

# UNIT 1

## Measures of Spread

Presented By :

Aditya Joshi

Assistant Professor

Graphic Era Deemed to be University, Dehradun

# Measures of Spread

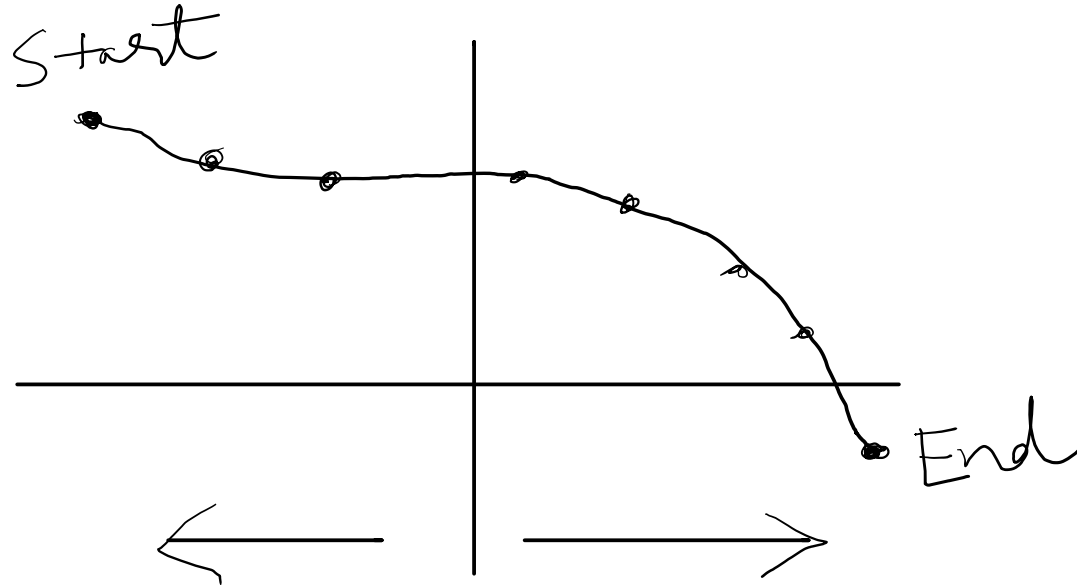
Measures of Spread is the important information tells you how spread out or varied your data set is.

Ex. Test scores that are in the 60-90 range might be expected while score in the 20-70 range might indicate a problem.

Range isn't the only measure of spread through.

# Range

- How spread out your data is
- It's the difference between the highest value and lowest value in data set



Q. Find the range of following set of numbers

10, 99, 87, 45, 67, 43, 45, 33, 21, 7, 65, 98

1. Sort the number in order smallest to largest
2. Subtract the smallest number from the largest number

Q. Range of these integers

14, -12, 7, 0, -5, -8, 17, -11, 19

**Standard Deviation:** range is about 4 times the standard deviation

$$\text{Standard deviation} = \frac{\text{Range}}{4}$$

- The standard deviation is another measure of spread in statistics. It tells you how your data is clustered around the mean.
- The most cases bulk of data can be formed pretty close to the mean.
- The result is that those erroneous outliers should have very little effect on your final statistic.

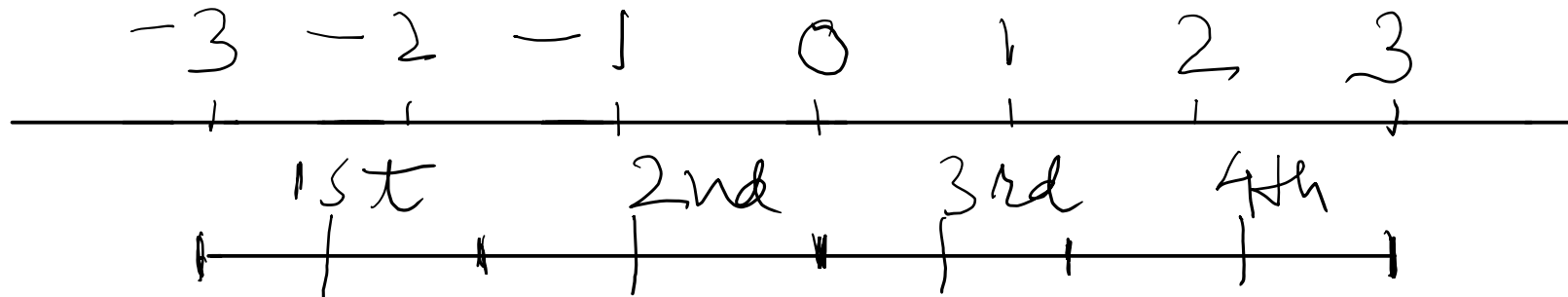
## Quartiles(Boundaries of the lowest, middle and upper Quarter of data):

Quartiles in stats are values that divide your data into quarters, they divide your data into four segments according to where the number fall in the number line.

The four quarters that divide a dataset into quartiles are

1. The lowest 25% of numbers.
2. The next lowest 25% of number (up to the median).
3. The second highest 25% of number (above the median).
4. The highest 25% of numbers.

A set of number -3, -2, -1, 0, 1, 2, 3 divided into four quartiles



Q. Divide the following dataset into quartiles even sets

2, 5, 6, 7, 10, 22, 13, 14, 16, 65, 45, 12

1. Put the dataset in order
2. Make groups by dividing total numbers of items in dataset by 4

Q. Uneven set of numbers

10, 13, 17, 19, 21

1. Put dataset in order find middle quartile which is **median**
2. find

25<sup>th</sup> Percentile

50<sup>th</sup> Percentile

75<sup>th</sup> Percentile

Q. Calculate the upper quartile Q3

27, 19, 5, 7, 6, 9, 15, 12, 18, 2, 1

1. Put in Order
2. Find median
3. Median of upper set of number

**$Q1 = 1/4 (n+1)\text{th term}$**

**$Q2 = 2/4 (n+1)\text{th term}$**

**$Q3 = 3/4 (n+1)\text{th term}$**

**Interquartile Range:** (where the middle fifty percent of your data is)

- The interquartile range is a measure of where the middle fifty is in a data set.
- Where a range is a measure of where the beginning and end are in a set.
- An interquartile range is a measure of where the bulk of the values lie.
- The interquartile range formula is the first quartile subtracted from the third quartile

$$\text{IQR} = Q3 - Q1$$



**Q. find the interquartile range of odd sets**

**1, 2, 5, 6, 7, 9, 12, 15, 18, 19, 27**

1. Put the numbers in order

2. find the median and divide data in equal set of lower and upper half from median

3. Find Q1 and Q3

Q1 is median of lower half of data

Q3 is median of Upper half of data

4.  $Q3 - Q1$

**Q. Find the interquartile range of even set of numbers  
3, 5, 7, 8, 9, 11, 15, 16, 20, 21**

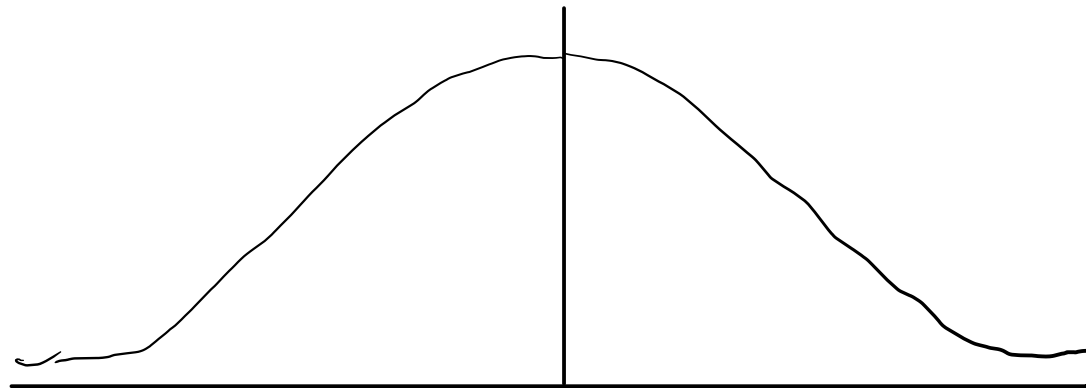
1. Put the number in order
2. Make a mark in center of data
3. Find Q1 and Q3
4.  $Q3 - Q1$

# Skewed (Does your data have mainly low, or mainly high values)

## Skewed Distribution:

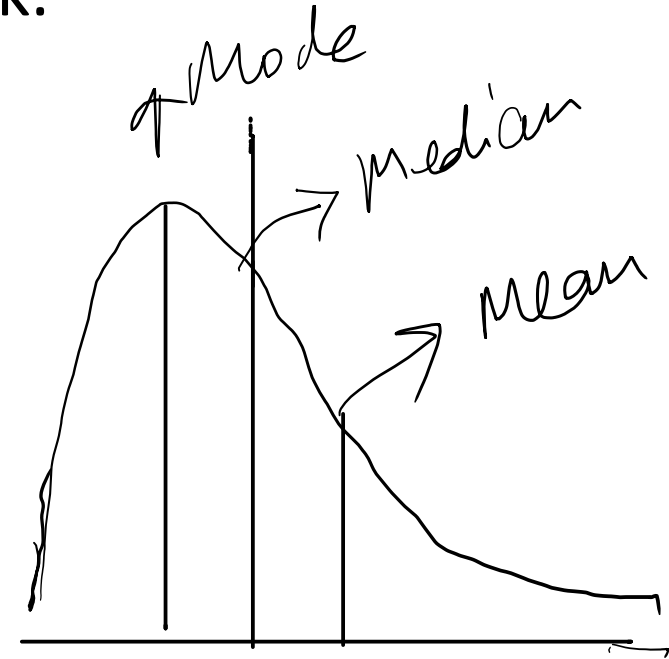
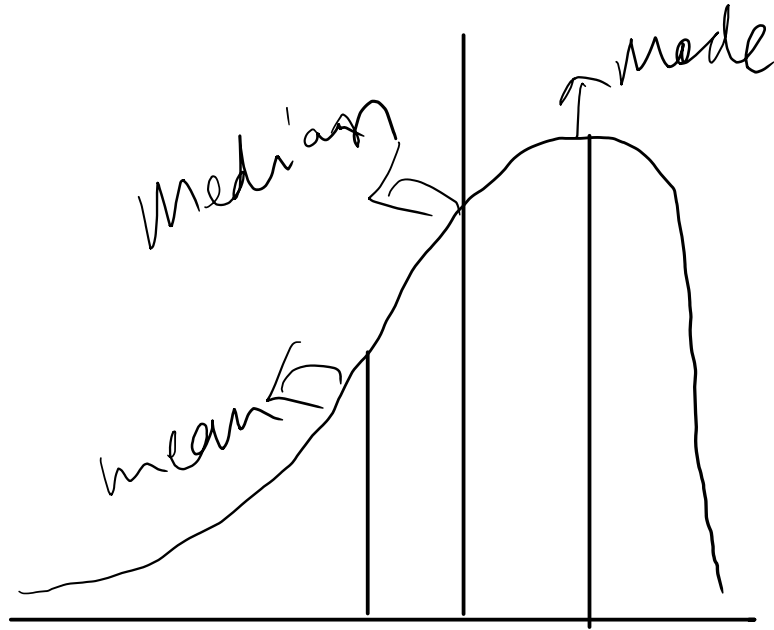
if one tail is longer than other , the distribution is skewed. These distribution are sometimes called asymmetric or asymmetrical distributions as they don't show any kind of symmetry. Symmetry means that one half of the distribution is mirror image of the other half.

For example the normal distribution is a symmetric distribution with no skew the tails are exactly the same



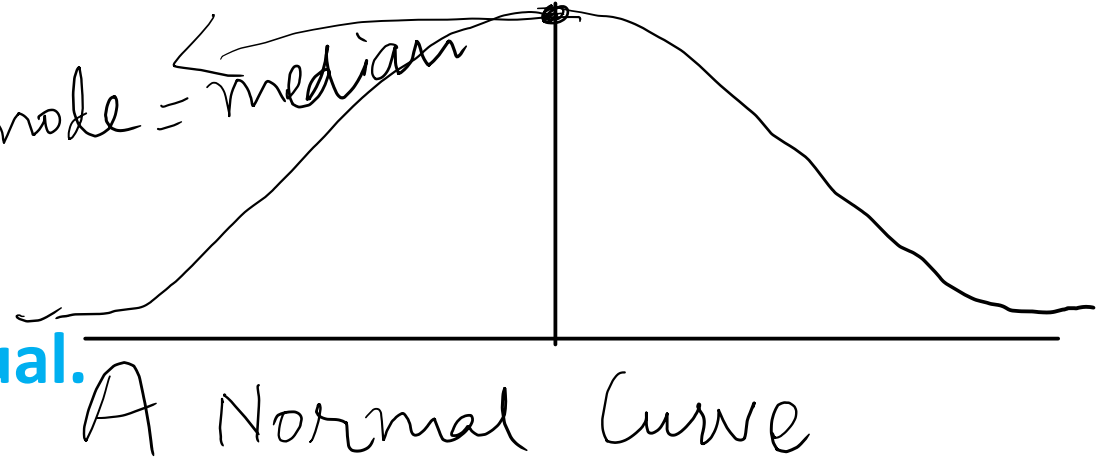
A Normal Curve

- **A left skewed distribution** has a long left tail this is also called negatively skewed distribution. That because there is a long tail in the negative direction on a number line. The mean is also to the left peak.
- **A right skewed distribution** has a long right tail also called positive skewed distribution because there is a long tail in positive direction of number line the mean is also to the right of peak.



- In a normal distribution the mean mode and the median are the same number while mean mode and median in a skewed distribution are different. It is some times called **bell curve**

*mean = mode = median*



### Properties of Normal distribution:

1. The mean, mode and median all are equal.
2. The curve is symmetric at the center
3. Exactly half of the value are to the left of center and half are to the right
4. The total area under the curve(AUC) is 1

**Kurtosis: (a measure of how much data is in the tails)**

**A positive value tells you that you have heavy tails(i.e lot of data in your tails)**

**A negative value tells you that you have light tails(i.e a little data in your tails)**