

Лекція 2

Тема: Обробка вимірних даних



Лекція 2

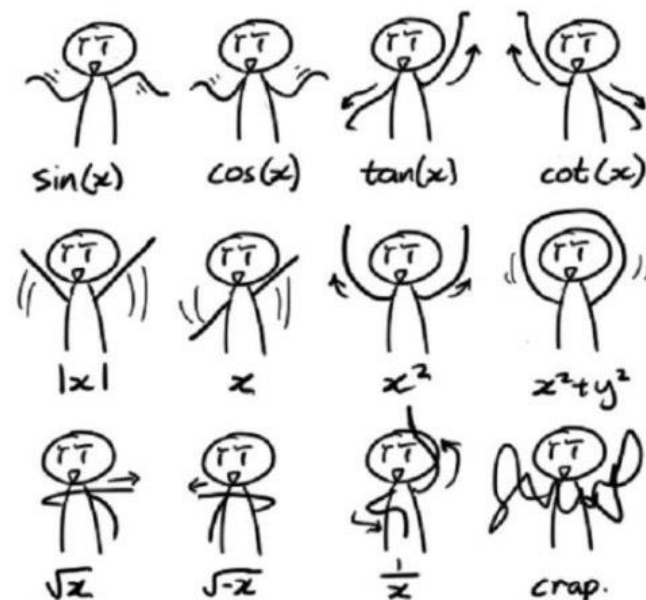
Тема: Обробка вимірених даних

1. Збирання даних

2. Оброблення даних та їх подання у зручній формі (у вигляді таблиці, графіків, діаграм).

3. Аналіз даних.

4. Висновки та рекомендації.



1.Збирання даних

Збирання даних є фундаментальним етапом у процесі роботи з вимірними даними. Від якості та точності зібраних даних залежить надійність подальшого аналізу та коректність висновків. Для збору вимірних даних використовуються різні інструменти залежно від типу вимірюваних величин, а самі дані можуть надходити з різних джерел, таких як прилади для вимірювання фізичних параметрів, автоматизовані системи або ручні методи.

Основною метою збирання даних є отримання точних, повних і надійних даних, які можуть бути використані для подальшого аналізу. Варто зазначити, що неправильно зібрані дані або з даними з високою похибкою можуть суттєво вплинути на кінцеві результати, що робить цей етап критично важливим.

Основні аспекти збирання даних:

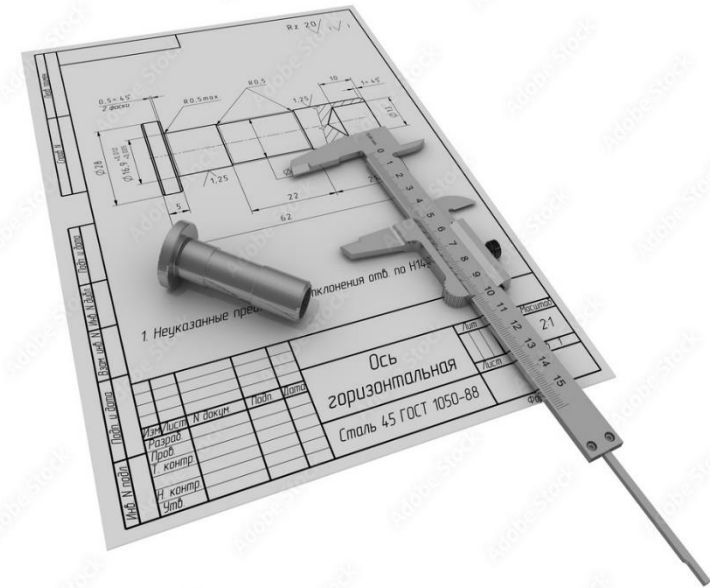
- Джерела даних. Це можуть бути експерименти, дослідження, вимірювальні пристрої, або інформація з автоматизованих систем моніторингу. Важливо переконатися, що джерела даних є надійними.
- Точність вимірювань. Похибки у вимірюваннях можуть суттєво вплинути на результати. Тому важливо використовувати точні методи вимірювання та перевіряти коректність приладів.
- Збереження даних. Дані повинні бути збережені у такій формі, яка дозволить легко працювати з ними надалі. Це можуть бути електронні таблиці, бази даних або хмарні сервіси.
- Попередня перевірка. Перед початком обробки даних їх перевіряють на наявність помилок або пропущених значень, щоб уникнути неправильних результатів.



Джерела та інструменти для збирання даних

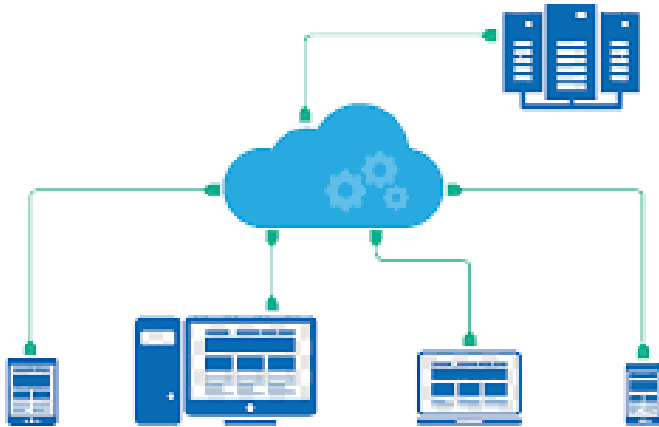
- **Штангенциркуль.** Цей інструмент використовують для вимірювання лінійних розмірів з високою точністю (наприклад, товщини або діаметра). Він дозволяє вимірювати до сотих частин міліметра.
- **Рулетка.** Для вимірювання довжини або відстаней використовується рулетка. Вона є універсальним інструментом для будівельних та геодезичних робіт, а також для домашнього використання.
- **Мікрометр.** Це інструмент, призначений для вимірювання дуже малих відстаней з високою точністю, зазвичай у межах мікрометрів. Використовується в точній механіці та машинобудуванні.
- **Термометр.** Для вимірювання температури застосовують різні типи термометрів, від класичних ртутних до сучасних електронних.
- **Ваги.** Використовуються для вимірювання маси предметів з різною точністю, від аналітичних ваг, що використовуються в лабораторіях, до побутових ваг.
- **Осцилограф.** Це прилад, який використовують для вимірювання електричних сигналів і їх візуалізації у вигляді хвиль. Він застосовується у фізиці, електроніці та телекомунікаціях.
- **Датчики і сенсори.** Сучасні технології дозволяють використовувати датчики для автоматичного збирання даних у різних системах, включаючи промислові процеси, метеорологію та контроль якості. Наприклад, датчики температури, вологості, тиску або швидкості вітру.

Adobe Stock | #53629676



Куди йдуть дані

- **Журнали вимірювань.** Ручні вимірювання, наприклад, за допомогою штангенциркуля або рулетки, часто записуються у спеціальні журнали або таблиці. Це можуть бути паперові бланки або електронні таблиці у програмі, наприклад, Excel.
- **Бази даних.** Автоматизовані системи або великі обсяги вимірювань часто зберігаються у базах даних. Це дозволяє легко організувати та управляти даними для подальшого аналізу.
- **Хмарні сервіси.** Сучасні системи збору даних можуть відразу передавати їх у хмарні сховища, що дозволяє мати доступ до даних з будь-якої точки світу і забезпечує зручне зберігання великих обсягів інформації.
- **Програмне забезпечення для аналізу.** Дані можуть бути безпосередньо інтегровані в спеціалізовані програми для подальшої обробки, наприклад, системи для статистичного аналізу або моделювання.



При **безпосередньому обліку** фактів статистичні дані отримують шляхом особливого обліку одиниць сукупності – огляду, виміру, зважування та записують на індивідуальні карти спостереження.

Документальний облік як первинний ґрунтується на систематичній реєстрації фактів, наприклад, у лікувально-профілактичних закладах. Такі дані з різних офіційних документів викопійовують у карту для вивчення.

Викопіювання даних в розроблений статистичний документ може бути застосоване, наприклад, для отримання інформації про групи осіб, які звертались за медичною допомогою, про самі медичні заклади, їхню діяльність, кадри та з інших питань відповідно до розробленої програми.

Використання технічних засобів обліку медичної інформації та її централізація оптимізують механізми подальшої обробки та аналізу.

Збір медико-статистичної інформації шляхом **опитування** проводять експедиційним та кореспондентським методами та у формі самореєстрації.

При **експедиційному методі** дослідник опитує хворого і з його слів самостійно заповнює карту дослідження, чим забезпечує контроль за правильністю відповідей.

При **самореєстрації** особа, яка обстежується, заповнює карту самостійно.

При **кореспондентському методі** дослідник розсилає карти для обстеження з відповідними вказівками до їх заповнення.

Заповнені карти (з відповідями на запитання) респондент відсилає на адресу дослідника.

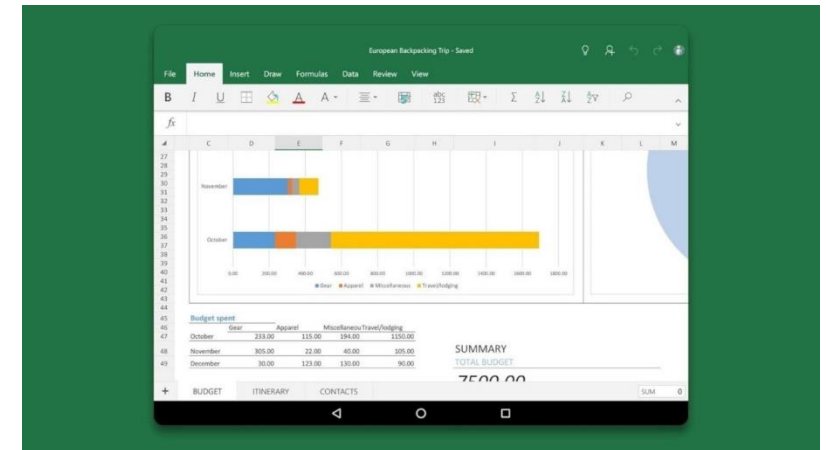
Анкетний метод використовується при неможливості безпосереднього спостереження за досліджуваним явищем.

2. Оброблення даних та їх подання у зручній формі (у вигляді таблиці, графіків, діаграм).

Після збирання даних настає етап їх оброблення. Обробка включає структурування, очищення та перетворення даних у зручний формат для аналізу. Однією з найважливіших частин цього етапу є представлення даних у зручній для сприйняття формі, найчастіше у вигляді таблиць або графіків.

Основні етапи оброблення даних:

- Очищення даних. Спочатку видаляють або коригують помилкові чи пропущені дані. Це включає усунення аномалій, викидів, дублікатів або некоректних записів, щоб дані відповідали реальним вимірюванням і були готові для аналізу.
- Структурування даних. На цьому етапі дані організовуються у логічно зрозумілу структуру. Зазвичай, це формат таблиць, який дозволяє легко виконувати сортування, фільтрацію або групування. Дані повинні бути організовані таким чином, щоб їх можна було швидко проаналізувати або візуалізувати.



Представлення даних у таблицях

Таблиці є основним інструментом для впорядкування числових та текстових даних. Вони допомагають легко організувати великі масиви інформації та робити їх зрозумілими. Стандартні електронні таблиці, такі як Microsoft Excel або Google Sheets, дозволяють:

- Організувати дані у рядки та стовпці. Це дозволяє відображати конкретні значення та робити порівняння між ними.
- Застосовувати функції для обчислень. Наприклад, обчислення середнього значення, суми або відсоткових співвідношень для великих обсягів даних.
- Сортувати та фільтрувати дані. Це допомагає виділити потрібні значення або категорії для аналізу.
- Групування даних. Об'єднання даних за категоріями або параметрами для зручного перегляду та аналізу тенденцій.

Візуалізація даних: діаграми та графіки

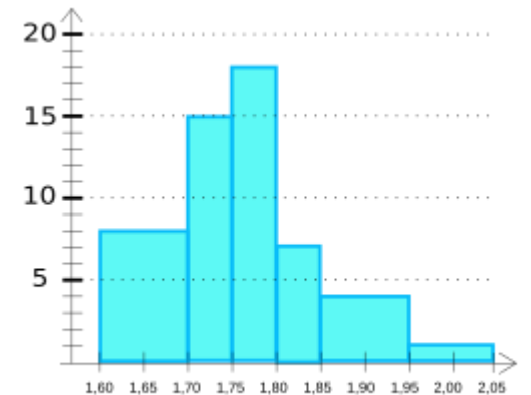
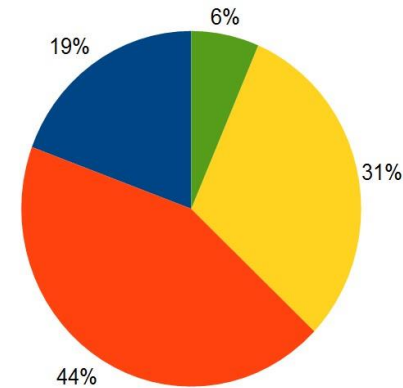
Графіки та діаграми дозволяють наочно відобразити дані та зробити їх легшими для інтерпретації. Це дозволяє виявити закономірності, порівняння або тренди, які важко помітити у вигляді сирих чисел.



Найпоширеніші види візуалізації:

- Лінійні графіки. Використовуються для відображення змін даних у часі. Наприклад, графік може показати, як змінюється температура протягом доби або як змінюються продажі за місяць.
- Стовпчикові діаграми. Допомогають порівнювати кілька категорій або значень між собою. Наприклад, можна порівняти обсяги продажів кількох продуктів за рік.
- Кругові діаграми. Відображають пропорції або частки кожної категорії від загального значення. Наприклад, кругова діаграма може показати, як різні джерела доходів складають загальний бюджет.
- Гістограми. Використовуються для відображення частот розподілу даних. Наприклад, можна відобразити, як часто трапляються певні значення у наборі даних (наприклад, кількість клієнтів з певним рівнем доходу).

Таблиці та діаграми дозволяють зробити складну інформацію зрозумілою, а також допомагають приймати рішення на основі наочної візуалізації ключових показників.



3. Аналіз даних.

Аналіз даних – це процес вивчення та обробки зібраної інформації для виявлення тенденцій, закономірностей, взаємозв'язків і аномалій. У цьому пункті розглянемо основні базові методи аналізу даних, а також використання формул у Excel та Google Таблицях, які є основними інструментами для обробки та аналізу великих обсягів даних.

Основні методи аналізу даних та формули в Excel/Google Таблицях

1. Описова статистика

Описова статистика допомагає виявити ключові характеристики даних, такі як середнє значення, медіана, мода, мінімум, максимум та розмах.

•**Середнє значення (Average):** Формула: =AVERAGE(діапазон) Наприклад, якщо потрібно обчислити середнє значення для діапазону комірок A2

, формула буде виглядати так:

=AVERAGE(A2:A10)

•**Медіана (Median):** Формула: =MEDIAN(діапазон) Медіана – це центральне значення у впорядкованому наборі даних.

Приклад:

=MEDIAN(A2:A10)

•**Мода (Mode):** Формула: =MODE(діапазон) Мода – це значення, яке зустрічається найчастіше у наборі даних.

Приклад:

=MODE(A2:A10)

•**Мінімум і максимум:** Формули для мінімального та максимального значень:

•=MIN(діапазон) для мінімального значення.

•=MAX(діапазон) для максимального значення.

Наприклад:

=MIN(A2:A10)

=MAX(A2:A10)

2. Кореляційний аналіз

Кореляція визначає ступінь зв'язку між двома змінними. Для цього можна використовувати функцію **CORREL**.

•**Кореляція (Correl):** Формула: =CORREL(діапазон_1, діапазон_2)

Наприклад, якщо ви хочете обчислити кореляцію між значеннями у стовпцях А і В:

=CORREL(A2:A10, B2:B10)

Кореляційний аналіз допомагає виявити, чи існує зв'язок між двома змінними і наскільки цей зв'язок сильний. Кореляція варіюється від -1 (сильна негативна) до 1 (сильна позитивна), де 0 вказує на відсутність кореляції.

3. Трендовий аналіз

Трендовий аналіз полягає у вивченні змін у часі для прогнозування майбутніх показників.

•**Лінійна регресія:**

Формула: =LINEST(діапазон_Y, діапазон_X)

Функція **LINEST** дозволяє створити рівняння прямої, яке найкраще описує дані.

Наприклад, якщо у вас є дані про продажі (стовпець В) і час (стовпець А), ви можете використовувати формулу:

=LINEST(B2:B10, A2:A10)

Для побудови графіку тренду можна скористатися функцією **Тренд** у самій програмі Excel або Google Таблицях. Для цього вибирають дані і використовують інструмент побудови графіку з трендовою лінією.

4. Стандартне відхилення

Стандартне відхилення показує, наскільки значення в наборі даних відхиляються від середнього.

•**Стандартне відхилення (STDEV.P або STDEV.S):**

•=STDEV.P(діапазон) для повної сукупності даних.

•=STDEV.S(діапазон) для вибіркової сукупності.

Наприклад:

=STDEV.S(A2:A10)

5. Сумарні значення

Часто при аналізі даних необхідно обчислити суму значень.

•**Сума (SUM):** Формула: =SUM(діапазон)

Наприклад:

=SUM(A2:A10)

6. Підрахунок кількості значень

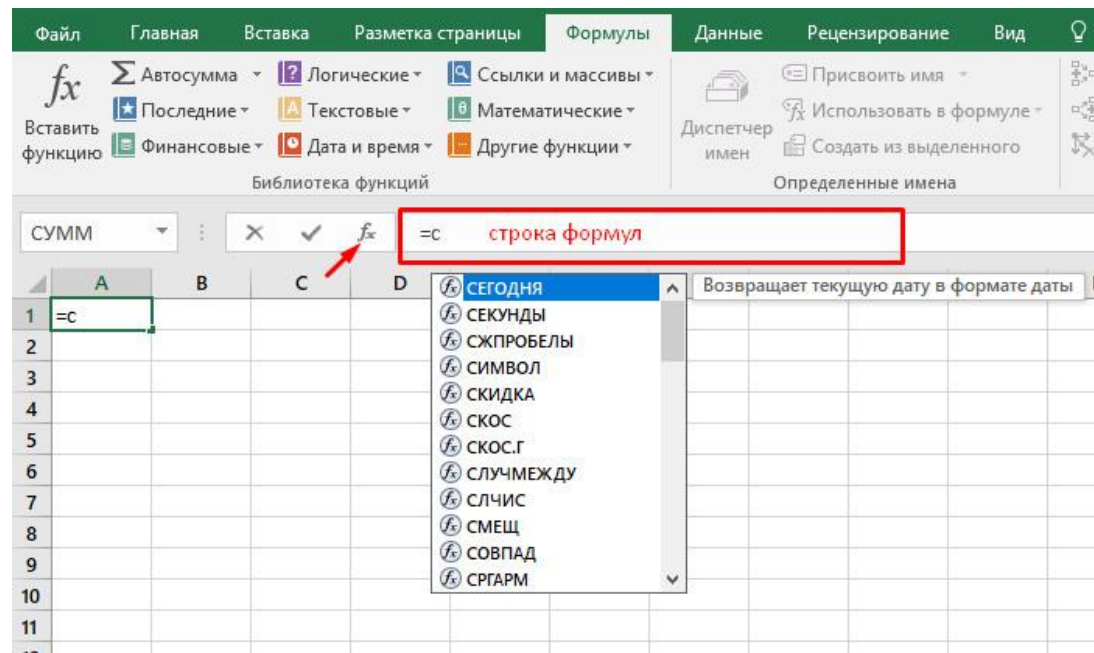
Ця функція допомагає визначити кількість числових значень у певному діапазоні.

•**COUNT:**

Формула: =COUNT(діапазон)

Наприклад:

=COUNT(A2:A10)



4. Висновки та рекомендації.

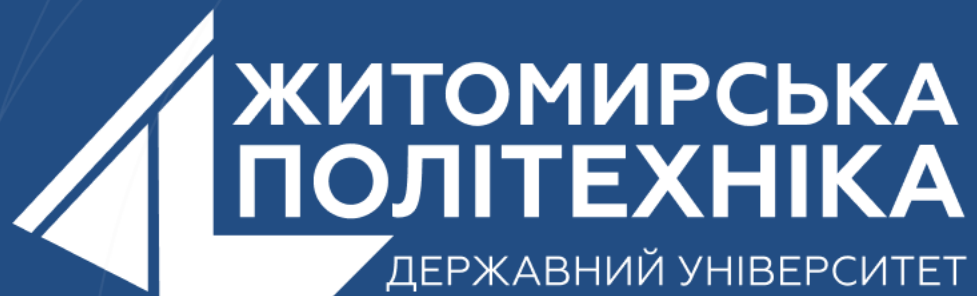
Висновки:

- 1.Точність та якість даних.** Основна частина обробки даних залежить від початкової точності збору даних. Використання якісних інструментів, таких як штангенциркулі, мікрометри, датчики та сенсори, дозволяє зібрати дані з мінімальними похибками. Це забезпечує достовірність подальших висновків і результатів аналізу.
- 2.Організація даних.** Обробка даних із використанням інструментів, таких як Excel або Google Таблиці, дозволяє ефективно організувати великий обсяг даних, очистити їх і провести базовий аналіз. Такі інструменти забезпечують швидкий доступ до функцій для розрахунку середнього значення, медіани, моди, стандартного відхилення та побудови трендів.
- 3.Кореляція та прогнозування.** Кореляційний аналіз є важливим інструментом для виявлення залежностей між змінними. Використання кореляції дозволяє визначити, чи впливають одні параметри на інші. Наприклад, при аналізі продажів і сезонних коливань можна виявити вплив температури на попит.
- 4.Візуалізація даних.** Графіки та діаграми дозволяють наочно виявити тренди та закономірності, що допомагає приймати більш обґрунтовані рішення. Це особливо корисно для бізнесу, де необхідно швидко виявити динаміку ринку, прогнозувати майбутні показники або проаналізувати ефективність кампаній.

Рекомендації:

- 1.Використання надійних інструментів для вимірювання.** Для зменшення похибок у даних рекомендується використовувати точні прилади для вимірювань, такі як електронні ваги, калібровані термометри або цифрові штангенциркулі.
- 2.Регулярне оновлення даних.** Для отримання актуальних результатів слід періодично оновлювати бази даних, переглядати старі вимірювання та коригувати їх за потреби.
- 3.Використання програмних засобів для автоматизації.** Рекомендується застосовувати спеціалізоване програмне забезпечення для обробки великих обсягів даних, автоматизації аналізу та побудови прогнозів.

   @ZTUEDUUA



- Розвиваємо лідерів
- Створюємо інновації
- Змінюємо світ на краще

