Data Types, Measures of Central Tendency, and Measures of Dispersion

# Data Types

## Overview

Data types are classifications that determine the possible values for a variable, the operations that can be performed on it, and the way it can be stored. Understanding data types is essential for data analysis, as it helps in choosing the right statistical methods and tools.

## Types of Data

1. Qualitative (Categorical) Data:  
- Nominal: Data that represents categories with no intrinsic ordering (e.g., gender, blood type).  
- Ordinal: Data that represents categories with a meaningful order but no fixed interval between them (e.g., rankings, satisfaction levels).  
  
2. Quantitative (Numerical) Data:  
- Discrete: Data that can only take specific values, often counts (e.g., number of children, number of cars).  
- Continuous: Data that can take any value within a range (e.g., height, weight, temperature).

# Measures of Central Tendency

## Overview

Measures of central tendency are statistical metrics that describe the center point of a dataset. They provide a single value that is representative of the entire dataset.

## Types

1. Mean (Arithmetic Average):  
- Calculated by summing all values and dividing by the number of values.  
- Sensitive to extreme values (outliers).  
  
2. Median:  
- The middle value when the data is ordered.  
- Not affected by outliers, making it a better measure for skewed distributions.  
  
3. Mode:  
- The most frequently occurring value in a dataset.  
- Useful for categorical data and identifying the most common category.

## Properties

Symmetric Distribution: Mean = Median = Mode  
Skewed Distribution:  
- Right Skewed: Mean > Median > Mode  
- Left Skewed: Mean < Median < Mode

# Measures of Dispersion

## Overview

Measures of dispersion describe the spread or variability of a dataset. They indicate how much the data values deviate from the central tendency.

## Types

1. Range:  
- Difference between the maximum and minimum values.  
- Simple but sensitive to outliers.  
  
2. Interquartile Range (IQR):  
- Difference between the first quartile (Q1) and third quartile (Q3).  
- Measures the spread of the middle 50% of the data.  
  
3. Variance:  
- Average of the squared differences from the mean.  
- Higher variance indicates more spread out data.  
  
4. Standard Deviation:  
- Square root of the variance.  
- Provides a measure of dispersion in the same units as the original data.  
  
5. Coefficient of Variation (CV):  
- Standard deviation divided by the mean, expressed as a percentage.  
- Useful for comparing variability between datasets with different units or means.

## Properties

Low Dispersion: Data points are close to the mean.  
High Dispersion: Data points are spread out over a wider range of values.