

On Importance of Normality Assumption in Using a T-Test: One Sample and Two Sample Cases

Srilakshminarayana Gali,
SDM Institute for Management Development,
Mysore, India.
E-mail: lakshminarayana@sdmimd.ac.in

Abstract

Over the last few decades, the researchers have been using statistical tools for data analysis that are based on the assumption of normality and other assumptions related to specific tools. Among other assumptions, the assumption of normality plays a vital role as most of the parametric tools and regression methods are based on this assumption. When the researchers have developed the theoretical aspects related to these methods, they have assumed that the population from where the sample has been drawn follows a normal law. But, in reality the population need not be always normal. The assumption of normality has to be tested using the sample drawn before these methods are adopted to draw conclusions about the population under study. When these methods are used for non-normal population, they may lead to unreliable results and also inferences with low power. For example, a t-distribution has been developed completely based on the assumption of normality. Other assumptions also have to be tested before the methods are adopted. If any of the assumptions is not satisfied by the data, then it is not advised to use the methods. Alternative non-parametric and robust statistical methods have been developed by the researchers to address these issues. But unfortunately, these methods are not as popular among the researchers as that of parametric methods. Even though few non-parametric methods are known their usage in the research is limited. Robust statistical methods which are more powerful than other methods are not being used by the researchers for data analysis. The present paper is an attempt to bring to the notice of the researchers the role of normality in each parametric method, importance of testing other assumptions before using standard statistical methods and introduce robust statistical methods that can give better results when the assumptions associated with the standard methods fail. The paper not only focuses on theoretical developments but also provides empirical evidence at each stage in support of robust statistical methods.

Key words: Data analysis, Normality, Parametric and Non-Parametric methods, Power of the test, Robust Statistical methods, t-distribution, Trimmed mean.

JEL Classification: C 19.

1. Introduction

The use of statistical tools, to analyze the data collected, is being given utmost importance in research studies. The main emphasis has been given on extracting appropriate information from the data, which was collected, to address predetermined objectives of the study. It is very important for any researcher to choose an appropriate tool for the same purpose. Another important aspect in the choice of an analytical tool is to look at how robust it is to the deviation of the assumptions. Taking this need into consideration, various statistical tools have been developed to model the data appropriately and draw valid inferences. In the process, it has been noted that few tools depend on different assumptions which are critical for their functioning. For example, a t-test functions well if the data has been collected from normal population. But, this assumption has to be tested using the sample drawn, before it is used to test the hypothesis constructed on the population mean. If the data do not support the assumption of normality and still if one uses it, then the inferences drawn need not be reliable. Also, the power of the test may be low for an appropriate alternative value against the null hypothesis value.

In reality population need not always be normal and can also exhibit a behavior that can be close to heavy tail phenomenon. The presence of outliers or extremes lead to a heavy tailed phenomenon and one has to choose a model that best fits the phenomenon. For example, a lognormal model or a stable Pareto model will be useful to model the stock market returns better than a normal model. Similarly, symmetric stable distributions have been used to model the heavy tail phenomenon. It is very important for one to examine the data properly and test for presence of extremes or outliers. The use of other non-normal models have not gained popularity due to the limitations associated with each of the model. One such limitation is the existence of finite moments. For a stable model with index greater than 1 and less than 2, the mean is finite but the variance is infinite. For index less than 1 the mean is not defined. But, the researchers look for a model for which all the moments are finite and can be used for further analysis. This is one another reason, to use either normal or lognormal model in most of the research studies. Another important limitation is that, only for few values of the index the density function of the stable law is well defined. For example, the density function is well defined for index equal to one (Cauchy), half (Levy), and index equal to two, one gets a normal law. For other values of the index, the density function is defined as infinite series and is difficult to use for further analysis. In spite these limitations, these models have gained importance in modelling of stock market returns and other economic, financial variables.

Most of the research studies, rely on the sample drawn out of the population, to draw valid inferences about the characteristics of the population (parameters) under study. The

behaviour of the sample characteristics (statistics) are studied using the standard sampling distribution like Student-t, Chi-Square and F distributions. These distributions have been developed based on the assumption that sample is drawn from a normal population. Also, the tests based on these distributions are highly sensitive to the assumption of normality. For example, a Chi-Square distribution, which is used to model the behavior of variance, is derived as square of standard normal, t-distribution is derived using standard normal and Chi-square, and F-distribution is derived using two independent Chi-square random variables. Though few comment on the robustness of these tests, in practical application these tests need not perform better in the absence of normality. In section 5 we presented few empirical examples in support of this statement.

One important aspect in any research is to estimate the characteristics of the population under study precisely and accurately. The parameters of interest include most of the descriptive statistics, such as mean, median, standard deviation, Skewness, and Kurtosis etc. The sample drawn is used to construct both point estimates and interval estimates for these unknown parameters of the population. The point estimates constructed gives the current position of these parameters whereas interval estimates gives how the values of these parameters fluctuate within the interval. Note that the interval estimates are usually constructed under the assumption of normality, at predefined level of confidence (99%, 95%, etc.). If the data do not satisfy the assumption of normality, the interval estimates constructed need not provide better values for the parameters. Most of the estimators that are being used to estimate the parameters possess desired properties of a good estimator and hence can be used appropriately as per the given situation. But, care has to be taken when the data has extremes or outliers. For example, mean is severely affected by extreme observations. In cases, where the data has extreme observations, it is advised to use a trimmed mean as an alternative to mean. One has to note that trimming has to be done at specific percentile points in order to preserve the property of asymptotic normality. It is well known that for non-normal populations, if variance is finite then central limit theorem can be used to approximate the distribution of mean to a normal law. Similar discussion holds good for a trimmed mean if the trimming is done at specific points. After trimming is done, one can use the same to estimate the population mean and construct interval estimates under normality.

The next step in any research is to construct suitable hypotheses on the parameters and test them either for possible rejection or non-rejection. The most frequently used tests are t-test, ANOVA (F-test), Chi-square test etc., which rely completely on the assumption of normality. If the assumption of normality is violated then one can question on the reliability of the inferences drawn using these tests. Hence, it is not advised to use these tests under non-normality and alternative methods either non-parametric or robust methods, have to be used

to draw valid inferences. There are few arguments and discussions that appear on different forums, groups, and papers, on the robustness of these methods, when few of the assumptions are violated. For example, there are arguments that t-test is robust to the violation of normality. But, the question is under what conditions it is robust? One have to look into the conditions and again check whether the data supports those assumptions. It may lead to a process that need not give reliable results. Few thoughts on this have been presented in the literature review and in the sub-sequent sections that cover the discussion on robustness of the methods. This is a very important discussion as it lays foundation to the use of alternative methods in the place of traditional methods.

The developments in the area of statistical inference include the non-parametric statistical methods that paved path to researchers choose an alternative tools when the parametric tools fail to satisfy any of the assumptions. One important aspect of many non-parametric procedures is that they are distribution free. These methods can be used as an alternative to parametric methods, when any of the underlying assumptions of parametric are violated. They are simple to understand and easy to apply in research. When one is dealing with either one sample or two sample problems, Wilcoxon Signed rank test and Mann-Whitney U test are appropriate alternatives to t-test. If the assumption of normality is violated, then using t-test may decrease the power in identifying the wrong null hypothesis correctly. In such cases, using these methods will increase the power. In section 6, we consider the one sample and two sample problems and demonstrate that under non-normality non-parametric methods produce better power than a t-test.

Over the last few decades, there has been a lot of discussion on the appropriate use of parametric, non-parametric methods and their robustness to violation of few assumptions. This has opened gates decades back to alternative methods named as “Robust Statistical methods”. The development of these methods is due to the presence of outliers in the data. It started with estimation and then it has moved on to the development of testing procedures, regression methods etc. One advantage the researchers can draw by using these methods, is to ensure that the type I error does not increase and at the same time power of the test is at desired level, against the alternative value. Under the section “Robust Statistical Methods”, we present the discussion on these methods. We confine to those methods based on trimming and discuss the advantages of trimming in detail. We only mention about other types of estimators and omit the details of discussion as they can be done on lines similar to that of trimmed mean.

Finally we would like to inform the readers that, taking few of the above statements made and few statements that were drawn from different sources, we present our views. This also forms as a motivating factor to re-look at traditional methods and compare them with

other modern methods. Testing the assumptions associated with each method and computing the power for each test will help the researcher to take appropriate decisions regarding the choice of the methods. Also, one has to note that a method sensitive to the assumptions should not be preferred for analysis. A method that can sustain the violation of assumptions should be chosen so that it makes researcher rely on the output of the same. Few aspects like sample size, power of the test, appropriateness of the statistical tool selected etc., have to be considered in any empirical study. When the study is based on an empirical sample, the researcher has to choose sufficing sample size that makes the study more reliable and also decrease the sampling error. At the same time, non-sampling errors have to be taken seriously and ensure that they do not contribute to the sampling error. These issues are important as they can increase the variance of the estimator and also lead to a non-normal situation. The main aim of the paper is to highlight these issues and mainly show empirically how variance contributes to non-normality and how it affects the power of the test.

2. Literature Review

The Scientists who make valuable contributions to their respect areas of research, though empirical studies rely on the sample data drawn. They ensure that appropriate sampling design has been used to collect the responses from the respondents. The data collected is cleaned and made ready from further processing using appropriate data management tools. In later stages, the same is processed using appropriate statistical tools to extract information. In the process of the study, researchers construct several hypotheses on the key characteristics of the population. These characteristics are estimated using the sample drawn and the hypotheses are tested using standard parametric, non-parametric or suitable testing procedures. Most of the parametric testing procedures have been developed based on the assumption of normality and any deviation from this assumption may affect their performance in producing reliable results. This section looks at various views expressed by scientists, some of them who have developed the methods, on the assumption of normality and consequences that arise due to violation of the same.

Right from its derivation (Abraham de Moivre, 1718) normal distribution has gained importance in further developments of the statistical methods, as well as occupied a prominent place in the history of statistics. Patel and Read (1982), in their book on normal distribution, discuss in detail about normal distribution and presents many of its important properties, characterizations, and, other approximations. A beginner who wishes to know about normal distribution is advised to start with this book and proceed onto other advanced books that talk about normal distribution.

One of the most important and most frequently used parametric test, the t-test is highly dependent on the assumption of normality. Student (1908), in his paper on the derivation of

probable error of mean, assumes that the population from where the sample has been drawn is normal and derives t-distribution curve to model sample means. He also mentions that, the conclusions drawn are not strictly applicable to populations known not to be normally distributed. If one questions on strict normality of the population, the answer is not necessarily. The population need not be strictly normal to use the t-distribution in the analysis. It is sufficient if the population is approximately normal. The distribution curve that has been derived by student is used to model the means obtained from samples drawn from a normal population. One has to note that if the population is strictly normal, then the distribution of mean will be normal. If the population is non-normal and population variance is known, then one can use central limit theorem to approximate the distribution of sample mean to normal. But, if the population variance is not known and estimated using the sample drawn, then one can use a t-distribution, assuming normality, to study the behavior of sample mean. Another important point that has been raised by student (1909) is on the randomness of the sample drawn and importance of normal population. He notes that, if one is interested in studying the behavior of the sample mean when the underlying population is non-normal, using a random sample will make the asymptotic behavior of sample mean to approach normal curve sooner as compared that of a non-random sample. Here, a random sample means an independent and identically distributed (i.i.d.) sample. In a non-random sample, the observations may be correlated and this delays the original curve of the distribution approaching the normal curve. This is very important when the sample is drawn from a non-normal population. Hence, a researcher has to ensure that the sample is a random.

In his paper, Fisher (1925) studies the applications of t-distribution. The distribution of the t-test statistic, for one sample and two sample problems, has been given in the paper under the assumption of normality. When one is interested in studying the behavior of difference of two sample means, a t-distribution can be used with respective degrees of freedom. But, it has been assumed that the two population variances are equal. Welch (1947) studies the behavior of difference of two sample means, when the equality of variances assumption is violated. He proposes a correction for degrees of freedom, so that the sampling distribution of the test statistics is t-distribution again. In the case of two sample means, it has been assumed that the populations independently follow normal distribution. In fact Welch (1938) shows that the actual t-test for difference of two means with sum of two sample sizes minus two, will function well under equality of variances. It do not perform well under unequal variances. Bartlett (1935) examines the effect of non-normality on t-distribution and confirms that t-test can still be used under moderate departures of normality but not severe departures, for testing differences in means of equal number of observations. Other works on studying the performance of these test, under non-normality, can be found in Pearson ES et.al. (1929),

Eden and Yates (1933). The notion of sample size greater than 30 has been drawn from the Paper of Eden and Yates (1933). Gayen (1949) studies the effect of Skewness and kurtosis on the t-distribution and shows that if the population is non-normal then t-distribution performs better for a population distribution specified by a number of terms of the Edge worth series. This indicates that t-distribution can perform better for specific distributions other than normal. But, all these are based on few approximations and may not fit well for practical purposes.

Boneau (1960) looks at the robustness of t-distribution under the violation of assumptions. He notes that if the sample sizes are equal or nearly so, the assumed population distributions are of same shape or nearly so, then t-test is robust to the violations of assumptions. But, in reality getting always populations that satisfies the conditions may not be possible and what researchers look at is for a method that can be used with comfort and ease for any situation. He ends the discussion with a suggestion to the researchers to search for a more powerful test when the conditions are not met.

Taking the above view points and discussion, it is very important for any researcher to identify appropriate statistical testing procedures, which can address the hypotheses constructed on the population characteristics, appropriately with better power. From the above discussion one can also note that those procedures that are free of distributional assumption, may give better options for researchers to test the hypotheses. A class of methods that are being considered as alternative testing procedures come from non-parametric statistics. These techniques have been very well received by the research community and are being used frequently when the parametric tests fail. These methods are also used to give appropriate conclusions, when the sample sizes are small and when the distribution of the population is unknown. These methods have been developed by making few assumptions related to the distribution such as symmetric about zero etc. When the assumption of normality is violated due to the presence of extremes/outliers, mean will not provide a better understanding of the location. In such cases, median gives a better understanding about the location as median is calculated using the central portion of the data and not effected by the fluctuations in the sampling. A median test issued as an alternative to a one sample t-test, where the hypotheses are constructed on the median and conclusions are drawn on median. Similarly, Wilcoxon (1945) developed an alternative procedure to a two sample (dependent) paired t-test and widely used with the name "Wilcoxon signed rank test". Mann-Whitney (1947) gave a testing procedure as alternative to t-test (independent samples), Kruskal-Wallis (1952) proposed a test which can be used as an alternative to ANOVA. If one is interested in testing the hypotheses on more than two related samples, then Friedman (1937)'s test can be used. In almost all the non-parametric tests, the data are ranked based on the magnitude and the same

is used in the computation of the test statistic. Also, another important aspect related to non-parametric methods is that, the test statistic approaches normality for large sample sizes. Most of the statistical software has included add-ins to compute these methods.

There are few discussions on the choice of a Wilcoxon's test or a t-test, based on the power of the test. Hodges and Lehmann (1956) studies the efficiency of some non-parametric competitors of the t-test. Boneau (1962) compares the power of U and t-test. Blair and Higgins (1980) compares the power of Wilcoxon's test to that of student's t-test under various non-normal distributions. Blair et.al. (1980) discusses on the relative power of Mann-Whitney U test and t-test. Taking few of these arguments, we compare the power of t-test with that of Wilcoxon test, for the data sets considered in this study. If any of the assumptions of the non-parametric methods are not satisfied, then the methods may not produce reliable results. Also, checking the assumptions such as symmetry etc. is not so easy to check. The same issues as in the case of normality will recur again in this case also.

The limitations of the non-parametric tests, have given raise to the next generation statistical methods that are more robust in their performance. These methods have been named as "Robust Statistical Methods". These methods include estimation of location, construction of intervals, procedures that can be used to test the significance of location, one sample, two sample tests etc. They are important in data analysis, due to their insensitive behavior towards the violation of assumptions. Huber (1964) studies methods to estimate the location and presents different methods to estimate the location. Huber (1972) also reviews different robust statistical methods in literature. Among other aspects, these methods functional well even the data is contaminated. That is, the data has many outliers and extremes observations. Trimming is one process that helps one to handle outliers and extremes. Using a trimmed mean in the place of usual mean will help the researchers to understand the behaviour of location appropriately. But, care has to be taken when trimming, so that the trimmed mean will be again normal for sufficient sample size. Stigler (1973) noted that, trimming appropriately at specified percentile points will ensure that the trimmed mean will be asymptotically normal. If one considers this mean and use it in t-test, there is every chance of increasing the power of the test. Showing this is a part of our objectives.

Turkey and McLaughlin (1963) studies the advantage of using trimmed mean, when the data are contaminated and far from normality. They also provide exact critical values for the resulting tailored t using Monte Carlo computation. This paper also discusses the advantages of using a winsorized mean. Huber (1964) provides an estimator that is between sample mean and sample median, which is asymptotically robust among all translation invariant estimators. Bickel (1965) provides an alternative method of trimming and winsorizing which encompasses the efficient estimates provided by Huber and also generalizes to higher

dimension. Huber (1972) reviews the proposed estimates for location and gives three important classes of estimates. He also discusses methods that can be used to assess the robustness of the estimates (Jack knifing and Influence curve). Hampel (1973) provides a survey on robust estimates that address the practical issues. Stigler (1977) compares eleven estimators using real data sets. He shows that light trimming provides some improvement over the sample mean, but that the sample mean itself compares favorably with many recent proposals. Probal and Sengupta (1993) proposes a modified version of M-estimator. Oosterhoff (1994) shows that, in the symmetric case trimmed means are better estimates of location than sample medians, unless the error distributions are sharply peaked at the Centre. Wilcox and Keselman (2003) briefly review some of the more fundamental problems with conventional methods based on means; provide some indication of why recent advances, based on robust measures of location (or central tendency), have practical value; and describe why modern investigations dealing with non-normality find practical problems when comparing means, in contrast to earlier studies. David and Vikki (2008) discusses the importance of robust statistical methods and also explain how they can improve the accuracy and power of the research. All these discussions help us in understanding the importance of estimating the location accurately and using the same in the process of testing the related hypotheses. In our paper we consider few of the arguments on trimming and use trimmed mean to check the power of a t-test.

The most frequently used parametric and non-parametric methods depends on the assumption of normality. There are views on their robustness in asymmetrical populations. Ghurye (1949) discusses the use of student's t-test in an asymmetrical population. Boneau (1960) studies the effect of violations of assumptions underlying the t-test. In 1962's paper, he also compares the power of the U test and t-test. Blair (1980) looks at the power of t-test and Wilcoxon statistics and shows that t-test has higher power in the case of normal populations. Ratcliffe (1968) discusses on the sample size required to decrease the effect of non-normality and use a t-test with an appreciable power. Blair and Higgins (1980) compares the power of Wilcoxon's rank sum statistic with t-test and shows that Wilcoxon's statistic has higher power under non-normal populations. Micceri (1989) considers 440 large sample achievement and psychometric measures and shows that the underlying assumption of normality and statistics based on the same appear fallacious. Also he shows other statistics used under robustness also fail under non-normal conditions. Rand Wilcox (2009), in his book on modern statistical methods, discusses in detail the advantages of using trimmed means and tests-based on trimmed means. This book also gives a complete fundamental discussion on advantages of using modern statistical methods. Keselman (2004) considers the

case of small samples and provides an improved version of a t-test that controls Type-I error, through a transformation.

We end this section with a statement that there are several other papers and reports in the literature, which are not included in this work, which can give more understanding about robust methods and their advantages in data analysis.

3. Objectives of the Study

1. To understand parametric methods used to test characteristics of one and two populations that depend on the assumption of normality.
2. To identify the limitations of the parametric methods when the assumption of normality is not satisfied by the sample drawn.
3. Identify alternative non-parametric methods when the assumption of normality is not satisfied.
4. To Identify Robust methods based on trimming as alternative to Parametric and Non-Parametric Methods.

4. Methodology

In this section, we discuss the methodology used to achieve the objectives stated.

The study objectives are achieved using an exploratory study under which the existing literature has been thoroughly reviewed to understand standard parametric procedures and their limitations. Attempts have been made to understand the later developments that dealt with non-normality and with violation of other assumptions. The focus is mainly on the t-test (one sample and two sample problems) and empirically the consequences of violation of assumptions has been studied and presented in section 5.

We also looked at papers that dealt with alternative approaches to t-test and the limitations of these methods are listed in the form of a table. An empirical comparison of the power of t-test and its alternative methods have also been presented. Later, we have searched for robust statistical methods that can be used when either parametric or non-parametric methods fail to provide satisfactory answers to several research questions. The search includes methods on estimation as well as testing.

The quality of the data collected and appropriate tools chosen for analysis, determines the quality of the inferences drawn on the population under study. The data considered in each of the sections is a generated data based on the discussion. For example, the data required to discuss the power related issues under non-normality has been generated from a Pareto distribution. The reason being, Pareto distribution is heavy tailed and is used in finance to model the stock marker behavior. Our attempt also includes generating data suitable for non-parametric methods and robust statistical methods. This approach has been adopted to make the reader understand the importance of the assumptions, the quality of the data and their

effect on the statistical methods. One can take our work as reference and compare the actual data with the type of the data generated and take appropriate decision in selecting the statistical method.

5. Parametric Statistical Methods Depending on the Assumption of Normality and their limitations

The most frequently used tests in research are t-test, F-test, and Chi-square test. It is important to note that all the three depends on the assumption of normality. In this section we present the discussion on these tests in detail and also discuss their limitations. In the process, we also touch upon other testing procedures being used to test the assumptions associated with each of these tests.

5.1. T-test- Independent and Dependent Samples

In most of the research studies, the variable under study will be either a quantitative or a quantified variable. In either of the cases, the parameter under consideration is average, denoted by μ . The sample mean has been shown as a best estimator of the population mean and the same is also used to test the hypotheses constructed on μ . The sample mean, denoted by \bar{x} , is a random quantity and possess a probability structure, which has to be studied or assumed, before using \bar{x} in the analysis. If the population variance, denoted by σ^2 , is known then one can use normal distribution to model the behaviour of \bar{x} . If σ^2 is unknown, then student-t distribution can be used as an alternative to normal distribution. Student-t distribution models \bar{x} better if the population is either exactly normal or approximately normal. When a hypothesis is constructed on μ , \bar{x} is used in the construction of the test statistic and the asymptotic distribution of the test statistic also happens to be t-distribution. If the sample do not satisfy the assumption of normality, then the results are not reliable and the power may be low against an appropriate alternative value of μ .

A t-test is being used, for one sample and two sample problems. A two sample problem can be further classified as dependent and independent sample problem. In a one sample problem, the population under study is assumed to be normal and in an independent two sample problem, both the populations involved in the study are assumed to be normal. One has to test the normality of both the populations independently. Whereas, in a dependent two sample problem, the difference of post and prior is assumed to be normal.

In the case of an independent two sample problem, an additional assumption, homoscedasticity of variances, is made before applying the test. Sawilowsky and Blair (1992) have shown that t test is relatively robust to violation of the normality assumption when the following four conditions hold: (a) variances are equal, (b) sample sizes are equal, (c) sample sizes are 25 or more per group, and (d) tests are two-tailed. This combination of conditions is not reflective of most real data analytic circumstances, where unequal sample sizes are

common and variances are often heterogeneous. Sometimes, a transformation of the original random variable may help the researchers in using a t-test. But, use of transformations is not frequently suggested due to the following reasons (a) transformations often fail to restore normality and homoscedasticity; (b) they do not deal with outliers; (c) they can reduce power; (d) they sometimes rearrange the order of the means from what they were originally; and (e) they make the interpretation of results difficult, as findings are based on the transformed rather than the original data. Another important aspect with respect to independent sample t-test is that, when “equality of population variances” assumption is not violated, the degrees of freedom is calculated as the sum of two sample sizes minus two. Also, the estimate for common variance can be calculated by pooling both the sample variances. When this assumption is violated, Welch (1945) introduced an alternate formula for degrees of freedom. This formula takes into consideration the sample variances in the calculation of degrees of freedom. Also, the variances are estimated separately using the sample variances and the sum of both the estimates are taken into consideration while constructing the test statistic.

5.2. Empirical look at t-test and importance of assumption of normality

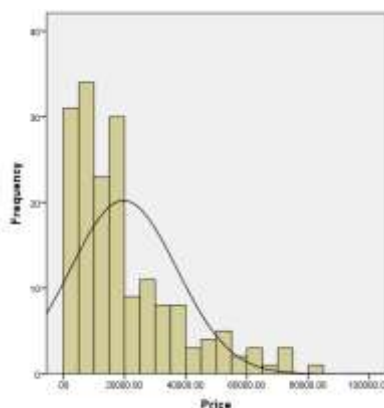
5.2.1 One sample t-test

We now look at an empirical example that takes into consideration of two different data sets. The first one satisfies the assumption of normality and the second do not satisfy the assumption of normality. In the first case, we show that the power will be high if the normality is satisfied and in the second case we show that the power will be low due to violation of normality. Data set one is prices of mobiles from different brands. The following table gives the summary statistics of the same.

Table: 1

Summary Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Price	176	1100.00	80500.00	19719.4886	17366.17112	1.384	1.423

Graph-1



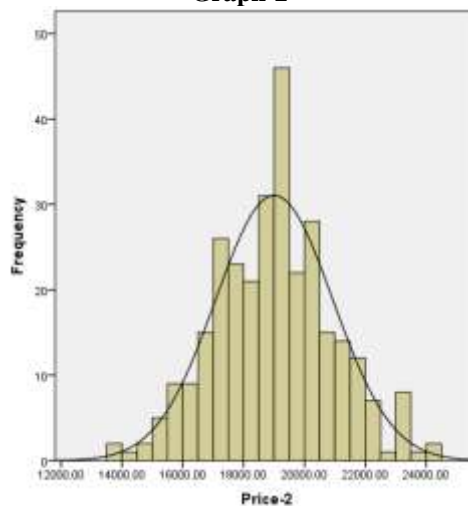
It is very obvious from the above graph that the variable price do not follow normal distribution. Moreover, from the histogram, one can infer that the variable follows a positively skewed distribution. Suppose that one wishes to test the hypothesis that the price of mobile is greater than or equal to Rs.23, 000. The result of a t-test is that the hypothesis is rejected at 5% level of significance (p -value= 0.0001). This indicates that the average price is less than Rs.23, 000. Now, if the test is of good power, then, a value less than Rs.23, 000 of average price, against null hypothesis value, should make the test reject null hypothesis with high power (close to one). For example, if one chooses the value as 22, 000 (taken from 95% confidence interval for population mean), then, the expectation is that the hypothetical value (Rs.23, 000) under null should be rejected with better power against this value. But, the power is 0.1874. This indicates that sample has constructed a test statistic that can reject the wrong null value correctly with a chance of 0.1874, which is very low. This is due to the violation of normality.

Now we consider the second data set. The variable under study is again the price of mobile. The following table give the summary of the data

Table: 2

Descriptive Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Price	300	13950.88	24283.73	19006.2856	1927.08010	.109	.035

Graph-2



From the above graph it is apparent that the variable price is approximately normal. Also, the Kolmogorov-Smirnov test reveals that the assumption of normality is satisfied by the data.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Price2 is normal with mean 19,006.286 and standard deviation 1,927.08.	One-Sample Kolmogorov-Smirnov Test	.816	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

This suffices the requirement stated in Student (1908) to apply the t-test. The null hypothesis value is average price is greater than or equal to Rs.19, 500. The hypothesis is rejected at 5% level of significance. The power of the test against an alternative null hypothesis value Rs.19, 000 (chosen from the confidence interval) is 0.9976. This is the difference between a testing procedure which has been constructed under normality and non-normality.

5.2.2 Independent sample t-test:

Now, we consider two situations to test the difference between two population means, when the samples have been drawn independently. In the first situation, we have considered the datasets that are generated from a normal distribution and in the second situation, datasets are generated from a non-normal distribution.

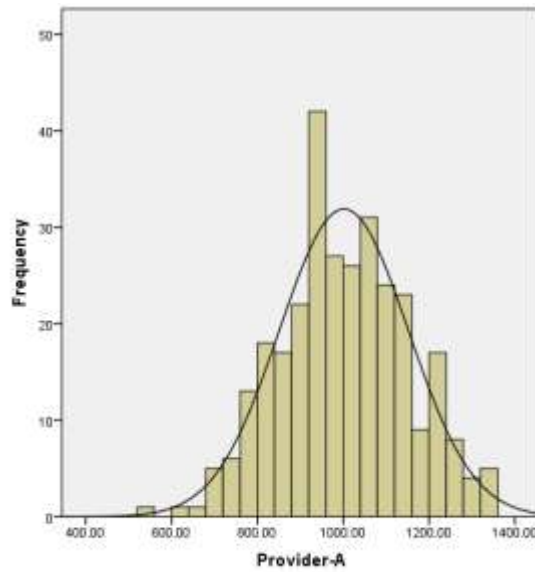
In the first situation, the data sets have been generated by considering the variable as registered mobile users for service provider A and service provider B. It is assumed that the users are independent and their experiences are independent. The time horizon considered is six months, across 300 outlets each.

The summary of the data sets generated are given in the following table

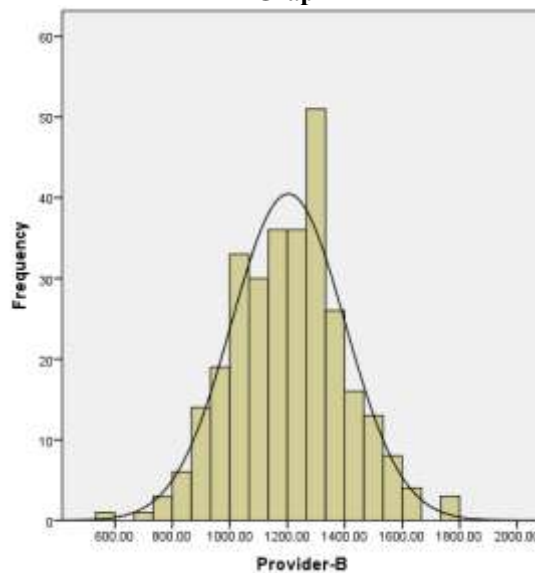
Table: 3

Summary Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Registered users-A	300	540.55	1354.64	1001.7029	150.14698	-.010	-.209
Registered users-B	300	551.78	1755.71	1203.5865	197.13716	.000	.168

Graph-3



Graph-4



The above graphs indicate that the data satisfy the assumption of normality. Also, the Kolmogorov-Smirnov test reveals that both the samples follow independently normal.

Table: 4

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Group1 is normal with mean 1,001.703 and standard deviation 150.15.	One-Sample Kolmogorov-Smirnov Test	.904	Retain the null hypothesis.
2	The distribution of Group2 is normal with mean 1,203.586 and standard deviation 197.14.	One-Sample Kolmogorov-Smirnov Test	.677	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Now, independent sample t-test is applied, to test the difference between the two population means and the following table gives the results.

Table: 5

Independent Samples Test					
	Levene's Test for Equality of Variances		t-test for Equality of Means		
	F	Sig.	t	d.f.	Sig. (2-tailed)
Equal variances assumed	17.547	.0001	-14.111	598	.0001
Equal variances not assumed			-14.111	558.553	.0001

One can note from the F-test that, equality of variances assumption is violated and using the Welch's (1968) formula, the degrees of freedom is calculated. From the test, it is apparent that the null hypothesis (there is no significant difference between the population means) is rejected. Note that, the data satisfies the assumption of normality and still the hypothesis is rejected. The difference between the two means is 201.9 and the standard error is 14.3068. This standard error is calculated by taking both the variances using the following formula

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

If sample variances are high, the standard error will be high and ultimately the distance between the sample mean and population mean will be high. For the example considered, the hypothesis is rejected due to high variance. That is, the data satisfies normality and still the hypothesis is rejected due to high variance. One expects that if the assumptions of the test are satisfied, the test may not reject the null hypothesis. If the hypothesis is rejected then, it is due to high variance, which increases the distance between the sample mean and population mean. This can be understood well from the formula for the confidence interval.

$$(\bar{X}_1 - \bar{X}_1) \pm t_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

This interval is said to contain the population mean and if the standard error is high then, the distance between the sample mean difference and the population mean difference will be high. For the example considered, the difference between the two sample means is -201.9 and the confidence interval for difference of two population means is (-230.0019, -173.7980) which is wider. One can note the difference between the sample mean and the two limits of the interval is high and this is because of high variance.

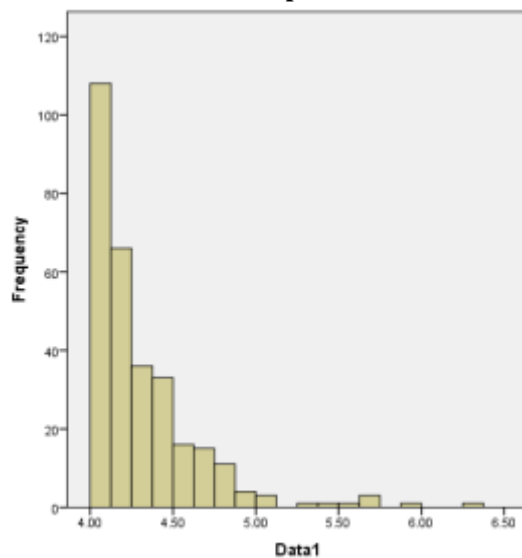
Now, we consider the second situation, where the data sets are generated from non-normal populations. The data has been generated from Pareto distribution and the variable considered is the returns of a given stock. Variable-1 is the returns of the stock A and

Variable-2 is the returns of the stock B. The following table gives the summary of the data sets generated.

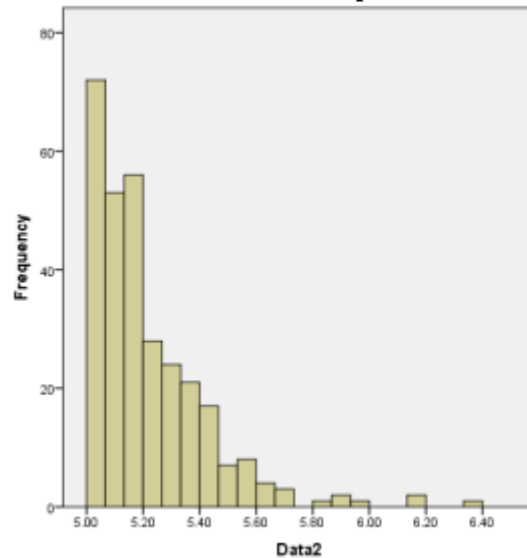
Table: 6

Summary Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Returns of Stock-A	300	4.00	6.37	4.3070	.33714	2.484	8.720
Returns of Stock-B	300	5.00	6.37	5.2166	.20658	2.008	5.931

Graph-5



Graph-6



The above graphs indicate that the returns do not follow the normal distribution. We apply the t-test to the above data and the following table gives the summary of the test.

Table: 7

Independent Samples Test					
	Levene's Test for Equality of Variances		t-test for Equality of Means		
	F	Sig.	t	df	Sig. (2-tailed)
Equal variances assumed	28.415	.0001	-39.848	598	.0001
Equal variances not assumed			-39.848	495.788	.0001

Since the data do not satisfy the assumption of normality, the results of the above test are not reliable. In this situation, the variances are low and one can expect that the hypothesis may not be rejected. But, the hypothesis is rejected because of non-normality. This means that, non-normality has played a role in rejecting the hypothesis. What does this mean? Suppose that the assumption of normality is satisfied by the data and the variances are low. Then, obviously the null hypothesis will not be rejected and the confidence interval will be narrow. For the situation-2, the confidence interval is (0.8651, 0.95485) and the sample mean

difference is 0.91. Unlike in situation-1, the difference between the sample mean and the limits of the interval is less. Hence, one can expect the null hypothesis to be not rejected. Still the hypothesis is rejected due to non-normality.

A summary of both the situations is presented in the following table

Table: 8

Situation-1	Situation-2
Mean of first variable is 1001.7 and that of second variable is 1203.58. The difference is 201.9.	Mean of first variable is 4.3 and that of second variable is 5.21. The difference is 0.91.
The data satisfies the assumption of normality.	The data do not satisfy the assumption of normality
The variance of first variable is 150.14 and that of second variable is 197.14. The standard error is 14.3068.	The variance of first variable is 0.3374 and that of second variable is 0.20658. The standard error is 0.02282.
Note that, the hypothesis is rejected as the sample variances are high and the standard error is also high.	Note that, the hypothesis is rejected though the sample variances are low.
The confidence interval for the population differences is (-230.0019, -173.7980).	The confidence interval for the population differences is (-0.95485, -0.8651).
The gap between the sample mean difference and the limits of the interval is high.	The gap between the sample mean difference and the limits of the interval is low.
The hypothesis in this case has been rejected due to high variance in spite of normality being satisfied by the sample data.	The hypothesis in this case has been rejected due to non-normality.
The power of the test for a value ($d = -180$) chosen within the confidence interval is high, as one expects	The power of the test for a value ($d = -0.9$) is high.

We provide further explanation on this. In statistical inference, a test statistic is usually developed under null hypothesis. In the examples considered, the null hypothesis is that the difference between the population means is zero and the sampling distribution of the test statistic is t-distribution. When the hypothesis is rejected in the first situation, it is due to higher variance. Whereas in the second situation, the hypothesis is rejected due to non-normality. In both the cases, one can choose an appropriate alternate value against the null hypothesis value, which is non-zero and note that the power will be high in both cases. This is because, now the sampling distribution of the test statistic is calculated under alternative hypothesis and the value chosen is sufficient to make the distribution of the test statistic asymptotically follow a t-distribution faster than in the previous case.

The last version of the t-test that is again more frequently used is the paired $-t$ -test. This test is mainly used when the samples are either dependent or matched. In this case, the difference between the two samples is assumed to follow a normal distribution. A similar discussion can be given as in the case of other version of t-test and the details are omitted.

Hence, from the above discussion, it very apparent that deviation from normality may lead to unreliable results and also either increases type-I error or decreases the power of the test.

6. Non-Parametric Statistical Methods

Non-parametric statistical methods are flexible to use when the structure of the population is unknown or the researcher is not sure whether the population is normal or non-normal. The researcher can make few basic assumptions regarding the population and carry out these tests. For example, one can assume that the distribution is symmetric. For most of the parametric tests, we have alternative non-parametric tests, which can be used for drawing valid conclusions, under the violation of assumptions of parametric tests. We only discuss the most frequently used non-parametric methods such as Wilcoxon signed rank test, Mann Whitney-Wilcoxon U test, which are alternatives to one sample and two sample problems.

6.1. Wilcoxon Signed Rank Test

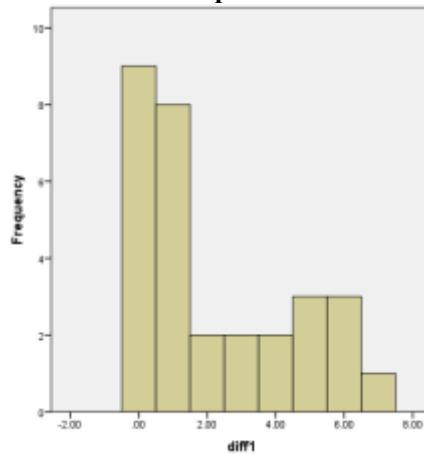
This test is used as an alternative to a paired t-test, to test whether the median difference between the pairs of observations is significantly different from zero or not. Under a paired t-test, average difference between the pairs is considered. But, under non-normal conditions or under the influence of extreme observations, one uses median to study the behavior of the location. The same is being used to construct the hypothesis related to the location and tested using Wilcoxon signed rank test. This test considers both the sign and magnitude of the differences. It can be understood easily and adopted in research with ease. It assumes that the data are paired, each pair is chosen randomly and independently, the data are measured at least on an ordinal scale. Once the data, satisfying these assumptions has been collected, the test can be applied to get the desired results. One can refer to Corder and Foreman (2009) to understand the procedure in a simple way. When the test is used, on the data drawn from the non-normal populations, one can expect that the power will be high as compared to that of a t-test for the same data.

We now consider a data set where the above mentioned statement with respect to the power of the test has been examined. The variable measured is the energy levels of the individuals before adopting a HR policy. The energy levels are measured as the additional time (in hours) they could work at the office. The following table gives the summary of the data set before and after implementation of the HR policy. The sample size is 30 and the duration of the study is one month.

Table: 9

Descriptive Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Before	30	1.00	13.00	5.2333	3.77545	.687	-.838
After	30	1.00	7.00	3.0333	1.73172	.842	.109

Graph-7



From the above graph, it is very clear that the data do not follow a normal distribution. The results of Kolmogorov-Smirnov test is as follows.

Table: 10

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of diff1 is normal with mean 2.200 and standard deviation 2.30.	One-Sample Kolmogorov-Smirnov Test	.029	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

We hence conclude that the data do not satisfy the assumption of normality. Applying a paired t-test will give the following results. The test considered here is a two-tailed test.

Table: 11

Paired Samples Test								
	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Before - After	2.2	2.2954	.4190	1.3428	3.0571	5.25	29	.0001

The null hypothesis (the average difference is zero) is rejected (p-value <0.05). The power of the test against the alternative value 1.4 is 0.89.

Now, we apply Wilcoxon signed rank test for the same data and calculate the power of the test for the same value. Note that, since the sample size is 30, one can use the normal approximation of the test statistic to calculate the power of the test.

Table: 12

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of diff1 equals 0.000	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

From the above table, one can note that the hypothesis is rejected. When the median difference chosen as 1, the power of the test is 0.977. This shows that the power of Wilcoxon in identifying the correct alternative difference between two median is high as compared that of the power calculated using a t-test. This is due to the effect of non-normality on the calculations.

Note that, the power is calculated based on the asymptotic behaviour of the test statistic, which is normal in this case. Also, one interesting aspect is that, since normality fails, the median difference selected in this case is different from that of mean difference selected in the previous case. If the population is normal, then the same difference yields a power which is similar. This will be demonstrated in the using the second data set that satisfies the assumption of normality.

6.2. Mann-Whitney U test

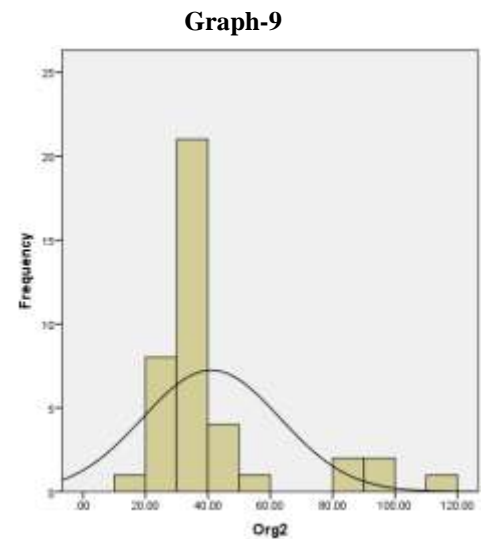
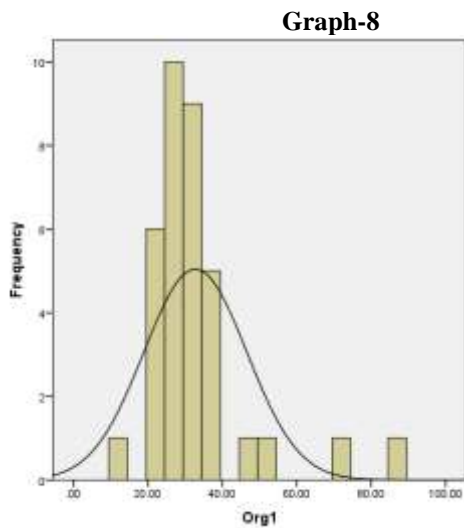
This test is used to test the difference between two independent population medians. This is an alternative testing tool to the independent samples t-test. The following are the assumptions of the test.

1. All the observations from both groups are independent of each other.
2. The responses are ordinal (i.e. one can at least say, of any two observations, which is the greater).
3. The distributions of both groups are equal under the null hypothesis, so that the probability of an observation from one population (X) exceeding an observation from the second population (Y) equals the probability of an observation from Y exceeding an observation from X . That is, there is a symmetry between populations with respect to probability of random drawing of a larger observation.
4. Under the alternative hypothesis, the probability of an observation from one population (X) exceeding an observation from the second population (Y) (after exclusion of ties) is not equal to 0.5. The alternative may also be stated in terms of a one-sided test, for example: $P(X > Y) + 0.5 P(X = Y) > 0.5$.

We consider another data set, where the variable under study is the number of likes to two independent ads advertised by two organizations. Here, we test the hypothesis that the distribution of the data is same across the two groups. The sample sizes are respectively 35 and 40. We first apply an independent sample t-test and then apply Mann-Whitney U test and compare the results. The following table gives the summary of the data set.

Table: 13

Descriptive Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Org1	35	12.00	85.00	32.7429	13.84652	2.444	7.065
Org2	40	14.00	110.00	41.1000	22.04983	2.065	3.349



From the above graphs, it is apparent that the data do not follow normal distribution and the following results of Kolmogorov-Smirnov test supports the same.

Table: 14

Hypothesis Test Summary				
	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Org1 is normal with mean 32.743 and standard deviation 13.85.	One-Sample Kolmogorov-Smirnov Test	.041	Reject the null hypothesis.
2	The distribution of Org2 is normal with mean 41.100 and standard deviation 22.05.	One-Sample Kolmogorov-Smirnov Test	.001	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

The results of the t-test are as follows.

Table: 15

	Levene's Test for Equality of Variances							
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Equal variances assumed	3.138	.081	-1.933	73	.057	-8.35714	-16.97529	.26101
Equal variances not assumed			-1.990	66.566	.051	-8.35714	-16.73966	.02538

The following is the result of the Mann-Whitney U test.

Table: 16

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Combined is the same across categories of Factor.	Independent-Samples Mann-Whitney U Test	.007	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

It is very interesting to see the results. T-test do not reject the hypothesis whereas Mann-Whitney U test rejects the hypothesis at the same level of significance 5%. Similar to the previous discussion, one can compute the powers and observe the performance of U-test in comparison with a t-test. The details are omitted.

Note that either of the tests, if the assumptions are violated then the results are not reliable as in the case of parametric methods. We now present few important robust methods as alternative tools to parametric and non-parametric methods, based on trimming.

7. Robust Statistical Methods

The first step in data analysis is estimating the key characteristics of the population with accuracy. In the presence of the outliers/extremes the estimates produced for location are misleading and also may affect the testing as well as inferences drawn. To overcome this problem, researchers have introduced the concept of trimming, which will reduce the effect of outliers/ extremes on the estimates of the location. Trimmed mean is used in the place of usual mean, so that the performance of t-test can be improved. But, the basic question is how and where the trimming has to done, so that the trimmed mean will be again asymptotically normal. Stigler (1972) gives a necessary and sufficient condition for the asymptotic normality of the trimmed mean. He proved that if trimming is done at proportions corresponding to

uniquely defined percentiles of the population distribution, then the limiting distribution is not normal. If one uses a trimmed mean, trimmed at different percentiles, then it may mislead and produce invalid results. This is true even with large samples. In our discussion, we use this and make an attempt to show that the power of a t-test will increase if one uses a trimmed mean in the place of usual mean.

7.1. Estimation and testing using trimmed mean

As mentioned in the beginning of this section, location can be estimated using trimmed mean. In this section, we present trimmed mean calculation using a sample data set. The variable measured is the number of visitors to a place, in a given year. The data has been generated from a heavy tailed distribution. Using this data, we compute the trimmed and apply t-test. We compare the powers before trimming and after trimming. The sample size is 100. The summary of the data set considered is as follows.

Table: 17

Descriptive Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Visitors	100	31.00	250.00	110.74	75.70152	.609	-1.198

Graph-10

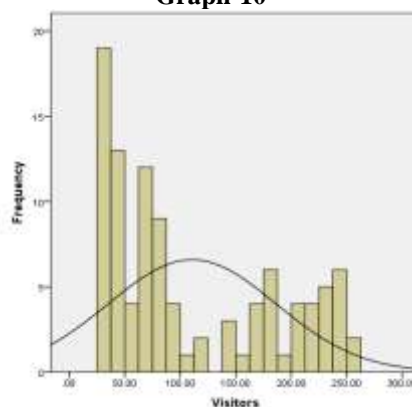


Table: 18

Hypothesis Test Summary

Null Hypothesis	Test	Sig.	Decision
The distribution of Data1 is normal with mean 110.740 and standard deviation 75.70.	One-Sample Kolmogorov-Smirnov Test	.001	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

From the above graph and tables, it is very clear that the data do not satisfy the assumption of normality also one can note that the variance is high. Applying a t-test may decrease the powers, as discussed earlier. The null hypothesis is that the population mean is less than or equal to 90 against it is greater than 90. The results of a t-test are as follows.

Table: 19
One-Sample Test

	Test Value = 90					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Data1	2.740	99	.007	20.74000	5.7192	35.7608

From the above table, one can note that the null hypothesis is rejected. The power of the test against an alternative value of 110 (which is close to sample mean) is **0.83**. As mentioned earlier, one reason for lower power could be non-normality.

Now, we apply the trimming process suggested by Stigler (1972). We choose different percentile point combinations as (10%, 70%), (10%, 80%), (20%, 80%), (20%, 70%) and study the behavior of the trimmed mean, in estimating the location and further testing on population mean. The following table gives the summary of test for normality

Table: 20
Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Ten_seventy is normal with mean 72.200 and standard deviation 35.33.	One-Sample Kolmogorov-Smirnov Test	.126	Retain the null hypothesis.
2	The distribution of Twen_seventy is normal with mean 79.620 and standard deviation 34.13.	One-Sample Kolmogorov-Smirnov Test	.097	Retain the null hypothesis.
3	The distribution of Ten_eighty is normal with mean 74.567 and standard deviation 37.29.	One-Sample Kolmogorov-Smirnov Test	.086	Retain the null hypothesis.
4	The distribution of Twen_eighty is normal with mean 96.733 and standard deviation 49.70.	One-Sample Kolmogorov-Smirnov Test	.021	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Note that, except the (20%, 80%), all the others satisfy the assumption of normality. We now proceed to test the hypothesis set previously.

Table: 21
One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Ten_seventy	60	72.2000	35.33031	4.56112
Twen_seventy	50	79.6200	34.13353	4.82721
Ten_eighty	60	74.5667	37.29333	4.81455

Table: 22

One-Sample Test						
	Test Value = 90					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Ten-seventy	-3.903	59	.000	-17.80000	-26.9268	-8.6732
Twenty-seventy	-2.150	49	.036	-10.38000	-20.0806	-.6794
Ten-eighty	-3.206	59	.002	-15.43333	-25.0672	-5.7994

The hypothesis has been rejected as in the previous case (non-normality). The powers of the test against the alternative value of 110 are respectively **0.9956, 0.9914, and 0.9920**. This is the difference between a trimmed mean and an untrimmed mean and also the effect of non-normality of the results. A similar discussion also can be made with respect to independent samples and paired samples.

8. Suggestions and Conclusion

Taking into consideration the above arguments, we suggest the researchers the following

1. Ensure that the normality assumption is satisfied by the sample.
2. Estimate the sample size appropriately, so that when a parametric method is used, the asymptotic behaviour of sample mean is preserved.
3. While using the t-test, either for a one sample or two sample ensure that all the assumptions associated with the test are satisfied.
4. Calculate the confidence interval for each parameter and choose a value from within the same as an estimate of the parameter.
5. Compute the power of the test and ensure that the power will at least same as confidence level.
6. Use non-parametric methods if any of the assumptions are not satisfied by the data.
7. If one is interested in using sample mean as an estimator of the location, then ensure that the sample is free of outliers/extremes. If the sample contains outliers, then using a trimmed mean will give better results.
8. Using a t-test using a trimmed mean will give better power and also better estimate to the location.

Finally, we conclude that using appropriate statistical test either for one sample or two samples will give an opportunity to the researcher to draw valid and reliable inferences with better power. Using either non-parametric or robust methods will increase the power of the test. One has to compute the power of the test along with other estimates. If a hypothesis is rejected, then it can also happen due to non-normality or due to high variance. A high

variance leads to non-normality and non-normality makes a t-test produce unreliable results. Hence, one has to take care while drawing the sample, so that the sampling error can be reduced and this may lead to normality. The objectives of the study are achieved using a data that has been generated. This may be a limitation of the study.

References

Patel, J., K., & Read C., B., 1982, '*Handbook of the normal distribution*', Marcel Dekker, INC.

Student 1908, 'The Probable error of mean', *Biometrika*, Vol. 6, No. 1, pp. 1-25.

Student 1909, 'The Distribution of the Means of Samples which are Not Drawn at Random', *Biometrika*, Vol. 7, No. ½, pp. 210-214.

Fisher, R., A., 1925, 'Applications of "student's" distribution', *Metron*, Vol. 5, pp. 90-104.

Welch, B., L., 1938, 'The significance of the difference between two means when the population variances are unequal', *Biometrika*, Vol. 29, No. ¾, pp. 350-362.

Welch, B., L., 1947, 'The Generalization of 'Student's' Problem when Several Different Population Variances are involved', *Biometrika*, Vol. 34, No. ½, pp. 28-35.

Bartlett, M., S., 1935, 'The Effect of Non-Normality on the t Distribution', *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 31, No. 2, pp. 223 – 231.

Pearson, E., S., & Adyanthāya, N., K., 1929, 'The Distribution of Frequency Constants in Small Samples From Non-Normal Symmetrical and Skew Populations', *Biometrika*, Vol. 21, No. ¼, pp. 259-286.

Eden, T., & Yates, F., 1933, 'On the validity of Fisher's z test when applied to an actual example of Non-normal data', *The Journal of Agricultural Science*, Vol. 23, No. 01, pp. 6-17.

Gayen, A., K., 1949, 'The Distribution of 'Student's t in Random Samples of any Size Drawn from Non-Normal Universes', *Biometrika*, Vol. 36, No. ¾, pp. 353-369.

Boneau, A., C., 1960, 'The effects of violations of assumptions underlying the t-test' *Psychological bulletin*, Vol. 37, No. 1, pp. 49-64.

Wilcoxon, F., 1945, 'Individual Comparisons by Ranking Methods', *Biometrics Bulletin*, Vol. 1, No. 6. , pp. 80-83.

- Mann, H. B., & Whitney, D., R., 1947, 'On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other', *The Annals of Mathematical Statistics*, Vol. 18, No. 1, pp. 50-60.
- Kruskal, W., H., & Wallis, A., W., , 1952, 'Use of Ranks in One-Criterion Variance Analysis', *Journal of the American Statistical Association*, Vol. 47, No. 260, pp. 583-621.
- Friedman, M., 1937, 'The use of ranks to avoid the assumption of normality implicit in the analysis of variance', *Journal of the American Statistical Association*, Vol. 32, No. 200, pp. 675-701.
- Hodges Jr, J., L., & E. L. Lehmann, E., L., 1956, 'The Efficiency of Some Nonparametric Competitors of the t -Test', *The Annals of Mathematical Statistics*, Vol. 27, No. 2, pp. 324-335.
- Blair, C., R., & Higgins, J., J., 1980, 'A comparison of the power of Wilcoxon's rank-sum statistic to that of Student's t statistic under various non-normal distribution', *Journal of Educational Statistics*, Vol. 5, No. 4, pp. 309-335.
- Boneau, A., C., 1962, 'A Comparison of the power of the U and t tests', *Psychological Review*, Vol. 69, No. 3, pp. 246-256.
- Blair, C., R., Higgins, J., J., & Smitley, W., D., S., 1980, 'On the relative power of the U and t tests', *British Journal of Mathematical and Statistical Psychology*, Vol. 33, pp. 114-130.
- Huber, P., J., 1964, 'Robust estimation of a location parameter', *Annals of Mathematical Statistics*, Vol. 35, pp. 73-101.
- Huber, P., J., 1972, 'Robust statistics: A review', *Annals of Mathematical Statistics*, Vol. 43, pp. 1041-1067.
- Stigler, S., M., 1973, 'The asymptotic distribution of trimmed mean', *The Annals of Statistics*, Vol. 1, No. 3, pp. 472-477.
- Tukey, J., W., & McLaughlin, D., H., 1963, 'Less Vulnerable Confidence and Significance Procedures for Location Based on a Single Sample: Trimming / Winsorization' *Sankhyā: The Indian Journal of Statistics*, Series A, Vol. 25, No. 3, pp. 331-352.
- Bickel, P., J., 1965, 'On some robust estimates of location', *Annals of Mathematical Statistics*, Vol. 36, pp. 847-58.

- Hampel, F., 1973, 'Robust Estimation: A Condensed Partial Survey', *Z. Wahrscheinlichkeitstheorie verw Geb*, Vol. 27, pp.87- 104.
- Stigler, S., M., 1977, 'Do Robust Estimators Work with Real Data?', *The Annals of Statistics*, Vol. 5, No. 6, pp. 1055-1098.
- Probal, C., & SenGupta, D., 1993, 'A note on robust estimation of location', *Statistics & Probability Letters*, Vol. 18, pp. 241-244.
- Oosterhoff, J., 1994, 'Trimmed mean or sample median?', *Statistics & Probability Letters*, Vol. 20, pp. 401-409.
- Wilcox, R., R., & Keselman, H., J., 2003, 'Modern Robust Data Analysis Methods: Measures of Central Tendency', *Psychological Methods*, Vol. 8, No. 3, pp. 254–274.
- David, M., E., & Vikki, M., M., 2008, 'Modern Robust Statistical Methods: An Easy Way to Maximize the Accuracy and Power of Your Research', *American Psychologist*, Vol. 63, No. 7, pp. 591– 601.
- Ghurye, S., G., 1949, 'On the Use of Student's t-Test in an Asymmetrical Population', *Biometrika*, Vol. 36, No. ¾, pp. 426-430.
- Ratcliffe, J., F., 1968, 'The Effect on the t Distribution of Non-Normality in the Sampled Population', *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Vol. 17, No. 1, pp. 42-48.
- Miceeri, T., 'The Unicorn, the Normal Curve, and Other Improbable Creatures', *Psychological Bulletin*, Vol. 105, No. 1, pp. 156-166.
- Wilcox, R., R., 2009, '*Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy*', 2nd Edition, Springer, New York.
- Sawilowsky, S., S., & Blair, C., R., 1992, 'A More Realistic Look at the Robustness and Type II Error Properties of the t Test to Departures From Population Normality', *Psychological Bulletin*, Vol. 111, No. 2, pp. 352-360