

# 1. Statistics :

⇒ What is statistics ?

- Statistics is the study of Collecting, Organising and analysing data. "For better decision Making".

⇒ Data : "Facts or piece of information"  
e.g: Height of Students {178cm, 280cm, 185cm}

## Statistics

↓  
• Descriptive Stats

↓  
• Inferential Stats

- 1) Descriptive Stats : It consists of organising and summarizing of data, with the help of :

pdf, Histogram, Box plot, Bar chart etc.

- 2) Inferential stats : It consists of using that you've measured to form conclusion with the help of hypothesis testing, P-Value, Z-test, t-test, Anova, Chi-square.



⇒ Data : 1) Population data  
2) Sample data

1) Population data refers to the entire group of individuals, objects, or events that share a common characteristics and are the focus of statistical study.

- "Entire group of interest"
- "Usually large (or infinite)"
- "Symbol  $\mu(\text{mu})$ "

2) Sample data :

- "Subset of the population"
- "Smaller and manageable"
- "Heights of 500 randomly chosen students"
- "Symbol  $\bar{x}$ "

⇒ Sampling techniques in stats : It is a methods used to select a subset (sample) from larger group (population)

1. Simple Random Sampling (SRS) :

- every individual has equal chance of being chosen.
- Selecting 100 students randomly from a university.



## 2. Stratified sampling :

- The population is divided into homogenous groups based on a characteristic (e.g. gender, age, income)
- Example : If a school has 60% boys and 40% girls the sample will maintain this ratio.

## 3. Systematic Sampling :

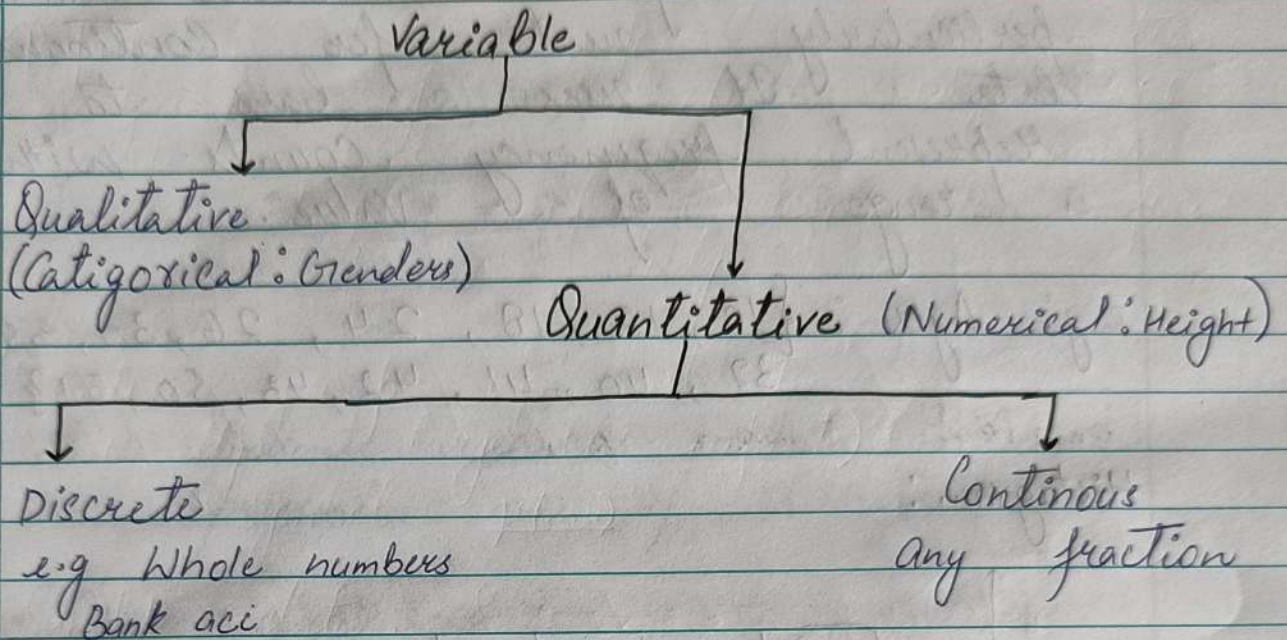
- A starting point is selected randomly then every  $K$ -th individual is chosen.
- Formula : 
$$K = \frac{\text{population size}}{\text{sample size}}$$
- Example : Selecting every 10th person from a list of 1000 students

## 4. Convenience sampling :

- Samples are taken from a group that is easily accessible.
- Example : Surveying student at nearby mall.



⇒ **Variable** : Property that store / take any value.

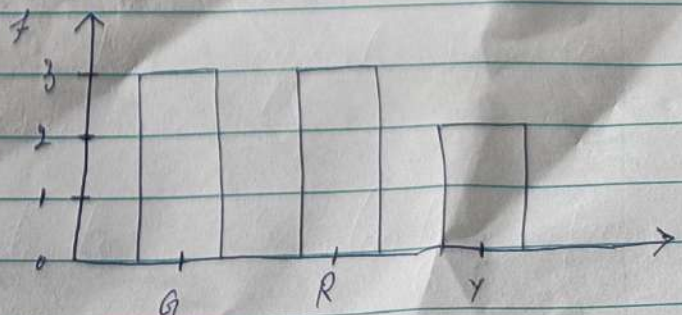


⇒ **Frequency distribution** : The amount of times element is in data

example : Sample data = Green, Red, Yellow, Green, Red, Yellow, Green, Red.

Colors	frequency
Green	3
Red	3
Yellow	2

Bar graph :





⇒ Measure of Central tendency : "Summarise a data set"

1) Mean : Arithmetic Average

$$x = \{1, 1, 2, 3, 3, 3, 4, 5, 5, 6\}$$

Formula :  $\mu = \frac{\sum_{i=1}^N x_i}{N}$  (Population mean)

Formula :  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$  (Sample mean)

$x_i$  represents each data point  
 $N$  represents no. of observations.

- It is sensitive to outliers.

2) Median : The middle element of data set

- It is resistant to outliers.

3) Mode : Most frequent value

⇒ Measure of dispersion : How spread out the data points are in data set. they help understand the variability or consistency of data. The main measure of dispersion are :

1) Variance ( $\sigma^2$  or  $s^2$ ) : Measures the average squared deviation from the mean.



- For population variance:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

- For sample variance:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

- 2) Standard deviation ( $\sigma$  or  $s$ ): The square root of variance, representing dispersion in the same units as the data.

- Formula:

$$\sigma = \sqrt{\sigma^2} \quad , \quad s = \sqrt{s^2}$$

- A higher standard deviation means more variability in the data.

- 3) Quartile range (Inter QR): The range of the middle 50% of data.

$$IQR = Q_3 - Q_1$$

- $Q_3$  (third quartile) and  $Q_1$  (first quartile) are 75th and 25th percentiles,
- Less sensitive to outliers.



⇒ Percentile : It is a value below which a certain percentage of observation lie.

• 99% percentile means the person has got better marks than 99% of student.

• Data Set :  $\{2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12\}$

Q# What is the percentile ranking of 10?

Percentile rank of  $X = \frac{\text{no. of elements below } X}{n} \times 100$

$$10 = \frac{16}{20} \times 100 = 80\% \text{ ile}$$

Q# What value exist at percentile ranking of 25% ile?

$$\text{value} = \frac{\text{percentile}}{100} \times (n+1) / (n-1) + 1$$

$$= \frac{25}{100} \times 21 = 5.25 \rightarrow \text{Index}$$

$$\text{Average} = \frac{5+5}{2}, 5$$

∴ value = 5