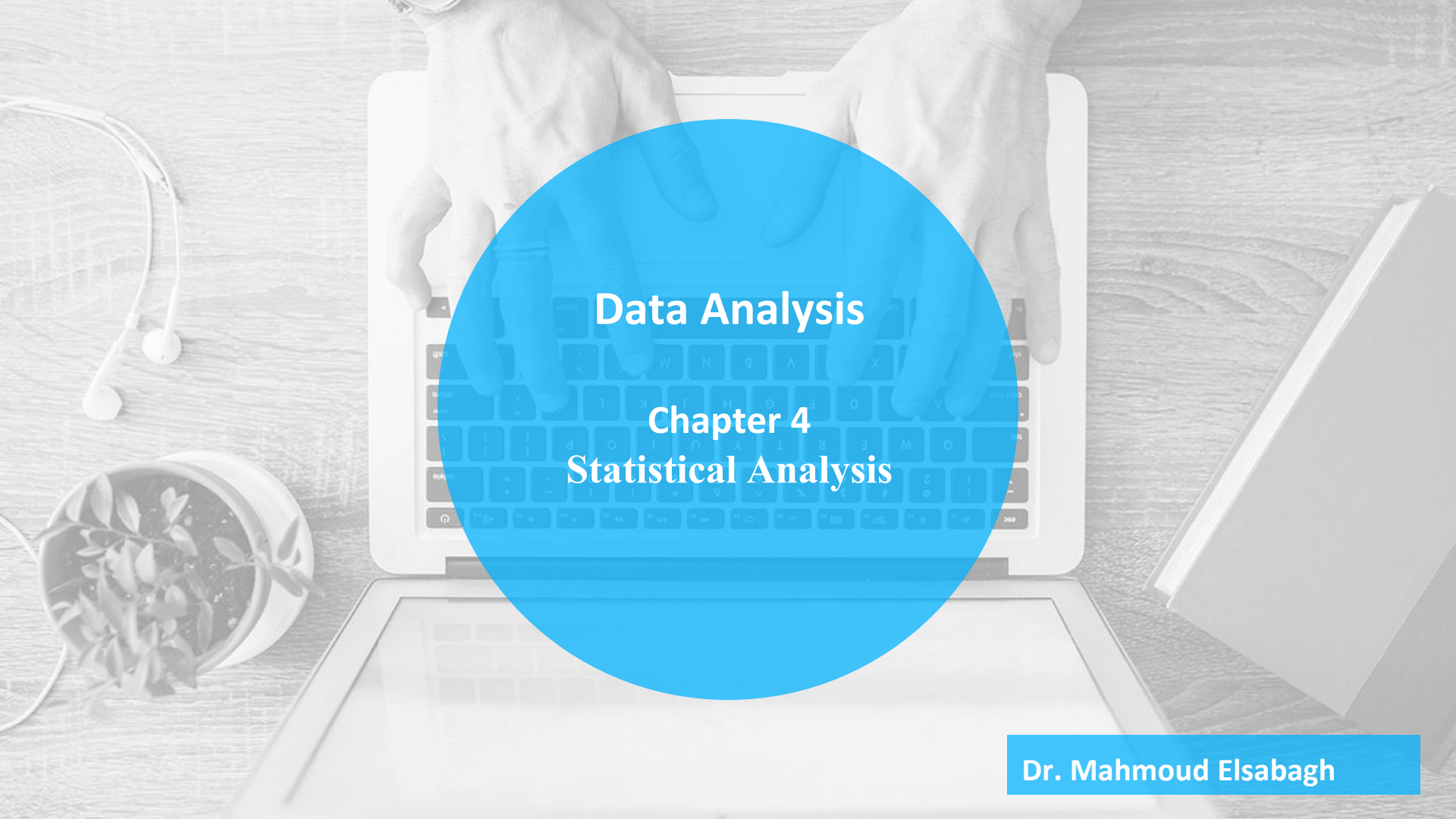
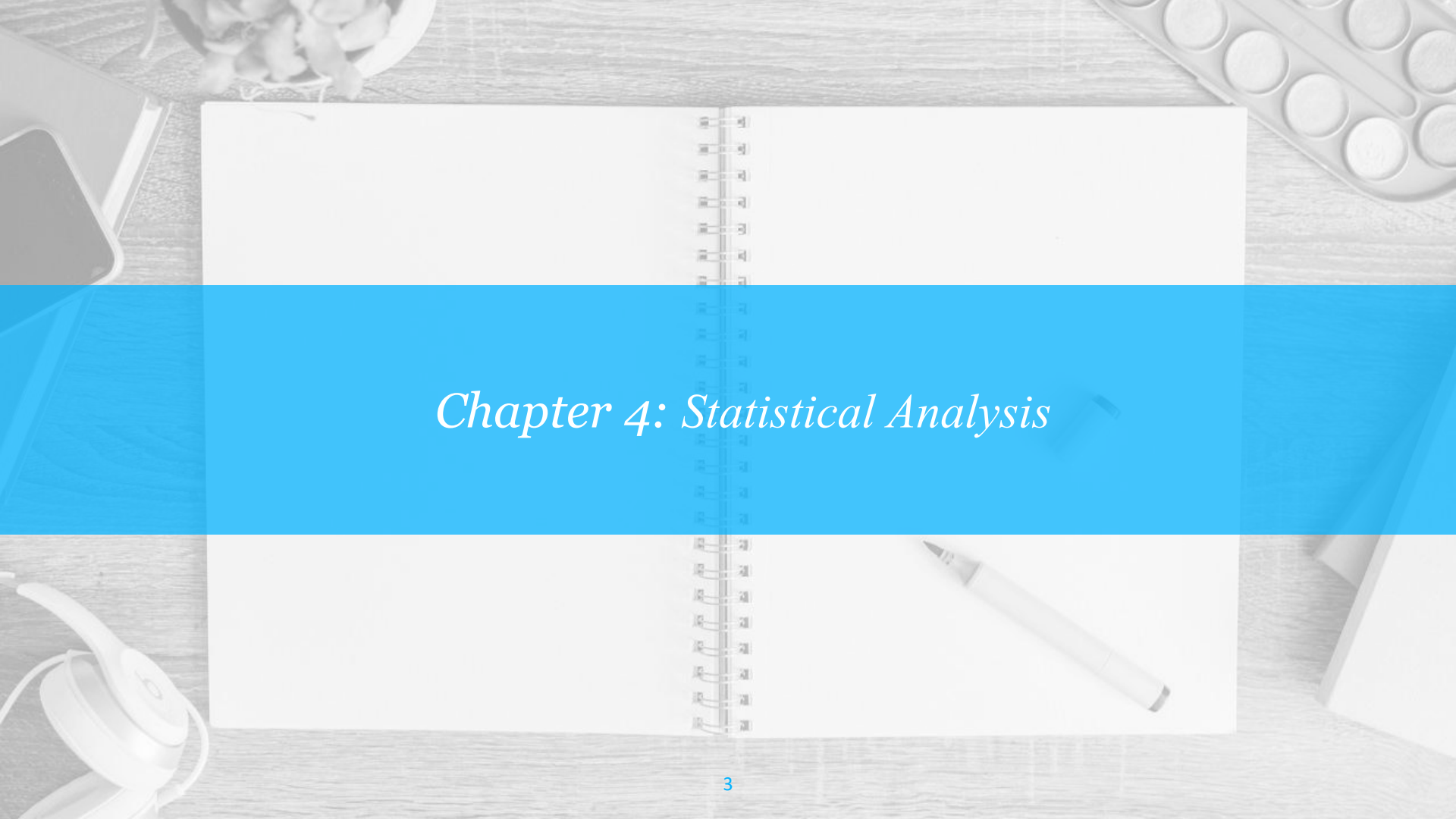


Data Analysis

Chapter 4
Statistical Analysis

Dr. Mahmoud Elsabagh

Contents



Chapter 4: Statistical Analysis

1. Introduction to Statistical Analysis

➜ Definition:

Statistical analysis is the process of collecting, exploring, organizing, and interpreting data using statistical methods to uncover patterns, test hypotheses, and support decision-making.

➜ Why It Matters in Data Analysis:

Helps summarize large datasets into meaningful insights.

Provides tools for comparing groups, predicting outcomes, and identifying relationships.

Essential for evidence-based decisions in business, healthcare, education, research, and AI.

➜ Analogy:

Think of statistics as the microscope of data — it allows us to see patterns and relationships that are invisible to the naked eye.

2. Types of Statistics

- A. Descriptive Statistics

➜ Goal: Summarize and describe the main features of a dataset.

➜ Key Measures:

Measures of Central Tendency: Mean, Median, Mode.

Measures of Dispersion: Range, Variance, Standard Deviation, Interquartile Range (IQR).

Data Distribution: Histograms, frequency tables, percentiles.

➜ Example: Average exam score of students in a class.

2. Types of Statistics

- B. Inferential Statistics

➜ Goal: Make predictions or generalizations about a population based on a sample.

➜ Key Concepts:

Hypothesis testing.

Confidence intervals.

Regression analysis.

ANOVA (Analysis of Variance).

➜ Example: Using a survey of 500 people to estimate the political preferences of a country.

👉 Tip: Use a diagram showing "Population → Sample → Inferential Statistics → Generalization."

3. Levels of Measurement in Data

- ➜ Nominal: Categories without order (e.g., gender, blood type).
- ➜ Ordinal: Categories with order but unequal intervals (e.g., satisfaction ratings).
- ➜ Interval: Ordered, equal intervals, no true zero (e.g., temperature in Celsius).
- ➜ Ratio: Ordered, equal intervals, true zero (e.g., weight, income).

⚠ Important: The level of measurement determines which statistical techniques can be applied.

4. Descriptive Statistics in Detail

A. Central Tendency

- ➜ Mean (average) = $\Sigma x / n$.
- ➜ Median = middle value when data is sorted.
- ➜ Mode = most frequent value.
- ➜ Example: Test scores: 60, 70, 70, 80, 90.

  Mean = 74.

  Median = 70.

  Mode = 70.

4. Descriptive Statistics in Detail

B. Dispersion

- ➜ Range = Max – Min.
- ➜ Variance = average squared deviation from the mean.
- ➜ Standard Deviation (SD) = square root of variance → shows spread.
- ➜ Coefficient of Variation (CV) = SD / Mean × 100.
- ➜ Example: If average income is $2000 with SD = $200 → most incomes lie between $1800–$2200.

5. Inferential Statistics in Detail

- A. Hypothesis Testing
- Steps:

  State null hypothesis (H_0) and alternative hypothesis (H_1).

  Select significance level (α, usually 0.05).

  Collect sample data.

  Calculate test statistic (t-test, z-test, chi-square, etc.).

  Compare with critical value → Accept or reject H_0.

- Example: Testing whether a new teaching method improves student performance compared to the old method.

5. Inferential Statistics in Detail

B. Confidence Intervals (CI)

➜ Definition: Range of values likely to contain the true population parameter with a given probability (e.g., 95% CI).

➜ Example: If average student GPA = 3.2, 95% CI = (3.0, 3.4) → We are 95% confident the true average GPA lies between 3.0 and 3.4.

5. Inferential Statistics in Detail

C. Regression Analysis

➜ Definition: Examines relationship between independent variable(s) and dependent variable.

➜ Types:

Simple Linear Regression ($Y = a + bX$).

Multiple Regression ($Y = a + b_1X_1 + b_2X_2 + \ldots$).

➜ Example: Predicting sales based on advertising spending and product price.

5. Inferential Statistics in Detail

D. Correlation

➜ Definition: Measures strength and direction of relationship between two variables.

➜ Correlation Coefficient (r): Ranges from –1 to +1.

$r > 0$: Positive correlation.

$r < 0$: Negative correlation.

$r = 0$: No correlation.

➜ Example: Hours studied vs. exam scores (likely positive correlation).

5. Inferential Statistics in Detail

E. ANOVA (Analysis of Variance)

➜ Definition: Compares means across three or more groups.

➜ Example: Testing if students' average grades differ across three different teaching methods.

6. Visualizing Statistical Data

➜ Charts for Descriptive Stats:

Histograms (distribution).

Pie charts (categories).

Boxplots (spread and outliers).

Scatterplots (relationships).

👉 You should learn when to use each visualization.

7. Common Mistakes in Statistical Analysis

- ➜ Misinterpreting correlation as causation.
- ➜ Using the wrong test for the data type.
- ➜ Ignoring outliers.
- ➜ Overgeneralizing from small samples.
- ➜ P-hacking (running many tests until one shows significance).

8. Practical Applications

- ➜ Healthcare: Analyzing clinical trial results to test drug effectiveness.
- ➜ Business: A/B testing for digital marketing campaigns.
- ➜ Education: Comparing student performance across teaching methods.
- ➜ AI/ML: Statistical foundations of algorithms (e.g., probability, regression).

Summary

- Two main branches: Descriptive (summarize data) and Inferential (draw conclusions).
- Core tools: Mean, Median, Mode, SD, Variance, CI, Hypothesis Testing, Correlation, Regression, ANOVA.
- Statistical analysis turns raw data into knowledge.
- Always ensure proper method selection, avoid bias, and interpret results carefully.

Thanks!

Any questions?