Univariate Analysis

Introduction to Univariate Analysis

- Analysis of a single variable
- Goal:
  - Describing the characteristics of the variable
  - Comparisons across different variables
  - Summarize and find patterns in the data

- Applications:
  - Statistics
  - Data Science
  - Research

Measures of Central Tendency

- Mean: The arithmetic average of all data points.
  - **Calculation**: Sum of all data points divided by the number of points.
  - Provides a quick snapshot of the average value.

- Median: The middle value in a sorted list of numbers.
  - **Calculation**: The middle value in an ordered data set.
  - Useful when the data set has outliers, as it is not affected by extreme values.

- Mode: The value that appears most frequently in the data set.
  - **Calculation**: The most frequently occurring value in the data set.
  - Helps identify the most common value in the data set.

Measures of Dispersion

- Variance: The average of the squared differences from the mean.
  - Indicates how much the values in the data set vary.

- Standard Deviation: The square root of the variance, indicating how spread out the values are.
  - Is the average distance from the mean.

  → Easier to interpret than the variance.

- Quantiles: Values that divide the data into equal-sized intervals.

- Range: The difference between the highest and lowest values.