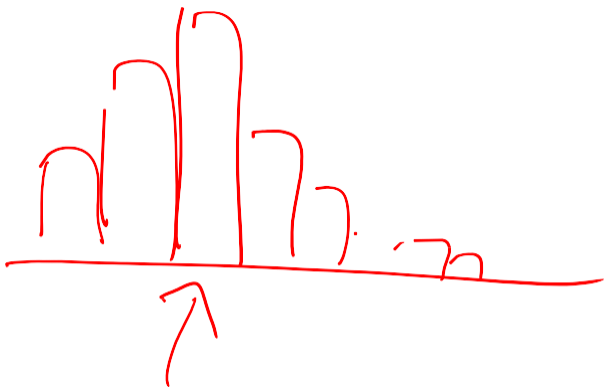
What is Central Tendency?

Whenever you measure things of the same kind, a fairly large number of such measurements will tend to cluster around the middle value. Such a value is called a measure of "Central Tendency". The other terms that are used synonymously are "Measures of Location", or "Statistical Averages".



Arithmetic Mean

Arithmetic Mean (called mean) is defined as the sum of all observations in a data set divided by the total number of observations. For example, consider a data set containing the following observations:

In symbolic form mean is given by $\bar{X} = \frac{\sum X}{n}$

"bar"

$\bar{X}$ = Arithmetic Mean

$\sum$ = adding

$\sum X$ = Indicates sum all X values in the data set

n = Total number of observations(Sample Size)

Arithmetic Mean -Example

The inner diameter of a particular grade of tire based on 5 sample measurements are as follows: (figures in millimeters)

565, 570, 572, 568, 585

Applying the formula $\bar{X} = \frac{\sum X}{n}$

We get mean = (565+570+572+568+585)/5 =572

Caution: Arithmetic Mean is affected by extreme values or fluctuations in sampling. It is not the best average to use when the data set contains extreme values (Very high or very low values).

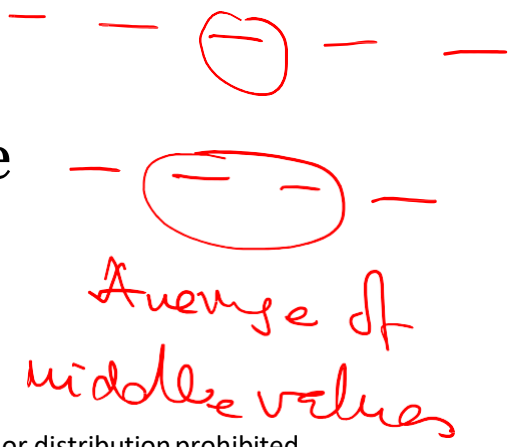
Median

Median is the middle most observation when you arrange data in ascending order of magnitude. Median is such that 50% of the observations are above the median and 50% of the observations are below the median.

Median is a very useful measure for ranked data in the context of consumer preferences and rating. It is not affected by extreme values (greater resistance to outliers)

$$\text{Median} = \frac{n+1}{2} \text{ th value of ranked data}$$

n = Number of observations in the sample



Median - Example

Marks obtained by 7 students in Computer Science Exam are given below: Compute the median.

45 40 60 80 90 65 55

Arranging the data after ranking gives

90 80 65 60 55 45 40

Median = (n+1)/2 th value in this set = (7+1)/2 th observation= 4th observation=60

Hence Median = 60 for this problem.

50th percentile

Mode

Mode is that value which occurs most often. It has the maximum frequency of occurrence. Mode also has resistance to outliers.

Mode is a very useful measure when you want to keep in the inventory, the most popular shirt in terms of collar size during festive season.

Mode -Example

The life in number of hours of 10 flashlight batteries are as follows: Find the mode.

340 350 340 340 320 340 330 330
340 350

340 occurs five times. Hence, mode=340.

Comparison of Mean, Median, Mode

Mean	Median	Mode
Defined as the arithmetic average of all observations in the data set.	Defined as the middle value in the data set arranged in ascending or descending order.	Defined as the most frequently occurring value in the distribution; it has the largest frequency.
Requires measurement on all observations.	Does not require measurement on all observations	Does not require measurement on all observations
Uniquely and comprehensively defined.	Cannot be uniquely determined under all conditions.	Not uniquely defined for multi-modal situations.

n even

multiple obs with same count

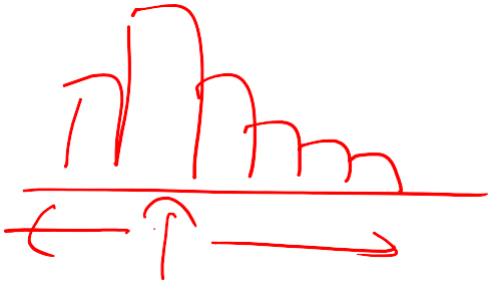
Comparison of Mean, Median, Mode Cont.

Mean	Median	Mode
Affected by extreme values. Can be treated algebraically. That is, Means of several groups can be combined.	Not affected by extreme values. Cannot be treated algebraically. That is, Medians of several groups cannot be combined.	Not affected by extreme values. Cannot be treated algebraically. That is, Modes of several groups cannot be combined.

Measures of Dispersion

Variation

In simple terms, measures of dispersion indicate how large the spread of the distribution is around the central tendency. It answers unambiguously the question " What is the magnitude of departure from the average value for different groups having identical averages?".



Range

Range is the simplest of all measures of dispersion. It is calculated as the difference between maximum and minimum value in the data set.

$$\text{Range} = X_{\text{Maximum}} - X_{\text{Minimum}}$$

$X_{max} = X_{min} \Rightarrow$ all obs are equal

Range-Example

Example for Computing Range

The following data represent the percentage return on investment for 10 mutual funds per annum. Calculate Range.

12, 14, 11, 18, 10.5, 11.3, 12, 14, 11, 9

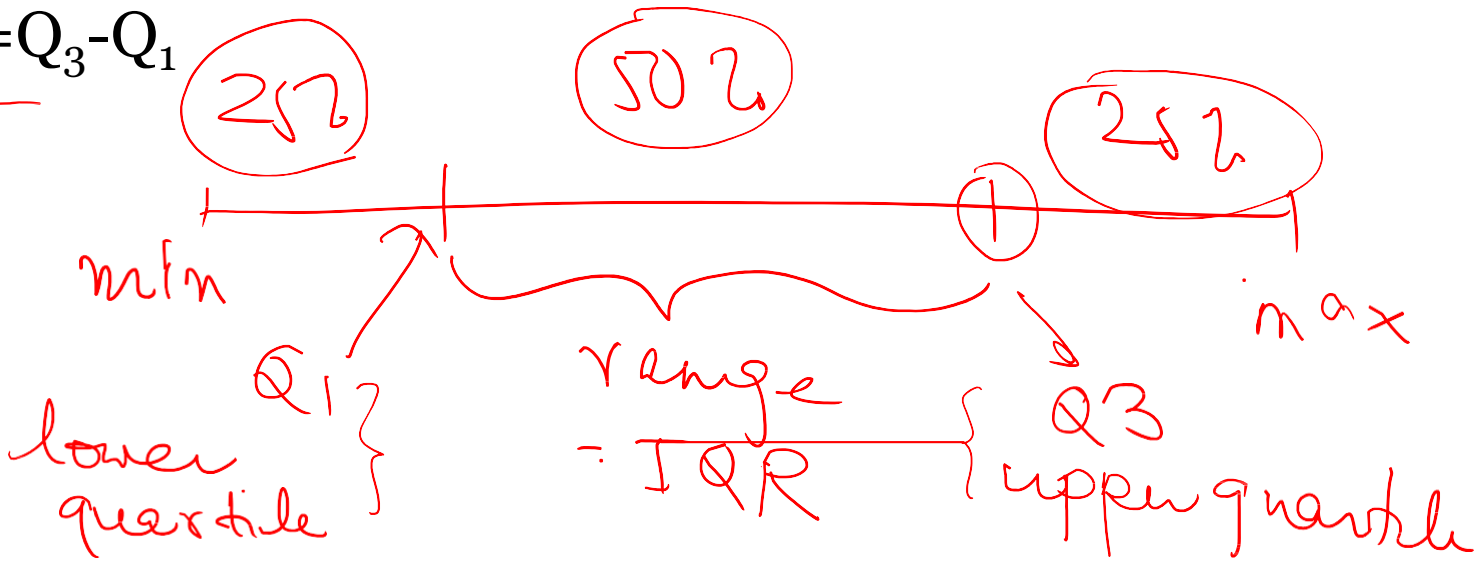
Range = $X_{Maximum} - X_{Minimum}$ = 18-9=9

Caution: If one of the components of range namely the maximum value or minimum value becomes an extreme value, then range should not be used.

Inter-Quartile Range(IQR)

IQR= Range computed on middle 50% of the observations after eliminating the highest and lowest 25% of observations in a data set that is arranged in ascending order. IQR is less affected by outliers.

$IQR = Q_3 - Q_1$



Interquartile Range-Example

The following data represent the annual percentage returns of 9 mutual funds.

Data Set: 12, 14, 11, 18, 10.5, 12, 14, 11, 9

Arranging in ascending order, the data set becomes

9, 10.5, 11, 11, 12, 12, 14, 14, 18

$IQR = Q_3 - Q_1 = 14 - 10.75 = 3.25$