Handout: Summary Statistics and Techniques in R

baharalsadat bahadori - 400897377

1. Introduction

Summary statistics help us understand a dataset quickly.
They show typical values, variation, and possible errors.
R provides simple tools to calculate these statistics in an efficient way.

---

2. Why Summary Statistics Matter

- Understand central values
- See how spread out the data is
- Detect outliers and missing values
- Identify wrong or inconsistent entries
- Understand distribution shape
- Prepare data for further analysis

---

3. Types of Summary Statistics

A. Measures of Central Tendency

- Mean
- Median
- Mode

B. Measures of Spread

- Range
- Variance
- Standard deviation

C. Distribution Shape

- Skewness
- Kurtosis

---

4. Techniques Used in R

A. Base R

- Built-in functions that calculate common summary values
- Does not require additional packages
- Suitable for small or simple datasets

B. dplyr (Tidyverse)

- Cleaner and easier to read
- Supports summarising multiple variables
- Handles missing values (`na.rm = TRUE`)
- Very useful for larger datasets

C. Grouped Summaries

- Used to compare categories (e.g., groups, genders, regions)
- Helps understand differences between categories

---

5. Student Exercise

Dataset: **45, 60, 75, 80, 90**

Tasks: - Find the mean
- Find the median
- Find the standard deviation
- Identify the minimum and maximum
- Identify the quartiles
- Write one sentence describing how spread out the scores are

Questions: - What does the mean tell you?
- Are the values close together or spread apart?
- What do the minimum and maximum show?

---

6. Solution (Summary)

- Mean = 70
- Median = 75

- Standard deviation 17
- Minimum = 45
- Maximum = 90
- Interpretation:
  The data has moderate spread and the scores gradually increase.

---

7. Conclusion

Summary statistics are important because they help us understand data clearly and quickly.
R provides simple methods using Base R and the dplyr package.
These techniques form the foundation for reliable and accurate data analysis.

---

8. References

Wickham, H., & Grolemund, G. (2017). *R for Data Science*.
Kabacoff, R. (2022). *R in Action* (3rd ed.).
R Core Team. (2024). *R: A Language and Environment for Statistical Computing*.
