

# *TYPES OF STATISTICS*

# Descriptive Statistics

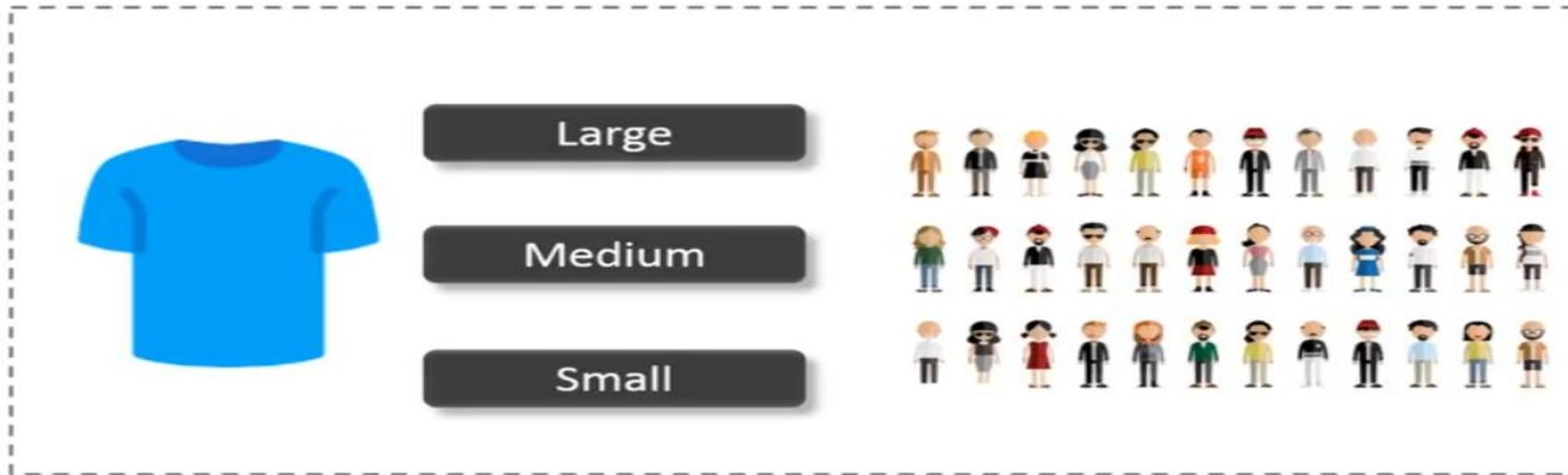
*Descriptive statistics uses the data to provide descriptions of the population, either through numerical calculations or graphs or tables.*



Descriptive Statistics is mainly focused upon the main characteristics of data. It provides graphical summary of the data.

# Inferential Statistics

*Inferential statistics makes inferences and predictions about a population based on a sample of data taken from the population in question.*



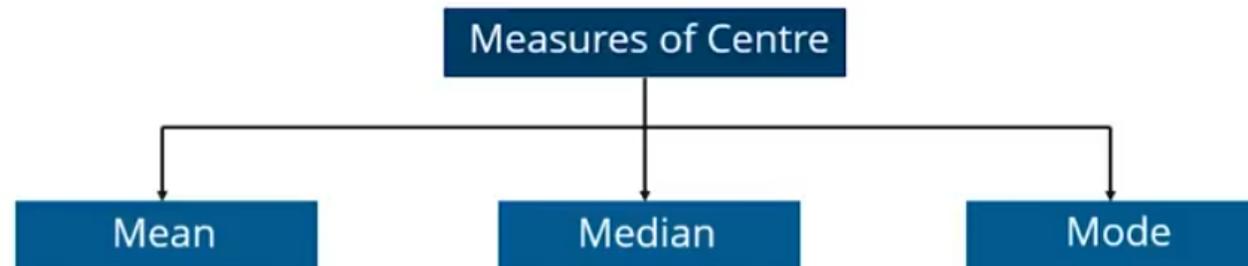
**Inferential statistics, generalizes a large dataset and applies probability to draw a conclusion. It allows us to infer data parameters based on a statistical model using a sample data.**

# Descriptive Statistics

*Descriptive statistics is a method used to describe and understand the features of a specific data set by giving short summaries about the sample and measures of the data.*

Descriptive statistics are broken down into two categories:

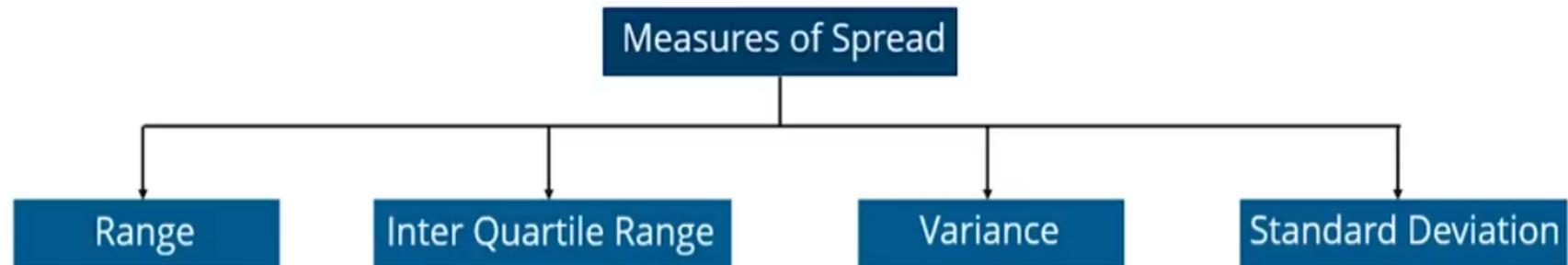
- **Measures of Central tendency**



*Descriptive statistics is a method used to describe and understand the features of a specific data set by giving short summaries about the sample and measures of the data.*

Descriptive statistics are broken down into two categories:

- **Measures of Variability (spread)**

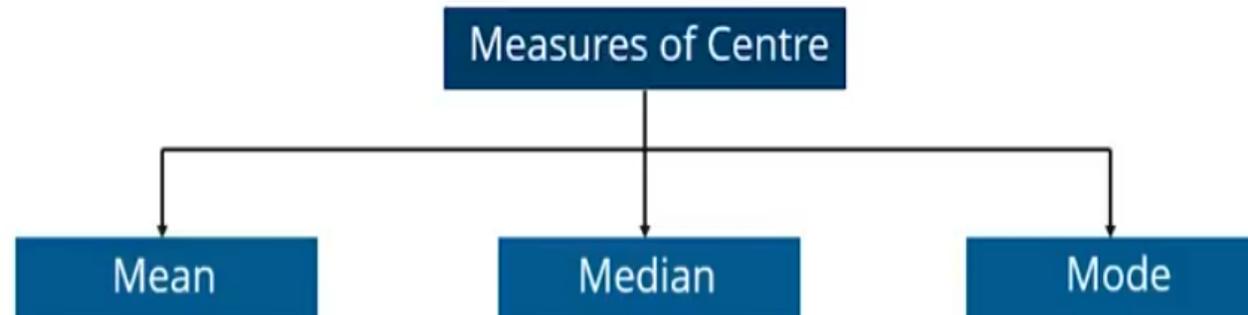


# Measures of Centre

*Descriptive statistics is a method used to describe and understand the features of a specific data set by giving short summaries about the sample and measures of the data.*

Descriptive statistics are broken down into two categories:

- **Measures of Central tendency**

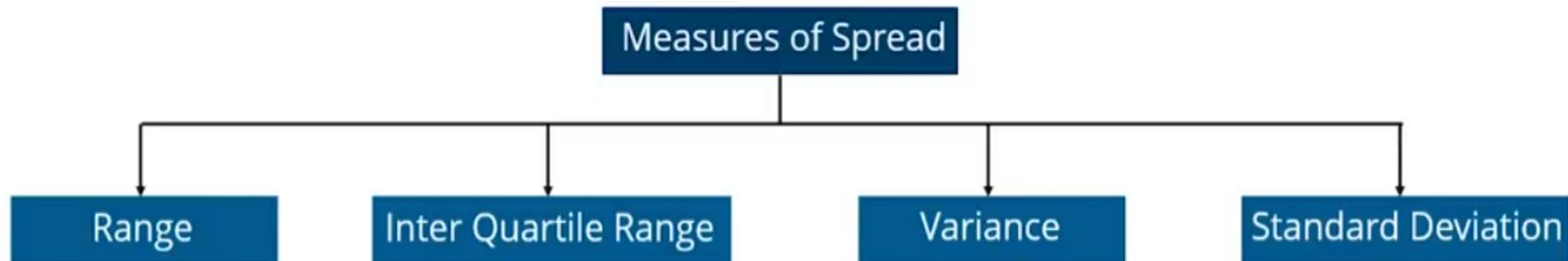


# Measures of Spread

*Descriptive statistics is a method used to describe and understand the features of a specific data set by giving short summaries about the sample and measures of the data.*

Descriptive statistics are broken down into two categories:

- **Measures of Variability (spread)**



# Mean

Here is a sample dataset of cars containing the variables:

- Cars,
- Mileage per Gallon(mpg)
- Cylinder Type (cyl)
- Displacement (disp)
- Horse Power(hp)
- Real Axle Ratio(drat)

Cars	mpg	cyl	disp	hp	drat
MazdaRX4	21	6	160	110	3.9
MazdaRX4_W AG	21	6	160	110	3.9
Datsun_710	22.8	4	108	93	3.85
Alto	21.3	6	108	96	3
WagonR	23	4	150	90	4
Toyata_11	23	6	108	110	3.9
Honda_12	23	4	160	110	3.9
Ford_11	23	6	160	110	3.9

## Mean

Measure of average of all the values in a sample is called Mean.

Here is a sample dataset of cars containing the variables:

- Cars,
- Mileage per Gallon(mpg)
- Cylinder Type (cyl)
- Displacement (disp)
- Horse Power(hp)
- Real Axle Ratio(drat)

Cars	mpg	cyl	disp	hp	drat
MazdaRX4	21	6	160	110	3.9
MazdaRX4_W AG	21	6	160	110	3.9
Datsun_710	22.8	4	108	93	3.85
Alto	21.3	6	108	96	3
WagonR	23	4	150	90	4
Toyota_11	23	6	108	110	3.9
Honda_12	23	4	160	110	3.9
Ford_11	23	6	160	110	3.9

## Mean

To find out the average horsepower of the cars among the population of cars, we will check and calculate the average of all values:

$$\frac{110 + 110 + 93 + 96 + 90 + 110 + 110 + 110}{8} = 103.625$$

# Median

Here is a sample dataset of cars containing the variables:

- *Cars*,
- *Mileage per Gallon(mpg)*
- *Cylinder Type (cyl)*
- *Displacement (disp)*
- *Horse Power(hp)*
- *Real Axle Ratio(drat)*

Cars	mpg	cyl	disp	hp	drat
MazdaRX4	21	6	160	110	3.9
MazdaRX4_W AG	21	6	160	110	3.9
Datsun_710	22.8	4	108	93	3.85
Alto	21.3	6	108	96	3
WagonR	23	4	150	90	4
Toyota_11	23	6	108	110	3.9
Honda_12	23	4	160	110	3.9
Ford_11	23	6	160	110	3.9

## Median

Measure of the central value of the sample set is called **Median**.

Here is a sample dataset of cars containing the variables:

- Cars,
- Mileage per Gallon(mpg)
- Cylinder Type (cyl)
- Displacement (disp)
- Horse Power(hp)
- Real Axle Ratio(drat)

Cars	mpg	cyl	disp	hp	drat
MazdaRX4	21	6	160	110	3.9
MazdaRX4_W AG	21	6	160	110	3.9
Datsun_710	22.8	4	108	93	3.85
Alto	21.3	6	108	96	3
WagonR	23	4	150	90	4
Toyata_11	23	6	108	110	3.9
Honda_12	23	4	160	110	3.9
Ford_11	23	6	160	110	3.9

## Median

To find out the center value of mpg among the population of cars, arrange records in *Ascending order*, i.e., **21, 21, 21.3, 22.8, 23, 23, 23, 23**

In case of even entries, take average of the two middle values, i.e.  $(22.8+23)/2 = 22.9$

# Mode

Here is a sample dataset of cars containing the variables:

- *Cars*,
- *Mileage per Gallon(mpg)*
- *Cylinder Type (cyl)*
- *Displacement (disp)*
- *Horse Power(hp)*
- *Real Axle Ratio(drat)*

Cars	mpg	cyl	disp	hp	drat
MazdaRX4	21	6	160	110	3.9
MazdaRX4_W					
AG	21	6	160	110	3.9
Datsun_710	22.8	4	108	93	3.85
Alto	21.3	6	108	96	3
WagonR	23	4	150	90	4
Toyata_11	23	6	108	110	3.9
Honda_12	23	4	160	110	3.9
Ford_11	23	6	160	110	3.9

## Mode

The value most recurrent in the sample set is known as Mode.

Here is a sample dataset of cars containing the variables:

- *Cars*,
- *Mileage per Gallon(mpg)*
- *Cylinder Type (cyl)*
- *Displacement (disp)*
- *Horse Power(hp)*
- *Real Axle Ratio(drat)*

Cars	mpg	cyl	disp	hp	drat
MazdaRX4	21	6	160	110	3.9
MazdaRX4_W AG	21	6	160	110	3.9
Datsun_710	22.8	4	108	93	3.85
Alto	21.3	6	108	96	3
WagonR	23	4	150	90	4
Toyota_11	23	6	108	110	3.9
Honda_12	23	4	160	110	3.9
Ford_11	23	6	160	110	3.9

### Mode

To find the most common type of cylinder among the population of cars, check the value which is repeated most number of times, i.e., *cylinder type 6*