ISBN: 978-81-948755-8-1

5. Statistical Techniques: An Overview

Abhishek Chakraborty

Assistant Professor,
Department of Management, Bhilai Institute of Technology,
Durg (C.G.), India.

Sunil Kushwaha

Research Scholar,
Department of Management, Bhilai Institute of Technology,
Durg (C.G.), India.

Sushma Singh

Research Scholar,
Department of Management, Bhilai Institute of Technology,
Durg (C.G.), India.

5.1 Introduction:

Regardless of where we stand on the matter of importance of Data Science, it's simply impossible to ignore the continuing importance of data, and our ability to analyze, organize, and interpret to draw conclusive decisions (Kazmier & J, 1968). It is important to understand the objective and purpose behind the various techniques, in order to know where, how and when to use them. It is important to assess the performance of various methods and techniques, to know how well or how badly it is working and solving the purpose of use (Pallant., 2013).

Additionally, this is an exciting research area, having important applications in psychology, literature, science, industry, and finance names a few but not limited to. Statistical way of thinking is to formulate problem and solve the problem statistically by collecting costly data (Yong & Pearce, 2013). There are two types of statistics: Descriptive and Inferential statistics

a. **Descriptive Statistics:** Summarizing information collected, and knowledge gathered by means of graphs like frequency tables, pie charts, bar charts, dot plots and histogram or by means of tables like mean, average, and correlation coefficients.
b. **Inferential Statistics:** Drawing conclusions about the population on the basis of limited number of sample. While using statistical techniques for data analysis following structured steps can be used:

5.2 Set Hypothesis:

The researches on any area are always based on certain assumptions which must be defined as statements. These statements are called hypothesis. There are two types of hypothesis, Null Hypothesis and Alternate Hypothesis.

The null hypothesis generally referred by H0, is the hypothesis which is checked for possible rejection under the assumption that it is true. Theoretically, a null hypothesis is set as no difference or status quo, until and unless it is proved wrong when sample us used.

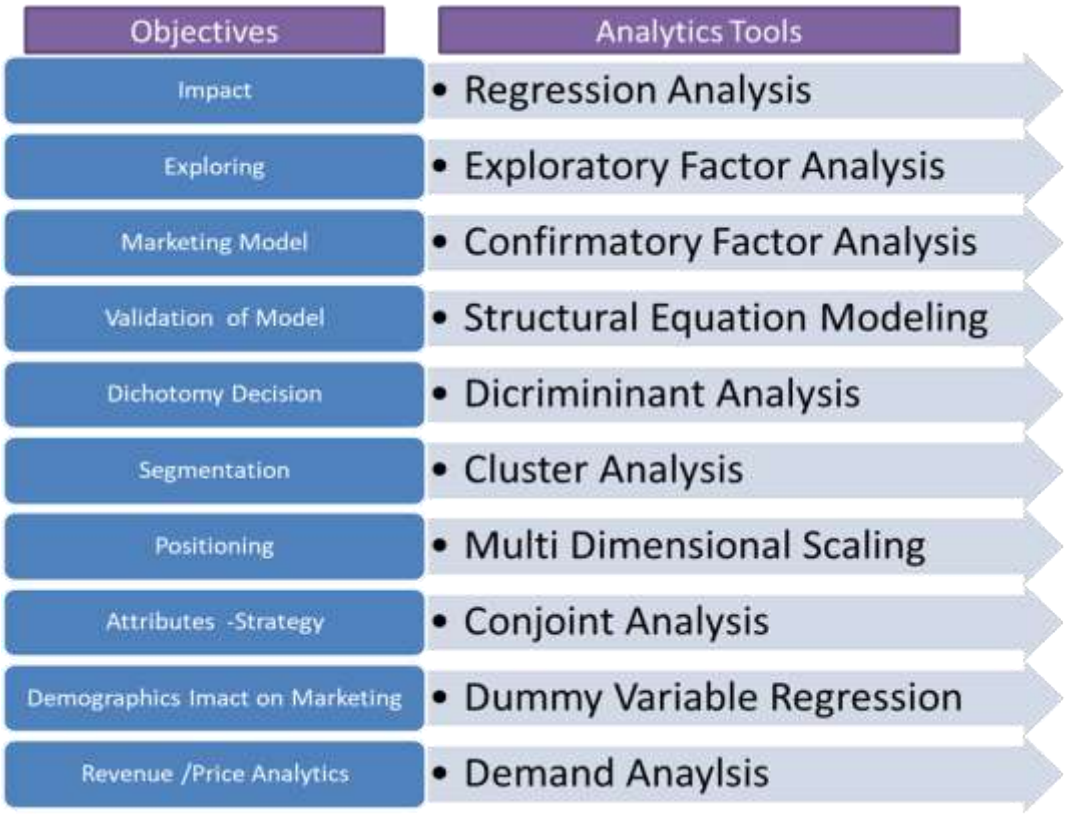
The alternative hypothesis, generally referred by H1, is opposite of the null hypothesis. In other words, when null hypothesis is found to be true, the alternative hypothesis must be false or when the null hypothesis is found to be false, the alternative hypothesis must be true.

5.3 Define Appropriate Statistical Test:

In research, most fundamental difficulty is selecting appropriate statistical tools. It is advised to select statistical test on the basis of data type, sample size, and the level of data may provide a benchmark for deciding the statistical test (Velicer & Jackson, 1990).

Apart from these, the statistics used in the study (mean, proportion, variance, etc.) must also be considered when a researcher decides on appropriate statistical test, which can be applied for hypothesis testing in order to obtain the best results.

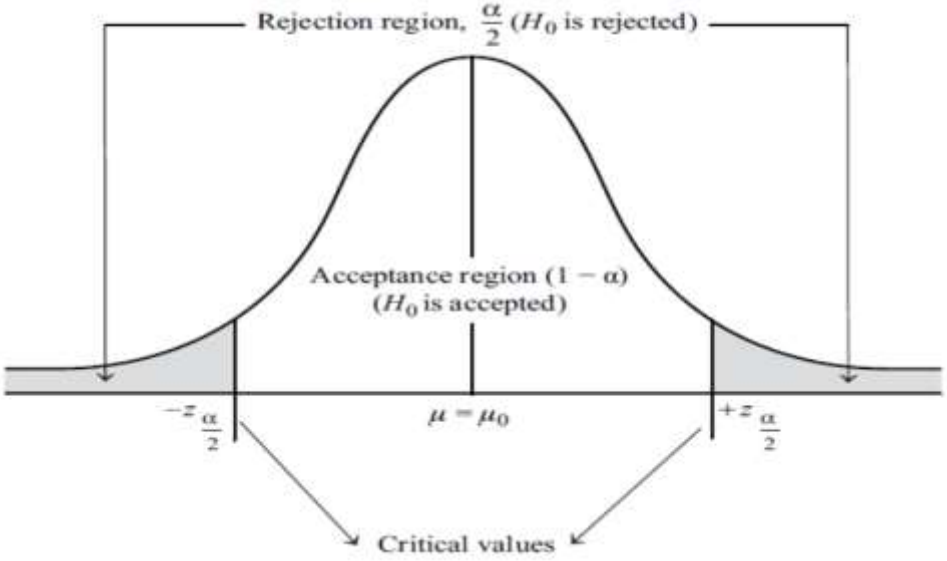
One another way of selecting appropriate techniques is analyzing the objective of research. Some authors compile the techniques on the basis of objectives set. Depending upon the objectives of the research, statistical tools and techniques are selected (Yong & Pearce, 2013). Following are compilation of analytical tools suggested on the basis of objectives set



5.4 Define Level of Significance:

The level of significance generally denoted by α is the probability, which is attached to a null hypothesis, which may be rejected even when it is true. The level of significance is also known as the size of the rejection region or the size of the critical region. The levels of significance which are generally applied by researchers are: 1%; 5%; 10%.

5.5 Define the Decision Rule:



On the basis of level of significance, decision rule is set to accept the hypothesis or reject it, for this rejection region and acceptance region is defined.

5.6 Collect the Sample Data:

In this stage of sampling, data are collected, and the appropriate statistics are calculated. The first four steps should be completed before collecting the data for the study. It is better not to collect the data first and then decide on the stages of hypothesis testing (Kothari & Garg, 2019).

5.7 Analyze the Data:

In this step, the researcher has to compute the test statistic. This involves selection of an appropriate probability distribution for a particular test. Depending upon number of dependent variables analysis approach can be Univariate analysis and multivariate analysis (Gorsuch, 1990).

Univariate Analysis:

If dependent variable under study is only one at a time, then Univariate analysis should be used and respective statistical techniques and tools to be used as discussed in following parts.

Multivariate Analysis:

If dependent variable under study is more than one at a time, then multivariate analysis should be used and respective statistical techniques and tools to be used as discussed in following parts.

At next level, depending upon type and nature of scale used to collect data, parametric and non-parametric test to be selected.

Parametric tests are those that make assumptions about the parameters of the population distribution from which the sample is drawn.

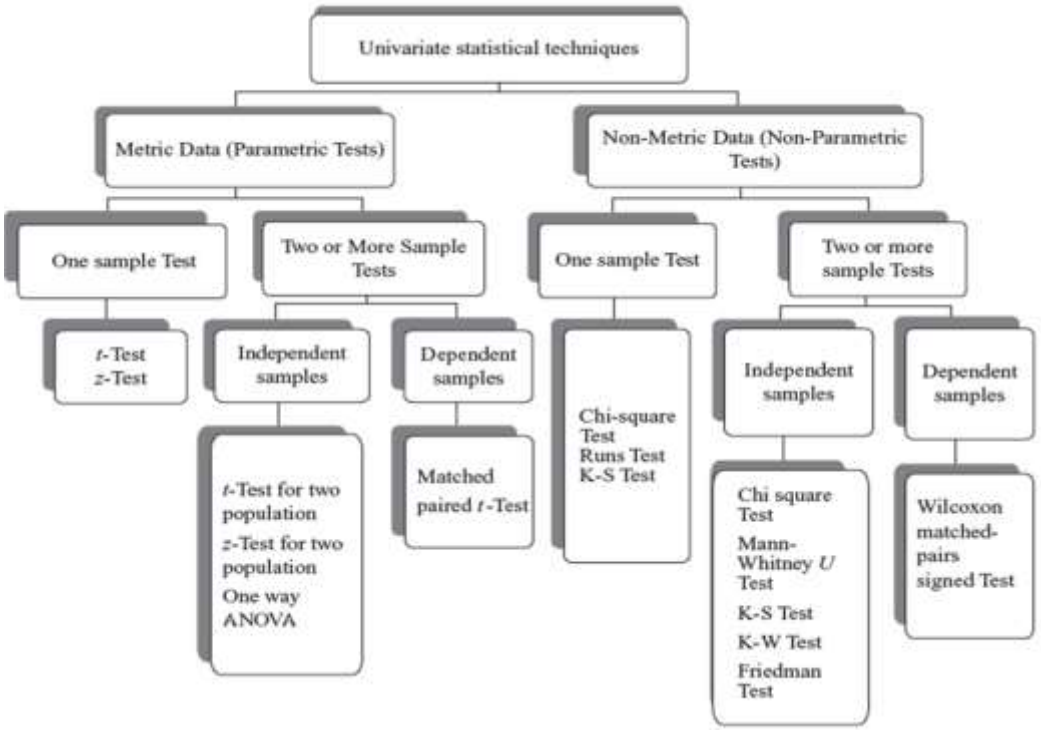
This is often the assumption that the population data are normally distributed. Non-parametric tests are "distribution-free" and, as such, can be used for non-Normal variables.

- **A Parametric test** is a test whose model requires and specifies certain conditions about the parameters of the population from which the sample is drawn. Such tests make certain assumptions about the nature of the underlying population like Normal Probability Distribution and their validity rests upon the validity of these assumptions. These tests are more powerful and strong in their assertions and are usually applicable when data is interval scale or Ratio Scale. These tests are very much rich and developed (Bajpai, 2017).
- **Non-Parametric tests** are also known as distribution free methods. These are the tests whose model does not specify conditions and assumptions about the parameters of the population; they lack parameters. These are widely used for nominal or ordinal data where no parametric tests are not available at all. However, they can also be used for Ratio or Interval Scale data as well. These tests are not very powerful and strong in their assertions. Non-parametric statistical tests are typically much easier to learn and apply than are parametric tests. These tests usually convert data into ranks (hence, such tests are also sometimes known as Rank Tests) or signs and thereby may lose some important information.

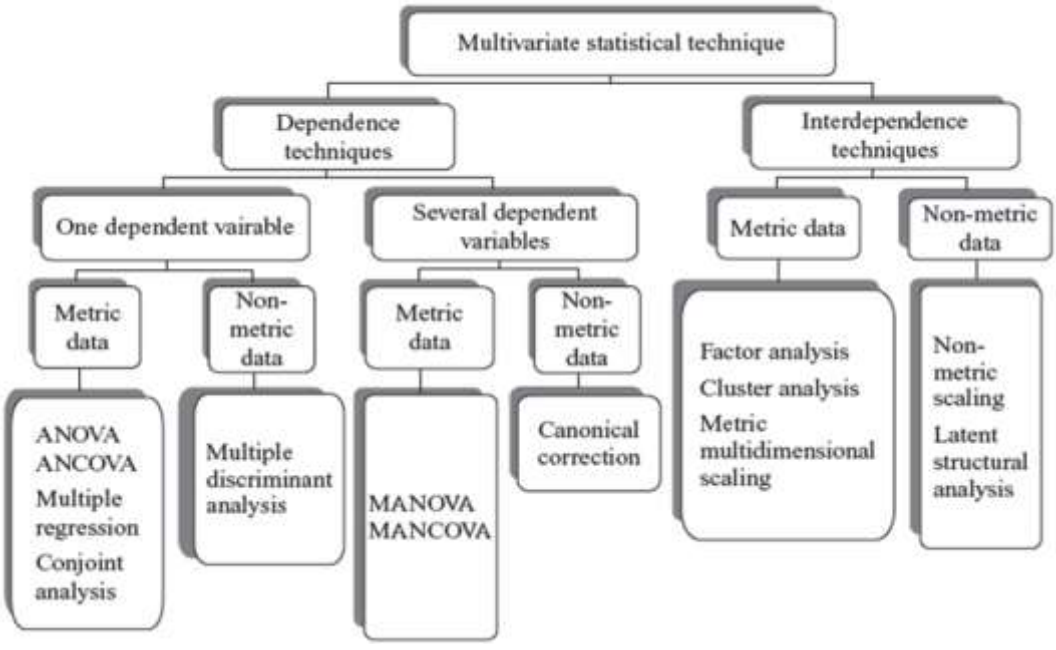
Following are flowcharts for selection of types of statistical tests,

a. First step select the type of dependent variable i.e., Univariate or multivariate.
b. Second step is to select type of test i.e., parametric or non-parametric test.
c. Third step is to select test on the basis of no. of samples i.e., one sample or more than one sample.
d. Fourth step is to select test depending upon interrelationship between samples i.e., dependent samples or independent samples.

Univariate Statistical Techniques:



Multivariate Statistical Techniques:



5.8 Arrive at a Statistical Conclusion and Implication:

In this step, the researchers draw a statistical conclusion. A statistical conclusion is a decision to accept or reject a null hypothesis. Statisticians present the information obtained using hypothesis-testing procedure to the decision makers. Decisions are made on the basis of this information. Ultimately, a decision maker decides that a statistically significant result is a substantive result and needs to be implemented for meeting the research goals.

The above steps are suggestive in nature which researchers, scholars and academicians can adopt to follow a structured step-by-step process of data analysis.

5.9 Bibliography:

1. Bajpai, N. (2017). Business Research Methods. Pearson Publications.
2. Field, A. (2000). Comparing several means: ANOVA (GLM 1). In F. A, Discovering Statistics Using SPSS for Windows, (1st ed., pp. 243-294). London: SAGE Publications Ltd.
3. Gorsuch, R. L. (1990). Common Factor-Analysis versus Component Analysis - Some Well and Little Known Facts. Multivariate Behavioral Research, 25 (1), 33-39.
4. Kaiser, H. (1974). An index of factorial simplicity. Psychometrika, 39 (1), 31-36.
5. Kazmier, & J, L. (1968). Probability learning as related to the response format used. Psychon. Science, 11 (6), 199-200.
6. Kothari, C. R., & Garg, G. (2019). Research Methodology. New Age International Publishers.
7. Littlejohn, A., Milligan, C., Fontana, R. P., & Margaryan, A. (2016). Professional Learning Through Everyday Work: How Finance Professionals Self-Regulate Their Learning. Vocations and Learning, 9 (1), 207–226.
8. MacCallum, R., Widaman, K., Zhang, S., & Hon. (1999). Sample size in factor analysis. Psychological Methods, 4, 84-99.
9. Pallant. J. (2013). SPSS survival manual: a step by step guide to data analysis using IBM SPSS (5th Ed.). Australia: Allen and Unwin.
10. Velicer, W., & Jackson, D. (1990). Component Analysis versus Common Factor-Analysis - Some Further Observations. Multivariate Behavioral Research, 25 (1), 97-114.
11. Yong, A. G., & Pearce, S. (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. Tutorials in Quantitative Methods for Psychology, 9 (2), 79-94.