

***В. С. Фетісов***

***Статистика & STATISTICA***  
***Частина І.***  
***Основи***

***Навчальний посібник***



Ніжинський державний університет  
імені Миколи Гоголя

**В. С. Фетісов**

**Статистика & STATISTICA**  
**Частина I.**  
**Основи**

*Навчальний посібник*

Ніжин  
2025

УДК 330.4:519.22:004.9  
Ф45

Рекомендовано Вченою радою  
Ніжинського державного університету імені Миколи Гоголя  
Протокол № 9 від 27.02.2025 р.

**Рецензенти:**

*Гладун О. М.* – заступник директора з наукової роботи Інституту демографії та проблем якості життя НАН України, доктор економічних наук, член-кореспондент НАН України.

*Гончар І. М.* – кандидат економічних наук, доцент кафедри статистики, інформаційно-аналітичних систем і демографії економічного факультету Київського національного університету імені Тараса Шевченка.

**Фетісов В. С.**

Ф45            Статистика & STATISTICA. Частина I. Основи : навч. посіб. Ніжин :  
НДУ ім. М. Гоголя, 2025. 164 с.

*Метою посібнику є дати уявлення, як саме провести статистичний аналіз максимально простою мовою. Оскільки коректний статистичний аналіз даних неможливий без знання статистичної теорії, посібник містить основи статистичної науки з конкретними зрозумілими прикладами. Разом із тим, здійснити статистичний аналіз даних просто неможливо без використання спеціалізованих статистичних пакетів, серед яких одним із беззаперечних лідерів є STATISTICA. Тому у посібнику розглядаються основи роботи зі статистичним пакетом STATISTICA.*

*Перша частина роботи містить теоретичні відомості щодо основних статистичних понять, групування даних, побудови графіків, а також базовий статистичний метод – описові статистики. При цьому все це супроводжуються прикладом реалізації у STATISTICA.*

*Посібник містить також лабораторні завдання для самостійного виконання за допомогою статистичного пакету STATISTICA.*

*Посібник адресований студентам, науковцям, викладачам і практикам, які під час навчання, роботи або практичних досліджень здійснюють статистичний аналіз даних, у тому числі з використанням спеціалізованого програмного забезпечення. Разом і з тим його можна використовувати і як навчальний посібник для роботи зі статистичним пакетом STATISTICA, так і як навчальний посібник зі статистики.*

**УДК 330.4:519.22:004.9**

# Зміст

\_Тос190769201

Умовні позначення.....	5
Передмова.....	6
1. Базові поняття.....	8
Базові поняття статистичного аналізу даних .....	8
STATISTICA: загальні положення.....	11
Робота із системою .....	12
Робота з даними.....	14
Операції з даними .....	31
Робочі книги .....	43
Загальний алгоритм проведення статистичного аналізу.....	44
Вибір модуля (метода статистичного аналізу).....	46
Дослідження даних. Виявлення помилок введення.....	49
2. Групування статистичних даних .....	53
Теорія .....	53
STATISTICA. Частотний аналіз (групування) .....	61
3. Статистичні графіки .....	69
Теорія .....	69
Типи графіків.....	71
Вибір типу графіка.....	84
Побудова графіків у STATIATICA .....	86
Гістограма.....	86
Діаграма розсіяння.....	90
Діаграма розмаху .....	92

Інструменти роботи з графіками .....	101
Редагування графіка.....	105
Редагування первинних даних графіка .....	107
4. Описові статистики.....	110
Теорія .....	110
Статистичні показники.....	110
Середні величини.....	110
Показники варіації.....	123
Характеристики форми розподілу.....	131
Розрахунок описових статистик у STATISTICA .....	136
Термінологічний словник.....	143
Лабораторні роботи .....	146
Література .....	163

## Умовні позначення

$A$	Коефіцієнт асиметрії
$D$	Дециль
$E$	Ексцес
$d$	Частка (доля кожної групи в загальному обсязі сукупності)
$f$	Частота
$h$	Інтервал
$\bar{l}$	Середнє лінійне відхилення
$t$	Кількість груп
$Me$	Медіана
$Mo$	Мода
$n$	Обсяг вибірки
$N$	Чисельність генеральної сукупності
$P$	Ймовірність
$Q, Q_i$	Квартиль.
$\sigma$	Стандартне відхилення (Середнє квадратичне відхилення)
$\sigma^2$	Дисперсія, загальна дисперсія
$\delta^2$	Міжгрупова дисперсія
$R$	Розмах варіації
$v$	(variable) Позначення змінної у таблиці <i>STATISTICA</i>
$V$	Коефіцієнт варіації
$x$	Значення кількісної ознаки. Факторна ознака.
$\bar{x}$	Середня величина
$\tilde{x}$	Вибіркова середня
$x'$	Середина інтервалу

## Передмова

Сучасний світ – це світ інформації. Завдяки Інтернет ми маємо доступ до величезних обсягів найрізноманітніших даних. Але ці дані становляться по справжньому корисними після того, як ми виявляємо існуючі в них закономірності, що дозволяє застосувати їх у практичній діяльності і досить часто напряму позначається на результатах діяльності нашого бізнесу. Розуміння сутності даних дозволяє обрати найкращий варіант напрямку розвитку бізнесу, що автоматично ставить вас у значно кращу ситуацію порівняно до такої, коли ви покладаетесь виключно на інтуїцію і випадок.

Зрозуміло, що розуміння сутності даних можна розкрити тільки після проведення їх *статистичного* аналізу. Аналіз даних потрібний у багатьох сферах людської діяльності. Він застосовується для дослідження процесів, що відбуваються в економіці, демографії, соціології, психології і т. ін., його проводять для вивчення життєвого рівня населення, громадської думки, для оцінки фінансових ризиків, страхової справи, у маркетингових дослідженнях.

Але для проведення статистичного аналізу даних необхідно мати знання основних принципів роботи з даними, статистичних методів обробки даних, знання умов, при яких можливо використовувати той чи інший метод та вміння застосовувати їх на практиці. Все це вирішує *статистика*. Статистика дає змогу компактно описати дані, провести їх класифікацію, виявити існуючі в них закономірності, використовуючи при цьому статистичні методи вивчення даних. Іноді навіть найпростіші статистичні методи аналізу дозволяють визначити типові риси та внутрішні відмінності даних і на підставі них здійснити прогнозування і прийняття рішення щодо розвитку певного явища або процесу.

Великі обсяги первинних статистичних даних просто фізично не дозволяють здійснювати вручну статистичний аналіз даних. Вкрай важко здійснити навіть просте групування кількох тисяч спостережень, виконати найпростіші розрахунки статистичних показників, не кажучи вже про застосування складних статистичних методів.

Але всі ці питання успішно вирішуються за допомогою спеціального класу програмного забезпечення – *прикладних статистичних пакетів*, які широко застосовуються під час роботи з первинними даними у найрізноманітніших галузях. Сьогодні на світовому ринку пропонується велика кількість таких пакетів, що вирішують різноманітні завдання статистичного аналізу даних. Залежно від завдань, які вони вирішують, їх можна поділити на дві великі групи: статистичні пакети загального призначення (універсальні пакети) і спеціалізовані програмні продукти. *Спеціалізовані пакети* реалізують один або кілька статистичних методів або методи, що використовуються в конкретній предметній області. Найчастіше вони орієнтовані на аналіз динамічних рядів, кореляційно-регресійний аналіз, факторний або кластерний аналіз. Застосовувати такі пакети доцільно в тих випадках, коли необхідно систематично вирішувати задачі у галузі, для якої призначений спеціалізований пакет, а можливостей пакетів загального призначення недостатньо. *Універсальні пакети* характеризуються відсутністю безпосередньої орієнтації на специфіку предметної області. Вони пропонують широкий діапазон статистичних методів, прагнучи як до найширшого їх охоплення. Незаперечними лідерами у цій категорії є *STATISTICA* і *SPSS*, які мають дуже схожі можливості.

Автор ставить за мету продемонструвати як саме здійснювати на практиці статистичний аналіз даних за допомогою пакету *STATISTICA*, вважаючи при цьому, що набуті при цьому навички допоможуть за необхідності достатньо швидко опанувати роботу і з конкуруючим пакетом *SPSS*. При цьому усі статистичні методи, що розглядаються у посібнику, супроводжуються прикладом їх реалізації у *STATISTICA*.

Посібник адресований студентам, науковцям, викладачам і практикам, які бажають опанувати вміння використовувати статистичні методи, в тому числі, використовуючи при цьому спеціалізоване програмне забезпечення. Разом і з тим його можна використовувати і для самостійного вивчення і як навчальний посібник.



# 1. Базові поняття

## ***Базові поняття статистичного аналізу даних***

Дані, які потрібні для проведення статистичного аналізу, одержують з різних джерел: зі статистичної звітності і реєстрів, спеціальним чином організованих досліджень тощо. Вони є базою для аналізу, на підставі якого формуються певні висновки для подальшого прийняття рішення. При цьому велику роль має обсяг даних, оскільки чим більше цей обсяг, тим більш чітко проявляється статистична закономірність.

Взагалі збирання даних є важливим процесом, що вимагає ретельної організації і планування, розгляд якого виходить за межі даного посібника. Зауважимо тільки, що зібрані дані мають відповідати таким умовам як вірогідність, своєчасність, доступність і т. ін.

Зібрані дані розглядаються як *статистичні сукупності*. ***Статистична сукупність*** – це множина елементів, поєднаних спільними умовами існування та розвитку. У свою чергу сукупності можуть бути ***генеральними*** або ***вибірковими***. Прикладом генеральної сукупності є *все* населення України під час проведення перепису населення. Вибіркова сукупність або просто вибірка це обмежена за чисельністю частка елементів генеральної сукупності. Прикладом вибіркової сукупності можуть бути студенти певної групи вищого навчального закладу. Кожний із студентів є ***елементом сукупності***. У психології, соціології елементи сукупності ще називають *респондентами*. Кількість елементів сукупності називають ***обсягом вибірки***.

При проведенні статистичного аналізу принципово важливою є вимога ***однорідності*** сукупності. Згадані вище студенти навчального закладу відрізняються один від одного спеціальністю, за якою вони навчаються, формою навчання, курсом і т ін., тобто однорідною її можна вважати лише з точки зору того, що всі студенти навчаються в тому самому

навчальному закладі, але вони досить суттєво відрізняються за формою і спеціальністю навчання, курсом і та ін. У цьому разі однорідною можна вважати навчальну групу. Студенти такої групи поєднані тим, що вони навчаються на тому самому факультеті за однаковою спеціальністю, мають ту саму форму навчання, курс, приблизно однаковий вік та інші властивості, які у статистиці прийнято називати **ознаками**. Отже самі студенти є носієм і джерелом інформації. Для студента ознакою може бути вік, зріст, стать, колір очей, навчальна група, спеціальність, оцінка з певної дисципліни та багато інших. Звичайно, такі ознаки мають своє значення для кожного студента, тобто змінюються від студента до студента, що у статистиці має назву *варіації*, а самі ознаки відповідно називаються *варіюючими* або просто **варіантами**. Одні ознаки є кількісними (зріст, вік), інші мають якісний вираз, наприклад, колір очей, стать. Їх відповідно називають *кількісними* і *атрибутивними* (описовими).

Статистична обробка ознак здійснюється за допомогою системи шкал. **Шкала** – це засіб кількісного вимірювання або впорядкування ознак.

Для вимірювання кількісних ознак застосовуються звичайна *метрична* шкала. Разом із тим для цього може бути також застосована **шкала інтервалів**. Інтервальні ознаки дають змогу не тільки впорядковувати об'єкти вимірювання, а й чисельно виразити та порівняти відмінності між ними. Якщо ознаки вимірюються в інтервальній шкалі, то можна кількісно зафіксувати ступінь вираженості. Наприклад, температура, виміряна в градусах Фаренгейта або Цельсія, утворює інтервальну шкалу. Іншими прикладами таких ознак є вік, вага, артеріальний тиск, коефіцієнт IQ, дохід, емоційна стійкість, дипломатичність і та ін.

Вимірювання і подальша статистична обробка атрибутивних ознак здійснюється завдяки застосуванню таких спеціальних шкал як *номінальна* і *порядкова*. **Номінальна шкала** – це шкала, за якою порядок розташування елементів ознаки не має значення. Наприклад, при проведенні анкетування «Чим Ви любите займатися у вільний час?» можливих відповідей може бути багато:

1. Займаюся спортом.
2. Провожу вільний час у колі друзів.
3. Відвідую культурні заходи і т. ін.

Зрозуміло, що для таких відповідей не можливо визначити порядок їх розташування або ранжування. У такому випадку прийнято позначати окремі значення цифрою «1» у разі відповіді «Так» і «0» у разі відповіді «Ні» (можливо, і навпаки).

Отже **номінальною ознакою** є така ознака, за якою елементи сукупності не можна упорядкувати у певному порядку. Номінальні ознаки називають також *категоріальними*. Зокрема, саме така термінологія використовується у *STATISTICA*.

При застосуванні **порядкової шкали** встановлюється *послідовність розташування* елементів. Значення ознаки такої шкали послідовно збільшується або зменшується. Наприклад, проводиться обстеження студентів щодо ступеня задоволеності майбутньою спеціальністю з можливими варіантами відповідей:

1. Повністю задоволений.
2. Задоволений.
3. Частково незадоволений.
4. Повністю незадоволений.

У цьому випадку кожній відповіді можна надати певний бал (ранг), поступово збільшуючи (або зменшуючи) його від відповіді до відповіді, наприклад від «5» до «2», що дозволить одержати уяву щодо порядку між одержаними значеннями.

Якщо значення для одного студента «5», а для іншого – «3», то можна вважати, що перший студент значно більше задоволений вибором спеціальності, якщо порівняти з другим студентом.

У свою чергу кількісні ознаки поділяються на *неперервні* та *дискретні*. *Неперервні* ознаки мають довільні значення в певних межах. Наприклад, зріст можна визначити з будь-якою точністю: 165 см, 165 см 3 мм, 165 см 3,5 мм і т. д. Разом з тим зрозуміло, що сукупність таких значень буде міститися в певному діапазоні. *Дискретні ознаки* – це ознаки, що набувають тільки окремих, ізольованих значень. Наприклад, оцінка може мати значення в межах від «2» до «5» і ці значення будуть чітко визначені «2», «3», «4» і «5».

### **STATISTICA: загальні положення**

Пакет *STATISTICA* – це універсальний пакет статистичного аналізу, в якому реалізовані основні математичні методи аналізу даних. Розробником пакету є фірма *StatSoft, Inc* (США). У 2014 р. ця фірма була поглинута корпорацією *Dell*, яка внесла пакет *STATISTICA* до складу власної лінійки програмного забезпечення проблематики великих даних.

*STATISTICA* дозволяє проводити різні процедури (модулі) обробки статистичних даних (в термінології програми – *аналізи*):

1. Розрахунок описових статистик.
2. Аналіз динамічних рядів й прогнозування.
3. Факторний аналіз.
4. Множинна регресія.
5. Дисперсійний аналіз.
6. Аналіз відповідностей.
7. Кластерний аналіз.
8. Дискримінантний аналіз і та ін.

Взагалі всі методи (аналізи) *STATISTICA* можна поділити на дві групи:

1. *Розвідувальний* аналіз. Статистичний аналіз з метою виявлення закономірностей в даних.

2. *Підтверджуючий* аналіз. Статистичний аналіз даних.

При цьому деякі статистичні методи можуть бути і розвідувальними і підтверджуючими. До таких методів, наприклад, належать дискримінантний і кластерний аналізи.

Крім загальних статистичних і графічних засобів *STATISTICA* має спеціалізовані модулі: для проведення соціологічних або біомедичних досліджень, вирішення технічних і, що дуже важливо, промислових завдань: карти контролю якості, аналіз процесів і планування експерименту.

За допомогою вбудованої мови програмування *STATISTICA BASIC* можна створювати рішення, які просто інтегруються до інших додатків.

Слід зауважити, що склад модулів відчутно розрізняється залежно від версії та типу ліцензії пакету. Наприклад, базова версія може додатково комплектуватися спеціалізованими модулями: *Power Analysis* (планування статистичних досліджень), *Neural Networks* (нейромережевий аналіз) і т. ін.

Перша версія пакета була створена в 1991 р. В роботі розглядається 12-та англомова версія програми.

### ***Робота із системою***

У загальному випадку робота із системою передбачає таку послідовність дій.

1. Визначити структуру даних.
2. Ввести первинні дані.
3. Провести дослідження даних на помилки.
4. За необхідності здійснити попереднє перетворення даних, наприклад групування або ранжування.
5. Розрахувати описові статистики.
6. Здійснити візуалізацію даних.
7. Застосувати конкретний метод аналізу.

## Початкове вікно програми

Початковий варіант роботи із системою визначається у налаштуваннях системи, доступ до яких здійснюється за командою **Home ▶ Options** (Параметри).

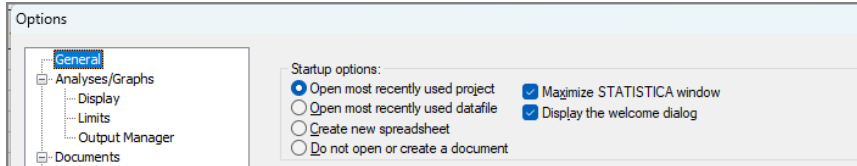


Рисунок 1.1. Налаштування початкового варіанта роботи із системою

Вікно параметрів пункту “General” (Загальні) містить групу “Startup option” (Варіант запуску), вибором однієї з опцій якої і визначається початковий варіант вікна програми під час її завантаження. Це може бути:

1. Open most recently used project. Відкриття проекту, з яким відбувалася робота під час останнього сеансу роботи з програмою.
2. Open most recently used datafile. Відкриття файлу, з яким відбувалася робота під час останнього сеансу роботи з програмою.
3. Create new spreadsheet”. Створення нового файлу даних.
4. Do not open or create a document. Завантажити програму без відкриття або створення нового файлу даних.

Наприклад, якщо в якості початкового варіанту роботи є створення нового файлу даних, то головне вікно міститиме новий документ системи.

Основну область вікна займає область для введення даних, що за принципом побудови дуже схоже з документом електронної таблиці.

## Налаштування (Параметри)

Найбільш швидкий варіант доступу до налаштувань програми здійснюється за командою **Home ▶ Options**.

### *General (Загальні)*

За замовчуванням система після її завантаження шукає документи у власній системній папці. Якщо ваші дані зберігаються в якомусь одному каталозі, то можна визначити її в параметрі “Default location” (Розміщення (файлів) за замовчуванням).

Оскільки ми працюємо в міжнародній системі одиниць (SI) встановіть прапорець для параметра “Use metric measurements” (Використовувати метричні вимірювання).

За замовчуванням система «дбає» про збереження ваших даних, автоматично зберігаючи їх кожні 10 хвилин. Але є можливість вказати власний інтервал часу збереження даних у групі “Auto save” (Автозбереження) в параметрі “Save recovery info every ... minutes” (Зберігати інформацію про відновлення кожні ... хвилини).

## *Робота з даними*

### **Таблиця даних**

Подання даних у програмі має табличний вигляд і зовні дуже схоже з електронною таблицею *Excel*. При цьому у рядках таблиці розташовуються *спостереження (Cases)*, а у стовпчиках – *змінні (Variables)*.

Заголовки рядків (спостереження) нумеруються арабськими цифрами, а заголовки стовпчиків містять ім'я змінної.

Проводячи аналогії з розглянутими вище статистичними поняттями саму *таблицю з даними* слід розглядати як *статистичну сукупність*, *рядок таблиці (спостереження)* – як *одиночку (елемент) цієї сукупності*, а *змінну* – як *ознаку* одиниці сукупності.

Data: fotbal\* (5v by 30c)

Анектування футбольних фахівців

	1 Тренє	2 Вік	3 Збірна	4 Стаж робс
1	No	37	Аргентина	12
2	No	38	Бразилія	6
3	No	39	Бразилія	2
4	No	39	Німеччина	4
5	No	40	Франція	5
6	Yes	44	Бразилія	10
7	Yes	45	Бразилія	4
8	Yes	45	Іспанія	5
9	No	45	Німеччина	10
10	No	45	Німеччина	8
11	No	50	Бразилія	15
12	No	51	Бразилія	15
13	Yes	54	Бразилія	21
14	Yes	55	Аргентина	20
15	Yes	55	Аргентина	16
16	Yes	56	Франція	18
17	Yes	58	Бразилія	25
18	Yes	60	Німеччина	33

Рисунок 1.2. Приклад введення анкетних даних

Наприклад, якщо до таблиці необхідно занести дані анкети, то окреме питання анкети розглядається як змінна, а окрема анкета з усіма запитаннями є спостереженням.

У загальному випадку для створення нової таблиці з даними слід виконати команду **Home** ▶ **New** (або **File** ▶ **New**) або натиснути відповідну кнопку на панелі стандартних інструментів. З'явиться вікно "Create New Document" (Створити новий документ),



в якому слід вибрати *таблицю даних (Spreadsheet)*. За замовчуванням новий файл даних створюється для 10 спостережень і 10 змінних, але, зрозуміло, що ці значення у вікні можна змінити.

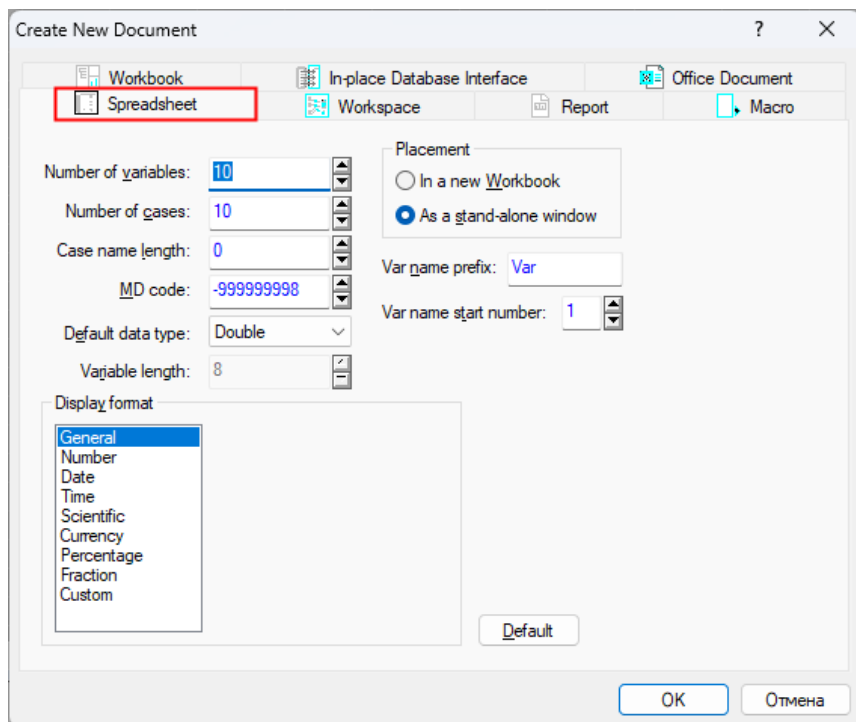


Рисунок 1.3. Завдання параметрів створення нової таблиці

### **Приклад 1.1**

Проведено анкетування 25 футбольних фахівців на тему «Яка збірна, на Ваш погляд, стане чемпіоном на майбутньому чемпіонаті світу?» До анкети було включено такі запитання:

## Структура даних анкети

Питання анкети	Ім'я змінної	Значення змінних
Чи є Ви діючим тренером («Так», «Ні»)	Тренер	«2» – так, «1» – ні,
Вік (повних років)	Вік	
Збірна, яка стане чемпіоном на майбутньому чемпіонаті світу.	Збірна	Збірна: Аргентина, Бразилія, Німеччина, Іспанія, Італія, Франція, інші.
Стаж роботи тренером (повних років)	Стаж роботи	
З якого віку тренує (повних років)	З якого віку тренує	

## Створення структури даних таблиці

Задати потрібну кількість спостережень та змінних для таблиці.

Для додавання змінної слід виконати команду **Data ▶ Variables ▶ Add** (Дані, змінні, додавання). Взагалі команда **Data ▶ Variables** містить основні дії зі змінними: додавання, вилучення, копіювання і переміщення. Всі ж дії зі змінними одержуємо після виклику контекстного меню заголовка змінної:

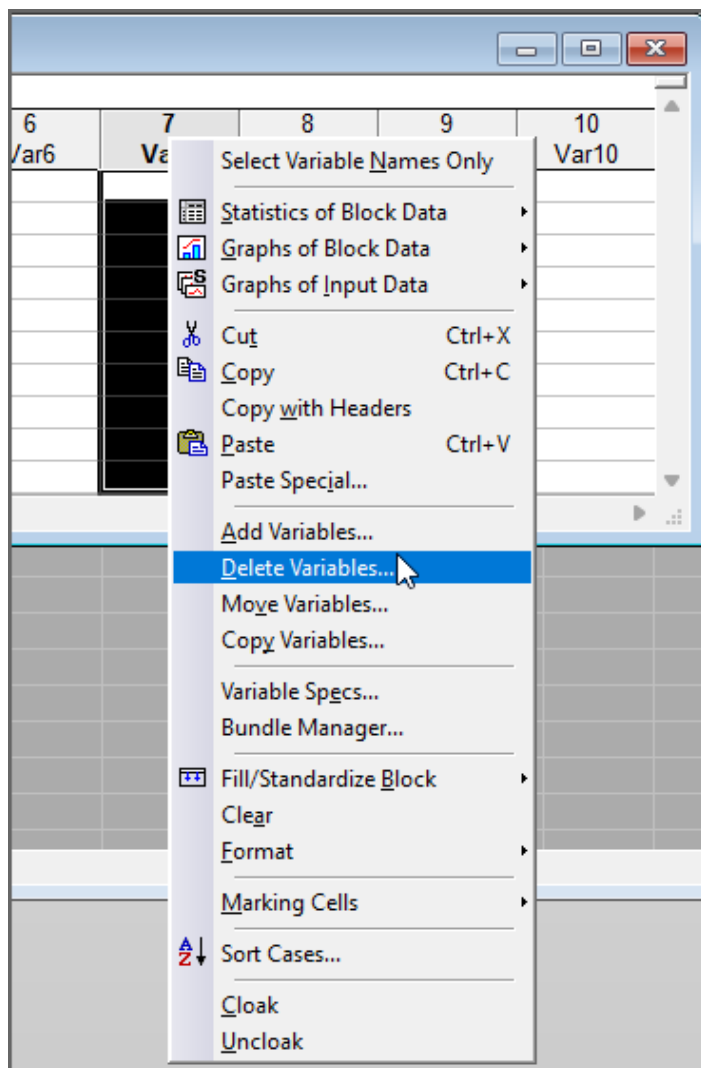



Рисунок 1.4. Контекстне меню роботи зі змінною

Як бачимо з наведеного рисунку, змінні можна додавати, вилучати, переміщувати та копіювати також за допомогою контекстного меню.

Якщо необхідно вилучити зайві змінні, то вибирається пункт “Delete Variables” (Вилучити змінні), а якщо додати – “Add Variables” (Додати змінні). У нашому прикладі 5 змінних, тому слід вилучити 5 зайвих. Після вибору пункту “Delete Variables” з’явиться вікно “Delete Variables”, в якому в полях “From Variable” (починаючи зі змінної) і “To Variable” (до змінної) слід вказати початкове і кінцеве значення імен змінних, які вилучаються і натиснути кнопку «ОК».  При додаванні нової змінної слід звернути увагу на ім’я змінної в полі “After” (після). Якщо у цьому полі відображається ім’я не тієї змінної, після якої необхідно додати нову змінну, його слід змінити ручним введенням нового імені з клавіатури. Це також можна зробити використовуючи список змінних, який з’являється після подвійного клацання в полі “After” або натискання клавіші <F2>, що демонструється на наступному рисунку.

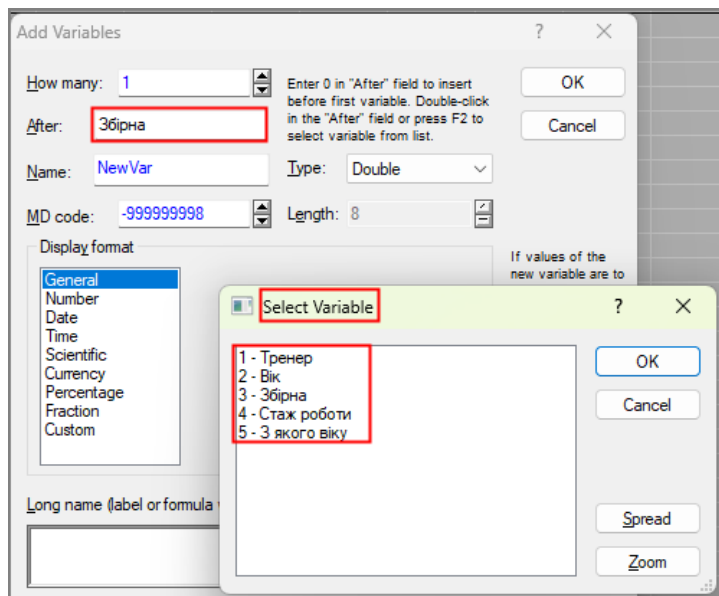


Рисунок 1.5. Список для вибору змінних

Зміну кількості спостережень краще всього робити за командою **Data ▶ Cases** (Дані – Спостереження) і вибираючи з її меню потрібну дію. У нашому прикладі 25 спостережень, тому необхідно додати ще 15 спостережень. Після вибору пункту “Add” (Додати) з’явиться вікно “Add Cases” (дати спостереження), в якому в полях “How many” (кількість спостережень) і “Insert after case” (вставити після спостереження) слід вказати кількість спостережень, які необхідно додати та вказати після якого саме спостереження слід здійснити вставлення нових, і натиснути кнопку «**OK**».

Для кожної таблиці даних можна ввести загальну додаткову інформацію, що зазвичай використовується для ідентифікації звітів. Для цього здійсніть подвійне натискання на полі, що знаходиться під заголовком вікна.

### **Визначення властивостей змінних**

Для задавання імені та інших властивостей змінної необхідно двічі натиснути на її імені у рядку заголовка. З’явиться вікно “Variable” (змінна), в якому і задаються властивості змінної.

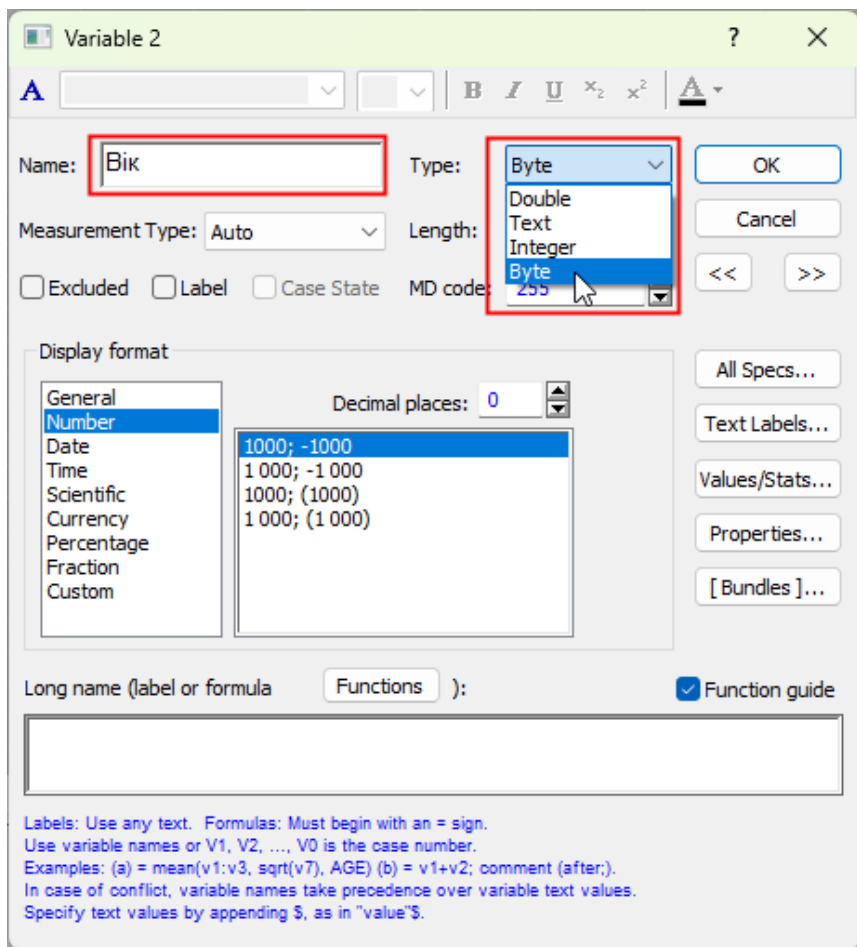


Рисунок 1.6. Завдання властивостей змінної

Система дає змогу швидко відобразити усі властивості однієї або всіх змінних. Для однієї змінної слід встановити курсор на стовпчик, який містить потрібну змінну, і виконати команду **Data ▶ Specs** (Дані – Характеристики), а для всіх змінних – **Data ▶ All Variables Specs** (Дані – Характеристики всіх змінних).

Розглянемо докладно, які саме властивості визначаються у вікні “Variable”.

Ім'я змінної визначається в полі “Name”. За замовчуванням змінні мають імена Var... (скорочення від англ. Variables – змінні). Звичайно, замість цього імені слід ввести зрозуміле.

### ***Тип даних***

Тип даних визначається в полі “Type”.

За потреби тип даних у таблиці можна відобразити у заголовку поруч з іменем змінної. Для цього слід виконати команду **View ▶ Variables Headers ▶ Display Types**. (Вигляд – імена змінних – відобразити типи даних).

Для аналізу даних змінні повинні бути саме у числовому вигляді, для чого застосовують *числовий* тип.

### ***Числовий тип даних***

У системі визначені такі числові типи: “Double” (64-бітні дійсні числа із плаваючою крапкою, тільки цей тип дозволяє вводити числа з комою), числовий), “Integer” (цілі числа в діапазоні від 2,147,483,648 до 2,147,483,647), “Byte” (цілі числа в діапазоні від 0 до 255). Залежно від можливої кількості знаків змінної користувач вибирає потрібний йому тип. Наприклад, для прикладу з футбольними фахівцями змінна «Чи є Ви діючим тренером» може мати тільки два значення «Так» і «Ні». Оскільки порядок розташування значень ролі не грає, то кодування здійснюється за номінальною шкалою, наприклад «2» – «Так», «1» – «Ні». Так як значення містять тільки одну цифру, то доцільно вибрати тип “Byte”.

Кількісні дані, як було сказано вище, можуть бути дискретними або неперервними і вимірюються за метричною шкалою. Для якісних ознак у разі необхідності їх вимірювання здійснюється їх кодування за номінальною або порядковою шкалою.

Деталізація типу даних здійснюється у полі “Display format” (Формат відображення). Наприклад, для числового формату додатково можна встановити кількість знаків після коми, формат “Currency” дає змогу відобразити дані у вигляді грошових сум і т. ін.

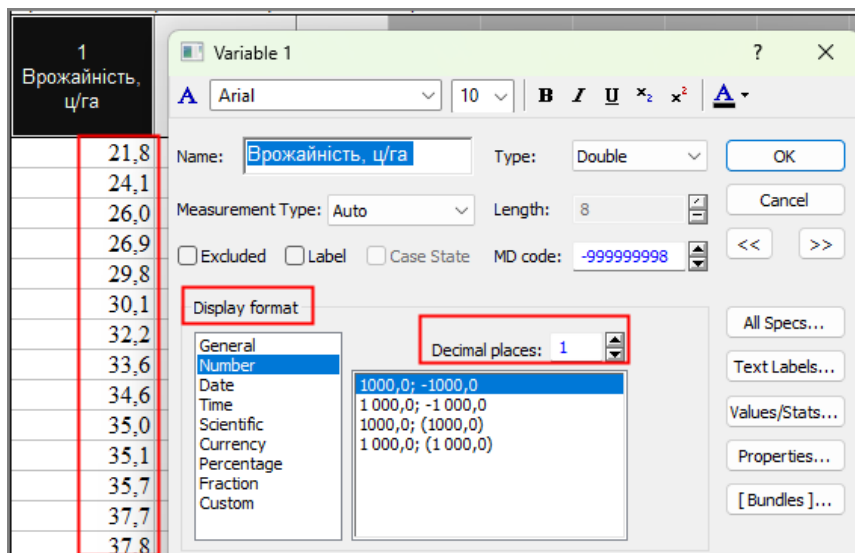


Рисунок 1.7. Завдання формату змінної

### ***Атрибутивний (категоріальний) тип даних***

Атрибутивні змінні в термінології програми називають *категоріальними*.

Якщо атрибутивна змінна не підлягає вимірюванню, то для неї застосовується тип “Text” (Текстовий). За такого типу даних користувач може визначити довжину (кількість знаків) для цієї змінної в полі “Length” (довжина).



Але при цьому врахуйте таке. Дійсно, можна (і потрібно), наприклад, представити значення змінної «Опис доходу» звичайним текстом: «Послуги від зберігання вантажу», «Транспортні послуги», «Реклама» і та ін. Але слід мати на увазі, що у процесі роботи з такого роду змінною може виникнути необхідність, наприклад, здійснити за нею аналіз, наприклад, дисперсійний, то для цього значення змінної повинні бути подані у вигляді даних числового типу.

Тому за можливістю намагайтеся надавати змінним числовий тип, використовуючи текстовий тип виключно для коментарів.

Отже, для того, щоб можна було виконувати статистичну обробку атрибутивних змінних, їм надають числовий тип і кодують числами. Звичайно, що для розуміння таких даних необхідно поставити у відповідність їхнім числовим



еквівалентам текстові значення. Користувачеві при цьому взагалі можна нічого робити, необхідно тільки подбати про те, щоб тип даних для змінної обов'язково був *числовий*. Але *перше* значення необхідно ввести саме у *текстовому* вигляді. При цьому під час запису *першого* текстового значення *STATISTICA* виводить діалогове вікно, в якому користувачеві буде необхідно обрати варіант, що дозволяє *автоматично* надавати текстовим значенням числові еквіваленти: “Enable Text Labels” (увімкнути текстові мітки). Надалі вводити дані у такому полі можна як у текстовому вигляді, так і у вигляді числового коду, який система буде автоматично надавати кожному текстовому значенню.

На наведеному рисунку система пропонує увімкнути текстові мітки (“Enable Text Label”) після введення першого значення змінної у текстовому вигляді.

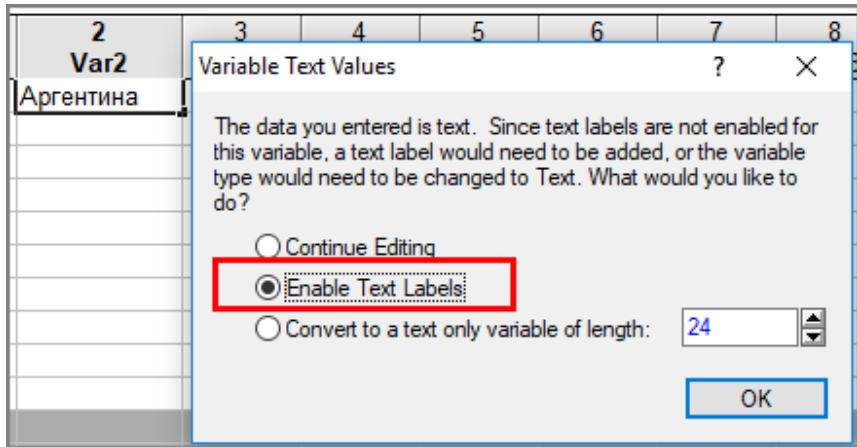



Рисунок 1.8. Ввімкнути текстові мітки

Але користувач може також і вручну створювати еквіваленти-коди для текстових значень. Це здійснюється так.

У вікні властивостей змінної слід натиснути кнопку «**Text Labels**» (текстові мітки). З’явиться вікно “Text Labels Editor” (*Редактор текстових міток*).

У стовпці “Text Label” ввести текстове значення змінної, а у стовбці “Numeric” – її числовий еквівалент.  При цьому *STATISTICA* не допустить наявності двох однакових текстових значень, вимагаючи під час збереження міток замінити конфліктні (дублюючі) записи.

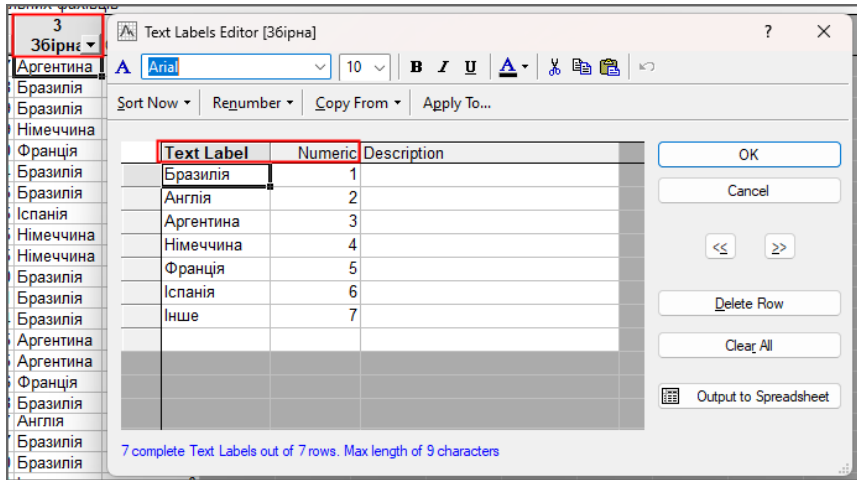


Рисунок 1.9. Редагування кодів текстових значень

Надалі наявність текстових еквівалентів числовим значенням дозволяє перемикати відображення даних у таблиці з числових значень на текстові і навпаки за командою **View ► Display Options ► Text Label**).

### Поле “Long name (label or formula)” (Довге ім’я або формула)

Залежно від потреби користувача це поле може бути або текстовим – для введення коментарів або числовим якщо воно містить розрахункову формулу.



Деякі змінні можна обчислити на основі значень інших змінних. Наприклад, площа прямокутника – це добуток його ширини та висоти. Для обчислення застосовують розрахункову формулу. Щоб ввести формулу, натиснути мишею в полі, що знаходиться під полем “Long name (label or formula)” та сформулювати формулу. При цьому слід врахувати таке.

1. Формула повинна починатися зі знаку «=».

2. Вона може містити імена змінних, математичні і логічні операції, функції тощо. Для визначення та підстановки до формули імені вбудованої функції слід натиснути кнопку «**Functions**».

*STATISTICA* має стандартний механізм запису імен змінних і операцій над ними.

Будь-яка змінна, незважаючи на те, яке ім'я їй надано користувачем, позначається літерою *v* з додаванням до неї її порядкового номеру в таблиці з даними. Так, перша за порядком змінна буде позначатися як *v1*, друга – *v2* і т. д.

Арифметичні оператори записуються так, як це прийнято, зокрема, у мовах програмування: +, -, / (ділення) і \* (множення).

У запису в разі необхідності можна використовувати будь-які логічні оператори, наприклад >, >=, != та оператори, наприклад and, or, not.

Кількість змінних і операцій над ними не обмежується, тому, наприклад, можна сформулювати такий вираз:  $v2 > 16$  and  $v2 < 280$  and  $v1=1$ .

У виразі можна використовувати вбудовані функції *STATISTICA*. Іноді доступ до функції надає специфічна кнопка, але, як правило, користувач має сам знати назви таких функцій, щодо переліку і призначення яких можна дізнатися з довідкової системи. Імена таких функцій, як правило, збігаються з іменами користувацьких функцій, що використовуються у мовах програмування, Excel, математичних і статистичних пакетах. Разом із тим, введення латинської літери ініціює

появу назв функцій, імена яких починаються з цієї літери або кількох перших літер.

З переліком функцій можна ознайомитися у довідковій системі програми. Зауважу тільки, що деякі функції дозволяють автоматично сформувати усі дані для змінної. Це буває корисним, наприклад, під час створення наборів даних. Так, функція **=RND()** заповнить змінну випадковими значеннями на інтервалі (0,1); **=RND(100)** – на інтервалі (0,100); **=RNDNORMAL(50)** – значеннями нормально розподіленої випадкової величини із середнім “0”, і стандартним відхиленням “50”.

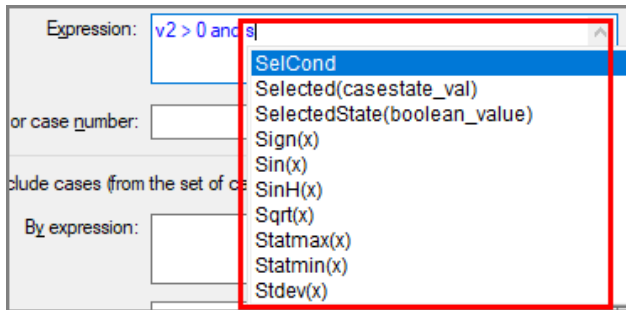


Рисунок 1.10. Вибір вбудованої функції

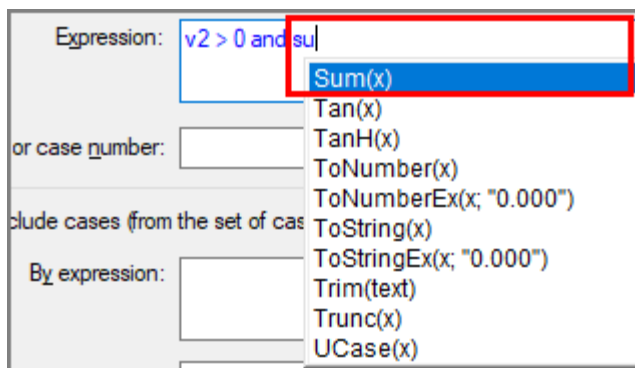


Рисунок 1.11. Продовження вибору вбудованої функції

У прикладі щодо анкетування 25 футбольних фахівців є запитання: «З якого року займаються тренуванням». Серед питань анкети є «Вік» і «Стаж роботи тренером». Отже значення змінної «З якого року займаються тренуванням» можна визначити як «Вік» (змінна  $v_2$ ) – «Стаж роботи тренером» (змінна  $v_4$ ).

Враховуючи вищесказане, значення змінної «З якого року займаються тренуванням» ( $v_5$ ) є різницею змінних  $v_2$  і  $v_4$ , а сама формула буде мати вигляд  $=v_2-v_4$ .

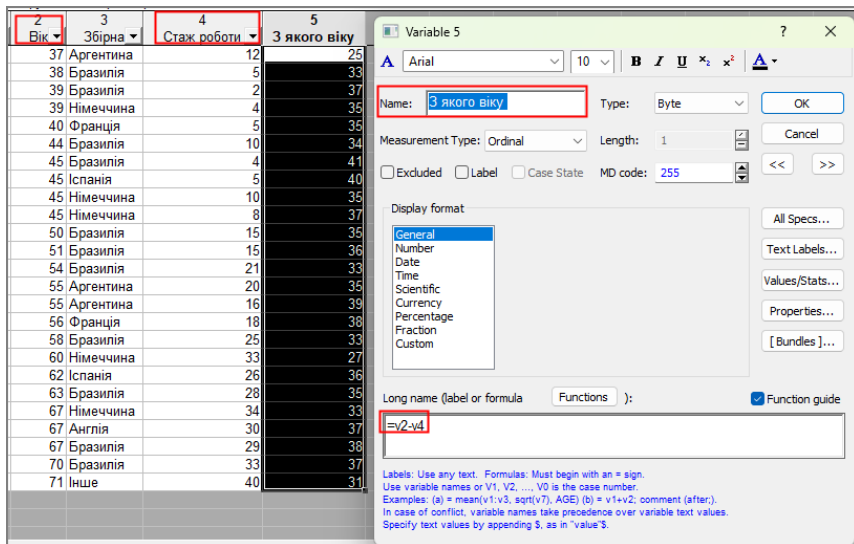


Рисунок 1.12. Формування розрахункової формули для змінної

За потреби розрахункову формулу можна відобразити у заголовку. Для цього слід виконати команду **View ► Variables Headers ► Display Longs Names** (Вигляд – імена змінних – відображати довгі імена).

### Шкали вимірювань (Measurement Type)

Вище було сказано, що статистична обробка ознак здійснюється залежно від типу шкали, які надають засіб кількісного вимірювання або впорядкування ознак.

У полі “Measurement Type” можна задати тип шкали.

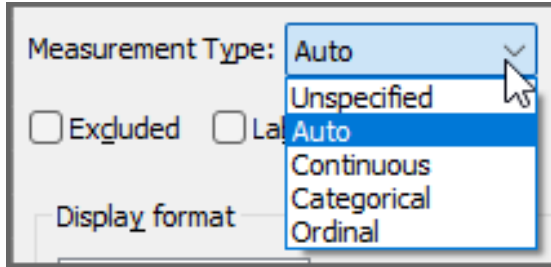


Рисунок 1.13. Вибір типу шкали

Шкала “Categorical” призначена для атрибутивних (категоріальних) змінних. В прикладі анкетування футбольних фахівців такий тип можна застосувати до змінної «Чи є Ви діючим тренером» та визначити для неї дві текстові мітки «Так» і «Ні».


Шкала “Continuous” призначена для неперервних змінних, які можуть мати необмежену кількість значень, в тому числі з дробовою частиною. Отже для таких змінних необхідно використовувати тип даних “Double”. Прикладом таких змінних є зріст, вага, довжина, температура.

Шкала “Ordinal” (порядкова) дозволяє розташувати градації змінної у певному порядку щодо зростання або зменшення інтенсивності властивостей.

Якщо необхідність визначення типу шкали не очевидна, то просто погоджуємося зі значенням за замовчуванням “Auto”.

## *Операції з даними*

### **Введення первинних даних**

Під час введення даних слід врахувати, що  Не необхідно вводити дані для змінних, значення яких можна




обчислити за розрахунковою формулою на основі значень інших змінних. При цьому значення у формулах після змін даних може бути розраховано автоматично або в ручному режимі. Щоб провести розрахунок у ручному режимі необхідно виконати команду **Data ► Recalculate** (Перерахувати дані) або натиснути клавішу <F9>. Різниця між цими діями в тому, що за першим варіантом (ручним) користувач одержує можливість встановлення перерахунку для певних клітинок з вибором змінної, а за другим здійснюється автоматичний перерахунок усіх даних таблиці.

Для автоматичного розрахунку даних за формулами при редагуванні даних слід здійснити налаштування системи:

1. Виконати команду **Home ► Options**. Відкриється вікно “Options”.
2. Вибрати пункт “Spreadsheet” (таблиці).
3. Встановити позначку поруч з полем-міткою “Auto-recalculate spreadsheet formulas when data change” (Виконувати автоматичний розрахунок даних за формулами при зміні даних).

Якщо текстовим даним поставити у відповідність числові еквіваленти, тобто створити текстові мітки, то дані можна вводити у текстовому форматі, а система надалі самостійно

перетворить ці дані в їх числові еквіваленти.  Можна взагалі попередньо не задавати для текстових значень числові еквіваленти, оскільки введення в клітинку текстового значення автоматично створює для неї числовий еквівалент.

Якщо дані повторюються, то їх можна копіювати/вставляти звичайними засобами (копіювання/вставка пункту головного меню **Edit** (Редагування) або за допомогою контекстного меню), при цьому буде скопійовано/ вставлено стільки значень змінних, скільки їх було виділено.

Система дозволяє також здійснювати автозаповнення, механізм якого повністю збігається з аналогічним механізмом в електронній таблиці *Excel*.

Якщо необхідно дублювати дані, то слід їх виділити і навести покажчик на невеличкий квадратик (маркер заповнення) у нижньому правому куті останньої правої клітинки, після чого він має набути хрестоподібний вигляд.

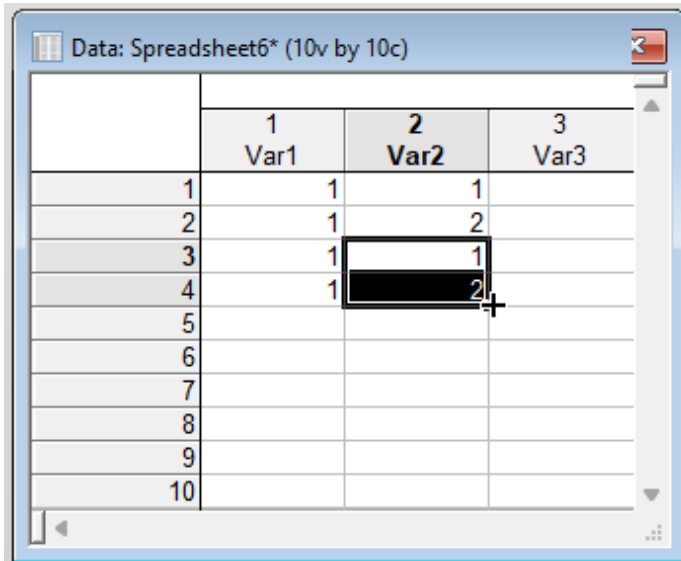


Рисунок 1.14. Демонстрація механізму автозаповнення

Натискаємо кнопку миші і протягуємо покажчик по діапазону клітинок, що слід заповнити.

Практично так само можна вводити в таблиці послідовність чисел. Для цього необхідно у двох суміжних клітинках ввести числа, за якими система визначить приріст (крок зміни), як різницю між цими двома числами. Надалі ці дві клітини виділяються, покажчик встановлюється на маркер заповнення другої клітинки і протягується по діапазону клітинок, які необхідно заповнити.

За необхідності змінити розмір (ширину) стовпчика, коли значення клітинки перевищує наявний розмір, здійснюємо

дану процедуру способами (методами), зазвичай прийнятими в електронних таблицях, тобто перетягуванням правої межі заголовка стовпчика або автоматично шляхом подвійного натискання на правій межі його заголовка.

## **Виділення даних**

Виділення даних буває потрібним, наприклад коли необхідно виконати аналіз не за усіма даними, а за якоюсь їх частиною, для визначення змінної в аналізі і т. ін.

Механізм виділення повністю тотожний аналогічному механізму в *Excel*. Так, для виділення:

- усіх значень змінної слід натиснути на її заголовку;
- кількох суміжних змінних – протягнути мишею на їх заголовках;
- кількох несуміжних змінних – натискати на їх заголовках, утримуючи натиснутою клавішу <Ctrl>.

Нарешті, для виділення всієї таблиці слід натиснути на клітинку, що знаходиться на перетинанні заголовку змінних і стовпчика з порядковими номерами (вона має назву *Info Box*).

## **Обмеження на роботу з даними**

Інколи виникає необхідність виключити зі статистичного аналізу спостереження, яке може бути помилковим, викидом, екстремальним значенням, але в будь-якому варіанті явно спотворює результат.

Усі дії щодо заборони / дозволу на роботу з окремими спостереженнями здійснюються з контекстного меню, що викликається *колонки з порядковим номером* спостереження.

Для виключення з подальших аналізів спостереження слід встановити на ньому курсор і вибрати з контекстного меню пункт “Off”, виконання команди в колонці з’явиться піктограма виключення з обробки:

10	отель аеропорту	15026,11
11	отель аеропорту	990490,00
	Off	17260,00
		0000,00

Рисунок 1.15. Виключення з розрахунків спостереження

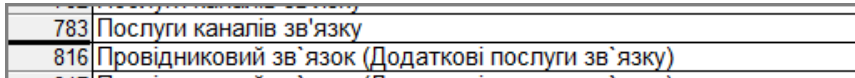
Повторне виконання цієї команди знову зробить спостереження доступним для аналізів. Так само здійснюється виключення з аналізу (а потім додавання) групи спостережень. Якщо ж необхідно зробити доступними для аналізу всі спостереження, то треба виділити всю таблицю та виконати команду “Off” контекстного меню таблиці.

У разі збереження таблиці з даними і в наступних сеансах роботи вилучені спостереження будуть виключатися з обробки.

*STATISTICA* має також можливість приховати спостереження, що здійснюється вибором з контекстного меню пункту “Cloak” (накидка). Але слід мати на увазі, що в цьому випадку приховані спостереження *не виключаються* з аналізів, незважаючи на те, що у таблиці вони відображатися не будуть. До речі, система сама буде нагадувати про це користувачеві. Але *STATISTICA* все ж таки надає можливість вимкнути приховані спостереження з аналізів. Для цього слід виконати **Data ▶ Auto Filter ▶ Mark Cloaked Cases Off** (Дані – Автоматичний фільтр – Вимкнути приховані спостереження). Після виконання цієї команди система позначає приховані спостереження на вимкнення і надалі користувачеві буде необхідним подбати про їх подальше повернення до розрахунків.

Вибір у контекстному меню пункту “Uncloak” (зняти накідку) робить приховані спостереження знову видимими у таблиці даних.

Приховати можна відразу групу спостережень, при цьому на місці прихованих спостережень у першому стовпчику таблиці з’являється жирна горизонтальна лінія.



783	Послуги каналів зв'язку
816	Провідниковий зв'язок (Додаткові послуги зв'язку)

Рисунок 1.16. Приховані спостереження у таблиці з даними

Для поновлення їх відображення слід виділити спостереження з обох боків цієї лінії і вибрати у контекстному меню пункту “Uncloak”. Але якщо невидимими є спостереження з самого початку або наприкінці таблиці, то зробити цього фізично не можна.

У цьому разі для поновлення видимості прихованих спостережень буде необхідно додати до таблиці перший (або останній) рядок, що і дозволить скористатися наведеним вище алгоритмом.

За завершенням процесу відновлення додане штучне спостереження вилучається.

Можна взагалі виключити з обробки відразу групу спостережень таблиці даних. Для цього у контекстному меню виконується команда **Selection Conditions ▶ Remove** (Умови відбору Видалити) або **Selection Conditions ▶ Replace with Selected Cases Excluded**, (Замінити на виключено вибрані випадки). До речі, виконання команди **Selection Conditions ▶ Replace with Selected Cases Included** дозволяє проводити аналіз з вибраними спостереженнями. Виконання команди **Selection Conditions ▶ Enable** відмінняє дію відбору.

*STATISTICA* має низку спеціальних засобів, що дозволяють обмежити дані в таблицях: фільтрація даних, створення вибірок, підмножин (Відносно двох останніх див. «Формування вибірових сукупностей у *STATISTICA*» в розділі «Вибірковий метод»).

### Фільтрація даних

*STATISTICA* має механізм фільтрації даних, зовні дуже схожий на той, що має *Excel*. Для доступу до цього механізму слід виділити змінні, для яких необхідна фільтрація даних, а потім виконати команду **Data ▶ Auto Filter ▶ Auto Filter**. Це призведе до появи у заголовку вибраних змінних кнопки списку, розкриття якого відображає меню з повним переліком значень для вибраної змінної і кількома додатковими діями: вибором режиму впорядкування даних (за зростанням або спаданням), зняттям фільтру (“All”)

Анектування футбольних фахівців				
	1 Трене ▼	2 Вік ▼	3 Збірна ▼	4 Стаж роботи ▼
1	2	37	Sort Ascending	1
2	2	38	Sort Descending	6
3	2	39	(All)	2
4	2	39	(Custom...)	4
5	2	40	Англія	5
6	1	44	Аргентина	0
7	1	45	Бразилія	0
8	1	45	Інше	0
9	2	45	Іспанія	10
10	2	45	Німеччина	8
11	2	50	Франція	15
12	2	51	(Missing)	15
13	1	54	(Non-missing)	0

Рисунок 1.17. Фільтрація даних

Користувач може також самостійно побудувати достатньо складний фільтр, що досягається шляхом вибору пункту “Custom” (користувацький).

Наприклад, для відбору спостережень, для яких змінна «Тренер» має значення «2», а стаж роботи становить більше 10 років, застосовуючи стандартний механізм запису імен змінних та операцій над ними, сформуємо у вікні введення, яке з’явиться після вибору пункту “Custom”, такий вираз:

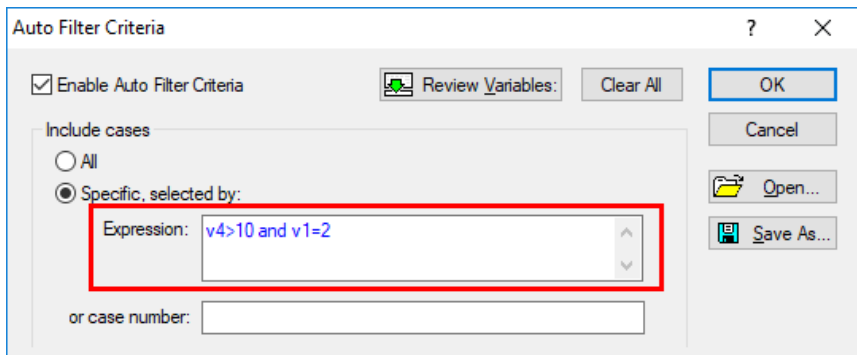


Рисунок 1.18. Створення власного складного фільтру

Слід мати на увазі, що збереження змін у таблиці даних призводить також до збереження фільтру, який буде застосований до даних у наступному сеансі роботи.

### Сортування спостережень

Спостереження у таблиці можна впорядковувати (сортувати) за значеннями однієї або відразу кількох змінних. Для цього слід виконати такі дії.

1. Виконати команду **Data ▶ Sort**. Відкриється вікно “Sort Options” (параметри сортування).
2. Визначити змінну (або змінні), за якою слід впорядкувати дані (Add Vars), напрямок сортування: за зростанням або за спаданням (Direction). Для змінних текстового типу можна

також вибрати варіант сортування (Sort By): за текстом або за числовим кодом, що відповідає текстовому опису змінної. Програма також надає можливість створити нову таблицю з (упорядкованими) даними, встановивши позначку біля дії “Create Spreadsheet” (Створити таблицю).

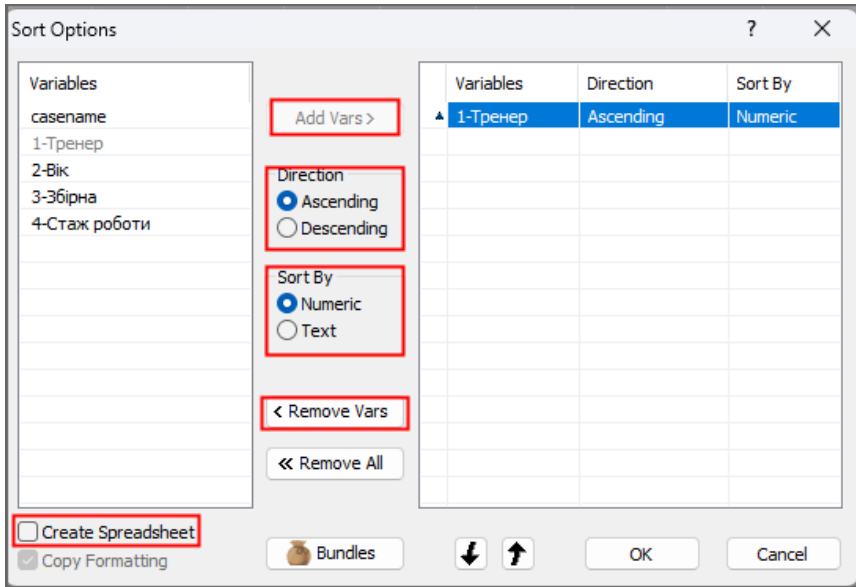


Рисунок 1.19. Сортування спостережень

3. Натиснути «ОК».

### Імпорт даних

Якщо дані були сформовані в інших статистичних пакетах, наприклад, SPSS або в електронній таблиці, то їх можна скопіювати до таблиці з даними. Для цього слід виконати команду **Home** ► **Open** або **File** ► **Open** та вибрати у стандартному вікні “Open”, наприклад, файл Excel.

Після вибору файлу Excel з’явиться вікно, в якому буде запропоновано здійснити вибір до таблиці всіх аркушів файлу або якогось конкретного.



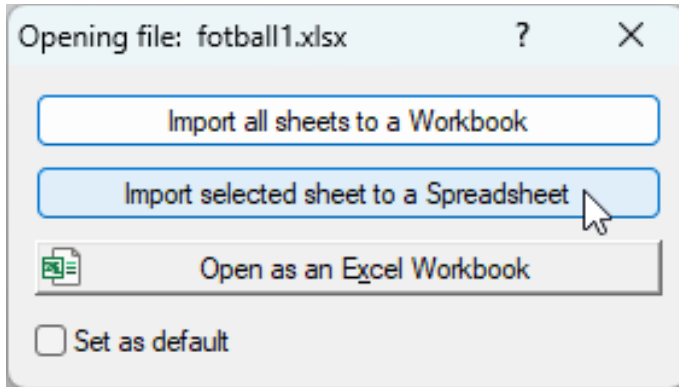



Рисунок 1.20. Варіанти імпорту даних

Після вибору потрібного аркуша *STATISTICA* запропонує визначитися з кількома пунктами:

1. Звідки брати імена змінних: з першого рядка (from first row) або першого стовпчика (from first column).  Якщо таблиця Excel не має заголовків, то необхідно зняти позначку з обох пунктів.

2. Чи необхідно використовувати наявний формат клітинок (import cell format).

3. Чи необхідно створювати текстові мітки для змінних числового типу.

4. Діапазон даних таблиці Excel, який *STATISTICA* визначає автоматично.

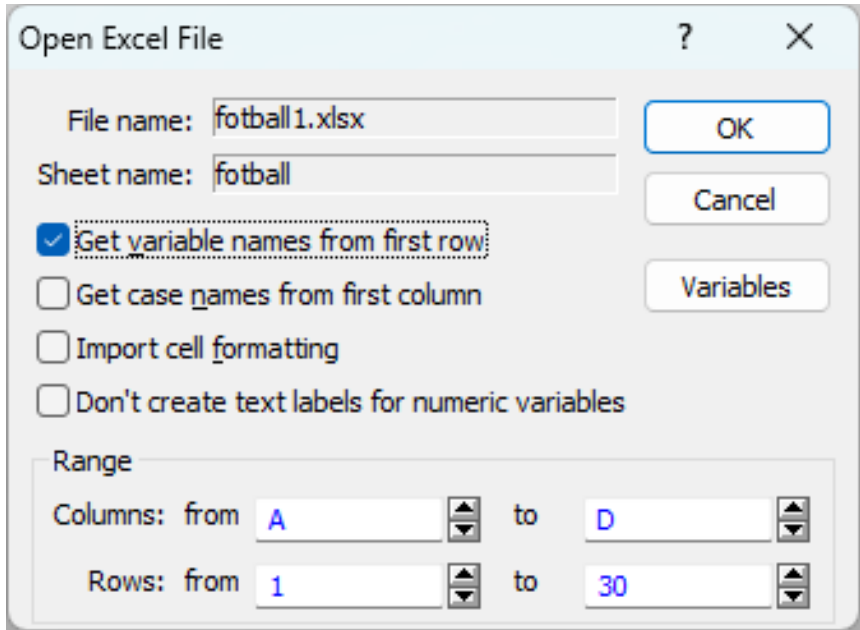


Рисунок 1.21. Імпорт даних з таблиці Excel

Натискаємо «OK» та отримуємо нову таблицю *STATISTICA*.

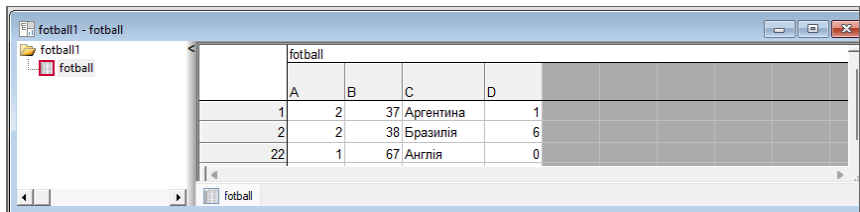


Рисунок 1.22. Результат імпорту даних з таблиці Excel

І якщо надалі зі змінною з текстового типу даних не передбачається виконання якихось аналізів, то на цьому процедура імпорту закінчується. Якщо ж надалі такий аналіз

планується, то необхідно перетворити тип даних з текстового в числовий:

1. Виділяємо стовпчик з текстовими даними, змінюємо тип даних: текстовий на числовий, наприклад, “Integer”.

2. Натискаємо кнопку «Text Labels», після чого *STATISTICA* автоматично сформує числові значення для такої змінної.

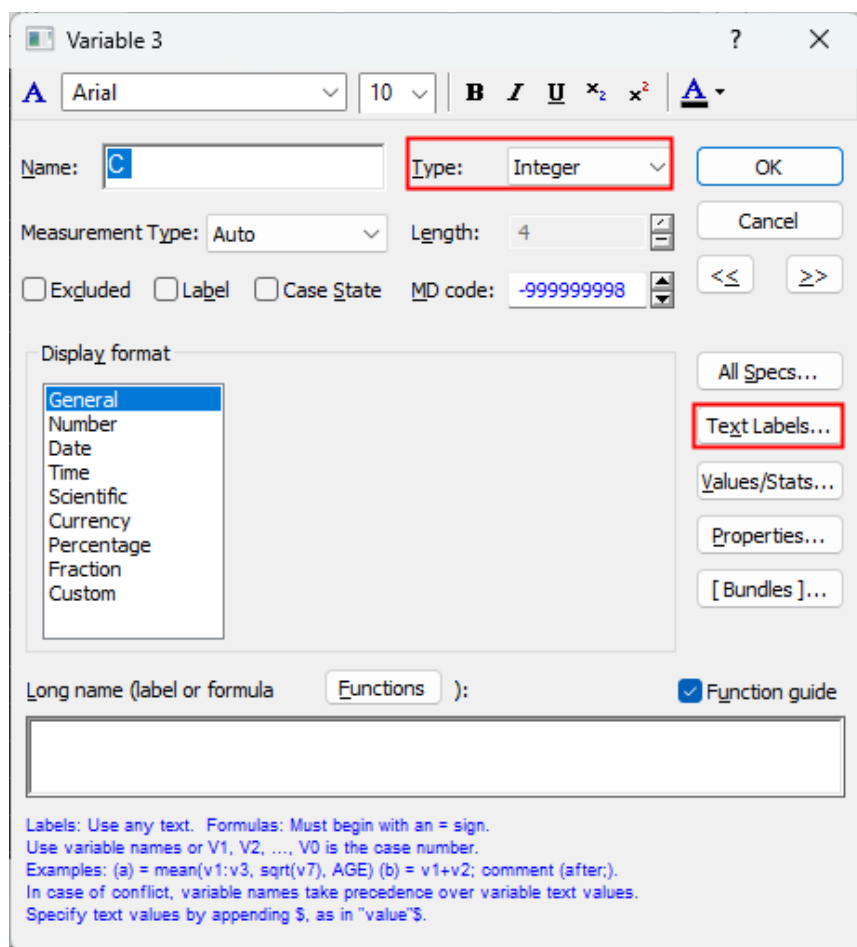


Рисунок 1.23. Формування числових значень для текстових змінних

### Збереження даних

Збереження файлів відбувається стандартним чином за командою **File ▶ Save**, **Home ▶ Save** або натисканням піктограми



Якщо необхідно зберегти існуючий файл на інше іншою назвою, виконуємо стандартну команду **File ▶ Save As** або **Home ▶ Save As**. До імені файлу *STATISTICA* автоматично додає розширення *STA*. Такий файл даних є набором файлів, оскільки він містить і автоматично зберігає інформацію про всі додаткові файли (графіки, звіти, програми), що використовуються з поточним набором даних. Крім цього зберегти таблицю даних можна також у форматі електронної таблиці *Excel*, у вигляді вебсторінки, текстового файлу і т. ін.

### Відкриття даних

За певних налаштувань під час завантаження системи автоматично відкривається останній файл, з яким відбувалася робота. Відкрити файл даних можна також під час роботи з системою за командою **Home ▶ Open** або **File ▶ Open**.

При цьому в робочій області може бути тільки один файл з даними.

### Робочі книги

Всі результати: графіки, діаграми, таблиці, отримані під час роботи з даними, записуються у *робочу книгу (Workbook)*.



При цьому результати будь-якого аналізу відкритої таблиці даних записуються в одну робочу книгу, але на окремому аркуші.

Кожна робоча книга міститиме два робочих поля: деревоподібне навігаційне поле, побудоване на зразок «Провідника», для навігації аркушами (результатами аналізу)

у лівій частині вікна та поле перегляду вмісту аркуша.

Від аркуша до аркуша можна також переходити за допомогою вкладок у нижній частині робочої книги.

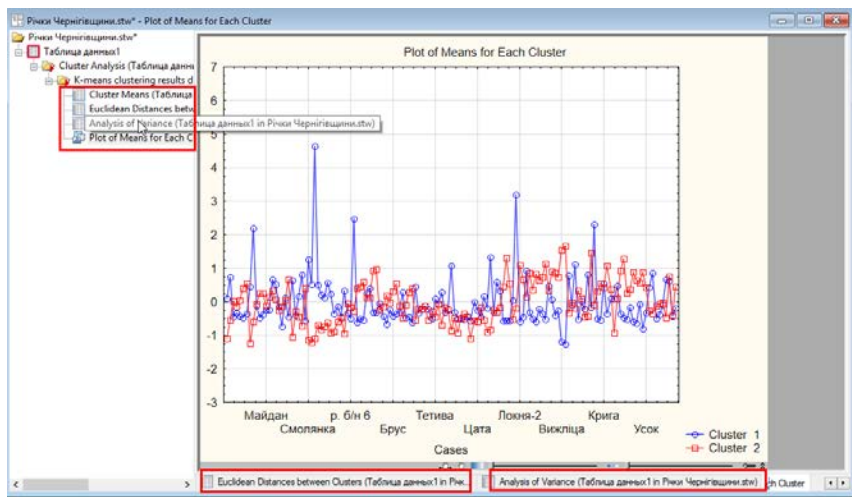


Рисунок 1.24. Приклад робочої книги з розрахунками

Самі робочі книги можна зберігати та редагувати, вставляти, вилучати, перейменовувати окремі аркуші. Для виконання певної дії слід викликати контекстне меню на назві аркуша лівої частини вікна і вибрати з нього потрібну дію.

Окремі аркуші книги можна «витягувати» з неї, після чого вони вилучаються з неї і з’являються у новому вікні. Для цього їх слід просто перетягнути мишею зі списку робочої книги за межі її вікна.

### **Загальний алгоритм проведення статистичного аналізу**

Як зазначалося раніше, *STATISTICA* містить кілька модулів, кожний з яких реалізує певний метод статистичного аналізу. Але всі вони виконуються за єдиним алгоритмом:

1. Вибір модуля (метода статистичного аналізу).
2. Вибір змінної для аналізу.
3. Налаштування модуля (визначення параметрів для застосування методу).
4. Ініціація розрахунку (проведення аналізу). Як правило, для цього використовується кнопка «**Summary**».
5. За необхідності збереження результатів розрахунку.
6. Завершення роботи з аналізом. Найбільш швидкий варіант цього – застосування опції “Close Analysis” (закрити аналіз) пункт меню кнопки «**Options**» у вікні налаштування.

Натискання кнопки «**Cancel**» здійснює повернення до попереднього вікна.

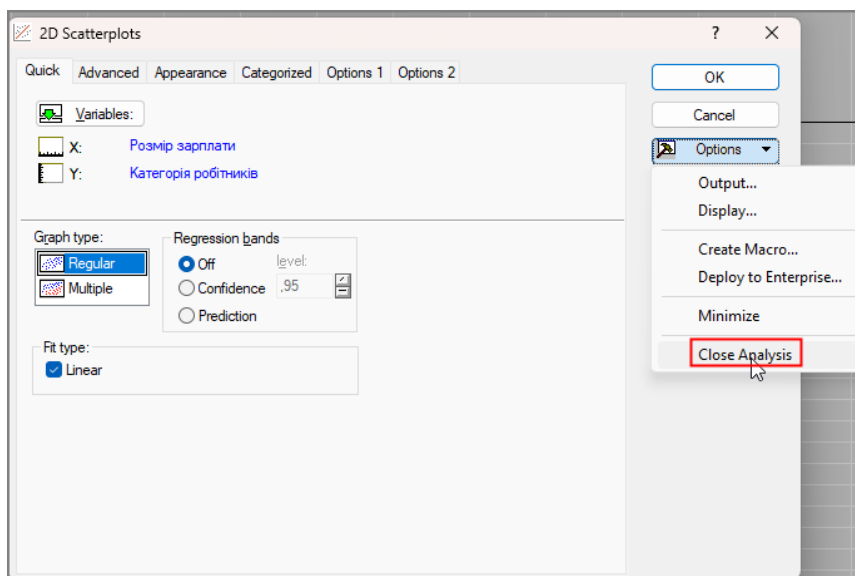



Рисунок 1.25. Приклад вікна аналізу

## *Вибір модуля (метода статистичного аналізу)*

Звернутися до будь-якого модуля можна кількома способами:

1. Використовуючи пункти вкладки **Statistics** (Аналізи).

2. Використовуючи панель інструментів **Statistics** якщо ви працюєте з класичним меню. Якщо панель інструментів **Statistics** відсутня, то виведення її на екран здійснюється стандартним чином: за командою **View ▶ Toolbar ▶ Statistics** або викликом контекстного меню області панелі інструментів і вибором з меню пункту “Statistics”.

3. Натисканням кнопки виклику меню для найбільш вживаних інструментів, що знаходиться у лівому куті панелі статусу  “Start menu of commonly used tools”.

## **Вибір змінної для аналізу**

Для проведення аналізу необхідно виділити змінну. Це може бути здійснено різними способами. За найпростішим варіантом це можна зробити так само, як це робиться в електронних таблицях, натиснувши на заголовку змінної. І, взагалі, засоби виділення даних таблиці збігаються з засобами, що використовуються в електронних таблицях. Наприклад, щоб виділити повністю рядок або стовпчик можна натиснувши на його заголовку, для виділення кількох несуміжних змінних натискають на їх заголовку, утримуючи натиснутою клавішу **<Ctrl>** і т. ін. Проте можна і не виділяти повністю змінну: достатньо просто встановити курсор у клітинці з потрібною змінною або виділити у рядку клітинки зі змінними, якщо показники необхідно розрахувати відразу для кількох змінних.

Але вибір змінної можна здійснити і пізніше у вікні модуля, використовуючи кнопку «**Variables**» (Змінні). Натискання цієї кнопки ініціює появу вікна “Select the Variables for the analysis” (Виберіть змінні для аналізу). Після вибору змінної праворуч від кнопки «**Variables**» відображається ім'я вибраної змінної.

Іноді достатньо несподівано у вікні вибору змінних не будуть відображатися деякі змінні таблиці даних.

*STATISTICA* прагне обмежити коло змінних, що пропонуються до вибору у вікні вибору, на підставі аналізу їх типів.

На перший погляд, це є цілком слушним, але іноді це робиться некоректно і може заплутати користувача. Наприклад, інколи програма ігнорує змінну, що має тип даних «цілий».

Наприклад, є дві змінні з однаковими даними, але різними типами даних:

1	2
NewVar Double	Зріст Integer
165,00	165,00
171,00	171,00
155,00	155,00
180,00	180,00

Рисунок 1.26. Змінні з однаковими даними і різними типами даних

У вікні вибору буде відображена тільки одна змінна з типом даних «Подвійний» (Double):



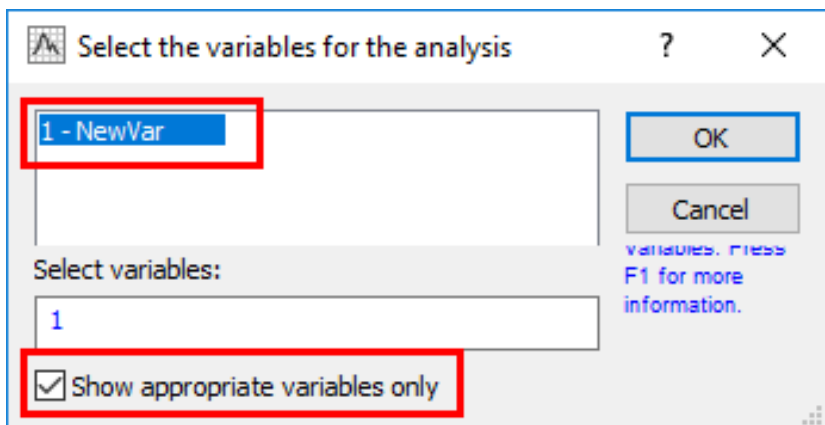


Рисунок 1.27. Вікно вибору змінної зі списку з «придатними» для аналізу змінними

Для усунення такої ситуації необхідно зняти прапорець поруч з полем-міткою “Show appropriate Variables only” (Показувати тільки придатні для аналізу змінні), після чого вікно вибору відобразить обидві змінні:

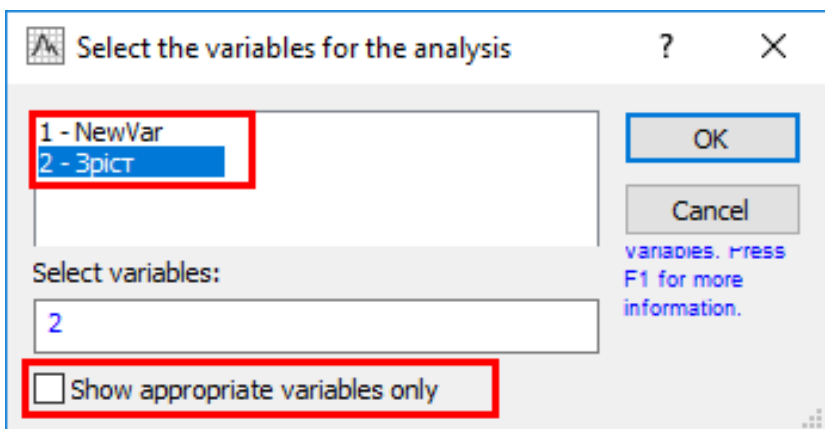


Рисунок 1.28. Вікно вибору змінної зі списку без урахування «придатних» для аналізу змінних

## *Дослідження даних. Виявлення помилок введення*

На першому кроці дослідження доцільно піддати докладному аналізу самі дані з метою виявлення помилок введення, а в багатьох випадках також здійснити перевірку закону розподілу даних на нормальність.

Найточніший спосіб перевірки даних (тобто значень всіх змінних) на помилки полягає у порівнянні даних, уведених у таблицю, з оригіналом (наприклад, з анкетною). Проте цей спосіб вимагає дуже багато часу, особливо, якщо обсяг даних занадто великий. Тому проводити таку трудомістку роботу доцільно тільки тоді, коли обсяг даних є невеликим. У загальному випадку рекомендується проводити частотний аналіз значень змінних. Результати частотного аналізу досить часто дозволяють виявити невірні значення. Наприклад, якщо змінна містить зріст в сантиметрах, то значення "400" при частотному аналізі явно свідчить про те, що в даних є помилка. Після проведення частотного аналізу це значення можна відшукати у таблиці даних і виправити. Отже, під час аналізу частотних таблиць особливу увагу треба звертати на максимальні і мінімальні значення. Проте, якщо замість віку 65 років було введено, наприклад, значення 56, то за допомогою таблиці частот цю помилку виявити неможливо.

Часто є можливість провести змістовний аналіз даних з використанням таблиць спряженості. Наприклад, якщо дані узяті з анкети, в якій було питання щодо сімейного стану (холостий/не заміжня, одружений/одружена, вдівець/вдова, розведений(а)), то, побудувавши таблицю спряженості для цього питання і питання на зразок: «Якщо у вас є сім'я, то чи прийнято у вас проводити відпустку окремо один від одного?», легко можна виявити, чи відповіли на нього тільки одружені опитувані.

Маючи певні практичні навички, за допомогою таких прийомів можна виявити велику кількість помилок введення. Всі такі помилки обов'язково повинні бути виправлені. Навіть

якщо спостережень кілька тисяч, то навіть одне суперечливе значення завдає шкоди дослідженню, тому що створюється враження, що робота зі збирання інформації виконана неякісно.

Дослідження даних може бути проведено і спеціальними засобами програми. Для виявлення помилок введення програма має спеціальний модуль «Перевірка даних», доступ до якого здійснюється за командою **Data ▶ Verify Data** (Дані – Перевірка даних). У вікні модуля можна задати для даних кілька умов, виконання яких (усіх або хоча б однієї) дозволяє вважати їх правильними. Кожна умова складається з двох частин. Перша частина – це заздалегідь визначені правила, наприклад, “Valid If” (Вірно, якщо), “Valid cases” (Вірні спостереження), “Invalid cases” (Не вірні спостереження). Друга частина є виразом для правила, що формується за такими правилами.

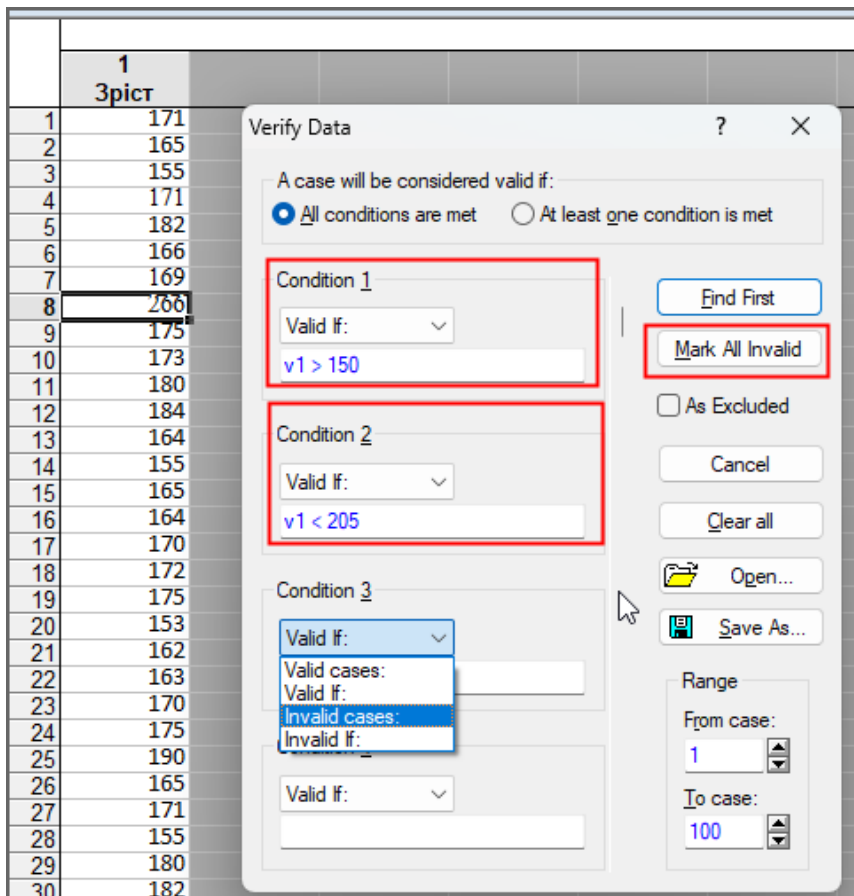


Рисунок 1.29. Завдання умов для перевірки на помилки даних

1. Ліва частина умови – це ім'я змінної.
2. Після імені змінної вводиться логічний оператор.
3. Права частина умови – це значення з яким порівнюється змінна. При цьому значенням може бути число, математичний вираз, функція тощо.

Наприклад, значення змінної «Зріст» з високою ймовірністю буде в межах від 150 до 205 см. Якщо у таблиці з

даними ця змінна має ім'я "v1", то це припущення можна виразити через дві умови :

1.  $v1 > 150$
2.  $v2 < 205$

Після формування умов також необхідно визначити:

1. *Діапазон перевірки.* За замовчуванням перевіряються всі спостереження, але можна задати перевірку на певному інтервалі спостережень.

2. *Характер перевірки.* Знайти відразу всі помилкові значення або робити це крок за кроком, відшукуючи спостереження з неправильними даними послідовно одне за одним. Дії виконуються відповідно з використанням інструментів "Mark All Invalid" (Відмітити всі невірні) і "Find First" (знайти перше). За першим варіантом усі неправильні значення виділяються червоним кольором.

	1
	Зріст
1	171
2	165
3	155
4	171
5	182
6	166
7	169
8	166
9	175
10	213
11	180

Рисунок 1.30. Приклад результатів перевірки на помилки даних

*STATISTICA* виділяє червоним кольором значення, на які слід звернути увагу. Наприклад, виділення червоним кольором коефіцієнта кореляції свідчить про те, що для певного рівня значущості між змінними визнається існування зв'язку.

## 2. Групування статистичних даних

### *Теорія*

Зібрані статистичні дані, як правило, підлягають *групуванню*. Первинні дані, навіть невеликого обсягу, дуже важко проаналізувати, не кажучи вже про реальні масиви даних, які являють собою великі сукупності, тому групування у статистиці, як правило, передує проведенню статистичного аналізу.

Уявіть просту ситуацію. необхідно здійснити аналіз успішності в навчальному закладі, наприклад, за семестр. Звичайно, всі оцінки містять відповідні документи – журнали. Але скільки часу займе навіть перегляд усіх оцінок? Більше того, все в кінцевому підсумку зведеться до підрахунку кількості одержаних оцінок за кожним балом. Такий процес як раз є прикладом найпростішого групування. *Групування* – це процес розподілу сукупності на групи з метою виявлення спільних рис і суттєвих відмінностей серед елементів сукупності. Групи будуються за певною ознакою, яка називається в цьому випадку *групувальною*. Групувальна ознака ще називається *варіантою* і позначається у статистиці літерою “х”.

Якщо групувальна ознака є дискретною, то побудова груп відбувається дуже просто, оскільки кожна група відповідає конкретному значенню ознаки.

*Приклад 2.1.* Було проведено опитування студентів навчальної групи з метою з'ясування їх орієнтованості на педагогічну роботу. Оцінювання проводилося за 5-бальною ранговою шкалою. Значення для максимальної орієнтованості має найвищий бал (ранг) – «5», для мінімальної – «2». Свою

орієнтованість на педагогічну роботу студенти оцінили так: 4, 4, 5, 5, 4, 2, 4, 5, 3, 3, 4, 4, 4, 5, 5, 4, 4, 4, 4, 3, 5, 3, 4, 3, 4.

Оскільки оцінка набуває значення від «2» до «5», то групування буде мати такий вигляд:

Таблиця 2.1

**Розподіл студентів початкової групи за ступенем орієнтованості на педагогічну роботу**

<i>Ступень орієнтованості, бали</i>	<i>Кількість студентів</i>
<i>x</i>	<i>f</i>
2	1
3	5
4	13
5	6
<b>Разом</b>	<b>25</b>

Після цього підраховується кількість студентів, що мають певну оцінку, і отримані значення заносять до відповідної клітинки другого стовпчика. Такі значення, що показують, скільки разів повторюються певні значення варіанти у групуванні, називаються *частотами* і позначаються літерою “*f*”.

Якщо групування складаються тільки зі значень груповальної ознаки та їх частот, то таке групування називається *рядом розподілу*. Таблиця 2.1 як раз і є типовим прикладом ряду розподілу. Графічним зображенням ряду розподілу є *крива розподілу*, що характеризує його тип і форму, про що докладно йтиметься в наступних розділах. Її ще називають *емпіричною кривою розподілу*, щоб підкреслити, що вона побудована за емпіричними (первинними) даними, на відміну від *теоретичної кривої*, яка «у чистому вигляді» відображає якийсь *теоретичний розподіл*, наприклад, нормальний розподіл.

При проведенні групування за неперервною кількісною ознакою будують *інтервали*, кількість яких визначає сам дослідник.

*Інтервал* – це різниця між двома сусідніми значеннями групувальної ознаки, які називають нижньою та верхньою межами інтервалу. Наприклад, при визначенні доходу можна створити інтервал «14200-16000», в якому «14200» буде нижньою, а «16000» – верхньою межами інтервалу.

Інтервали бувають *рівними* і *нерівними*, *закритими* і *відкритими*. *Закриті* інтервали – це інтервали на зразок наведеного вище, вони мають і нижню, і верхню межі. *Відкриті* інтервали мають тільки одну з меж, наприклад, «До 20 років», «10000 грн і більше». У першому випадку відсутня нижня межа, у другому – верхня.

Процес побудови груп є *суб'єктивним*, він залежить від людини, яка здійснює групування. Тому як кількість груп, так і розміри інтервалів може бути різними для однієї тієї ж сукупності при здійсненні групування різними людьми. А це, у свою чергу, може привести до того, що за тими самими даними можуть бути отримані зовсім різні і навіть протилежні висновки. Тому під час проведення групування необхідно на базі аналізу сутності явища, що вивчається, визначати спільні риси окремих елементів сукупності і тільки тоді поєднувати їх в (однотипні) групи.

При виділенні груп слід прагнути до того, щоб до кожної групи потрапило достатньо велика кількість елементів сукупності.

Суто математично кількість інтервалів прийнято визначати за формулою Стерджеса:

$$m = 1 + 3,322 \lg(n) \quad (2.1)$$

де  $n$  – кількість одиниць сукупності.



Для *рівних* інтервалів *ширину* (*крок*) інтервалу визначають за формулою:

$$h = \frac{x_{max} - x_{min}}{m} \quad (2.2)$$

де

$h$  – ширина інтервалу,

$x_{max}$  і  $x_{min}$  – відповідно найбільше і найменше значення групувальної ознаки.

**Приклад 2.2.** Студенти навчальної групи університету мають зріст 165, 171, 155, 180, 182, 166, 158, 150, 164, 173, 180, 184, 154, 155, 165, 164, 170, 172, 175, 153, 165, 163, 170, 175, 190. необхідно побудувати групування із п'яти рівних інтервалів.

**Розв'язання.**

Знаходимо серед значень ознаки «Зріст» максимальне та мінімальне значення – «190» і «150». Це і будуть  $x_{max}$  і  $x_{min}$ . Оскільки необхідно побудувати п'ять груп, маємо:

$$h = \frac{190 - 150}{5} = 8 \text{ см}$$

Нижня межа першого інтервалу збігається зі значенням  $x_{min}$ . Для визначення верхньої межі першого інтервалу слід додати до значення  $x_{min}$  крок інтервалу:  $150+8=158$ . Нижня межа кожного наступного інтервалу збігається з верхньою межею попереднього, а для визначення верхньої межі поточного інтервалу до її нижньої межі додається крок інтервалу. Наприклад, для другого інтервалу нижня межа дорівнює верхній межі першого інтервалу, а її верхня межа розраховується як  $158+8=166$ . Наступним кроком будується групування, першим стовпчиком якого завжди є межі групувальної ознаки.

Надалі необхідно визначити, до якого саме інтервалу належить те чи інше значення, Наприклад, значення «165» припадає на інтервал «158-166», «171» – на інтервал «166-174»

і т. д. Підсумовані значення є частотами і заносяться до графі «Кількість студентів, осіб.».

Таблиця 2.2

**Розподіл студентів навчальної групи університету за зростом. Варіант 1**

<i>Зріст, см.</i>	<i>Кількість студентів, осіб</i>	<i>Частка</i>	<i>Кумулятивна частота</i>
<i>x</i>	<i>f</i>	<i>d</i>	<i>S</i>
150-158	6	0,24	6
158-166	7	0,28	13
166-174	5	0,20	18
174-182	5	0,20	23
182-190	2	0,08	25
Разом	25	1,00	x

При цьому може виникнути питання до якого саме закритого інтервалу слід відносити значення, що припадає на межу інтервалів, наприклад «158», яке одночасно є верхньою межею першого інтервалу і нижньою межею другого. В принципі це значення може бути віднесено до будь-якого з цих інтервалів, але, як правило, за наявності закритих інтервалів приймають, що верхня межа – це включно, нижня – це виключно.

Разом із тим включення значення до інтервалу може бути визначено словом для першого та (або) останнього інтервалу, якщо він є *відкритим*. [3, ст.66] Про це свідчать слова «до» у першому інтервалі та (або) «більше» в останньому інтервалі. Оскільки для наведеного прикладу першій інтервал «150-158», то його можна записати як «До 158». Це вказує на те, що *нижня межа закритого* інтервалу буде включати значення «включно», а верхня – «виключно», чим і слід керуватися при

віднесенні значення, яке припадає на межу інтервалів. У даному випадку значення «158», що припадає на межу першого та другого інтервалів, слід віднести до другого інтервалу «158-166», тоді таблиця набуде такого вигляду. До речі, саме за таким варіантом будуються групування у *STATISTICA*, про що йтиметься пізніше.

Таблиця 2.3

**Розподіл студентів навчальної групи університету за зростом. Варіант 2**

<i>Зріст, см.</i>	<i>Кількість студентів, осіб</i>	<i>Частка</i>	<i>Кумулятивна частота</i>
<i>x</i>	<i>f</i>	<i>d</i>	<i>S</i>
До 158	5	0,20	5
158-166	7	0,28	12
166-174	6	0,24	18
174-182	4	0,16	22
182-190	2	0,08	24
190 і більше	1	0,04	25
Разом	25	1,00	x

Разом з частотами для групування можна визначити ще кілька статистичних показників.

*Частка* – це доля кожної групи в загальному обсязі сукупності.

$$d_i = \frac{f_i}{\sum f_i} \quad (2.3)$$

Частка може бути використана для дослідження структури сукупності і визначається відносною величиною або відсотком.

Наприклад, визначимо в таблиці 2.2 частку студентів, які мають зріст 150-158 см:

$$d_i = \frac{f_i}{\sum f_i} = \frac{6}{25} = 0,24 \text{ або } 24 \%$$

Загальна сума часток у ряді розподілу завжди дорівнює “1”.

**Кумулятивна частота (частка)** – це накопичена частота (частка), яка є звичайним підсумком частот (часток). Для першого інтервалу накопичена частота збігається з її частотою (часткою), для другого – це підсумок частот (часток) першого та другого інтервалів, для третього – підсумок частот (часток) першого, другого та третього інтервалів і под. Значення кумулятивної частоти для останнього інтервалу буде збігатися із загальним обсягом сукупності. Так для таблиці 2.2 кумулятивна частота для другого інтервалу (150-158 см) – це сума частот першого та другого:  $6+7=13$ , для третього інтервалу до одержаної суми додається частота третього інтервалу:  $13+5=18$  і под.

### **Комбінаційні групування**

Такі групування є комбінацією кількох (як правило – двох) ознак, кожна з яких подається звичайним групуванням. У цьому випадку групи, що виділені за однією ознакою, поділяються на підгрупи за значеннями другої ознаки, третьої ознаки і под. В результаті одержуємо таблицю, де рядки подаються одним групуванням, а стовпчики – іншим, а клітинки містять значення, що належать одночасно певним групам як в одному, так і в іншому групуванні.

У прикладних статистичних дослідженнях комбінаційні групування часто використовуються для аналізу взаємозв'язків між ознаками. Такі групування містять комбінаційні розподіли сукупностей за двома ознаками – факторною  $x$  і результативною  $y$  і мають назву *таблиці взаємної спряженості* (їх також називають *таблицями кростабуляції*). Якщо *стохастичний* зв'язок між ознаками існує, то розподіли змінюються від групи до групи. Аналіз зв'язків буде розглянуто у наступних частинах посібника.

Таблиця 2.4

**Розподіл населення регіону за віковими групами та за статтю**

<i><b>Вікові групи (повних років)</b></i>	<i><b>Чисельність жінок</b></i>	<i><b>Чисельність чоловіків</b></i>
До 1 року	97	93
1-5	395	379
...	...	...

### **Статистичні таблиці**

Результати статистичного зведення подаються у вигляді *статистичних таблиць*, при побудові яких прийнято дотримуватися певних правил. Можна виділити основні правила:

1. Таблиця повинна мати загальний заголовок, бокові та верхні заголовки. Вони повинні бути сформульовані за можливістю стисло, але при цьому зрозуміло розкривати зміст таблиці.

2. Числові дані для певного стовпчика повинні мати однакову точність представлення, наприклад, «2» знаки після коми, незважаючи на те, що деякі дані можуть мати меншу кількість десяткових знаків або взагалі не мати дробової частини.

3. Якщо клітинка не заповнюється, то в ній проставляється символ «х».

4. Якщо відомості відсутні, то у клітинці проставляється тире.

5. Таблиця повинна мати підсумковий рядок.


## **STATISTICA. Частотний аналіз (групування)**

Для побудови групування слід виконати такі дії:

1. Виділити змінну (або змінні), для якої необхідно здійснити групування. Зауважимо, що вибір змінної можна буде зробити і пізніше. Якщо буде виділено кілька змінних, то для кожної змінної у робочій книзі буде створено окремий аркуш з групуванням.

2. Завантажити модуль “Bases Statistics/Tables” (Основні статистики і таблиці). Це може бути здійснено кількома способами:

- Виконати команду **Statistics ▶ Bases Statistics/Tables**.

- На панелі “Statistics” натиснути кнопку  «**Bases Statistics/Tables**».

- Натиснути кнопку  “Start menu of commonly used tools” і вибрати з меню **Statistics ▶ Bases Statistics/Tables**.

З’явиться вікно “Basic Statistics and Tables:”.

3. Вибрати у поточному вікні пункт “Frequency tables” (Таблиці частот) і натиснути «**OK**», що приведе до появи однойменного вікна.

4. Якщо змінна групування раніше не була вибрана, то натискаємо кнопку «**Variables**» (змінні) і вибираємо групувальну ознаку.

5. Групування даних здійснюється за параметрами, що знаходяться у вікні модуля на вкладці “Advanced” (Розширені) у групі “Categorization methods for tables & graphs” (Методи категоризації для таблиць і графіків). За замовчуванням створюється окрема група для кожного значення змінної (тобто будується дискретний ряд розподілу), що визначається встановленням значення перемикача в положення “All distinct value” (Всі відмінні значення). При цьому для описових ознак встановлення прапорця поруч з полем-міткою “with text label” (з текстовою міткою) дозволяє вивести назви груп (текстові мітки) змінної.

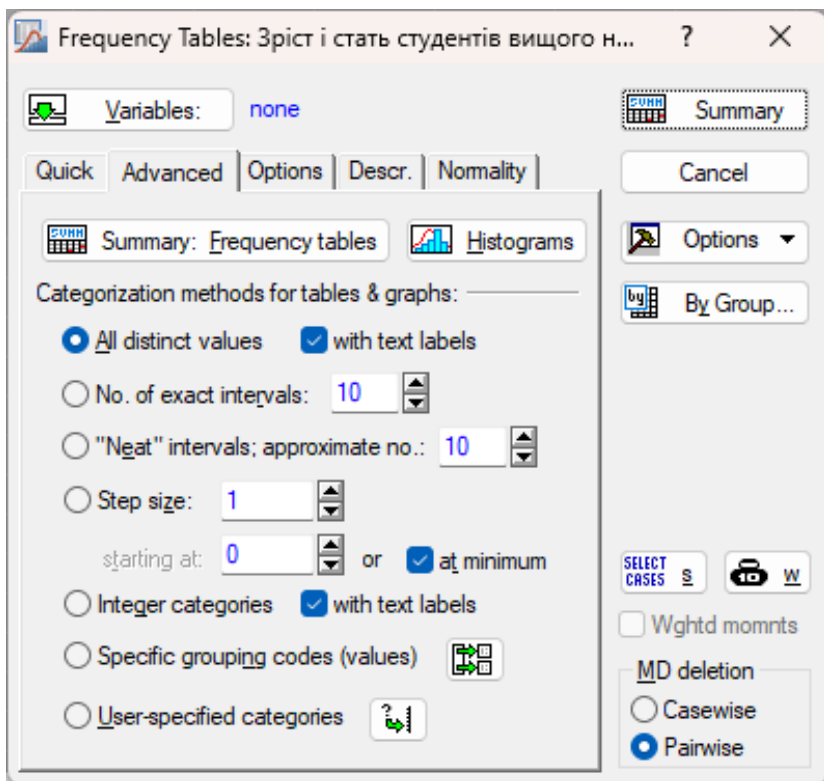



Рисунок 2.1. Вікно модуля «Частотний аналіз»

Але можна застосувати і інші параметри побудови групування:

- Якщо необхідно побудувати групування *тільки* для *цілих* значень змінної, *відкидаючи* спостереження, що містять дробові значення змінної, то слід встановити перемикач вибору методу групування в положення “Integer Categories” (Цілочисельні категорії).

- Для неперервних ознак групування будується, як правило, за інтервалами. При цьому користувач має можливість використовувати такі параметри:

- 5.1. Задати кількість рівних інтервалів в полі “No. of exact intervals” (Кількість точних (рівних) інтервалів).
- 5.2. Побудувати групування, де межі інтервалів є “круглими” (кратними десяти) значеннями, для чого використовується параметр “Neat” intervals; approximate no”.  При цьому слід мати на увазі, що у такий спосіб фактична кількість інтервалів може відрізнятись від заданої кількості.
- 5.3. Групування можна здійснити шляхом визначення ширини інтервалу, встановивши перемикач в положення “Step size” (крок шагу) і задати в полі праворуч від нього крок інтервалу. Початкове значення автоматично визначається системою як мінімальне з усіх значень, але й може бути задано користувачем в полі “starting with” (почати з). Для застосування першого варіанту достатньо встановити прапорець поруч з полем-міткою “at minimum” (з мінімального значення), а для другого – ввести потрібне значення. Якщо це значення не нульове, то для завдання початкового значення застосуйте один з цих варіантів.

6. За замовчуванням у модулі розраховуються групові і кумулятивні частоти, та проценти. Інші показники, що можна розраховувати, визначаються на вкладці “Options”.

7. Натискання кнопки «**Summary**» (Підсумки) ініціює побудову таблиці частот. Те саме можна зробити також на вкладці “*Quick*” (швидкий) натиснувши кнопку «**Summary**», натискання ж кнопки «**Histograms**» ініціює графічну побудову ряду розподілу.

Побудуємо групування для розподілу студентів навчальної групи за ступенем орієнтованості на педагогічну роботу з прикладу 2.1.



Завантажимо модуль “Bases Statistics/Tables” і виберемо у вікні модуля пункт “Frequency tables” (Таблиці частот). Незважаючи на те, що таблиця містить тільки одну змінну, система буде вимагати вибрати змінну для групування. Це «Оцінка студента», що має «цілий» тип даних. Оскільки за замовчуванням створюється окрема група для кожного значення змінної, то залишаємо всі значення такими, якими вони є за замовчуванням. Оскільки ознака не є описовою, то прапорець поруч з полем -міткою “with text label” можна зняти, але нічого не зміниться, якщо цього і не робити.

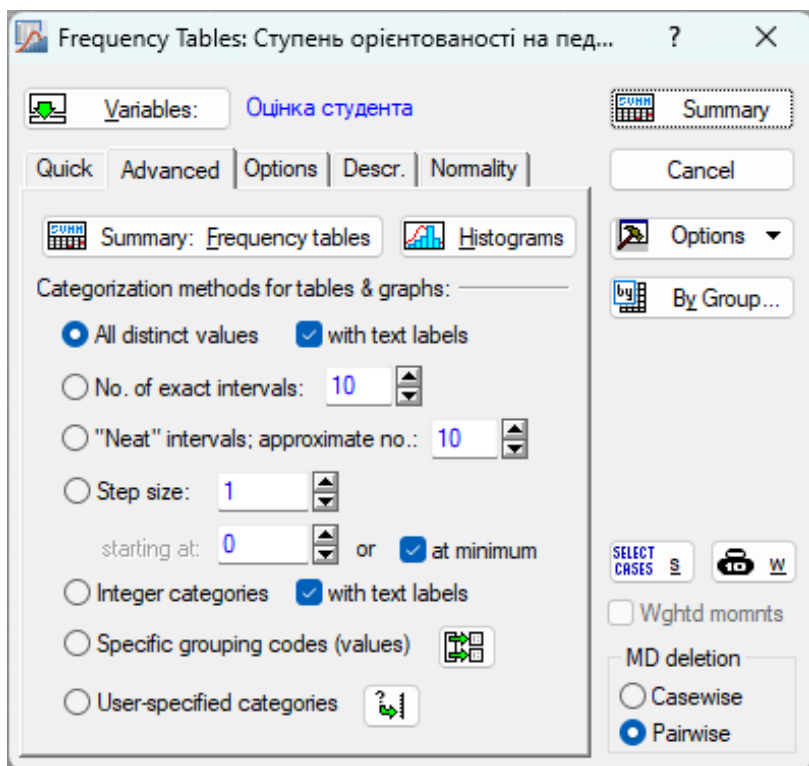
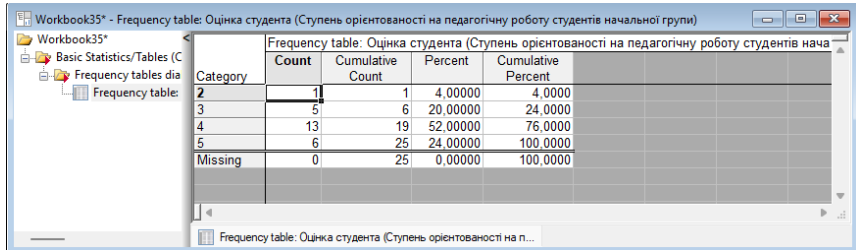


Рисунок 2.2. Вікно вибору параметрів модуля «Частотний аналіз» для дискретного ряду розподілу

Натиснувши кнопку «**Summary**» одержуємо таблицю частот.



Category	Count	Cumulative Count	Percent	Cumulative Percent
2	1	1	4.00000	4.00000
3	5	6	20.00000	24.00000
4	13	19	52.00000	76.00000
5	6	25	24.00000	100.00000
Missing	0	25	0.00000	100.00000

Рисунок 2.3. Результати модуля «Частотний аналіз» для дискретного ряду розподілу

Одночасно розраховуються кумулятивні частоти, відсотки і накопичені відсотки.

Побудуємо групування за даними таблиці 2.2 для розподілу студентів навчальної групи університету за зростом, в якій первинні дані згруповані за інтервалами.

Потрібні для цього основні налаштування знаходяться на вкладці “*Advanced*”.

Отже, необхідно побудувати групування з п’яти рівних інтервалів. Оскільки крок інтервалу дорівнює «8», встановлюємо перемикач в положення “*Step size*” (крок інтервалу) і задаємо в полі праворуч від нього значення «8». Оскільки початкове значення першого інтервалу становить «150», то для визначення нижньої границі першого інтервалу можна або ввести це вже відоме мінімальне значення в полі “*starting with*” (почати з) або перекласти визначення мінімального значення на систему, встановивши прапорець поруч з полем-міткою “*at minimum*” (з мінімального значення).

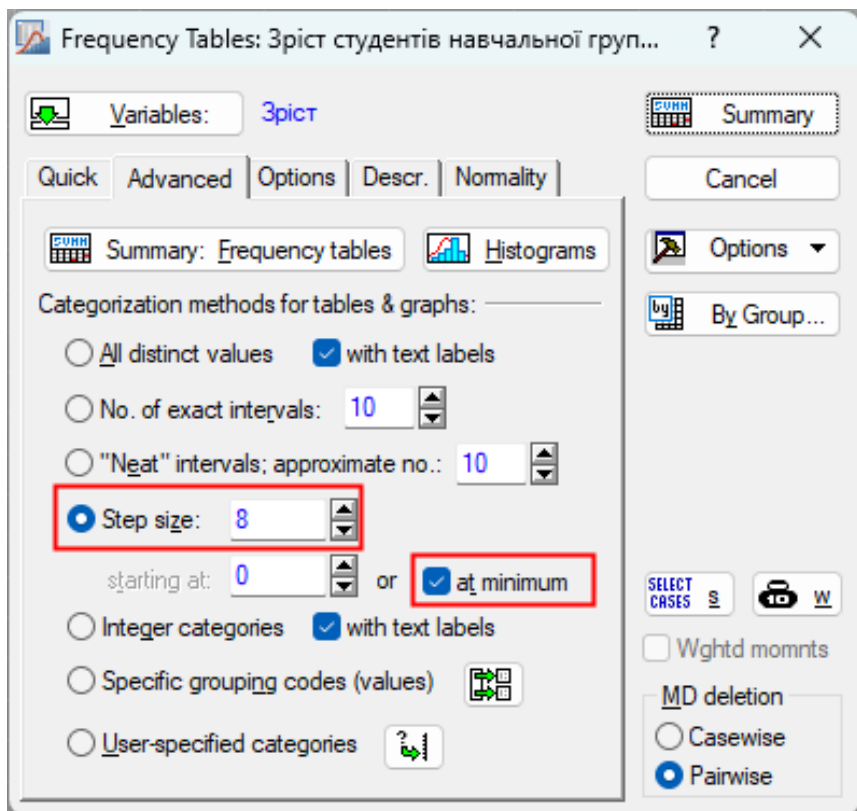


Рисунок 2.4. Вікно вибору параметрів модуля «Частотний аналіз» для інтервального ряду розподілу


Натиснувши кнопку «**Summary**», одержуємо таблицю частот.

From	To	Count	Cumulative Count	Percent	Cumulative Percent
150.00	<=x<158.00	5	5	20.00000	20.00000
158.00	<=x<166.00	7	12	28.00000	48.00000
166.00	<=x<174.00	6	18	24.00000	72.00000
174.00	<=x<182.00	4	22	16.00000	88.00000
182.00	<=x<190.00	2	24	8.00000	96.00000
190.00	<=x<198.00	1	25	4.00000	100.00000
Missing		0	25	0.00000	100.00000

Рисунок 2.5. Результати модуля «Частотний аналіз» для інтервального ряду розподілу

Як бачимо, одержана таблиця збігається з таблицею 2.3, оскільки побудова групувань у *STATISTICA* здійснюється так, що нижня межа закритого інтервалу буде включати значення «включно», а верхня – «виключно», що є прийнятним в американській статистиці.

Якщо ж ми бажаємо побудувати групування за принципом верхня межа – це включно, нижня – це виключно, то необхідно виконати спеціальні дії

1. На вкладці “*Advanced*” встановлюємо перемикач в положення “*User-specified categories*” (Визначені користувачем категорії (інтервали) і натискаємо кнопку , що знаходиться праворуч від перемикача, в результаті з’явиться вікно “*Define Categories*” (Визначити групи (інтервали)).

2. Залишаємо правило для всіх категорій “*Include if:*” (включати до групи якщо). Формуємо умови для кожної категорії, в якості яких і будуть наші інтервали. Наприклад, перший інтервал «150-158» описуємо як  $v1 \leq 158$ , другий інтервал «158-166»:  $v1 > 158 \& v2 \leq 166$ , а останній інтервал «182 і більше» –  $v1 > 182$ .

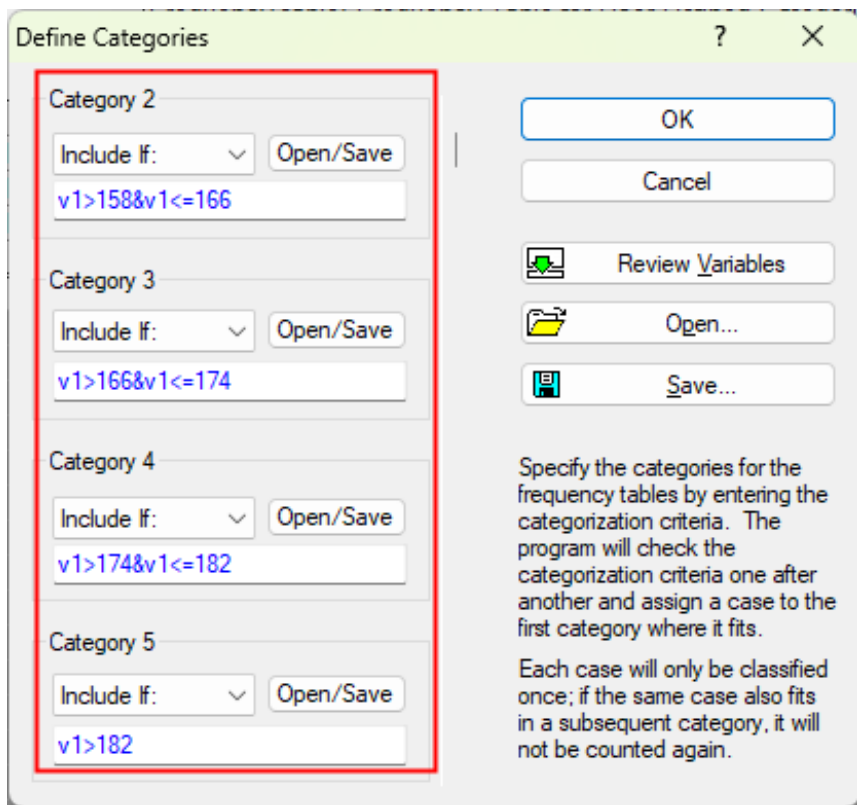


Рисунок 2.6. Формування користувачем власного групування



Такий варіант дозволяє створювати групування будь-якої складності, використовуючи як закриті, так і відкриті інтервали, як рівні так не рівні інтервали.

3. Натискаємо кнопку «**ОК**» і одержуємо таблицю частот, що збігається з таблицею 2.2.

Category	Frequency table: Frequency Table for User-Defined Ca			
	Count	Cumulative Count	Percent	Cumulative Percent
Include v1<=158	6	6	24,00000	24,0000
Include v1>158&v1<=166	7	13	28,00000	52,0000
Include v1>166&v1<=174	5	18	20,00000	72,0000
Include v1>174&v1<=182	5	23	20,00000	92,0000
Include v1>182	2	25	8,00000	100,0000
Not selected	0	25	0,00000	100,0000

Рисунок 2.7. Результати модуля «Частотний аналіз» для визначеного користувачем групування

Якщо необхідно буде і надалі будувати групування із саме такими інтервалами, то можна запам'ятати створені умови. У вікні “Define Categories” (Визначте категорії) натискаємо кнопку «Save» і зберігаємо умови в файлі який розміщуємо у потрібному місці. Файл зберігається з розширенням INI.

У подальшому для застосування збережених умов у вікні “Define Categories” достатньо просто натиснути кнопку «Open» і вибрати збережений файл.

### 3. Статистичні графіки

#### Теорія

Людині притаманно краще сприймати інформацію візуально, ніж у вигляді статистичних таблиць, а тим більше у вигляді незгрупованих первинних даних. Багато закономірностей важко виявити безпосередньо за даними, проте вони чітко можуть проявлятися з графіка. Класичним прикладом цього є наявність на гістограмі двох піків (*бімодальність* розподілу), що, як правило, є свідченням неоднорідності сукупності.

*Побудова графіків*, тобто здійснення *візуалізації даних* дуже часто передує безпосередньому застосуванню статистичних методів, а іноді взагалі є обов'язковим етапом дослідження.

Функцію візуального відображення статистичних даних відіграють статистичні графіки та діаграми. **Статистичний графік** – це засіб подання статистичних даних за допомогою різноманітних графічних знаків та символів. Незважаючи на велику кількість типів графіків, усі вони містять п'ять елементів:

1. *Поле графіка*. Площина, де розташовується графік.
2. *Графічний образ*. Безпосередньо сам графік, побудований за допомогою ліній, точок та інших графічних знаків.
3. *Просторові орієнтири*. Вони визначають порядок розміщення графічних знаків. Цей порядок визначається системою координат. Як правило, це прямокутна (декартова) система координат. Разом із тим, для відображення даних, яким притаманна циклічність (сезонність), використовують полярну систему координат.
4. *Масштабні орієнтири*. Це масштаб, масштабна шкала і масштабний знак.
5. *Експлікація*. Загальна назва графіка, підписи уздовж осей та інші словесні пояснення щодо графіка.

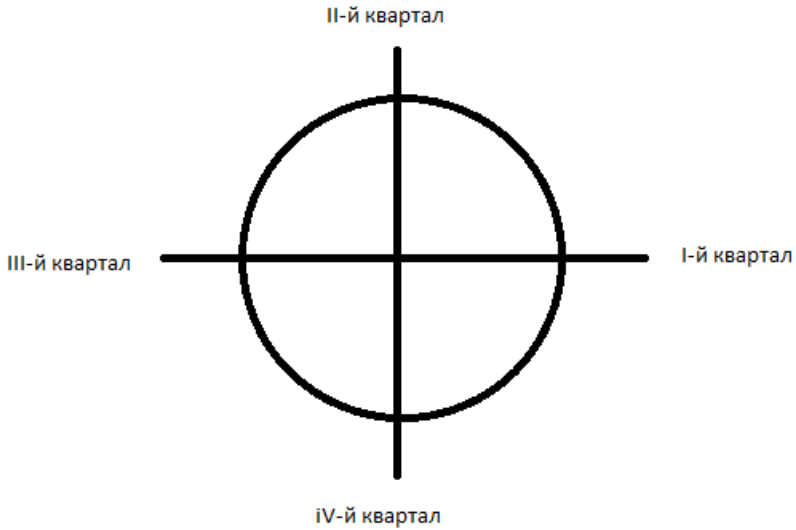


Рисунок 3.1. Полярна система координат

Графіки широко використовуються для зображення даних рядів розподілу. Основна мета такого зображення – надати уяву щодо структури сукупності, характер і форму її розподілу. Найпростішими видами такого роду графіків є однорядкові та одностовпчикові. Основних їх призначенням є відображення структури сукупності.

### **Типи графіків**

1. *Полігон (лінійна діаграма або лінійний графік, графік залежності X-Y, Line Char)* – графік за точками, коли точки з'єднуються прямими лініями, а самі точки (вузли) не виділяються. Використовується для представлення дискретних рядів розподілу.



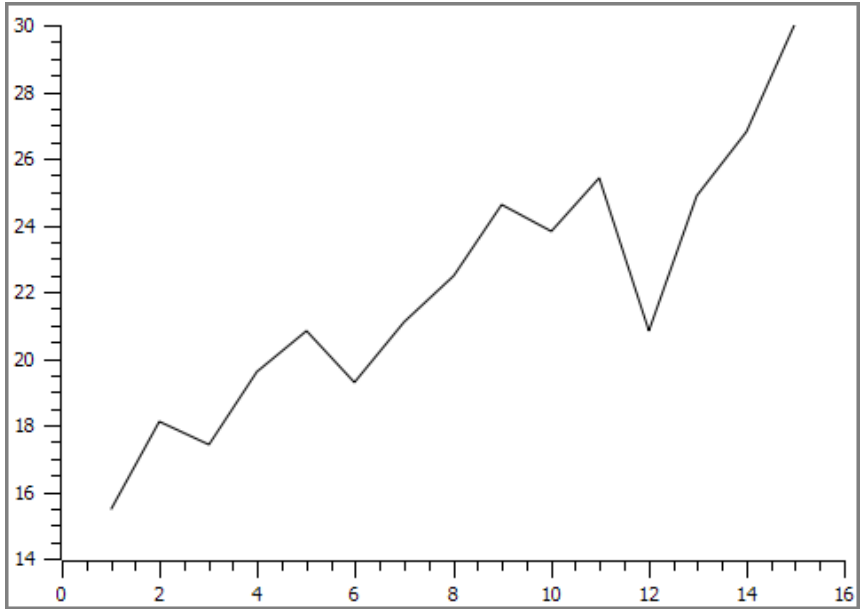


Рисунок 3.2. Полігон

2. *Діаграма розсіяння (діаграма кореляції, поле кореляції, кореляційне поле, Scatter Plot)* – графік за точками, коли точки не з’єднуються. Використовується для відображення даних для двох змінних, одна з яких є факторною, а інша – результативною. За допомогою даної діаграми здійснюється графічне подання пар даних у вигляді множини точок («хмари») на координатній площині. Докладніше про цей тип графіка дивись далі.

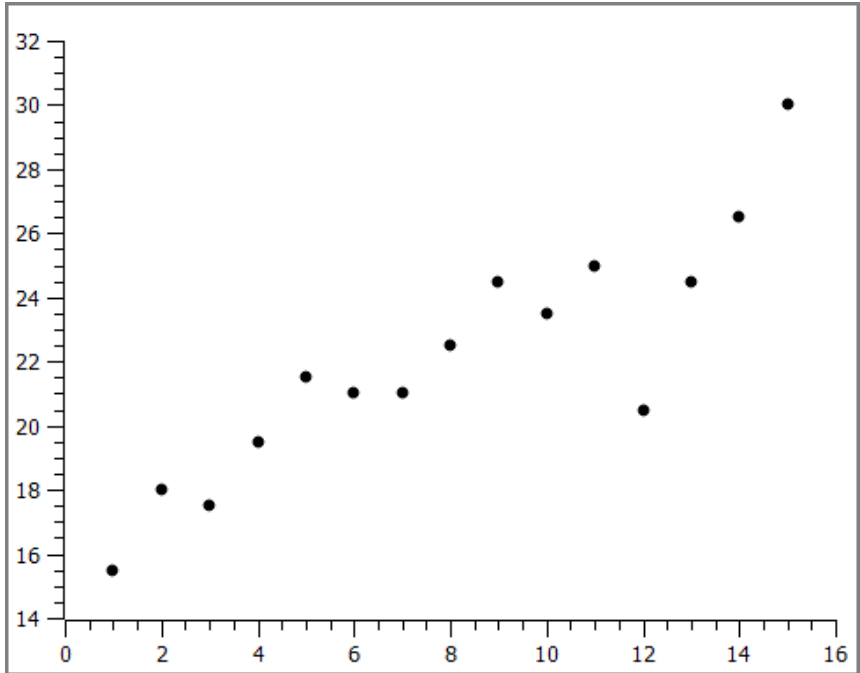


Рисунок 3.3. Діаграма розсіяння

3. *Лінійний графік трасувального типу (Line+Symbol).*  
Графік за точками, коли точки виділяються і з'єднуються прямими лініями.

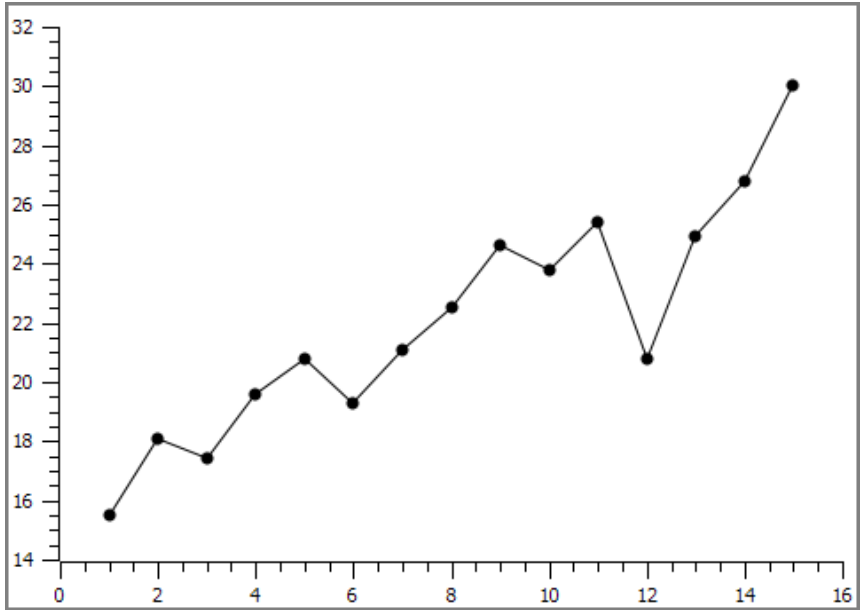


Рисунок 3.4. Лінійний графік

4. *Графік сплайна (Spline)*. Графік за точками, коли точки з'єднуються сплайном, зазвичай це агрегатна функція, що збігається з функціями простішої природи на кожній ділянці розбиття області визначення графіка.

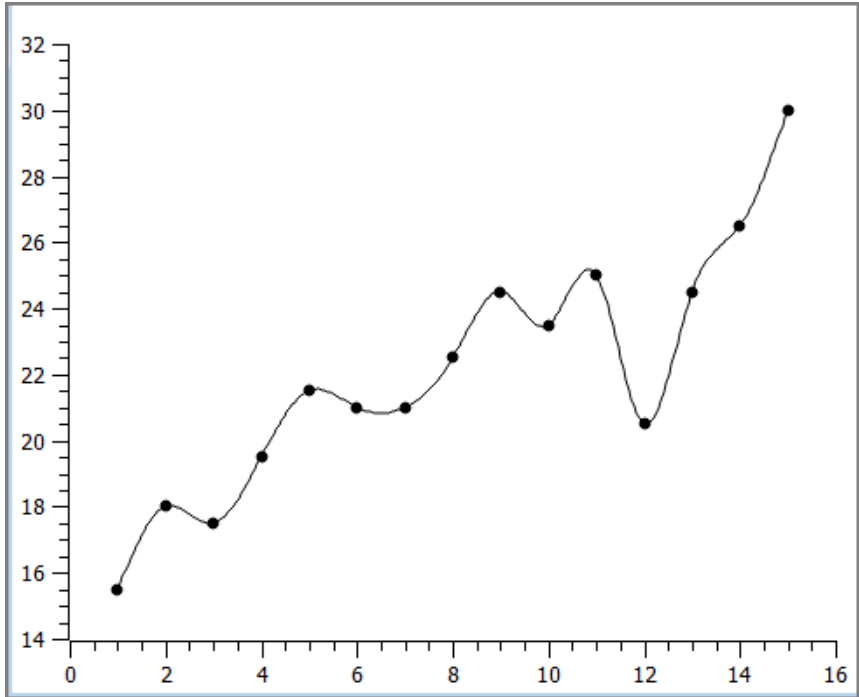


Рисунок 3.5. Сплайн

5. *Ступінчата вертикальна діаграма (Vertical Steps)*. Точки з'єднуються горизонтальними і вертикальними лініями, внаслідок чого створюються сходинки. Якщо повернути її на 180 градусів, то одержимо *ступінчасту горизонтальну діаграму (Horizontal Steps)*.

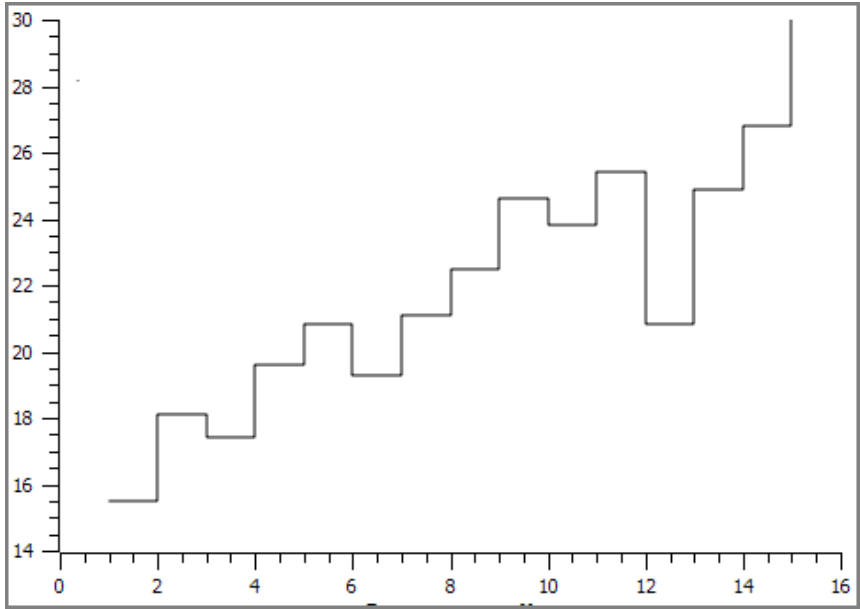


Рисунок 3.6. Ступінчата вертикальна діаграма

6. *Стовпчикова діаграма (Vertical Bar)* – графіки у вигляді вертикальних стовпчиків.

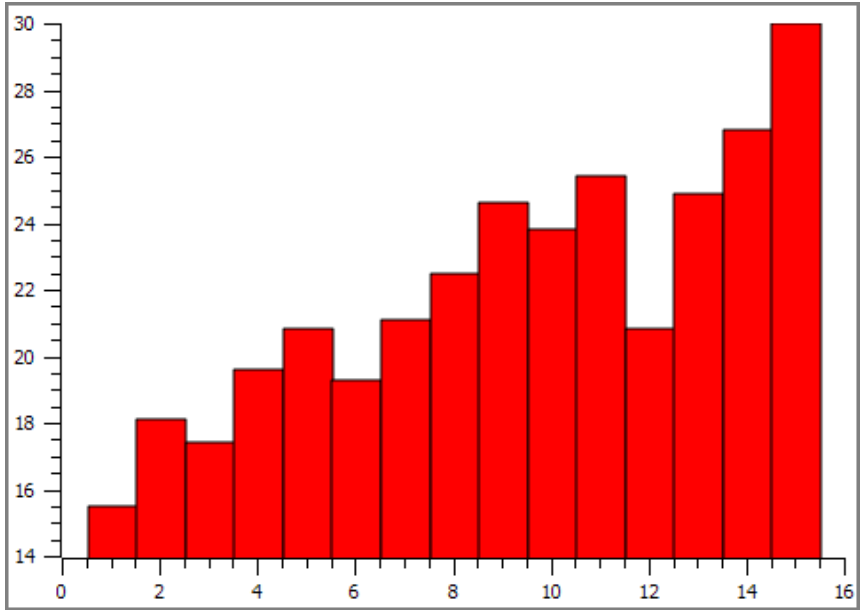


Рисунок 3.7. Столпчикова діаграма

7. Різновид столпчикової діаграми – графік у вигляді вертикальних ліній, що закінчуються у точках графіка, які позначаються колом (*Vertical Drop Lines*).

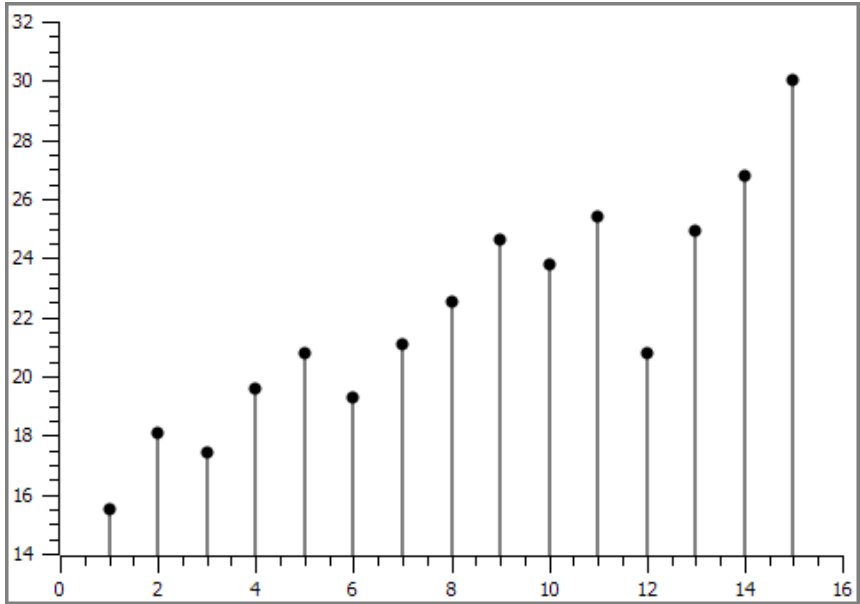


Рисунок 3.8. Графік у вигляді вертикальних ліній

8. *Стрічкова діаграма (Horizontal Bar)* – графіки у вигляді горизонтальних стрічок.

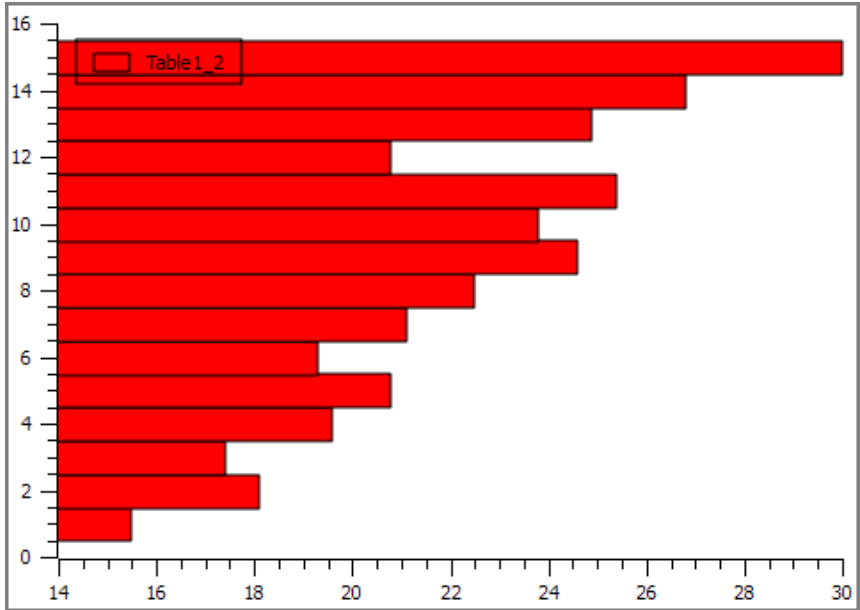


Рисунок 3.9. Стрічкова діаграма

9. *Діаграма-область* (Area). Область під графіком зафарбовується або заштриховується.



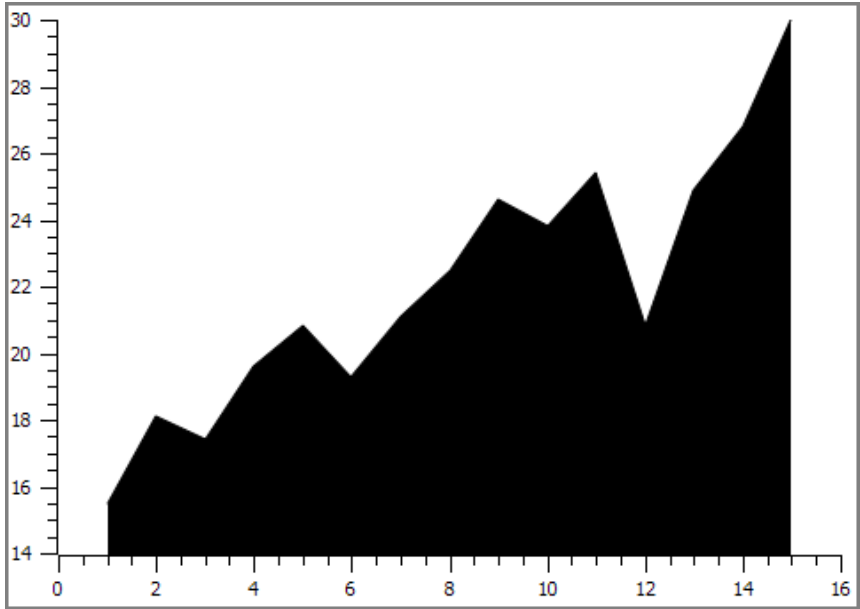


Рисунок 3.10. Діаграма-область

10. *Секторна* або *кругова діаграма (Pie Chart)*. Така діаграма поділена на кілька сегментів-секторів, площа кожного з яких пропорційна їх частці.

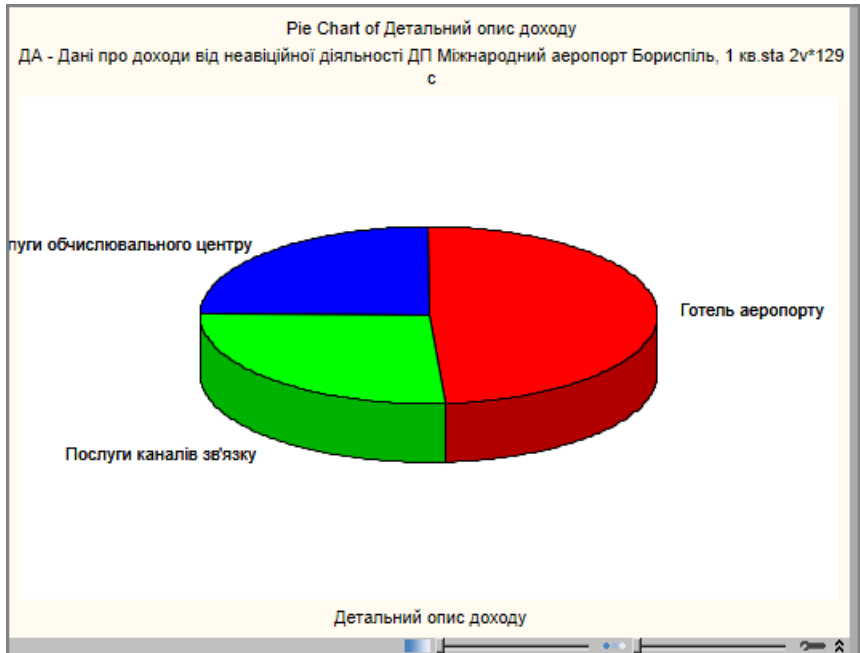


Рисунок 3.11. Секторна (кругова) діаграма

11. Набір векторів з можливістю задавати координати початку і кінця вектору, також можна задавати координати початку, його довжину і кут повороту.

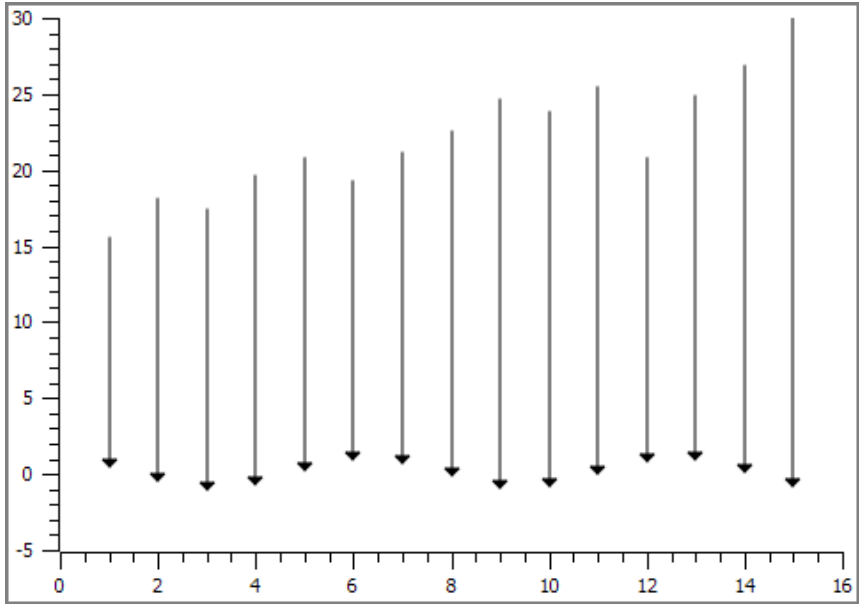


Рисунок 3.12. Діаграма у вигляді набору векторів

12. *Коробкова діаграма (коробка з вусами, скринька з вусами, Box Plot, box-and-whisker diagram)*. Інформація такого графіка є дуже змістовною і корисною. Він водночас відображує кілька величин, що характеризують ряд розподілу: мінімальне та максимальне значення, середню або медіану, перший кuartиль (або 25 процентиль) та третій кuartиль (75 процентиль). Отже коробкові діаграми не тільки відображують основні характеристики розподілу, але можуть бути використані також і для оцінки розмаху варіації та асиметрії. Докладніше про цей тип графіка дивись далі.

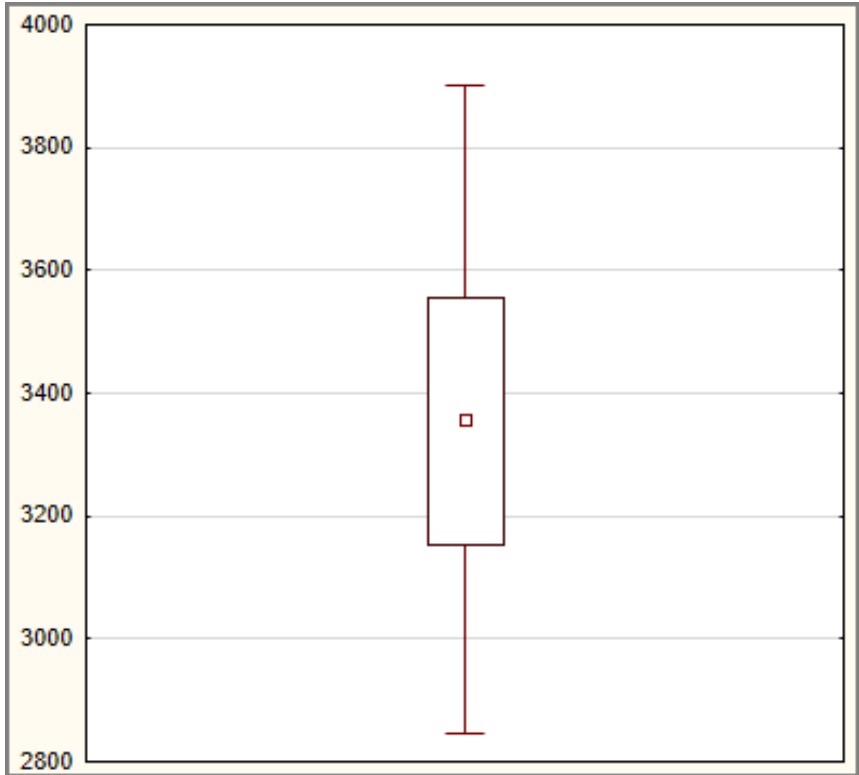


Рисунок 3.13. Коробкова діаграма

13. *Гістограма (Histogram)*. Використовується для представлення даних інтервальних рядів розподілу. Під час її побудови на осі ординат відкладають частки або частоти, а на осі абсцис – значення ознаки. Докладніше про цей тип графіка дивись далі.

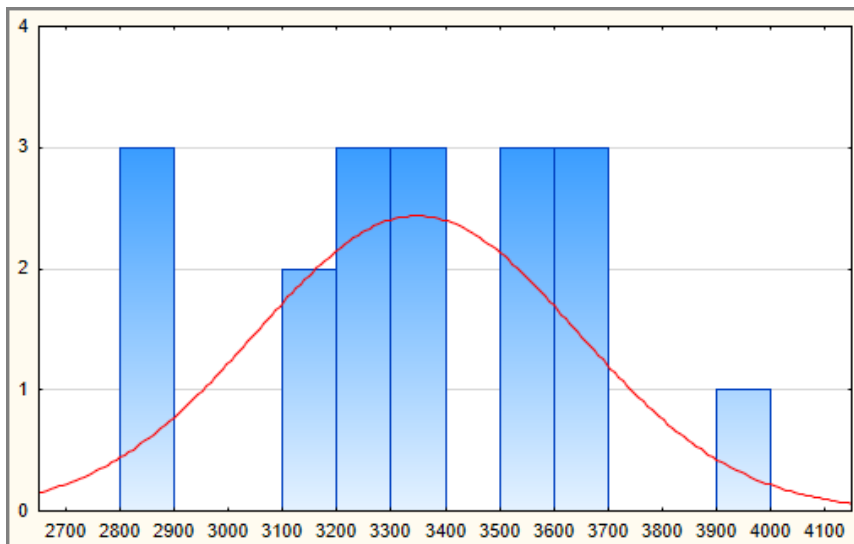


Рисунок 3.14. Гістограма

### ***Вибір типу графіка***

Під час побудови графіка перед дослідником постає питання вибору потрібного типу, оскільки правильно вибраний тип допомагає прийняти правильне рішення, а неправильно – може привести до хибних висновків.

Вибір типу графіка залежить від мети дослідження. На наш погляд, можливі такі варіанти..

1. Для *порівняння*, наприклад, цін на однакові товари або рейтингу кандидатів на вибори від різних партій. З цією метою можна застосувати:

- стрічкову діаграму;
- стовпчикову діаграму;
- діаграму розсіювання;
- кругова діаграму;
- полігон;

2. Для відображення *розподілу даних, виявлення аномальних відхилень* і та ін. Наприклад, розподіл часу очікування відповіді в Call-центрі, розподіл помилок, виявлених під час тестування програмного забезпечення. З цією метою можна застосувати:

- гістограму;
- стовпчикову діаграму;
- діаграму розсіювання;
- полігон;
- коробкову діаграму.

3. Для відображення *структури явища, тобто співвідношення частин одного цілого*. Наприклад, співвідношення жінок і чоловіків у місті. З цією метою можна застосувати:

- стовпчикову діаграму;
- кругова діаграму.

4. Для *вивчення тенденції, тобто динаміки зміни явища*. Наприклад, зміни обсягу виробництва певного товару на протязі кількох років. З цією метою можна застосувати:

- стовпчикову діаграму;
- лінійну діаграму.


5. Під час вивчення *відхилень значень від певного нормативу*. З цією метою можна застосувати:

- стрічкову діаграму;
- стовпчикову діаграму;
- лінійну діаграму.

6. Для *оцінки наявності або відсутності кореляції між двома величинами, тобто при вивченні взаємозв'язків*:

- діаграму розсіювання;
- лінійну діаграму.

## Побудова графіків у STATISTICA

Зважаючи на важливість візуального представлення даних, доступ до інструментів побудови графіків містить один з пунктів головного меню **Graphs** (Графіка). Дістатися до них можна також або з панелі “Graphs” або з меню кнопки  (засоби, що найчастіше використовуються).

Візуальні методи в STATISTICA діляться на категоризовані і не категоризовані. У категоризованих методах використовуються різні способи групування даних.

Слідуйте золотому правилу аналізу даних: групуйте дані, розбивайте їх на однорідні групи, завдяки чому їх закономірності стають більш очевидними.

Найпростішими і водночас найуживанішими статистичними графіками є *гістограма* і *діаграма розсіяння*.

### Гістограма

Гістограма є одним з найбільш вживаних видів графіків. Широке використання гістограми пояснюється її інформативністю: з її вигляду дуже просто визначити, які діапазони значень змінної зустрічаються найчастіше, наскільки сильно вони розрізняються між собою, наскільки щільно вони групуються біля середньої тощо.

Гістограма будується для *однієї* змінної.

Для побудови гістограми слід здійснити одну з дій.

1. Звернутися до модуля побудови гістограми:

- виконати команду **Graphs ▶ Histogram**;
- на панелі “Graphs” натиснути кнопку «**2D Histograms**»;
- натиснути кнопку засобів, що найчастіше використовуються, і вибрати з меню **Graphs ▶ 2D Graphs ▶ Histograms**.

Будь-яка з цих дій приведе до появи вікна “*2D Histograms*”.

2. Натиснути кнопку «**Variables**» (Змінні) і вибрати змінну (змінні).

3. За необхідності визначити загальний заголовок гістограми на вкладці “*Options 1*” слід зняти позначку біля поля-мітки “Display default title” (Відображати стандартний заголовок) і полі “Custom title” (Заголовок) ввести власний заголовок.

4. На вкладці “*Quick*” в полі “Categories” (Категорії) якщо слід визначити потрібні групи.

5. В разі потреби перевірки розподілу на нормальність необхідно виконати такі дії:

- На вкладці “*Advanced*” (Додатково) для розрахунку критеріїв перевірки на нормальність Колмогорова-Смірнова і Шапіро-Уїлка у групі “Statistics” встановити прапорець біля найменувань критеріїв.

- На вкладці “*Quick*” для відображення теоретичних частот вибрати пункт “Normal” у групі “Fit Type” (Тип підгонки).

6. Разом із гістограмою можна розрахувати також базові описові статистики. Для цього на вкладці “*Advanced*” у групі “Statistics” слід встановити позначку поруч з полем-міткою “Descriptive Statistics” (Описові статистики).



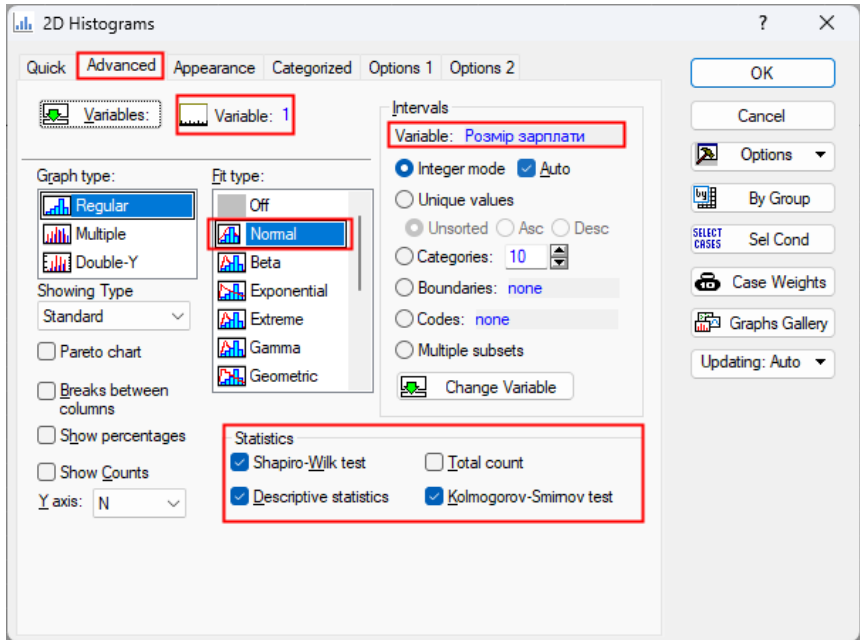


Рисунок 3.15. Вікно побудови гістограми

7. Натиснути кнопку «OK».

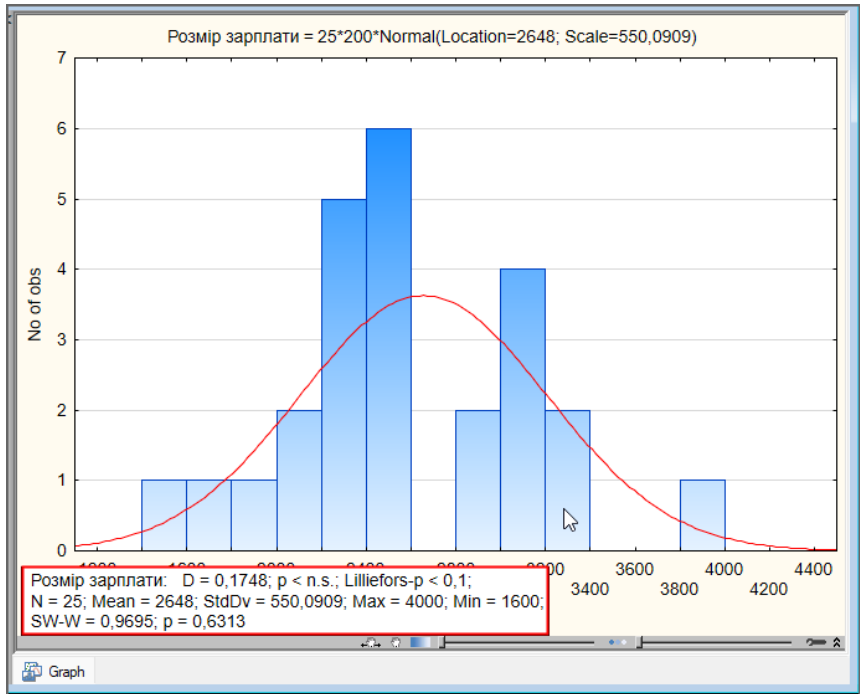


Рисунок 3.16. Гістограма

Для друку графіка слід викликати контекстне меню даного графіка і вибрати в ньому пункт “Print Graph” або натиснути стандартну для друку комбінацію функціональних клавіш <Ctrl>+<P>. В разі потреби графік можна зберегти. При цьому він може бути збережений як у спеціальному форматі *STATISTICA* з розширенням *STG*, так і у багатьох поширених графічних форматах, зокрема *GIF*, *JPEG*, *TIFF*, *BMP*.

Гістограму можна побудувати і в модулі “Descriptive Statistics:” (Описові статистики) аналізу “Bases Statistics/Tables” (Основні статистики і таблиці):

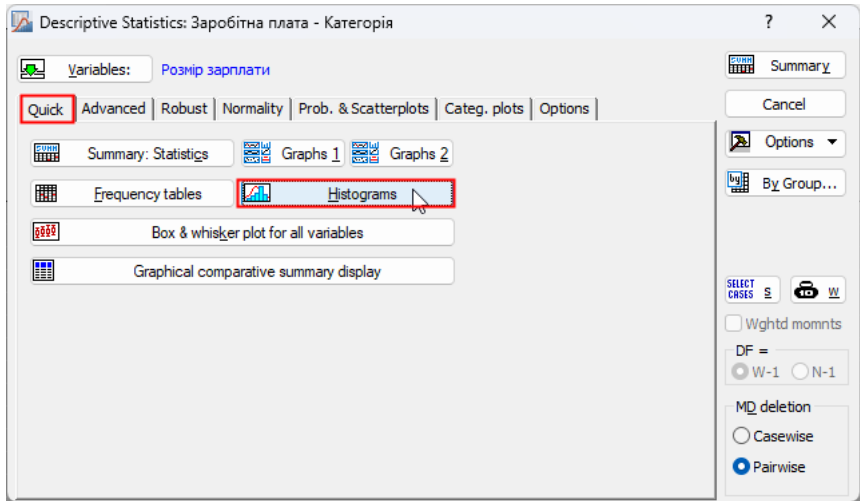


Рисунок 3.17. Доступ до побудови гістограми в модулі «Описові статистики»

### Діаграма розсіяння

Діаграма розсіяння (діаграма кореляції, поле кореляції, scatterplot) використовується для відображення даних двох змінних, одна з яких є чинником, а інша – наслідком. За допомогою діаграми розсіяння будується графічне відображення пар даних у вигляді множини точок («хмари») на координатній площині, що дозволяє оцінити зв'язок між двома змінними.

Алгоритм побудови графіка:

1. Здійснити одну з дій:

- Виконати команду **Graphs ▸ Scatterplots**.
- На панелі “Graphs” натиснути кнопку “2D Graphs”.
- Натиснути піктограму засобів, що найчастіше використовуються, і вибрати пункт меню **Graphs ▸ Scatterplots**.

Після виконання будь-якої з цих дій з'явиться вікно “2D Scatterplots”.

2. На вкладці “Quick” вибрати змінні, за якими будується графік. Для цього слід натиснути кнопку «Variables» (Змінні),

що призведе до появи вікна “Select Variables for Scatterplot” (Виберіть змінні для діаграми розсіяння). У лівому списку вікна вибирається факторна (незалежна) змінна, а у правому – результативна, що залежить від змінної-фактору.

3. На цій же вкладці знаходиться група показників “Regression bands” (діапазони регресії), за допомогою яких задається рівень довірчого інтервалу для лінії регресії.

4. На вкладці “Advanced” можна задати розрахунок коефіцієнта кореляції  $r$  і коефіцієнта детермінації  $r^2$ .

5. Закрити послідовно вікна натисканням «ОК», після чого відкриється вікно робочої книги з побудованим графіком.

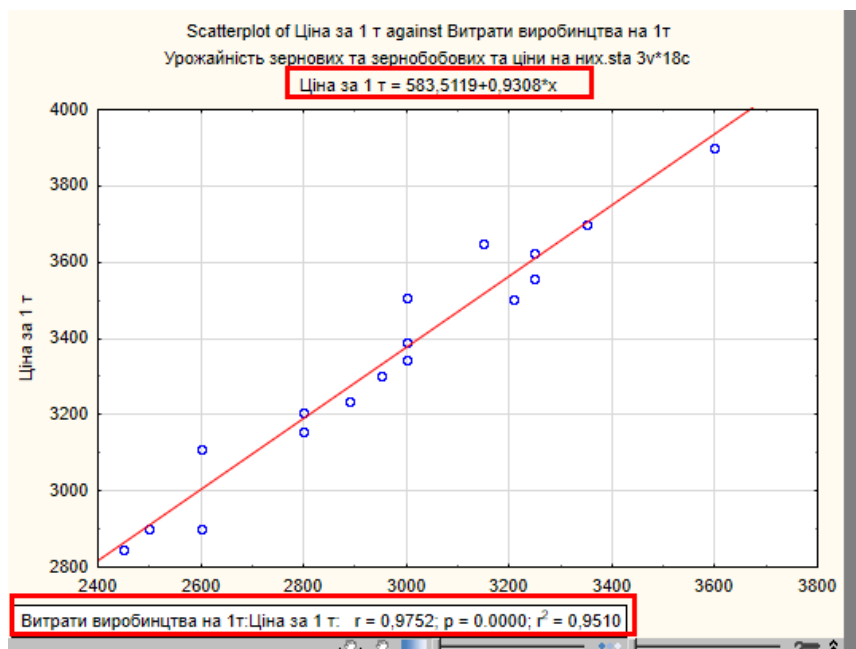


Рисунок 3.18. Діаграма розсіяння

Червона лінія регресії графіка дає уявлення щодо лінійності зв'язку між змінними: якщо точки близько

розташовані біля лінії, то можна говорити щодо наявності лінійної залежності між змінними.

У верхній частині графіка відображається рівняння регресії.

Спливаюче вікно у нижній частині вікна містить коефіцієнт кореляції  $r$  і коефіцієнт детермінації  $r^2$ .

### *Діаграма розмаху*

Таку діаграму ще називають *коробка з вусами*, *скринька з вусами* (*Box Plot*, *Box & Whisker Plot*).

Інформація графіка є дуже змістовною і корисною. У загальному випадку на ньому водночас графічно відображується кілька статистичних показників, що характеризують первинні дані: мінімальне та максимальне значення, середня або медіана, перший квартиль (або 25 перцентиль) та третій квартиль (75 перцентиль). Отже такі діаграми не тільки відображують основні характеристики розподілу, але можуть бути використані також і для оцінки розмаху варіації та асиметрії. Основою діаграми є вертикальний або горизонтальний прямокутник, нижній бік якого (або лівий бік, якщо прямокутник розташовано горизонтально) є першим квартилем ( $Q_1$ ), а верхній (правий) його бік є третім квартилем ( $Q_3$ ). Перший і третій квартилі також називають відповідно нижнім і верхнім квартилем. Таким чином висота (або довжина) прямокутника дорівнює міжквартильному інтервалу ( $IQR$ ). Невеличкий квадрат у площі прямокутника відображає середню арифметичну або медіану. Особливістю графіка є наявність «вусів» (*Whisker*), якими є вертикальні або горизонтальні лінії, довжина яких відповідає вибраному значенню показника розкидання даних (це може бути максимум і мінімум, стандартне відхилення, дисперсія, квартилі) або точності оцінки генеральних параметрів (стандартна похибка, довірчий інтервал).

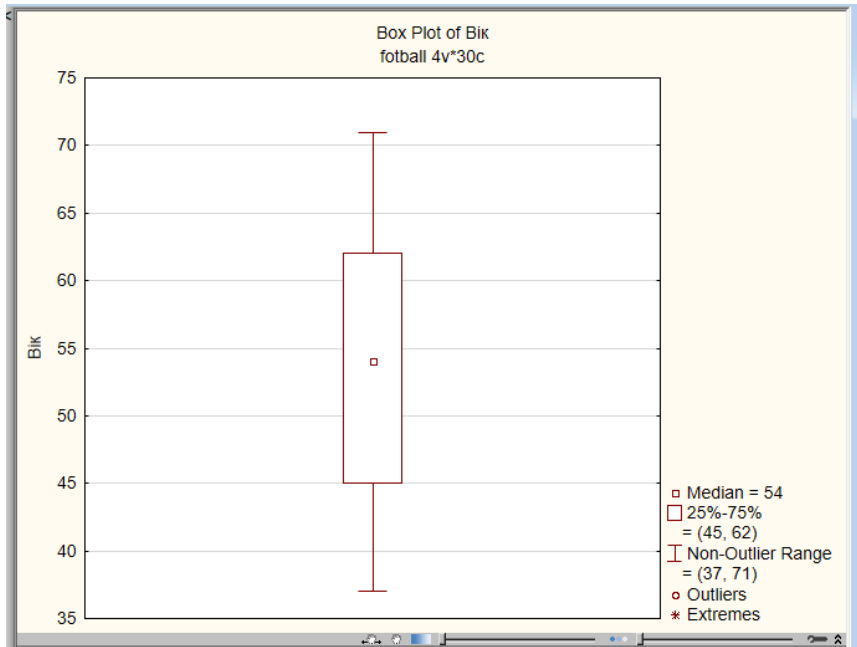


Рисунок 3.19. Діаграма розмаху

Діаграма використовується у багатьох статистичних методах, зокрема для візуалізації описових статистик, оцінки змін у часі або між різними групами, у кореляційному аналізі тощо. Відповідно її побудову можна здійснити у різних аналізах. Якщо ж будувати графік сам по собі, то необхідно виконати певні дії::

1. Звернутися до модуля побудови діаграми розмаху, наприклад за командою **Graphs ▶ 2D Graphs ▶ Box Plots....** Після чого з'явиться вікно "2D Box Plots" (2-вимірна діаграма розмаху).

2. Визначити змінну, за якою будується графік, для чого на вкладці "*Quick*" слід натиснути кнопку «**Variables**» (Змінні) і у вікні "Select Variables for Scatterplot" (Виберіть змінні для діаграми розсіяння) вибрати потрібну змінну, використовуючи

залежну змінну (“Depended Variable”). Але досить часто визначаються дві змінні: залежна і змінна, за якою здійснюється групування (Grouping Variable). Такий підхід зокрема використовується, коли необхідно об’єднати однотипні значення змінної за групами. Наприклад, залежною змінною може бути заробітна плата співробітників, а змінною, за якою здійснюється групування – категорії співробітників: керівники, службовці, робітники.

3. На вкладці “Quick” відображається невеличке зображення графіка із показником центру розподілу (середня арифметична (mean) або медіана (median) з тими значеннями показника розкидання даних, що будуть застосовані до «вусів» діаграми. У групі “Middle point” (Центральна точка) можна швидко змінити показники центру розподілу: зі списку “Value” (Значення) вибирається показник центру розподілу, а значення зі списку “Style” визначає, яким чином буде відображатися центр з графіка: у вигляді точки або лінії.

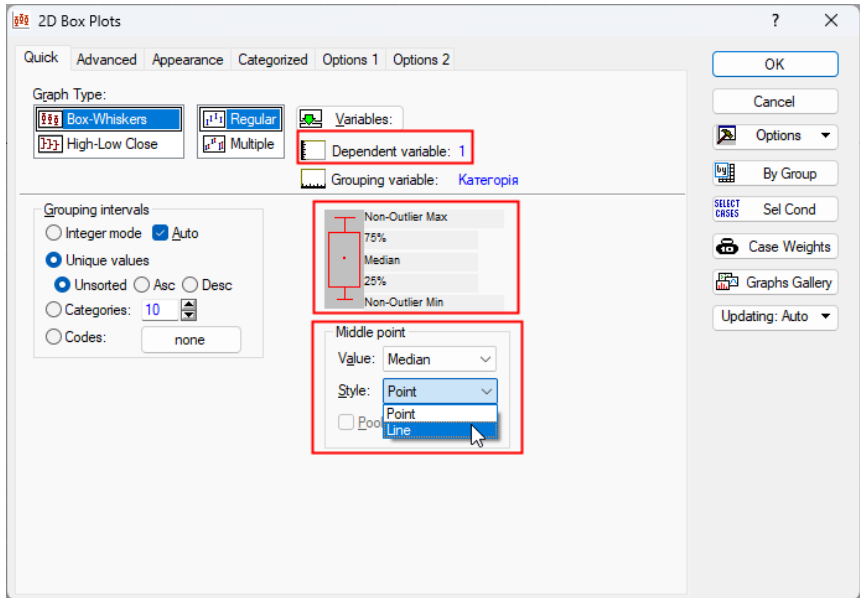


Рисунок 3.20. Вікно побудови діаграми розмаху



Вибір центру розподілу, а також інших елементів діаграми залежить від того, чи є він *нормальним*, а це в свою чергу впливає на перелік елементів діаграми.

Які саме значення слід вибрати демонструє наступна таблиця.



Таблиця 3.1

Значення графічних елементів діаграми розмаху

Графічний елемент	Значення графічного елемента залежно від типу розподілу	
	Нормальний	Відмінний від нормального
Вуси	Стандартна похибка / 95 % довірчий інтервал	Розмах / Non-Outlier Range
Коробка	Стандартне відхилення	Міжквартильний розмах
Центр	Середня арифметична	Медіана

4. Змінити показник центру розподілу можна також на вкладці “Advanced” (Додатково). На цій вкладці також знаходяться налаштування, за допомогою яких задаються інші елементи діаграми:

- Група “Box” (Розмах) призначена для визначення основи прямокутника («коробки») діаграми. За замовченням його нижній бік є 1-м квартилем ( $Q1$ ), а верхній є 3-м квартилем ( $Q3$ ), що визначається у списку “Value” (Значення) (значенням “Percentiles” (Процентілі) і числовим значенням «25» в полі “Coefficient”.

- Група “Whisker” (Вуси) (визначає статистичний показник, що буде задіяний з графіка у вигляді вусів. Він вибирається зі списку “Value”.

- Група “Outliers” (Викиди) дозволяє відображати або ні так звані точки-викиди, тобто значення змінної, що суттєво відрізняються у більший або менший бік порівняно з іншими значеннями вибірки. Для відключення відображення викидів у списку “Out. & Extremes” слід вибрати значення “Off”.

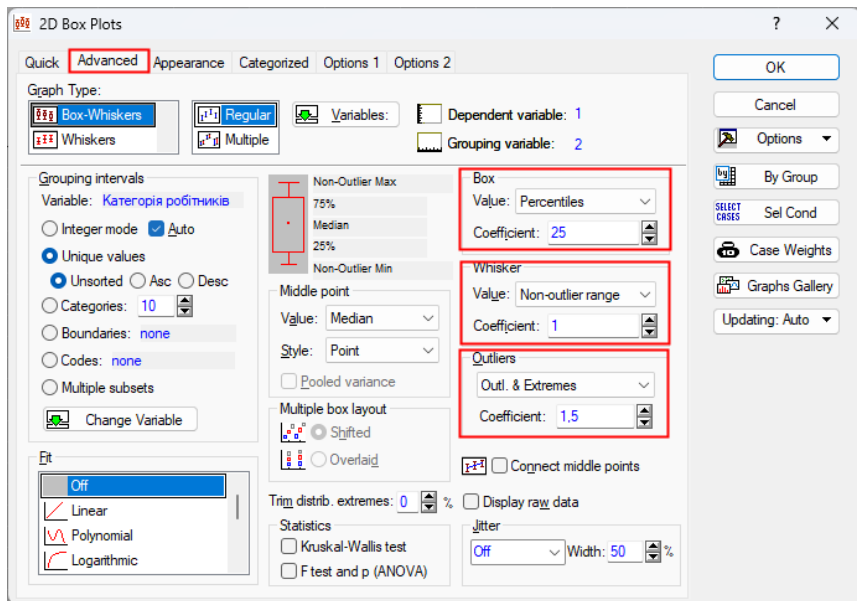


Рисунок 3.21. Вибір параметрів для діаграми розмаху

5. За наявності змінної, за якою здійснюється групування, можна обмежити кількість груп. Для цього на вкладці “Quick” у групі “Grouping intervals” (Інтервали групування) слід виконати такі дії:

- встановити перемикач для групи “Grouping intervals” в положення “Codes” (Коди);
- клацнути кнопку «None», після чого з’явиться вікно “Category Codes” (Коди категорій);
- для вибору груп натисніть у ньому кнопку «Zoom», що приведе до появи вікна з переліком груп;

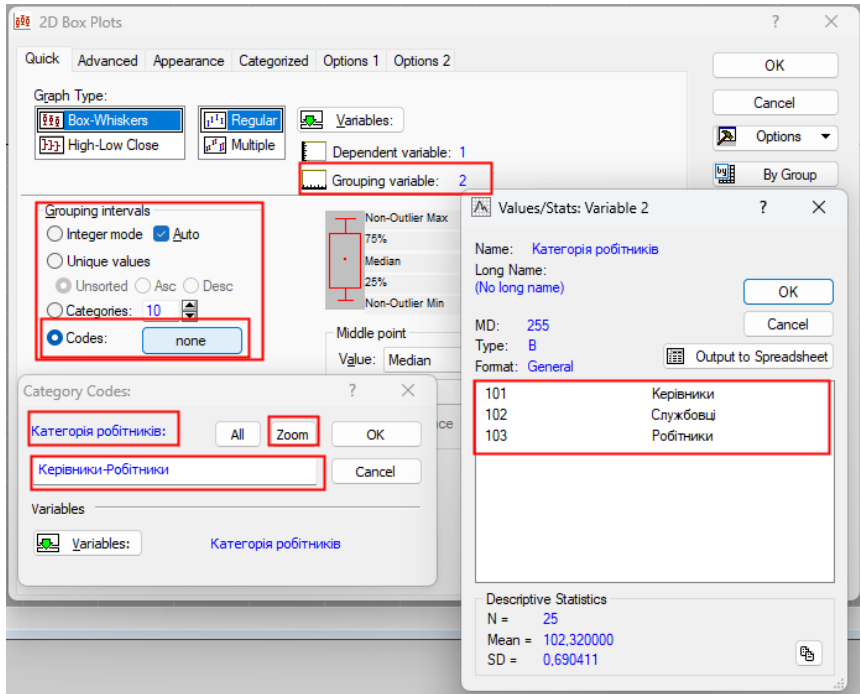


Рисунок 3.22. Вибір груп для діаграми розмаху

- здійснити вибір потрібних груп, натискаючи на їх найменуванні;
- послідовно закрити всі відкриті вікна натисканням кнопки «ОК».

6. Натиснути кнопку «ОК».

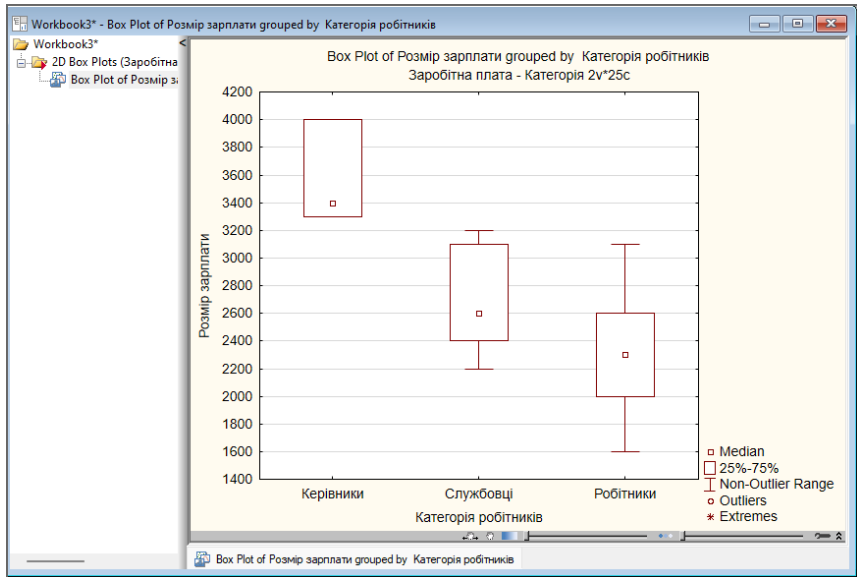


Рисунок 3.23. Діаграма розмаху с трьома групами

Як було зауважено вище, діаграма використовується у багатьох статистичних методах, зокрема для візуалізації описових статистик.

Побудувати її досить просто. Для цього на вкладці "Quick" слід натиснути кнопку «**Box & whisker plot for all Variables**» (діаграма розмаху для всіх змінних) і вибрати потрібні зміни. Але аналіз надає можливість побудувати діаграму для окремих груп (звісно, якщо вони є), для чого слід натиснути кнопку «**ByGroup**» (За групами) і вибрати у однойменному вікні потрібні групи.

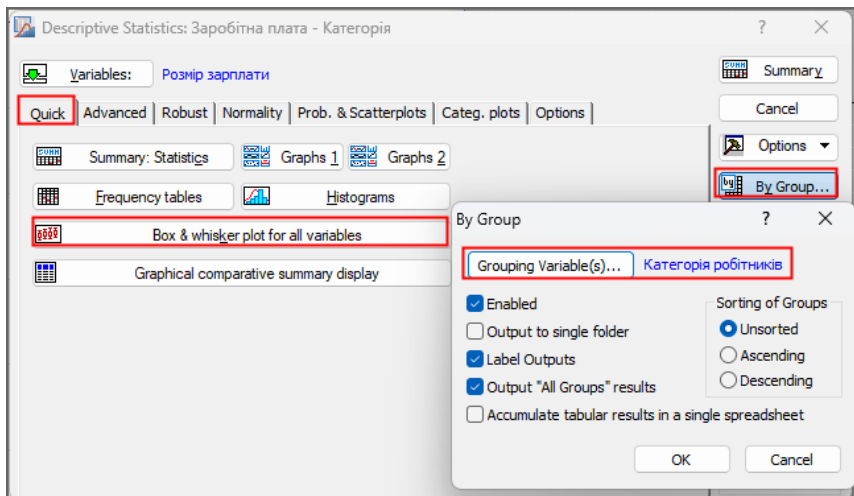


Рисунок 3.24. Доступ до діаграми розмаху в модулі «Описові статистики»

У програмі передбачено кілька форм діаграми розмаху, що визначається на вкладці "Options":

1. Медіана / Квартилі / Розмах
2. Середнє значення / Помилка середнього / Стандартне відхилення.
3. Середнє значення / Стандартне відхилення / Інтервал 1,96 стандартного відхилення.
4. Середнє значення / Помилка середнього / Інтервал 1,96 помилки середнього.

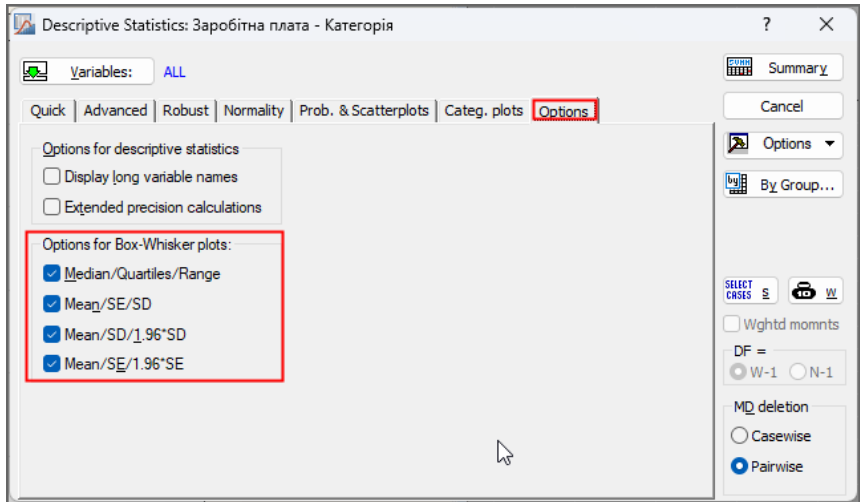


Рисунок 3.25. Визначення показників для діаграми розмаху

### *Інструменти роботи з графіками*

Вікно з графіком має власну панель інструментів, доступ до якої після побудови графіка здійснюється з пункту головного меню за командою **Edit** (Редагування). Вона містить інструменти для рисування, налаштування параметрів і зовнішнього вигляду графіка, інструменти керування об'єктами у вікні, наприклад, зв'язування, вбудовування графікові графічних об'єктів.

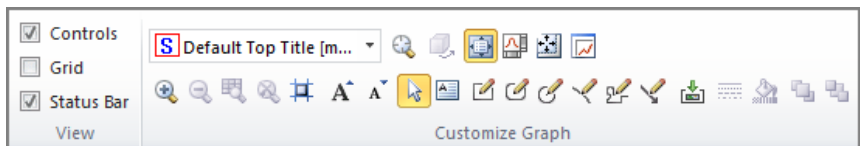





Рисунок 3.26. Панель інструментів для роботи з графіками

Розглянемо найбільш вживані інструменти керування графіками.

Після створення графік відображається в режимі  “Fit to windows”, коли розміри графіка автоматично підлаштовуються під розміри вікна. Інший режим – відображення фактичного розміру графіка визначається інструментом  “Display Graph of the actual size” (Відобразити графік фактичного розміру).

Визначити потрібні розміри вікна дозволяє інструмент налаштування розмірів . Після вибору цього інструмента з'являється вікно “Document Size and Scaling” (Розміри і масштаб документа), в якому і задаються потрібні розміри. При цьому одиниця вимірювання розмірів не відповідає системі одиниць, що задається у налаштуваннях системи параметром “Use metric measurements” (Використовувати метричні вимірювання). Тому за потреби зміни розмірів замінійте одиницю вимірювання з дюймів на сантиметри.

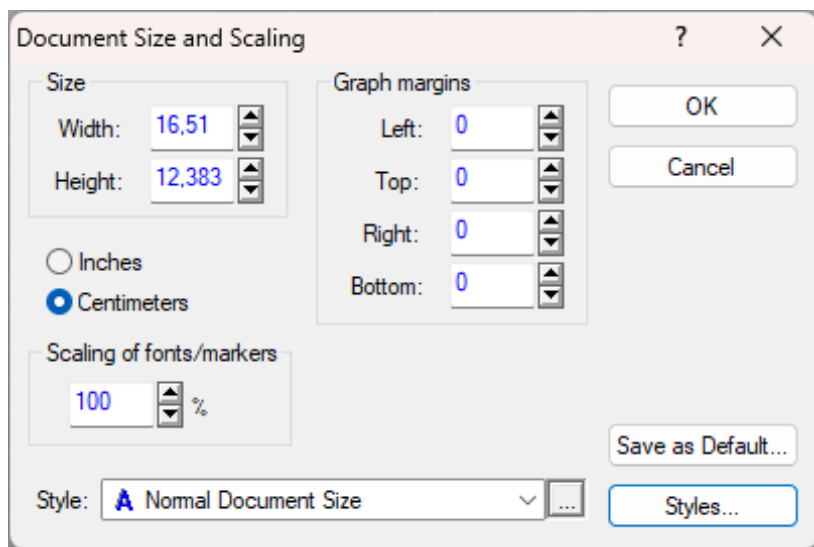



Рисунок 3.27. Визначення одиниці виміру розмірів графіка

Є можливість збільшити або зменшити будь-яку частину графіка, ці дії здійснюється за допомогою інструментів збільшення / зменшення: . Після вибору потрібного інструменту слід виділити частину графіка, яка цікавить дослідника, протягнувши по ній мишею.

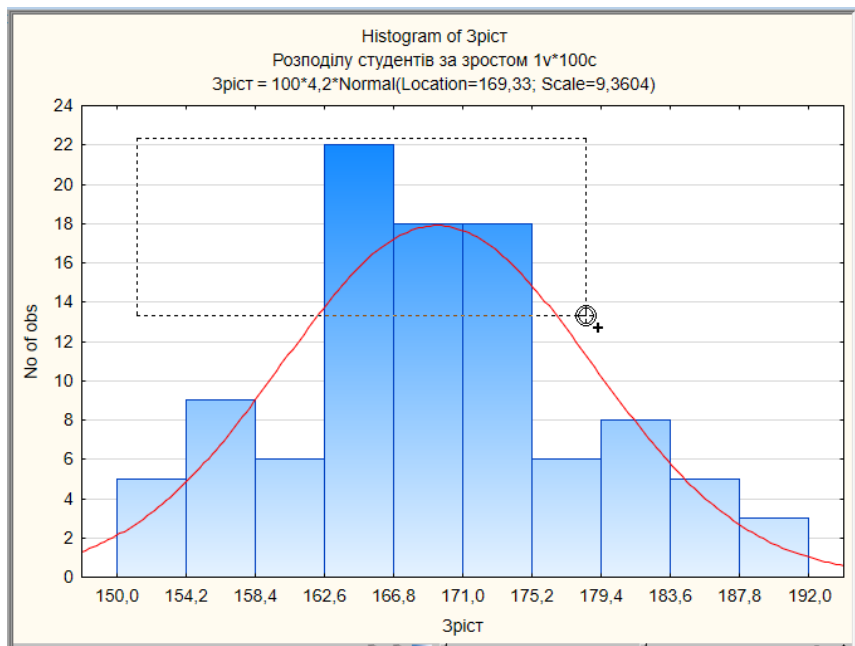





Рисунок 3.28. Застосування механізму збільшення / зменшення ділянки графіка

При цьому потрібну частину графіка можна виділити у вигляді прямокутника, еліпса тощо за допомогою відповідних інструментів: .

Зручно збільшувати графік, наприклад, з метою масштабування графіка для одержання більш точних значень з графіка.



Для відміни роботи в режимі масштабування слід застосувати інструмент  “Reset zoom” (повернути масштаб графіка до розмірів за замовчуванням).

Кілька інструментів дозволяють нанести на графік додаткову інформацію: текст, низку графічних примітивів – прямокутники, еліпси і взагалі вставити інший об’єкт. Наприклад, додавання до графіка тексту здійснюється за допомогою інструмента “Текст” . Після вибору цього інструмента слід встановити курсор в поле графіка і натиснути кнопку миші, що призведе до появи на цьому місці текстового об’єкта “Custom Text” (текст користувача), який надалі редагується звичайним чином із застосуванням стандартних інструментів форматування.

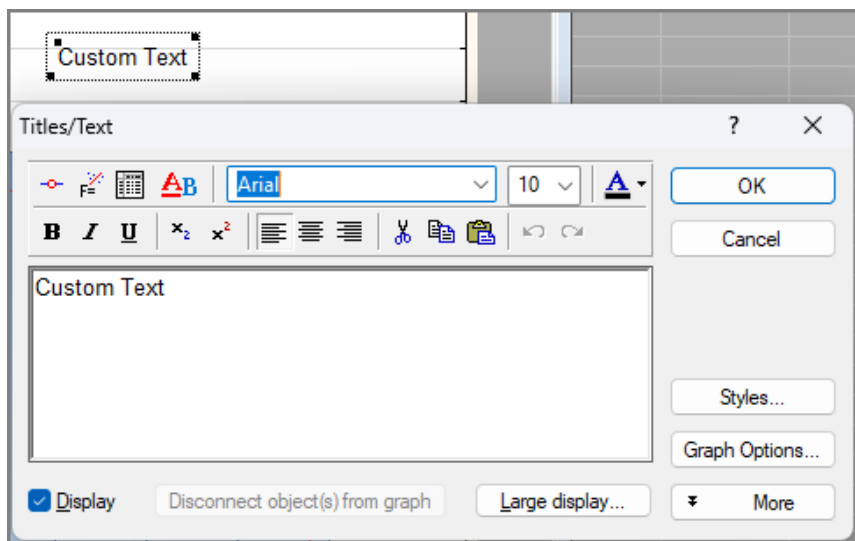



Рисунок 3.29. Застосування інструменту «Текст»

## Редагування графіка

Побудованому графіку можна надати більш зручний вигляд шляхом редагування його елементів. Редагуванню підлягають усі його елементи: загальна назва, заголовки осей і та ін. Для вибору об'єкта застосовується інструмент

“Покажчик” . Він дозволяє виділити об'єкт простим натисканням на нього. Подвійне натискання на об'єкті відкриває вікно його редагування. Перехід до редагування графіка можна здійснити також іншим шляхом. Для цього необхідно викликати контекстне меню на об'єкта. Як правило, він містить пункт “Properties” (властивості), виклик якого відкриває вікно з властивостями об'єкта.

Щоб дістатися до властивостей усіх об'єктів графіка слід двічі натиснути на ньому. Відкриється вікно “Graph Options” (параметри графіка). Визначимо деякі дії редагування.

1. Змінити загальний заголовок графіка: “Graph” – “Titles/Text”.

2. Змінити назви осей: “Axis” – “Title”.

3. Змінити початкове значення шкали осі та вибрати новий числовий інтервал шкали (крок) вкладка “Axis” – “Scaling”.

- Перейти в режим ручного редагування: зі списку “Mode” (Режим) вибрати елемент “Manual” (ручний), після чого стануть доступними поля “Minimum:” і “Maximum:”, в яких задаються початкове і кінцеве значення числової шкали.

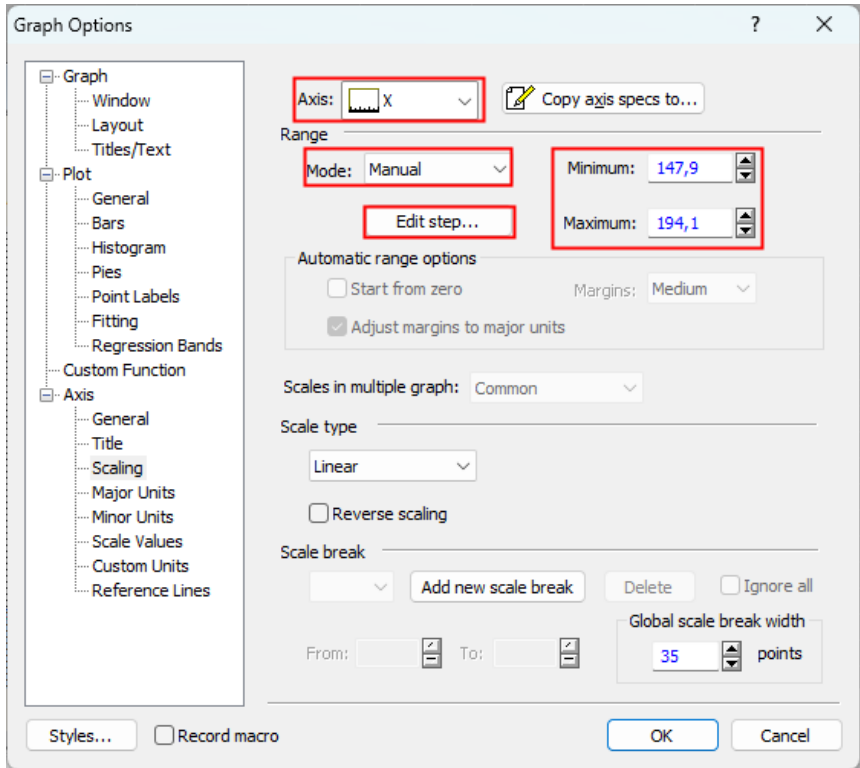


Рисунок 3.30. Ручний варіант зміни масштабу осей

- Для зміни величини кроку інтервалу слід натиснути кнопку «**Edit Step**» (змінити крок). Автоматично відбудеться перехід на вкладку “Major Units” (основні одиниці). Так само зі списку “Mode” слід вибрати елемент “Manual” і в полі “Step Size” (величина кроку) задати його нове значення. Для повернення до режиму редагування числової шкали слід натиснути кнопку «**Edit Range**» (змінити діапазон).

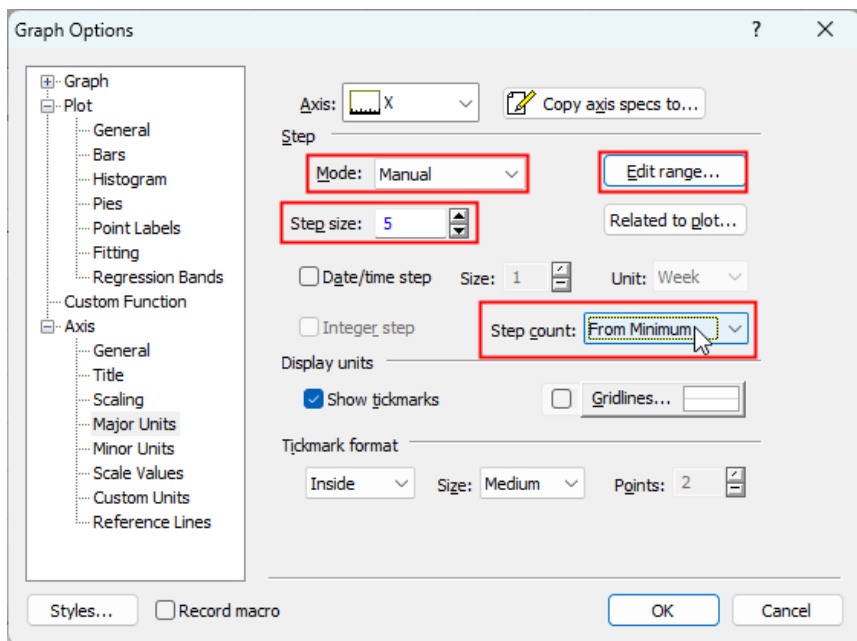


Рисунок 3.31. Зміна величини кроку інтервалу

- За закінченням редагування натиснути кнопку «ОК».

### *Редагування первинних даних графіка*

Як правило, первинні дані не бувають «у чистому вигляді»: серед них зустрічаються помилкові та аномальні значення, викиди, пропуски і т. ін., які спотворюють форму графіка. Проте програма має спеціальний засіб, за допомогою якого з графіка можна вилучати окремі дані. Для цього слід виконати команду **Format ▶ Graph Data** (Форматування даних графіка). З'явиться вікно з даними всіх спостережень, у якому достатньо просто змінити значення даних або викликати в рядку з даними контекстне меню певного спостереження і вибрати, наприклад пункт “Cut” (Вирізати).



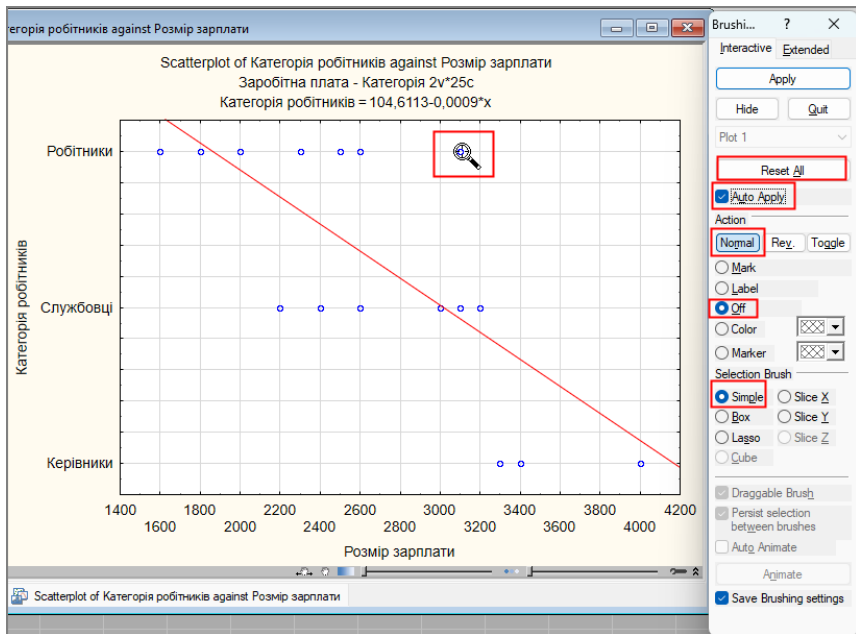


Рисунок 3.33. Вилучення точки з графіка

Для скасування змін слід натиснути кнопку «**Reset All**» (відмінити все).

І, нарешті, ще одним варіантом редагування графіка є безпосереднє редагування даних. Для вилучення конкретного значення з графіка слід викликати контекстне меню даного значення і вибрати з нього пункт "Exclude/Include" (виключити/включити). Внаслідок виконання цих дій відбудеться вилучення значення з графіка, а сам графік автоматично зміниться з урахуванням здійсненого вилучення. Повторне застосування цього пункту спричинить повернення вилученого значення.

Сама таблиця з даними буде відображати дії вилучення.

⊗	9	2600	Робітники
⊗	10	2500	Робітники
	11	2600	Службовці
⊗	12	2300	Робітники

Рисунок 3.34. Відображення вилучених точок у таблиці з даними

При цьому, як раніше було зауважено, вилучені спостереження у таблиці даних виключаються з подальших аналізів як у поточному сеансі роботи, так і у наступних, доки користувач знову не підключить їх до обробки (див. розділ «Операції з даними»).

## 4. Описові статистики

### *Теорія*

#### *Статистичні показники*

За допомогою *описових статистик* визначаються найбільш загальні властивості емпіричних даних, які дають загальне уявлення відносно значень, що набуває змінна. До них належать середня, вибіркова дисперсія, стандартне відхилення, медіана, мода, максимальне та мінімальне значення, розмах варіації та квартилі. Якщо описові статистики застосовують до генеральної сукупності, то їх ще називають *параметрами* сукупності.

#### *Середні величини*

Середні величини дають можливість охарактеризувати сукупність за допомогою лише одного узагальнюючого

показника. Їх широко вживають економісти, політики, соціологи. Усім добре відомі поняття «середня заробітна плата», «середній екзаменаційний бал» і т. ін.

Середня величина тільки тоді має сенс, якщо вона розрахована за *якісно однорідною* сукупністю.

Так, якщо розрахувати дохід для групи осіб, що складається з кількох пенсіонерів та кількох менеджерів з високою заробітною платою, то отримана величина середнього доходу явно буде мати суто абстрактний характер.

**Приклад 4.1.** необхідно визначити середню величину рівня доходу за наступними даними. Пенсіонери А, Б вказали, що їхній дохід становить 4600 грн та 5400 грн відповідно. Підприємці В, Г задекларували дохід на рівні 65000 грн і 75000 грн відповідно.

#### **Розв'язання**

Навіть і не маючи уявлення про математичний апарат для розрахунку середньої величини, її просто визначити суто логічно:

$$\begin{aligned} \text{Середній дохід} &= \frac{4600 + 5400 + 65000 + 75000}{4} \\ &= 37500 \text{ грн} \end{aligned}$$

Бачимо, що розрахована величина неспівставна з доходом пенсіонерів і, разом з тим, дуже мала для фактичного рівня доходу підприємців. Отже, визначення середньої величини має сенс саме для *якісно однорідної* сукупності. Таким чином, **середня величина** – це узагальнюючий показник, що характеризує типовий рівень варіюючої ознаки в якісно однорідній сукупності. Вона відображує те спільне, характерне, що об'єднує всю масу елементів, тобто статистичну сукупність.

Слід зауважити, що середня величина (як, до речі, і будь-який інший статистичний показник) поєднує кількісну і якісну



визначеність ознаки. *Кількісна визначеність* – це конкретне числове значення показника, а *якісна визначеність* зумовлюється суттю явища і відображається назвою показника: врожайність пшениці, кількість студентів тощо. Тому *обов'язково* поруч з числовим значенням статистичного показника вказується його назва.

## Середня арифметична

Під час статистичного аналізу використовують різні види середніх величин. Найбільш поширена – це *середня арифметична*. Залежно від того, згруповані чи незгруповані первинні статистичні дані, для їх розрахунку застосовують дві формули.

Для незгрупованих даних, тобто за первинними даними:

$$\bar{x} = \frac{\sum x}{n} \quad (4.1)$$

де  $\bar{x}$  – середня величина,

$x$  – значення окремих варіант,

$n$  – кількість елементів (обсяг) сукупності.

Така середня величина має назву *простий середньої арифметичної величини*.

Розрахуємо таку величину для даних з прикладу 2.1:

$$\bar{x} = \frac{4 + 4 + 5 + \dots + 3 + 4}{25} = \frac{100}{25} = 4,0 \text{ бали}$$

Для згрупованих даних формула має вигляд:

$$\bar{x} = \frac{\sum xf}{\sum f} \quad (4.2)$$

де  $f$  – частоти.

Частоти в цій формулі виступають в ролі так званої *статистичної ваги*, коли кожна з варіант множиться на частоту, що їй відповідає. Такий добуток у статистиці має назву *зважування*. У цьому випадку середня називається *середньою арифметичною зваженою*.

Розрахуємо таку величину для даних з таблиці 2.1:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{2 \cdot 1 + 3 \cdot 5 + 4 \cdot 12 + 5 \cdot 7}{25} = \frac{100}{25} = 4,0 \text{ бали}$$

Зверніть увагу, що величина простої середньої, розрахованої вище для прикладу 4.1, збігається з величиною середньої арифметичної зваженої для таблиці 4.1. Це так і повинно бути, тому що, незважаючи на те, що розрахунок проводився за різними формулами, дані були *ті самі* і розрізнялися тільки за їх характером: у першому випадку вони були незгруповані, а в другому – згруповані.

Для розрахунку середньої величини у випадку незгрупованих даних використовується формула середньої арифметичної простої, а у випадку згрупованих даних – формула середньої арифметичної зваженої, значення яких за тими самими даними збігаються.

Дещо умовного характеру набуває розрахунок середньої величини з інтервального ряду розподілу. У цьому разі для розрахунку використовують формулу середньої арифметичної зваженої, але замість значень окремих варіант використовують величину середини інтервалу:

$$\bar{x} = \frac{\sum x' f}{\sum f} \quad (4.3)$$

де  $x'$  – середина інтервалу.

Середина інтервалу розраховується як напівсума двох його меж. Наприклад, середина першого інтервалу для таблиці 2.2 розподілу студентів навчальної групи університету за зростом розраховується таким чином:

$$x' = \frac{150 + 158}{2} = \frac{308}{2} = 154 \text{ см}$$

Модифікуємо дані таблиці 2.2 і виконаємо розрахунок середньої арифметичної:

Таблиця 4.1

Техніка розрахунку середньої арифметичної зваженої за даними з інтервального ряду

<i>Зріст, см</i>	<i>Кількість, осіб</i>	<i>Середина інтервалу, см</i>	<i>Розрахункові дані, см</i>
<i>x</i>	<i>f</i>	<i>x'</i>	<i>fx'</i>
До 158	6	154	924
158-166	7	162	1134
166-182	10	174	1740
182 і вище	2	190	380
Разом	25	x	4178

Під час розрахунку середини другого та третього інтервалів жодних проблем не виникає, оскільки вони мають обидві межі. Так, для другого інтервалу середина визначається як  $(158+166):2 = 162$  см.

Проблема розрахунку середини інтервалу виникає у випадку відкритих інтервалів, коли відсутня одна з меж. Вона вирішується так: крок відкритого інтервалу умовно визначається таким самим, як у закритого інтервалу, що з ним межує.

Оскільки крок другого інтервалу, що межує з першим, дорівнює «8» (166 – 158), тому і для першого інтервалу встановлюється такий саме крок («8») і, відповідно, нижня межа його буде  $158 - 8 = 150$ . Для останнього інтервалу крок дорівнює «16» (182-166), тому його верхня межа становить  $182+16= 198$ . Після визначення середин інтервалів розрахунок

середньої арифметичної зваженої здійснюється так само, як і в дискретному ряді розподілу:

$$\bar{x} = \frac{\sum x'f}{\sum f} = \frac{4178}{25} \approx 167 \text{ см}$$

Середня величина – показник іменований. Вона має таку саму одиницю виміру, що і варіююча ознака.

Середня арифметична має певні математичні властивості, які позначаються на її використанні. Найважливіші з них такі:

1. Алгебраїчна сума відхилень усіх варіант від середньої дорівнює «0»:

$$\sum (x - \bar{x})f = 0 \quad (4.5)$$

2. Якщо кожному варіанту зменшити або збільшити на якусь постійну величину  $A$ , то середня зміниться відповідно на ту саму величину:

$$\sum \frac{(x \mp A)f}{\sum f} = \bar{x} \mp A \quad (4.6)$$

3. Якщо кожному варіанту розділити чи помножити на якусь постійну величину  $A$ , то середня зменшиться або збільшиться у стільки ж разів.

$$\frac{\sum Ax f}{\sum f} = A\bar{x} \quad \frac{\sum \frac{x}{A} f}{\sum f} = \frac{\bar{x}}{A} \quad (4.7)$$

4. Якщо частоту кожної групи зменшити в те саме число разів, то середня при цьому не зміниться:

$$\frac{\sum x \frac{f}{A}}{\sum \frac{f}{A}} = \bar{x} \quad (4.8)$$

Ця властивість дозволяє зробити висновок, що середня залежить не від абсолютної суми частот, а від їх співвідношення у сукупності, тобто від питомої ваги кожної варіанти у сукупності. Тому, якщо абсолютні частоти замінити їх частками, то розрахунок середньої можна записати так:

$$\bar{x} = \sum x d_i \quad (4.9)$$

де  $\sum d = 1$ .

5. Сума квадратів відхилень варіант від середньої арифметичної менша, ніж сума квадратів відхилень варіант від будь-якої іншої величини:

$$\sum (x - \bar{x})^2 f = \min \quad (4.10)$$

### Середнє гармонійне (гармонічне)

Середня арифметична величина – найбільш вживана середня величина, але існують і інші різновиди середньої величини. Так, якщо підсумуванню підлягають не самі варіанти, а обернені їм числа:  $\frac{1}{x_1}, \frac{1}{x_2}, \dots$ , то середнє значення ознаки обчислюють за допомогою *середньої гармонійної*. На практиці цей вид середніх застосовується, наприклад, при обчисленні середніх державних цін.

Формула для її розрахунку також залежить від того, згруповані чи не згруповані дані.

Формула середньої гармонійної простої для незгрупованих даних:

$$\bar{x} = \frac{n}{\sum \frac{1}{x}} \quad (4.11)$$

Формула середньої гармонійної зваженої для згрупованих даних:

$$\bar{x} = \frac{\sum xf}{\sum \frac{xf}{n}} \quad (4.12)$$

**Приклад.** Типовим прикладом використання середньої гармонійної є розрахунок середньої врожайності сільськогосподарських культур, за наявності даних про врожайність ( $x$ ) та валовий (загальний) збір ( $xf$ ) сільськогосподарських культур, коли для визначення середньої врожайності не вистачає значення площі, з якої збирається врожай.

Таблиця 4.2  
Валовий збір і врожайність за господарствами

<i>Господарство</i>	<i>Врожайність, ц/га</i>	<i>Валовий збір, ц</i>
	$x$	$xf$
1	100	100000
2	120	180000
3	140	140000

Але ця проблема просто вирішується шляхом ділення валового збору культури на врожайність, а сам розрахунок має такий вигляд:

$$\bar{x} = \frac{100000 + 180000 + 140000}{\frac{100000}{100} + \frac{180000}{120} + \frac{140000}{140}} = \frac{420000}{3500} = 120 \text{ ц/га}$$

З даними з таблиці не важко розрахувати посівну площу для кожного господарства (прямі дані), наприклад для першого вона становить  $\frac{100000}{100} = 1000$  га, для другого – 1500 га,

третього – 1000 га. Підставимо ці дані до формули середньої арифметичної зваженої:

$$\bar{x} = \frac{100 * 1000 + 120 * 1500 + 140 * 1000}{1000 + 1500 + 1000} = 120 \text{ ц/га}$$

Як бачимо, розрахунок за прямими даними призвів до аналогічного результату.

Отже, середнє гармонійне – це обернена величина до середньої арифметичної. Вона застосовується, коли дані, що виступають вагою  $f$ , відсутні, але відомі варіанти  $x$  й добутки  $xf$ .

### Структурні (позиційні) середні

На практиці застосовують ще дві особливі середні величини, що називають *структурними (позиційними) середніми*.

*Мода* (позначається як  $Mo$ ). Це *варіанта*, який в ряду розподілу відповідає найбільша частота. За наявності тільки одного максимального значення ряд розподілу має назву одномодального. Але ряд розподілу може мати і кілька модальних значень, у цьому разі він називається полімодальним.

У дискретному ряді розподілу мода легко відшукується візуально. Відшукаємо моду для таблиці 2.1 розподілу студентів начальної групи за ступенем орієнтованості на педагогічну роботу. Бачимо, що найбільша частота становить «12», їй відповідає варіанта, тобто оцінка «4». Таким чином,  $Mo = 4$  бали.

В інтервальному ряді розподілу мода наближено розраховується за формулою:

$$Mo = x_0 + h \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \quad (4.13)$$

де  $x_0$  – нижня межа модального інтервалу,  
 $h, f_m$  – ширина і частота модального інтервалу,  
 $f_{m-1}, f_{m+1}$  – частоти попереднього та наступного інтервалів відносно модального.

Якщо модальний інтервал збігається з першою групою, то вважається, що мода знаходиться в межах цього інтервалу.

*Модальний* інтервал – це інтервал, що містить найбільшу частоту. Він також просто визначається візуально.

Розрахуємо моду для таблиці 2.2 розподілу студентів навчальної групи університету за зростом. Бачимо, що найбільшу частоту «7» містить інтервал «158-166», який відповідно і є модальним. Ширина цього інтервалу становить  $166-158=8$ . Частоти інтервалів, що межують з ним, становлять «5» і «6». Звідси мода:

$$Mo = x_0 + h \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} = 158 + 8 \cdot \frac{7 - 5}{(7 - 5) + (7 - 6)} \approx 163,3 \text{ см}$$

Отже, середній зріст студентів навчальної групи університету становить приблизно 163,3 см.

*Медіана* (позначається  $Me$ ) – це варіанта, що ділить впорядкований ряд розподілу на дві рівних частини за чисельністю сукупності.

Для розрахунку медіани використовуються кумулятивні частоти.

Якщо медіанний інтервал збігається з першою групою, то медіана не визначається.

Відшукаємо медіану у дискретному ряду розподілу для даних з таблиці 2.1 розподілу студентів початкової групи за



ступенем орієнтованості на педагогічну роботу. Перш за все розрахуємо за даними таблиці кумулятивні частоти та запишемо їх до наступної таблиці.

Таблиця 4.3

**Техніка визначення медіани в дискретному ряду розподілу**

<i>Ступінь орієнтованості, бали</i>	<i>Кількість студентів</i>	<i>Кумулятивні частоти</i>
<i>x</i>	<i>f</i>	
2	1	1
3	5	6
4	12	18
5	7	25
Разом	25	x

Визначимо центр розподілу, для чого суму частот поділимо на «2»:  $\frac{25}{2} = 12,5$ . Зі стовпчика «Кумулятивні частоти» бачимо, що центр припадає на варіанту (оцінку) «4». Їй відповідає кумулятивна частота «18», яка перевищує отримане вище значення центру розподілу «12,5». Таким чином  $Me = 4$  бали.

В інтервальному ряду розподілу медіана визначається за формулою:

$$Me = x_0 + h \cdot \frac{\frac{\sum f}{2} - S_{m-1}}{f_m} \quad (4.14)$$

де  $x_0$  – нижня межа медіанного інтервалу;

$h$  і  $f_m$  – ширина і частота медіанного інтервалу;

$S_{m-1}$  – кумулятивна частота в інтервалі, що передує медіанному.

Розрахуємо медіану в інтервальному ряду розподілу з таблиці 2.2., що містить дані про розподіл студентів навчальної групи університету за зростом.

Знову визначимо центр розподілу, для чого суму частот поділимо на «2»:  $\frac{25}{2} = 12,5$ . Зі стовпчика «Кумулятивні частоти» бачимо, що центр припадає на інтервал «166-174», який і є медіанним. Йому відповідає кумулятивна частота «18», яка перевищує отримане вище значення центру розподілу «12,5».

Таблиця 4.4

**Розподіл студентів навчальної групи університету за зростом**

<i>Зріст, см.</i>	<i>Кількість студентів, осіб</i>	<i>Кумулятивна частота</i>
<i>x</i>	<i>f</i>	<i>S</i>
150-158	5	5
158-166	7	12
166-174	6	18
174-182	4	22
182-190	3	25
Разом	25	x

Таким чином

$$Me = x_0 + h \cdot \frac{\frac{\sum f}{2} - S_{m-1}}{f_m} = 166 + 8 \cdot \frac{\frac{25}{2} - 12}{6} = 166,7 \text{ см}$$

Отже, половина студентів навчальної групи університету має зріст до 166,7 см., а друга половина – вище 166,7 см.

Слід зазначити, що мода і медіана є не умовними величинами, як середня величина, а реальними, тому вони використовуються на практиці.

На величину цих важливих величин не впливають значення варіант, не характерних даній сукупності, наприклад, надмірно малі чи надмірно великі. При обчисленні ж середньої арифметичної до уваги беруться значення всіх без винятку варіант. Саме через це мода та медіана в окремих випадках мають свої переваги перед середньою арифметичною й використовуються при розв'язанні практичних завдань. Наприклад, саме на модальний розмір орієнтуються виробники взуття під час планування випуску асортименту взуття за розмірами.

Для характеристики структури ряду розподілу обчислюють також інші структурні характеристики, що називають *квантилі розподілу*.

**Квартилі (Q)** – це значення варіант, які ділять упорядкований ряд за обсягом на чотири рівних частини.

Перший квартиль:

$$Q_1 = x_0 + h \cdot \frac{0,25 \sum f - S_{Q_1-1}}{f_{Q_1}} \quad (4.15)$$

де  $x_0$  – нижня межа інтервалу, що містить перший квартиль (інтервал визначається за накопиченою частотою; це перший інтервал, значення в якому перевищує 25 %);

$h$  і  $f$  – ширина і частота інтервалу, що містить перший квартиль;

$S_{Q_1-1}$  – кумулятивна частота в інтервалі, що передує інтервалу, який містить перший квартиль.

Третій квартиль:

$$Q_3 = x_0 + h \cdot \frac{0,75 \sum f - S_{Q_3-1}}{f_{Q_3}} \quad (4.16)$$

де  $x_0$  – нижня межа інтервалу, що містить третій квартиль (інтервал визначається за накопиченою частотою; це перший інтервал, значення в якому перевищує 75 %);

$h$  і  $f$  – ширина і частота інтервалу, що містить третій кватиль;  
 $S_{Q_3-1}$  – кумулятивна частота в інтервалі, що передує інтервалу, який містить третій кватиль.

**Децилі (D)** – це значення варіант, які ділять упорядкований ряд за обсягом на десять рівних частин.

$$D = x_0 + h \cdot \frac{0,1 \cdot i \cdot \sum f - S_{D-1}}{f_D} \quad (4.17)$$

де  $i$  – порядковий номер дециля,

$S_{D-1}$  – кумулятивна частота в інтервалі, що передує децильному.

### Показники варіації

Значення кожної ознаки змінюється в певних межах від елемента до елемента або для того самого елемента протягом певного часу. Такі зміни називаються коливанням або *варіацію* ознаки, а сама **варіація**, тобто коливання, мінливість будь-якої ознаки, є властивістю статичної сукупності.

Як вище було зазначено, середня величина тільки тоді відображає типовий рівень ознаки, коли вона розрахована за якісно однорідною сукупністю. Тому резонно уявити, що разом з розрахунком середньої величини доцільно визначити міру однорідності сукупності, щоб знати, наскільки можна довіряти розрахованій середній величині.

Чим менша варіація, тим однорідніша сукупність, отже, тим більш надійна й типова розрахована за її значенням середня величини.

На практиці розраховують показники варіації.

*Розмах варіації.*

$$R = x_{max} - x_{min} \quad (4.18)$$

де  $x_{max}$  і  $x_{min}$  – найбільше і найменше значення ознаки.

Показник характеризує межі, в яких змінюється значення ознаки. Надійність цього показника невелика, оскільки він базується на двох крайніх значеннях ознаки, які можуть бути випадковими. Наприклад, у групі студенти зазвичай мають зріст в межах від 165 см у дівчат до 185 у чоловіків і значення розмаху буде 20 см, тобто хоч і достатньо великим, але достатньо нормальним. Але уявімо, що в групі є «дюймовочка», зріст якої 150 см, і «баскетболіст» на зріст 200 см, то відразу значення розмаху буде дорівнювати 50 см, що дозволяє зробити висновок, що студенти групи дуже значно відрізняються за зростом.

Тому на практиці цей показник використовується зрідка, для грубої оцінки варіації ознаки. Широко ж використовуються з метою оцінки варіації такі показники як *середнє лінійне відхилення* і *середнє квадратичне відхилення*. Вони базуються на сумі відхилень від середньої величини за модулем або у квадраті відповідно, враховуючи властивість середньої величини, згідно якої алгебраїчна сума відхилень усіх варіант від середньої дорівнює нулю.

Формули для розрахунку цих показників, як і у випадку із середніми величинами, залежать від того, згруповані чи не згруповані первинні дані.

Формули для розрахунку *середнього лінійного відхилення*:

а) За незгрупованими даними:

$$\bar{l} = \frac{\sum |x - \bar{x}|}{n} \quad (4.19)$$

де  $n$  – кількість елементів сукупності,

$\bar{x}$  – середня величина.

б) За згрупованими даними:

$$\bar{l} = \frac{\sum |x - \bar{x}| \cdot f}{\sum f} \quad (4.20)$$

Формули для розрахунку *середнього квадратичного відхилення* (інша назва – *стандартне відхилення*):

а) За незгрупованими даними:

$$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{n}} \quad (4.21)$$

б) За згрупованими даними:

$$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2 \cdot f}{\sum f}} \quad (4.22)$$

Стандартне відхилення у статистиці є широко вживаним показником.

Ці показники – *іменовані* величини, що мають таку саму одиницю виміру, що і варіююча ознака.

Вони ідентичні за своїм змістом і лише дещо відрізняються за числовим значенням.

У симетричному розподілі математичний зв'язок між ними має вигляд:

$$\sigma = 1,25\bar{l} \quad (4.23)$$

Сутність цих показників у тому, що вони показують, на скільки одиниць у середньому відхиляються індивідуальні значення ознаки від середньої величини.

Розрахуємо показники варіації за даними прикладу 2.1 (опитування студентів навчальної групи з метою з'ясування їх орієнтованості на педагогічну роботу) для перших 5 студентів.

*Розмах варіації*:  $R = x_{max} - x_{min} = 5 - 4 = 1$ .

Показник свідчить про досить однорідний склад сукупності, тобто рівень орієнтованості на педагогічну роботу студентів приблизно однаковий. Але якщо розрахувати цей показник для усіх студентів з прикладу 2.1, то розмах варіації  $R = 3$ , оскільки серед студентів є і такий, що має найнижчу оцінку – «2», тобто значення показника суттєво зміниться.

Техніку обчислення середнього лінійного і квадратичного відхилення за незгрупованими даними демонструється у таблиці 4.5. Таблиця містить оцінки, що одержали студенти під час проведення опитування.

Для виконання розрахунків у таблиці необхідно розрахувати середню величину:

$$\bar{x} = \frac{4 + 4 + 5 + 5 + 4}{5} = 4,4 \text{ бала}$$

Таблиця 4.5

**Техніка обчислення середнього лінійного і квадратичного відхилення за незгрупованими даними**

<i>Прізвище студента</i>	<i>Оцінка</i>	<i>Розрахункові дані</i>		
		$ x - \bar{x} $	$(x - \bar{x})$	$(x - \bar{x})^2$
Власова В.	4	0,4	-0,4	0,16
Дубонос Т.	4	0,4	-0,4	0,16
Макаренко Ю.	5	0,6	0,6	0,36
Федорець А.	5	0,6	0,6	0,36
Романець Ю.	4	0,4	-0,4	0,16
<b>Разом</b>	<b>x</b>	<b>2,4</b>	<b>x</b>	<b>1,20</b>

$$\bar{l} = \frac{\Sigma|x-\bar{x}|}{n} = \frac{2,4}{5} = 0,48 \text{ балів}$$

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{1,20}{5}} = 0,49 \text{ балів}$$

Таким чином, середній бал за опитаними студентами становить 4,4 бали, а середнє відхилення, що показує, як у середньому відхиляються індивідуальні оцінки від середньої арифметичної, становить 0,48 балів за середнім лінійним

відхиленням і 0,49 балів за середнім квадратичним відхиленням.

Техніку обчислення середнього лінійного і квадратичного відхилення за згрупованими даними з таблиці 2.2 розподілу студентів навчальної групи навчального закладу за зростом демонструє таблиця 4.6. Для виконання подальших розрахунків необхідно розрахувати середню величину:

$$\bar{x} = \frac{\Sigma x' f}{\Sigma f} = \frac{4194}{25} \approx 168 \text{ см}$$

Таблиця 4.6

**Техніка обчислення середнього лінійного і  
квадратичного  
відхилення за згрупованими даними**

<i>Зріст, см</i>	<i>Кіль- кість, осіб</i>	<i>Середина інтервалу</i>	<i>Розрахункові дані</i>				
			$x'f$	$(x' - \bar{x})$	$(x' - \bar{x})^2$	$(x' - \bar{x})^2 f$	$ x' - \bar{x} f$
До 158	5	154	770	-	196	980	70
158-166	7	162	1134	-6	36	252	42
166-174	6	170	1020	2	4	24	12
174-182	4	178	712	10	100	400	40
182 і вище	3	186	558	18	324	972	54
<b>Разом</b>	<b>25</b>	<b>x</b>	<b>4194</b>	<b>x</b>	<b>x</b>	<b>2628</b>	<b>218</b>

Підставимо дані з таблиці до відповідних формул:

$$\bar{l} = \frac{\Sigma |x - \bar{x}| \cdot f}{\Sigma f} = \frac{218}{25} \approx 8,7 \text{ см}$$



$$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2 \cdot f}{\sum f}} = \sqrt{\frac{2628}{25}} \approx 10,2 \text{ см}$$

Таким чином, середній зріст студентів навчальної групи університету становить 168 см, а середнє відхилення, що показує, як у середньому відхиляються індивідуальній зріст студентів від середньої арифметичної, становить 8,7 см за середнім лінійним відхиленням і 10,2 см за середнім квадратичним відхиленням.

Показниками середнього лінійного та квадратичного відхилення не можна скористатися для порівняння розміру варіації різнойменних ознак однієї і тієї ж сукупності чи міри варіації однієї і тієї ж ознаки для сукупностей, що мають різні середні.

Для цього використовується показник, що характеризує *міру варіації*. Загальний його вигляд можна подати так:

$$\begin{aligned} \text{Коефіцієнт варіації} & & (4.24) \\ &= \frac{\text{Іменований показник варіації}}{\text{Середня величина}} \end{aligned}$$

Підставимо розглянуті вище показники варіації до цієї формули і отримаємо наступні показники (їх іноді називають *відносні показники варіації*):

$$\text{Лінійний коефіцієнт варіації } V_{\bar{l}} = \frac{\bar{l}}{\bar{X}} \quad (4.25)$$

$$\text{Квадратичний коефіцієнт варіації } V_{\sigma} = \frac{\sigma}{\bar{X}} \quad (4.26)$$

$$\text{Коефіцієнт осціляції } V_R = \frac{R}{\bar{X}} \quad (4.27)$$

Саме за допомогою коефіцієнтів варіації здійснюється порівняння ступеню варіації однієї тієї ж ознаки в різних сукупностях.

При цьому найчастіше з цією метою використовується квадратичний коефіцієнт варіації.

Вважається, що сукупність є однорідною, а середня – типовою, якщо квадратичний коефіцієнт варіації не перевищує 33 %.

У прикладі 4.1 була розрахована величина середнього рівня доходу для сукупності, що складалася з двох пенсіонерів і двох підприємців. Розрахуємо за цими даними середнє квадратичне відхилення.

$$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{\frac{(4600-37500)^2+(5400-37500)^2+\dots}{4}} \approx 32693 \text{ грн,}$$

а потім – квадратичний коефіцієнт варіації:

$$V_{\sigma} = \frac{\sigma}{\bar{X}} = \frac{32692}{37500} \approx 0,872 \text{ або } 87,2 \%$$

Суто логічно зрозуміло, що до сукупності були включені явно різнорідні елементи, а розрахунок квадратичного коефіцієнта варіації тільки дав цьому математичне підтвердження, оскільки отримане значення відчутно перевищує 33 %.

Розрахуємо квадратичний коефіцієнт варіації для таблиці 4.6:

$$V_{\sigma} = \frac{\sigma}{\bar{X}} = \frac{10,2}{168} \approx 0,061 \text{ або } 6,1 \%$$

У цьому випадку значення квадратичного коефіцієнта варіації значно менше за 33 %. Це дає нам право вважати сукупність, за якою проводився розрахунок, однорідною, а отримане значення середнього зросту студентів – типовим.

Одним з найбільш важливих показників не тільки серед

показників варіації, але й взагалі серед усіх статистичних показників є **дисперсія**. Дисперсія використовується для вимірювання варіації, під час аналізу зв'язків, перевірки статистичних гіпотез тощо.

Базою для її обчислення служить середнє квадратичне відхилення. Звідси маємо знов-таки дві формули для розрахунку дисперсії:

За незгрупованими даними:

$$\sigma^2 = \frac{\Sigma(x - \bar{x})^2}{n} \quad (4.28)$$

За згрупованими даними:

$$\sigma^2 = \frac{\Sigma(x - \bar{x})^2 \cdot f}{\Sigma f} \quad (4.29)$$

Дисперсію можна також розрахувати і так:

$$\sigma^2 = \overline{x^2} + \bar{x}^2 \quad (4.30)$$

На відміну від розглянутих вище показників варіації, дисперсія – величина не іменована.

Техніка розрахунку дисперсії практично ідентична розглянутій вище в таблицях 4.5 і 4.6.

Варіація притаманна і номінальним ознакам. При цьому, якщо можливі тільки два значення ознаки, що взаємно виключають одне одного, таку варіацію називають *альтернативною*. Так само називається і сама ознака. Отже, **альтернативна ознака** – це така ознака, значення якої притаманні одній частині сукупності і не притаманні іншій.

Дисперсія альтернативної ознаки розраховується за формулою:

$$\sigma^2 = d_1 d_2 = d_1(1 - d_1) \quad (4.31)$$

де  $d_1$  і  $d_2$  – частки елементів сукупності, яким притаманна або не притаманна ознака.

У тому випадку, коли номінальна ознака набуває не два, а більше значень, тобто елементи сукупності розподіляються на три або більше груп, то оцінка варіації є узагальнюючою дисперсією:

$$\sigma^2 = d_1 d_2 \cdots d_n = \prod_{i=1}^n d_i \quad (4.32)$$

де  $d_i$  – частки групи.

### *Характеристики форми розподілу*

Крива розподілу є графічним зображенням ряду розподілу і характеризує його тип і форму. Розрізняють емпіричні і теоретичні криві розподілу. *Емпірична* – будується за фактичними даними. *Теоретична* крива розподілу відображає закономірність певного типу розподілу в чистому вигляді, тобто в тому випадку, коли на розподіл не впливають випадкові чинники.

Зауважимо, що форма розподілу дає уяву щодо однорідності сукупності. Якщо сукупність більше однієї вершини, то це свідчить про неоднорідний склад сукупності, тобто вона містить різнотипні дані. І для одержання коректних результатів дослідник має подбати про усунення цієї проблеми шляхом відокремлення однорідних груп.

Одним з таких типів розподілу є *нормальний розподіл (закон Гауса)*. Він найчастіше застосовується при вивченні природних явищ завдяки його математичним властивостям. Наприклад відомо, що випадкові похибки в економічних рядах, радіотехніці і та ін. мають приблизно нормальний розподіл; за його допомогою можна наближено описати зріст дорослих людей тощо. До того ж вимога нормальності розподілу передуює застосуванню багатьох статистичних методів.

Для нього виконується **правило трьох сигм**, згідно якого варіація окремих значень ознаки знаходиться в межах  $\pm 3\sigma$  від середньої. При цьому в межах  $\bar{x} \pm \sigma$  знаходиться приблизно 68,3 % усіх значень, в межах  $\bar{x} \pm 2\sigma$  – 95,4 %, а в межах  $\bar{x} \pm 3\sigma$  – 99,7 %.

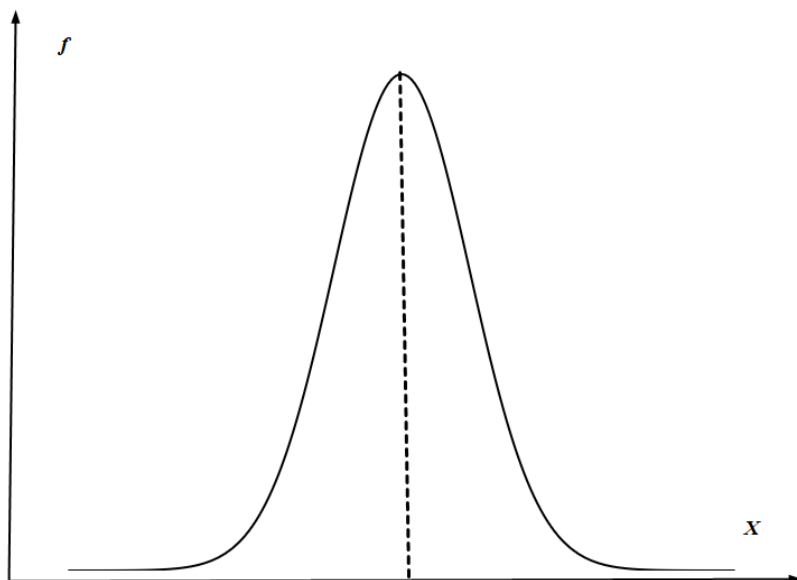


Рисунок 4.1. Графік нормального розподілу

Для нормального розподілу всі три показники центру розподілу збігаються:  $\bar{X} = Me = Mo$ .

Нормальна крива розподілу зображує статистичну сукупність, в якій відхилення індивідуальних значень ознаки від середньої в один бік зустрічаються так само часто, як відхилення в інший бік. Вона описується відповідною функцією і є симетричною: її праве крило абсолютно ідентичне лівому. При цьому характеристики центру розподілу (середня, мода, медіана) мають однакові значення.

Разом із тим, слід зауважити, що для суспільних явищ нормальна крива в чистому вигляді зустрічається зрідка. Частіше всього зустрічаються асиметричні криві розподілу, які мають право – або лівосторонній напрямок.

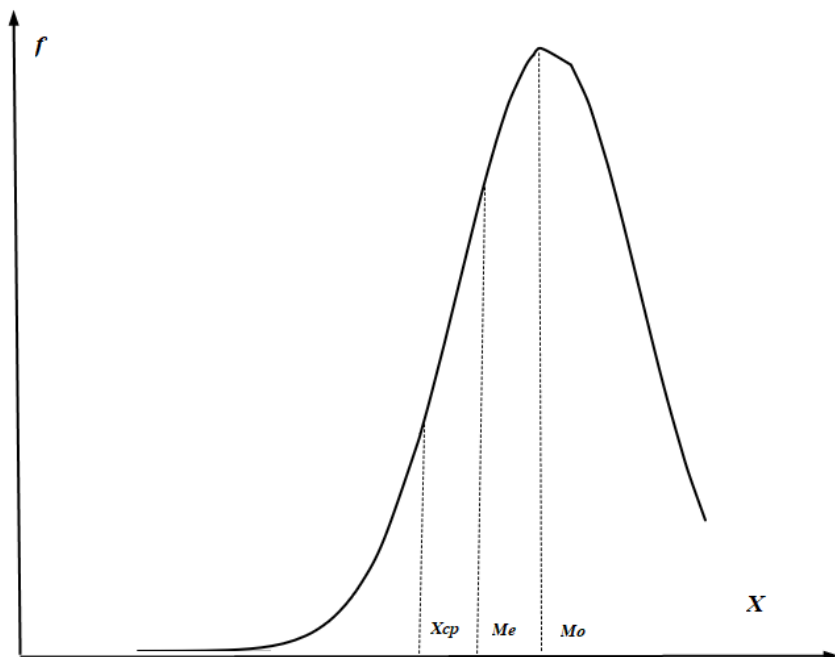


Рисунок 4.2. Лівостороння асиметрія

Для лівосторонньої асиметрії справедливою є нерівність  $\bar{X} < Me < Mo$ .

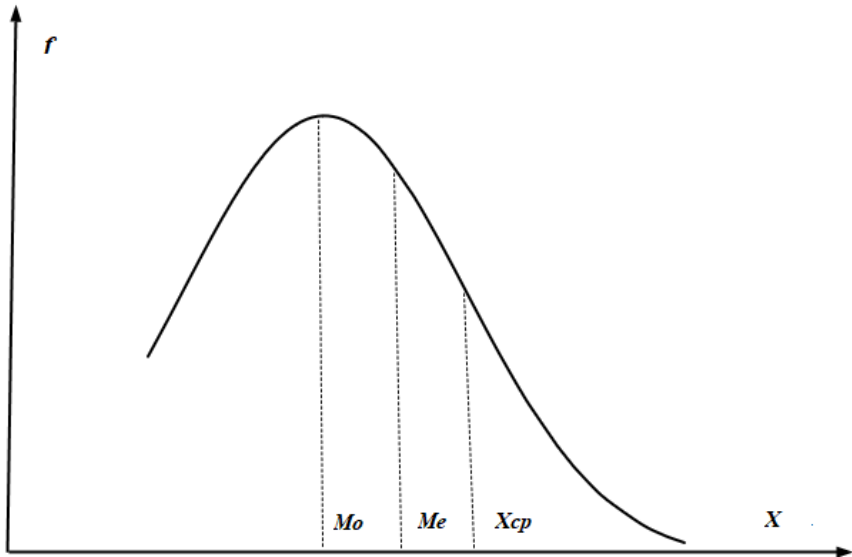


Рисунок 4.3. Правостороння асиметрія

Для правосторонньої асиметрії справедливою є нерівність  $\bar{X} > Me > Mo$ .

Нормальний розподіл характеризується двома параметрами: середньою і стандартним відхиленням. Середнє значення задає розташування кривої на числовій осі, а стандартне відхилення – ширину кривої, виступаючи як масштаб виміру.

Для еталонного розподілу значення середньої становить «0», а середнє відхилення «1».

Форма кожного ряду розподілу характеризується двома властивостями: мірою і напрямом скошеності і ступенем зосередженості елементів сукупності навколо центра розподілу.

Найпростішим показником, що дає уявлення про напрям асиметрії і міру скошеності в розподілі є **коефіцієнт асиметрії**

$$A = \frac{\bar{X} - Mo}{\sigma} \quad (4.33)$$

де  $\bar{X}$  – середня арифметична,  $Mo$  – мода,  $\sigma$  – середнє квадратичне відхилення.

Як узагальнюючи характеристики розподілу використовуються **центральні моменти розподілу**. За допомогою невеликої їх кількості можна описати будь-який розподіл. **Момент розподілу** – це середня арифметична  $k$ -го ступеня відхилення індивідуальних значень ознаки від середньої:

$$\mu_k = \frac{\sum (X - \bar{X})^k f}{\sum f} \quad (4.34)$$

Ступінь  $k$  визначає порядок моменту.

**Ексцес** – це показник, який характеризує крутизну, гостровершинність кривих розподілу. Для його вимірювання використовується момент 4-го порядку, який називається **коефіцієнтом ексцесу**.

$$E = \frac{\mu_4}{\sigma^4} - 3 \quad (4.35)$$

де  $\mu_4$  – центральний момент 4-го порядку,  $\sigma^4$  – середнє квадратичне відхилення, зведене до четвертого ступеня.

На практиці ексцес і коефіцієнт асиметрії використовуються для перевірки закону розподілу на нормальність (про що йде мова в розділі «Перевірка відповідності емпіричних даних нормальному розподілу»).



1. В симетричному (нормальному) розподілі  $A = 0$ , при правосторонній асиметрії  $A > 0$ , при лівосторонній  $A < 0$ . При цьому напрям асиметрії протилежний напрямку зміщення вершини розподілу. Якщо ця вершина зміщена ліворуч, то спостерігається правостороння асиметрія і навпаки.

2. При  $E > 0$  – крутизна вище нормального (гостровершинне), при  $E = 0$  – нормальна,  $E < 0$  – нижче нормального (плосковершинне). До речі, поправка «-3» наприкінці формули коефіцієнта ексцесу використовується як раз для того, щоб для нормального розподілу він становив «0».

## **Розрахунок описових статистик у STATISTICA**

Найбільш вживані описові статистики, а саме: середню арифметичну, медіану, стандартне відхилення, перший і третій квартилі, а також обсяг сукупності і мінімальне та максимальне значення можна швидко визначити за допомогою так званих *блокових статистик* (*Statistics of Block Data*), що здійснюється за таким алгоритмом.

1. Виділити потрібні дані. Це може бути будь-який діапазон даних, в якості якого може бути діапазон клітинок, одна або кілька змінних (стовпчиків) або кілька спостережень (рядків).

2. Виконати команду **Statistics ► Statistics of Block Data ► Block Columns (Block Rows) ►...** Якщо потрібно розрахувати всі можливі показники з меню, то слід вибрати останній його пункт “All” (всі), якщо – якийсь конкретний, то вибрати потрібний пункт меню.

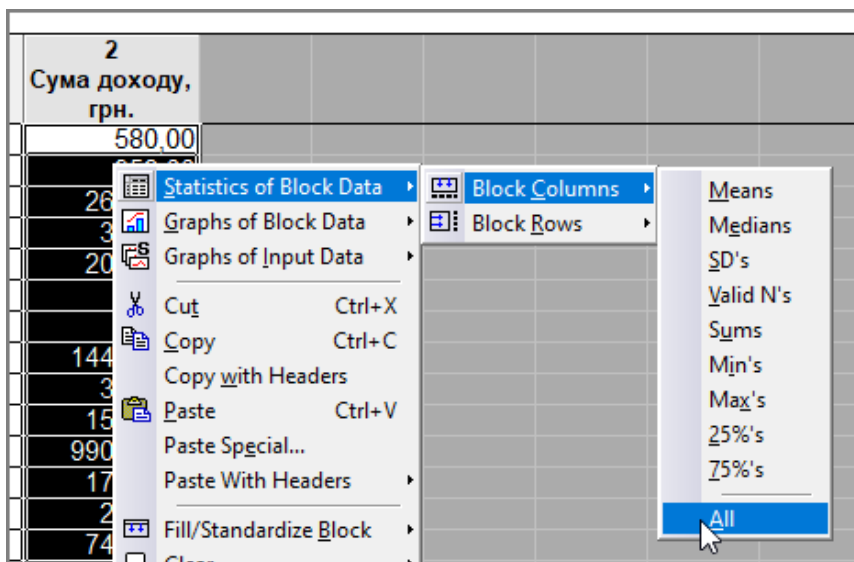


Рисунок 4.4. Доступ з контекстного меню до бокових статистик

Ще простіше ті самі дії можна виконати, якщо викликати контекстне меню виділеного діапазону даних і вибрати послідовно пункти **Statistics of Block Data** ► **Block Columns (Block Rows)** ► ...

Розрахунок всіх описових статистик здійснюється в модулі “Bases Statistics/Tables”, для чого слід виконати такі дії.

1. Виділити змінну (або змінні), для якої слід розрахувати показники.

2. Завантажити модуль “Bases Statistics/Tables” (Основні статистики і таблиці»). З’явиться вікно “Bases Statistics/Tables:”.

3. Вибрати зі списку вікна пункт “Descriptive Statistics” (описові статистики).

4. Натиснути кнопку «**Summary**». З’явиться вікно “Descriptive Statistics”.

5. За замовчуванням розраховуються середня, стандартне відхилення, а також визначаються обсяг вибірки, максимальне і мінімальне значення. Якщо необхідно розрахувати інші описові статистики, то слід перейти на вкладку “Advanced” (Додатково) і встановити позначку біля показників, що необхідно розрахувати.

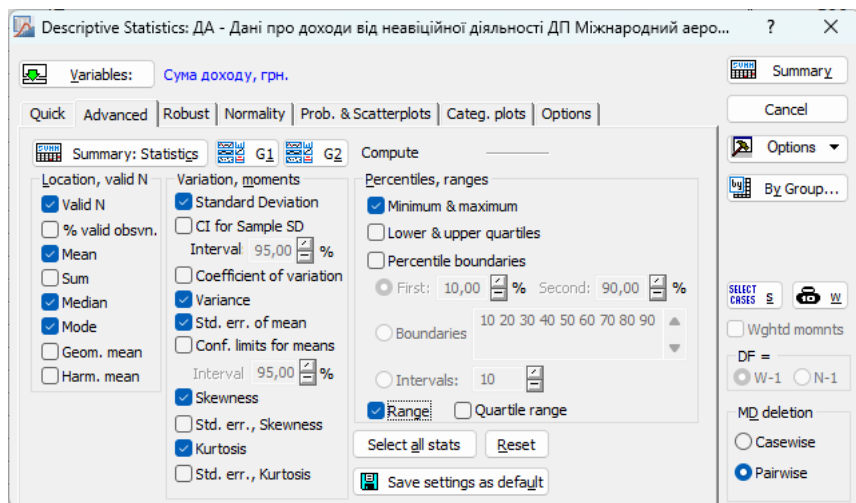


Рисунок 4.5. Вибір описових статистик для розрахунку

6. Розрахунок показників ініціюється натисканням кнопки «Summary», після чого з'явиться вікно з таблицею результатів на ім'я “Descriptive Statistics”. Розрахунок також ініціюється натисканням кнопки «Summary: Descriptive statistics», яка доступна на будь-якій вкладці.

Variable	Valid N	Mean	Median	Mode	Frequency of Mode	Minimum	Maximum	Range	Variance	Std Dev.	Std. Err.
Сума доходу, грн.	122	52011.14	2345.160	1500.000	15	5.100000	3551338	3551333	1.087333E+11	329747.4	2

Рисунок 4.6. Результати розрахунку описових статистик

Variable	Std Dev	Confidence SD 95.000%	Confidence SD +95.000%	Coef Var	Standard Error	Skewness	Std Err Skewness	Kurtosis	Std Err Kurtosis
Сума доходу, грн.	329747.4	293827.3	375752.3	633.9938	29032.64	9.744835	0.213214	101.4849	0.423335

Рисунок 4.7. Результати розрахунку описових статистик.  
Продовження

Після розрахунку вікно з результатами можна просто закрити, а можна зберегти у форматі *STW*. Надалі до збережених результатів можна буде звернутися у будь-який час, не звертаючись при цьому до первинних даних.

При переході до іншого режиму вікно “Descriptive Statistics” не закривається, а згортається на панель стану. В разі потреби до нього можна надалі в будь-який момент звернутися, навіть якщо закрити вікно з результатами. При цьому у вікні зберігаються всі встановлені раніше параметри.



Якщо вікно з результатами аналізу не закривати, то спроба повторного звернення до модуля “Bases Statistics/Tables” призведе до появи вікна-попередження, в якому вказується, що аналіз такого типу вже виконується, і буде запропоновано або продовжити цей аналіз або завантажити новий.

За потреби під час розрахунку описових статистик можна виконати додаткові дії:

1. Здійснити групування даних. Інструмент (кнопка) «**Frequency tables**» (Таблиці частот) дозволяє побудувати таблиці частот. Алгоритм побудови групування було описано вище.

2. Побудувати графік за допомогою інструмента (кнопки) «**Histograms**» (Гістограми). При цьому на вкладках “Normality” (*Нормальність*) і “Prob. & Scatterplots” (*Діаграми*) можна задати додаткові параметри побудови гістограм і діаграм розсіяння, наприклад, побудову 3-вимірної гістограми “3D histograms” (3-В гістограми).

3. Розрахувати критерії для перевірки нормальності Колмогорова-Смірнова і Шапіро-Уїлка, а також одержати теоретичні частоти для нормального розподілу.

## Розрахунок середніх для категоріальних змінних

Для категоріальних змінних достатньо часто буває необхідним розрахувати показники відразу для кожної групи. *STATISTICA* має механізм, що дозволяє задати необхідність розрахунку показників для таких змінних.

Це здійснюється так.

В правій частині вікна аналізу “Descriptive Statistics” слід натиснути кнопку «**By Group**», що приведе до появи однойменного вікна.

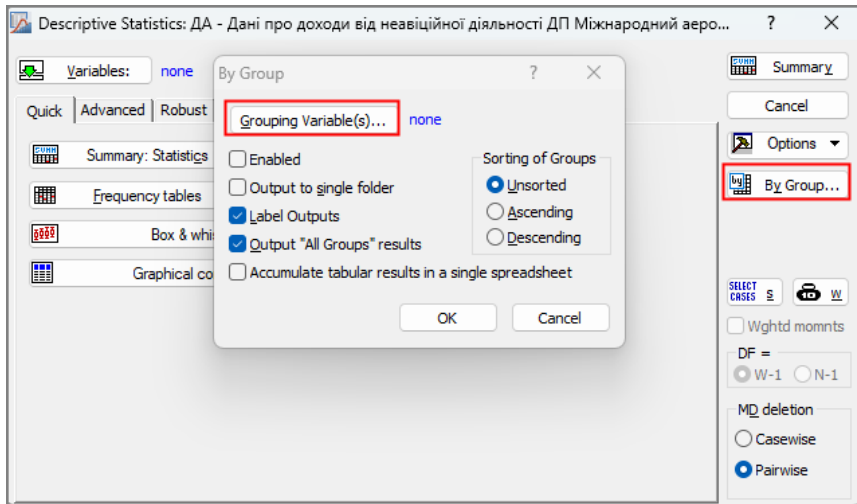


Рисунок 4.8. Вибір груп для категоріальних змінних

У цьому вікні знаходиться кнопка «**Grouping Variable(s)**» (групуюча змінна/і), натискання якої призводить до появи вікна вибору категоріальної змінної.

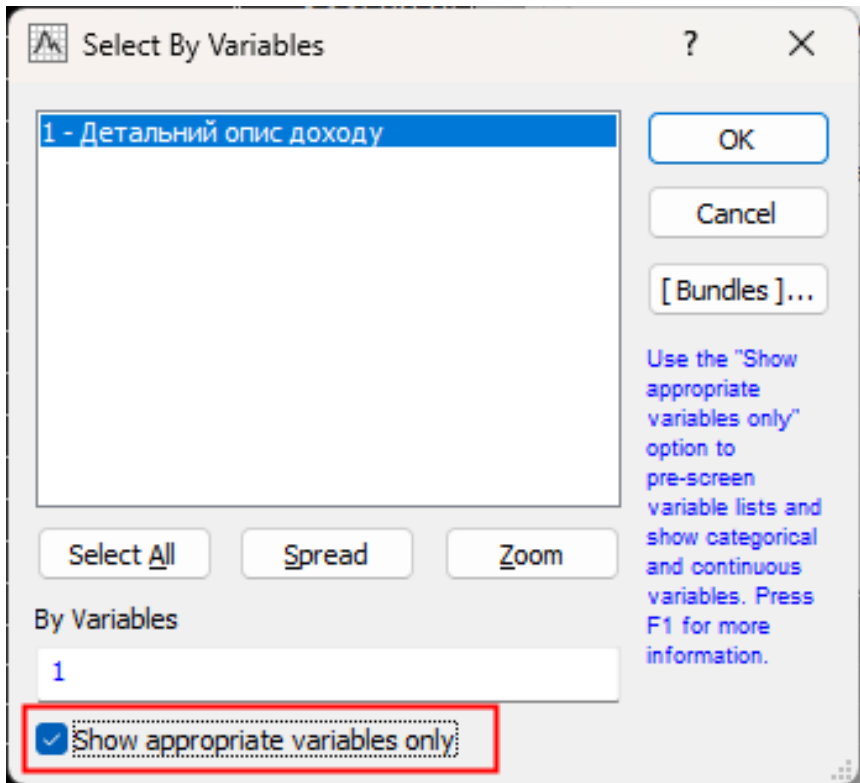


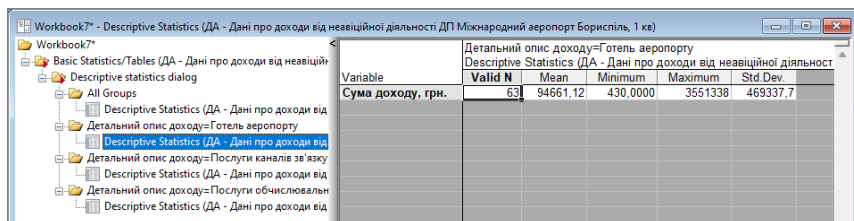
Рисунок 4.9. Вибір конкретної групи для категоріальних змінних

Зверніть увагу на наявність позначки біля поля-мітки “Show appropriate Variables only” (Показувати тільки відповідні змінні). *STATISTICA* здійснює інтелектуальний аналіз переліку змінних і пропонує користувачу вибір з тих змінних, які, на її погляд, є найбільш придатними у цьому випадку. Не завжди такий вибір є вдалим. Тому, якщо в запропонованому переліку змінна відсутня, то слід зняти прапорець біля вище вказаного поля-мітки, після чого буде відображений весь перелік змінних. Але для даного прикладу

цей аналіз як раз спрацював вдало, залишивши дійсно тільки атрибутивну змінну і відкинувши числову.

Після вибору групувальної змінної закриваємо вікно вибору цієї змінної і натискаємо кнопку «**Summary**».

Описаний механізм дозволяє одержати показники як в цілому для змінної, так і для кожної її групи:



The screenshot shows the SPSS 'Descriptive Statistics' dialog box and its output table. The dialog box is titled 'Descriptive Statistics (ДА - Дані про доходи від неавіаційної діяльності ДП Міжнародний аеропорт Бориспіль, 1 кв)'. The output table is titled 'Детальний опис доходу=Готель аеропорту' and shows the following data:

Variable	Valid N	Mean	Minimum	Maximum	Std. Dev.
Сума доходу, грн.	63	94661.12	430.0000	3551338	469337.7

Рисунок 4.10. Результати розрахунку описових статистик для кількох категоріальних змінних

## Термінологічний словник

**Адитивність** (від лат. Additivus – додатковий). Властивість величин, яка полягає в тому, що значення величини, яке відповідає цілому об'єкту, дорівнює сумі значень величин, що відповідають його частинам.

**Альтернативна ознака.** Ознака, значення якої притаманні одній частині сукупності і не притаманні іншій.

**Асиметрія (Skewness).** Показник, що характеризує ступінь зміщення ряду розподілу відносно середнього значення за величиною і напрямком. У симетричній кривій розподілу коефіцієнт асиметрії дорівнює нулю, якщо цей коефіцієнт значно відрізняється від «0», розподіл є асиметричним.

**Варіанта** кількісної ознаки. Конкретні числові значення, які вона набуває.

**Генеральна сукупність.** Сукупність, що містить всі без винятку елементи за якими здійснюється статистичне дослідження.

**Групування.** Процес розподілу сукупності на групи з метою виявлення спільних рис і суттєвих відмінностей серед елементів сукупності.

**Децилі.** Значення варіант, які ділять упорядкований ряд за обсягом на десять рівних частин.

**Дисперсія.** Величина розсіювання значень ознаки відносно середнього значення.

**Експлікація.** Загальна назва графіка, підписи уздовж осей та інші словесні пояснення щодо графіка.

**Екссес (Kurtosis).** Показник, що характеризує ступінь концентрації значень ознаки навколо середнього значення і є своєрідною мірою крутизни кривої. У кривій нормального розподілу екссес дорівнює нулю. Якщо екссес більший нуля – крутизна розподілу вище нормального (гостровершинний розподіл), якщо менший, то крива розподілу характеризується плосковершинністю.



**Інтервал.** Різниця між двома сусідніми значеннями групувальної ознаки, які називають нижньою та верхньою межами інтервалу.

**Категоріальна ознака.** Інша назва номінальної ознаки.

**Квартилі.** Значення варіант, які ділять упорядкований ряд за обсягом на чотири рівних частини. Наприклад,  $Q_1$  – перший квартиль, що містить 25 % спостережень.

**Крива розподілу.** Графічне зображення ряду розподілу.

**Кумулятивна частота (частка).** Накопичена частота (частка), яка є наростаючим підсумком частот (часток).

**Медіана.** Варіанта, що ділить ряд розподілу на дві рівних частини за чисельністю сукупності. Медіана є другим квартилем.

**Мода.** Варіанта, який в ряду розподілу відповідає найбільша частота.

**Номінальна ознака.** Ознака, за якою елементи сукупності не можна упорядкувати у певному порядку. Номінальні ознаки називають також категоріальними.

**Номінальна шкала.** Шкала вимірювання якісних ознак, за якою порядок розташування її елементів не має значення.

**Огіва.** Графічне зображення ряду розподілу за накопиченими частотами або частками.

**Полігон.** Графік за точками, коли точки з'єднуються прямими лініями, а самі точки (вузли) не виділяються.

**Порядкова ознака.** Якісна ознака, яка може бути ранжована в спадному або висхідному порядку.

**Порядкова шкала.** Шкала вимірювання якісних ознак, за якою встановлюється порядок розташування елементів ознаки. Значення ознаки такої шкали послідовно збільшується або зменшується.

**Процентиль.** Значення ознаки, що поділяють впорядковану сукупність на 100 рівних частин.

**Ранг.** Порядковий номер значень ознаки, розташованих у порядку зростання або зменшення.

**Рангова шкала.** Інша назва порядкової шкали.

**Розмах варіації.** Різниця між найменшим і найбільшим значенням ознаки в сукупності.

**Ряд розподілу.** Групування, що складається тільки зі значень групувальної ознаки і їх частот.

**Середнє гармонійне.** Величина обернена до середньої арифметичної.

**Середнє гармонічне.** Інша назва середнього гармонійного.

**Середня величина.** Узагальнюючий показник, що характеризує типовий рівень варіюючої ознаки в якісно однорідній сукупності.

**Сплайн** (від англ. spline – планка, рейка). Стосовно графіка: агрегатна функція, що збігається з функціями більш простої природи на кожному елементі розбиття області визначення графіка.

**Стандартне відхилення середньої.** Величина, на яку відрізняється середнє значення вибірки від середнього значення генеральної сукупності за умови, що розподіл є наближеним до нормального.

**Статистичний графік.** Засіб подання статистичних даних за допомогою різноманітних графічних знаків та символів.

**Статистична сукупність.** Множина елементів, поєднаних спільними умовами існування та розвитку.

**Статистичний показник.** Узагальнююча характеристика явища, в який поєднується кількісна і якісна їх визначеність.

**Частоти.** Значення, що показують, скільки разів повторюються певні значення варіанти у групуванні.

**Шкала.** Засіб кількісного вимірювання та / або впорядкування ознак.

## Лабораторні роботи

### *Загальні вказівки до виконання:*

1. «Індивідуальний номер» – Ваш номер у навчальному журналі групи (курсу).
2. Всі роботи зберігаються у власній папці.

### *Лабораторна робота №1*


#### **Тема. Знайомство з програмою. Введення даних**

**Мета.** Навчитися створювати файл даних, вводити дані спостереження, а також здійснювати їх первинний аналіз.


#### **Вказівки до виконання:**

Щоб одержати список змінних у вікні додавання нової змінної можна здійснити подвійне клацання мишею в полі “After” або натиснути клавішу <F2>.

#### **Завдання**

1. Завантажте *STATISTICA*.
2. Створіть файл даних (електронну таблицю) для двох змінних та п’яти спостережень.
3. Створіть змінні «Чисельність персоналу» та «Обсяг продукції, млн грн» використовуючи дані таблиці 4. Тип даних – цілий (Integer). Зверніть увагу на ім’я змінної, яке створює система за замовчуванням.  Якщо під час введення даних виникнуть проблеми, то з’ясуйте причину їх появи (проаналізуйте типи даних) і виправіть помилки.
4. Додайте після змінної «Обсяг продукції, млн грн» змінну «Обсяг продукції в розрахунку на одного працюючого, грн».
5. Створіть формулу для автоматичного обчислення змінної «Обсяг продукції в розрахунку на одного працюючого, грн» за існуючими змінними.
6. Відобразіть у заголовку даних імена змінних разом:
  - з типами даних;
  - з формулами.

7. Відобразіть таблицю змінних з усіма їх властивостями.
8. Введіть загальну інформацію про таблицю з даними: «Дані про обсяг продукції і чисельність персоналу», а потім Ваше прізвище і групу. Застосуйте до тексту курсив і напівжирне накреслення.
9. Виконайте налаштування системи для того, щоб дані у клітинках з формулами при зміні даних, на підставі яких вони обчислюються, не розраховувалися автоматично, тобто змінювалися тільки в ручному режимі.
10. Введіть в електронну таблицю для змінних «Чисельність персоналу» та «Обсяг продукції, млн грн» частину даних з таблиці 1 та переконайтеся, що значення змінної «Обсяг продукції в розрахунку на одного працюючого, грн» не розраховуються.
11. Розрахуйте дані змінної «Обсяг продукції в розрахунку на одного працюючого, грн» вручну.
12. Виконайте налаштування системи для того, щоб дані у клітинках з формулами при зміні даних, на підставі яких вони обчислюються, розраховувалися автоматично.
13. Введіть в електронну таблицю для змінних «Чисельність персоналу» та «Обсяг продукції, млн грн» дані з таблиці Л.4, що залишилися, і переконайтеся, що значення змінної «Обсяг продукції в розрахунку на одного працюючого, грн» під час введення даних розраховуються автоматично.
14. Впорядкуйте значення змінної «Чисельність персоналу» за спаданням, а потім за зростанням. Для повернення до початкового розташування даних достатньо здійснити відміну виконаної дії (сортування).
15. Здійсніть налаштування системи для автоматичного збереження даних, інтервал кількості хвилин для збереження задайте на власний розсуд.
16. Збережіть таблицю з даними з довільним іменем (але надалі у лабораторних роботах вона буде мати назву «Обсяг продукції»).

17. Закрийте таблицю з даними.
18. Створіть нову таблицю з даними.
19. Створіть змінну "Заробітна плата робітників". Тип даних – байтовий (Byte).
20. За наведеними нижче даними визначте потрібну кількість спостережень і додайте їх в таблицю з даними.
21. Введіть дані щодо заробітної плати: 13300, 12600, 13100, 12000, 12400, 13200, 12500, 22000, 12600, 12500, 20600, 20300, 12600, 13200, 24000, 13400, 21400, 12300, 20200, 13100, 31000, 13000, 23000, 18000, 40000.  Якщо під час введення даних виникнуть проблеми, то з'ясуйте причину їх появи (проаналізуйте типи даних) і виправіть помилки.
22. У модулі перевірки даних (**Data ► Verify Data**) здійсніть аналіз даних з метою виявлення невірних даних. Визначте з цієї метою дві умови, згідно з якими значення змінної мають міститися в інтервалі від 10000 до 31000 включно.
23. Якщо під час аналізу такі значення будуть знайдені, то замініть їх на найближчі до них значення умови.
24. Збережіть таблицю з даними:
  - у вигляді робочої книги *STATISTICA* (ім'я таблиці – довільне, але надалі у лабораторних роботах вона буде мати назву «Заробітна плата»);
  - у вигляді Web-сторінки;
  - як файл електронної таблиці Excel.
25. Додайте змінну «Категорія робітників». У вікні властивостей змінної задайте такі її властивості:
  - Тип даних – байтовий.
  - Натисніть кнопку «**Text Labels**» (текстові мітки) і у вікні редактора текстових міток створіть три текстові мітки (значення у полі «Число» (Numeric) залишайте таким, яким його створює система автоматично, наприклад, «101»):
    - 25.1. Керівники.
    - 25.2. Службовці.
    - 25.3. Робітники.

26. Введіть дані для змінної «Категорія робітників»: 101, 102, 103, 103, 102, 102, 103, 102, 103, 103, 102, 103, 103, 102, 102, 101, 103, 103, 102, 102, 102, 102, 103, 103, 101. Загальна кількість даних має збігатися із загальною кількістю даних змінної «Заробітна плата робітників».

27. Перемкніть відображення даних з числових значень на текстові і навпаки.

28. Збережіть таблицю з даними.

Таблиця Л.4

### Початкові дані до теми «Введення даних»

Чисельність персоналу	Обсяг продукції, млн грн
890	200,05
845	188, 18
1130	307, 77
925	224, 24
700	112, 11



#### Контроль знань та навичок

Після виконання лабораторної роботи студент повинен **знати**:

1. Як називають файл з даними *STATISTICA*?
2. Що являє собою «спостереження» у таблиці з даними?
3. Що являє собою «змінна» у таблиці з даними?
4. Як створити нову таблицю з даними?
5. Яке ім'я за замовчуванням надає система новій змінній?
6. Чи можна змінити розмірність вже створеної таблиці із даними, додаючи або вилучаючи спостереження та змінні?
7. Як вилучити змінну?
8. Як додати нову змінну у конкретному місці?
9. Як додати або вилучити спостереження?
10. Що таке «властивості» змінної?

11. Які типи даних є у *STATISTICA*?
12. У чому відмінність між числовими типами даних?  
Наведіть приклади.
13. Як можна ввести не числові дані?
14. Що таке «Текстові мітки»?
15. Як відобразити таблицю змінних з усіма їх властивостями?
16. Яким чином можна відобразити у клітинці з ім'ям змінної тип даних і розрахункові формули?
17. Для чого призначена загальна інформація про таблицю з даними?
18. Як ввести загальну інформацію про таблицю з даними?
19. Коли доцільно подати змінну у вигляді формули?  
Наведіть приклад.
20. З яких елементів складається розрахункова формула для змінної?
21. Як для змінної створити розрахункову формулу?
22. Як відобразити не числові (текстові) дані у текстовому вигляді або вигляді їх числових еквівалентів?
23. Які дії слід застосувати для того, щоб дані в клітинках із формулами обчислювалися автоматично?
24. Які дії слід застосувати для того, щоб дані в клітинках із формулами обчислювалися в ручному режимі?
25. Як здійснити сортування (упорядкувати) значення змінної?
26. Як здійснити аналіз даних, накладаючи для цього умови на значення даних?
27. В яких форматах можна зберігати файл з даними *STATISTICA*?

Після виконання лабораторної роботи студент повинен **уміти:**

1. Створити нову таблицю з даними.
2. Змінити розмірність таблиці з даними.

3. Вилучити або додати нову змінну в потрібному місці.
4. Додати або вилучити спостереження.
5. Відобразити таблицю змінних з усіма їх властивостями.
6. Відобразити у клітинці з ім'ям змінної тип даних і розрахункові формули.
7. Ввести загальну інформацію про таблицю з даними.
8. Створити для змінної розрахункову формулу.
9. Ввести текстові дані; створювати для них текстові мітки.
10. Відобразити текстові дані у текстовому вигляді або у вигляді їх числових еквівалентів.
11. Здійснити налаштування системи так, щоб дані в клітинках із формулами обчислювалися автоматично або в ручному режимі.
12. Відсортувати (упорядкувати) значення змінної.
13. Здійснити аналіз даних, накладаючи для цього умови на значення даних.
14. Зберегти файл з даними у вигляді Web-сторінки та в форматі електронної таблиці Excel.

## *Лабораторна робота №2*

### **Тема. Частотний аналіз. Побудова графіків**

**Мета.** З'ясувати принципи групування даних і застосування їх для аналізу даних.

#### **Завдання**

1. Завантажте таблицю з даними «Заробітна плата».
2. Визначте мінімальне та максимальне значення для змінної, здійснивши впорядкування (сортування) змінної.
3. За одержаними значеннями вручну розрахуйте кількість груп за формулою Стерджеса.
4. Вручну розрахуйте ширину (рівного) інтервалу.
5. Зверніться до модуля “Frequency tables” (Частотний аналіз) пункт меню “Statistics” різними шляхами (після появи



вікна “Frequency tables” у перших двох випадках просто закривайте його):

- через команду основного меню;
- використовуючи панель інструментів “Statistics” (за відсутності панелі виведіть її використовуючи стандартну команду з пункту головного меню “View”;

- застосуванням кнопки  «Start menu», (Викликати меню засобів, що найчастіше використовуються).

6. Відкрийте вікно модуля “Frequency tables” пункт меню “Statistics”. Виконайте у модулі наведені нижче дії. Всі одержані результати під час виконання цих дій зберігайте в *одній* робочій книзі, по закінченню виконання усіх дій збережіть на носії саму робочу книгу. Виконайте такі дії:

- Побудуйте таблицю частот і гістограму не змінюючи параметрів модуля. З’ясуйте, які саме показники (окрім частот) розраховуються системою автоматично за замовчуванням. Перейдіть на вкладку “Options” і перегляньте, які ще статистики можна розрахувати в модулі.

- Побудуйте таблицю частот і гістограму за розрахованою Вами кількістю груп. Проаналізуйте відмінності цих даних від попередніх.

- Побудуйте таблицю частот і гістограму за розрахованим Вами кроком інтервалу; як початкове значення першого інтервалу виберіть мінімальне з усіх значень.

- Перевірте розподіл на нормальність за критерієм Шапіро-Уїлка. Для цього перейдіть на вкладку “Normality” (Нормальність), встановіть позначку для поля “Shapiro-Wilk’s W test” (Критерій Шапіро-Уїлка) і натисніть кнопку «Test for normality» (Тест на нормальність).

7. Зверніться до модуля “2D Histograms” пункт меню “Graphs” (Графіка). Виконайте у модулі наведені нижче дії. Всі створені діаграми під час виконання цих дій зберігайте в *одній* робочій книзі, по закінченню виконання усіх дій збережіть на диску саму робочу книгу. Виконайте такі дії:

- Побудуйте гістограму не змінюючи параметрів модуля. Визначте, які спостереження система включила до кожного інтервалу (для цього достатньо навести курсор на стовпчик гістограми).

- Побудуйте гістограму за розрахованою Вами кількістю груп.

- Відобразить на діаграмі відсотки. Для цього на вкладці “Advanced” (Додатково) встановіть прапорець поруч з полеміткою “Show percentages” (Показати відсотки).

- Перевірте розподіл на нормальність за критерієм Шапіро-Уїлка.

- Побудуйте графічно кумуляту розподілу. Для цього на вкладці “Advanced” (Додатково) зі списку “Showing Type” (Тип, що відображається) виберіть “Cumulative” (Кумулятивний)

- Побудуйте графічно огіву розподілу. При її побудові на осі абсцис відкладають накопичені частоти або частки, а на осі ординат – межі інтервалів варіаційного ряду). Для цього на вкладці “Options 2” (Параметри 2) зі списку “X-Y Axes position” “Позиції осей X-Y” виберіть “Reversed” (Зворотний).

8. З'ясуйте, в якому режимі відображається графік. Перемкніть відображення в інший режим.

9. Перейдіть до режиму редагування графіка і виконайте такі дії:

- змініть загальну назву графіка на Гістограма для змінної «Заробітна плата робітників»;

- знайдіть місце, де можна змінити максимальне та мінімальне значення на осі значень і встановіть мінімальне значення “10000”. Поясніть зміну вигляду графіка.

10. Додайте в таблицю з даними після змінної «Категорія робітників» ще дві змінні: «Значення заробітної плати», «Частоти». Тип даних – числовий.

11. Застосуйте модуль “Bases Statistics/Tables” (Основні статистики і таблиці») (пункт “Frequency tables”) для побудови таблиці частот змінної “Заробітна плата робітників”. Таблицю

будуйте таким чином, щоб для кожного значення змінної створювалася окрема група.

12. Скопіюйте з таблиці частот стовпчик “Category” (Категорії), що містить значення заробітної плати і вставте їх у змінну



“Значення заробітної плати”. Не звертайте увагу, що кількість спостережень для цієї змінної буде менше, ніж кількість спостережень для інших змінних.

13. Скопіюйте з таблиці частот стовпчик “Count” (Частота), що містить розраховані значення часто і вставте їх у змінну «Частоти».

14. Використовуючи дані змінних «Значення заробітної плати» та «Частоти» побудуйте полігон розподілу. Для цього виконайте такі дії:

- Виконайте команду **Graphs ▶ 2D Graphs ▶ Line Plots (Variables)**, (Графіки, 2М Графіки, Лінійні графіки (за змінними)) що приведе до появи вікна “2D Line Plot” (2В Лінійні графіки).

- На вкладці “Advanced” (Додатково):

- 14.1. У групі “Graph Type” (Тип графіку) виберіть “XY Trace”.

- 14.2. Натисніть кнопку «Variables» (Змінні) та визначте значення по осі X («Значення заробітної плати») і Y («Частоти»).

- 14.3. Зі списку “Display point” (Відображати точки) виберіть “On”.

15. Перетягніть зі списку вікна робочої книги назву графіка за межі вікна. Збережіть нове вікно з графіком у якомусь графічному форматі. За допомогою контекстного меню даного вікна можна відобразити перелік дій, які можна здійснити з його вмістом.

16. Збережіть вікно з графіком:

- у графічному форматі *STATISTICA*;
- у форматі GIF;
- у форматі BMP.

17. Збережіть файл даних.



## Контроль знань та навичок

1. Як можна звернутися до конкретного модуля?
2. Для чого призначений модуль “Frequency tables” (Частотний аналіз)?
3. Як називають вікно виведення результатів у *STATISTICA*?
4. Яка структура вікна виведення результатів?
5. Як побудувати таблицю частот?
6. Як задати ширину (крок) інтервалу для побудови групування?
7. Які можна задати початкові значення для першого інтервалу групування?
8. Як задати кількість груп для побудови групування?
9. Які показники можна одержати у модулі “Frequency tables”?
10. Які варіанти побудови групувань надає користувачу модуль “Frequency tables”?
11. Як здійснити перевірку розподілу на нормальність за допомогою критерію Шапіро-Уїлка у модулі “Frequency tables”?
12. Як побудувати гістограму у модулі “Frequency tables”?
13. Які способи побудови гістограми має програма?
14. Як побудувати гістограму у модулі “2D Histograms”?
15. Як задати кількість інтервалів групування у модулі “2D Histograms”?
16. Як у модулі “2D Histograms” визначити спостереження, які потрапили до конкретного інтервалу діаграми?
17. Як у модулі “2D Histograms” відобразити відсотки на діаграмі?
18. Як здійснити перевірку розподілу на нормальність за допомогою критерію Шапіро-Уїлка у модулі “2D Histograms”?
19. Що таке кумулята розподілу?
20. Як побудувати графічно кумуляту розподілу у модулі “2D Histograms”?
21. Що таке огіва?

22. Як побудувати огіву у модулі “2D Histograms”?
23. Які є режими відображення графіка?
24. В якому режимі за замовчуванням відображається графік?
25. Як в режимі редагування графіка змінити загальну назву графіка, назви його осей?
26. Як в режимі редагування графіка змінити максимальне та мінімальне значення на осі абсцис?
27. Які результати аналізу можна зберігати в одній робочій книзі?
28. Як у файл з даними додати нову змінну?
29. У яких форматах можна зберегти графік?

Після виконання лабораторної роботи студент повинен **уміти:**

1. Звернутися до конкретного модуля різними способами.
2. Побудувати таблицю частот.
3. Змінити ширину (крок) інтервалу для побудови групування.
4. Визначити початкові значення для першого інтервалу групування.
5. Задати кількість груп для побудови групування.
6. Побудувати групування різними способами у модулі “Frequency tables” (Частотний аналіз).
7. Здійснити перевірку розподілу на нормальність за допомогою критерію Шапіро-Уїлка у модулі “Frequency tables”.
8. Побудувати гістограму.
9. Задати кількість інтервалів групування у модулі “2D Histograms”.
10. Визначити спостереження, які потрапили до конкретного інтервалу діаграми.
11. Відобразити відсотки на діаграмі.
12. Здійснити перевірку розподілу на нормальність за допомогою критерію Шапіро-Уїлка у модулі “2D Histograms”.

13. Побудувати графічно кумуляту розподілу.
14. Побудувати огіву.
15. Змінити режими відображення графіка.
16. Здійснити редагування графіка, зокрема змінити загальну назву графіка, назви його осей.
17. У режимі редагування графіка змінити максимальне та мінімальне значення на осі абсцис.
18. Додати нову змінну у файл з даними.
19. Зберегти результати аналізу в одній робочій книзі.
20. Зберегти графік у різних форматах.

### *Лабораторна робота №3*

#### **Тема. Описові статистики**

**Мета.** З'ясувати принципи розрахунку описових статистик та використання їх для аналізу даних.

**Вказівки до виконання:**

1. Визначення переліку додаткових статистик, що необхідно розрахувати, в модулі “Descriptive Statistics” (Описові статистики) аналізу “Bases Statistics/Tables” (Основні статистики і таблиці) здійснюється на вкладці “Advanced” (Додатково).


2. Англomовні скорочення на діаграмі розмаху:

- SE – Standard error of estimate Standard (Стандартна похибка оцінювання).
- SD – Standard Deviation (Стандартне відхилення).

**Завдання**

1. Завантажте таблицю з даними «Обсяг продукції».
2. Застосуйте блокові статистики для змінних «Чисельність персоналу» та «Обсяг продукції, млн грн» з метою визначення:
  - середніх величин;
  - всіх описових статистик.
3. Завантажте таблицю з даними «Заробітна плата».
4. Застосуйте блокові статистики з метою визначення середньої величини для перших десяти спостережень.

5. Застосуйте модуль “Histograms” пункт меню “Graphs” (Графіка) для розрахунку базових описових статистик. З’ясуйте, які саме статистики при цьому автоматично розраховуються. Для розрахунку статистик встановіть позначку поруч з полем-міткою “Descriptive Statistics” (Описові статистики) на вкладці “Advanced” (Додатково) у групі “Statistics”.

6. Застосуйте модуль “Descriptive Statistics” (Описові статистики) аналізу “Bases Statistics/Tables” (Основні статистики і таблиці) для розрахунку базових описових статистик, що розраховуються за замовчуванням. З’ясуйте, що це за статистики.  Виконаний аналіз і всі аналізи, що будуть здійснені надалі, запам’ятовують в одній робочій книзі.

7. Виконайте в цьому модулі такі дії:

- Розрахуйте моду (Mode) та медіану (Median). Порівняйте значення середньої величини, моди та медіани. Вмійте пояснити їх сутність. Проаналізуйте результати аналізу та визначте, що являє собою показник “Frequency of Mode” (Частота моди).

- Розрахуйте перший (нижній) та останній (верхній) квартилі (Lower, Upper Quartile), перший та останній децилі. Дайте пояснення сутності цих показників.

- Розрахуйте показники варіації: “Range” (Розмах варіації), “Variance” (дисперсія), “Standard Deviation” (стандартне відхилення), і “Std. err. of mean” (стандартна похибка середнього значення). Дайте пояснення сутності показників розмах варіації та дисперсія.

- Розрахуйте характеристики форми розподілу “Skewness” (Асиметрія) і “Kurtosis” (ексцес). На їх підставі зробіть висновки щодо напрямку асиметрії і міри скошеності; вмійте обґрунтувати свої висновки.

- Побудуйте гістограму, під час її побудови задайте для неї розраховану Вами кількість інтервалів. Нанесіть на гістограму напис з Вашим висновком відносно напрямку асиметрії і міри скошеності.

- Побудуйте діаграму розмаху, для чого слід натиснути кнопку «**Box & Whisker plot for all Variables**» (Діаграма розмаху для всіх змінних»). Визначте на вкладці “Options”, які опції для діаграми можна застосовувати в модулі.

8. Побудуйте діаграми розмаху в модулі “2D Box Plots” (2В Діаграми розмаху). Зберігайте в одній робочій книзі всі побудовані надалі діаграми.

- У вікні налаштувань модуля виконайте:

8.1. Визначте, яка змінна буде змінною групування для змінної «Заробітна плата робітників». Задайте ці змінні.

8.2. З’ясуйте, який саме показник центру розподілу відобразатиметься з графіка, а також форму (стиль) його відображення.

8.3. Побудуйте діаграму.

8.4. Здійсніть аналіз діаграми:

8.4.1. Вмійте пояснити, що відображає кожний з її елементів.

8.4.2. Вмійте пояснити, які саме налаштування і як саме вони формують вигляд діаграми: її вуси, основу коробки, наявність викидів.

8.4.3. Поясніть, чому одна з діаграм не має вусів.

8.4.4. З’ясуйте, як можна визначити, які спостереження складають ту чи іншу групу.

8.5. Дайте відповідь на питання: чим відрізняються можливості побудови діаграми розмаху в модулі від тих, що має користувач в модулі “Descriptive Statistics” (Описові статистики) аналізу “Bases Statistics/Tables” (Основні статистики і таблиці)?

- Змініть стиль відображення центру розподілу на «ризку». Побудуйте діаграму. Здійсніть її аналіз.

- Побудуйте діаграму, вибравши як показник центру розподілу середню арифметичну. Поясніть наявність на «вусах» окремих діаграм точок-викидів.

- Побудуйте таку саму діаграму, але без викидів.



- Побудуйте діаграму для категорій «Службовці» і «Робітники».

9. З'ясуйте, які саме графічні елементи повинна мати діаграма розмаху для змінної «Заробітна плата робітників». Обґрунтуйте свій висновок.



### Контроль знань та навичок

Після виконання лабораторної роботи студент повинен **знати**:

1. Що таке «описові статистики»? Назвіть найбільш вживані.
2. Що таке «блокові статистики»?
3. Які описові статистики можна розрахувати, використовуючи блокові статистики?
4. Як розрахувати описові статистики, використовуючи блокові статистики, в тому числі для групи спостережень?
5. В яких модулях можна розрахувати описові статистики?
6. Який модуль аналізу спеціально призначений для розрахунку описових статистик?
7. Що у статистиці відносять до показників центру розподілу?
8. Як в модулі “Descriptive Statistics” (Описові статистики) аналізу “Bases Statistics/Tables” (Основні статистики і таблиці) розрахувати показники центру розподілу?
9. Призначення показника «Мода».
10. Призначення показника «Медіана».
11. Що являють собою перший (нижній) та останній (верхній) кватилі?
12. Як в модулі “Descriptive Statistics” аналізу “Bases Statistics/Tables” розрахувати перший (нижній) та останній (верхній) кватилі?
13. Що являють собою децилі?
14. Як в модулі “Descriptive Statistics” аналізу “Bases Statistics/Tables” розрахувати децилі?
15. Що у статистиці відносять до показників варіації?

16. Як в модулі “Descriptive Statistics” аналізу “Bases Statistics/Tables” розрахувати показники варіації?
17. Яка сутність показника «Розмах варіації»?
18. Яка сутність показника «Стандартне відхилення»?
19. Яка сутність показника «Стандартна похибка (середнього значення)»?
20. Які є характеристики форми розподілу?
21. Як на підставі характеристики форми розподілу зробити висновки щодо напрямку асиметрії і міри скошеності розподілу?
22. Для чого призначена діаграма розмаху?
23. З яких графічних елементів складається діаграма розмаху?
24. На підставі чого визначається перелік графічних елементів діаграми розмаху?
25. Як в модулі “Descriptive Statistics” аналізу “Bases Statistics/Tables” побудувати діаграму розмаху?
26. Як побудувати діаграму розмаху в модулі “2D Box Plots” (2В Діаграми розмаху)?
27. У чому основна відмінність діаграми розмаху між такою, що будується в модулі “Descriptive Statistics” і такою, що будується в модулі “2D Box Plots”?
28. Що таке «незалежна змінна» і «змінна групування» для неї?
29. Які налаштування в модулі “2D Box Plots” формують вигляд діаграми розмаху?
30. Як можна визначити, які спостереження складають тут чи іншу групу діаграми розмаху?
31. Як можна змінити стиль відображення центру розподілу на діаграмі розмаху?
32. Що являють собою «викиди» на діаграмах розмаху?
33. Як можна відключити відображення викидів на діаграмах розмаху?
34. Як побудувати діаграму розмаху тільки для окремих груп змінної групування?

Після виконання лабораторної роботи студент повинен **уміти:**

1. Розрахувати «блокові статистики».
2. Розрахувати описові статистики за допомогою блокових статистик, у тому числі для групи спостережень.
3. Використати модуль “Descriptive Statistics” (Описові статистики) аналізу “Bases Statistics/Tables” (Основні статистики і таблиці) для розрахунку показників центру розподілу.
4. Використати модуль “Descriptive Statistics” аналізу “Bases Statistics/Tables” для розрахунку квантилів.
5. Використати модуль “Descriptive Statistics” аналізу “Bases Statistics/Tables” для розрахунку децилів.
6. Використати модуль “Descriptive Statistics” аналізу “Bases Statistics/Tables” для розрахунку показників варіації.
7. На підставі характеристики форми розподілу зробити висновки щодо напрямку асиметрії і міри скошеності розподілу.
8. Визначити перелік графічних елементів діаграми розмаху.
9. Використати модуль “Descriptive Statistics” аналізу “Bases Statistics/Tables” для побудови діаграми розмаху.
10. Побудувати діаграму розмаху в модулі “2D Box Plots” (2В Діаграми розмаху).
11. Визначити, які спостереження складають ту чи іншу групу діаграми розмаху.
12. Змінити стиль відображення центру розподілу на діаграмі розмаху.
13. Відключити відображення викидів на діаграмах розмаху.
14. Побудувати діаграму розмаху для окремих груп змінної групування.

## Література

1. Загальна теорія статистики : підручник / за ред. А. В. Непрана, І. А. Дмитрієва; авт. кол.: І. А. Дмитрієв, О. А. Дмитрієва, О. М. Гіржева, А. В. Непран, Н. О. Бірченко, А. А. Воронкова, Н. В. Чуйко. – Харків : ПП Іванченка, 2022. – 720 с.
2. Єріна А.М., Пальян З.О. Статистика: підручник. / А. М. Єріна, З. О. Пальян – К.: КНЕУ, 2010. – 351 с.
3. Мормоза А. Т. Теорія статистики: підручник. / А. Т. Мормоза – 2-е вид. перероб. та доп. – К.: «Центр учбової літератури», 2013. – 592 с.
4. Фетісов В. С. Математичні та статистичні пакети: навчально-методичний посібник. / В. С. Фетісов – Ніжин: Видавець ПП Лисенко М.М., 2011. – 324 с.
5. Фетісов В. С. Пакет статистичного аналізу даних STATISTICA: навчальний посібник. / В. С. Фетісов – Ніжин: Видавництво НДУ ім. М.Гоголя, 2018. – 102 с.
6. Фетісов В. С. Побудова групувань з використанням пакета STATISTICA. – Статистика України, 2019, № 1, с. 121-129.
7. Фетісов В. С. Прикладна статистика: навчальний посібник. / В. С. Фетісов – Ніжин: Видавництво НДУ ім. М. Гоголя, 2019. – 115 с.
8. Фетісов В.С. Прикладні пакети статистичної обробки: лабораторний практикум / В. С. Фетісов – Ніжин: Видавництво НДУ ім. М. Гоголя, 2010. – 27 с.
9. StatSoft. Офіційний сайт. [Електронний ресурс] Режим доступу: <https://www.statistica.com/en/software/statistica-evaluation>

Навчальне видання

**В. С. Фетісов**

**Статистика & STATISTICA**  
**Частина I.**  
**Основи**

*Навчальний посібник*

Технічний редактор – І. П. Борис  
Верстка, макетування – В. М. Косяк

---

Підписано до друку	Формат 60x84/16	Папір офсетний
Гарнітура Times New Roman	Ум. друк. арк. 9,53	Тираж ел. вид.
Замовлення №	Обл.-вид. арк. 5,99	

---



Ніжинський державний університет  
імені Миколи Гоголя.

м. Ніжин, вул. Воздвиженська, 3-А  
(04631) 7-19-72

E-mail: [vidavn\\_ndu@ukr.net](mailto:vidavn_ndu@ukr.net)

[www.ndu.edu.ua](http://www.ndu.edu.ua)

Свідоцтво суб'єкта видавничої справи  
ДК № 2137 від 29.03.05 р.