Seminar 2

Calculating the Mode and the Median

Mode

The **mode** is the value or values that appear most frequently in a dataset. It is the most common value(s) in the data.

Steps to calculate the mode:

1. **Sort the data** (optional): Sorting the data helps in visualizing the frequency of each value.
2. **Count the frequency** of each unique value in the dataset.
3. **Identify the value(s) that occur the most frequently**: The mode is the value(s) with the highest frequency.

Example:
Consider the following data representing the number of books read by students:

$$3, 4, 4, 6, 7, 4, 8, 6, 3, 9$$

- **Step 1:** Sort the data (optional): $3, 3, 4, 4, 4, 6, 6, 7, 8, 9$
- **Step 2:** Count the frequency of each value:
  - 3 appears twice
  - 4 appears three times
  - 6 appears twice
  - $7, 8, 9$ each appear once
- **Step 3:** The value 4 appears the most (three times), so the **mode** is 4.

Special Cases:

- If all values occur with the same frequency or only once, the dataset has **no mode**.
- If more than one value occurs with the highest frequency, the dataset is **multimodal**.

Median

The **median** is the middle value of a dataset when it is ordered from smallest to largest. It separates the dataset into two equal halves.

Steps to calculate the median:

1. **Order the data** from smallest to largest.
2. If the number of observations n is **odd**, the median is the value at position $\frac{n+1}{2}$ in the ordered data.
3. If the number of observations n is **even**, the median is the average of the two middle values at positions $\frac{n}{2}$ and $\frac{n}{2} + 1$.

Example 1 (Odd number of observations):
Data: $5, 3, 9, 6, 4$

- **Step 1:** Order the data: $3, 4, 5, 6, 9$
- **Step 2:** The number of observations $n = 5$, which is odd. The median is at position $\frac{5+1}{2} = 3$.
- **Step 3:** The median is the 3rd value, which is 5.

Example 2 (Even number of observations):
Data: $2, 7, 4, 9, 3, 6$

- **Step 1:** Order the data: $2, 3, 4, 6, 7, 9$
- **Step 2:** The number of observations $n = 6$, which is even. The median is the average of the 3rd and 4th values: 4 and 6.
- **Step 3:** The median is:

$$\text{Median} = \frac{4+6}{2} = 5$$

Special Cases:
If there are repeated values, the steps for finding the median remain the same. You just use the positions in the ordered data, regardless of whether some values are repeated.

Summary

- **Mode**: The most frequent value(s) in the dataset.
- **Median**: The middle value in an ordered dataset. For an odd number of observations, it is the middle value, and for an even number, it is the average of the two middle values.

Problem 1: Basic Descriptive Statistics

You are given the following dataset representing the ages of participants in a study:

$$18, 21, 24, 20, 19, 23, 25, 18, 22, 20, 24, 26, 19, 21, 22$$

1. Compute the **mean**, **median**, and **mode** of the dataset.
2. Find the **range**, **variance**, and **standard deviation**.
3. Interpret what these descriptive statistics tell you about the dataset.

Problem 2: Percentiles and Quartiles

Consider the following test scores of 20 students:

$$45, 67, 78, 88, 54, 72, 61, 80, 92, 53, 77, 84, 69, 65, 70, 89, 90, 74, 81, 73$$

1. Find the **25th percentile (Q1)**, **50th percentile (median)**, and **75th percentile (Q3)**.
2. Calculate the **interquartile range (IQR)** and explain its significance.
3. Identify any **outliers** using the $1.5 \times \text{IQR}$ rule.

Problem 3: Grouped Data

The table below shows the frequency distribution of the number of books read by a group of students over the summer:

1. Estimate the **mean** number of books read using the midpoint method.
2. Find the **mode** and **median** for the grouped data.
3. Compute the **variance** and **standard deviation**.

Books Read	Frequency
0 - 2	5
3 - 5	8
6 - 8	12
9 - 11	6
12 - 14	3

Table 1: Books Read by Students

Problem 4: Box Plot and Interpretation

The following data represents the weekly hours spent studying by a group of students:

$$4, 6, 8, 3, 5, 7, 9, 10, 2, 6, 8, 5, 6, 11, 7$$

1. Construct a **box plot** for the data.
2. Identify the **median**, **quartiles**, and any potential **outliers**.
3. Explain what the box plot reveals about the distribution of the data.

Problem 5: Visualization and Summary Measures

The following dataset represents the heights (in cm) of students in a class:

$$162, 167, 170, 155, 165, 160, 168, 172, 163, 159, 166, 169, 173, 161, 164$$

1. Create a **histogram** to represent the height distribution.
2. Construct a **stem-and-leaf plot**.
3. Calculate the **mean**, **median**, and **standard deviation** of the dataset.
4. Describe the **shape** of the distribution.

Problem 6: Comparison of Two Data Sets

Two different groups of students take two versions of a test. Their scores are as follows:
Group A: 85, 78, 92, 88, 75, 90, 95, 80, 85, 87
Group B: 72, 65, 80, 75, 82, 68, 78, 83, 70, 76

1. Compute the **mean**, **median**, **range**, **variance**, and **standard deviation** for both groups.
2. Which group has more variability? Justify your answer.
3. Based on the descriptive statistics, which group appears to have performed better?

Problem 7: Real-World Data Analysis

The monthly income (in dollars) of employees at a small company is given below:

$$3000, 3500, 4000, 3800, 4200, 3900, 4500, 4300, 5000, 5500, 6000, 7000$$

1. Calculate the **mean**, **median**, and **mode**.
2. Determine the **range**, **variance**, and **standard deviation**.
3. If the CEO earns $20,000 per month, how does this outlier affect the mean and median?

Discussion Questions

1. How does the **shape** of a dataset (e.g., skewness) affect the relationship between the mean and median?
2. Why might the **median** be a better measure of central tendency than the **mean** in some cases?
3. What are the limitations of **range** as a measure of variability, and how does the **standard deviation** improve on this?