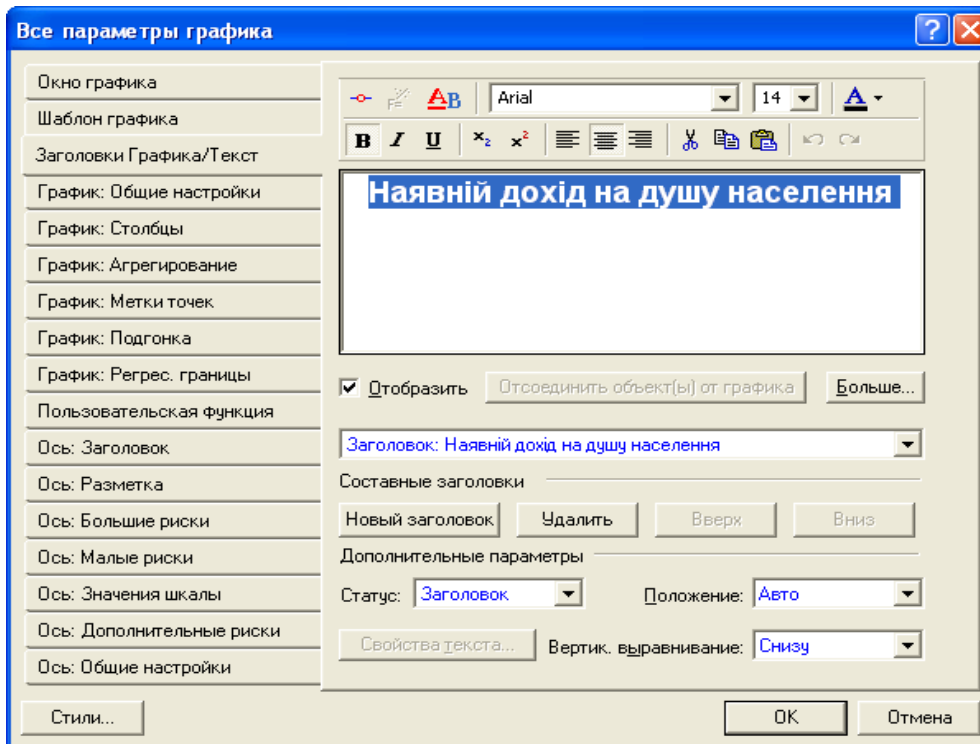


Рис. 3.10. Вікно зміни назви та формату заголовка

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. 3.11).

4. На вкладці "Axis Titles" ввести назву вісі X, колір шрифту зелений, змінити шрифт, його розмір. Аналогічно ввести назви вісі Y (вікно Axis) (рис. 3.12).

5. На вкладці "Axis Scaling" змінити діапазон вісі значень наявного доходу (вісь Y), потім натисканням кнопки "Edit step" змінити ціну розподілу цієї вісі (рис. 3.13).

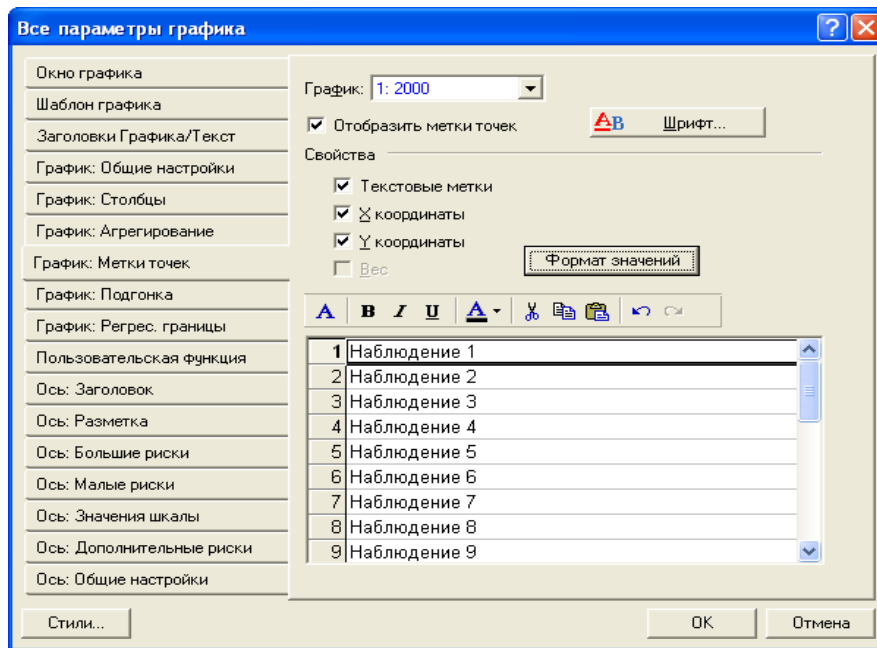


Рис. 3.11. Зміна властивостей графіка змінної

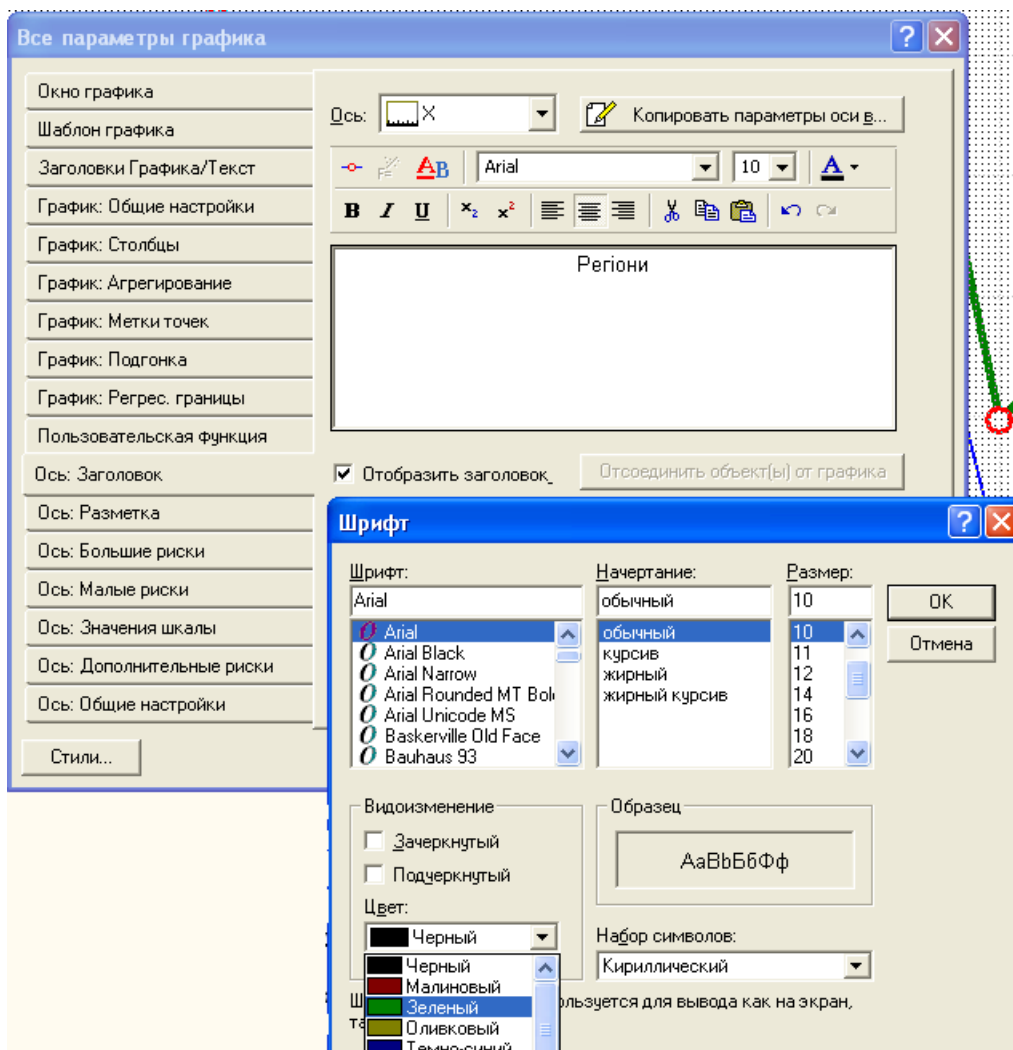


Рис. 3.12. Додавання назви осей

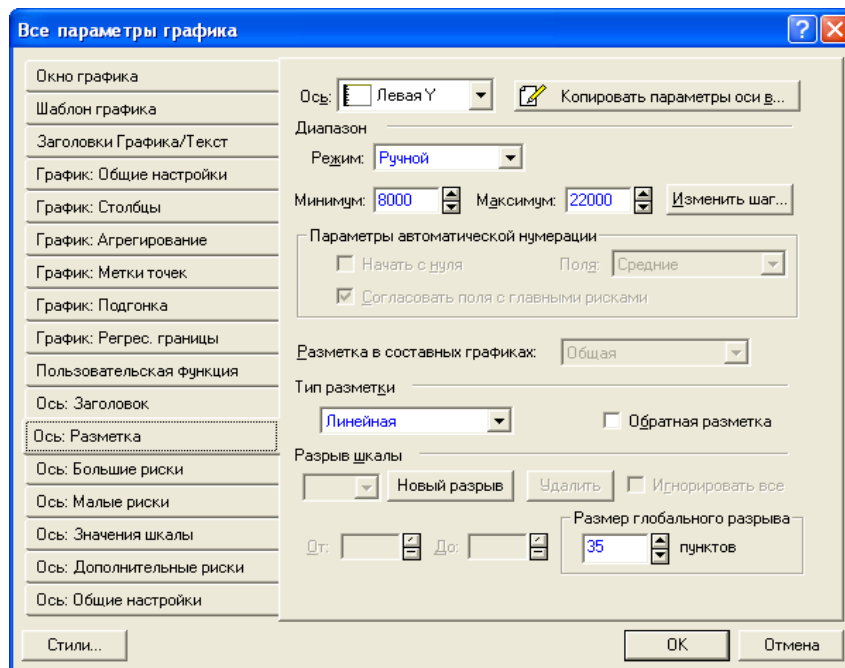


Рис. 3.13. Зміна діапазону та шагу осей

6. На вкладці "Axis Major Units" натисканням кнопки "Gridlines" змінити тип лінії сітки вісі Y і вісі X (змінити тип і колір) (рис. 3.14).

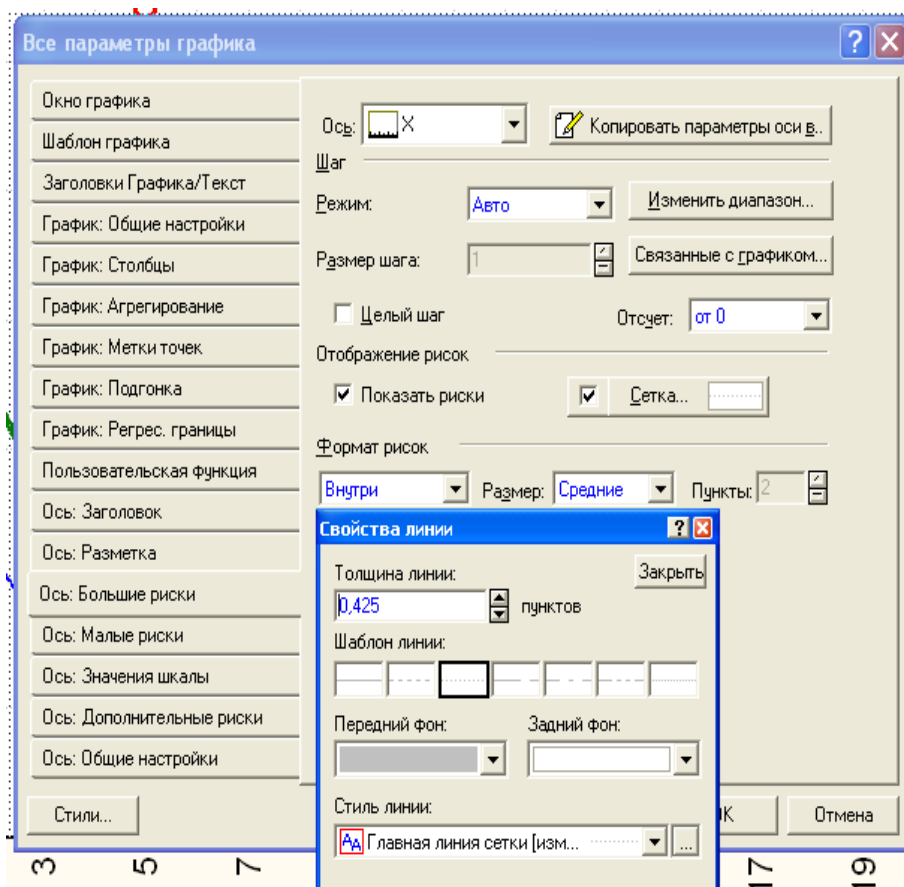


Рис. 3.14. Зміна формату основної сітки графіка

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. 3.15).

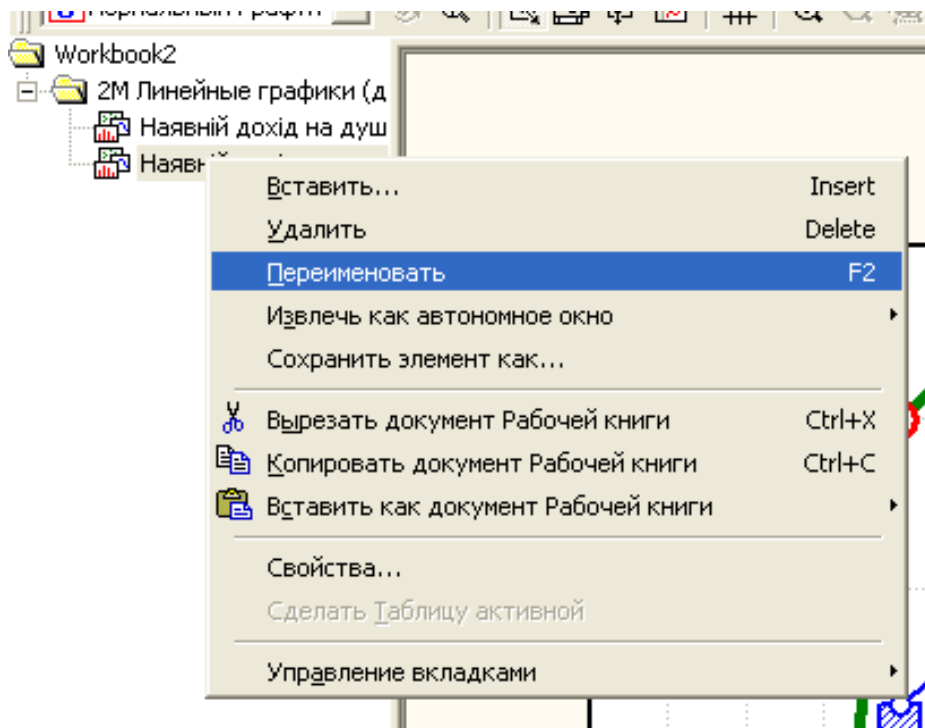


Рис. 3.15. Переименовання графіка

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

Лабораторна робота № 4 "Групування даних"

Мета – опанування навичок щодо групування даних.

Завдання – вивчити основні методи побудови дискретних та інтервальних рядів розподілу за допомогою модулів Descriptive statistics та Frequency Tables.

Методичні вказівки

На першому етапі обробки даних часто виникає необхідність у їхньому угрупованні. Угруповання дозволяє представити первинні дані в компактному вигляді, виявити закономірності варіювання досліджуваної ознаки.

Перш за все необхідно виявити вид наступного групування. В STATISTICA можливо побудувати як інтервальний, так й дискретний ряд розподілу в залежності від вибраних вхідних даних.

Кількість класів можна приблизно намітити, користуючись наступними інтервалами:

при кількості спостережень 25 – 40 – 5 – 6 класів,

при кількості спостережень 40 – 60 – 6 – 8 класів,

60 – 100 – 7 – 10,

100 – 200 спостережень – 8 – 12,

більше 200 спостережень – 10 – 15 класів.

Або розрахувати можливу кількість груп за допомогою формули Стреджеса:

$$m = 1 + \frac{2}{3} \sqrt{N}$$

де N – кількість об'єктів сукупності (спостережень).

Для побудови гістограми і таблиць частот використовується група кнопок Normality (Нормальність) вікна Descriptive statistics (Описові статистики) вкладки Basic Statistics and Tables (Основні статистики та таблиці).

У вкладці Normality знаходяться дві опції Categorization (Угрупування), що дозволяють задати число інтервалів угруповання або встановити величину інтервалу, рівну цілому числу. Число класів (інтервалів) угруповання даних устанавлюється за допомогою лічильника перемикача Number of intervals (Число інтервалів). Якщо активувати перемикач Integer intervals (categories) (Категорії), то класи (інтервали) угруповання будуть представляти цілі числа.

Наприклад, маємо дані про розподіл 20 робітників за рівнем кваліфікації (табл. 4.1).

Таблиця 4.1

Розподіл робітників за рівнем кваліфікації

№	Кваліфікація робітників	№	Кваліфікація робітників
1	5	11	4
2	6	12	5
3	4	13	3
4	2	14	4
5	3	15	3
6	3	16	4
7	2	17	5
8	6	18	4
9	5	19	5
10	4	20	1

Побудуємо дискретний ряд розподілу робітників за рівнем кваліфікації (розрядам) за допомогою опції Categorization (Угрупування) – Integer intervals (categories) (Категорії).

Перш за все необхідно сформулювати файл даних. У вкладці Normality вибрати необхідні змінні (кнопка **Variables** – Змінні). Після цього активувати Integer intervals (categories) (Категорії) та натиснути кнопку Frequency Tables (Таблиця частот) (рис. 4.1).

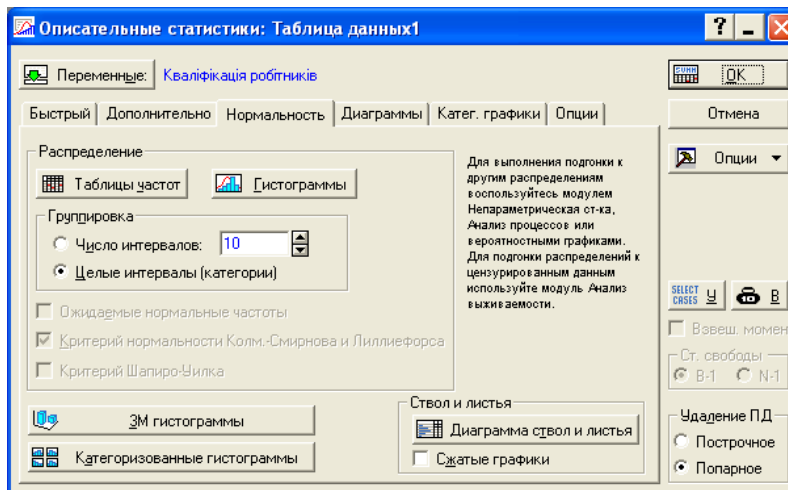


Рис. 4.1. Побудова дискретного ряду розподілу

У результаті програма формує наступне групування (табл. 4.2).

Таблиця 4.2

Угрупування робітників за рівнем кваліфікації

Таблиця частот: Кваліфікація робітників (Таблиця даних 1)						
	Частота	Кумул.	Відсоток	Кумул. %	% всіх	Кумул. %
1	1	1	5,00000	5,0000	5,00000	5,0000
2	2	3	10,00000	15,0000	10,00000	15,0000
3	4	7	20,00000	35,0000	20,00000	35,0000
4	6	13	30,00000	65,0000	30,00000	65,0000
5	5	18	25,00000	90,0000	25,00000	90,0000
6	2	20	10,00000	100,0000	10,00000	100,0000
Пропущ.	0	20	0,00000		0,00000	100,0000

З табл. 4.2 видно, що найбільша кількість робітників має 4 рівень кваліфікації.

Представимо розподіл змінних на гістограмах. Для цього призначена кнопка Histograms (Гістограми) вікна Normality (рис. 4.2).

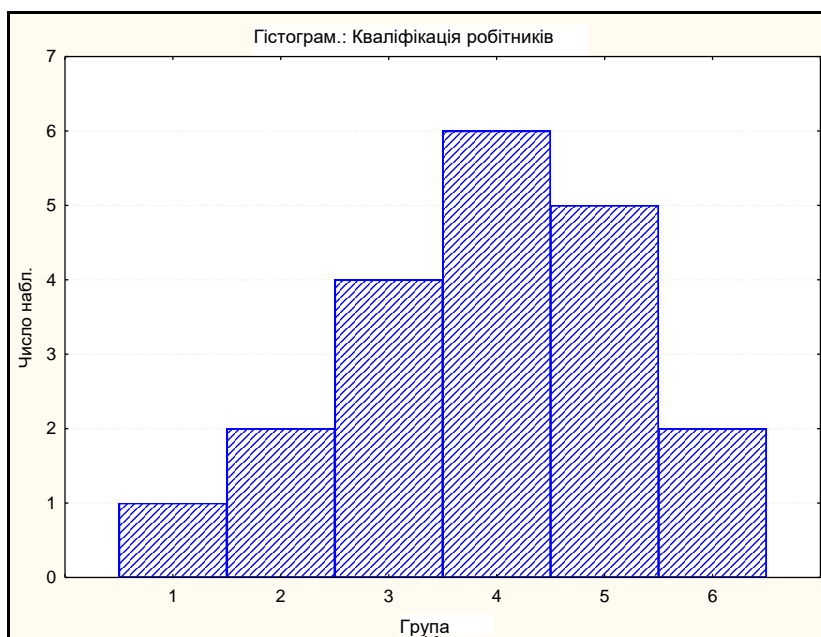


Рис. 4.2. Гістограма розподілу робітників за рівнем кваліфікації

Побудуємо інтервальний ряд розподілу підприємств за вартістю основних засобів (табл. 4.3).

Таблиця 4.3

Розподіл підприємств за вартістю основних засобів

Номер підприємства	Вартість основних засобів, тис. грн	Номер підприємства	Вартість основних засобів, тис. грн
1	480	11	1 000
2	690	12	4 073
3	1 750	13	1 014
4	3 420	14	586
5	1 146	15	905
6	1 205	16	1 746
7	1 308	17	3 140
8	364	18	529
9	2 120	19	2 861
10	1 332	20	407

У вкладці Normality вибрати необхідні змінні (кнопка **Variables** – Змінні). Після цього активувати Number of intervals (Число інтервалів). Розрахуємо необхідне число інтервалів за формулою Стерджеса:



За формулою Стерджеса розподіл повинен проводитись за 3 інтервалами. Проте модуль Описової статистики не дозволяє розбиття менше ніж на 5 інтервалів. Застосуємо саме таку кількість груп (рис. 4.3).

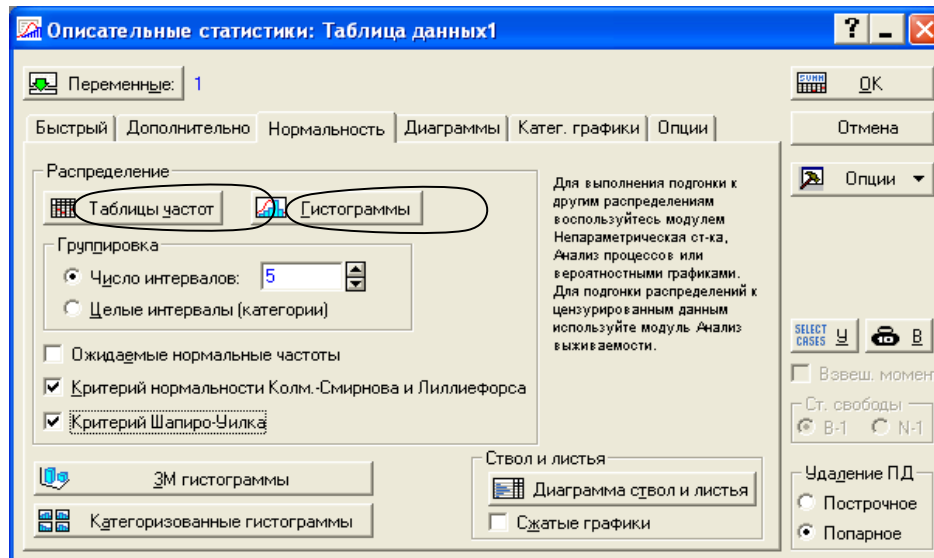


Рис. 4.3. Введення необхідної кількості інтервалів

Для побудови інтервального ряду необхідно натиснути кнопку Frequency Tables (Таблиця частот) (табл. 4.4).

Таблиця 4.4

Інтервльний ряд розподілу підприємств за вартістю основних засобів

Таблиця частот: Вартість основних засобів, тис. грн (Таблиця даних 1)						
	Частота	Кумул.	Відсоток	Кумул. %	% всіх	Кумул. %
$-1000,00 < x \leq 0,000000$	0	0	0,00000	0,0000	0,00000	0,0000
$0,000000 < x \leq 1000,000$	8	8	40,00000	40,0000	40,00000	40,0000
$1000,000 < x \leq 2000,000$	7	15	35,00000	75,0000	35,00000	75,0000
$2000,000 < x \leq 3000,000$	2	17	10,00000	85,0000	10,00000	85,0000
$3000,000 < x \leq 4000,000$	2	19	10,00000	95,0000	10,00000	95,0000
$4000,000 < x \leq 5000,000$	1	20	5,00000	100,0000	5,00000	100,0000
Пропущ.	0	20	0,00000		0,00000	100,0000

Також можливо побудувати гістограму розподілу підприємств за вартістю основних засобів. На гістограму при необхідності можна накласти щільність нормального розподілу, перевірити близькість розподілу до нормального виду за допомогою критеріїв Колмогорова – Смирнова, Лілієфорса; обчислити статистику Шапіро – Уїкса. Для цього в групі опцій необхідно встановити прапорець напроти відповідних статистик. Значення статистик показуються прямо на гістограмах (рис. 4.3).

Можливість побудови інтервального ряду саме за трьома групами представляє модуль Analysis – далі – Frequency Tables.

У вкладці Advanced (Додатково) задайте No of exact intervals: 3 (число рівних інтервалів 3) (рис. 4.4).

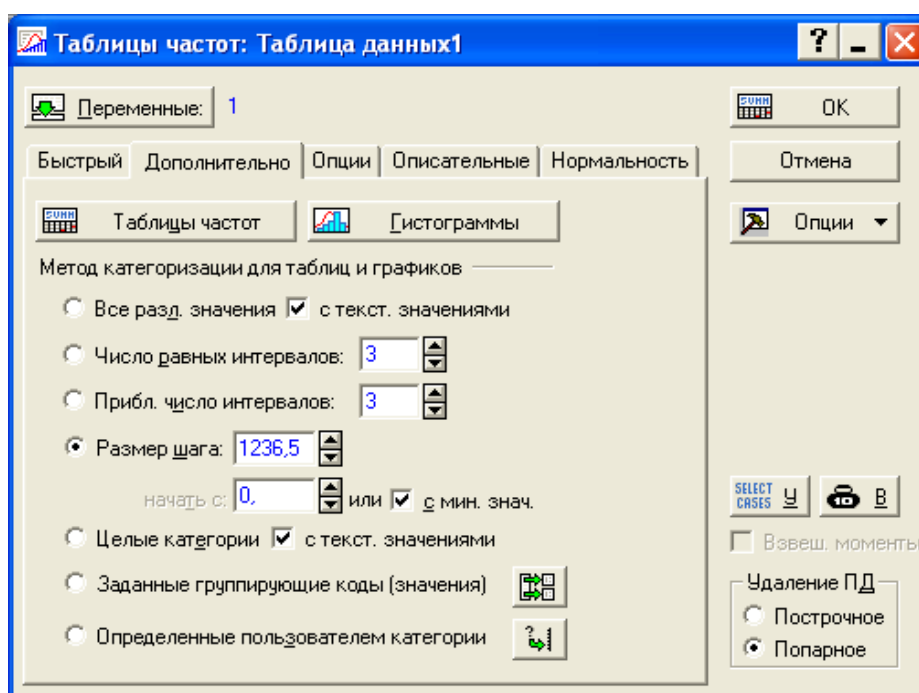


Рис. 4.4. Побудова рівних інтервальних рядів

У поле Options відзначте Cumulative frequencies (Опції – Накопичені частоти) – ОК.

Таблиця 4.5

Інтервальний ряд розподілу регіонів за вартістю основних засобів, 3 групи

Таблиця частот: Вартість основних засобів, тис. грн (Таблиця даних 1)		
	Частота	Кумул.
$-563,250 < x \leq 1291,250$	11	11
$1291,250 < x \leq 3145,750$	7	18
$3145,750 < x \leq 5000,250$	2	20

Якщо є необхідність почати з мінімального значення та встановити заданий розмір кроку, активуємо перемикач Step Size (Розмір кроку). Розрахуємо крок за допомогою формули:

$$h = \frac{B - A}{m - 3}$$

Застосуємо розмір кроку 1 236,5 (рис. 4.5). Результат побудови інтервального ряду з заданим кроком представлений в табл. 4.6.

Таблиця 4.6

Побудова інтервального ряду розподілу з вибраним розміром інтервалу

Таблиця частот: Вартість основних засобів, тис. грн (Таблиця даних 1)		
	Частота	Кумул.
364,0000 ≤ x < 1600,500	13	13
1600,500 ≤ x < 2837,000	3	16
2837,000 ≤ x < 4073,500	4	20
4073,500 ≤ x < 5310,000	0	20
Пропущ.	0	20

Різні види графіків побудови інтервального розподілу можливо застосувати в меню Graphs. Далі 2D Graphs – Histograms, переходимо на вкладку Advanced та встановіть: ім'я змінної, Graph Type: Regular, Fit Type: off (без підбора), число інтервалів групування Categories, та, наприклад, виведення відсоткового значення кожної групи за допомогою перемикача Show percentages (Проценти) (рис. 4.5). Також в цій вкладці можливо вибрати, як буде виглядати вісь Y – абсолютні значення (N), відсотки (%) або те і інше. На графік можна вивести Descriptive statistic (Описові статистики) та Total count (Число спостережень). Для побудови гістограми ініціюйте кнопку ОК.

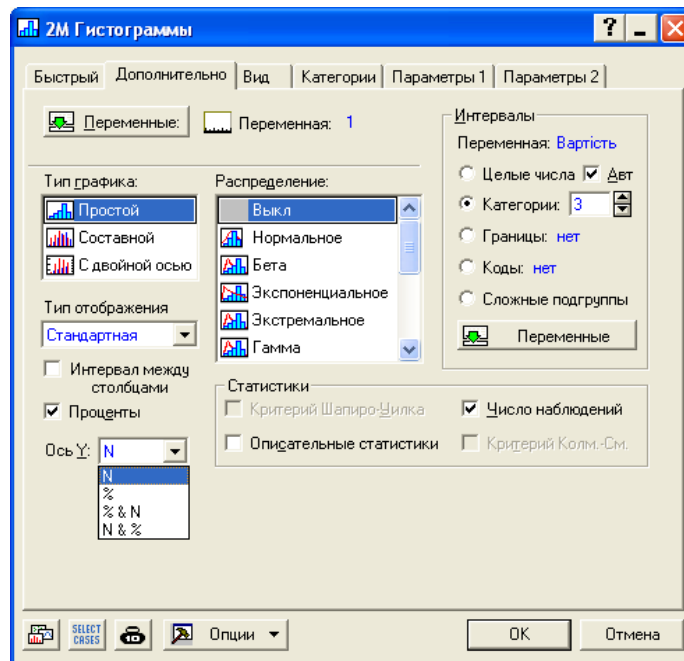


Рис. 4.5. Побудова інтервального розподілу за допомогою гістограми

Лабораторна робота № 5 "Розрахунок характеристик варіаційних рядів"

Мета – закріплення теоретичного та практичного матеріалу, придбання навичок аналізу рядів розподілу в STATISTICA.

Завдання – необхідно провести аналіз ряду розподілу в модулі Descriptive statistics ППП STATISTICA. Зробити висновки щодо варіації та форми розподілу даних.

Методичні вказівки

У статистиці існує три способи характеристики центру розподілу – мода, медіана і середнє.

Мода – найбільш проста характеристика центра розподілу. Модою для будь-якої змінної є група або клас значень, куди потрапляють більшість випадків. Якщо дві або більше груп містять однакове найвище число випадків, то змінна має "мультимодальний" розподіл.

Медіана. Якщо всі значення певної змінної розташувати в порядку зростання, то медіаною називатиметься значення, яке ділить цей впорядкований ряд навпіл. Коли у нас парна кількість випадків, то медіана розраховується як середнє арифметичне двох серединних значень.

Середнє – ця характеристика центра розподілу використовується найчастіше. Ця характеристика є просто середньою арифметичною всіх значень розподілу (тобто сума всіх значень змінної, розділена на загальну кількість випадків). Цей параметр можна назвати "центром тяжіння" розподілу.

Середнє і медіана слугують основою багатьох статистичних процедур, проте характеристики центра розподілу також корисні і самі по собі як спосіб опису даних. Проте завжди потрібно пам'ятати, що середнє є найбільш "могутньою", ефективною характеристикою центра розподілу, а найменш характерною характеристикою є мода.

У номінальних шкалах можливе використання тільки моди, оскільки для застосування інших характеристик необхідно, щоб дані були впорядковані. Кодування категорій номінальною змінною є абсолютно довільним і порядок категорій не має ніякого сенсу. Єдиний спвісіб узагальнити значення номінальної змінної – це знайти значення, що найбільш часто зустрічається (модальне).

У порядкових шкалах можна використовувати моду, проте ефективнішою характеристикою є медіана, оскільки вона враховує порядок розташування категорій. В більшості випадків при роботі з порядковими даними середнє не використовується, оскільки відстань між категоріями значень може бути різною. Проте середнє арифметичне все ж таки розраховується для деяких типів порядкових змінних, особливо для тих, в яких використовується проранжирувана шкала, наприклад, від "повністю згоден" (кодується 1) до "повністю незгодний" (кодується 5). Хоча неможливо довести, що дистанції між категоріями на цій шкалі є однаковими, середнє арифметичне використовується для того, щоб порівняти декілька об'єктів, які оцінювалися за однією шкалою.

На відміну від попередньої шкали, **інтервальна шкала** дозволяє не тільки упорядкувати прояв властивостей змінної, що вивчається, або об'єкта, але і розрахувати різницю (інтервал) між цими проявами.

З даними цього типу можна використовувати всі три характеристики центра ряду розподілу, але перевага віддається середньому арифметичному і медіані, оскільки вони більшою мірою використовують властивості даних. Найбільш могутньою характеристикою є середня арифметична, оскільки при її обчисленні беруться абсолютні значення змінних, а не тільки їх порядок. Якщо в наших даних існують екстремальні

значення, зазвичай використовується медіана. Оскільки на значення медіани не впливає ступінь віддаленості найвищих і найнижчих значень, ця характеристика є стійкою по відношенню до екстремальних значень як зверху, так і знизу. Екстремальні значення можуть негативно впливати на середнє арифметичне, перетягуючи його в тому або іншому напрямі. Коли змінна має багато дискретних значень (наприклад, отримувана зарплата), мода не буде ефективною характеристикою центра розподілу, оскільки в модальну категорію входить дуже невелика частина випадків.

Як правило, краще використовувати декілька характеристик центра розподілу в тих випадках, коли це можливо. Таким чином ми можемо не тільки досліджувати положення даних, але і, спираючись на відмінності в значеннях характеристик, зробити припущення про форму розподілу. Так, при нормальному розподілі мода, медіана і середнє арифметичне співпадають.

Інформація про доцільність використання різних характеристик центра розподілу в номінальних, порядкових, інтервальних та відносних шкалах представлена в табл. 5.1.

Характеристики центра розподілу описують тільки один аспект – місцеположення випадків. Проте в тому випадку, якщо у нас всі значення змінної не є однаковими, розподіл характеризується також ступенем розмаху, розпиленості, розосередженості значень. Також цей аспект розподілу можна використовувати для оцінки того, наскільки купчасто згруповані значення змінної.

Таблиця 5.1

**Відповідність характеристик центра розподілу
для кожного типу даних**

Шкала вимірювання	Відповідна характеристика центра розподілу		
	мода	медіана	середнє
номінальний	+		
порядковий	+	+	
інтервальний	+	+	+
відносний	+	+	+

Існує багато способів характеристики варіації, серед них: максимум, мінімум, дисперсія і середньоквадратичне відхилення.

Максимум (Maximum) – найбільше значення тієї або іншої змінної. Використовуючись окремо, ця характеристика нічого не говорить про форму розподілу.

Мінімум (Minimum) – найнижче значення в розподілі, і, як і максимум, використовується тільки у поєднанні з іншими характеристиками.

Розмах (Range) – різниця між максимальним і мінімальним значенням. Проте і ця характеристика нічого не говорить про форму розподілу і може ввести нас в оману в тому випадку, якщо у нас є хоч би одне крайнє значення.

Квартільний розмах (Quartile range) – розмах значень змінної при ігноруванні верхніх і нижніх 25 % випадків. Ця характеристика надійніша, ніж звичайний розмах, оскільки екстремальні значення виключаються і досліджуються тільки середні 50 %. З іншого боку, це є і обмеженістю цієї характеристики, оскільки, за визначенням, ми ігноруємо 50 % даних.

Дисперсія (Variance) – це перший параметр, який дозволяє оцінити загальний розмах значень. Дисперсія показує, наскільки купчасто розташовані значення змінної навколо середнього значення.

Середньоквадратичне відхилення (Standard deviation) – характеризує відхилення кожного індивідуального значення ознаки від середнього значення і тим самим визначає однорідність статистичного ряду. Ця характеристика варіації особливо важлива, оскільки вона, у поєднанні з середнім значенням, використовується в багатьох статистичних процедурах.

Ексцес (Kurtosis) – ступінь гостроти піку розподілу випадкової величини. Коефіцієнт ексцесу нормального розподілу дорівнює нулю. Він позитивний, якщо пік розподілу біля середньої гострий, і від'ємний, якщо пік гладкий.

Асиметрія (Skewness) – параметр, який характеризує асиметрію розподілу випадкової змінної. Якщо розподіл симетричний відносно середньої, то коефіцієнт асиметрії дорівнює нулю. Якщо він більше нуля, то в розподілі існує правостороння асиметрія, та, навпаки, якщо він менше нуля, то існує лівостороння асиметрія.

Як і характеристики центра розподілу, вибір параметрів варіації, для застосування в тому або іншому конкретному випадку залежить від рівня вимірювання досліджуваною змінною. Параметри варіації також різняться за їх потужністю (або корисністю). Дисперсія і середньоквадратичне відхилення вважаються найбільш могутніми, оскільки вони використовують більше інформації про дані, ніж інші параметри. У табл. 5.2 представлено узагальнення цієї інформації.

Відповідні характеристики варіації для кожного типу даних

Шкала вимірювання	Відповідні характеристики варіації					
	Мінімум	Максимум	Розмах	Квартільний розмах	Дисперсія	Середньоквадратичне відхилення
Номінальний						
Порядковий	+	+	(+)	+		
Інтервальний	+	+	+	+	+	+
Відносний	+	+	+		+	+

Номінальні дані – неможливо оцінити ступінь розмаху, оскільки відсутній порядок категорій.

Порядкові дані – можливе використання мінімуму, максимуму, розмаху і квартильного розмаху. Проте кількість категорій в порядковій змінній зазвичай дуже невелика, і ці параметри дають недостатньо інформації.

Інтервальні і відносні дані – можливе використання всіх параметрів варіації. Найбільш важливим є дисперсія і середньоквадратичне відхилення, оскільки вони використовують більше інформації про варіацію даних. Можна використовувати і інші характеристики, хоча при використанні мінімуму, максимуму і розмаху екстремальні значення з обох кінців розподілу впливають на ряд розподілу. Хоча, з іншого боку, екстремальні значення можуть серйозно впливати і на дисперсію і середньоквадратичне відхилення, особливо коли ми маємо справу з невеликими масивами даних.

Розглянемо розрахунок показників ряду розподілу на прикладі динаміки абсолютного приросту валового внутрішнього продукту України.

Відкриваємо файл з даними.

Вибираємо команду меню Statistics (Статистики) Basic Statistics/Tables (Базові статистики та таблиці) (рис. 5.1) Descriptive Statistics (Дескриптивні статистики).

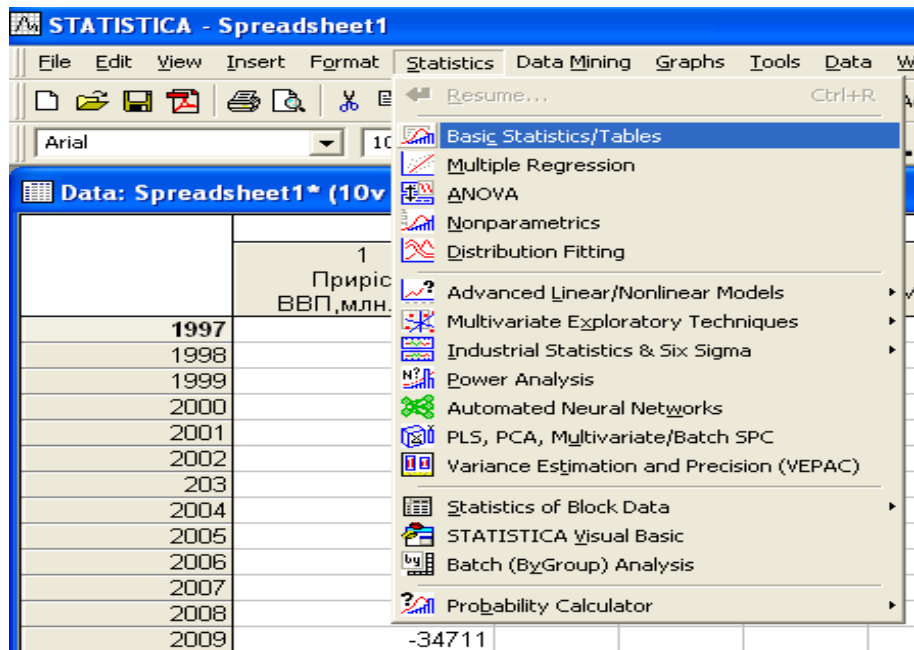


Рис. 5.1. Меню **Statistics (Статистики) Basic Statistics/Tables (Базові статистики та таблиці)**

Відкриється діалогове вікно Descriptive Statistics (рис. 5.2).

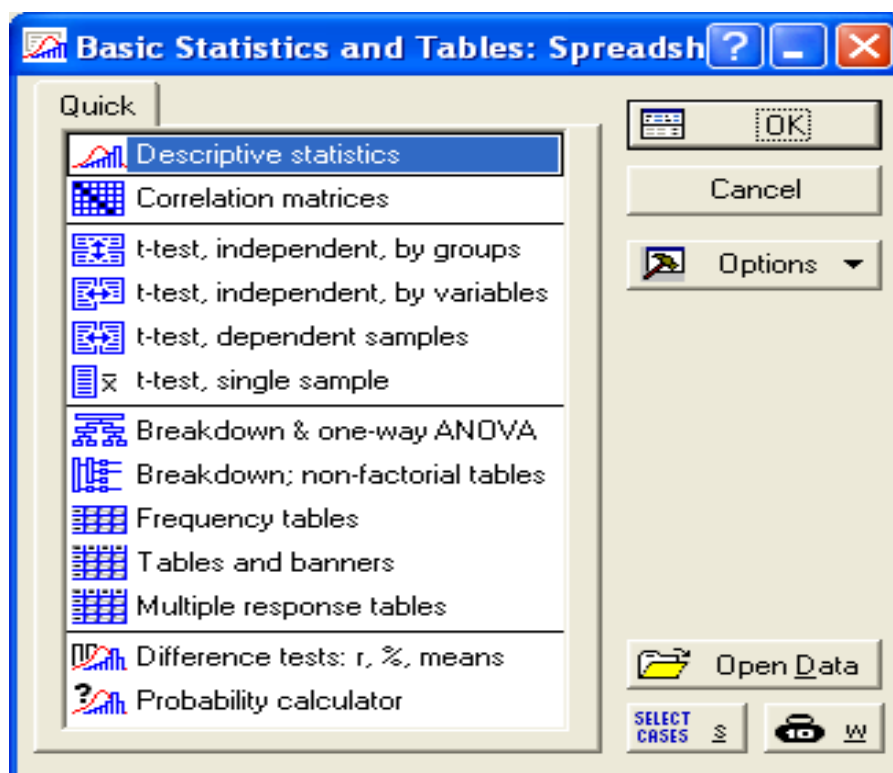


Рис. 5.2. Діалогове вікно: **Basic Statistics/Tables (Базові статистики та таблиці)**

Переносимо змінну в список змінних для аналізу, ініціюючи кнопку Variables (змінні) та обираючи змінну, і натискаємо кнопку ОК (рис. 5.3).

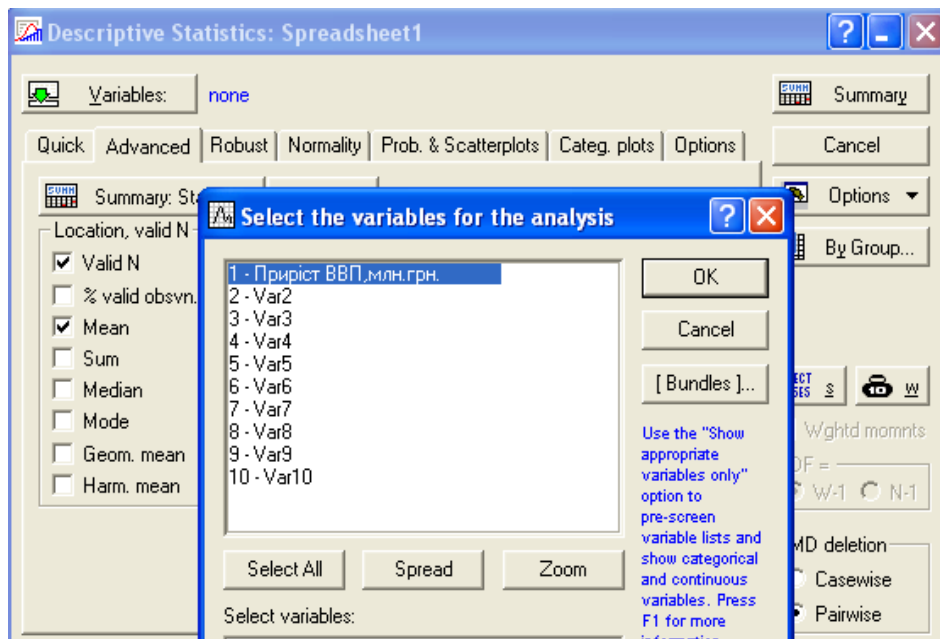


Рис. 5.3. **Діалогове вікно: Variables (змінні)**

Після цього маємо можливість задати перелік необхідних для аналізу характеристик (рис. 5.4).

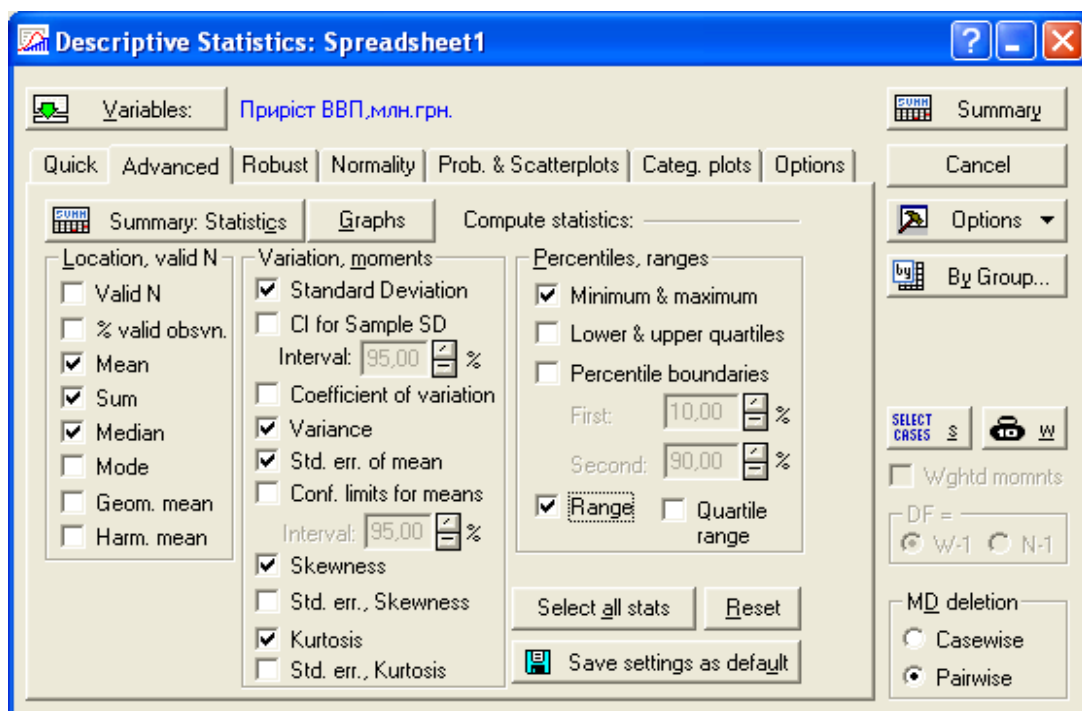


Рис. 5.4. **Діалогове вікно: Descriptive Statistics (Дескриптивні статистики)**

Тут можна задати обчислення наступних статистичних характеристик: середнього значення; суми; стандартного відхилення; стандартної помилки; дисперсії; мінімуму; максимуму; розмаху; ексцесу; асиметрії.

Натискаємо на кнопці Summary (Результат). Результат буде показаний у вікні перегляду:

Descriptive Statistics (Spreadsheet1)											
Variable	Mean	Median	Sum	Minimum	Maximum	Range	Variance	Std.Dev.	Standard Error	Skewness	Kurtosis
Приріст ВВП,млн.грн.	63986,62	39628,00	831826,0	-34711,0	227325,0	262036,0	5,220032E+09	72249,79	20038,49	1,158524	1,099713

Рис. 5.5. Результат обчислення Descriptive Statistics (Дескриптивні статистики)

З рис. 5.5 видно, що мінімальний рівень приросту ВВП – -34 711, а максимальний – 22 7325. Середній рівень ВВП України складав 63 986,62 млн. грн. В цілому усі значення відхиляються від середнього в той або інший бік на 72 249,79 млн грн. Це свідчить про значну варіацію серед отриманих оцінок. Коефіцієнт асиметрії дорівнює – 1,158, що говорить про правосторонню асиметрію. Коефіцієнт ексцесу свідчить про наявність піків розподілу.

Щоб наочно побачити форму розподілу, необхідно побудувати гістограму з накладанням на неї кривої нормального розподілу, ініціювавши вкладку Normality, кнопку Histograms (рис. 5.6, 5.7).

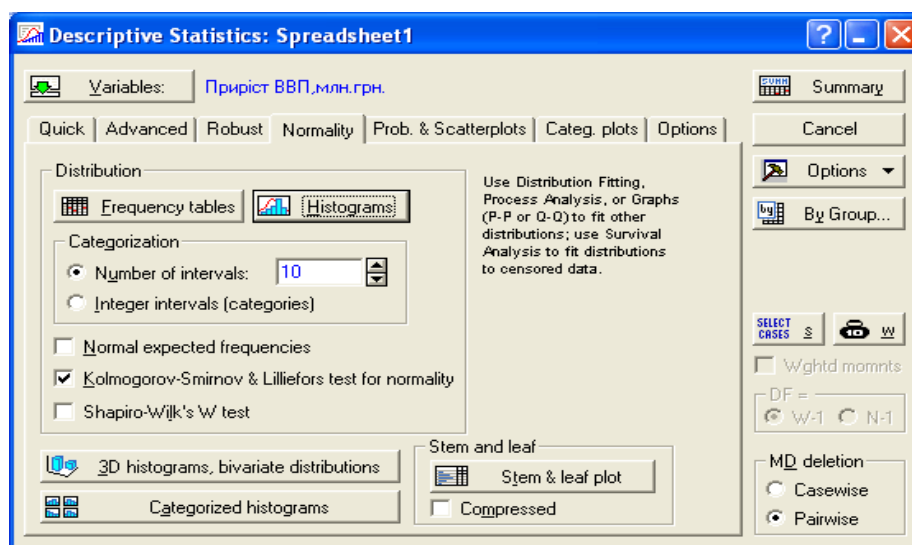


Рис. 5.6. Діалогове вікно: Descriptive Statistics (Дескриптивні статистики) вкладка Normality

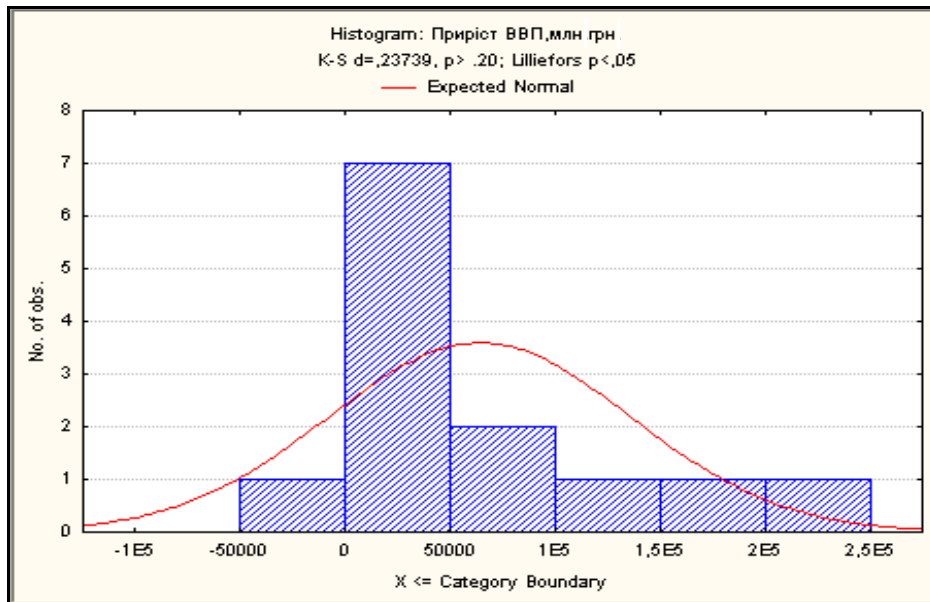


Рис. 5.7. Гістограма з накладанням на неї кривої закону нормального розподілу

Лабораторна робота № 6 "Дисперсійний аналіз"

Мета – придбання студентами навичок роботи щодо виявлення впливу на статистичну ознаку за допомогою модуля ANOVA в пакеті STATISTICA.

Завдання – вивчити основні методи проведення дисперсійного аналізу.

Методичні вказівки

Дисперсійний аналіз використовується в статистичних дослідженнях з метою виявлення впливу на статистичну ознаку, що вивчається та, яка не має кількісного виразу. Суть дисперсійного аналізу полягає в розкладанні загальної варіації результуючого показника на компоненти, які відображають вплив чинника (чинників), що вивчається (що вивчаються), і неврахованих чинників.

Залежно від кількості досліджуваних чинників моделі дисперсійного аналізу прийнято підрозділяти на одно-, дво- та багатофакторні.

За допомогою одно факторного дисперсійного аналізу досліджується вплив одного чинника. Дані для проведення аналізу отримують в процесі проведення спеціальним чином організованих досліджень. Початковими даними є значення результуючої ознаки, зафіксовані при m

рівнях чинника. Кількість спостережень при кожному рівні чинника дорівнює n .

За допомогою багатofакторного дисперсійного аналізу досліджується вплив на результуючий показник двох або більше чинників. Чинники встановлюють, проводячи процедуру багатofакторного дисперсійного аналізу, при цьому варіація результуючої ознаки розкладається на компоненти, що відображають вплив на неї досліджуваних і неврахованих чинників. Початкові дані можна фіксувати на m рівнях чинника А і на n рівнях чинника В. Загальна кількість груп спостережень дорівнює mn . Для спрощення в кожній групі фіксується по одному спостереженню, тому загальне число спостережень буде дорівнювати $N = mn$.

При проведенні однофакторного дисперсійного аналізу перевіряється статистична гіпотеза:

$$H_0: a_1 = a_2 = \dots = a_m \text{ або } H_1: a_1 \neq a_2 \neq \dots \neq a_m,$$

де a_1, a_2, \dots, a_m – генеральні середні значення результуючої ознаки, зафіксовані при першому, другому, ..., m -ому рівні фактора.

Для проведення дисперсійного аналізу використовується модуль ANOVA ППП Statistica. Вихідні дані для проведення аналізу представлені в табл. 6.1.

Таблиця 6.1

Початкові дані для проведення дисперсійного аналізу

Район	Обсяг реалізації продукції, у.о.			
	Пн.	Вт.	Ср.	Чт.
"А"	14	15	14,8	15
"Б"	14,4	14,9	14,9	15,5
"В"	14,2	15,2	14,6	15,4
"Г"	14,5	15	14,7	15,2

Даними є значення обсягу реалізованої продукції, отриманої в чотирьох торгових місцях та розташованих в чотирьох районах міста за чотири дні тижня. Умовно райони міста назвемо "А", "Б", "В", "Г". Перевіряється вплив розташування торгового місця на обсяг реалізованої продукції.

Створимо файл розмірності 3x16. Перша змінна файла буде призначена для фіксації району розташування торгового місяця, друга – для фіксації дня тижня (знадобиться надалі для проведення двофакторного дисперсійного аналізу), третя змінна – для фіксації обсягу реалізованої продукції (рис. 6.1).

У пункті меню "Статистика" виберіть пункт ANOVA (від analyze of variance), який призначений для проведення процедури як однофакторного, так і багатофакторного ТАК. У результаті буде відкрито стартове вікно модуля (рис. 6.2).

	1 Розміщення торгового місця	2 День тижня	3 Обсяг РП
1	А	Пн.	14
2	А	Вт.	15
3	А	Ср.	14,8
4	А	Чт.	15
5	Б	Пн.	14,4
6	Б	Вт.	14,9
7	Б	Ср.	14,9
8	Б	Чт.	15,5
9	В	Пн.	14,2
10	В	Вт.	15,2
11	В	Ср.	14,6
12	В	Чт.	15,4
13	Г	Пн.	14,5
14	Г	Вт.	15
15	Г	Ср.	14,7
16	Г	Чт.	15,2

Рис. 6.1. Початкові дані для проведення дисперсійного аналізу

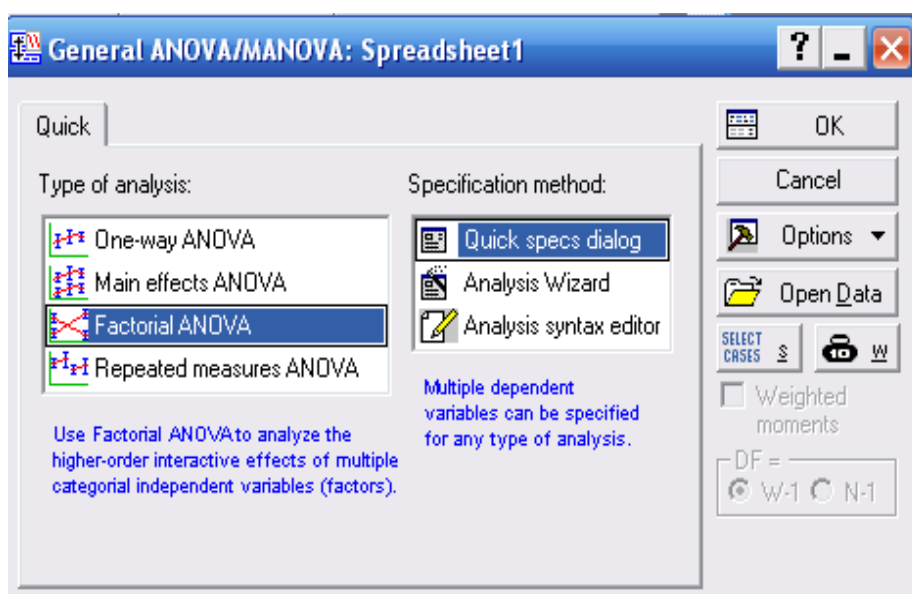


Рис. 6.2. Стартове вікно модуля ANOVA

Спочатку проведемо процедуру однофакторного дисперсійного аналізу. Для цього встановимо тип аналізу One-way ANOVA і метод Quick specs dialog. Відкриється вікно установки змінних та опцій аналізу (рис. 6.3).

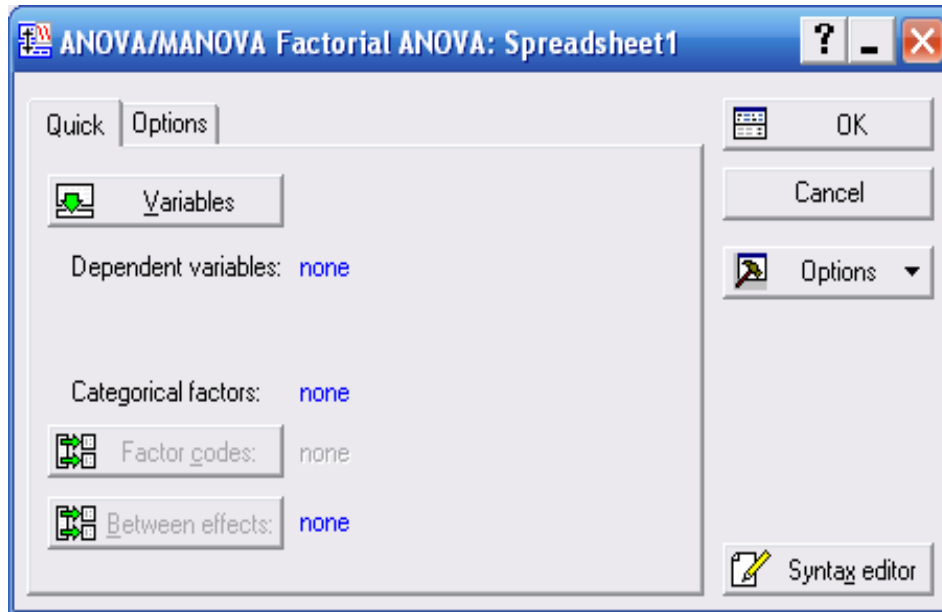


Рис. 6.3. Вікно установки змінних та опцій аналізу

У діалоговому вікні виберемо факторну змінну "Розміщення торгового місця", а результуюча ознака – "Обсяг реалізованої продукції" (рис. 6.4).

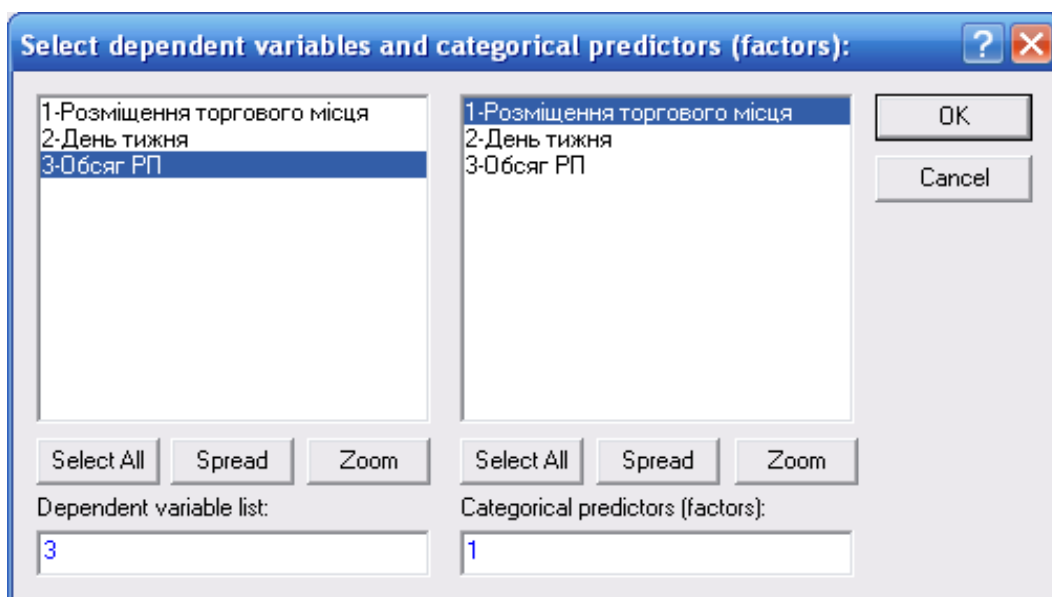


Рис. 6.4. Вікно вибору змінних для аналізу

Після вибору досліджуваних ознак відкривається вікно результатів аналізу (рис. 6.5).

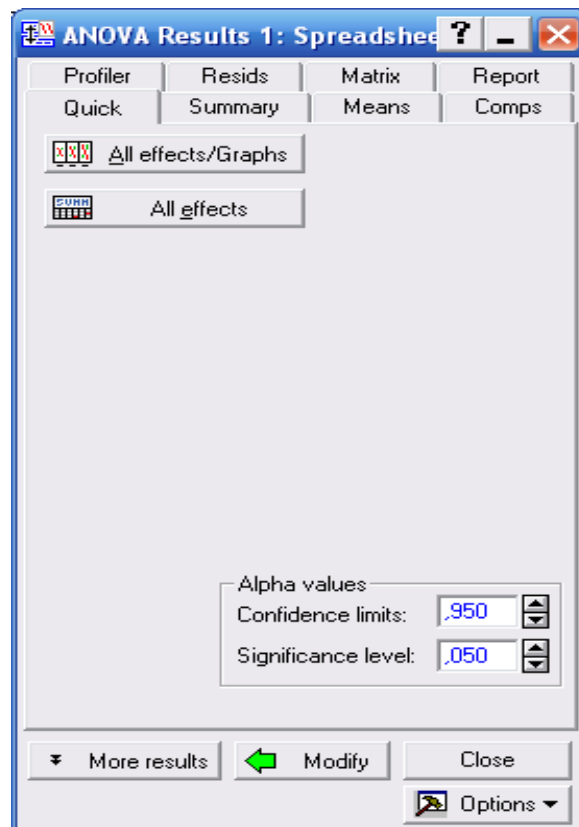


Рис. 6.5. Вікно результатів аналізу

Залишаємо рівень значущості "Significance level" рівним $\alpha=0,05$ для розрахунку компонент варіації і перевірки гіпотези з використанням статистики Фішера (рис. 6.6).

Univariate Tests of Significance for ОбсягПП (Spreadsheet1)					
Sigma-restricted parameterization					
Effective hypothesis decomposition					
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	3519,456	1	3519,456	16978,28	0,000000
Розміщення торгового місця	0,107	3	0,036	0,17	0,913350
Невраховані фактори	2,488	12	0,207		

Рис. 6.6. Вікно результатів однофакторного дисперсійного аналізу

1. Система розрахувала компоненти "Сума квадратів" (SS) $Q_1^2 = 0,107$ та $Q_2^2 = 2,488$ загальної варіації Q^2 , які обумовлені чинником "Розташування торгових місць" та неврахованих чинників.

2. Визначила число мір свободи (Degr. of Freedom) кожної компоненти. Зокрема, для компоненти варіації, обумовленою факторною змінною, число мір свободи дорівнює $m-1=3$, а для компоненти, обумовленою неврахованими чинниками, число мір свободи дорівнює $m(n-1)=12$.

3. Розрахувала величини "Середній квадрат" (MS) $Q_1^2 / (m-1) = 0,0357$ та $Q_2^2 / (m(n-1)) = 0,2073$ відповідно.

4. Розрахувала величину статистики Фішера, $F = \frac{Q_2^2 / (m(n-1))}{Q_1^2 / (m-1)} = 5,81$ та визначила вірогідність $p=0,913350$, при правдивій гіпотезі H_0 , величина F може прийняти такі ж або більші значення. Оскільки ця вірогідність істотно більше рівня значущості $\alpha=0,05$, то гіпотезу про значущість впливу чинника "Розташування торгових місць" на обсяг реалізованої продукції можна не приймати.

Проведемо аналіз багатофакторного дисперсійного аналізу (на прикладі двофакторного).

У стартовому вікні модуля ANOVA встановимо тип аналізу Main effects ANOVA. У відкритому діалоговому вікні виберемо результуючу ознаку і змінні чинники (рис. 6.7).

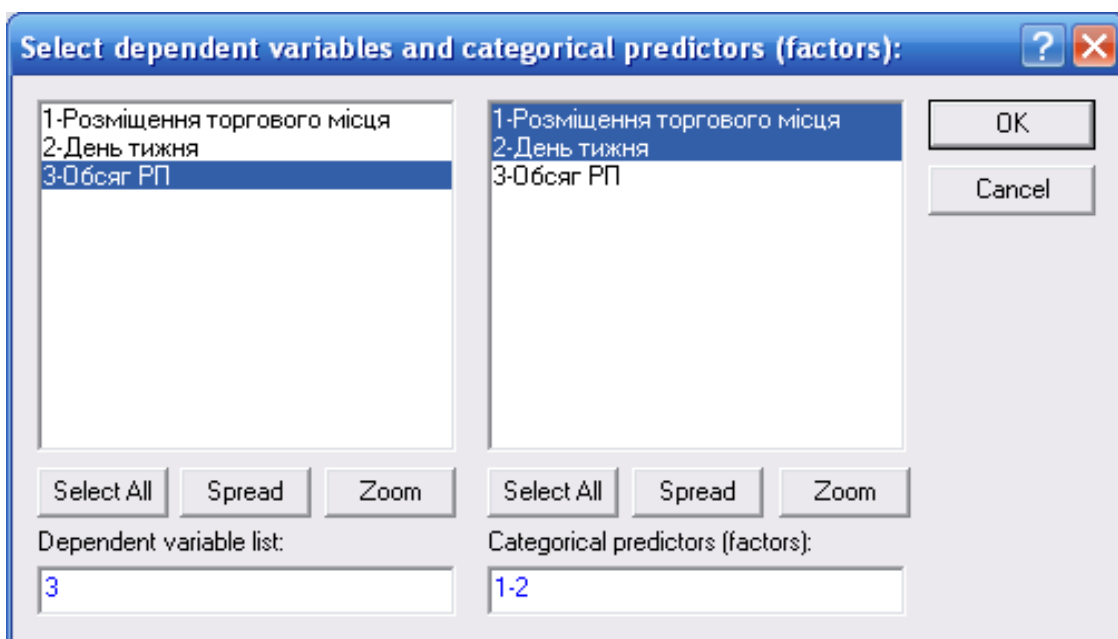


Рис. 6.7. Вікно результатів двофакторного дисперсійного аналізу

На рис. 6.8. представлені установки рівнів змінних чинників.

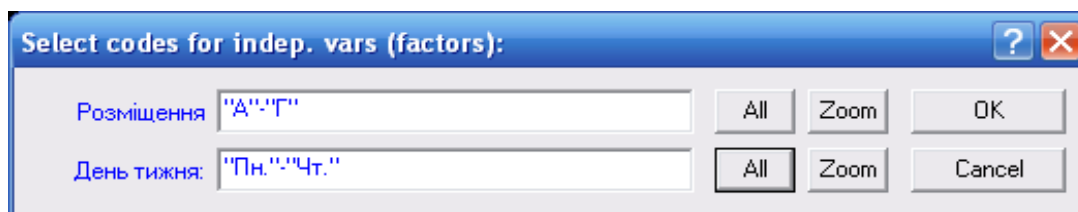


Рис. 6.8. Вікно установки рівнів змінних чинників

У відкритшомуся вікні результат аналізу залишаємо рівень значущості рівним $\alpha=0,05$, після чого відкривається вікно із результатами розрахунку компонентів варіації та перевірки гіпотез з використанням статистик Фішера (рис. 6.9).

Effect	SS	Degr. of Freedom	MS	F	p
Intercept	3519,456	1	3519,456	110897,5	0,000000
Розміщення торгового місця	0,107	3	0,036	1,1	0,390317
День тижня	2,202	3	0,734	23,1	0,000145
Невраховані чинники	0,286	9	0,032		

Рис. 6.9. Вікно результатів двофакторного дисперсійного аналізу

1. Система розрахувала компоненти "Сума квадратів" $Q_1^2=2,202$ та $Q_2^2=0,286$ загальній варіації Q^2 обумовлені вибраними і неврахованими чинниками.

2. Визначила число мір свободи кожної компоненти. Зокрема для компоненти варіації, обумовленою змінною "Розташування торгових місць", число мір свободи дорівнює $m-1=3$, для компоненти, обумовленою змінною "День тижня", число мір свободи дорівнює $n-1=3$, а для компоненти, обумовленою неврахованими чинниками, число мір свободи дорівнює $(m-1)(n-1)=9$.

3. Розрахувала величини "Середній квадрат $\overline{Q_A} = 0,07$ та $\overline{Q_B} = 0,000145$ відповідно.

4. Розрахувала значення статистик Фішера

$F_A = \frac{\overline{Q_A}}{\overline{Q_B}} = 1$; $F_B = \frac{\overline{Q_B}}{\overline{Q_A}} = 2$ та визначила

відповідні значенням F_A і F_B при вірогідності $p_A=0,390317$ і $p_B=0,000145$ того, що при справедливості гіпотез H_0 величини F_A і F_B можуть прийняти такі ж або більші значення.

Оскільки вірогідність p_A істотно більше, а вірогідність p_B істотно менше рівня значущості $\alpha=0,05$, то гіпотезу про значущість впливу чинника "Розташування торгових місць" на обсяги реалізованої продукції можна не приймати, а гіпотезу про значущість впливу чинника "День тижня" на обсяги реалізованої продукції необхідно прийняти.

Лабораторна робота № 7 "Вибірковий метод"

Мета – закріплення теоретичного та практичного матеріалу за темою вибіркоче спостереження, проведення відбору даних в модулі Data / Subset.

Завдання – необхідно провести вибірку даних в модулі Data / Subset ППП STATISTICA та дати економічну інтерпретацію отриманим результатам.

Методичні вказівки

Відбір даних – це вибір спостережень за певними критеріями. Для цього в STATISTICA існує дві принципові можливості:

- 1) вибір спостережень за певною умовою (логічному виразу);
- 2) відбір випадкової вибірки спостережень з файла даних.

Вибір спостережень за певною умовою

Проведемо частотний аналіз змінної "Наявний дохід на душу населення". Вчинемо наступним чином.

Завантажуємо файл в редактор даних.

Вибираємо в меню команди Дані (Data) Підмножина/Випадковий вибір (Subset /Random Sampling) (рис. 7.1).

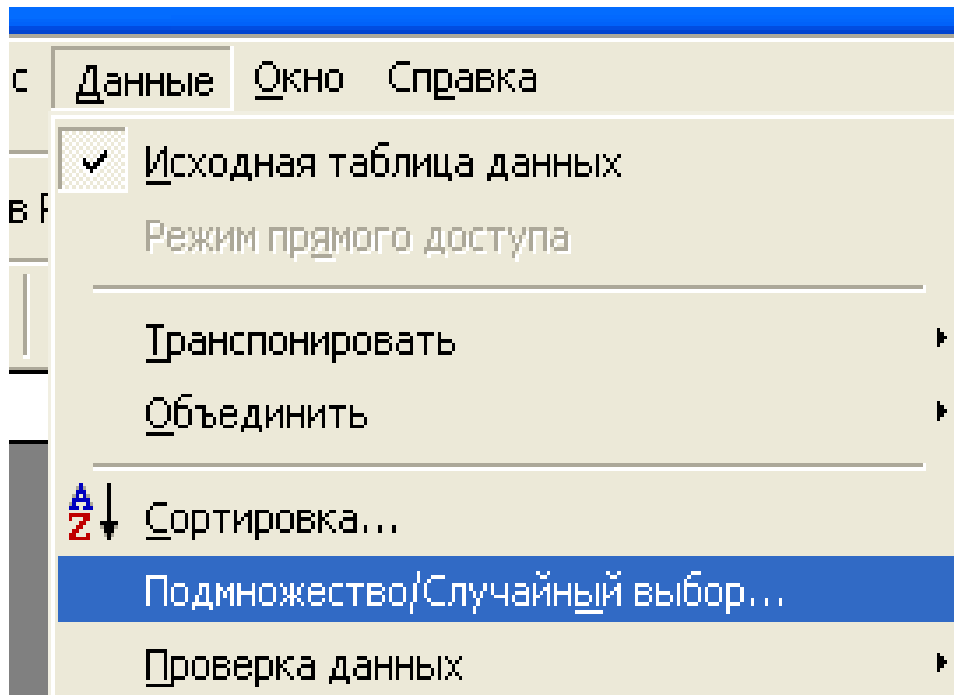


Рис. 7.1. Меню Дані (Data)

Відкриється діалогове вікно Вибір умов (Select Cases) (рис. 7.2). За замовчуванням в цьому вікні діалогу вибраний пункт All cases (Всі спостереження, тобто в розрахунку приймають участь всі спостереження вихідного ряду).

Далі вибираємо пункт Завдати умови вибору – Enable selection conditions (Якщо виконується умова) і клацаємо на кнопці Задані – Specific, Selected by.

Це діалогове вікно розділене на наступні частини:

- включити спостереження в Аналіз/Графік;
- виключити спостереження з Аналізу/Графіка.

Кожна частина містить строку умови або номера спостереження. Під умовою розуміється функція, яка віддзеркалює необхідні значення для вибору.

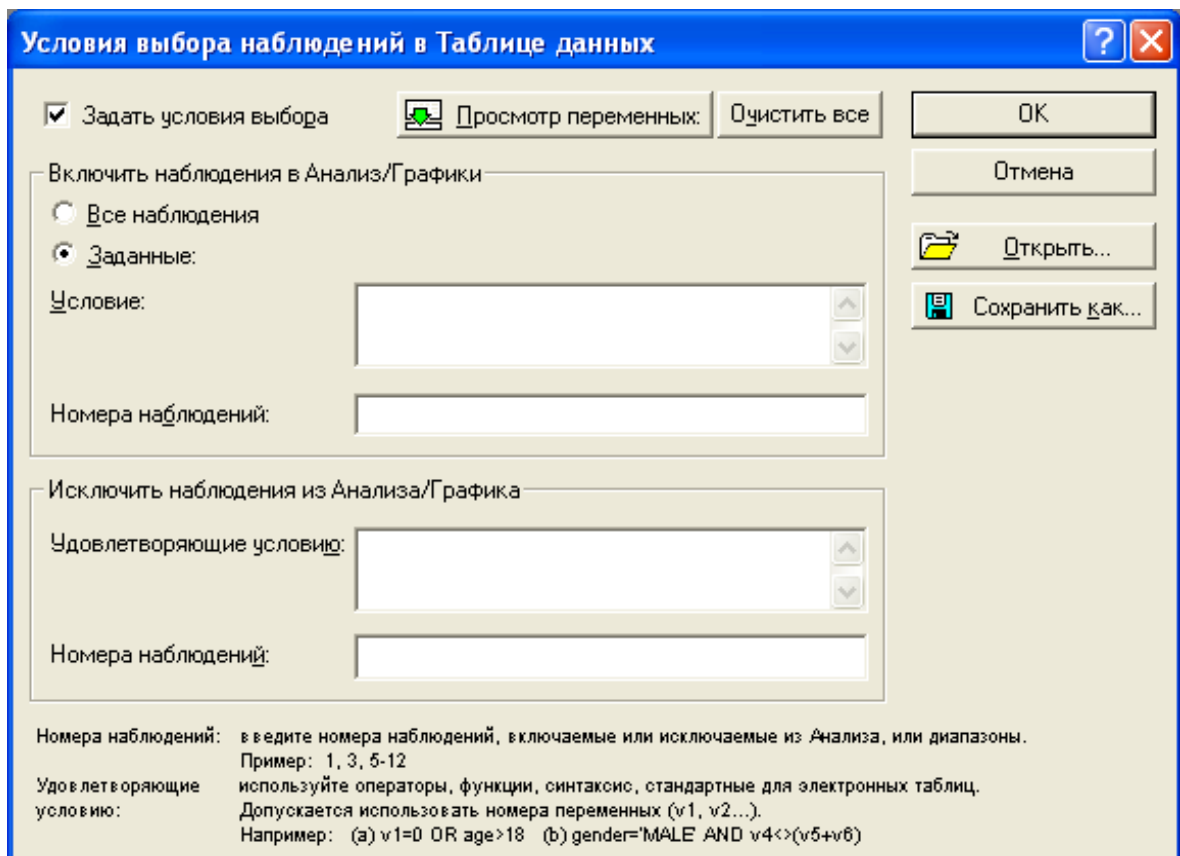


Рис. 7.2. Диалоговое окно Вибір спостережень (Select Cases)

Наприклад, зробимо вибірку з умовою включення до неї регіонів з наявним доходом на душу населення більше ніж 1 500 грн (рис. 7.3).

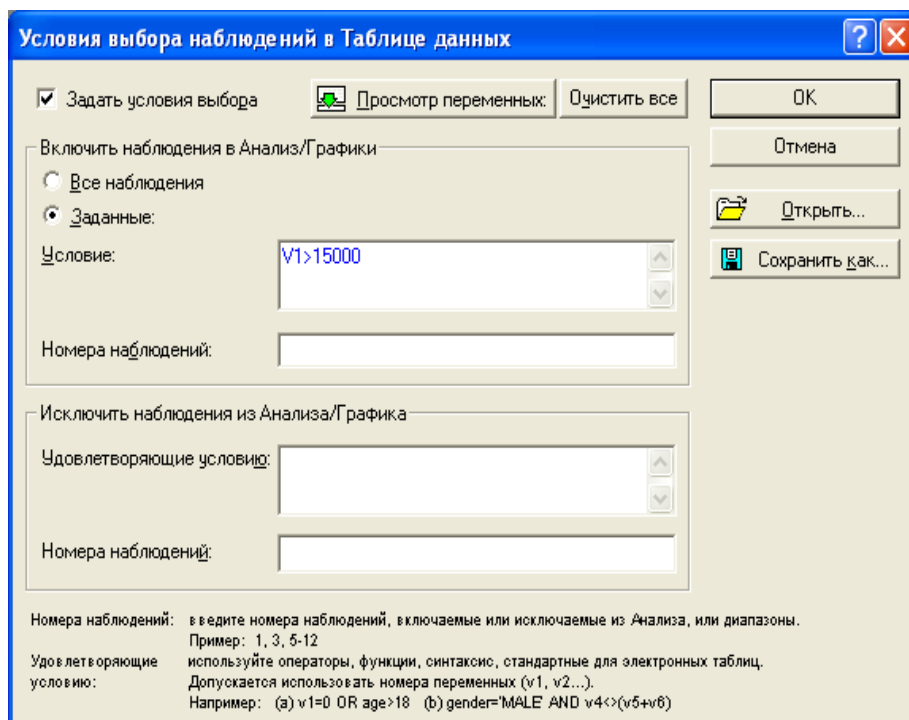


Рис. 7.3. Введення умови включення спостережень у вибірку

Програма STATISTICA також дозволяє формувати випадковий відбір об'єктів вибірки. Для цього в головному вікні меню Дані (Data) Підмножина/Випадковий вибір (Subset /Random Sampling), на вкладці Simple Sampling (Простой выбор) Активуємо перемикач Simple random sampling (Проста випадкова вибірка) (рис. 7.4).

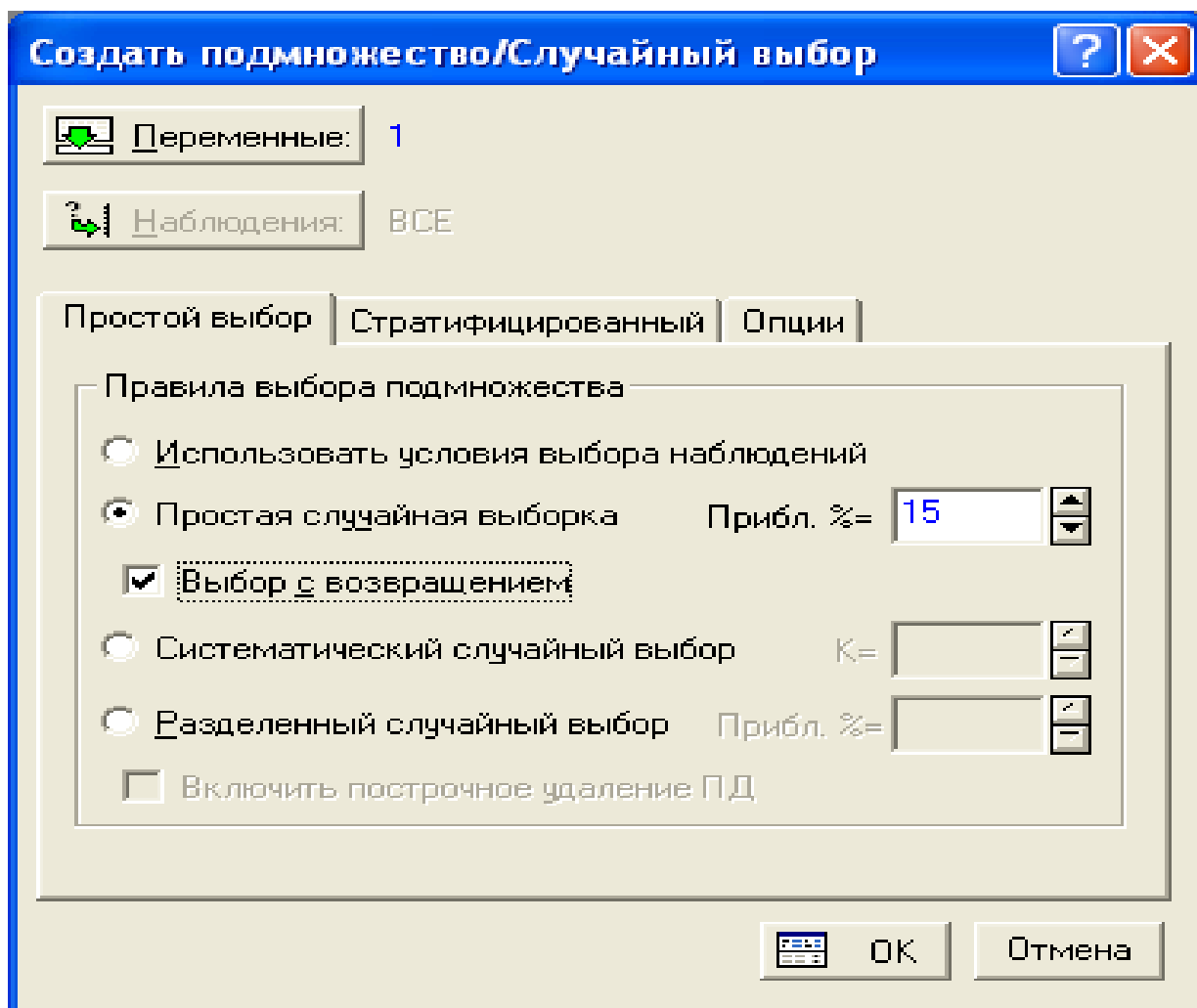


Рис. 7.4. Проста повторна випадкова вибірка

Прапорець "Вибір з поверненням" вмикається при необхідності повторної вибірки, тобто в даному випадку при виборі дані будуть повернені в сукупність, як наслідок можуть бути вибрані повторно.

Приблизний відсоток вибірки (Approximate %) – це очікувана доля об'єктів вибіркової сукупності в загальній, генеральній сукупності.

Наступним видом випадкового відбору є систематична вибірка, яка враховує відбір за введеною кількістю разів (K) (рис. 7.5).

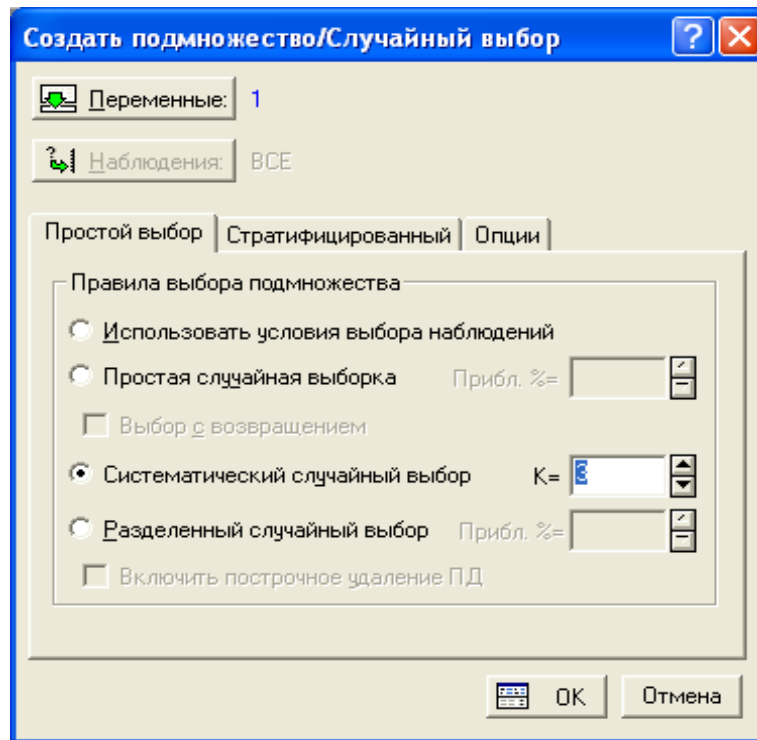


Рис. 7.5. Систематичний випадковий відбір

Останній вид вибірки – це розділений випадковий відбір. Він, на відміну від простої випадкової вибірки, в результаті формує два файли, в одному з яких містиця вибіркова сукупність, в іншому об'єкти, що не ввійшли до вибірки (рис. 7.6).

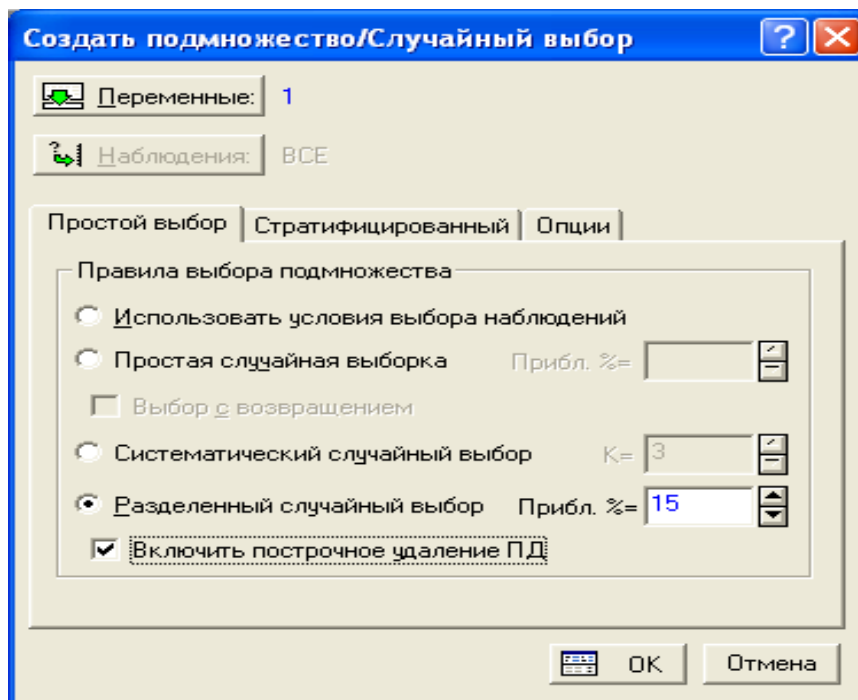


Рис. 7.6. Розділений випадковий відбір

Рекомендована література

Халафян А. А. STATISTICA 6. Статистический анализ данных / А. А. Халафян. – М. : Бином-Пресс, 2007. – 508 с.

Боровиков В. Программа STATISTICA для студентов и инженеров / В. Боровиков. – М. : КомпьютерПресс, 2001. – 304 с.

Вуколов Э. А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL / Э. А. Вуколов. – М. : ФОРУМ, 2008. – 464 с.

Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров ; под ред. В. Э. Фигурнова. – 3-е изд. – М. : ИНФРА-М, 2003. – 544 с.

НАВЧАЛЬНЕ ВИДАННЯ

**Лабораторний практикум
з модуля 1 "Методи описової статистики
та статистичний аналіз рядів розподілу"
навчальної дисципліни
"СТАТИСТИКА"
для студентів напряму підготовки
6.030506 "Прикладна статистика"
денної форми навчання**

Укладачі: **Раєвська** Олена Валентинівна
Горохова Ольга Іванівна
Чанкіна Ірина Володимирівна та ін.

Відповідальний за випуск **Раєвська О. В.**

Редактор **Бутенко В. О.**

Коректор **Мартовицька-Максимова В. А.**

План 2012 р. Поз. № 411.

Підп. до друку

Формат 60x90 1/16. Папір MultiCopy. Друк Riso.

Ум.-друк. арк. 3,75. Обл.-вид. арк. 4,69. Тираж

прим. Зам. №

Видавець і виготівник – видавництво ХНЕУ, 61166, м. Харків, пр. Леніна, 9а

*Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи
Дк № 481 від 13.06.2001 р.*

**Лабораторний практикум
з модуля 1 "Методи описової статистики
та статистичний аналіз рядів розподілу"
навчальної дисципліни**

"СТАТИСТИКА"

**для студентів напряму підготовки
6.030506 "Прикладна статистика"
денної форми навчання**