PCED

Module 3 -Statistical Analysis

Topics

1. Descriptive Statistics
2. Inferential Statistics
3. Key Concepts to Review
4. Preparation Tips

Descriptive Statistics

Measures of Central Tendency

Mean: The average of a dataset.
Median: The middle value when the data is ordered.
Mode: The most frequently occurring value(s).

Measures of Variability

Range: Difference between the maximum and minimum values.
Variance: The average of squared differences from the mean.
Standard Deviation (SD): The square root of the variance, indicating spread around the mean.

Inferential Statistics

Simple Regression Analysis

Understand the relationship between two variables:
- Independent variable (predictor).
- Dependent variable (response).

Linear regression equation: $y=mx+c$, where:
- y: Dependent variable.
- x: Independent variable.
- m: Slope (rate of change).
- c: Intercept (value of y when x=0).

Basics of correlation:

Correlation coefficient (r): Measures the strength and direction of the linear relationship between two variables (range: -1 to +1).

Key Concepts to Review

Population vs. Sample:

Population: Entire group being studied.
Sample: Subset of the population used for analysis.

Hypothesis Testing:

Null hypothesis (H0): Assumes no effect or relationship.
Alternative hypothesis (Ha): Assumes an effect or relationship exists.
P-value: Probability of observing the data given that H0 is true (commonly compared to a significance level, e.g., 0.05).

Confidence Intervals:

Range within which the true population parameter is likely to fall.

Preparation Tips

Practice calculating mean, median, mode, variance, and standard deviation from sample datasets.

Use example problems to apply simple regression concepts, such as predicting y given x.

Review scatterplots to visually interpret relationships between variables.

Understand when and how to use t-tests or z-tests (even if they’re not explicitly required, familiarity helps).

Examples

1. Find the mean, median, mode, and standard deviation for the dataset:
   4,8,6,5,3,7,9,124,8,6,5,3,7,9,12

2. For the dataset above, calculate the range and interpret what it tells you about variability.

More on : - https://www.khanacademy.org/math/statistics-probability

Descriptive Statistics Problem

Dataset:

$$4, 8, 6, 5, 3, 7, 9, 12$$

1. Mean

The formula for the mean is:

$$\text{Mean} = \frac{\text{Sum of all data points}}{\text{Number of data points}}$$

$$\text{Mean} = \frac{4 + 8 + 6 + 5 + 3 + 7 + 9 + 12}{8} = \frac{54}{8} = 6.75$$

2. Median

The median is the middle value of an ordered dataset.

- Arrange the data:

$$3, 4, 5, 6, 7, 8, 9, 12$$

- Since the dataset has an even number of values ($n = 8$), take the average of the two middle numbers:

$$\text{Median} = \frac{6 + 7}{2} = 6.5$$

2

3. Mode

The mode is the most frequently occurring value.

- Each number appears once, so there is **no mode**.

4. Range

The range is the difference between the maximum and minimum values:

$$\text{Range} = \text{Maximum} - \text{Minimum} = 12 - 3 = 9$$

5. Variance and Standard Deviation

Variance formula:

$$\text{Variance} = \frac{\sum (x_i - \bar{x})^2}{n}$$

Standard Deviation:

$$\text{SD} = \sqrt{\text{Variance}}$$

- Calculate deviations:

$$x_i - \bar{x} = \{-2.75, 1.25, -0.75, -1.75, -3.75, 0.25, 2.25, 5.25\}$$

- Square deviations:

$$\{7.5625, 1.5625, 0.5625, 3.0625, 14.0625, 0.0625, 5.0625, 27.5625\}$$

- Sum of squared deviations:

$$59.1875$$

- Variance:

$$\text{Variance} = \frac{59.1875}{8} = 7.3984$$

- Standard Deviation:

$$\text{SD} = \sqrt{7.3984} \approx 2.72$$