## Analysis

| | | | |
|---|---|---|---|
| **Data Analysis** means examinig, sorting, categorising, comparing and evaluating the coded data | **Types of Analysis** | **Descriptive Statistics**-techniques used to summarize and display numerical data.This provides a general understanding of trends in the data | These include: central tendency, variability skewness, ANOVA/MANOVA, correlation, regression, canonical analysis |
| **Processing of Data**<br>Preparing data for analysis | Editing-processing raw data, detect errors<br>Coding-assigning symbols to responses to be put into categories<br>Classification-reduce to groups<br>Tabulation-arranging in logical order<br>Using percentages | **Inferential Statistics** or Statistical Analysis- draw conclusions, generalize results from sample to the population, find meaningful relationship from data, and reduce possibility of error | These include: hypothesis testing (parametric, nonparametric tests), estimaition of parameter values |

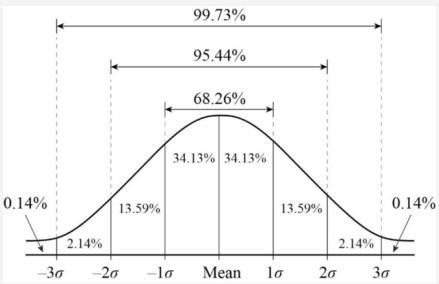
## Measures of Central Tendency

| | |
|---|---|
| **Mean** | arithmetic average of distribution of numbers |
| **Median** | middle score in an ordered distribution |
| **Mode** | most frequently occuring score ina distribution |

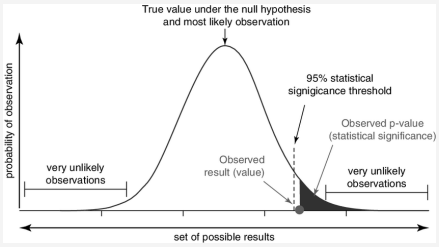
## Distribution of Data

| | |
|---|---|
| **Normal Probability Curve (NPC/NDC)** | special type of density curve that is bell shaped<br>describes tendency of most data to normally cluster around the middle |
| **Skewness** | non symmetrical data<br>collection of data on either side of the curve |
| **Kurtosis** | peaked or flat distribution of data |

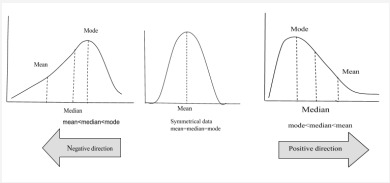
## NPC



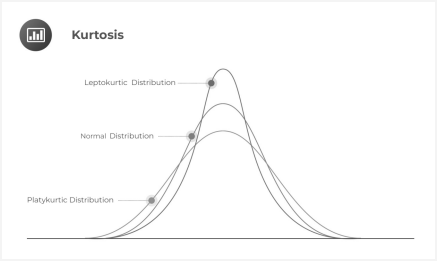

## NPC.1



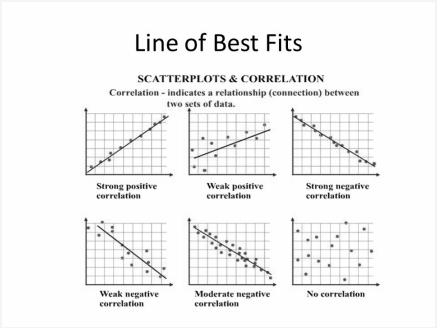

## skewness




## kurtosis




## correlation




## Measures of Relationship

**Univariate** (one variable( **Bivariate** (two variables) **Multivariate** (more than two variables)

## Measures of Relationship (cont)

**one way ANOVA**- analysis of variance which is one directional, x - y

**Index Number** - measure of relative change in magnitude of a variable (change in price of commodity in the span of a year)

**Time series analysis** - observationof a phenomenon over a period of time (trend analysis)

**Simple Correlation**- determine the strength and direction of relationship between two variables

**Simple Regression**-study cause and effect relationship, determination of statistical relationship between two/more variables, used for prediction of future values

Two way ANOVA

**Multiple Regression and Multiple correlation**

**Multiple discriminant analysis**-tech to distinguish datasets obf particular charac

**MANOVA**

**Canonical Analysis**-determining relationship between two sets of variables simultaneously

*The strength of the relationship will always range between +1.00 and -1.00 If the number is closer to +1.00 or -1.00, it indicates a strong correlation between the variable. The closer the number is to 0, the weaker the relationship becomes*

bivariate contd.

**Coefficient of Association**- indicates strength of relationship between variables

**Coefficient of Contingency**- indicates whether the IV and DV are dependent or independent of each other

mulitvariate contd.

**Factor Analysis**-data reduction system

**Cluster Analysis**-used to classify objects into groups where objects in one group are more similar to each other and different from objects in other groups

*correlation does not prove causation*

*correlation can be studied through: Charles Spearman's coefficient OR Karl Pearson's coefficient*

## Statistical Significance

Used to determine whether the differences in the data set are significant or not, i.e whether the differences are real and not caused due to random variations of the experiment. It gives us a probability that the results were caused by chance and not by experimental manipulation

**Type I error**-we accept Ho when it is false

**Type II error**- we reject Ho when it is true

Probability is denoted by ***p*** indicating the difference due to chance

For ex. If $p < 0.05$, it means that there is a 5 out of 100 probability of result being due to chance OR 95% certain that results were real and not due to chance

## Measures of Variability/Dispersion

| | |
|---|---|
| **Variance** | measure of how much values in a dataset differ from the mean<br>the amount of dispersion of scores |
| **Range** | difference between values of extreme items(highest and lowest scores) |

## Measures of Variability/Dispersion (cont)

| | |
|---|---|
| **Standard Deviation** | average distance between the scores and the mean OR<br>avergae squared deviations from the mean scores in a distribution |

## Inferential Statistics

**Point estimate**- a single value, best estimate of a parameter

**Interval estimate**-a range of plausible values of a parameter

**Parametric test**-specifies certain conditions about parameter of population, stronger than nonparametric tests
normally distributed data
ex. z test, t test, F test

**Nonparametric tests**-does not specify any conditions, *distribution free statistics*, data does not fall under NPC
ex. Man Whitney U test, Kendall's tau, chi-square test