RAMA UNIVERSITY

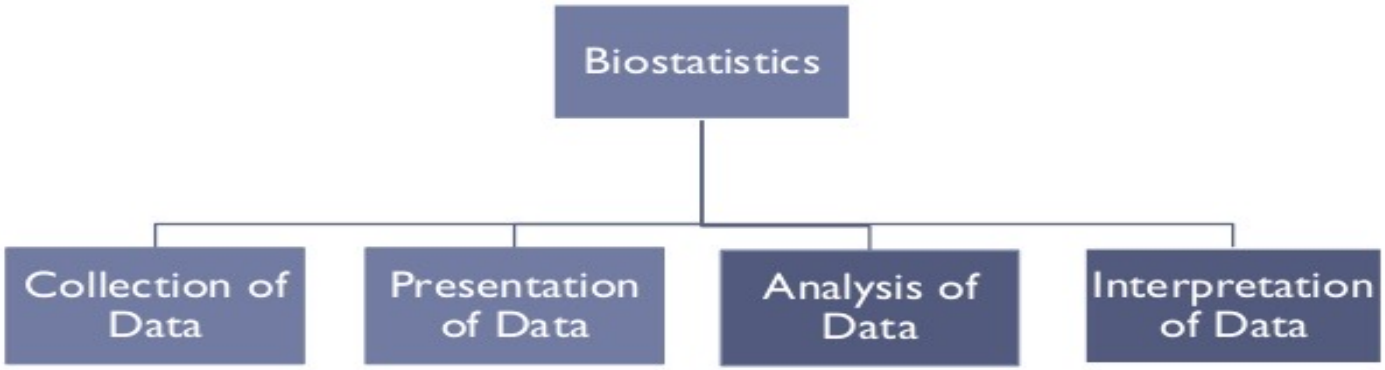
www.ramauniversity.ac.in

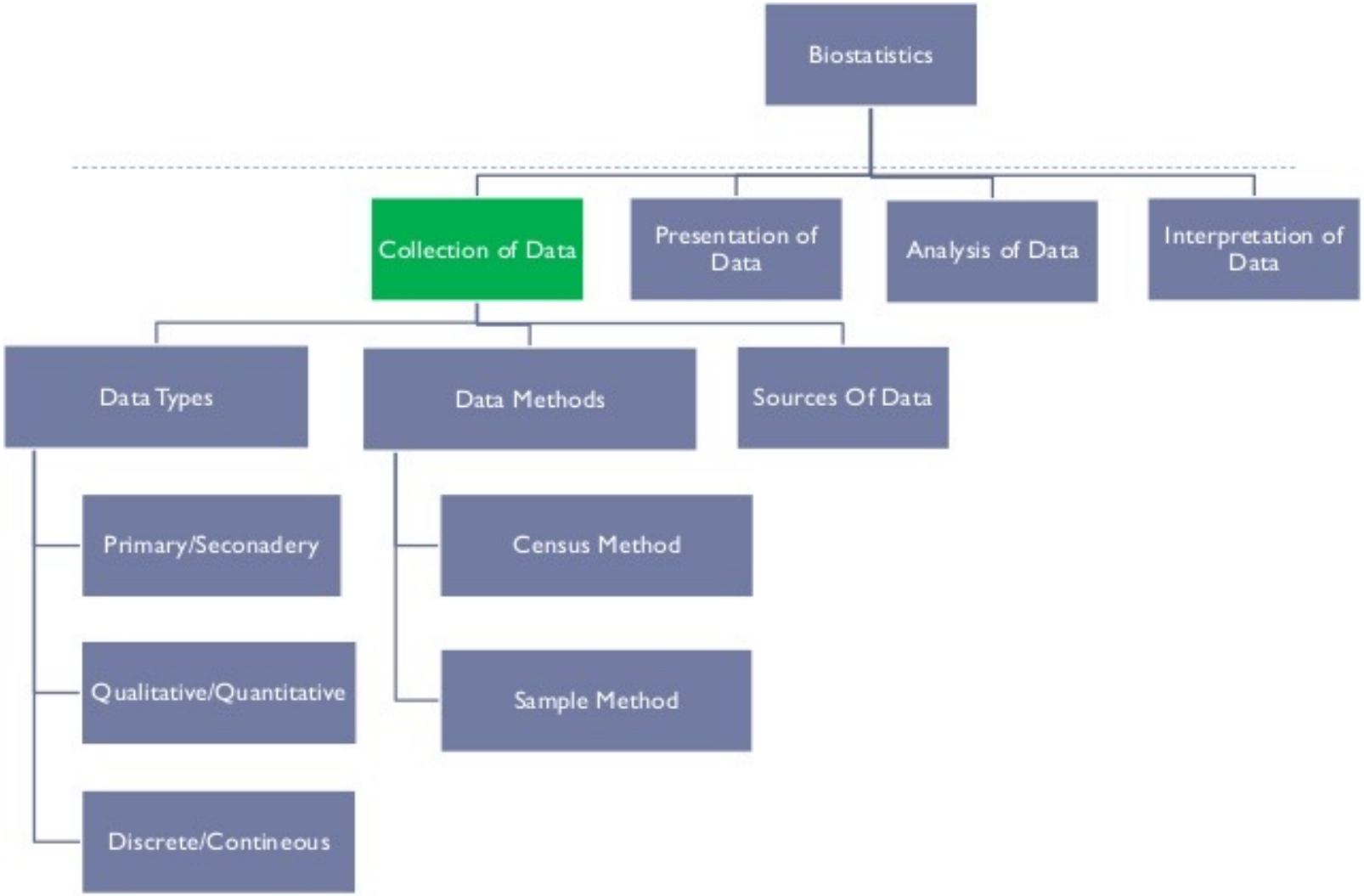
FACULTY of Engineering & Technology

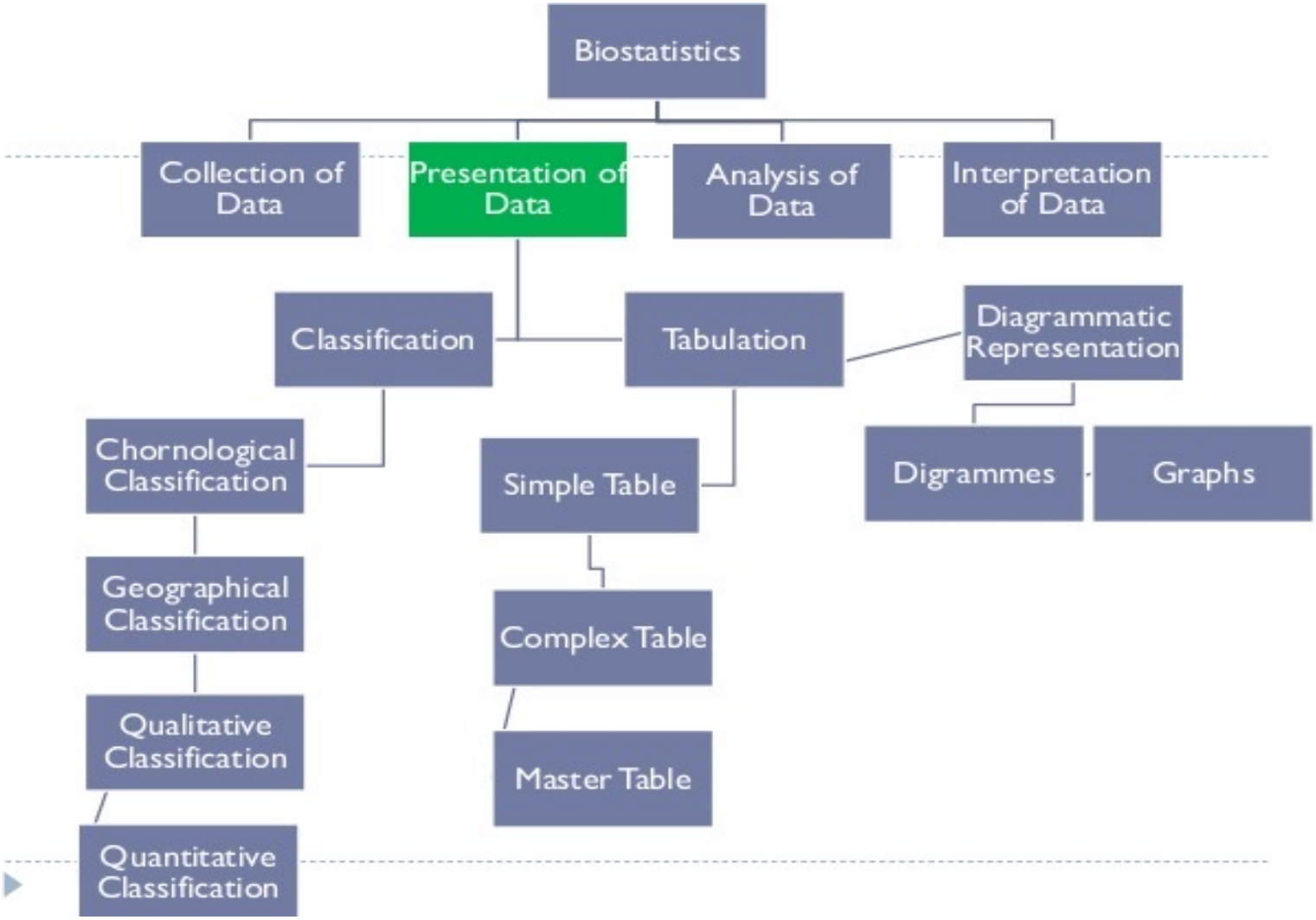
Biostatistics BTPE-103

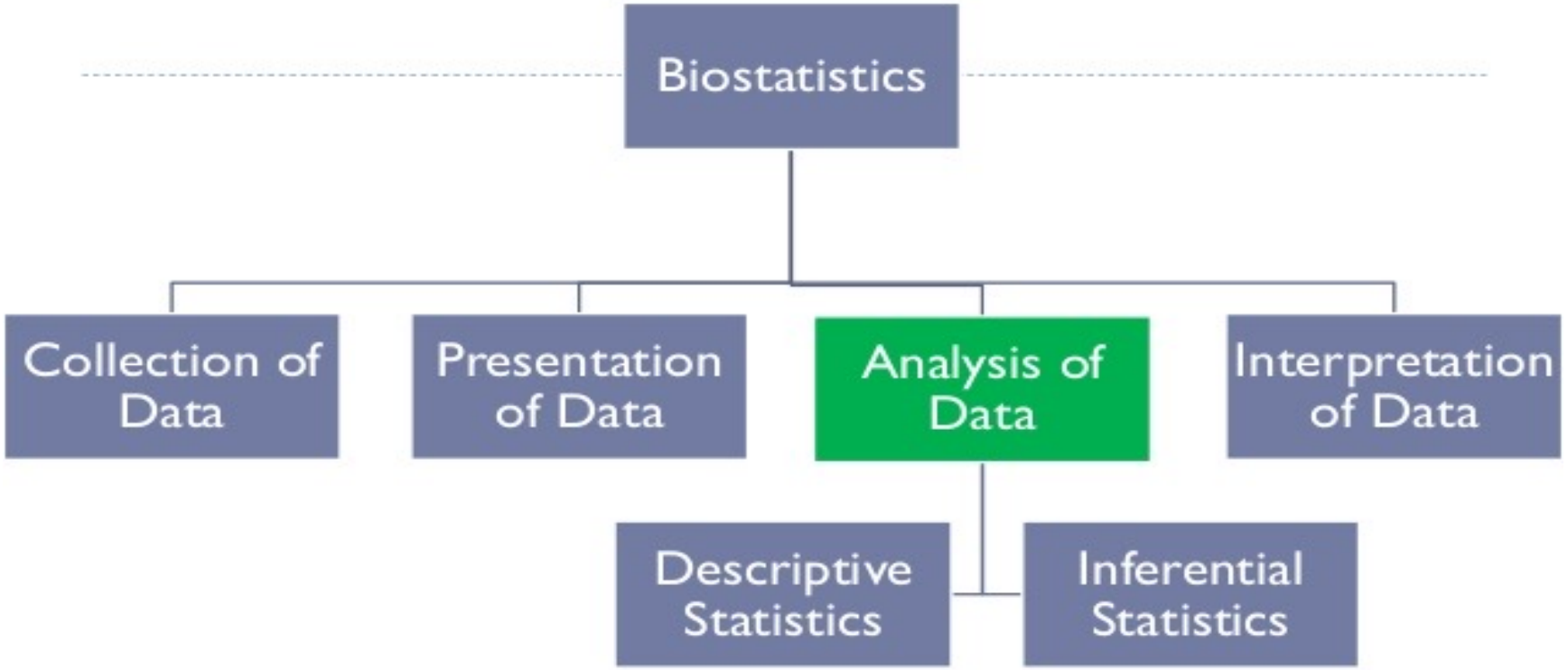
Submitted by-
Namrata Singh
Assistant Professor,
Department of Biotechnology

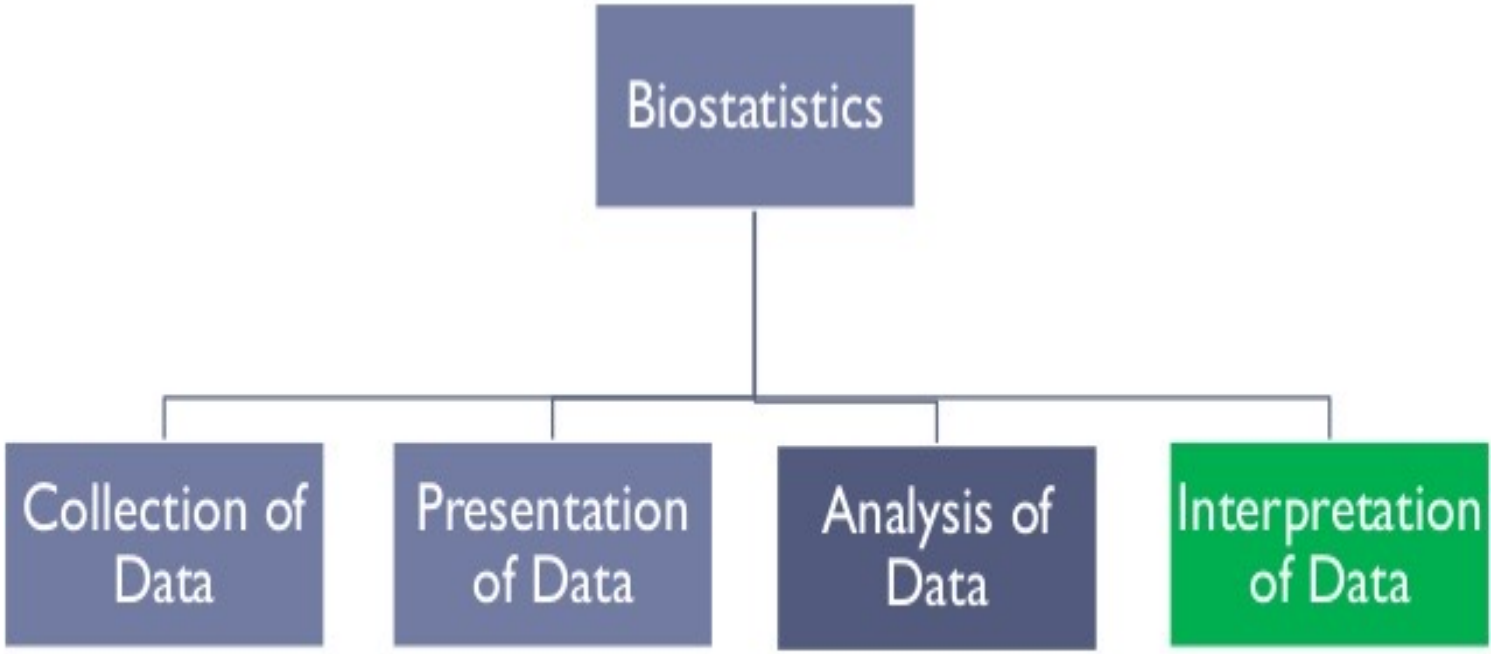
Unit-1


Biostatistics
Collection of Data
Presentation of Data
Analysis of Data
Interpretation of Data


Biostatistics
Collection of Data
Presentation of Data
Analysis of Data
Interpretation of Data
Data Types
Data Methods
Sources Of Data
Primary/Seconadery
Qualitative/Quantitative
Discrete/Contineous
Census Method
Sample Method


Biostatistics
Collection of Data
Presentation of Data
Analysis of Data
Interpretation of Data
Classification
Tabulation
Diagrammatic Representation
Chornological Classification
Simple Table
Digrammes
Graphs
Geographical Classification
Complex Table
Qualitative Classification
Master Table
Quantitative Classification


Biostatistics
Collection of Data
Presentation of Data
Analysis of Data
Interpretation of Data
Descriptive Statistics
Inferential Statistics


Biostatistics
Collection of Data
Presentation of Data
Analysis of Data
Interpretation of Data

MEASURES OF CENTRAL TENDENCY

- Definition:
  - In statistics, is a central value or a typical value for a probability distribution.
  - It is occasionally called an average or just the center of the distribution.

MEASURES OF CENTRAL TENDENCY

- The most common measures of central tendency are:
  - The arithmetic mean
  - The median
  - The mode

MEASURES OF CENTRAL TENDENCY

- Arithmetic mean (or simply, mean)
  - The sum of all measurements divided by the number of observations in the data set.
- Median
  - The middle value that separates the higher half from the lower half of the data set.
- Mode
  - The most frequent value in the data set.

Introduction

Dispersion measures how the various elements behave with regards to some sort of central tendency, usually the mean.

Most measures of dispersion have the same units as the quantity being measured.

In other words, if the measurements are in metres or seconds, so is the measure of dispersion. Such measures of dispersion include:

Range

Inter Quartile Range

CoVariance

Standard Deviation

Absolute Deviation.

Standard Deviation:

* A measure of the dispersion of a set of data from its mean. The more spread apart the data, the higher the deviation. Standard deviation is calculated as the square root of variance.

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

* Standard deviation is a statistical measurement that sheds light on historical volatility.
* For example, a volatile stock will have a high standard deviation while the deviation of a stable blue chip stock will be lower. A large dispersion tells us how much the return on the fund is deviating from the expected normal returns.

Coefficient of Variance

* The coefficient of variation represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from each other.

$$\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Expected Return}}$$

* In the investing world, the coefficient of variation allows you to determine how much volatility (risk) you are assuming in comparison to the amount of return you can expect from your investment. In simple language, the lower the ratio of standard deviation to mean return, the better your risk-return tradeoff.

Range and Inter Quartile Range

Range:

- Range is defined simply as the difference between the maximum and minimum observations.

$$\text{Range} = (\text{Highest Score} - \text{Lowest Score})$$

- Range is quite a useful indication of how spread out the data is, but it has some serious limitations. This is because sometimes data can have outliers that are widely off the other data points. In these cases, the range might not give a true indication of the spread of data.

Inter-quartile Range

•The **interquartile range** (IQR) is a measure of variability, based on dividing a data set into quartiles.

•Quartiles divide a rank-ordered data set into four equal parts. The values that divide each part are called the first, second, and third quartiles; and they are denoted by Q1, Q2, and Q3, respectively.

❖Q1 is the "middle" value in the *first* half of the rank-ordered data set.

❖Q2 is the median value in the set.

❖Q3 is the "middle" value in the *second* half of the rank-ordered data set.

The interquartile range is equal to Q3 minus Q1.

$\text{IQR} = Q_3 - Q_1$

SKEWNESS

http://pages.intnet.mu/cueboy

Skewness

Skewness is a measure of symmetry – it determines whether there is a concentration of observations somewhere in particular in a distribution.

© Rajesh Gunesh – Jul 2008

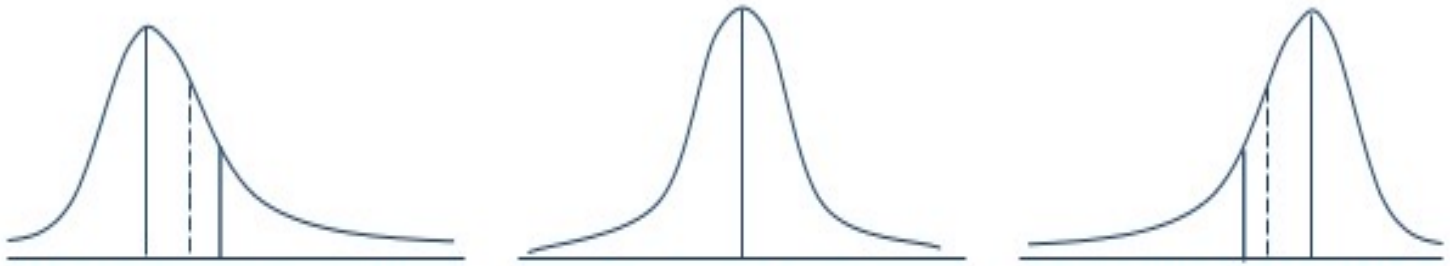
MEASURES OF SKEWNESS

http://pages.intnet.mu/cueboy

© Rajesh Gunesh – Jul 2008

Skewness

Skewness is a measure of **symmetry** – it determines whether there is a concentration of observations somewhere in particular in a distribution. If most observations lie at the **lower end** of the distribution, the distribution is said to be **positively skewed**. If the concentration of observations is towards the *upper end* of the distribution, then it is said to display **negative skewness**. A **symmetrical** distribution is said to have zero skewness.

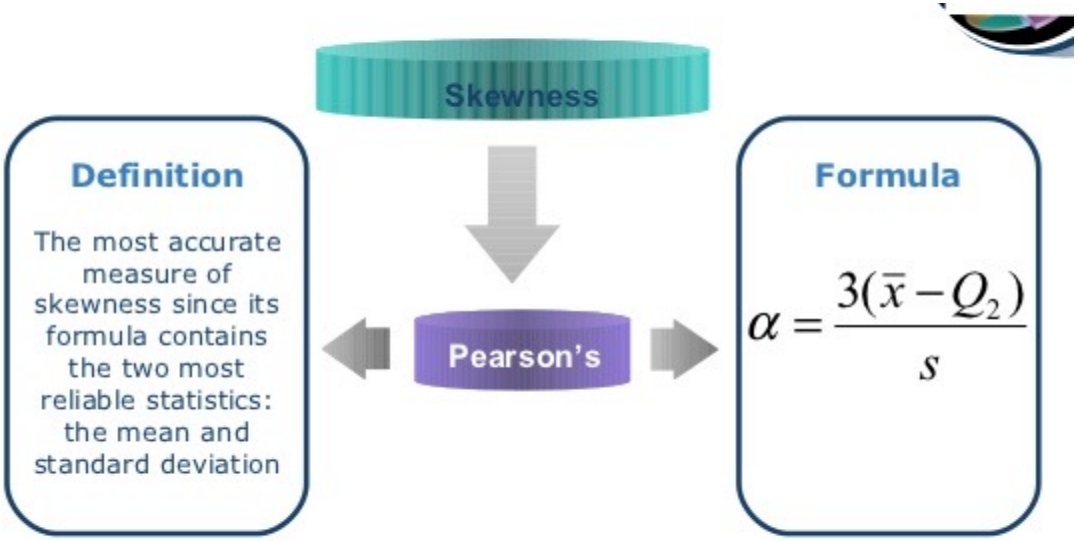


- The vertical bars on each diagram indicate the respective positions of the mean (bold), median (dashed) and mode (normal). In the case of a symmetrical distribution, the mean, median and mode are all equal in values (for example, the normal distribution).

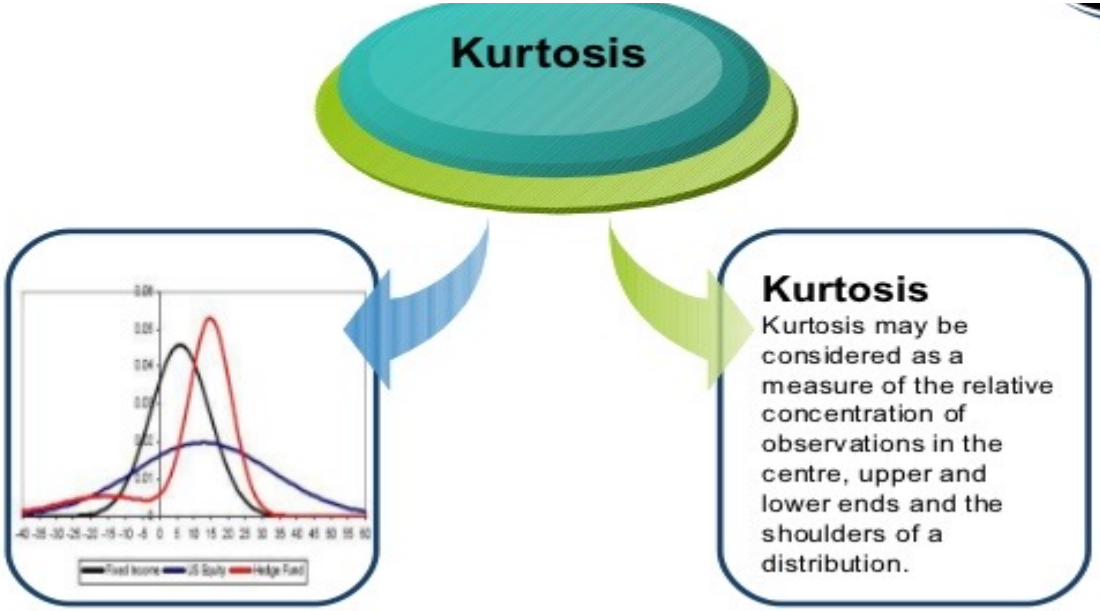
Skewness –Pearsonn Coefficient

This is the most accurate measure of skewness since its formula contains two of the most reliable statistics, the **mean** and **standard deviation**.

Pearson coefficient



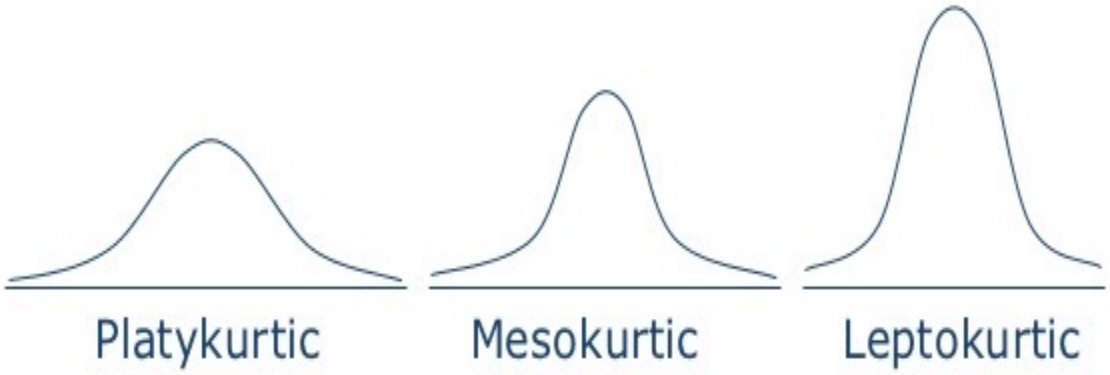
Kurtosis


Kurtosis
Kurtosis
Kurtosis may be considered as a measure of the relative concentration of observations in the centre, upper and lower ends and the shoulders of a distribution.

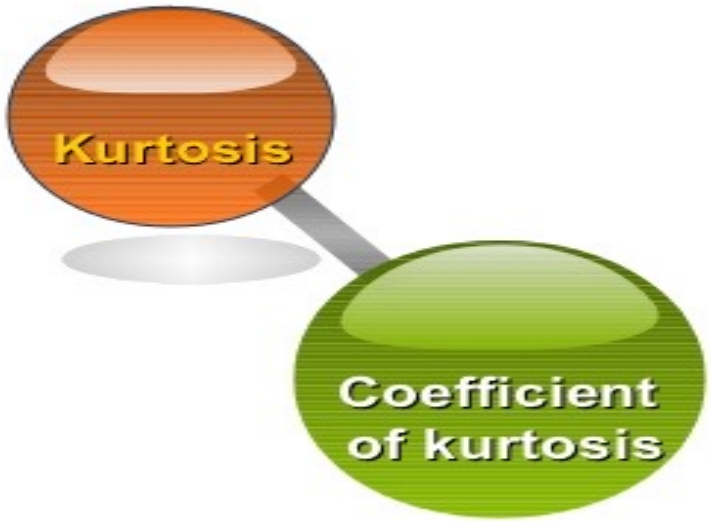
Kurtosis

It is customary to subtract 3 from the coefficient of kurtosis for the sake of reference to the normal distribution. A negative value would indicate a **platykurtic** curve whereas a positive coefficient of kurtosis indicates a **leptokurtic** distribution. A value close to 0 means that the distribution is **mesokurtic**, that is, close to the normal.

Kurtosis



© Rajesh Gunesh – Jul 2008

A MEASURE OF KURTOSIS

http://pages.intnet.mu/cueboy