

# Statistics

25 May 2024 08:22

Data: It is information and facts.

Sample Vs Population

## Statistics

Collating, Organizing and analysing Data.

### Types of Statistic:

**Descriptive statistics** : Organizing and summarizing the data. (Mean, Median and Mode)

**Inferential statistics** : To make Predictions on Sample data from population data (T-test, Z-test etc)

### Types of Data:

1. Numerical data
  - a. Discrete (Whole Number)
  - b. Continuous Data (Floating)
2. Categorical Data
  - a. Nominal : No order
  - b. Ordinal: Order / Rank

### Scales of Measurement:

1. Nominal scales: No order
2. Ordinal scales: Order but no difference
3. Interval scales: Order and Difference but Ratio cannot justify. Doesn't start with zero.
4. Ratio scales: Order, Difference and Ratio justifies. Starts with zero.

Ex:

- a. Length of the River - 4
- b. IQ - 100, 120, 200 - 3
- c. Marital status: Married, Unmarried - 1

### Measure of central tendency:

**Mean: Not good with Outliers**

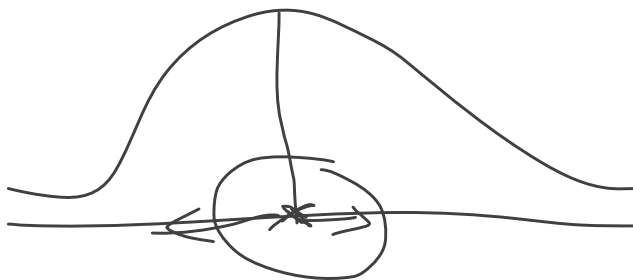
**Median: Better with Outliers**

**Mode: Missing Values**

1,2,3,4,5 - Mean : 3

**Outlier: Noisy data**

1,2,3,4,5,100 Mean:

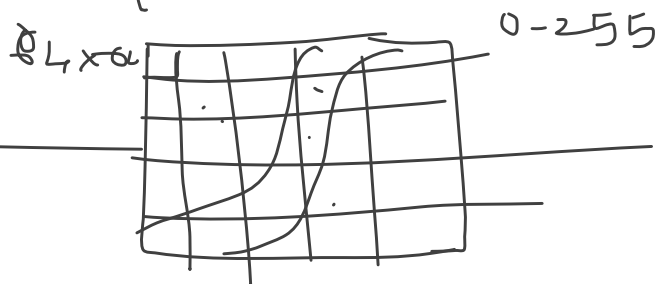


**Random Variable:** Mapping of the output on the basis of random experiment.

X=1 H  
0 T

$$X = \begin{cases} 1 & \text{if H} \\ 0 & \text{if T} \end{cases}$$

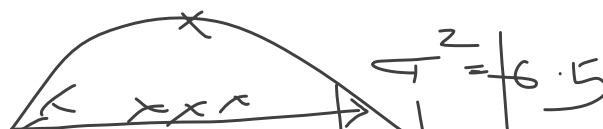
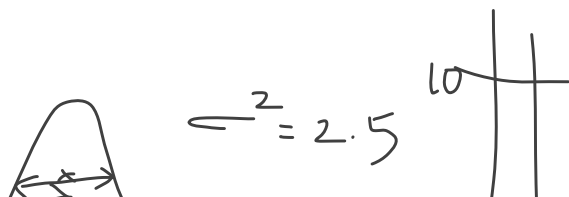
$$\begin{cases} X + Y = 2 \\ 2X - Y = 10 \end{cases}$$

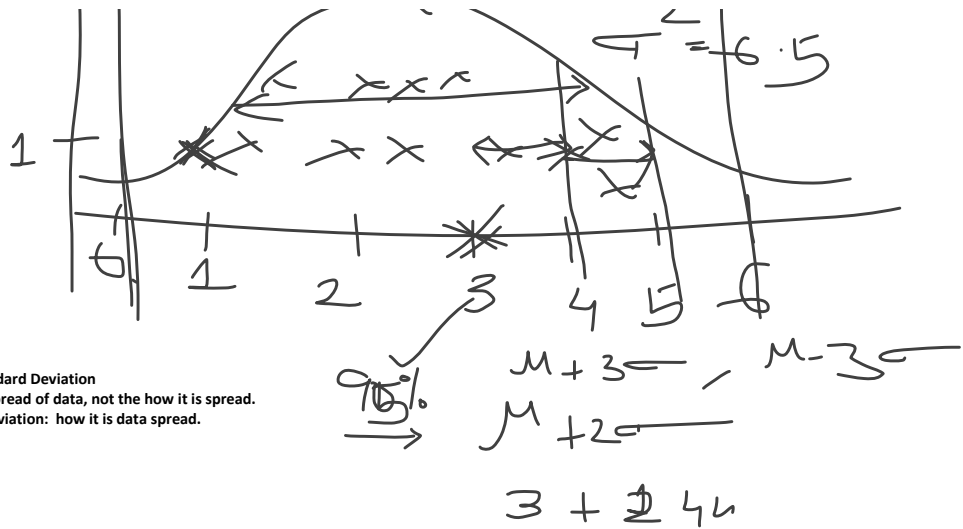
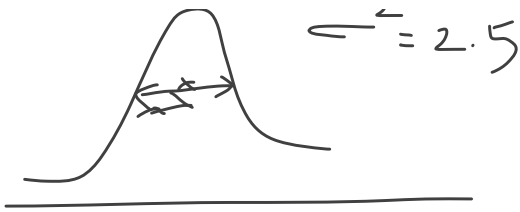


Set:

Union  
Intersection  
Subset  
Super Set  
Difference

Population  $\rightarrow \sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} \rightarrow \text{Mean}$   
 $\rightarrow \text{No. of Data!}$





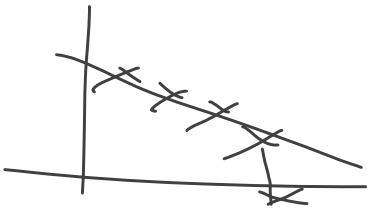
Bessel Correction:  
Population  
Sample

Variance and Standard Deviation  
Variance : Spread of data, not the how it is spread.  
Standard Deviation: how it is data spread.

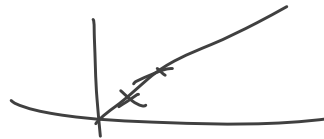
Covariance and Correlation:

Covariance : Relation between two variable X and Y

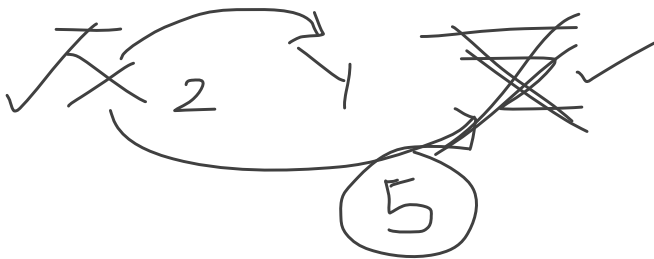
$$\text{Cov} = \sum_{i=1}^N \frac{(\underline{x}_i - \bar{x})(\underline{y}_i - \bar{y})}{n - 1} \quad \underline{\underline{+ve}}$$



-ve



+ve



Correlation:

$$\rho_{x,y} = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y} \quad \left( \begin{matrix} \leftarrow -1, 0, 1 \rightarrow \end{matrix} \right)$$