
msn
India
WELCOME

TEST OF SIGNIFICANCE
&
ITS APPLICATIONS

PRESENTED

CONTENTS (PART-I)

- Statistics, Biostatistics
- Variable
- Types of Variables
- Frequency Distribution
- Parameters Of Frequency Distribution
- Normal Distribution
- Research, Designing a research protocol
- Hypothesis, Process of hypothesis testing

CONTENTS (PART-II)

- Tests of Significance

  Parametric Tests

- t- test
- Chi-square test
- F-test
- ANOVA
- Coefficient of Correlation & Regression
- Non Parametric tests

STATISTICS

- *The word **‘ Statistics’** has come from an Italian word ‘ Statista’ which means a number that describes a property of a set of data.*

- ***Statistics*** *is the ‘ Science which deals with collection, organization, summarization, analysis, interpretation & presentation of data.’*

- ***Biostatistics*** *implies‘ Statistical methods & techniques applied to biological problems.’*

JOHN GRAUNT

(1620-1674)

Father of Health Statistics

DIFFERENCE BETWEEN MATHEMATICS & STATISTICS

$$y = mx + b$$

- *Here, m & b are constant.*
- *Y intercept, i.e. the amount of change in y for a unit increase in value of x.*
- *In mathematics, one of the variable either x/y is unknown , while the formula & constant are known.*
- *In statistics, reverse is true, i.e. the variables, x& y are known for all observations & investigator wishes to know the relationship between the variables.*

USE OF STATISTICS (DENTISTRY)

- *To assess the state of oral health in the community .*
- *To determine the availability & utilization of dental care.*
- *To determine success / failure of specific oral health care programmes / to evaluate the programme action.*

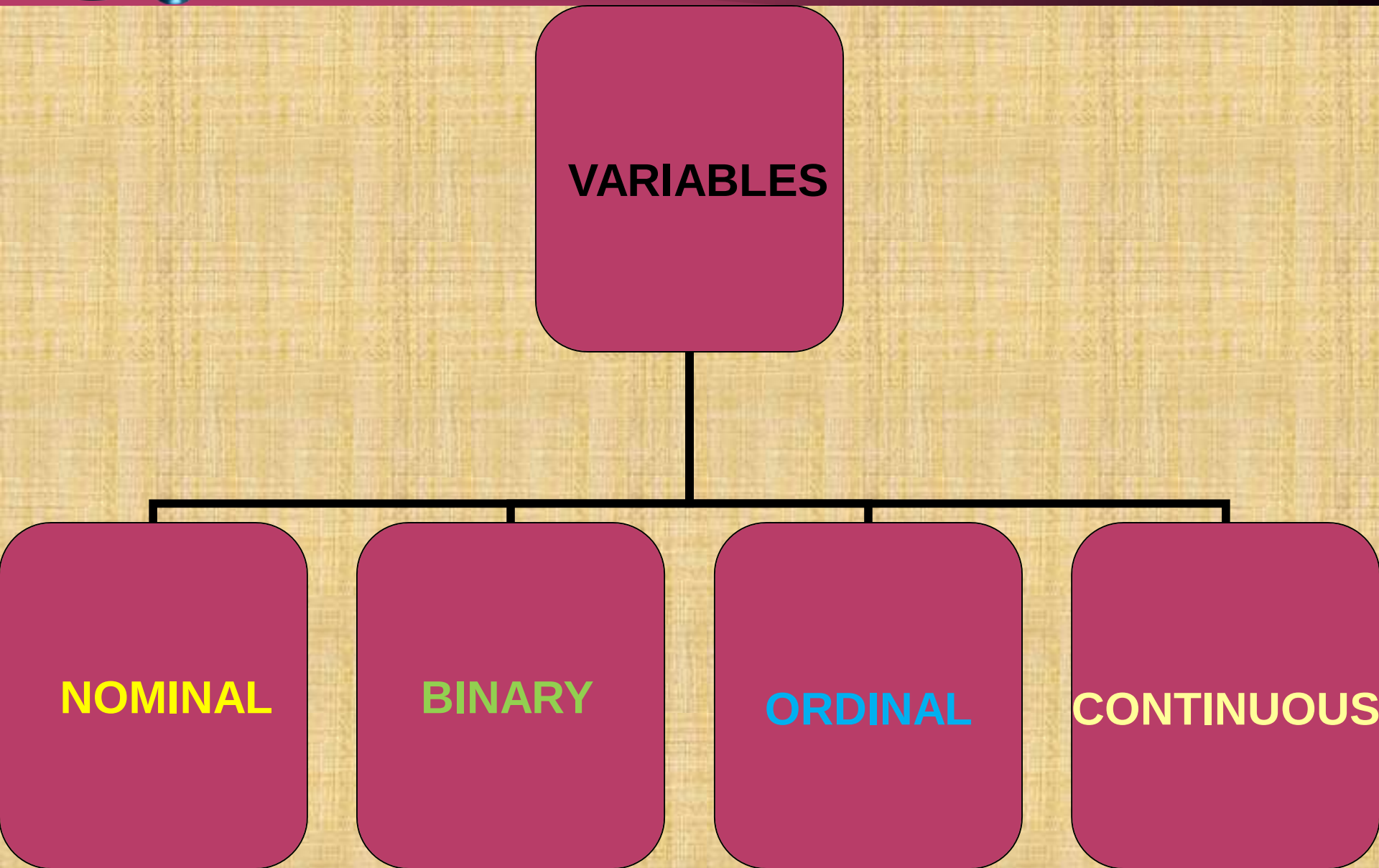
VARIABLE

- *Measure of a single characteristic.*

- *Statistics enable the Investigators to :*

1. *Describe the pattern of variation in single variables.*
2. *Describe when observed differences are real differences.*
3. *Determine the pattern & strength of association between variables.*

TYPES OF VARIABLES



NOMINAL VARIABLES

- *They are 'naming' / 'categorical' variables that have no measurement scale.*

- *E.g. Blood groups: O, A, B & AB*

DICHOTOMOUS VARIABLES

- *The word Dichotomous has come from a Greek word meaning 'cut into two'.*
- *Thus these are those variables which has only 2 levels.*
- *E.g. Well/ sick, living/ dead, normal/ abnormal, etc.*

- *Binary variables & Nominal variables are called as 'Discrete variables' because the different categories are completely separate from each other.*

ORDINAL VARIABLES

- *Many types of medical data can be characterised in terms of more than 2 values & have a clear implied direction from better to worse, but the data are not measured on a measurement scale.*

- *E.g. The amount of swelling in a patient's leg is estimated by the clinician & is usually reported as none/ +1, +2, +3,+4 pitting edema*

CONTINUOUS VARIABLES

- *Medical important data, which are measured on continuous (dimensional) measurement scales.*

- *E.g. patient's height, weight, systolic & diastolic blood pressure & serum glucose levels.*

RATIO VARIABLES

- *If a continuous scale has a true 0 point, variables derived from it are called as ratio variables.*
- *E.g. Kelvin temperature scale is a ratio scale, because 0 degrees on this scale is absolute 0.*

- *Centigrade temperature scale is a continuous scale but not a ratio scale, because 0 degrees on this scale does not mean absence of heat.*

UNITS OF OBSERVATION

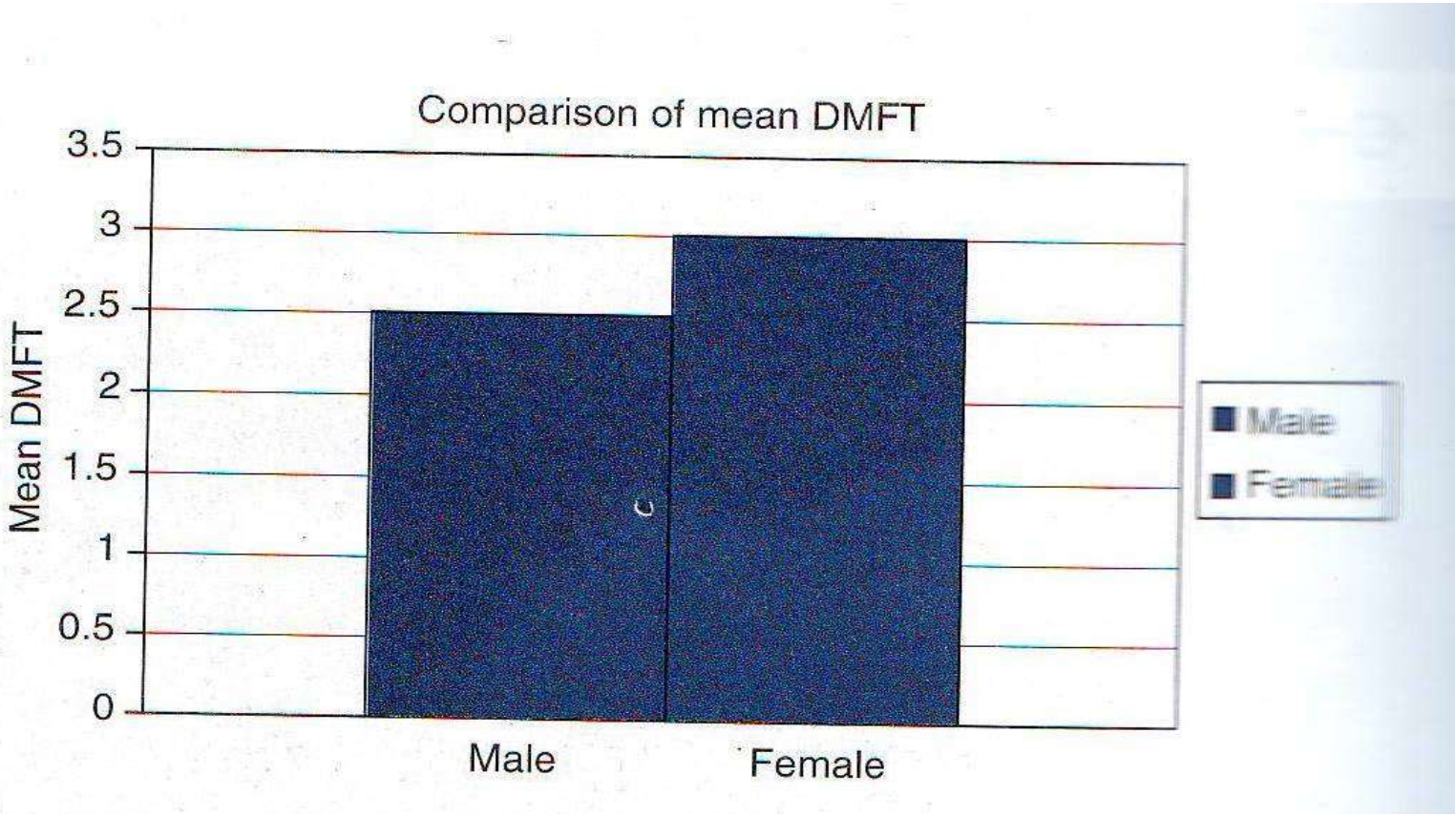
- *The units of observation is the person/ thing from which the data came.*

- *E.g. Units of observation in medical studies are persons, animals & cells etc.*

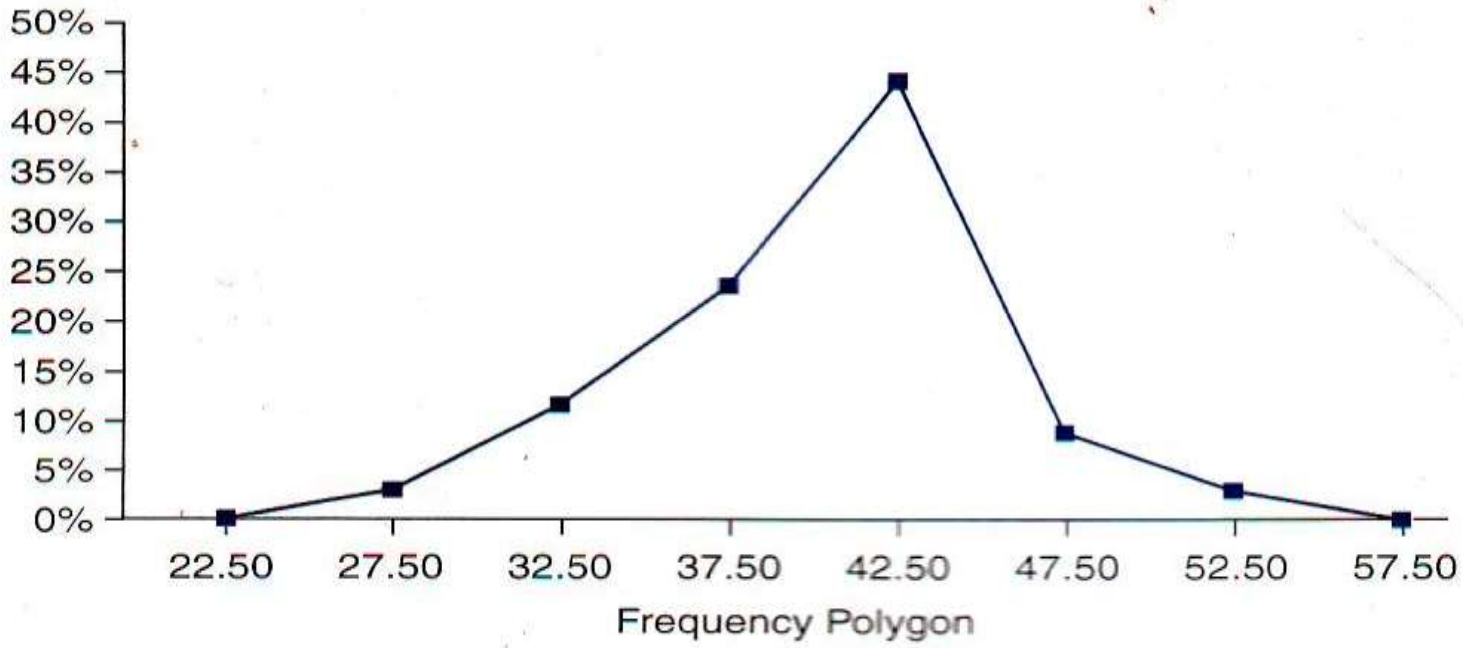
FREQUENCY DISTRIBUTIONS

- *Observations of one variable may be shown visually by putting the value on one axis (usually the horizontal axis) & putting the frequency with which it appears on the other axis (usually the vertical / y-axis).*

BAR CHART


Comparison of mean DMFT
Mean DMFT
3.5
3
2.5
2
1.5
1
0.5
0
Male
Female
Male
Female

FREQUENCY POLYGON



RANGE OF A VARIABLE

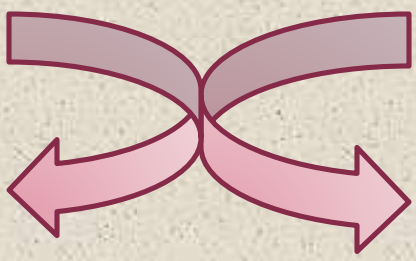
- *Distance between the lowest & highest observation is called the range of variable.*

REAL & THEORETICAL FREQUENCY DISTRIBUTION

- *Real Frequency distributions are those obtained from the actual data & theoretical frequency distributions are calculated using certain assumptions.*

PARAMETERS OF FREQUENCY DISTRIBUTION

- *Frequency Distributions are defined by 2 types of parameters:*



MEASURES OF CENTRAL TENDENCY

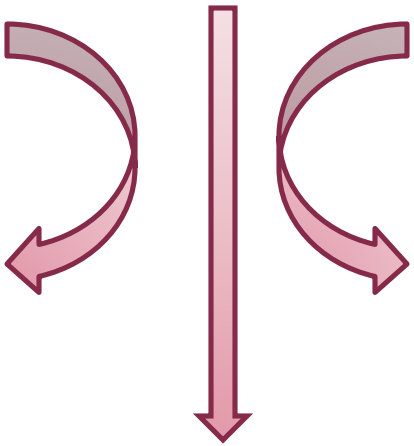
MEASURES OF DISPERSION

MEASURES OF CENTRAL TENDENCY

- *They are the values, which lie between 2 extreme observations(i.e., the smallest & the largest observations), of the distribution & give us an idea about the concentration of the values in the central part of the distribution.*

- *Most types of medical data tend to clump in such a way that the density of observed values is greatest near the centre of the distribution.*

MEASURES OF CENTRAL TENDENCY



MEAN

MODE

MEDIAN

REQUISITES OF MEASURE OF CENTRAL TENDENCY

- *It should be rigidly defined i.e. it should be clear & unambiguous so that it leads to one & only one interpretation by different persons.*

- *It should be based on all observations .i.e. in its computation, the entire set of data at our disposal should be used & there should not be any loss of information resulting from not using the available data.*

- *It should be least affected by fluctuations of sampling.*

MEAN

- *It is the average value / the sum of all the observed values divided by the total number of observations (N):*

$$Mean = \bar{x} = \frac{Sum\ (xi)}{N}$$

Here, (i) means, for each individual observation

ADVANTAGES OF MEAN

- *It is rigidly defined & simple to calculate.*
- *It is based on all observations.*
- *Mean of a sample is an unbiased estimator of the mean of population from which it came.*

DISADVANTAGES OF MEAN

- *Main disadvantage is that it is very much affected by extreme observations.*

- *It can't be used if we are dealing with qualitative characteristics which cant be measured quantitatively such as intelligence, honesty, etc.*

- *It can't be obtained if a single observation is missing/ lost unless we drop it out & compute the mean of remaining observations.*

MEDIAN

- *It is the middle observation when data have been arranged in order from lowest to highest value.*

- *According to L.R.Connor, ‘ The median is that value of the variable which divides the group in two equal parts , one part comprising all the values greater & the other, all the values less than median.*

CALCULATION OF MEDIAN

- *If the number of observations is odd, then the median is the middle value after the observations have been arranged in ascending / descending order of magnitude.*

- *In case of even number of observations, median is obtained as the mean of the two middle observations after they are arranged in ascending / descending order of magnitude.*

MERITS OF MEDIAN

- *It is rigidly defined, easy to understand & easy to calculate .*
- *Since median is a positional average, it is not affected at all by extreme observations*

- *It is the only average to be used while dealing with qualitative characteristics which cannot be measured quantitatively but still can be arranged in ascending/ descending order of magnitude.*

  e.g. To find out the average intelligence, average honesty , etc among a group of people.

MODE

- *Most commonly observed value, i.e. the value that occurs most frequently in a data set.*

BIMODAL DISTRIBUTION :

- *2 Modes are close enough together to be considered part of the same central cluster.*

MERITS OF MODE

- *Mode is easy to calculate & understand.*

- *Mode is not at all affected by extreme observations & it is preferred to mean while dealing with extreme observations.*

DEMERIT OF MODE

- *Mode is not rigidly defined .It is ill defined if the maximum frequency is repeated / if the maximum frequency occur either in the very beginning / at the end of the distribution / if the distribution is irregular.*

MEASURES OF DISPERSION

DISPERSION

- *The term dispersion refers to ‘ Scatteredness’.*

- *It gives us an idea about the homogeneity (compactness) / heterogeneity (scatter) of the distribution.*

OBJECTIVES OF MEASURES OF DISPERSION

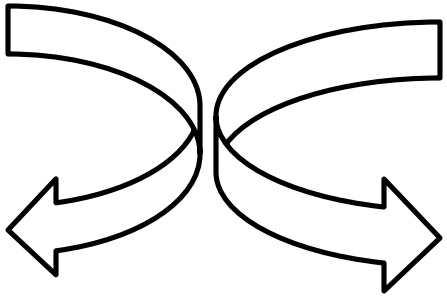
- *To find out the reliability of an average.*

  If the dispersion is small, it means the given data values are closer to the central value & hence the average may be regarded as reliable as it provides a good estimate of the corresponding population average.

- *To control the variation of the data from the central value.*

  Measure of dispersion helps us to determine the causes & nature of variation, so as to control variation.

MEASURES OF DISPERSION



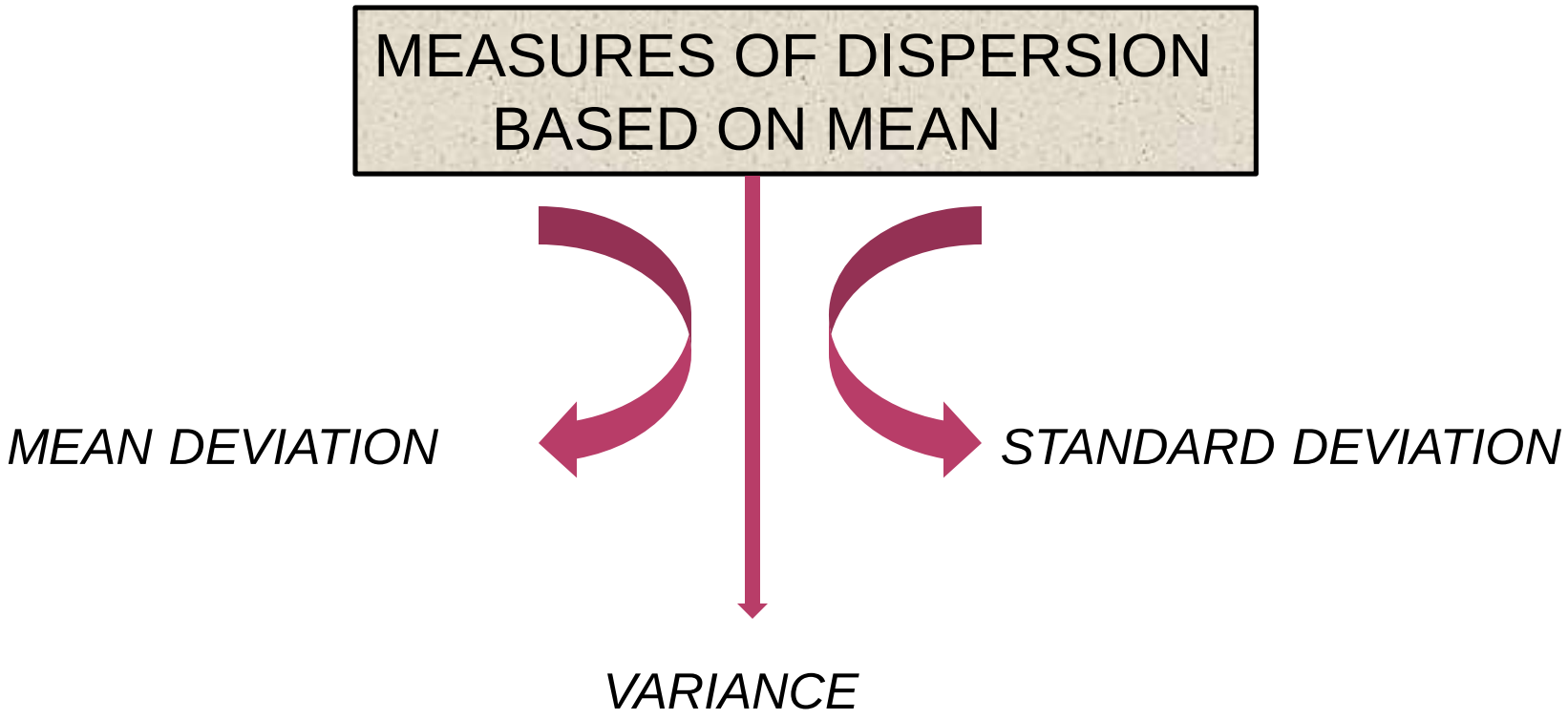
Based on Percentiles

Based on Mean

PERCENTILES

- *They are percentage of observations below the point indicated when all of the observations are ranked in descending order.*

- *75^{th} Percentile is the point below which 75% of observations lie.*


MEASURES OF DISPERSION BASED ON MEAN
MEAN DEVIATION
STANDARD DEVIATION
VARIANCE

MEAN DEVIATION

- *Average deviation from the mean.*

- *Mean Deviation = Sum* $\dfrac{\sum \left(\left| xi - \bar{x} \right| \right)}{N}$

VARIANCE

- *It is the sum of the squared deviations from the mean, divided by number of observations minus 1.*

$$VARIANCE = \frac{\sum (xi - \bar{x})\,2}{N-1}$$

- *The Numerator of the variance, i.e. the sum of the squared deviations of the observations from the mean is an important entity in statistics.*

- *In the denominator, dividing by N-1, (degree of freedom) instead of dividing by N, is necessary for the sample variance to be an unbiased estimator of the population variance.*

STANDARD DEVIATION

- *It is the square root of the variance.*

$$= \sqrt{\frac{\sum (xi - \bar{x})2}{N-1}}$$

NORMAL DISTRIBUTION

- *Normal Distribution is also called as 'Gaussian Distribution', named after Johann Karl Gauss.*

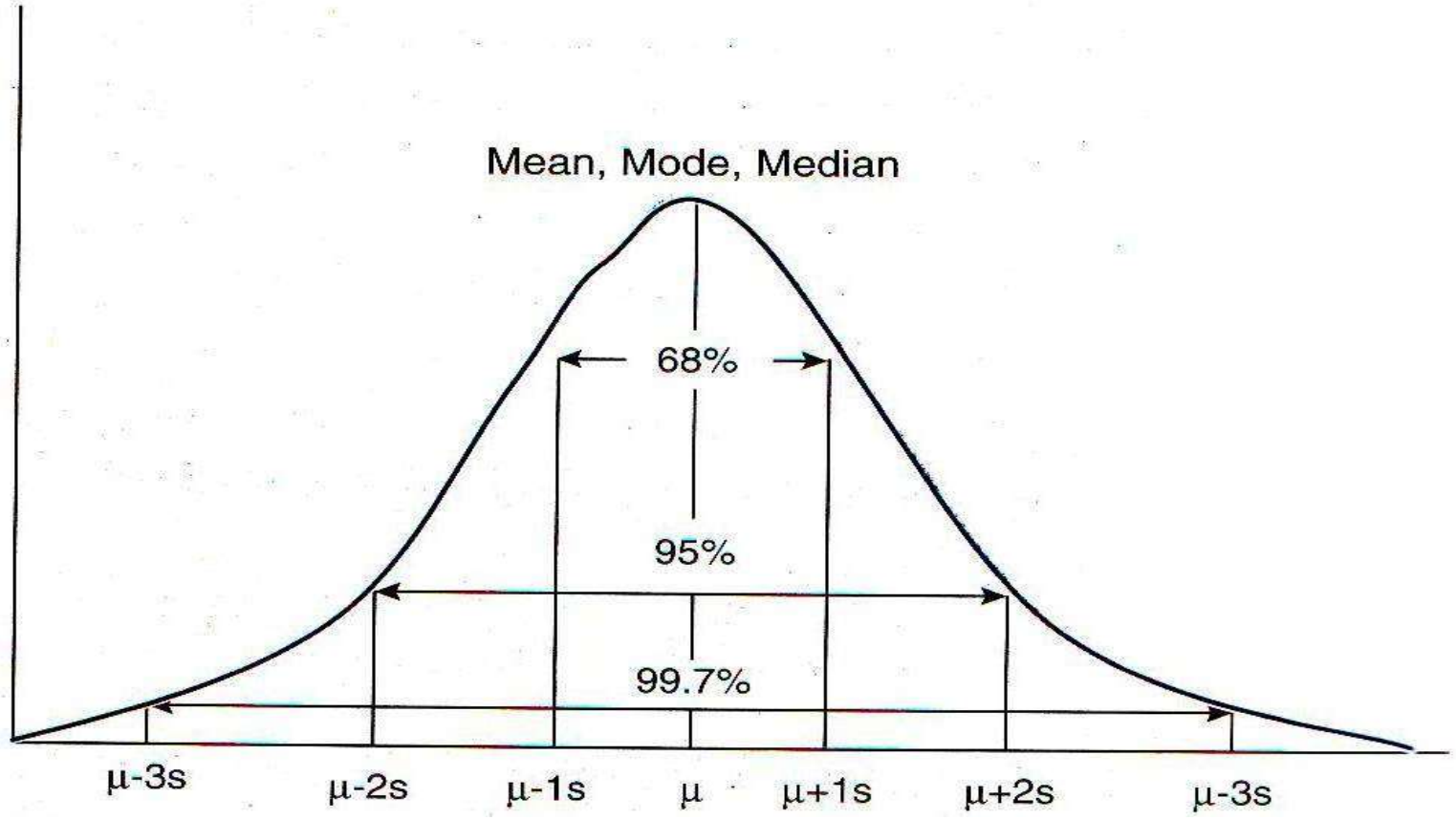
- *Normal Distribution is one of the most frequently occurring distributions in bio-medical & dental research.*

- *It is a population frequency distribution.*

PROPERTIES OF NORMAL DISTRIBUTION

- *It is characterized by a Normal Curve ,which is:*
- *Bell Shaped.*
- *Unimodal*
- *The curve is symmetric around the mean of the distribution.*
- *The height of the curve is maximum at the mean & all the three measures of central tendency i.e. mean, median & mode coincide.*
- *Tail never touches the base line theoretically.*

NORMAL CURVE

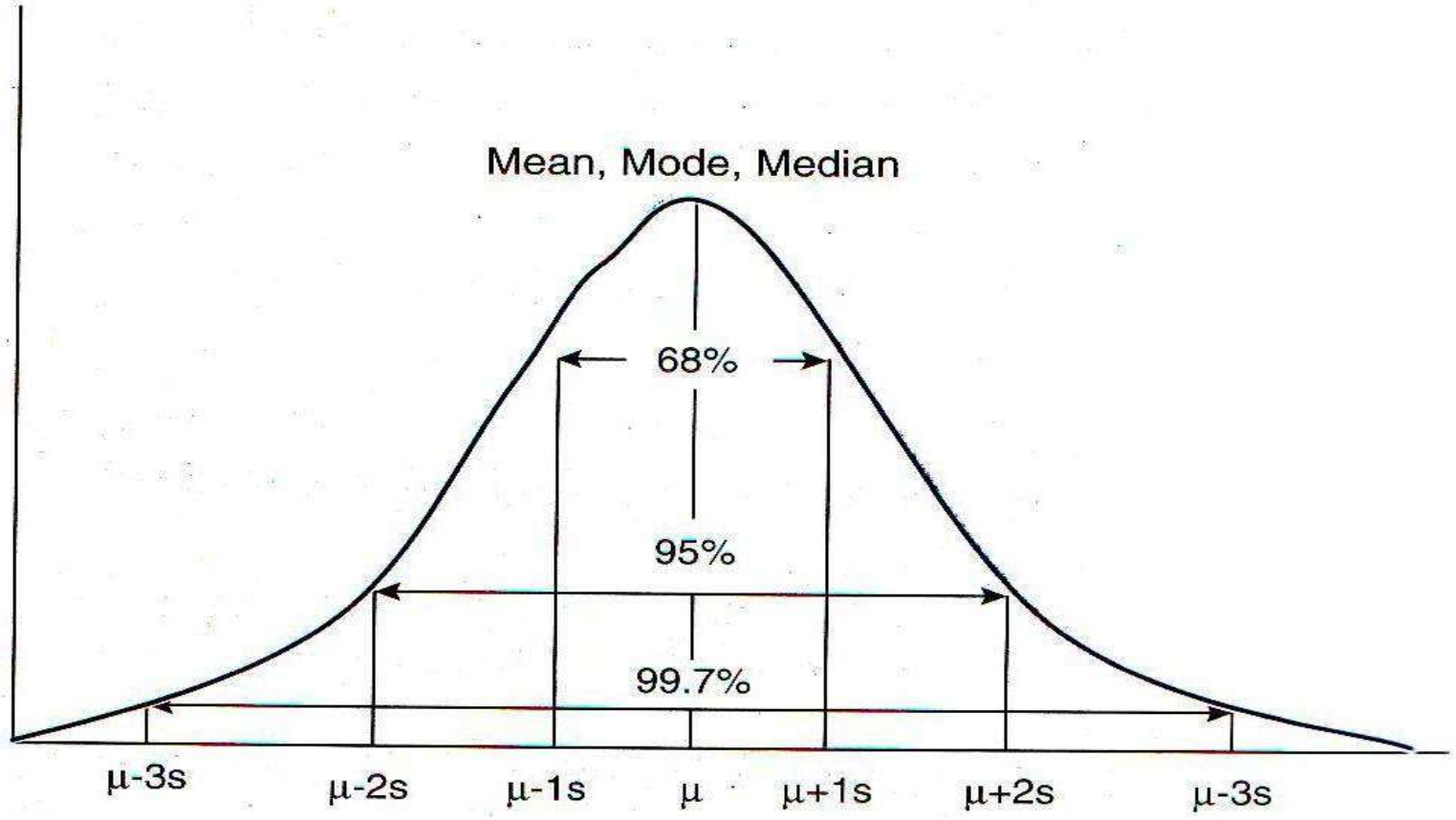


- *Normal Curve depends on only two parameters:*

- *Population Mean*
- *Population Standard Deviation.*

- *In order to know the area under the normal curve, normal distribution has been standardized to the standard normal distribution, in which population mean is 0 & population SD= 1.*
- *The area under the normal curve can be segmented starting with mean in the centre(on the x-axis) & moving by increments of 1 SD above &below the mean.*

NORMAL CURVE



- *Normal Distribution is arithmetically expressed as :*
- *Mean ± 1SD = 68.27% of observations will be covered within this limit*

.

- *Mean ±2SD = 95.45% is covered &*

- *Mean ± 3SD =99.73% is covered.*

HYPOTHESIS

- *A Hypothesis can be defined as a tentative prediction/ an assumption that is made before investigation regarding the outcome under study.*

- *It is made because it can be tested scientifically using statistical procedure.*

NULL HYPOTHESIS

- *It is the hypothesis , which states that there is no real difference between means/ proportions of the groups being compared / that there is no real association between 2 continuous variables.*

- *It is the hypothesis which states that there is no significant difference between 2 values.*

- *The difference which occurs is due to chance.*

- *E.g. Suppose a group of patients had received routine periodontal scaling& were given a new pain relieving drug, Drug A & recorded their number of minutes to 100 % pain relief.(minutes to pain relief is a continuous variable).*
- *Suppose a similar sample of patients tested another drug , Drug B, in the same way.*
- *Investigator want to know , if one group of patients experienced total pain relief more quickly than the other group.*

- *In this case, Null Hypothesis states that " There is no difference in time to 100 % pain relief between the 2 pain relieving drugs A & B.*
- *Null Hypothesis is one of No Difference, No Effect, No association.*

- *Symbolically is expressed as : Ho.*

- *NULL HYPOTHESIS*

$$Ho = \overline{X}\,A = \overline{X}\,B$$

- *ALTERNATE HYPOTHESIS*

$$H1 = \overline{X\,A} \neq \overline{X\,B}$$

ALTERNATE HYPOTHESIS

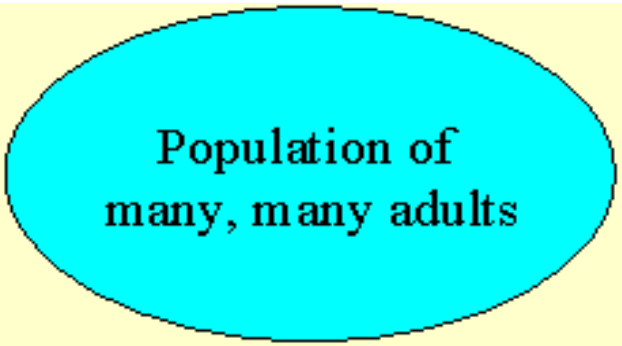
- *It is the hypothesis , which states that there is a real difference / association between means / proportions of groups being compared.*

- *It is the hypothesis which state that there is a significant difference.*

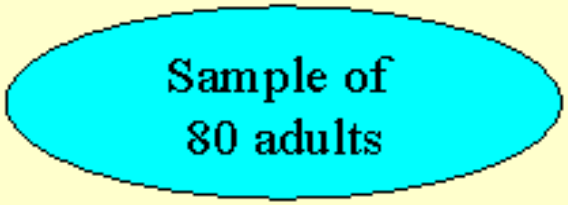
TYPE-I ERRORS
&
TYPE-II ERRORS

- **Type I error**: The null hypothesis is rejected when it is true.
- **Type II error**: The null hypothesis is not rejected when it is false.
- There is always a chance of making one of these errors. But, we will want to minimize the chance of doing so!

Decision	Truth	
	Null hypothesis	Alternative hypothesis
Do not reject null	OK	TYPE II ERROR
Reject null	TYPE I ERROR	OK



Is average adult body temperature 98.6 degrees? Or is it lower?



Average body temperature of 80 sampled adults is 98.4 degrees.

- Specify hypotheses.
  - H_0: μ = 98.6 degrees
  - H_A: μ < 98.6 degrees
- Make initial assumption: μ = 98.6 degrees
- Collect data: Average body temp of 80 sampled adults is 98.4 degrees. How likely is it that a sample of 80 adults would have an average body temp as low as 98.4 if the average body temp of population was 98.6?

- **Type I error** here is claiming that average body temp is lower than 98.6 when in fact it really isn't.
- **Type II error** here is failing to claim that the average body temp is lower than 98.6 when it is.

ALPHA LEVEL - α

- *Maximum probability of making an error that the investigator is willing to accept.*

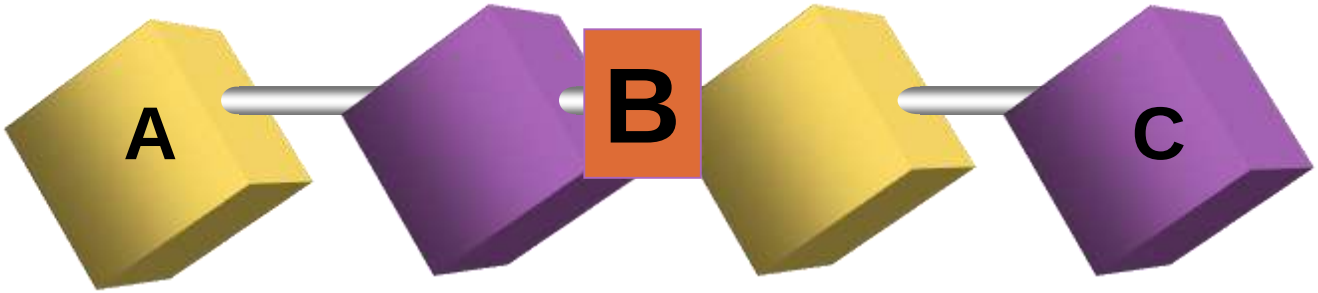
- *Usually, the alpha level is set at p=0.05, i.e. the investigator is willing to run a 5% risk of being in error.*

PROCESS OF HYPOTHESIS TESTING

- *Hypothesis testing is done by various tests of statistical significance.*

- *Significance testing involves following basic steps :*

PROCESS OF SIGNIFICANCE TESTING



ASSERTING NULL HYPOTHESIS

A

ESTABLISHING ALPHA LEVEL / LOS

B

REJECTING/ FAILING TO REJECT NULL HYPOTHESIS

C

- *STEP 1:*

Set the null hypothesis

- *STEP 2 :*

Choose the level of significance

While testing the hypothesis, experimenter makes 2 types of errors.

TYPE I ERROR

TYPE II ERROR

- *The experimenter does not like to commit these types of errors / will see to it that both are committed with least chance.*
- *When we try to minimize one type of error, other type of error will increase & vice versa.*
- *So, we fix the probability of committing type I error/ Level of Significance.*

Usually , LOS is kept as 5% ,i.e. experimenter runs a risk of making wrong decision only in 5 out of 100 cases.

- *STEP 3:*

Compute the value of test statistics

- *STEP 4 :*
- *Obtain the table value at given L.O.S & for the number of degree of freedom.*

- *STEP 5 :*
- *Compare calculated value with table value.*

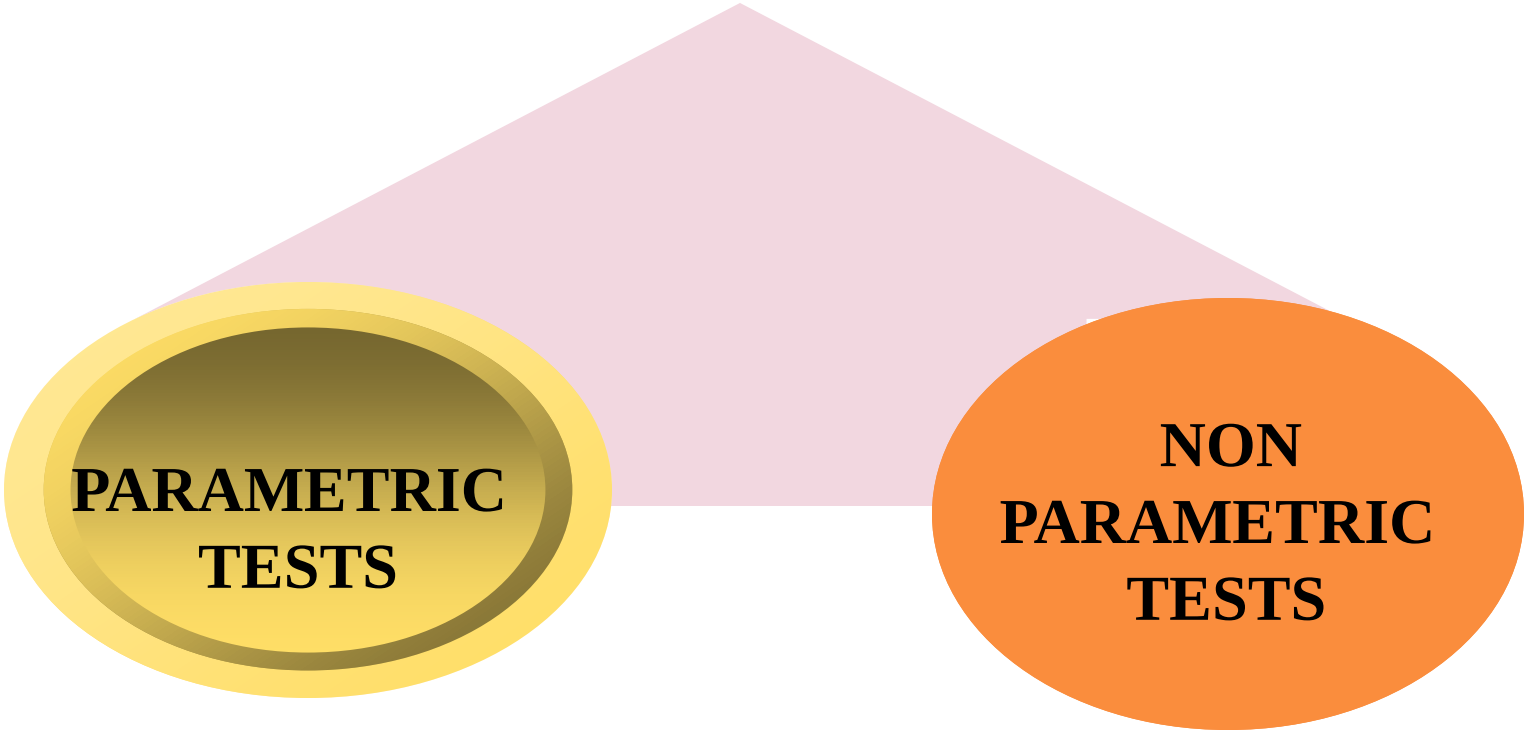
- *STEP 6 :*

Draw the conclusion.

Reject Ho if Calculated Value > Table Value

Accept Ho if Calculated Value < Table Value

STATISTICAL TESTS



PARAMETRIC TESTS

- *They are those tests, which assume certain properties of parent population from which we draw samples.*
- *t-test*
- *Chi-square test*
- *F-test*
- *Z-test*
- *Karl`s pearsonn Correlation & Regression analysis*

NON PARAMETRIC TESTS

- *They are those tests which do not depend on any assumption about the parameters of parent population.*

- *Kruskal- Wallis Test.*
- *Fisher-Irwin Test.*
- *Sign Test.*

PARAMETRIC TESTS

BASIS-t TEST

- *T-Test is based on t-distribution.*
- *It is a sampling distribution derived from the parent normal distribution.*
- *It is a symmetric distribution about its mean*
- *Like in all symmetric distribution, mean , median & mode coincide.*

APPLICATIONS OF t-Test

- *It is applied to small samples (of size below 30) in 2 situations to find the significance of difference between 2 means:*

- *As Unpaired t-test*

- *As Paired t-Test*

INDICATIONS OF t-TEST

- *If 2research samples come from 2 groups e.g. a group of men & a group of women, unpaired t-test is used.*

- *If 2 samples come from the same group e.g. pretreatment & post treatment values for the same study subjects , the paired t-test is used.*

UNPAIRED t-TEST

- *FORMULA USED IS :*

$$t = \frac{\overline{x1} - \overline{x2}}{SE}$$

Here, $\overline{x1}$ *&* $\overline{x2}$ *are 2 sample means & SE is the standard error of difference.*

PAIRED t-TEST

- *FORMULA USED IS :*

$$t = \frac{\bar{x}}{SE}$$

Here, $\bar{x}$ is the mean of the difference of each set of paired observation & SE is the standard error of difference.

CHI-SQUARE TEST

- *'Chi', is a Greek word which refers to the twenty second letter of the Greek alphabet namely 'x'.*

- *Thus the test symbol for Chi-Square is x2.*

- *It is used for categorical data only.*

PRE-REQUISITES

- Random Samples
- Qualitative Data
- Lowest observed frequency not less than 5.

USE OF CHI-SQUARE TEST

- *This test offers a method of testing the significance difference between 2 proportions.*

- *This test represents a useful method of comparing experimentally obtained results with those to be expected theoretically on some hypothesis.*

FORMULA OF CHI-SQUARE TEST

- $X2 = \sum \frac{(O-E)\,2}{E}$

Here, $\sum$ *= Summation*

O= Observed Value

E= Expected Value

- *E.g. An investigator wish to compare the proportion of caries free children living in a district whose water supply is fluoridated to the proportion of caries free children living in a non fluoridated district.*

- *In each district, the investigator would count the number of caries free & not caries free children.*

- *The research question involves two categorical variables : Caries status of child (caries free / not) & the fluoridation status of the district (yes/ no) .*

Table 14-4. Calculation of χ^2 for the Association of Absence of Caries with Fluoridation of W

Category	*O*	*E*	*(O – E)*	*$(O - E)^2$*	*$(O - E)^2/E$*
Caries free, fluoridated water	310	$\frac{500 \times 510}{1000} = 255$	$310 - 255 = 55$	$55^2 = 3025$	$3025/255 = 11$
Not caries free, fluoridated water	190	$\frac{500 \times 490}{1000} = 245$	$190 - 245 = -55$	$-55^2 = 3025$	$3025/245 = 12$
Caries free, water not fluoridated	200	$\frac{500 \times 510}{1000} = 255$	$200 - 255 = -55$	$-55^2 = 3025$	$3025/255 = 11$
Not caries free, water not fluoridated	300	$\frac{500 \times 490}{1000} = 245$	$300 - 245 = 55$	$55^2 = 3025$	$3025/245 = 12$
					$\chi^2 = \sum = 48$

- *Chi square test is considered a measure of association between two categorical variables.*

- *Thus the Null Hypothesis is one of " No association" between 2 variables.*
- *Data is displayed in a Contingency Table.*
- *Number of rows & Number of columns in the table are determined by number of categories in the two variables of interest.*

- *The arithmetic for Chi square test is based on discrepancy between observed frequencies (O) & (E) the frequencies that one would expect if the Null Hypothesis were true.*
- *To calculate E for each O (FORMULA):*

  E= (Row total x Column total / Grand total).

- *Ho for this e.g. states that there is no difference of proportion of caries free children in fluoridated & non fluoridated areas.*

- *p=0.05*
- *df=1,*

  df for a 2 x 2 contingency table is always 1.

CONCLUSION

- *Table value of Chi square test = 3.84*
- *Calculated value= 48.42*
- *Calculated value >> Tabulated value.*
- *Thus, we reject Ho & it states that there is a significant association between caries status & fluoridation of water.*

INTERPRETATION OF CHI-SQUARE TEST

- *If the calculated value of Chi-Square test, is less than the corresponding tabulated value, then it is said to be non-significant at the required level of significance .*

  i.e. the difference between 'O' &'E' , is due to chance.

- *If the calculated value of Chi-Square test is greater than the tabulated value, it is said to be significant.*

  i.e. the difference between 'O' &'E' cannot be attributed to chance & we reject the Null Hypothesis.

CORRELATION
&
REGRESSION ANALYSIS

CORRELATION

- *If 2 random variables are such that , when one gets changed the other will do so in a related manner, they are said to be correlated.*

  i.e. , if 2 quantities vary in such a way that movements in one are accomplished by movements in the other, these quantities are correlated.

- *Variables which are independent are not correlated.*

- *As with Chi-square test, correlation analysis quantifies the relationship between variables, but the scale of measurement needed for correlation is continuous.*
- *E.g. In a sample of elderly xerostomia patients, does the number of root caries increase with increasing amounts of sugar in the diet.(servings/ day)*
- *In a sample of patients from a large dental school clinic, is the number of years of education inversely associated with the number of decayed teeth, as a portion of total DMFT index.*

- *In both of these instances, data are recorded on a single group of subjects & each subject contributes a pair of measures (number of servings/ day & number of root caries, number of years of education & D/ DMFT).*

KARL PEARSON'S COEFFIECIENT OF CORRELATION

- *It refers to measure of association.*
- *Denoted by the symbol 'r' & is referred to as 'r' value.*
- *This value varies from – 1 to +1, going through 0.*

- *A finding of -1 indicates that the 2 variables have a perfect negative linear relationship.*
- *A finding of +1 indicates that they have a perfect positive linear relationship.*
- *0 indicates that 2 values are totally independent of each other.*

- *Correlation coefficient is calculated using the raw values of the (X, Y) pairs of data.*

- *In 1st e.g. r=+ ve , i.e* ↑ *X* ⟶ ↑ *Y.*

- *In 2nd e.g. r = -ve , i.e X & Y vary in opposite direction,* ↑ *X* ⟶ ↓ *Y & vice versa.*

- $$r = \frac{\sum (xi - \bar{x})(yi - \bar{y})}{\sqrt{\sum (xi - \bar{x})^2 \sum (yi - \bar{y})^2}}$$

- *In the formula of 'r', Numerator is the sum of covariance.*
- *The covariance is the product of the deviation of an observation from the mean of the 'x' variable multiplied by the same observation's deviation from the mean of 'y' variable.*

- *Denominator is the square root of the sum of the squared deviations from the mean of the x variable multiplied by the sum of the squared deviations from the mean of the y variable.*

- *Phillips 1978, Dawson –Saunder & Trapp 1994, the value of t can be calculated from the following formula :*

- $$t = r \frac{\sqrt{N-2}}{\sqrt{1-r2}}$$

Here, df = N-2

MERITS & DEMERITS OF PEARSONIAN'S COEFFICIENT

- *In order to measure the degree of relationship, Karl's Pearson's method is most popular.*
- *This coefficient summarizes not only the degree of correlation, but also the direction, i.e. it is positive/ negative.*
- *Coefficient value is affected by extreme items.*

REGRESSION

- *The term Regression means, the act of returning/ going back to normal.*

- *This term was 1st used by British Bio-mathematician Sir Francis Galton (1877) while he was studying the relationship between height of father & sons.*

LINES OF REGRESSION

- *It is the line which gives the best estimate of one variable for any given value of the other variable.*

- *In case of 2 variables, x & y, we will have 2 lines of regression, one of y on x & other of x on y.*

USEFULNESS OF REGRESSION

- *Regression is more important in Biostatistics than correlation , because it estimates the value of one variable (unknown), if we are knowing the value of other variable.*

Regression equations are thus useful for prediction of unknowing phenomenon.

HOW DOES REGRESSION ANALYSIS ESTIMATES THE UNKNOWN VALUE FROM KNOWN ?

- *If we wish to know in an individual case, value of one variable, knowing the value of the other . We calculate regression- coefficient of one measurement to the other.*

- $y = \bar{y} + b\,(x - \bar{x})$,

$here, \bar{y} = mean\ of\ y1, y2, y3 \ldots\ldots yn.$

$\bar{x} = mean\ of\ x1, x2, x3 \ldots\ldots xn.$

$$b = \frac{\sum (x-\bar{x})(y-\bar{y})}{(x-\bar{x})2}$$

Value of b is called as Regression Coefficient of y upon x

MULTIVARIABLE ANALYSIS

TYPES

1. ANOVA- One Way & N-Way
2. ANCOVA
3. MULTIPLE LINEAR REGRESSION
4. LOGISTIC REGRESSION
5. LOG – LINEAR ANALYSIS

MULTI VARIABLE MODELS

- *Statistical models that have one outcome variable but includes > than one independent variables.*

- *These models are more attractive to investigators because they seem more true to life than the single-variable models.*

USES OF MULTVARIABLE STATISTICS

- *Multivariable analysis uses all of the observations in the analysis & is therefore more efficient & robust.*

- *Multivariable techniques also make it possible to determine whether there is an interaction between variables.*

  E.g. Freeman et al 1983, in a large blood pressure survey in Connecticut, found that in African-American people under 50 years of age, hypertension was more likely to occur in men than in women.

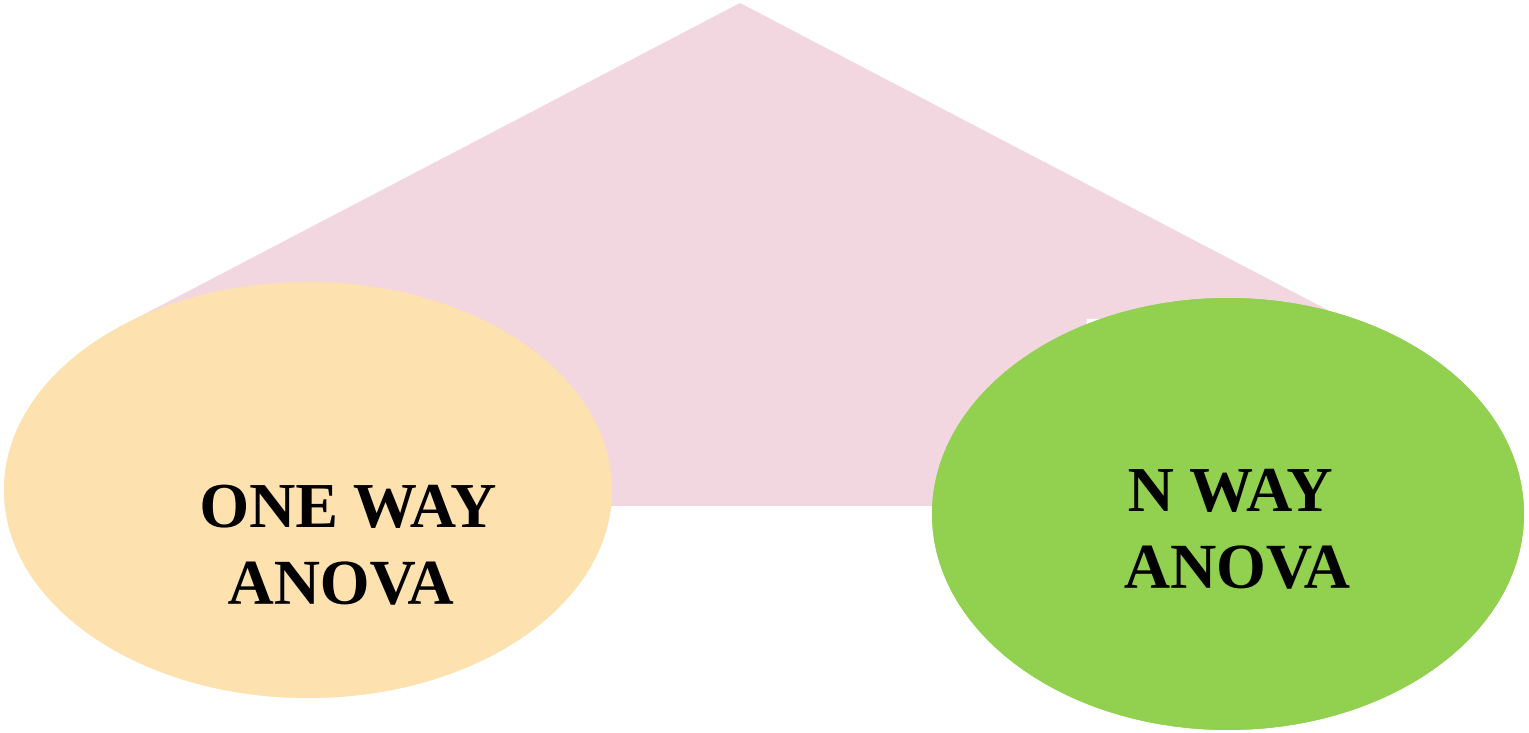
After age 50, that trend is reversed & Hypertension was more likely to occur in women than in men.

Thus , there was an interaction between age & gender in explaining the prevalence of Hypertension.

- *Multivariable analysis was also used by investigators in the Framingham study to develop prediction equations for the 8 year risk of developing Cardio Vascular Disease in people with various combination of following factors : Smoking, Increased cholesterol levels, hypertension, glucose intolerance & left ventricular hypertrophy.*

ANALYSIS OF VARIANCE

ANOVA



If the design includes Only 1 independent variable

If the design includes >than1 independent variable

ONE WAY ANOVA / F-TEST

Suppose a team of investigators wanted to study the effects of drugs A & B on blood pressure .They will randomly allocate hypertensive patients into 4 treatment groups:

- *Those taking drug A alone*
- *Those taking drug B alone*
- *Those taking drug A+B in combination*
- *Those taking placebo*

- *The Student t-test could be used to compare each pair of groups, but this would require 6 different t-tests: each of the 3 drug groups (A, B, & AB) versus the placebo group.*

- *The best approach in such cases is to begin by performing a F-TEST, which is the 1st step of ANOVA.*

- *The F-TEST , is a kind of 'super t-test' that allows the investigators to compare more than 2 means simultaneously.*

- *Investigators would measure systolic blood pressure before & after treatment in each patient & calculate a difference score. i.e.(Post treatment systolic blood pressure- Pre treatment systolic blood pressure) for each study subject.*
- *This difference score is outcome variable.*
- *They would then calculate a mean difference score for each of the four treatment groups, so that these mean scores could be compared in a test of statistical significance.*

F-TEST

- *F-Test is developed by Sir Ronald Fisher.*

- *He developed 2 measures of variance of the observations.*

- *One is called as the Between Group Variance & is based on the variation between different groups.*
- *Other is called as Within Group Variance & is based on the variation within each group.*

- *Ratio of these 2 groups of variance is expressed as :*

- $$F\text{-}Ratio = \frac{Between\ groups\ variance}{Within\ Groups\ variance}$$

$$= \frac{Between\ groups\ mean\ square}{Within\ groups\ mean\ square}$$

INTERPRETATIONS OF F-TEST

- *If the F-ratio is fairly close to 1.0, the 2 estimates of variance are similar & the null hypothesis which states, that all of the means came from the same underlying population is not rejected.*

- *If the ratio is much larger than 1, there must have been some force, attributable to group differences, pushing the means apart & the null hypothesis of no difference is rejected.*

IMPORTANCE OF F-TEST

- *Evaluation of source of variance.*
- *Comparison of more than 2 means.*
- *To study the design of experiments involving the effects of multiple factors.*

N-WAY ANOVA

- *It is used when the study includes more than 1 independent variables.*

- *E.g. N-way ANOVA procedures were used in the study performed by Finkelstein et al 1994, to determine whether supplementing Gonadotropin – releasing hormone with Parathyroid hormone would reduce the Osteoporosis causing effect of GnRH . In this study, the investigators used ANOVA, to examine the effects of treatment & other independent variables on the bone loss induced by estrogen deficiency.*

REFRENCES

- *Epidemiology, Biostatistics, & Preventive Medicine*
  2nd Edition

  James F.Jekel , David L.Katz, Joann G Elmore
- *Textbook of Preventive & Social Medicine*

  18th Edition

  K.Park
- *Fundamentals of Statistics*

  6th Edition

  S.C.Gupta

- *Freeman , D.H, Jr et al.The Prevalence distribution of hypertension : Connecticut adults, 1978-1979.*

  Journal of Chronic Disease, 36, 171-181, 1983

- *Finkelstein , J.S.et al .Parathyroid hormone for the prevention of bone loss induced by estrogen deficiency.*

  New England Journal of Medicine, 331, 1618-1623, 1994.

- *Textbook of Research Methodology*

  2nd Edition,

  C.R.Kothari
- *Fundamentals in Biostatistics*

  6th Edition

  Mahajan.
- *Biostatistics*

  2nd Edition

  Narendra. Kr Dutt

- *Jong`s Community Dental Health*
  Vth Edition , G.M.Gluck & Warren .M.Morganstein

- *Community Oral Health*
  Cynthia .M.Pine


THANK YOU…