Data Analysis

Chapter 1
Introduction to Data Analysis

Dr. Mahmoud Elsabagh

Contents

Chapter 1: Data Analysis

1. Opening & Motivation

➜ 👉 *"How many of you check the weather before leaving home?"*

👉 *"How many of you compare prices online before buying something?"*

➜ *Each of these actions involves data analysis.*

Weather forecasts are based on historical and real-time weather data.

Price comparisons use stored data from e-commerce websites.

➜ Message: **Data analysis is not just for scientists – it's part of daily life.**

2. What is Data Analysis?

➡ Formal definition:

"Data Analysis is the process of systematically applying statistical, logical, and computational techniques to describe and evaluate data."

➡ Simplified:

It's about making sense of numbers, text, or observations so we can make better decisions.

➡ Example:

Raw data: "Scores of 30 students in a math test"

Analysis: "Average score = 72, Highest = 95, Lowest = 45"

Insight: "Class is strong overall, but some students need support."

3. Why is Data Analysis Important?

➜ **Decision Making:** Companies decide where to open new stores based on population and sales data.

➜ **Identifying Patterns:** Doctors detect diseases from medical records and lab results.

➜ **Predicting the Future:** Banks analyze transaction data to predict loan risks.

➜ **Detecting Problems:** Fraud detection in credit cards (unusual spending).

➜ **Optimizing Performance:** Athletes use performance data to improve training.

➜ 💡 *"In the 21st century, data is called the 'new oil.' Those who know how to use it can power the future."*

4. The Data Explosion

→ Every second:

  Google processes 100,000+ searches.

  Facebook users upload 150,000+ photos.

  Twitter posts 6,000+ tweets.

→ In 2025, it's estimated **175 zettabytes of data will exist** (1 zettabyte = 1 trillion GB).

→ Meaning: *We live in a data-driven world. Without analysis, this data is useless.*

→ (Show students a chart of "data growth over the years.")

5. The Data Analysis Process

Step-by-step breakdown:

➜ **Data Collection** – gather raw information. (e.g., survey results, sensors, sales logs)

➜ **Data Cleaning** – remove duplicates, fill missing values, fix errors.

➜ **Exploration (EDA)** – understand the structure, patterns, and trends.

➜ **Modeling** – use statistical or machine learning models to test hypotheses.

➜ **Interpretation** – turn results into meaningful insights.

➜ **Reporting** – communicate results through visuals and stories.

➜ 💡 Analogy: *Think of cooking a meal.*

➜ Collect ingredients → Clean & cut → Cook → Taste → Serve → Eat.

➜ Data analysis is exactly like this process!

6. Types of Data

→ **Qualitative (Non-numeric):**

Nominal (no order): e.g., Gender = Male/Female/Other.

Ordinal (ordered): e.g., Customer satisfaction = Poor, Good, Excellent.

→ **Quantitative (Numeric):**

Discrete: Countable numbers (e.g., Number of books).

Continuous: Measurements (e.g., Height, Weight, Time).

💡 Example for students:

→ If I ask for your **favorite color** → Nominal data.

→ If I rank movies as *bad, average, excellent* → Ordinal data.

→ If I count the number of classes you take → Discrete data.

→ If I measure your height in cm → Continuous data.

7. Scales of Measurement

➜ **Nominal:** Labels/categories (car brand, eye color).

➜ **Ordinal:** Order matters, but distances between categories are unknown (race rankings: 1st, 2nd, 3rd).

➜ **Interval:** Numbers with meaningful differences, but no true zero (temperature in °C).

➜ **Ratio:** Has a true zero, so ratios make sense (salary, distance, age).

💡 Example to explain:

➜ Temperature = 0°C doesn't mean "no temperature."

➜ But Salary = $0 means "no money."

8. Real-Life Applications of Data Analysis

→ **Business:** Amazon recommends products based on previous purchases.

→ **Healthcare:** AI models analyze X-rays to detect cancer.

→ **Sports:** Coaches use data to analyze performance.

→ **Education:** Universities use data to track student progress and dropouts.

→ **Government:** Data is used in national census and urban planning.

9. Tools for Data Analysis

➡ **Beginner tools:** Microsoft Excel, Google Sheets.

➡ **Programming languages:**

Python (Pandas, NumPy, Matplotlib, Seaborn).

R (ggplot2, dplyr).

➡ **Visualization tools:** Tableau, Power BI.

➡ **Big data tools:** Apache Spark, Hadoop.

10. Challenges in Data Analysis

→ Missing or incomplete data.

→ Dirty data (inaccuracies, duplicates).

→ Large datasets (Big Data).

→ Privacy and ethical concerns (e.g., misuse of personal data).

→ Human bias in interpreting results.

→ 💡 Example: If a dataset has only male participants, results may not generalize to women.

11. Case Study: Student Performance

➜ Dataset: Scores of 50 students in Math, Science, English.

➜ Analysis Steps:

Calculate averages and identify top-performing subjects.

Find correlations (e.g., students good at Math are also good at Science).

Visualize with bar charts and scatter plots.

➜ Insight: Helps teachers improve weak areas and support students.

Summary

➔ Data analysis = turning raw data into insights.

➔ It is used everywhere – business, healthcare, education, government.

➔ Different types of data exist: qualitative & quantitative.

➔ A structured process ensures reliability.

➔ Tools vary from simple (Excel) to advanced (Python, Power BI, AI).

Thanks!

Any questions?