THE INTERNATIONAL JOURNAL OF BUSINESS & MANAGEMENT

The Role of Normality on One Sample and Two sample T-Tests

Srilakshminarayana Gali
Assistant Professor
Shri Dharmasthala Manjunathshwara Institute for Management Development,
Chamundi Hill Road, Siddhartha Nagar, Mysore, Karnataka, India

Abstract:
The present paper is an attempt to bring to the notice of the researchers, the role of normality inusing a t-test and show them empirically the effect of violation of normality on the inferences drawn on population mean, based on a t-test. The main discussion is on power of the test and how power decreases due to violation of normality. The data considered has been generated in accordance with the objectives and discussion has been provided accordingly. One can use this paper for a sample data which falls under similar situations. We also discuss the advantages of trimming in non-normal situations and how this can improve the power of the t-test. This paper is mainly written to caution the researchers on using a t-test under non-normal environments and introduce trimming in t-test.

Keywords: *Normality, power of the test, t-distribution, t-test, trimmed mean*

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 another reason, to use either normal or lognormal model in most of the research studies. Another important limitation is that, only for a 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 behavior 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.

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.

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 a sufficient 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, through 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 for 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 the consequences that arise due to violation of the same.

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 the 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 question is on the 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 the 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 a 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 behaviour 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, researchers 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 (1938) 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. Infact, 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 does 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.

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 behavior 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.

There are views on the robustness of t-test in asymmetrical populations. Ghurye (1949) discusses the use of student's t-test in an asymmetrical population. Ratcliffe (1968) discusses on the sample size required to decrease the effect of non-normality and use a t-test with an appreciable power. 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.

The references of the present paper have been chosen based on the objectives of the study. This paper is mainly meant to look at the historical developments and hence lists few paper of that time accordingly.

3. Objectives of the Study

1. To check the sensitivity of a t-test to normality assumption in terms of the power of the test.
2. To present the advantages of having a sample that satisfies the assumption of normality-one sample and two sample.
3. To discuss the process of trimming and present the effect of trimming on the performance of a t-test.

4. Methodology

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. An empirical comparisons of the power of t-test under normality and non-normality using the data generated has been considered. 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. T-Test –One Sample and Two Sample Problems: An Empirical Study

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 byσ^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. 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 (1938) 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.

5.1. 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.

Summary Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Price	176	1100	80500	19719.488	17366.17	1.384	1.423

Table 1

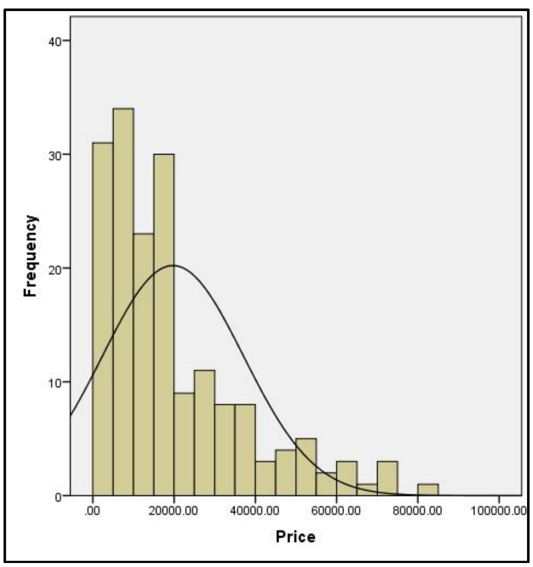

Figure 1

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. The following data gives the power of the t-test for different alternative values of population mean

µ1	Power	µ1	Power	µ1	Power
17500	0.994152782	19000	0.91867216	20000	0.737903223
17600	0.992796323	19100	0.906668247	20500	0.60095589
18000	0.984169146	19200	0.893391371	21000	0.450036281
18200	0.977189346	19500	0.845464748	21500	0.306145951
18500	0.961927259	19800	0.784981362	22000	0.187429629

Table 2

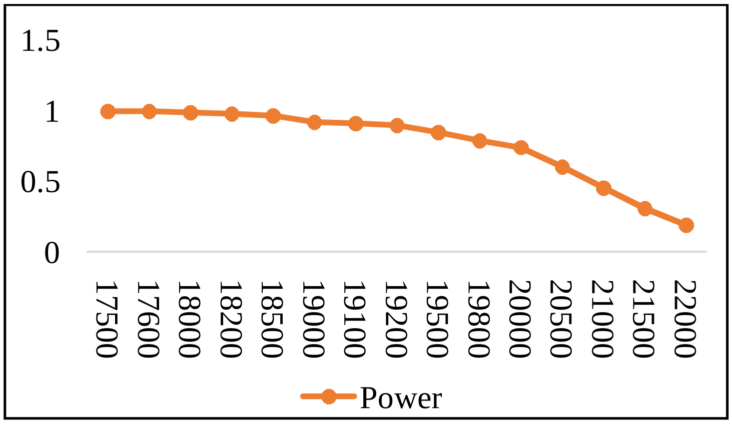

Figure 2

One can note the decrease in the power of the test. Even for values close to mean, the power less than 90%. All 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 gives the summary of the data

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

Table 3

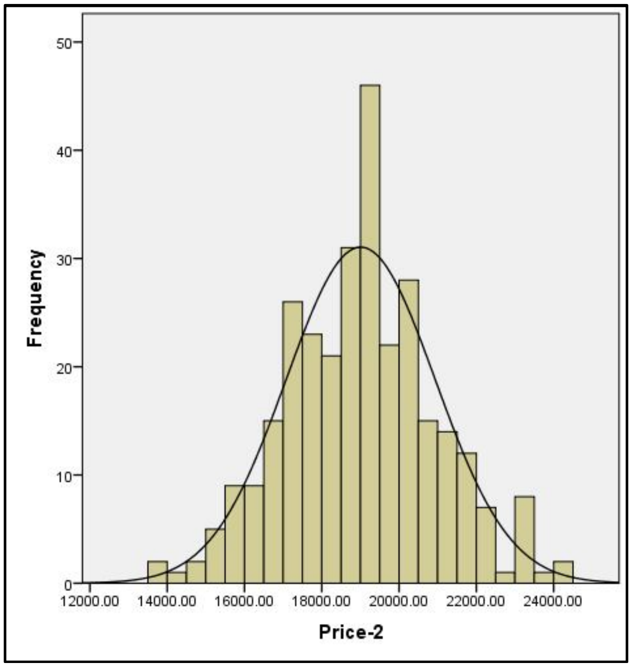

Figure 3

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	.616	Retain the null hypothesis.

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

Table 4

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. This suffices the requirement stated in the 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. The following table gives the power for different alternative values of population mean. One can note that the power increases for values close to the sample mean and also do not fluctuate more as in the case of non-normal sample.

µ1	Power	µ1	Power	µ1	Power
18800	0.999997409	19000	0.997618935	19225	0.794054974
18850	0.999981787	19100	0.973656937		
18900	0.999890904	19150	0.932122128		
18950	0.999445671	19200	0.851893957		

Table 5

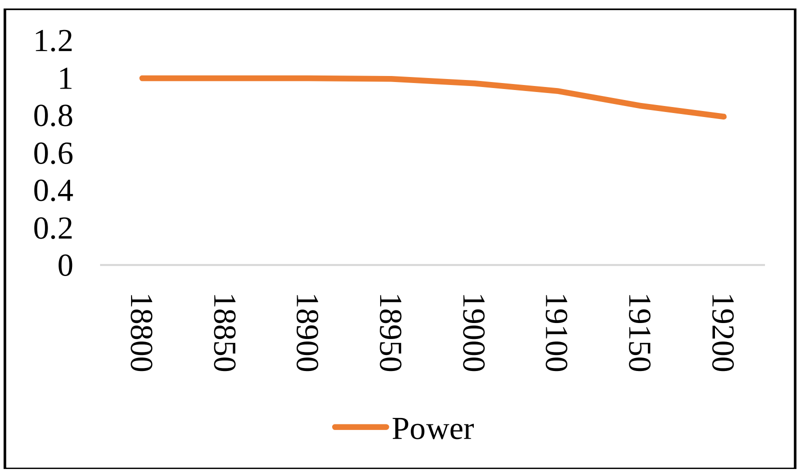

Figure 4

5.2. 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

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

Table 6

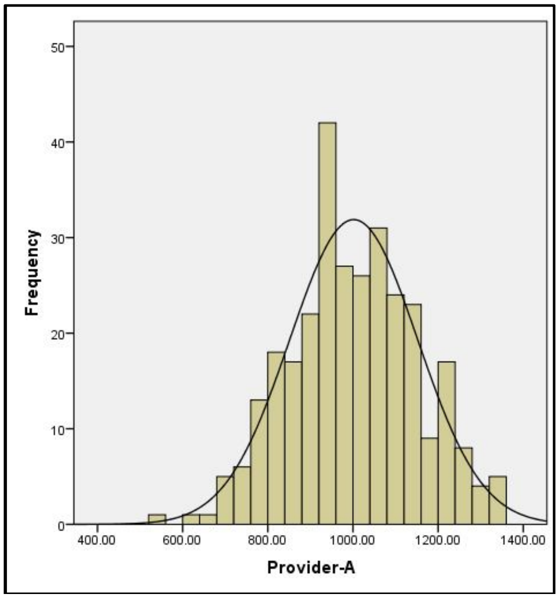

Figure 5

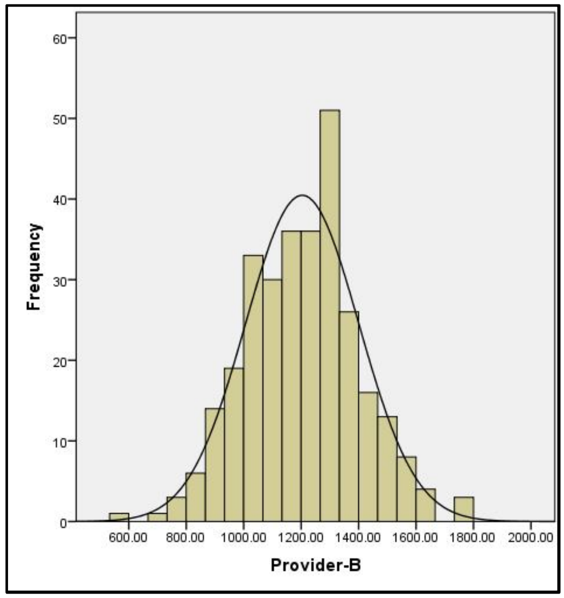

Figure 6

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.

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.

Table 7

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

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

Table 8

One can note from the F-test that, equality of variances assumption is violated and using the Welch's (1938) 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.

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

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.

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

Table 9

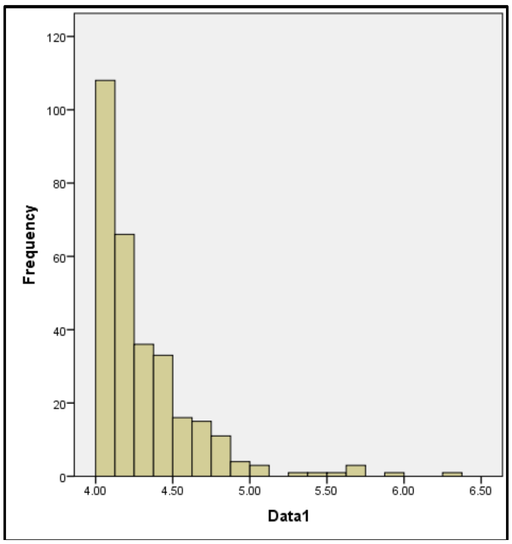

Figure 7

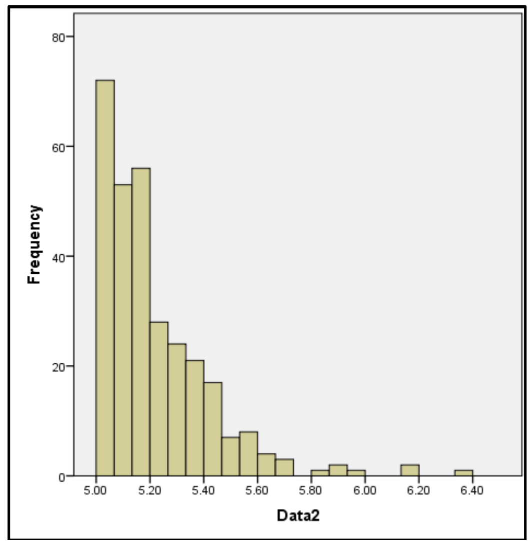

Figure 8

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.

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

Table 10

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

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= -185) chosen within the confidence interval is high, as one expects	The power of the test for a value (d= -0.9) is high.

Table 11

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. 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.

5.3. T-test Based on Trimming

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 be 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.

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.

Descriptive Statistics

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

Table 12

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	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.

Table 13

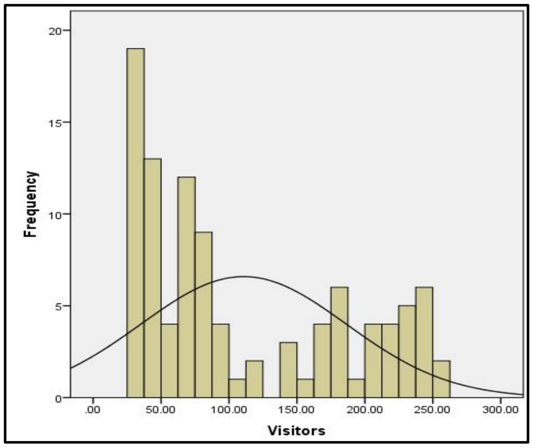

Figure 9

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 men is less than or equal to 90 against it is greater than 90. The results of a t-test are as follows.

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

Table 14

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

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.

Table 15

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

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

Table16

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

Table 17

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. The details are omitted.

6. Suggestions and Conclusion

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

1. While using the t-test, either for a one sample or two sample ensure that all the assumptions associated with the test are satisfied.
2. Estimate the sample size appropriately, so that when a parametric method is used, the asymptotic behaviour of the sample mean is preserved.
3. Calculate the confidence interval for each parameter and choose a value from within the same as an estimate of the parameter. Compute the power of the test and ensure that the power will at least same as the confidence level.
4. If one is interested in using the 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.
5. A t-test using a trimmed mean will give better power and also better estimates the location.

Finally, we conclude that using appropriate statistical test either for one sample or two sample 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.

7. References

i. Bartlett, M., S., (1935). The Effect of Non-Normality on the t Distribution. Mathematical Proceedings of the Cambridge Philosophical Society, 31 (2), pp. 223 – 231.
ii. Boneau, A., C., (1960). The effects of violations of assumptions underlying the t-test. Psychological bulletin, 37 (1), pp. 49-64.
iii. 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, 23 (1), pp. 6-17.
iv. Fisher, R., A., (1925). Applications of “student’s” distribution. Metron, 5, pp. 90-104.
v. Gayen, A., K., (1949). The Distribution of `Student's t in Random Samples of any Size Drawn from Non-Normal Universes. Biometrika, 36(¾), pp. 353-369.
vi. Ghurye, S., G., (1949). On the Use of Student's t-Test in an Asymmetrical Population. Biometrika, 36(¾), pp. 426-430.
vii. Miceeri, T., (1989).The Unicorn, The Normal Curve, and Other Improbable Creatures. Psychological Bulletin, 105(1), pp. 156-166.
viii. Pearson, E., S., & Adyanthāya, N., K., (1929). The Distribution of Frequency Constants in Small Samples from Non-Normal Symmetrical and Skew Populations. Biometrika, 21(¼), pp. 259-286.
ix. 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), 17(1), pp. 42-48.
x. 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. PsychologicalBulletin, 111 (2), pp. 352-360.
xi. Stigler, S., M., (1973). The asymptotic distribution of trimmed mean. The Annals of Statistics, 1(3), pp. 472-477.
xii. Student (1908). The Probable error of mean. Biometrika, 6(1), pp. 1-25.
xiii. Student (1909).The Distribution of the Means of Samples which are Not Drawn at Random. Biometrika, 7(½), pp. 210-214.
xiv. Welch, B., L., (1938). The significance of the difference between two means when the population variances are unequal. Biometrika, 29(¾), pp. 350-362.
xv. Wilcox, R., R., (2009). Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy. 2nd Edition, Springer, New York.