Q)) What is mean, median, mode

R) Mean shows the avg. Mean is also called the avg. Is only applicable for interval and ratio data. Affected by each value.

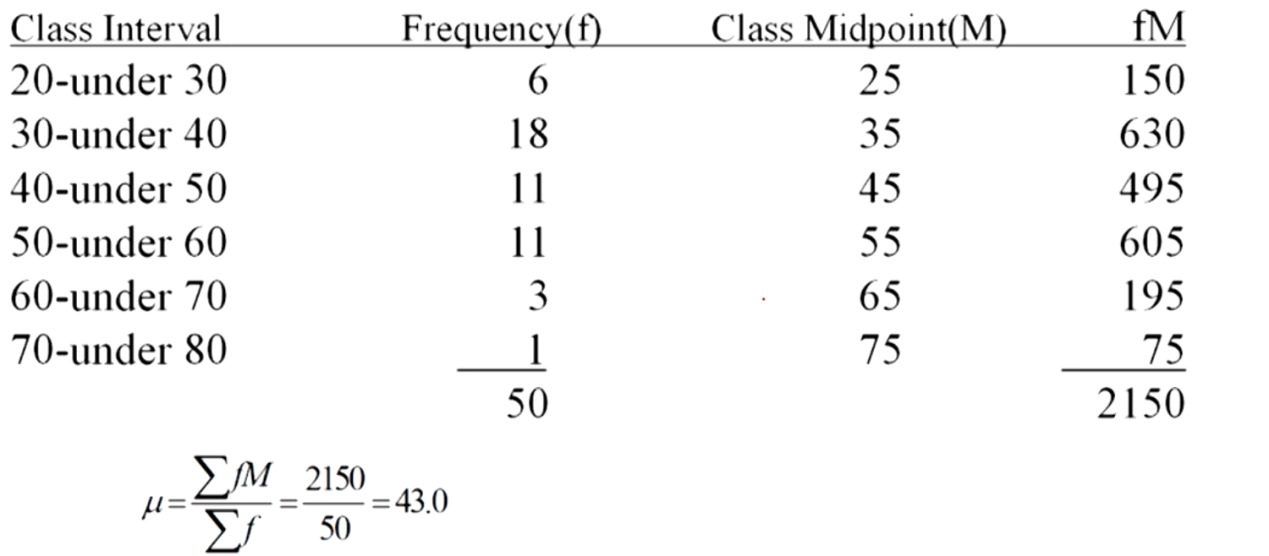
Median is the middle value. Applicable for ordinal, interval, ratio data. Uaffected by extremely large and extremely small values.

Mode is the most occuring value.

Q) What is the use of central tendency.

R) To represent the characteristic of a group of data through a single number.

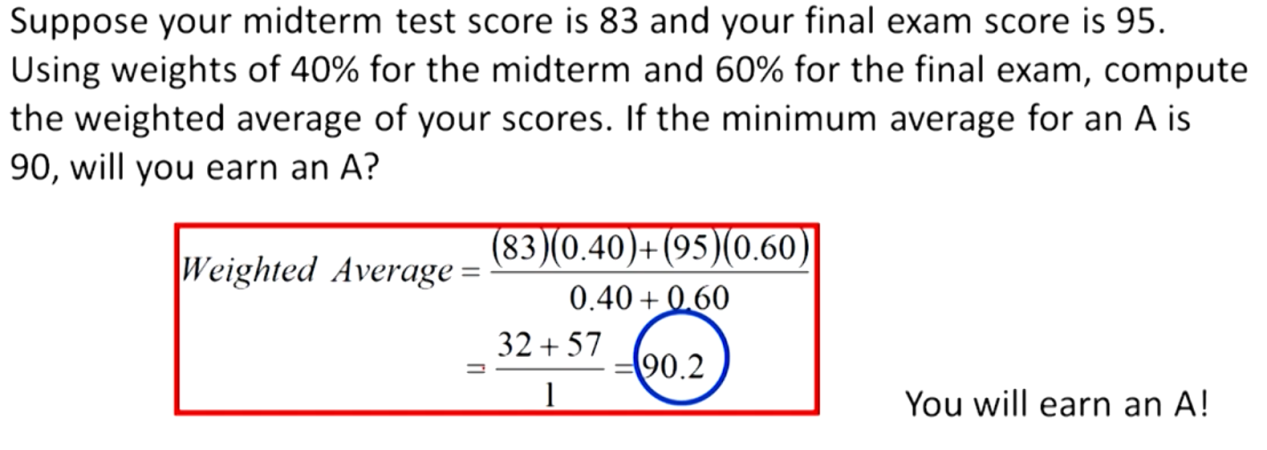
1. Different types of central tendency.
2. Arithematic mean, Weighted mean, median, Percentile.
3. What is despersion and its types.
4. Skewness, kurtosis, range, variance, Standard score, varience.
5. What is arithematic mean.
6. It is the sum of all entities divided by number of entities. Applicable for interval and ratio data and not for nominal or ordinal data.
7. Symbols to represent the population mean and sample mean.
8. Population mean is shown by mue and sample mean by x(bar).
9. What is the mean of grouped data.
10. It is the submission of frequency(Midpoint)/total frequency



1. What is the weighted avg
2. When certain values contain more weightage than others(are of more value).

Then we go for weighted avg.

Eg.



1. What is median.
2. Arrange the data in ASCENDING ORDER FIRST.

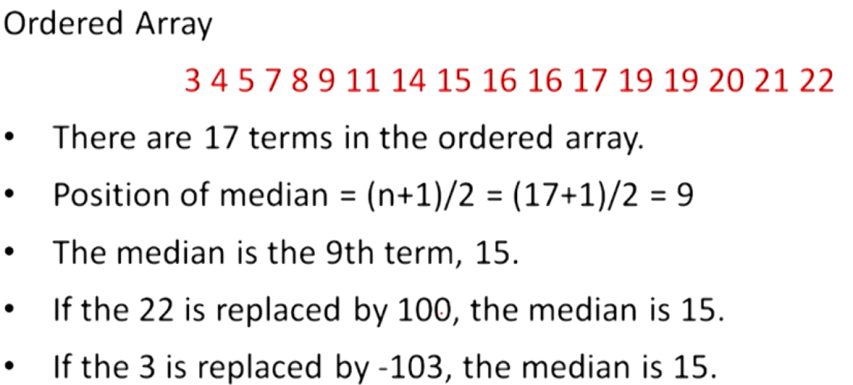
Method 1-

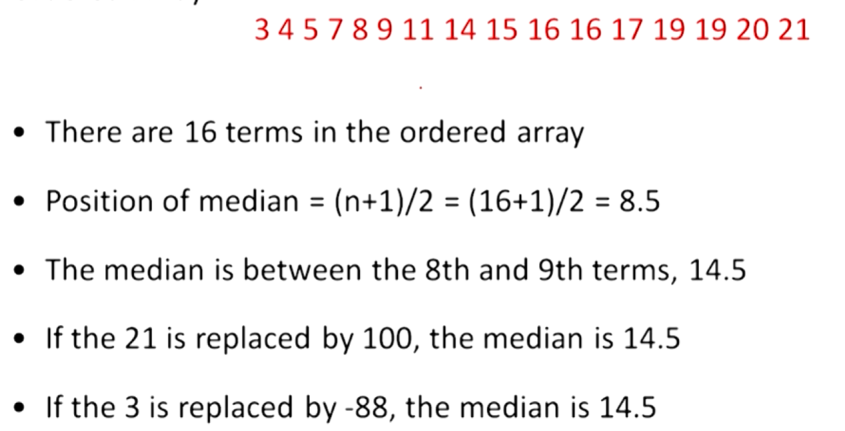
For odd number of terms.--> arithematic mean of middle two terms.

For even number of terms.--> It is the middle term.

Method 2- \*\*(**Preferred)\*\***

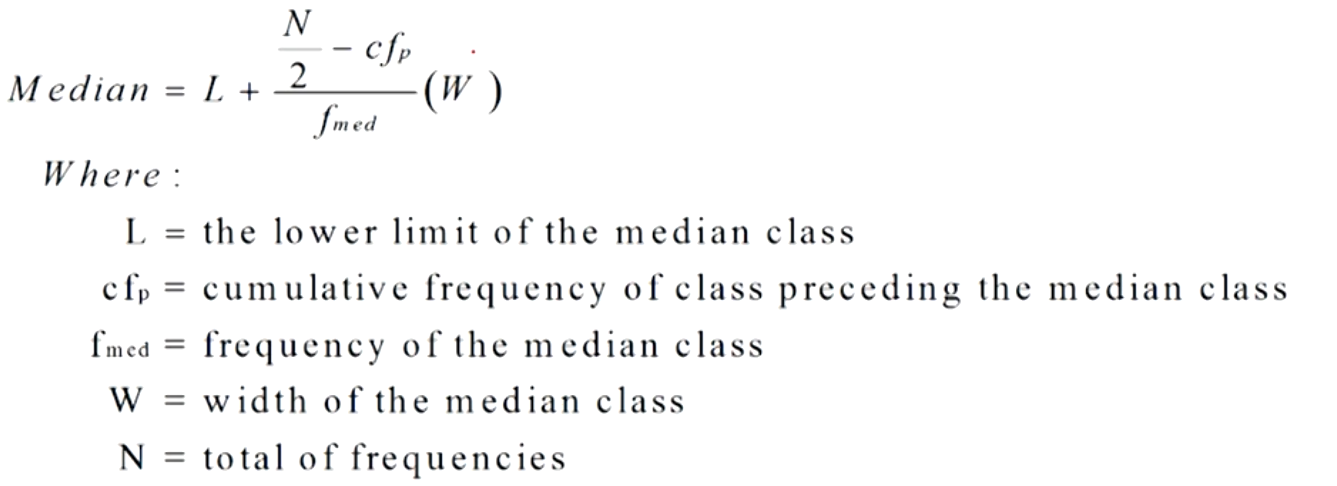
Median’s position is an ordered array given by (n+1)/2



**

*Looking at the above scenerio we can notice that even large change in the values dosent affect the median.*

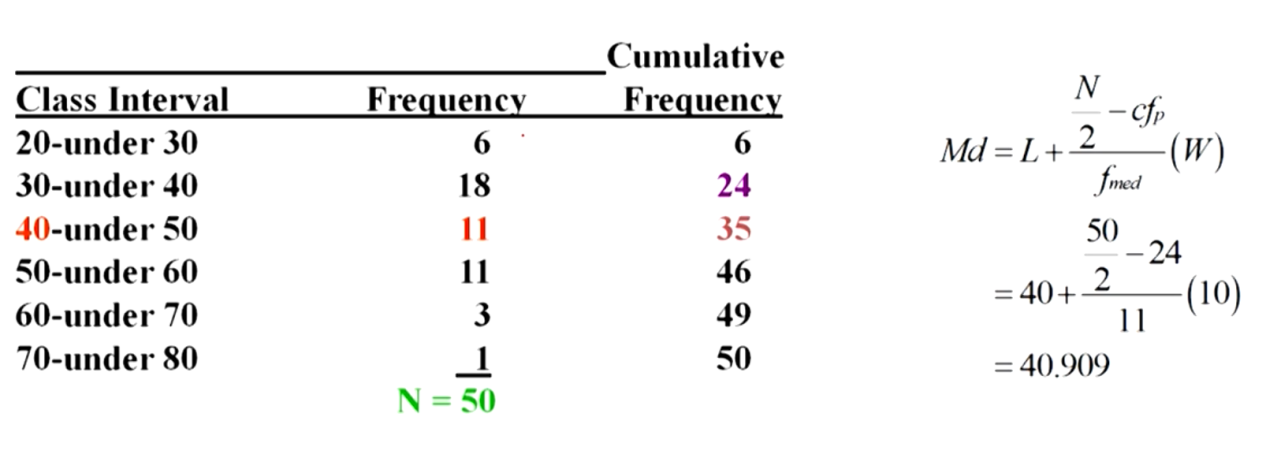
1. How do we find the median of grouped data.



Step 1) Find the total frequency.

Step 2) Divide the total frequency by 2 and look for the Cumulative Frequency column for the range it lies in.

Pick the Class Interval corresponding to that frequency.



Q)What is Mode.

R) It is the most frequently occuring value in a dataset.

It is applicable to all levels of data measurement-- nominal, ordinal, interval and ratio.

Q)What is the different types of modes available that dataset can be divided in.

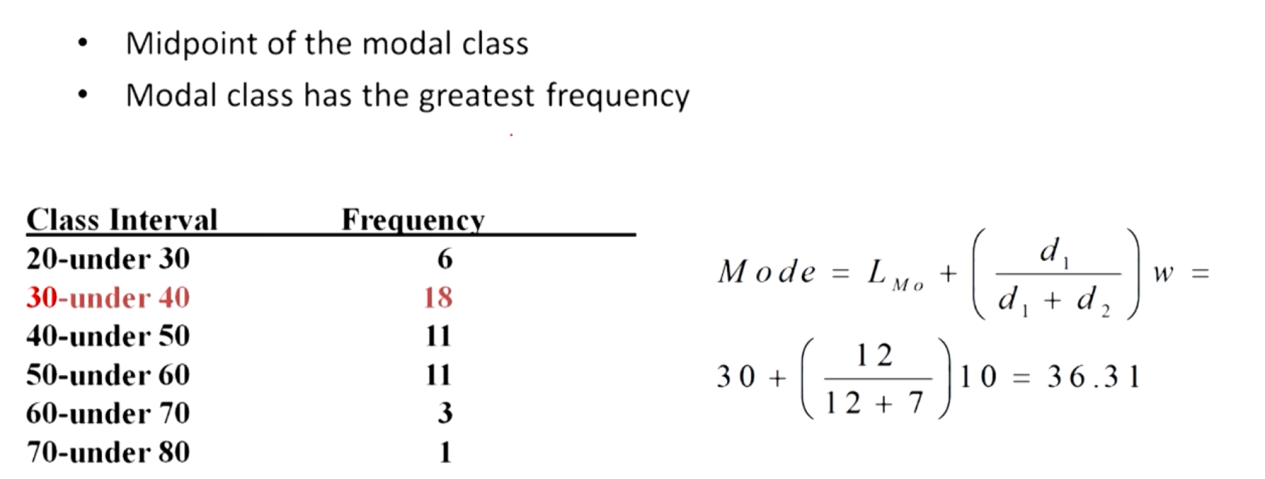
R) Bimodal-- Dataset that have two modes. Multimodal-- Dataset that contains more than two modes.

Q)How can we find mode of grouped data.

R) Step1) Find the mode class i.e. class having the highest frequency.

Step 2) d1= diff b/w frequency of mode class and the prev. Class.

Step 3) d2= diff b/w frequency of mode class and the later Class.



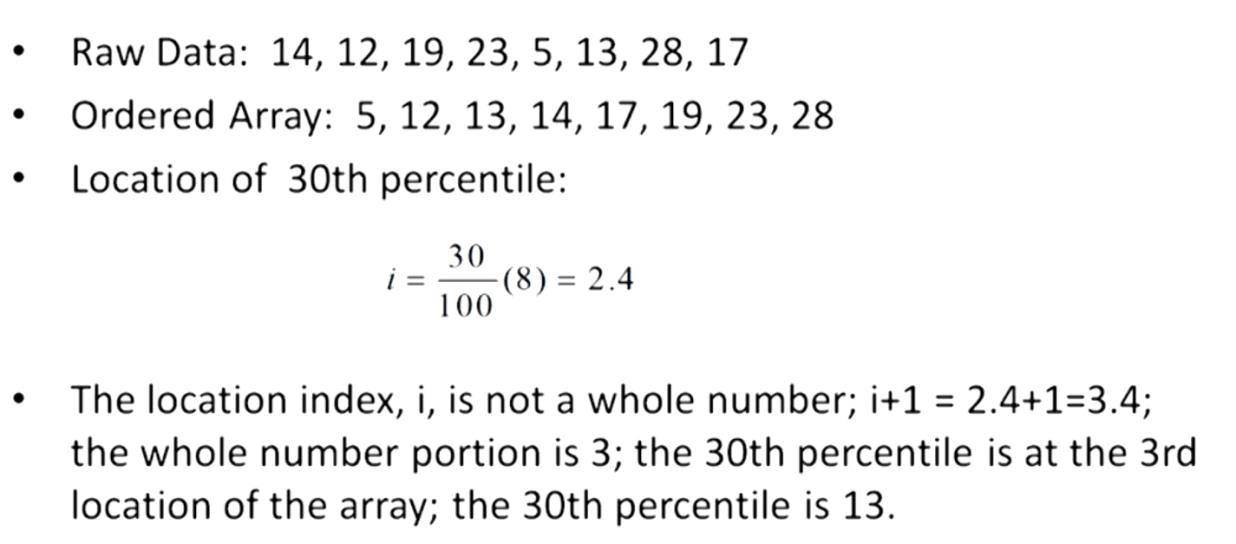
1. When to use mean, median, mode.
2. When data is squed we should follow

--- Go for median as a central tendancy.

When data is not sqwed(a bell-shaped curved)

--- Use any mean, median, mode.

1. What is the relation between median and percentile.
2. Median and 50th percentile means same value.



1. What is Dispersion.
2. Dispersion is same as Variability.

Measure of variablility describe the spread or the dispersion of the data.

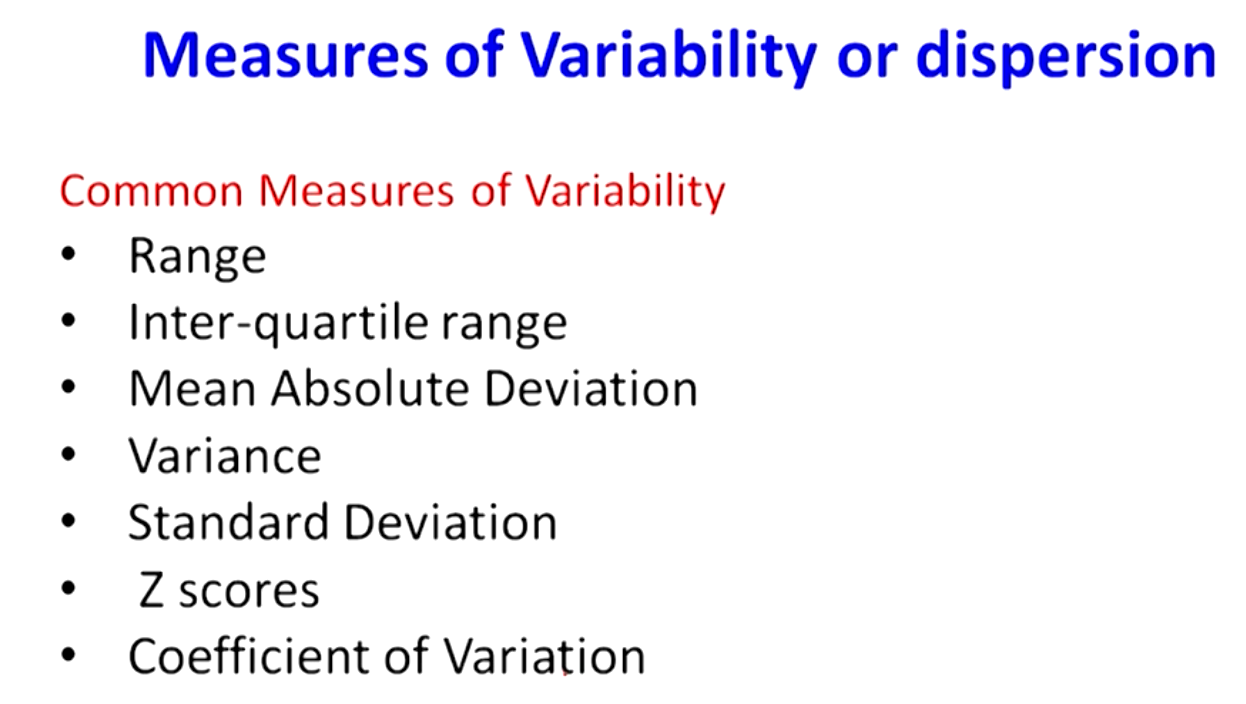
Reliablility of measures of central tendency is calculated by dispersion.

The quality of mean is explained by its variability.

Eg. The mean is same in the given below case even when the quality of mean is different.

The second dataset has more variability.





1. What is range and how will you find it.
2. It is the difference between the largest and the smallest value in the dataset.

RANGE=LARGEST-SMALLEST.

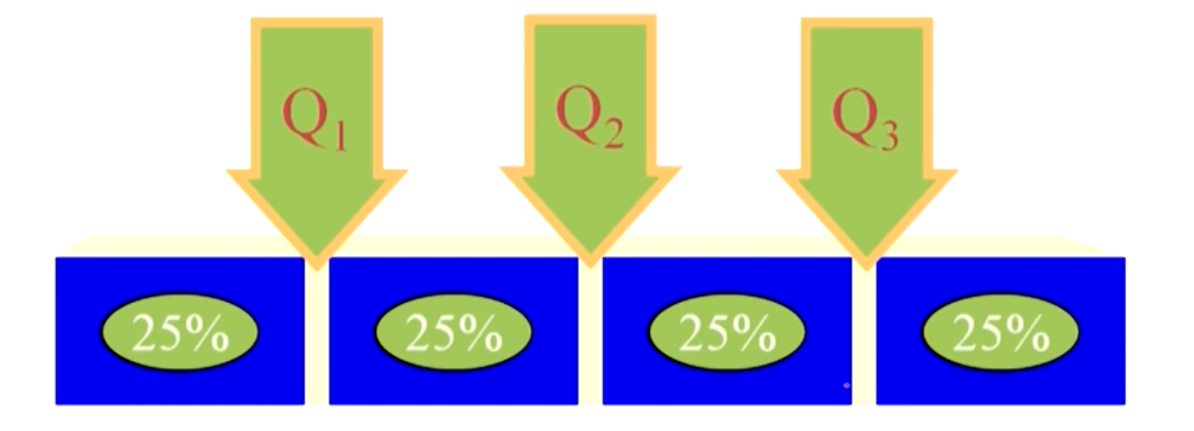
Problem-- Ignores all the values except the two extremes.

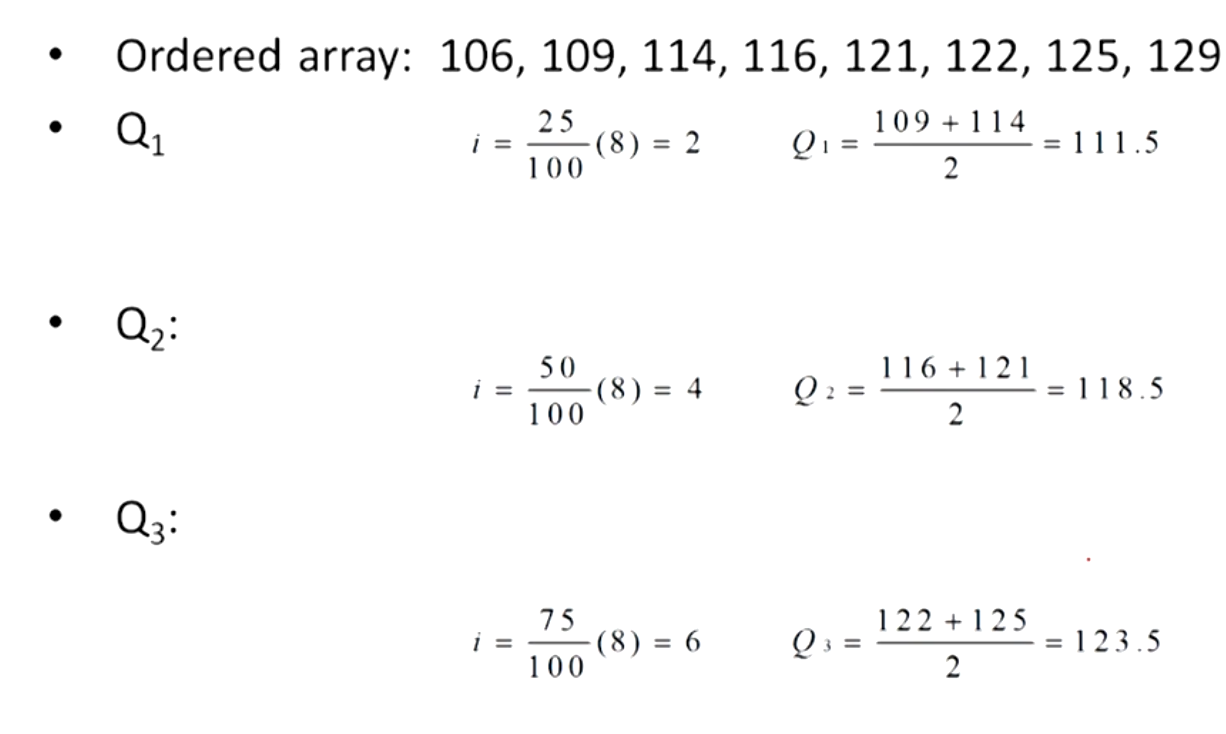
1. What is Quartiles.
2. It is the measure of central tendency that divides a group of data into four subgroups.

Q1:: means 25% percentile.

Q2:: means median.

Q3:: means 75% percentile.





1. What is Inter-quartile range.
2. Range of values between the first and the third quartile.

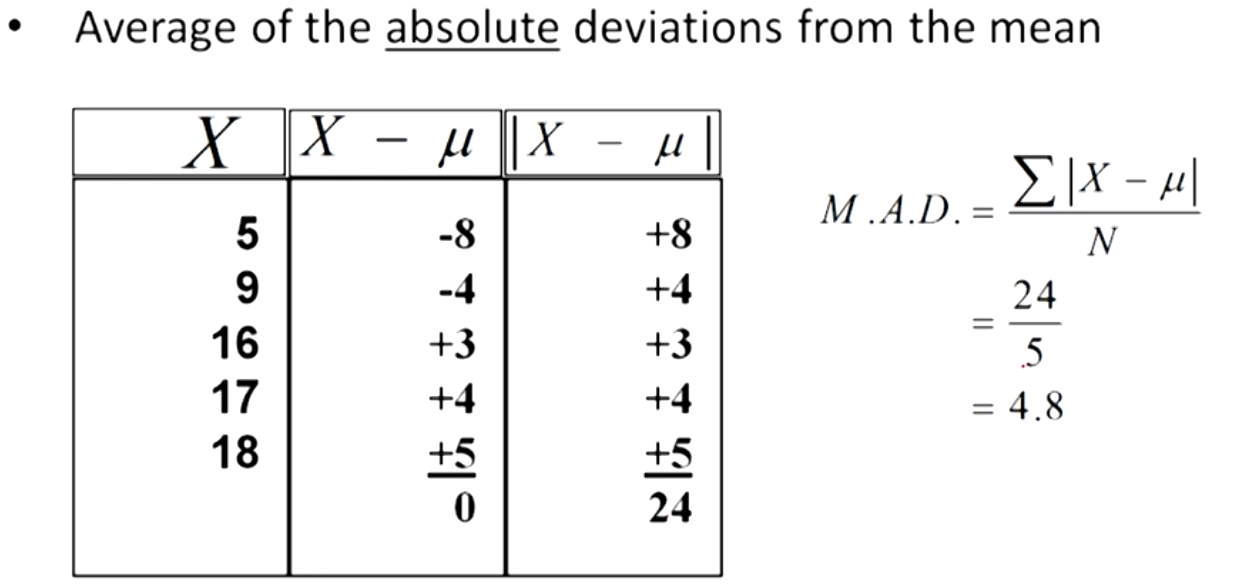
Less influenced by extremes.

Inter-quartile = Q3-Q1

1. How to find the deviations from the mean, what is absolute deviation and what is mean squared deviation.
2. First find the mean. Then find the deviations from the mean by subtracting the values of the mean from the dataset.

**Absolute deviation:**

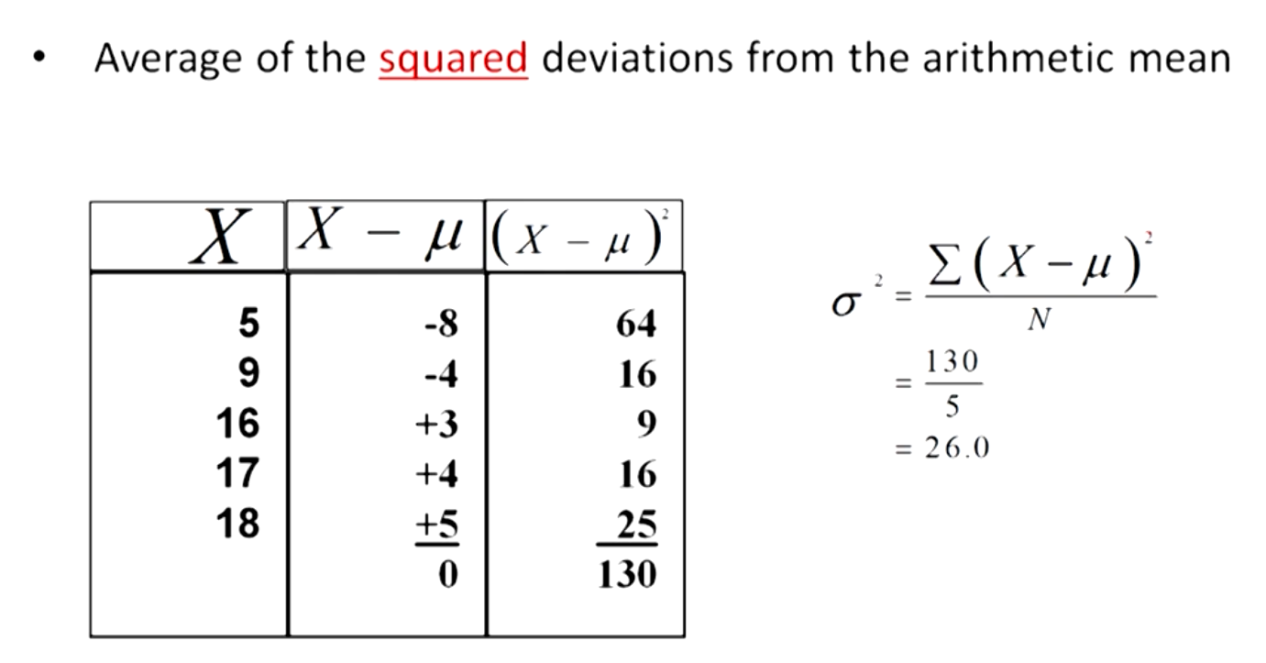
**Advantage--** it removes the negative numbers.



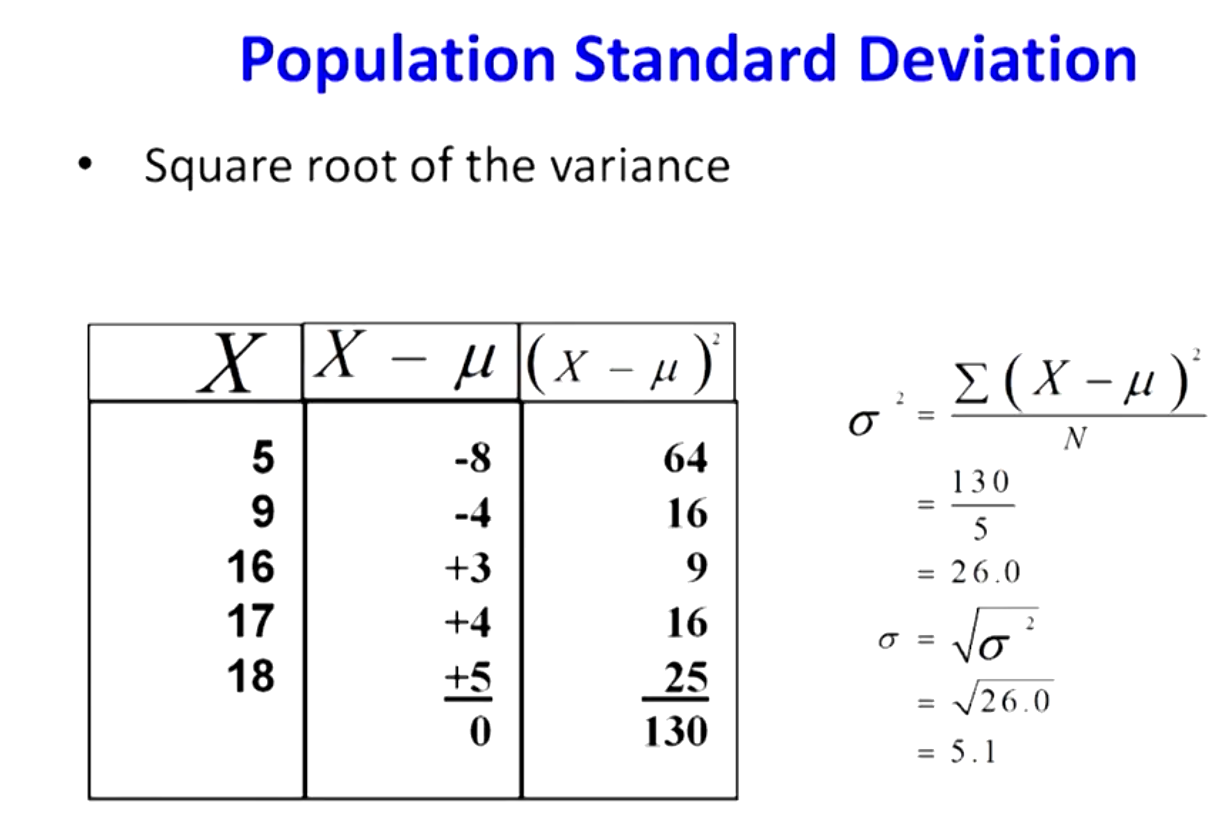
**Mean square deviation:**

**Advantage--** It removes the negative numbers. It also makes the deviation much larger. I.e the number that is further becomes more further so they are clearly distinguishable.

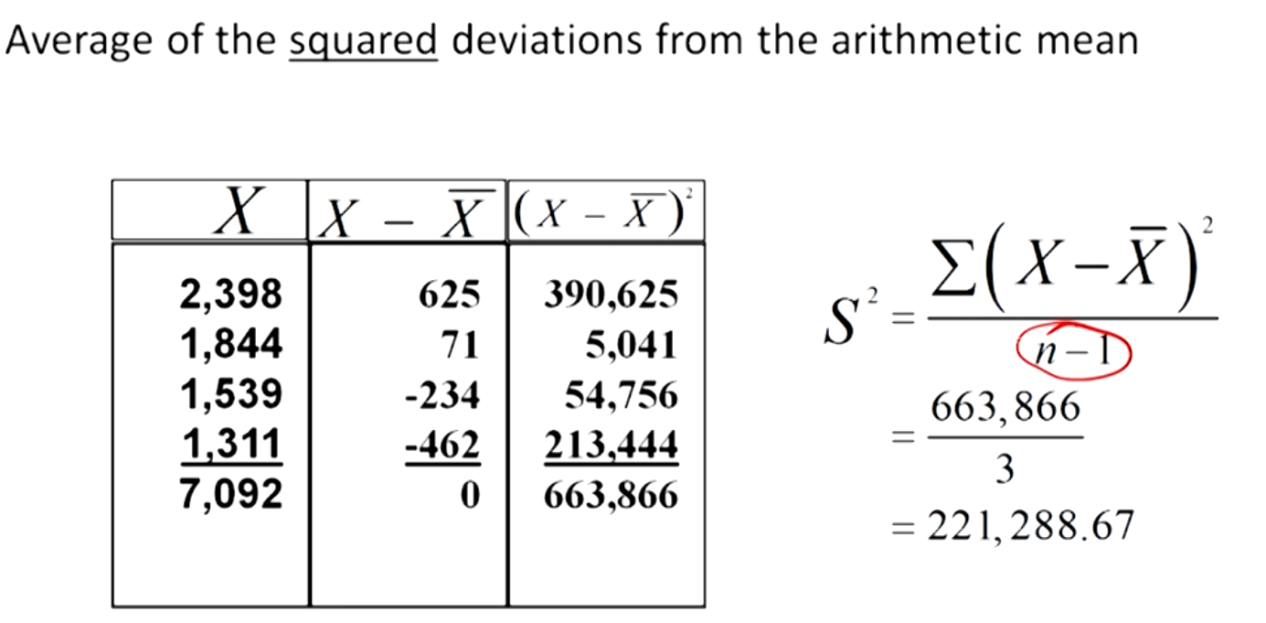
So it gives higher penalty for higher deviation.



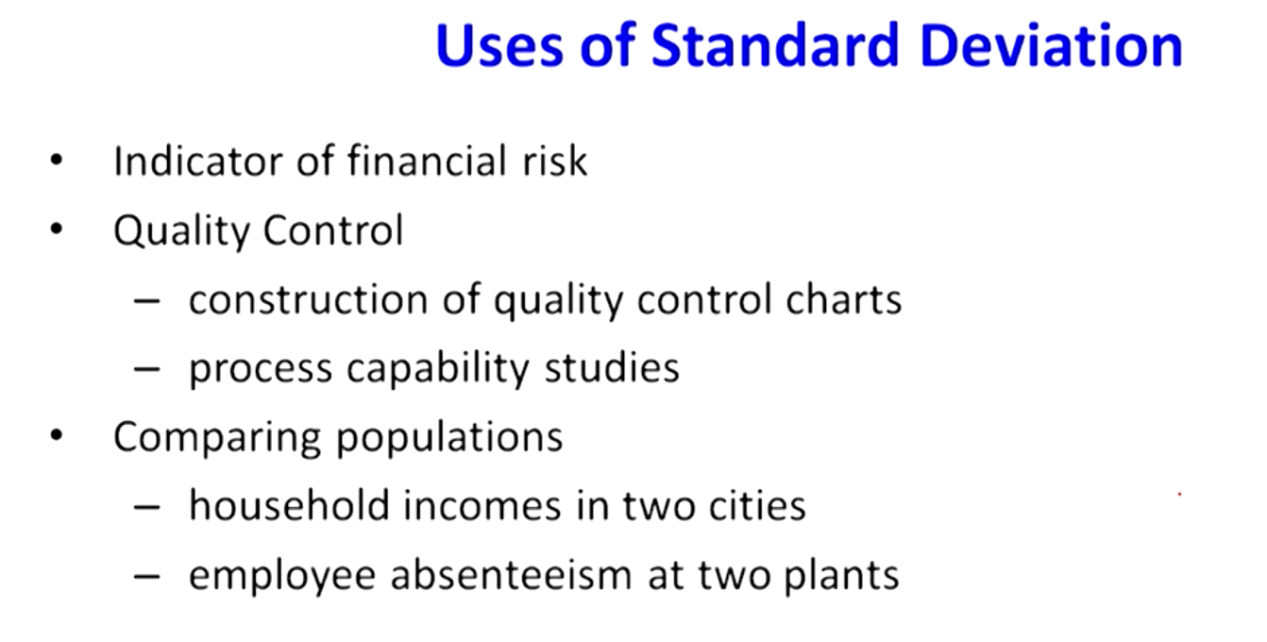
1. What is the best method to calculate the variability of the population
2. We find the root of the mean square deviation because if not done so there will be huge difference between the numbers.



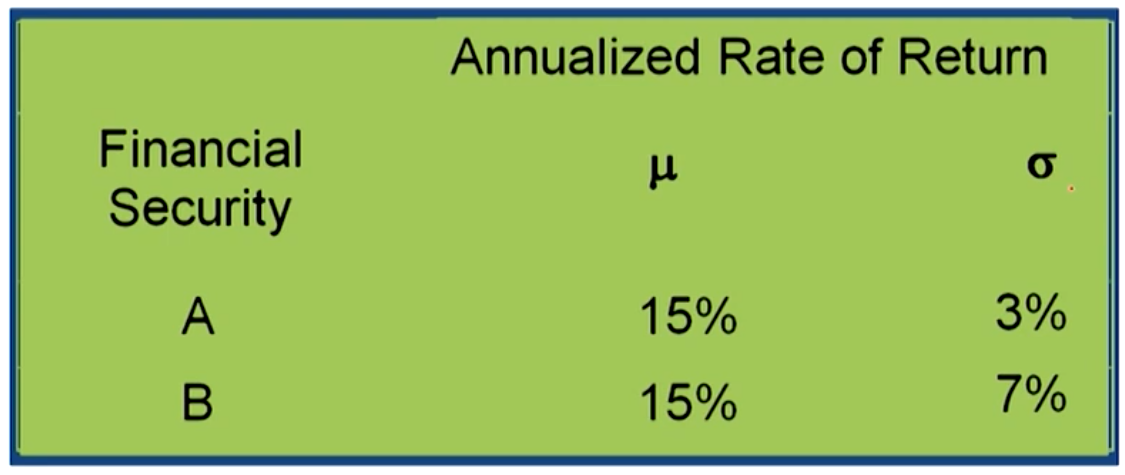
1. How to calculate the sample variance.
2. To make the variance as the unbiased estimator we divide it by (n-1)



1. Relation between standard deviation and variance.
2. SD=squareroot(variance)
3. Uses of standard deviation.



1. How is sd. Deviation used in financial risk analysis.



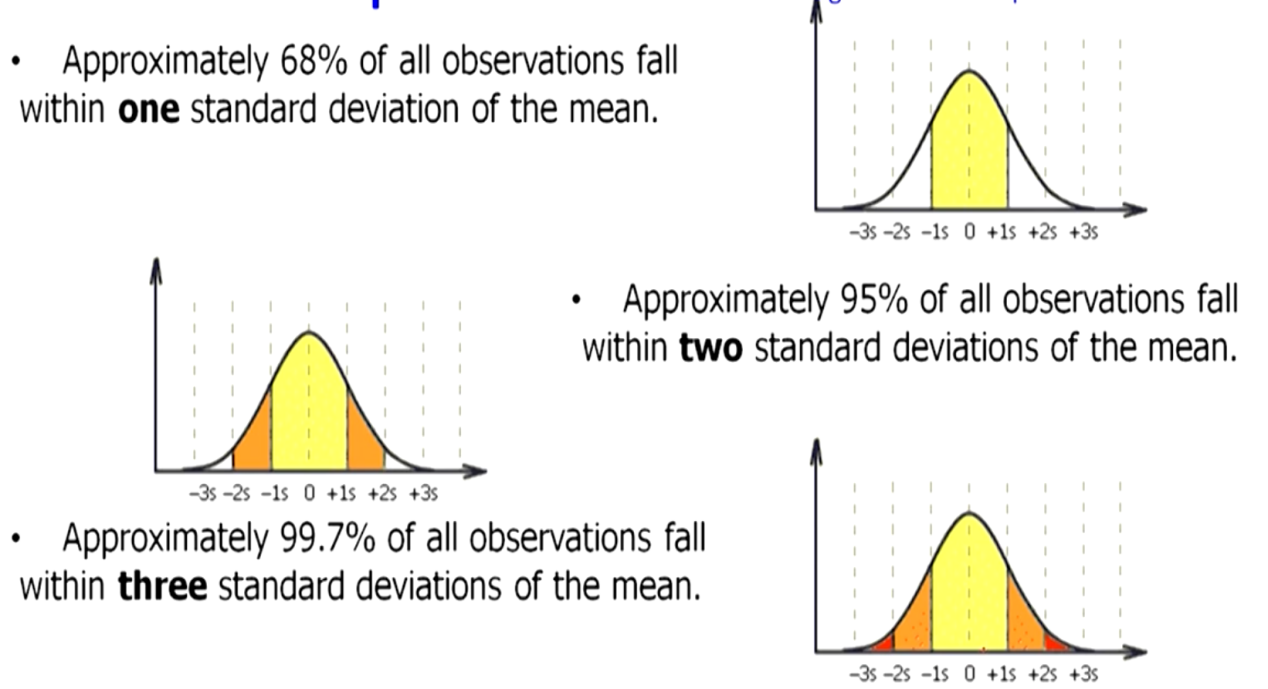
More standard deviation means more risk.

1. What is the relation between prediction and probability.
2. The prediction and probability go hand in hand. Science we cannot predict the exact value of an occurrence we have to calculate its probability.
3. Usage of bell curves in 6sigma.
4. **Properties of standard deviation for the bell shaped curve.**

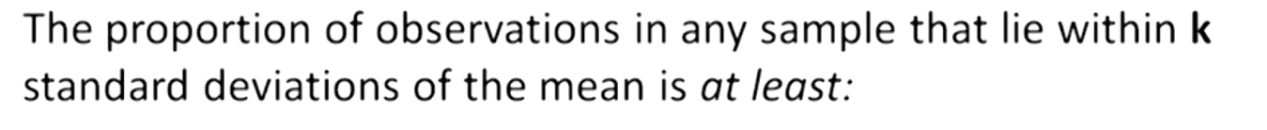
**For the bell shaped curves the normal distribution is the father of all the distributions.** From the mean 0 if you travel 1sigma on both the sides of the bell curve you can cover 68% of all the observations.

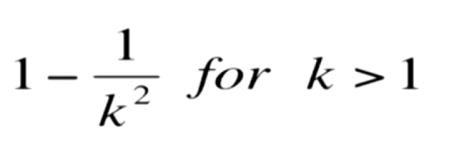
From the mean 0 if you travel 2 sigma on both the sides of the bell curve you can cover 95% of all observations.

If you travel 3sigma distance on either side from the mean of the normal distribution you can cover 99.3% of all the mean.



1. When the phenomenon does not follow a bell shaped curve what approach can you take to check the distribution.
2. The general interpretation of the standard deviation can be calculated from Chebysheff’s Theorem which applies to all shapes of histograms not just bell shaped curves.

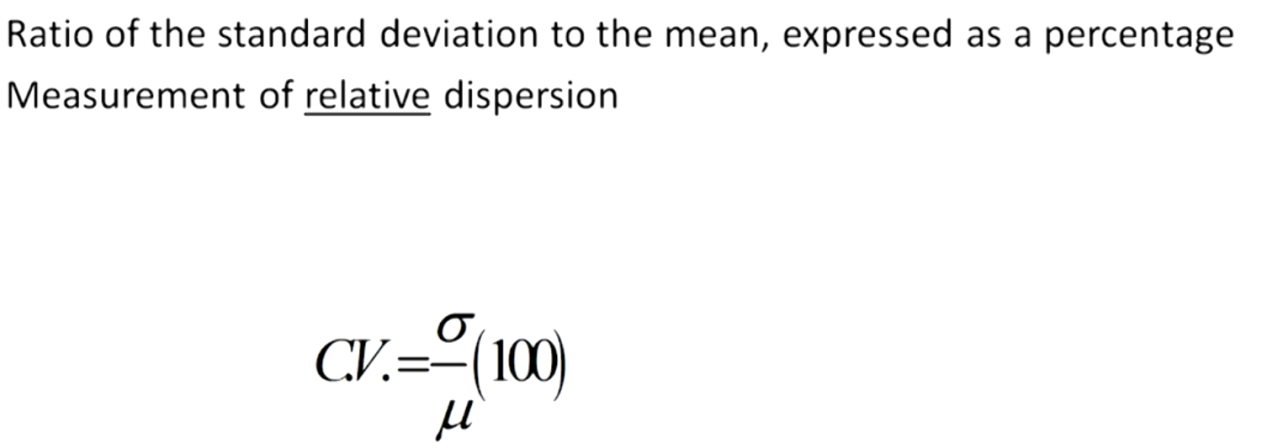


Here ‘k’ represents the value of standard deviation.

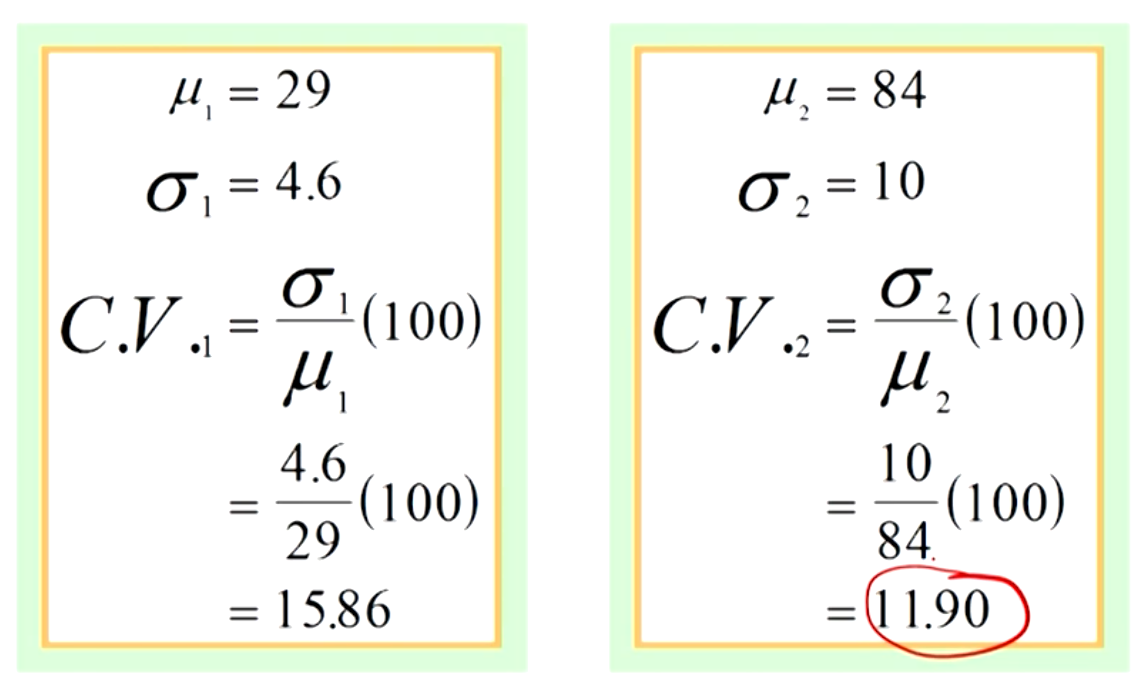
1. Find the value of 2 sd. deviation using chebysheff’s theorem.
2. For k=2, the theorem states that 3/4 of all observations lie within 2 standard deviation of the mean. 3/4 means 75% of all the data.

In comparison to normal standard deviation which is 95% of all the standard deviation.

1. Coefficient of Variation.

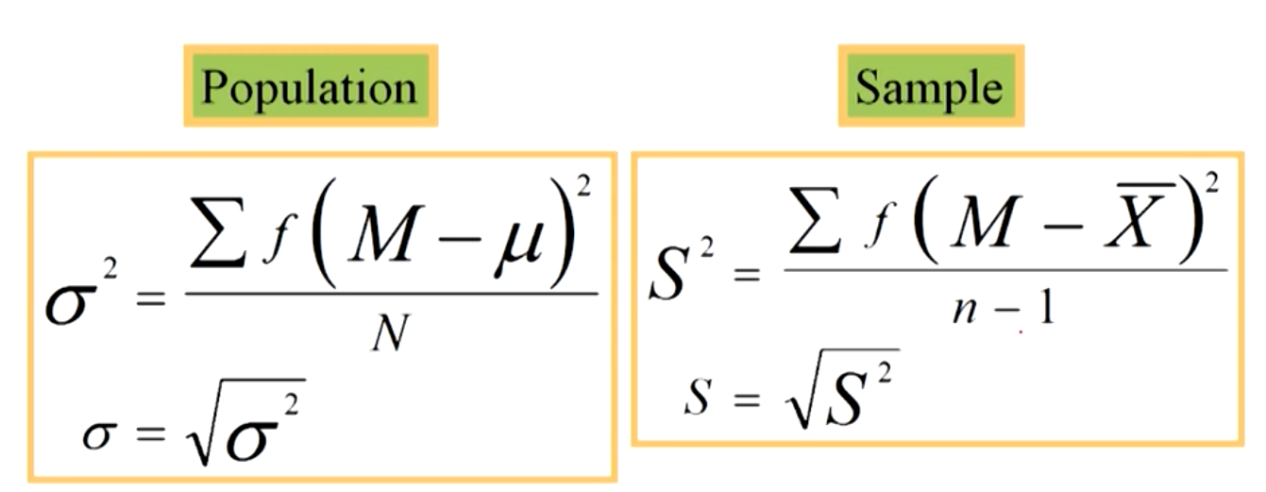


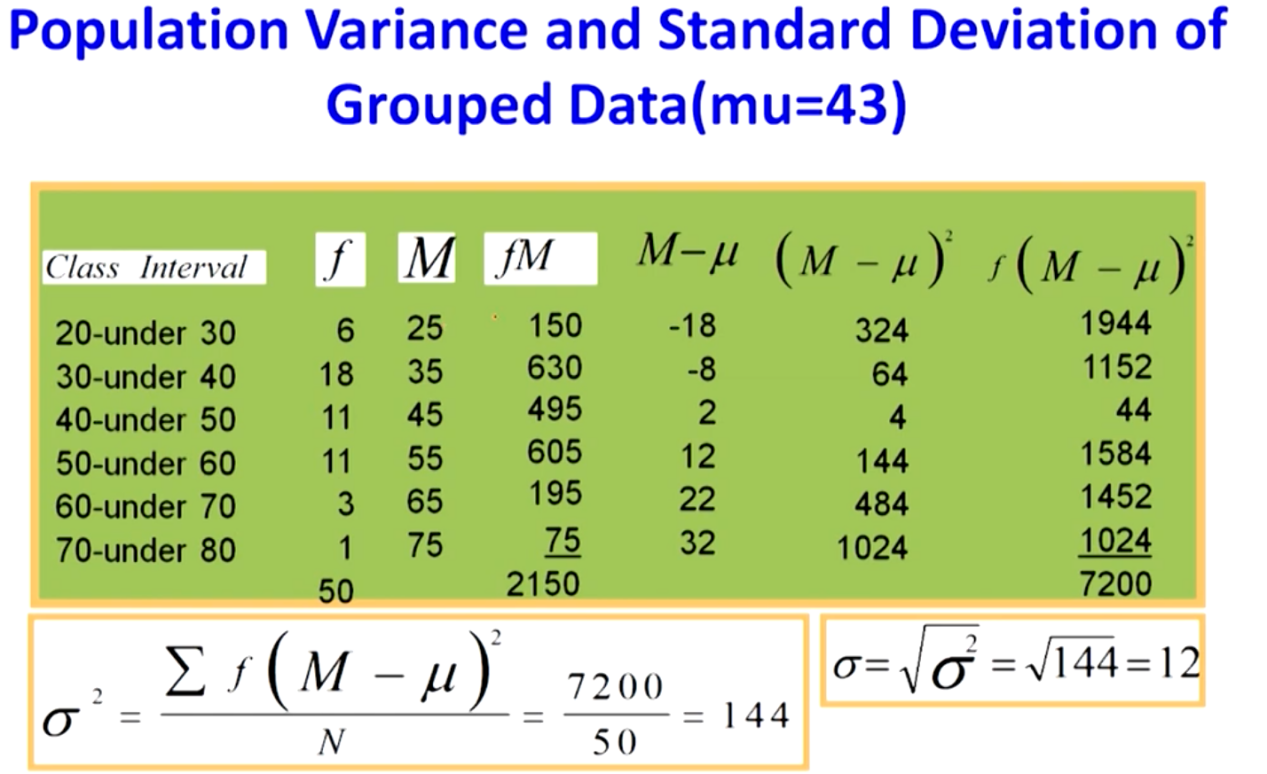
Eg. Suppose there are two stocks 1&2. We want to chose which is better.



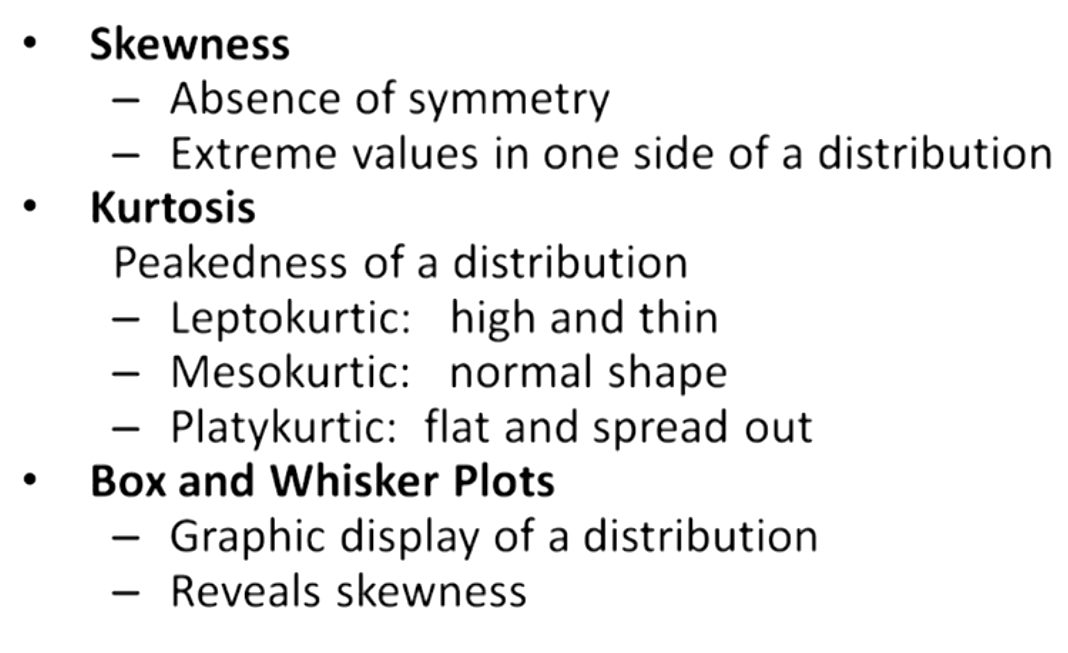
Here mean of first is greater but its standard deviation is less. So which is the right option. In this scenerio we take out the coefficient of variance to find out the better option. **The lesser the value of coefficient of variance the better the result.**

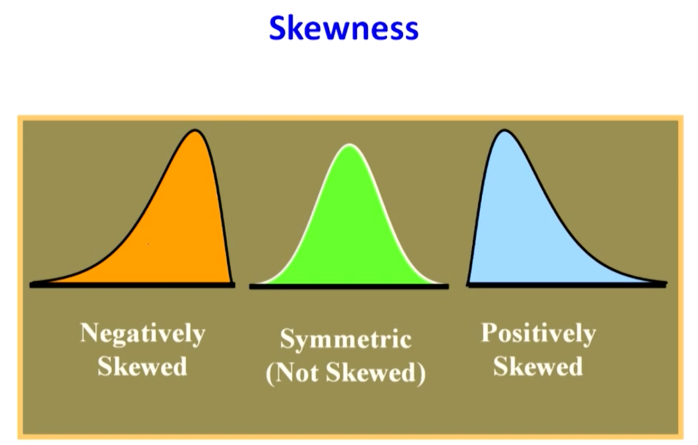
1. Find the variance and the standard deviation of the grouped data.

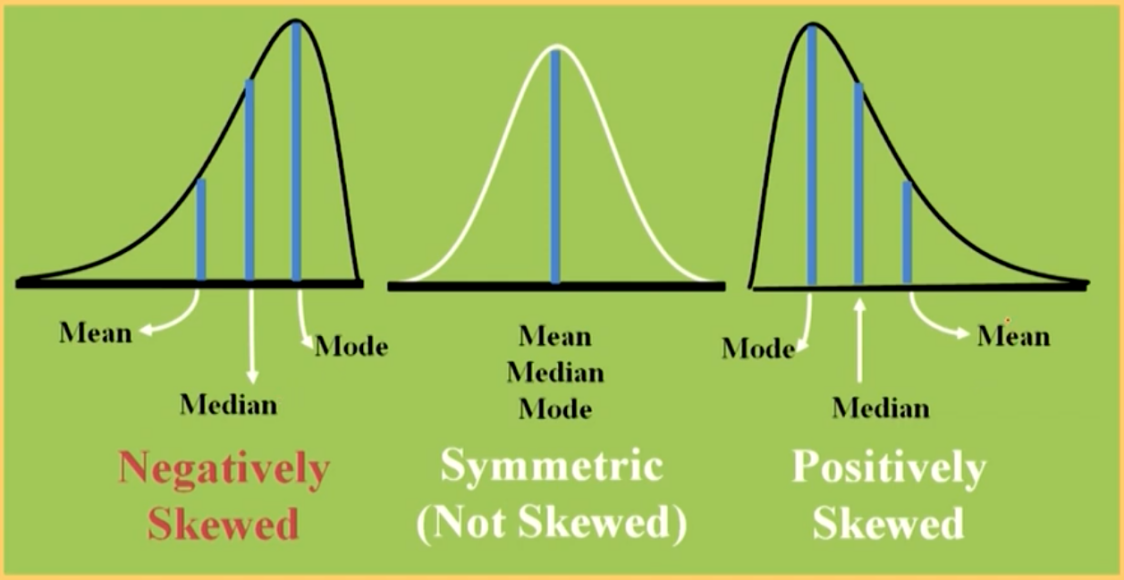
****

****

1. **What is the meaning of skewness, Kurtosis, Box and wisker plots.**

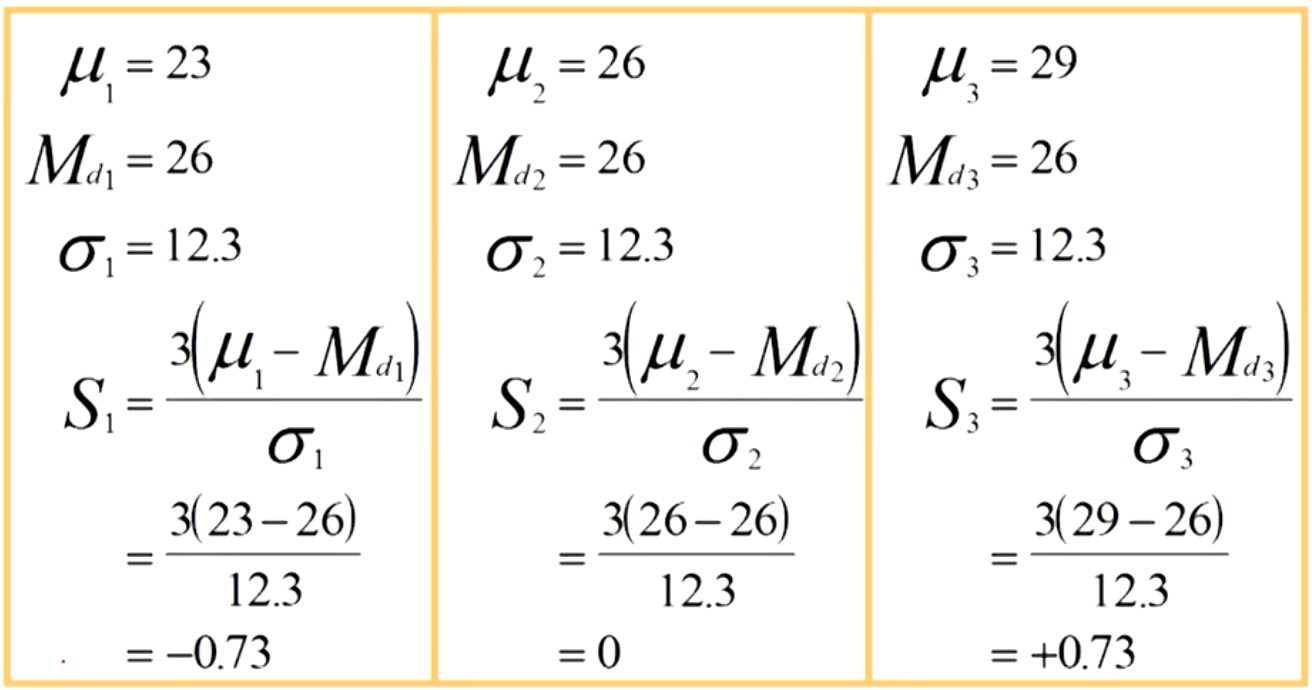
****

****

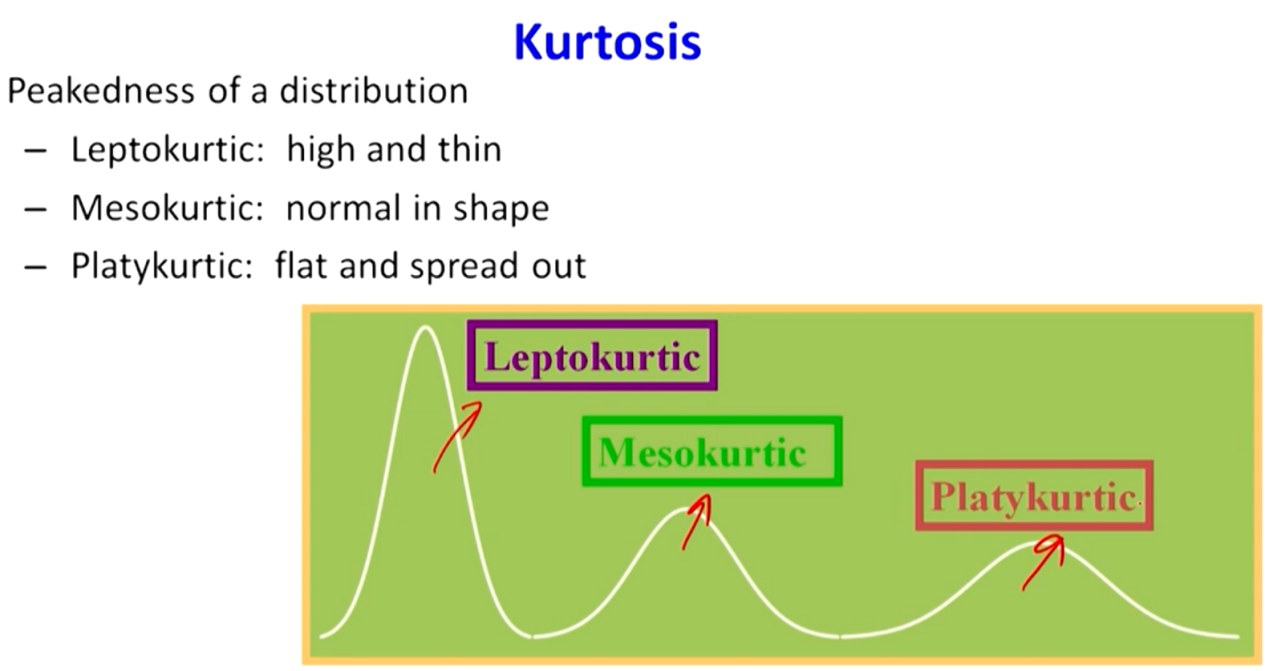
****

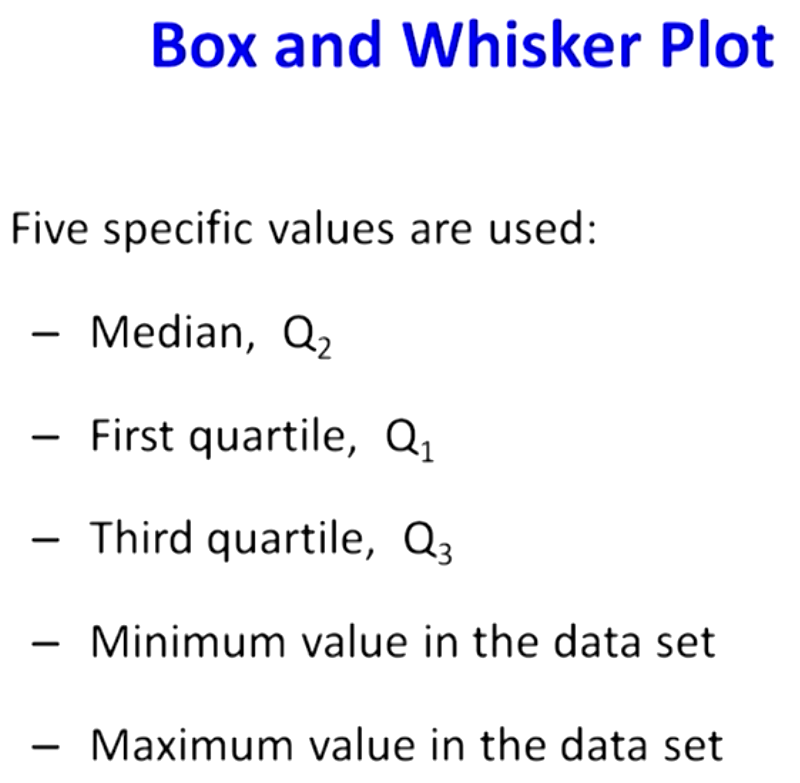
1. What is the use of coeff of skewness. Give example

Coeff of skewness to calculate if the distribution is +vely, -vely, or not skewed.



Here Md1 is median 1.

****

****