loresowhat

11 Basic Statistical Concepts for Data Analysts

swipe

loresowhat

Sampling Techniques

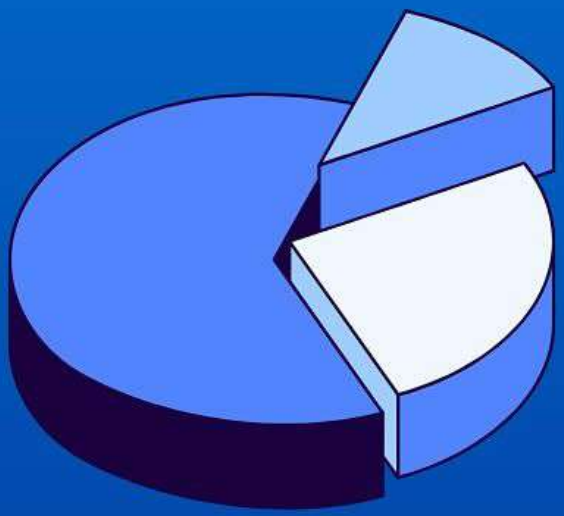
- **Simple Random Sampling:** Every member of the population has an equal chance of being chosen.
- **Stratified Sampling:** Dividing the population into distinct subgroups (strata) and sampling from each one.
- **Cluster Sampling:** Splitting the population into clusters, randomly selecting clusters, and sampling all members within these selected clusters.
- **Systematic Sampling:** Selecting every nth member from the population.

swipe

loresowhat

Types of Variables

- **Nominal:** Categories without a specific order (e.g., gender, colour).
- **Discrete:** Countable values (e.g., number of students).
- **Continuous:** Values within a range (e.g., height, weight).

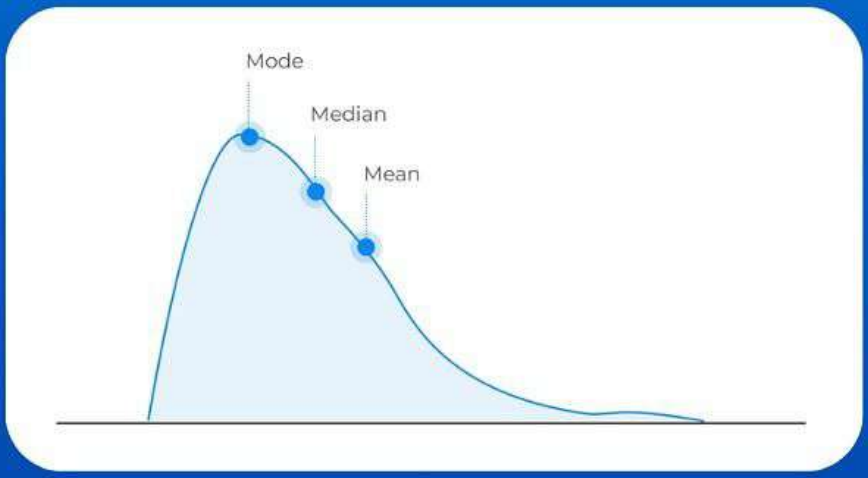


swipe

loresowhat

Measures of Central Tendency

- **Mean:** The average of all data points.
- **Median:** The middle value when data points are ordered.
- **Mode:** The most frequently occurring value.



swipe

loresowhat

Measures of Dispersion

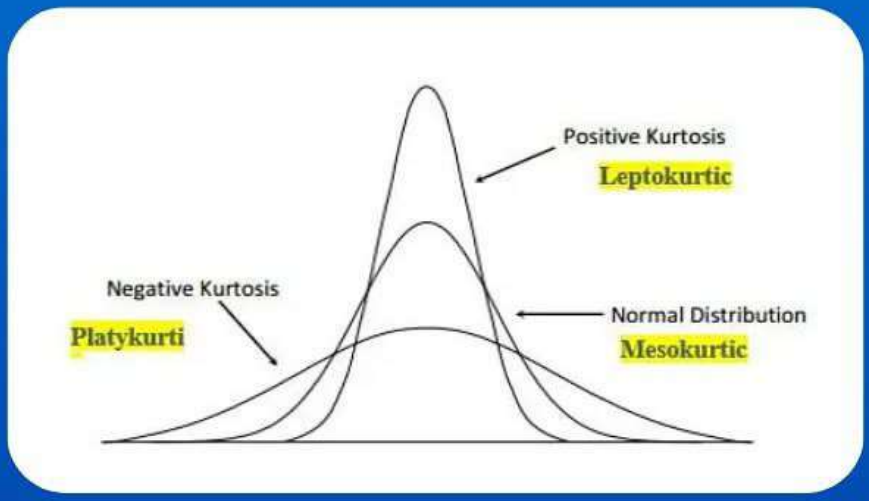
- **Range:** The difference between the highest and lowest values.
- **Interquartile Range (IQR):** The range of the middle 50% of the data.
- **Variance and Standard Deviation:** Measures of how much data points deviate from the mean.
- **Z-Score:** Indicates how many standard deviations a data point is from the mean.
- **Correlation:** Shows the relationship between two variables and how they change together.

swipe →

loresowhat

Skewness and Kurtosis

- **Skewness:** Measures the asymmetry of the distribution.
- **Kurtosis:** Measures the “peakedness” or “tailedness” of the distribution.



swipe

loresowhat

Other Concepts

Population and Sample: Understanding the difference between the entire group of interest (population) and the specific subset you actually analyse (sample).

Parameters and Sample Statistics: Distinguishing between measures that describe the whole population (parameters) and those that describe only the sample (sample statistics).

Estimators: Learning about methods or statistics that estimate population parameters based on sample data.

Normal Distribution: A symmetric, bell-shaped curve where most data points are concentrated around the mean.

Confidence Intervals: A range of values from the sample likely to contain the true population parameter.

swipe →