IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Applications Of Statistical Tools In Management And Research

***Mr. Anantha Murthy N K Research Scholar, Assistant Professor, CMR CBS, Bangalore, India**

****Dr. John Adaikalam, Associate Professor, Department of Management studies, Annamalai University, Chidambaram Tamil Nadu, India**

***** Dr. M .P.Hrisihikesh, Director, CMR-CBS Bhuvanagiri, OMBR Layout, Banasawadi, Bangalore-560043**

**Abstract**: Statistical tools used for the study include planning, designing, collecting data, analyzing, drawing meaningful interpretation and reporting of the research findings in Management. The statistical analysis of the data gives meaning for the data selected and thereby breathing life into a lifeless data. The results and inferences are precise only if proper statistical tests are used. This paper will try to focus the reader with the basic research tools that are utilized while conducting various studies. This article covers a brief outline of the variables, an understanding of quantitative and qualitative variables and descriptive statistics. Establishment of relationship and strength of relationship through regression modelling. An idea of the sampling and sample size concept to estimate the required no of samples for the study. Testing of hypothesis procedure framework and significance of statistical errors leads to decision making process with the help of parametric and non-parametric tests used for data analysis.

Key words: statistical tools, variables descriptive statistics, Regression Modelling, parametric tests and non-parametric tests, Analysis of variance and Decision Making process

**INTRODUCTION:** Statistics is a branch of science that deals with the collection, organization, analysis of data and drawing of inferences from the samples to the whole population. This requires a proper design of the study, an appropriate selection of the sample size and choice of a suitable statistical tests. An adequate knowledge of statistics is necessary for proper designing of an empirical study or through experiment. Improper statistical methods may result in wrong conclusions which may lead to unethical practice.

**Objectives of the study:**

1. To explore the applications of basic statistical tools for the analysis.
2. To explore its importance and significance in data analysis.

**Limitations**

1. This paper does not focus the application of Multivariate tools.
2. This paper focuses the application of statistical tools for social science and management Subjects.

**VARIABLES:** Variable is a characteristic that varies from one individual member of population to another individual. Variables such as height and weight are measured by some type of scale, convey quantitative information and are called as quantitative variables. Sex and eye color give qualitative information and are called as qualitative variables. Whereas Quantitative variables Quantitative or numerical data are subdivided into discrete and continuous measurements. Discrete numerical data are recorded with whole number such as 0, 1, 2, 3, (Integers), whereas continuous data can assume any value in the given intervals .Observations that can be counted constitute the discrete data and observations that can be measured constitute the continuous data. Examples of discrete data are number of graduates per family, no of articles published in a journal. Similarly, examples of continuous data are the Monthly income of families, time required to finish a project etc.

**Scale of Measurement:**

The variables can be measured in the hierarchical scale of increasing precision can be used for observing and recording the data which is based on categorical or Nominal , ordinal, interval and ratio scales .

Categorical or nominal variables are unordered. The data are merely classified into categories and cannot be arranged in any particular order. If only two categories exist (as in gender male and female), it is called as a dichotomous (or binary) data. The data with the variables related to occupation, geographical areas like Urban, semi Urban and rural are examples of categorical variables. Ordinal variables have a clear ordering between the variables. However, the ordered data may not have equal intervals. Examples are the salaries of the Central government scientists, Professors, Doctors and managers of the organization are related to ordinal scale. Interval variables are similar to an ordinal variable, except that the intervals between the values of the interval variable are equally spaced. A good example of an interval scale is the Fahrenheit degree scale used to measure temperature. With the Fahrenheit scale, the difference between 75° and 80° is equal to the difference between 85° and 90°: The units of measurement are equal throughout the full range of the scale. Ratio scales are similar to interval scales, in that equal differences between scale values have equal quantitative meaning. However, ratio scales also have a true zero point, which gives them an additional property. For example, the distance of A city from B city is 20 Kms whereas the distance of C city from B city is double the distance of A city, i.e. 40 Kms with zero as standard reference.

**Likert scale**: A Likert Scale is a type of rating scale used to measure attitudes or opinions. With this scale, respondents are asked to rate items on a level of agreement. The following table-1 shows the likert scale using in different scenario.

| Points | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|
| Agreement | Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree |
| Frequency | Always | Often | Sometimes | Rarely | Never |
| Importance | Very Important | Important | Moderately Important | Slightly Important | Unimportant |
| Quality | Excellent | Good | Fair | Poor | Very Poor |
| Likelihood | Definitely | Probably | Possibly | Probably Not | Definitely Not |

Table-1: Likert Scales points at different scenario.

**Reliability and Validity:** After Data collected through Designed Questionnaire as a primary source in order to strengthen the statistical significance of scales used for the variables in analysis. So there is an essential requirement of reliability and validity for the data concerned to the survey.

Reliability is a measure of the stability or consistency of test scores. It is the ability for a test or research findings to be repeatable. For example, a medical thermometer is a reliable tool that would measure the correct temperature each time it is used. In the same way, a reliable math test will accurately measure mathematical knowledge for

every student who takes it and reliable research findings can be replicated over and over. Very common statistical tool can use to measure reliability is Cranach's alpha: measures internal Consistency of reliabity for tests with multiple possible answers shown in the table-2.

| Cronbach's alpha | Internal consistency |
|---|---|
| $\alpha \geq 0.9$ | Excellent |
| $0.9 > \alpha \geq 0.8$ | Good |
| $0.8 > \alpha \geq 0.7$ | Acceptable |
| $0.7 > \alpha \geq 0.6$ | Questionable |
| $0.6 > \alpha \geq 0.5$ | Poor |
| $0.5 > \alpha$ | Unacceptable |

Table-2: Different levels of internal consistency

Validity simply means that a test or instrument is accurately measuring what it's supposed to. For example Predictive validity tells how well a certain measure can predict future behavior. One of the most common uses for predictive validity is in University Admissions. Grade Point Average, SAT/ACT scores and other criterion are used to predict a student's likely success in higher education.

**DESCRIPTIVE STATISTICS:** Descriptive statistics try to describe the relationship between variables in a sample or population. Descriptive statistics provide a summary of data in the form of mean, median and mode, range, standard deviation. Descriptive statistics the extent to which the observations cluster around a central location is described by the central tendency and the spread towards the extremes is described by the degree of dispersion. Measures of central tendency are mean, median and mode. Mean (or the arithmetic average) is the sum of all the scores divided by the number of scores. Mean may be influenced profoundly by the extreme variables. For example, the average stay of home quarantining for patients in who is having symptoms of CoVID-19 is 14 days. The extreme values are called outliers. The formula for the mean is

Mean, x (bar) = $\sum x/ n$ where x = each observation and n = number of observations. Median is defined as the middle of a distribution in a ranked data (with half of the variables in the sample above and half below the median value) or it is also called positional average. If we rank the data and after ranking, group the observations into percentiles, we can get better information of the pattern of spread of the variables. In percentiles, we rank the observations into 100 equal parts. We can then describe 25%, 50%, 75% or any other percentile amount. The median is the 50th percentile. The interquartile range will be the observations in the middle 50% of the observations about the median (25th–75th percentile). While mode is the most frequently occurring variable in a distribution. Range defines the spread, or variability, of a sample. It is described by the minimum and maximum values of the variables. Variance is a measure of how spread out is the distribution. It gives an indication of how close an individual observation clusters about the mean value. The variance of a population is denoted by $\sigma^2$.

**Coefficient of variation (CV):** It is used to measure the consistency level of the variables used in the data analysis with the help of mean and standard deviation. Lower the CV value more in the consistency level. The formula for the CV computation is given by

**CV= (Standard deviation/Mean)*100**

**Visual presentation of data:** Data visualization is the graphical representation of information and data. By using visual elements like Bar charts, Pie- charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

## CORRELATOION AND REGRESSION:

Correlation Analysis:
Correlation analysis is applied in quantifying the association between two continuous variables, for example, a dependent and independent variable or among two independent variables. The study of correlation undergoes in three methods
1. Scatter plot.
2. Karl Pearson's method
3. Spearman's Rank Method

**Regression Analysis and Modelling:**
Regression analysis refers to assessing the relationship between the outcome variable and one or more variables. The outcome variable is known as the dependent or response variable and the risk elements, and cofounders are known as predictors or independent variables. The dependent variable is shown by "y" and independent variables are shown by "x" in regression analysis.
Regression modelling can be used to study the impact the relationship between the dependent and independent variables in view of cause and effect relationship which uses further the applications of ANOVA test discussed further. The coefficient of determination gives the significance how well the model is the best to make predictions. The most common interpretation of the coefficient of determination is how well the regression model fits the observed data. For example, a coefficient of determination of 65% shows that 65% of the data fit the regression model. Generally, a higher coefficient indicates a better fit for the model.

Comparison between Correlation and Regression with respect to meaning, variable nature ,usage and objective can be shown in Table-3

| Basis | Correlation | Regression |
|---|---|---|
| Meaning | A statistical measure that defines co-relationship or association of two variables. | Describes how an independent variable is associated with the dependent variable. |
| Dependent and Independent variables | No difference | Both variables are different. |
| Usage | To describe a linear relationship between two variables. | To fit the best line and estimate one variable based on another variable. |
| Objective | To find a value expressing the relationship between variables. | To estimate values of a random variable based on the values of a fixed variable. |

Table-3: Comparison of Correlation and Regression

**Inferential statistics:**

Branch of statistics allows to make predictions ("inferences") from that data. With inferential statistics, data take from samples and make generalizations about a population. Inferential statistics takes data from a sample and makes inferences about the larger population from which the sample was drawn. Because the goal of inferential statistics is to draw conclusions from a sample and generalize them to a population, we need to have confidence

that our sample accurately reflects the population. This requirement affects our process. At a broad level, we must do the following:

1. Define the population we are studying.
2. Draw a representative sample from that population.
3. Analyze the sample that with minimum sampling error.

**Population and sample:**

The collection of individuals for the study of characteristics (Mean and standard deviation) of the population in a survey is termed as population and population size denoted by N. whereas the sample is represented of the population considered for the study. The sample size is denoted by n. the mean and standard deviation values for the population and sample are calculated by using the following formula.

| Population Mean | Sample Mean |
|---|---|
| $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$ | $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ |
| $N$ = number of items in the population | $n$ = number of items in the sample |

$$\text{Population Standard Deviation} = \sigma x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\text{Sample Standard Deviation} = \sigma x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

For Example: the computation of Sample mean and sample standard deviation based of daily wages of five employees of a company is shown in table-4.

| | |
|---|---|
| Daily wages( Rs) of employees | 90, 90, 70, 70, 80 |
| Mean Daily wage | (90+90 +70+ 70+ 80)/5 = Rs. 80 |
| Variance | $(90-80)^2+(90-80)^2+(70-80)^2+(70-80)^2+(80-80)^2$ ]/(5-1) =100 |
| Standard Deviation | = Square root of (100)= Rs.10 |

Table-4: Illustration of computation of sample mean and standard deviation.

**Sampling:** Sampling is a technique of selecting individual members or a subset of the population to make statistical inferences from them and estimate characteristics of the whole population. Different sampling methods are widely used by researchers in market research so that they do not need to research the entire population to collect actionable insights. It is also a time-convenient and a cost-effective method and hence forms the basis of any research design. Sampling techniques can be used in a research survey software for optimum derivation.

For example, if a drug manufacturer would like to research the adverse side effects of a drug on the country's population, it is almost impossible to conduct a research study that involves everyone. In this case, the researcher decides a sample of people from each demographic and then researches them, giving him/her indicative feedback on the drug's behavior.

Sampling is of two types – probability sampling and non-probability sampling. Probability sampling is a sampling technique where a researcher sets a selection of a few criteria and choose samples randomly from a larger population based on the theory of probability. All the members have an equal opportunity to be a part of the sample with this selection parameter. This sampling method considers every member of the population and forms samples based on a fixed process.

For example, in a population of 500 members, each member will have a 1/500 chance of being selected to be a part of a sample. Probability sampling eliminates bias in the population and gives all members a fair chance to be included in the sample.

There are four types of probability sampling techniques:

**Simple random sampling**: One of the best probability sampling techniques that helps in saving time and resources, is the Simple Random Sampling method. It is a reliable method of obtaining information where every single member of a population is chosen randomly, merely by chance. Each individual has the same probability of being chosen to be a part of a sample.

For example, in an organization of 100 employees, if the HR team decides on conducting team building activities, it is highly likely that they would prefer picking chits out of a bowl. In this case, each of the 100 employees has an equal opportunity of being selected.

**Cluster sampling:** Cluster sampling is a method where the researchers divide the entire population into sections or clusters that represent a population. Clusters are identified and included in a sample based on demographic parameters like age, sex, location, etc. This makes it very simple for a survey creator to derive effective inference from the feedback.

For example, if the central government wishes to evaluate the number of people living in Delhi, they can divide it into clusters based on states such as UP, Bihar, Karnataka, Tamil nadu Andrapradesh, Kerala, Madhya Pradesh etc. This way of conducting a survey will be more effective as the results will be organized into states and provide required data.

**Systematic sampling:** Researchers use the systematic sampling method to choose the sample members of a population at regular intervals. It requires the selection of a starting point for the sample and sample size that can be repeated at regular intervals. This type of sampling method has a predefined range, and hence this sampling technique is the least time-consuming.

For example, a researcher intends to collect a systematic sample of 50 people in a population of 1000. He/she numbers each element of the population from 1-1000 and will choose every 20th individual to be a part of the sample (Total population/ Sample Size = 1000/500 = 20).

**Stratified random sampling:** In this method in which the researcher divides the population into smaller groups that don't overlap but represent the entire population. While sampling, these groups can be organized and then draw a sample from each group separately.

For example, a researcher looking to analyze the characteristics of people belonging to different annual income divisions will create strata (groups) according to the annual family income. Example – less than Rs.20,000, Rs.21,000 – Rs.30,000, Rs.31,000 to Rs.40,000, Rs.41,000 to Rs.50,000, etc. By doing this, the researcher concludes the characteristics of people belonging to different income groups. Marketers can analyze which income groups to target and which ones to eliminate to create a roadmap that would bear fruitful results.

**Non-probability sampling:** The researcher chooses members for research at random. This sampling method is not a fixed or predefined selection process. This makes it difficult for all elements of a population to have equal opportunities to be included in a sample. The non-probability method is a sampling method that involves a

collection of feedback based on a researcher or statistician's sample selection capabilities and not on a fixed selection process. In most situations, the output of a survey conducted with a non-probable sample leads to skewed results, which may not represent the desired target population. But, there are situations such as the preliminary stages of research or cost constraints for conducting research, where non-probability sampling will be much more useful than the other type.

Four types of non-probability sampling explain the purpose of this sampling method in a better manner:

**Convenience Sampling**: This Method involves collecting a sample from somewhere convenient to the mall, or local school, or church. Sometimes also called accidental sampling or opportunity sampling.

**Purposive Sampling**: Here the researcher chooses a sample based on their knowledge about the population and the study itself. The study participants are chosen based on the study's purpose.

**Quota Sampling**: Defined as a non-probability sampling method in which researchers create a sample involving individuals that represent a population. For example, divide the population by the state they live in, income or education level, or sex. The population is divided into groups (also called strata) and samples are taken from each group to meet a quota.

**Snowball Sampling**: Snowball sampling or chain-referral sampling is defined as a non-probability sampling technique in which the samples have traits that are rare to find. This is a sampling technique, in which existing subjects provide referrals to recruit samples required for a research study. For example, studying the level of customer satisfaction among the members of a Metro city club, it is extremely difficult to collect primary data sources unless a member of the club agrees to have a direct conversation with researcher and provides the contact details of the other members of the club.

**Sample size:** A 'sample size' is the number of individuals required to conduct survey. A large sample size can yield better and more accurate study results. Research is a multi-step process that allows to achieve the desired results if all the steps are conducted in a systematic manner.

The most important of all the steps is determining how many responses are actually required to be able to derive a conclusive result. Larger sample sizes tend to increase the precision of your study.

**Population Size:** Population size is how many people actually fit to demographic data. For example, to get information on doctors residing in Bangalore. The population size is the total number of doctors in Bangalore city. Population size doesn't always have to be that big as even smaller population sizes can still give you accurate results.

**Confidence Level:** Confidence level tells how sure can be that the data is accurate. It is expressed as a percentage and aligned to the confidence interval. For example, the confidence level is 95% then it is most likely that the results will be 95% accurate with allowable error of 5%.

**The Margin of Error (Confidence Interval):** When it comes to surveys, there's no way to be 100% accurate. Confidence intervals tells how far off from the population means willing to allow the data to fall. A margin of error describes how close the reasonably expect a survey result to fall relative to the true population value.

**Standard of Deviation:** Standard deviation is the measure of the dispersion of a set of data from its mean. It measures the absolute variability of a distribution. The higher the dispersion or variability, the greater the standard deviation and the greater the magnitude of the deviation.

The two methods are usually applied in determining the sample size.

1. Based on mean values

$$n = \left(\frac{Z\sigma}{E}\right)^2$$

- Where n= sample size
- Z is the value from the table of probabilities of the standard normal distribution for the desired confidence level (e.g., Z = 1.96 for 95% confidence)
- E is the margin of error that the investigator specifies as important from a practical standpoint.
- σ is the standard deviation of the outcome of interest.

2. Based on proportion on proportion values

$$n = (z/e)^2 pq$$

- Where n= sample size
- Z is the value from the table of probabilities of the standard normal distribution for the desired confidence level (e.g., Z = 1.96 for 95% confidence)
- E is the margin of error that the investigator specifies as important from a practical standpoint.
- p=Probability of success
- q=probability of failure

**Testing of Hypothesis:**

In inferential statistics, data are analyzed from a sample to make inferences in the larger collection of the population. The purpose is to answer or test the hypotheses. A hypothesis (plural hypotheses) is a proposed explanation for a phenomenon used in the study. Hypothesis tests are thus procedures for making rational decisions about the reality of observed effects. Probability is the measure of the likelihood that an event will occur. Probability is quantified as a number between 0 and 1 (where 0 indicates impossibility and 1 indicates certainty). In Testing of hypothesis, the term 'Null hypothesis' (H0) denotes that there is no relationship (difference) between the population variables in question. Alternative hypothesis (H1 and Ha) denotes that a statement that there is a relationship between the variables is expected to be true. The P value (or the calculated probability) is the probability of the event occurring by chance if the null hypothesis is true. The P value is a numerical between 0 and 1 and is interpreted by researchers in deciding whether to reject or retain the null hypothesis.

If P value is less than the arbitrarily chosen value (known as α or the significance level), the null hypothesis (H0) is rejected actually null hypotheses (H0) is actually true state, this is known as a Type I error. However if the null hypothesis (H0) is accepted, if null hypotheses (H0) is actually false state, this is known as a Type II error shown in the table-5.

| | | Truth | |
|---|---|---|---|
| | **Hypothesis Testing** | **The Null Hypothesis Is True** | **The Alternative Hypothesis Is True** |
| Research | **The Null Hypothesis Is True** | Accurate | Type II Error |
| | **The Alternative Hypothesis Is True** | Type I Error | Accurate |

Table-5: Types of Error.

Decision making process in testing procedure based on p-values is given in the table-6.

| P-Values | Significant | Decision |
|---|---|---|
| <0.01 | Result is highly significant | Reject (null hypothesis) H0 |
| ≥0.01 but <0.05 | Result is significant | Reject (null hypothesis) H0 |
| Value ≥0.05 | Result is not significant | Do not reject (null hypothesis) H0 |

Table-6: Decision making process

**PARAMETRIC AND NON-PARAMETRIC TESTS**: Numerical data (quantitative variables) that are normally distributed are analyzed with parametric tests. Two most basic prerequisites for parametric statistical analysis are:

• The assumption of normality which specifies that the means of the sample group are normally distributed

• The assumption of equal variance which specifies that the variances of the samples and of their corresponding population are equal.

Normal distribution or Gaussian distribution: Most of the data variables values usually cluster around a central value, with symmetrical positive and negative deviations about this point. The standard normal distribution curve is a symmetrical bell-shaped. In a normal distribution curve, about 68.26% of the scores are within 1 SD of the mean. Around 95.44% of the scores are within 2 SDs of the mean and 99.73% within 3 SDs of the mean.

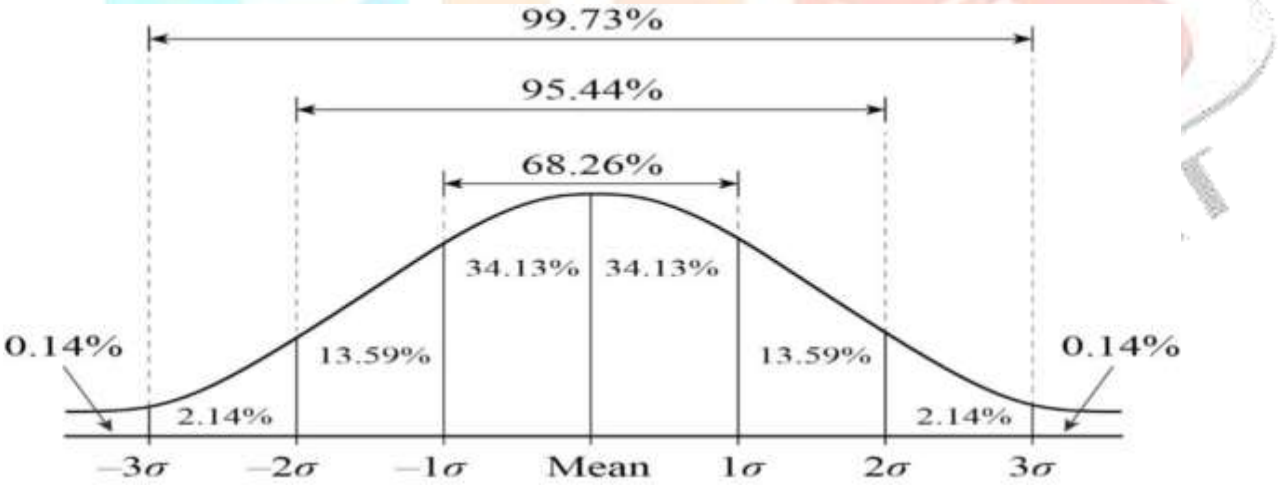


Figure-1: Area Coverage of Normal Distribution

However, if the distribution of the sample is skewed towards one side or the distribution is unknown due to the small sample size, non-parametric. Statistical techniques are used. Non-parametric tests are used to analyses ordinal and categorical data. Parametric tests assume that the data are on a quantitative (numerical) scale, with a normal distribution of the underlying population. The samples have the same variance (homogeneity of variances). The samples are randomly drawn from the population, and the observations within a group are independent of each other.

Testing of hypothesis procedure consists of the following steps.
Frame the null hypothesis.
Frame the Alternative hypothesis.
Based upon the information Compute the Statistic.
Refer the table for significant value.
Take the decision for rejecting/accepting the null hypothesis

The commonly used parametric tests are Z-test, Student's t-test, analysis of variance (ANOVA).

**A Z-test** is a statistical test to determine whether two population means are different when the variances are known and the sample size is large (n>=30). It can be used to test hypotheses in which the z-test follows a normal distribution-test is used for usually,

1. To test the significance of population Mean of the group (One sample test)
2 .To test the significance of population Means of two groups (Two sample test)
3. Test the significance of single population proportion and to test the significance of equality of two population proportions.

**Student's t-test**, in statistics, a method of testing hypotheses about the mean of a small sample (Sample size n<30) drawn from a normally distributed population when the population standard deviation is unknown. In 1908 William Sealy Gusset, an Englishman publishing under the pseudonym Student, developed the t-test and t distribution. (Gusset worked at the Guinness brewery in Dublin and found that existing statistical techniques using large samples were not useful for the small sample sizes that he encountered in his work). The t distribution is a family of curves in which the number of degrees of freedom (the number of independent observations in the sample minus one) specifies a particular curve. As the sample size (and thus the degrees of freedom) increases, the t distribution approaches the bell shape of the standard normal distribution. In practice, for tests involving the mean of a sample of size greater than 30, the normal distribution is usually applied. Student's t-test Student's t-test is used to

1. Test the significance of single population mean.
2. Test the significance of equality of means of two populations (independent samples).
3. Test the significance of equality of means of two populations( dependent samples)

**Analysis of variance (ANOVA)** is the technique of separation of variance ascribable to one group of causes from the variance ascribable to other group. It is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study. ANOVA uses F-test statistic for decision about acceptance or rejection of the null hypothesis. Using a standard ANOVA in this case is not appropriate because it fails to model the correlation between the repeated measures: The data violate the ANOVA assumption of independence. Hence, in the measurement of repeated dependent variables, repeated measures ANOVA should be used.

**Non-parametric tests** when the assumptions of normality are not met, and the sample means are not normally, distributed parametric tests can lead to erroneous results. Non-parametric tests (distribution-free test) are used in such situation as they do not require the normality assumption. Non-parametric tests may fail to detect a significant difference when compared with a parametric test. That is, they usually have less power. As is done for the parametric tests, the test statistic is compared with known values for the sampling distribution of that statistic and the null hypothesis is accepted or rejected. The types of non-parametric analysis techniques and the corresponding parametric analysis techniques are listed below.

**1-sample sign test**. Use this test to estimate the median of a population and compare it to a reference value or target value.

**1-sample Wilcoxon signed rank test.** This test, can be used to estimate the population median and compare it to a reference/target value. However, the test assumes the data comes from a symmetric distribution (like the Cauchy distribution or uniform distribution).

**Friedman test.** This test is used to test for differences between groups with ordinal dependent variables. It can also be used for continuous data if the one-way ANOVA with repeated measures is inappropriate (i.e. some assumption has been violated).

**Goodman Kruska's Gamma test**: a test of association for ranked variables.

**Kruskal-Wallis test.** Use this test instead of a one-way ANOVA to find out if two or more medians are different. Ranks of the data points are used for the calculations, rather than the data points themselves.

**Mann-Whitney test.** Use this test to compare differences between two independent groups when dependent variables are either ordinal or continuous.

**Mood's Median test.** Use this test instead of the sign test when the two samples are in independent nature.

**Spearman Rank Correlation:** Use to find a correlation between two sets of data which is measured in ordinal scale.

**Chi-square test**: This test compares the frequencies and tests whether the observed data differ significantly from that of the expected data if there were no differences between groups (i.e., the null hypothesis).

Conditions for applying $\chi^2$-test
The sample observations should be independent.
$\sum O=\sum E$
The total frequency should be large (>50).
No Expected frequencies should be less than 5.

**SOFTWARES AVAILABLE FOR STATISTICS, SAMPLE SIZE CALCULATION AND POWER ANALYSIS**: Many statistical software are available currently for data analysis. The following software are using for the data analysis are listing below

1. SPSS Stands for Statistical Package for the Social Sciences (IBM corporation), which is user friendly are available in many versions.

2. Statistical Analysis System ((SAS - developed by SAS Institute North Carolina, United States of America),

3. R (designed by Ross Ihaka and Robert Gentleman from R core team) which is available from open source.

4. Minitab (developed by Minitab Inc.) which is user friendly and working with the knowledge of MS-Excel.

5. Stata (developed by StataCorp) and the MS Excel (developed by Microsoft).

**CONCLUSION:** Based on the Application of statistical tools discussed in the paper, it is very important that a researcher knows the concepts of the basic statistical methods used for conduct of a research study. This will help to conduct an appropriately well-planned and designed study leading to valid and reliable results. Inappropriate use of statistical techniques may lead to faulty conclusions, inducing errors and undermining the significance of the article. Bad statistics may lead to bad research, and bad research may lead to unethical practice. Hence, an adequate knowledge of statistics and the appropriate use of statistical tests are important. An appropriate knowledge about the basic statistical methods will brings good results and will go a long way in improving the research designs and producing quality research work projects undertaken which can be utilized by the above statistical tools for formulating with the evidence-based guidelines.

**REFERENCES**

1. Winters R, Winters A, Amedee RG. Statistics: A brief overview. Ochsner J 2010; 10:213-6.

2. Bhattacharyya, G. K., and R. A. Johnson, (1997). Statistical Concepts and Methods, John Wiley and Sons, New York.

3. Dixon, W. J. and Massey, F.J. (1969). Introduction to Statistical Analysis, McGraw-Hill, New York.

4. Das S, Mitra K, Mandal M. Sample size calculation: Basic principles. Indian J Anaesth 2016; 60:652-6.

5. Altman DG, Bland JM. Parametric v non-parametric methods for data analysis. BMJ 2009; 338:a3167.

6. Berwick V, Cheek L, Ball J. Statistics review 10: Further nonparametric methods. Crist Care 2004; 8:196-9.

7. Basic statistical tools in research and data analysis, Zulfiqar Ali, S Bala Bhaskar , 2016 Indian Journal of Anesthesia | Published by Walters Kluwer - Medknow

8. Guenther, W. C. (1977). "Sampling Inspection in Statistical Quality Control", Griffin's Statistical Monographs, Number 37, London.

9. Hollander, M. and Wolfe, D. A. (1973). Nonparametric Statistical Methods, John Wiley & Sons, New York.

10. Moore, D. S. (1986). "Tests of Chi-Square Type". From Goodness-of-Fit Techniques (D'Agostino & Stephens eds.).

11. Miller, R. G., Jr. (1981). Simultaneous Statistical Inference, Springer-Verlag, New York.

12. Scheffe, H. (1959). The Analysis of Variance, John Wiley, New-York.

13. Duncan, D. B. (1975). "t-Tests and Intervals for Comparisons suggested by the Data" Biometrics, 31, pages 339-359.

14. Ryan, T. A. (1960). "Significance Tests for Multiple Comparisons of Proportions, Variances, and Other Statistics", Psychological Bulletin, 57, pages 318-328.

15. Research Methodology – Methods & Techniques 2 ed, Kothari C. R. – Vishwa Prakashan – New Delhi 1990.

16. Business Research Methods (2014)-12th Edition –Donald R Cooper and Pamela s. Schindler- The McGraw-Hill/Irwin Series.