STATISTICS
THESIS

SIMULATION STUDY ON NON-PARAMETRIC TEST STATISTICS

By, Nilabjanayan Bera
STUG/154/16
June 2020

Contents

1 INTRODUCTION

The choice of statistical test has a profound impact on the interpretation of data. Understanding this choice is important to draw inference about population characteristics. The question often arises whether to use non-parametric or parametric tests. Using non-parametric tests instead of parametric tests brings about two questions:

1. What happens if the non-parametric test is used when the parametric assumptions are met?
2. What happens when the parametric assumptions are not met?

To answer these questions, one must first discuss the underlying goal of the study. Usually one is interested in measures of location such as the mean. One can test if the treatment (experimental condition) has an effect (location shift) on the population under study. In the non-parametric case, equivalents the location statistic is the median.

The assumptions for the non-parametric test are weaker than those for the parametric test, and it has been stated that when the assumptions are not met, it is better to use the non-parametric test. However, real data are rarely exactly normal. Does this mean that one should never use the t test? In many data sets seen in various fields in the real world,there often exist several observations that differ from the others, the so-called outliers. One must also then consider what is the best summary statistic for central tendency. That is, there should be some concept of robustness to assess the properties of the estimators themselves. Robustness, in one sense, refers to the insensitivity of the estimator to outliers or violations in underlying assumptions. In the sample if we let any one of the observations get arbitrarily large, the mean will become arbitrarily large. The median, which is commonly used when data are skewed or there exist outliers, is defined as the central value in a distribution where above and below lie an equal number of values.

Now we are interested in using various parametric and non-parametric tests for both **Single Sample Location Problem** and **Two Sample Location Problem** over the same data sets and compare their power curves to study their effectiveness in various problems.

2 SOME IMPORTANT TERMS

2.1 Location Parameter

Location parameter is the abscissa of a location point and may be a measure of central tendency, such as mean,median etc. Alternatively In statistics, a location parameter of a probability distribution is a scalar- or vector-valued parameter x_0, which determines the **"location"** or **shift** of the distribution. A location parameter can also be found in families having more than one parameter, such as location–scale families. In this case, the probability density function or probability mass function will be a special case of the more general form:

$$f_{x_0,\theta}(x) = f_\theta(x - x_0)$$

where x_0 is the location parameter, θ represents additional parameters, and f_θ is a function parametrized on the additional parameters. To understand it let take an example and explain it. Suppose $f(x|\mu, \sigma)$ is a normal distribution with mean μ and variance σ i.e. $N(\mu, \sigma)$ and $f(x|\mu, \sigma)$ is defined as,

$$\begin{aligned} f(x|\mu, \sigma) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(- \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\Big) \\ &= f(x - \mu|\sigma) \end{aligned}$$

So here μ is the location parameter.

2.2 Parametric Tests

A parametric test is a hypothesis testing procedure based on the assumption that observed data are distributed according to some distributions of well-known form (e.g., normal, Bernoulli, and so on) up to some unknown parameter(s) on which we want to make inference (say the mean, or the success probability etc.).

Example: The normal family of distributions all have the same general shape and are parameterized by mean and standard deviation. That means that if the mean and standard deviation are known and if the distribution is normal, the probability of any future observation lying in a given range is known. Suppose that we have a sample of 99 test scores with a mean of 100 and a standard deviation of 1. If we assume all 99 test scores are random observations from a normal distribution, then we predict there is a 1% chance that the 100^{th} test score will be higher than 102.33 (that is, the mean plus 2.33 standard deviations), assuming that the 100^{th} test score comes from the same distribution as the others. Parametric statistical methods are used to compute the 2.33 value above, given 99 independent observations from the same normal distribution.

2.3 Non-parametric Tests

This is a technique of testing that do not rely on data belonging to any particular parametric family of probability distributions.
This is a distribution free method, which do not rely on assumptions that the data are drawn from a given parametric family of probability distributions. As such it is the opposite of parametric statistics.In this testing procedure it is not possible to assume the explicit functional form of the

parent distributions. Here statistics are based on fractiles, ranks, grades, concordance, discordance etc.
We mainly discuss here about different type of non-parametric test statistic and focus on their nature like power-curve and consistency.

2.4 QQ-Plot

In statistics, a Q–Q (quantile-quantile) plot is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. First, the set of intervals for the quantiles is chosen. A point (x, y) on the plot corresponds to one of the quantiles of the second distribution (y-coordinate) plotted against the same quantile of the first distribution (x-coordinate). Thus the line is a parametric curve with the parameter which is the number of the interval for the quantile.

If the two distributions being compared are similar, the points in the Q–Q plot will approximately lie on the line y = x. If the distributions are linearly related, the points in the Q–Q plot will approximately lie on a line, but not necessarily on the line y = x. Q–Q plots can also be used as a graphical means of estimating parameters in a location-scale family of distributions.

A Q–Q plot is used to compare the shapes of distributions, providing a graphical view of how properties such as location, scale, and skewness are similar or different in the two distributions. Q–Q plots can be used to compare collections of data, or theoretical distributions. The use of Q–Q plots to compare two samples of data can be viewed as a non-parametric approach to comparing their underlying distributions. A Q–Q plot is generally a more powerful approach to do this than the common technique of comparing histograms of the two samples, but requires more skill to interpret. Q–Q plots are commonly used to compare a data set to a theoretical model. This can provide an assessment of "goodness of fit" that is graphical, rather than reducing to a numerical summary. Q–Q plots are also used to compare two theoretical distributions to each other. Since Q–Q plots compare distributions, there is no need for the values to be observed as pairs, as in a scatter plot, or even for the numbers of values in the two groups being compared to be equal.

2.5 Shapiro-Wilk Test

The Shapiro–Wilk test is a test of normality in frequentist statistics. It was published in 1965 by Samuel Sanford Shapiro and Martin Wilk.

Theory: The Shapiro–Wilk test tests the null hypothesis that a sample $x_1, x_2, x_3, \ldots, x_n$ came from a normally distributed population. The test statistic is

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

where $x_{(i)}$ (with parentheses enclosing the subscript index i; not to be confused with x_i) is the ith order statistic, i.e., the ith-smallest number in the sample; $\bar{x} = (x_1 + \cdots + x_n)/n$ is the sample mean. The coefficients a_i are given by:

$$(a_1, \ldots, a_n) = \frac{m^\intercal V^{-1}}{C}$$

where C is a vector norm.

$C = \|V^{-1}m\| = (m^\intercal V^{-1} V^{-1} m)^{1/2}$ and the vector m,

$m = (m_1, \ldots, m_n)^\intercal$ is made of the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution; finally, V is the covariance matrix of those normal order statistics.

There is no name for the distribution of w. The cutoff values for the statistics are calculated through Monte-Carlo simulations.

Interpretation: The null-hypothesis of this test is that the population is normally distributed. Thus, if the p value is less than the chosen alpha level, then the null hypothesis is rejected and there is evidence that the data tested are not normally distributed. On the other hand, if the p value is greater than the chosen alpha level, then the null hypothesis that the data came from a normally distributed population can not be rejected (e.g., for an alpha level of .05, a data set with a p value of less than .05 rejects the null hypothesis that the data are from a normally distributed population).

Like most statistical significance tests, if the sample size is sufficiently large this test may detect even trivial departures from the null hypothesis (i.e., although there may be some statistically significant effect, it may be too small to be of any practical significance); thus, additional investigation of the effect size is typically advisable, e.g., a Q–Q plot in this case.

2.6 Kolmogorov-Smirnov Test

Suppose $X_1, X_2, X_3, \ldots, X_n$ are n random sample independently drawn from the continuous distribution function F then, if we want to test

$$H_0: \ F(x) = F_0(x) \quad vs$$
$$H_1: \ F(x) \neq F_0(x)$$

Kolmogorov Smirnov Test sataistic is given by,

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)|$$

Where $F_n(x)$ is the empirical distribution function for n independent and identically distributed ordered observations X_i. It can be defined as,

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[-\infty,x]}\big(X_i\big) \text{ i.e.}$$

$$= \begin{cases} 0 & x < X_{(1)} \\ \frac{1}{n} & X_{(1)} \leq x < X_{(2)} \\ \vdots & \vdots \\ \frac{i}{n} & X_{(i)} \leq x < X_{(i+1)} \\ \vdots & \vdots \\ 1 & x \geq X_{(n)} \end{cases}$$

Now, we have large sample test,

$$\lim_{n\to\infty} P(D_n \leq \frac{3}{\sqrt{n}}) = L(z)$$

$$\text{where, } L(z) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1} \exp\big(-2i^2z^2\big)$$

Reject H_0 in favour of H_1 for small values of D_n.

3 ONE SAMPLE LOCATION TESTS

3.1 Sign Test:

3.1.1 Description:

Let $(X_1, X_2, X_3, \ldots, X_n)$ be a independently drawn random sample from a population having distribution function $F(x - \theta)$.

$$F \in \Omega_0 = \big\{F, \text{ F is absolutely continuous, } F(0) = \frac{1}{2}\big\}$$

Suppose to test:

$$H_0 : \theta = 0 \text{ vs}$$
$$H_1 : \theta > 0$$

Here both H_0 and H_1 are nonparametric hypothesis. [If $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$ then replace observation as $X_i - \theta_0 = Y_i$ instead of X_i.]
Define the test statistic as

$$S = \sum_{i=1}^{n} I(X_i)$$

where,

$$I(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & o.w. \end{cases}$$

Reject H_0 in favour of H_1 for large values of the statistic.

$$S \sim \ Bin(n, p)$$

where,

$$p = P(X_1 > 0) = 1 - P(X_1 < 0) = 1 - F(-\theta)$$

Here as we can see that p depends upon F and θ. But under the null $p = P(X_1 > 0) = \frac{1}{2}$, and S is independent of the underlying distribution. And,

$$S \sim \ Bin(n, \frac{1}{2}) \ ; [\text{ Under } H_0]$$

Hence S is distribution free under H_0. Reject H_0 if $S > S_\alpha$, where S_α is the upper α-point from $Bin(n, \frac{1}{2})$.

3.1.2 Size α test:

The test function is defined as,

$$\phi = \begin{cases} 1 & if\ S > k \\ a & if\ S = k \\ 0 & if\ S < k \end{cases}$$

where $a \in [0, 1)$ and k ($k \in \mathbb{I}$) is such that

$$P_{H_0}(S > k) \leq \alpha < P_{H_0}(S \geq k)$$

Thus if this is size α test then,

$$E_{H_0}(\phi) = \alpha$$
$$\Rightarrow P_{H_0}(S > k) + a\ P_{H_0}(S = k) = \alpha$$

3.1.3 Large Sample test:

Define, $\tilde{S} = \frac{S}{n}$. So,

$$Z = \frac{\sqrt{n}\ \left(\tilde{S} - \frac{1}{2}\right)}{\frac{1}{2}} \overset{\mathrm{H_0}}{\sim} AN(0, 1)$$

Reject H_0 at size α if the observed $Z > Z_\alpha$ where Z_α is the upper α-point from $N(0, 1)$.
For our own ease we consider one sided test in our simulation and tests whether the power of the test statistic goes to one or not as θ deviates from the null hypothesis.

3.2 Wilcoxon Signed Rank Test

The Wilcoxon One Sample Signed-Rank test test is the non parametric version of the one sample t test. It is based on ranks and because of that, the location parameter is not here the mean but the median. This tool can be used in case you have doubts that the normality assumption necessary to apply the t test does not hold.

3.2.1 Description:

Now,let $(X_1, X_2, X_3, \ldots, X_n)$ be independently drawn random sample of size n from a population having continuous distribution function $F(x - \theta)$.
Suppose we have to test:

$$H_0 : F(a - x) + F(a + x) = 1 \text{ vs}$$
$$H_1 : F(a - x) + F(a + x) > 1$$

If we shift the location of the observations as$Y_i = X_i - a$,then we have new set of observations $(Y_1, Y_2, Y_3, \ldots, Y_n)$.
Then H_0 reduces into

$$F(-y) + F(y) = 1$$

Now the updated testing problem is:

$$H_0 : \xi_{\frac{1}{2}} = 0 \text{ vs}$$
$$H_1 : \xi_{\frac{1}{2}} > 0$$

Under null, i.e. when $\{\theta \; or \; \xi_{\frac{1}{2}} = 0\}$,

$$F(x) \in \Omega_{null} = \big\{F(x)\big|F(0) = \frac{1}{2}, \text{ F is absolutely continuous, } F(x) = 1 - F(-x)\big\}$$

Arrange $|X_i|$'s in increasing order (ignore ties)and assign ranks $1, 2, 3, \ldots, n$. Keep tracking their signs also.

$$\begin{aligned} \text{Define } W^+ = & \text{ Sum of ranks of those observations which are positive} \\ = & \sum_{i=1}^{n} Y_i \end{aligned}$$

where,

$$Y_i = \begin{cases} i & if \; x > 0 \\ 0 & o.w. \end{cases}$$
$$i = 1, 2, 3, \ldots, n$$

If the null is true,then the observations will be more or less symmetrically distributed around 0.If the alternative is true then we have more positive observations and with larger magnitude.Therefore,a right tailed test based on W^+ is appropriate.

3.2.2 Large Sample test:

$$
\begin{aligned}
E_{H_0}(W^+) &= E\left(\sum_{i=1}^{n} Y_i\right) \\
&= \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} iP(Y_i = i) \\
&= \sum_{i=1}^{n} \frac{1}{2} \times i \\
&= \frac{n(n+1)}{4}
\end{aligned}
$$

$$
\begin{aligned}
V_{H_0}(W^+) &= \sum_{i=1}^{n} V(Y_i) + 0 \\
&\text{(co-variance terms vanishes due to independence)} \\
&= \sum_{i=1}^{n} \left[E(Y_i{}^2) - \frac{i^2}{4}\right] \\
&= \sum_{i=1}^{n} \left[\frac{i^2}{2} - \frac{i^2}{4}\right] \\
&= \frac{n(n+1)(2n+1)}{24}
\end{aligned}
$$

So,

$$
Z = \frac{W^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \overset{\mathrm{H_0}}{\sim} AN(0,1)
$$

Here,test based on Z is asymptotically distribution free.We reject H_0 at level α if $Z_{obs.} > Z_\alpha$,where Z_α is the upper α point of N(0,1).
For our own ease we consider one sided test in our simulation and tests whether the power of the test statistic goes to one or not as the median i.e. $\xi_{\frac{1}{2}}$ deviates from the null hypothesis.

4 Two Sample Location Problem

Let there be two populations 'A' and 'B' having their respective distribution functions F and G(assume continuous).
A random sample of size n, $(X_1, X_2, X_3, \ldots, X_n)$ is drawn independently from the 'A' population and another independently drawn random sample of size m, $(Y_1, Y_2, Y_3, \ldots, Y_m)$ from 'B' population and two populations are independent.
Suppose to test:

$$
\begin{aligned}
&H_0 : F = G \;\text{ vs} \\
&H_1 : \text{B is at the right of A}
\end{aligned}
$$

Alternatively we can form the two sample location problem as,

$$
\text{independent} \begin{cases} X_1, X_2, X_3, \ldots, X_n \overset{i.i.d}{\sim} F(x) \\ Y_1, Y_2, Y_3, \ldots, Y_m \overset{i.i.d}{\sim} G(x) = F(x-\theta) \;\&\; \theta > 0 \end{cases}
$$

$F(x)$ and $G(x)$ are both continuous distribution. So the hypothesis can be re-written as,

$$H_0 : F(x) = G(x) \text{ vs}$$
$$H_1 : \begin{cases} F(x) \geq G(x) \; \forall \; x \\ F(x) > G(x) \text{ for some } x \text{ with positive probability} \end{cases}$$

i.e. we are to check if $X <_{s.t.} Y$ or not.
$X <_{s.t.} Y$ means Y is stochastically larger than X, alternatively we can say that $1-F(x) < 1-G(x)$, where F and G are the distribution functions of X and Y respectively.

4.1 Wilcoxon Rank Sum Test

4.1.1 Steps:

The following steps are needed to evaluate Wilcoxon Rank Sum Test:

- Draw the samples
- Combine two samples and arrange the observations in increasing order
- Arrange ranks $(1, 2, 3, \dots, N = (m + n))$

4.1.2 Test Statistics:

Define,

$$W = \text{ sum of ranks of second sample observation.}$$

If H_1 is true then ordered arrangement of sample observation is expected to begin with larger run of X's or with larger runs of Y values. On the other hand under H_0, X's and Y's are expected to be thoroughly mixed up. So,W tends to be large under H_1 as compared to that under H_0.Thus a right tailed test is appropriate.

4.1.3 Null Distribution of W:

If there is no tie, then the ranking for the drawn sample would be a permutation of $1, 2, 3, \dots, N$. Second sample ranks under H_0 substitute a simple r.s. of size m drawn without replacement from $(1, 2, 3, \dots, N)$. Here the test statistics is exactly distribution free. Now,

$$E_{H_0}\left(\frac{W}{m}\right) = \text{Population mean}$$
$$= \frac{N+1}{2}$$

Then,

$$E_{H_0}(W) = \frac{m(m+n+1)}{2}$$

$$V_{H_0}\left(\frac{W}{m}\right) = \frac{N-m}{N-1}\ \frac{(\text{population variance})}{m}$$
$$= \frac{N-n}{N-1}\ \frac{N^2-1}{12}$$
$$= \frac{mn(m+n+1)}{12}$$

Define,

$$Z = \frac{W - \frac{m(m+n+1)}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} \overset{\mathrm{H_0}}{\sim} AN(0,1)$$

Test based on Z would be asymptotically normally distributed and we reject H_0 at approximate size α if $Z_{(obs.)} > Z_\alpha$.

4.2 Median Test

4.2.1 Steps:

The following steps are needed to evaluate Median Test:

- Combine the samples
- Find the sample median (M) for the combined data.
- Count second sample observations which exceed M

Suppose M is the k-th order statistics based on N observations $(N = m+n)$. For large m,n we get $k \approx \frac{N}{2}$

4.2.2 Test Statistics:

	$\leq M$	$> M$	Total
$1st$ sample	$k-m+t$	$N-k-t$	n
$2nd$ sample	$m-t$	t	m
Total	k	$N-k$	N

Now we can easily write that

$$P_{H_0}(T=t) \;=\; \frac{\binom{N-k}{t}\binom{k}{m-t}}{\binom{N}{m}}$$

$t \in \{max(0, m-k), \ldots, min(m, N-k)\}$
Right tailed test based on T is appropriate here.
p-value $= P_{H_0}(T \geq t)$. Here the exact sampling distribution of T under H_0 does not depend on explicit form of F and G. So the test is exactly distribution free.

Alternative Approaches:
If one arrange m number of Y obs 1st then the given test statistics can alternatively written as the following form,

$$\begin{aligned} M = & \sum_{i=1}^{m} \frac{1}{2}\left[sign\left(R_i - \frac{N+1}{2}\right)+1\right] \\ & = \text{number of Y observation greater than the sample median and increased by} \\ & \frac{1}{2} \text{ if an Y observation is the sample median} \end{aligned}$$

The distribution of the test statistic M follow the same distribution as the T.

4.2.3 Large Sample Test

If we imagine being greater than M is success then the success probability for the combined sample is $p = \frac{N-k}{N}$.Now $T \sim Bin(m, p)$.
Therefore,

$$Z = \frac{T-mp}{\sqrt{mp(1-p)}} \overset{\text{H}_0}{\sim} AN(0,1)$$

We reject H_0 at level α if $Z_{(obs.)} \geq Z_\alpha$

5 Idea About Simulation

First of all we know about the parametric test statistic for location parameter when the underlying distribution is Normal. But if the information about the Normality of the data is not known to us then we can not ensure the test statistic we are using are good or not. Then the non-parametric test procedure comes in the mind. So we all see in the above for both the cases one sample and two sample the test statistic is distribution free under the null hypothesis. We are going to justify our claim practically. Then we will see how the power curve will be affected by the deviation of the normality and compare it with the parametric version of location problem assuming data is from normal distribution(for both the cases when variance is known and variance is unknown). Hence our simulation study can be divided into following points:

- We simulate 's' (say 1000) number of test statistic under H_0 and plot the 'Q-Q plot'. This step can be varied for different distributions and if the test statistic is distribution free then the 'Q-Q plot' of the test statistic will be similar.
- We will check for the normality for the large sample using 'Q-Q plot' for each of the test statistic.
- Then we will plot the power function of the test statistic for various location or more specifically deviating location from the null hypothesis. Compare the power function with each other for different distributions.
- Lastly, we will check whether the test will be size α test or not and the corresponding observations related to our simulation.

[Here we use 'R' software for simulation study and drawing the corresponding graphs.]

6 Brief Description Of Parametric Test Procedure

6.1 Single Sample Test

Suppose we have a $N(\mu, \sigma^2)$ population.We are interested in testing $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1 \quad (\mu_1 > \mu_0)$.Now we draw a sample with independent sample observations $(x_1, x_2, x_3, \ldots, x_n)$ of size n from the population.Define

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})$$

For single sample test for mean,test statistics are

$$\tau = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \sim N(0,1) \quad \text{[when } \sigma \text{ known]}$$

(We reject H_0 at level α if $\tau_{(obs.)} > \tau_\alpha$)

$$t = \frac{\sqrt{n}(\bar{x} - \mu)}{s} \sim t_n \quad \text{[when } \sigma \text{ unknown]}$$

(We reject H_0 at level α if $t_{(obs.)} > t_{\alpha;n}$)
t_n= t distribution with n degrees of freedom.

6.2 Two Sample Test

Suppose we have two normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ .We are interested in testing $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 > 0 \quad (\mu_1 > \mu_2)$.Now we draw one sample from each population with independent sample observations i.e. $(x_{11}, x_{12}, x_{13}, \ldots, x_{1n_1})$ of size n_1 from the first population and $(x_{21}, x_{22}, x_{23}, \ldots, x_{2n_2})$ of size n_2 from the second population.Define

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad i = 1, 2$$

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \mu_i) \quad i = 1, 2$$

For two samples test statistics are

$$\tau = \frac{(x_1 - x_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \quad \text{[when } \sigma \text{ known]}$$

(We reject H_0 at level α if $\tau_{(obs.)} > \tau_\alpha$)

$$t = \frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{n_1+n_2} \quad \text{[when } \sigma \text{ unknown]}$$

(We reject H_0 at level α if $t_{(obs.)} > t_{\alpha;(m+n)}$)
$t_{n_1+n_2}$= t distribution with $(n_1 + n_2)$ degrees of freedom.

7 Distribution Free

The non-parametric test statistics is distribution free i.e. if we plot the test statistic of different non-parametric test varying the parent distribution from where the data is generated, then the

'Q-Q plot' will look alike. Under this basic theory we first see for one-sample location test statistic and then two sample location test statistics.

7.1 Sign Statistic

Let us see the 'Q-Q plot' of the Sign Statistic.

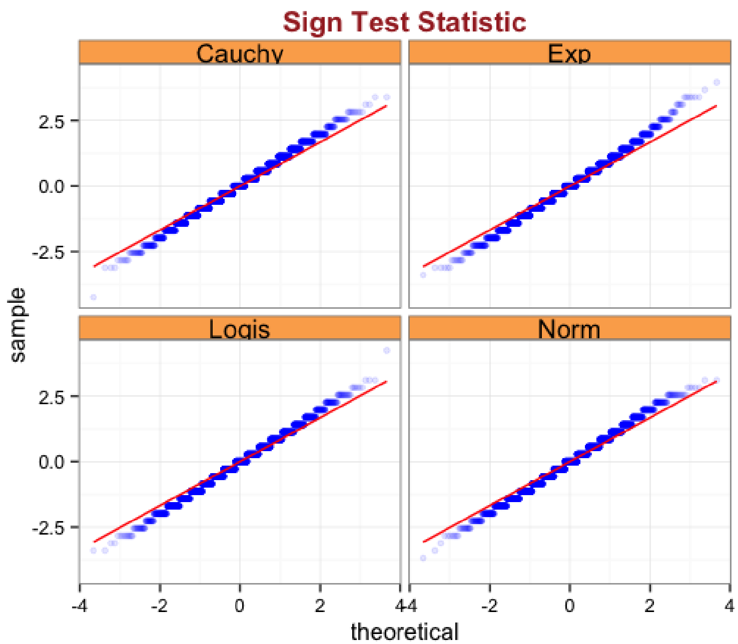


Figure 1: Sign Test Statistic value and 'Q-Q plots' for different distributions.

The 'Q-Q plot' (figure 1) of the sign test statistic looks like similar for all distributions, so it can be concluded that sign-test statistic is distribution free.

Generate Data: We first select a distribution and draw 'm'(Here 50) number of sample with location parameter 0 and then apply sign test on these observations to get one sign statistic value. We replicated the procedure of getting the sign statistic a few number of times to get the 'Q-Q plot' of the sign-statistic. We vary those steps for different distributions i.e. Cauchy, Exponential(Exp), Logistic(Logis) and Normal(Norm) distribution. Here we use 'SIGN.test()' in *R* library **BSDA** and it returns the statistics as well as the p-value of the test.

7.2 Wilcoxon Signed Rank Test Statistics

The 'Q-Q plot' (figure 2) of the wilcox signed rank test statistic looks like similar for all distributions, so it can be concluded that wilcoxon signed rank test statistic is distribution free.

Generate Data: We first select a distribution and draw 'm'(Here 50) number of sample with location parameter 0 and then apply wilcoxon signed rank test on these observations to get one wilcoxon signed rank statistic value. We replicated the procedure of getting the wilcoxon signed rank statistic a few number of times to get the 'Q-Q plot' of the wilcoxon signed rank statistic. We vary those steps for different distributions i.e. Cauchy, Exponential(Exp), Logistic(Logis) and Normal(Norm) distribution. Here we use 'wilcox.test()' (One Sample) in *R* and it returns the statistics as well as the p-value of the test.

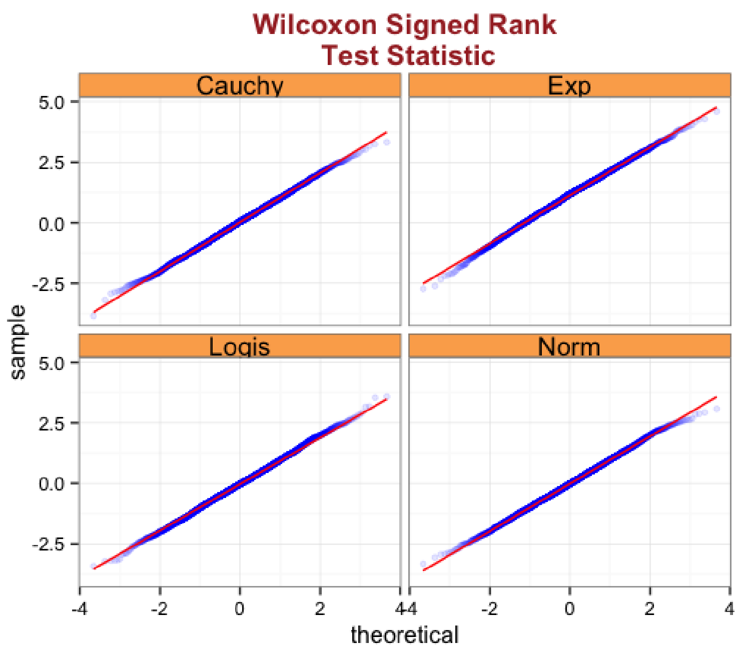


Figure 2: Wilcoxon Signed Rank Test Statistic value and 'Q-Q plots' for different distributions.

7.3 Wilcoxon Rank Sum Test Statistics

We do the same thing for two sample location problem. Here we generate the value of the statistics for 'm' and 'n' (Here n=50 and m=50) and plot the 'Q-Q plot' of the corresponding distributions. We can see that the 'Q-Q plot' (figure 3) looks like similar for all the distributions. So the test statistic is distribution free.

Generate Data: We first select a distribution and draw 'm'(Here 50) and 'n'(Here 50) number of sample with location parameter 0 (for both the sample) and then apply wilcoxon rank sum test on these observations to get one wilcoxon rank sum statistic value. We replicated the procedure of getting the wilcoxon rank sum statistic a few number of times to get the 'Q-Q plot' of the wilcoxon rank sum statistic. We vary those steps for different distributions i.e. Cauchy, Exponential(Exp), Logistic(Logis) and Normal(Norm) distribution. Here we use 'wilcox.test()' (Two Sample) in *R* and it returns the statistics as well as the p-value of the test.

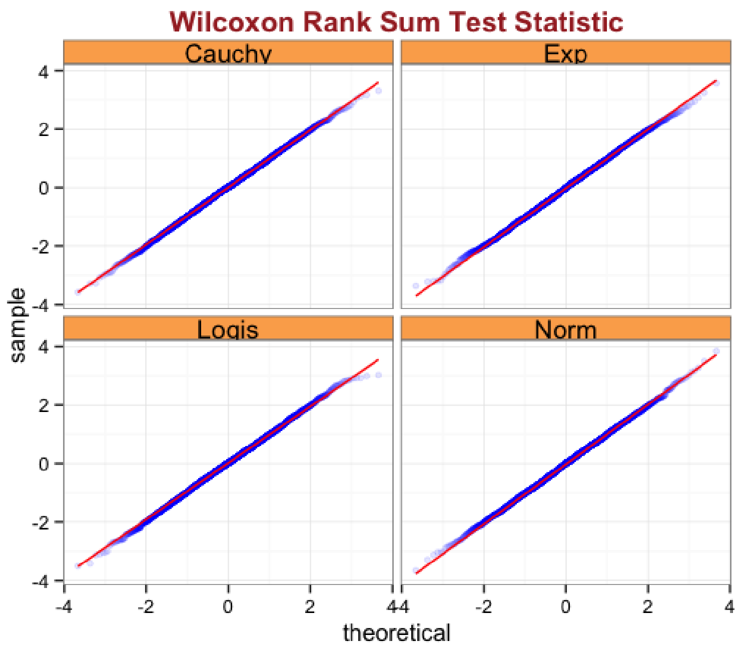


Figure 3: Wilcoxon Rank Sum Test Statistic value and 'Q-Q plots' for different distributions.

7.4 Median Test Statistics

Here we do the same thing as we did for Wilcoxon Rank Sum test statistics. Here we generate the value of the statistics for 'm' and 'n' (Here n=50 and m=50) and plot the 'Q-Q plot' of the corresponding distributions. We can see that the 'Q-Q plot' (figure 4) looks like similar for all the distributions. So the test statistic is distribution free.

Generate Data: We first select a distribution and draw 'm'(Here 50) and 'n'(Here 50) number of sample with location parameter 0 (for both the sample) and then apply Median test on these observations to get one median statistic value. We replicated the procedure of getting the Median statistic a few number of times to get the 'Q-Q plot' of the Median statistic. We vary those steps for different distributions i.e. Cauchy, Exponential(Exp), Logistic(Logis) and Normal(Norm) distribution. Here we use 'mood.medtest()' in *R* library 'RVAideMemoire' and it returns the statistics as well as the p-value of the test.

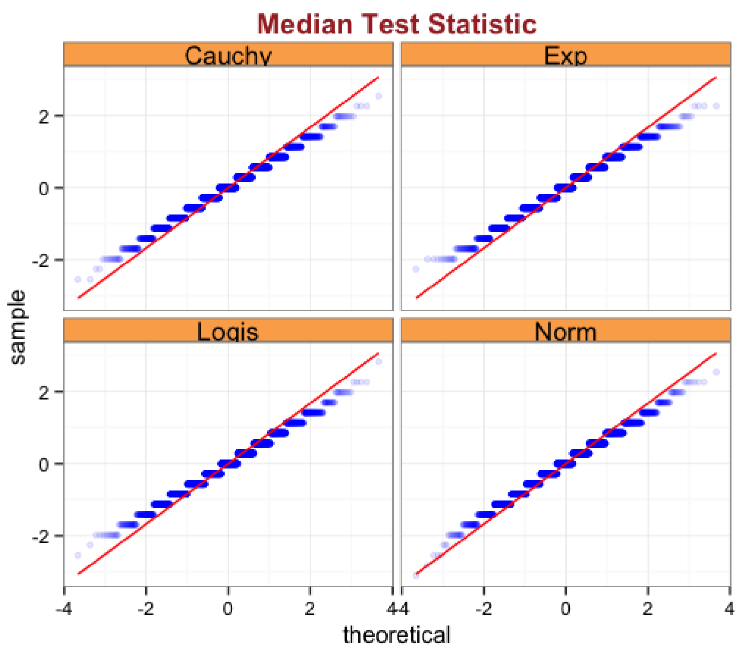


Figure 4: Median Test Statistic value and 'Q-Q plots' for different distributions.

8 Normality Approaches For The Test Statistic

In this part we want to check weather the test statistic follows normality in the large sample or not. For two sample location problem validity of normal assumption holds for what values or more specifically what relations of m (1st sample size) and n (2nd sample size). In this section will briefly discuss about it. At first we will look at the large sample graph of the test statistic and compare with it the standard normal distribution. We perform two test to support our answer one is 'Shapiro-Wilk' and another is 'Kolmogorov-Smirnov' test. The 'Shapiro-Wilk' test gives us if the data is normal with some mean and variance or not normal where as 'Kolmogorov-Smirnov' test gives us if the data is standard normal or not. Lets us begin with one sample location test statistic.

8.1 Sign Test Statistic

At first we look at the normality assumption on the sign-statistic. Here we mainly generated the statistic value for a given data of size 'n' which is generated by a given distribution. Here we test for the normality under the null assumption so we generate data from the distribution of location parameter, θ, equals to 0 and perform the test a few number of times.

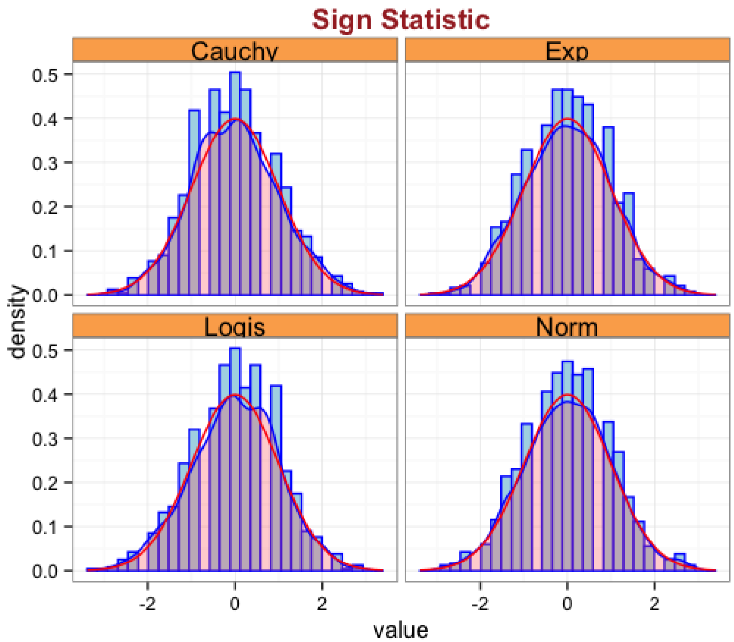


Figure 5: Validity of Normality assumption on Sign Test Statistic

In figure 5 red line represents the standard normal distribution and the blue line represents the empirical distribution function generated by the given data or the value of the sign-statistic. So from the figure 5 it is clear that for the four distributions i.e. Cauchy, Exponential(Exp), Logistic(Logis) and Normal(Norm) distribution the sign statistic approaches to the normality when the sample sizes is large. Here the value of the sample (n) is 50.

For the shake of more clarity we uses two test as told earlier and the p-value of the test statistic is given below in the table 1.

Distribution	**Shapiro-Wilk Test p-value**	**Kolmogorov-Smirnov Test p-value**
Normal	0.3405	0.27
Cauchy	0.6874	0.9639
Logistic	0.6874	0.1122
Exponential	0.1234	0.5441

Table 1: Table shows the p-value for different normality test for Sign Test Statistic Value

Since all the p-values for all the distributions are greater than level of significance α (here 0.05), so we will accept the null hypothesis which is the data is from standard normal distribution.

8.2 Wilcoxon Signed Rank Test Statistic

At first we look at the normality assumption on the wilcoxon signed rank statistic. Here we mainly generated the statistic value for a given data of size 'n' which is generated by a given distribution. Here we test for the normality under the null assumption so we generate data from the distribution of location parameter, $\xi_{\frac{1}{2}}$, equals to 0 and perform the test a few number of times.

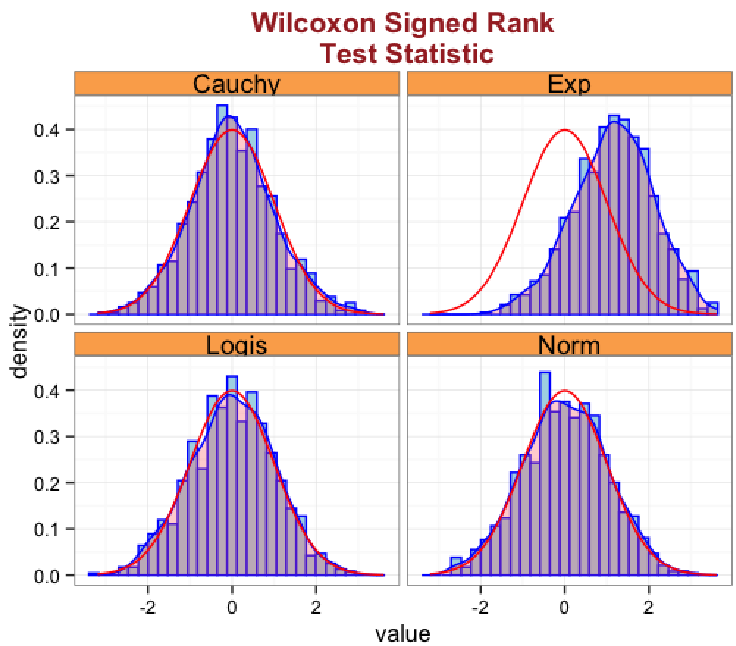


Figure 6: Validity of Normality assumption on Wilcoxon Signed Rank Test Statistic

In figure 6 red line represents the standard normal distribution and the blue line represents the empirical distribution function generated by the given data or the value of the wilcoxon signed rank statistic. So from the figure 6 it is clear that for the four distributions i.e. Cauchy, Exponential(Exp), Logistic(Logis) and Normal(Norm) distribution the sign statistic approaches to the normality when the sample sizes is large. Here the value of the sample (n) is 50.

For the shake of more clarity we uses two test as told earlier and the p-value of the test statistic is given below in the table

Distribution	Shapiro-Wilk Test p-value	Kolmogorov-Smirnov Test p-value
Normal	0.635	0.5487
Cauchy	0.9804	0.8643
Logistic	0.2582	0.7112
Exponential	0.9546	0.0006709

Table 2: Table shows the p-value for different normality test for Wilcoxon Sign Rank Test Statistic Value

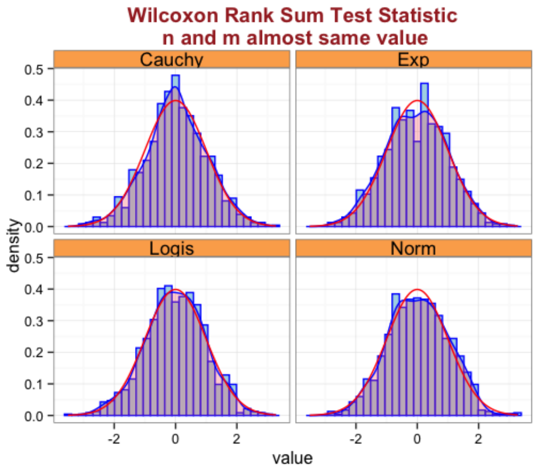
Since all the p-values for all the distributions except the one for KS test for exponential distribution are greater than level of significance α (here 0.05), so we will accept the null hypothesis for Cauchy,Normal and Logistic distributions which is the data is from standard normal distribution.As earlier I said in case of KS test for the exponential distribution p vale is lesser than the level of significance,we reject the null hypothesis for this.But for the Shapiro-wilk test on exponential data the p value is greater than the level of significance.Therefore, for exponential data we reject the null hypothesis for Kolmogorov-Smirnov test at 5% level of significance but accept the null for Shapiro-Wilk test at 5% level.Hence we understand that the normality assumption of the Wilcoxon Signed Rank test statistics for exponential distribution was not accurate. For exponential data under large sample observation Wilcoxon Signed Rank test statistics follows normal distribution with some other mean and variance except mean=0 and variance=1 i.e. doesn't follow standard normal.

In wilcoxon signed rank test the general null hypothesis is $F(a-x)+F(a+x) = 1$ i.e. we look at the shape of the corresponding distribution and want to know if the distribution is symmetric or not. As a result this test is doesn't provide a good result for skewed distributions. Exponential is a skewed distribution,so the asymptotic standard normal approximation for the exponential data in wilcoxon signed rank test is not accurate.

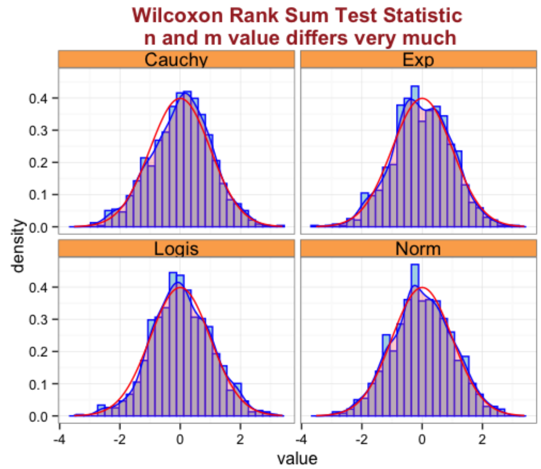
8.3 Wilcoxon Rank Sum Test Statistic

Coming to the two sample location problem we test the normality for mainly two types of m and n and then see what is the effect on the test statistic only for large n, for both m and simultaneously large. The test procedure is done under null, so we draw sample from a distribution with same location parameter.

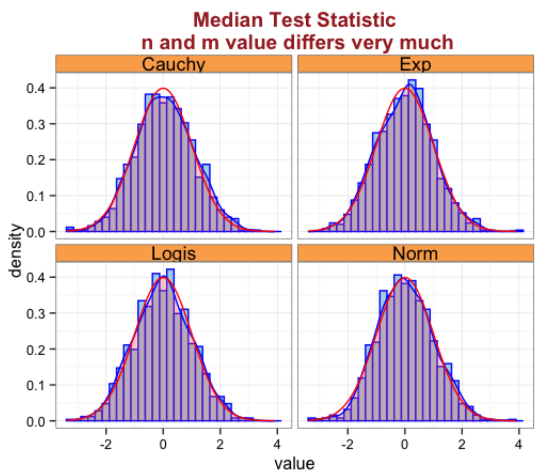
At first let us look at the test statistic when both m and n are sufficently large. From the figure 7a, it can be easily concluded that for both m and n large values the value of the wilcoxon rank sum test statistic follow normality. Coming to the second and third plot (figure 7b & 7c), when the value of m and n differs significantly or if one of the sample is very large then also it goes to normality. So for both the cases if it follow normality. Here we generated data for $m = 50$ & $n = 50$ for the first case, $m = 10$ & $n = 90$ for the second one and $m = 90$ & $n = 10$ for the third one.



(a) n=50 and m=50



(b) n=10 and m=90



(c) n=90 and m=10

Figure 7: Validity of Normality assumption on Wilcoxon Rank Sum Test Statistic

For the shake of more clarity we uses two test as told earlier and the p-value of the test statistic is given below in the table 3.

n and m value	**Distribution**	**Shapiro-Wilk Test p-value**	**Kolmogorov-Smirnov Test p-value**
$n = 50$ and $m = 50$	Normal	0.5806	0.9639
	Cauchy	0.9391	0.8643
	Logistic	0.5385	0.27
	Exponential	0.8667	0.5487
$n = 10$ and $m = 90$	Normal	0.09927	0.27
	Cauchy	0.1935	0.9639
	Logistic	0.913	0.3927
	Exponential	0.9314	0.9639
$n = 90$ and $m = 10$	Normal	0.07134	0.27
	Cauchy	0.828	0.1122
	Logistic	5998	0.9639
	Exponential	0.2827	0.7112

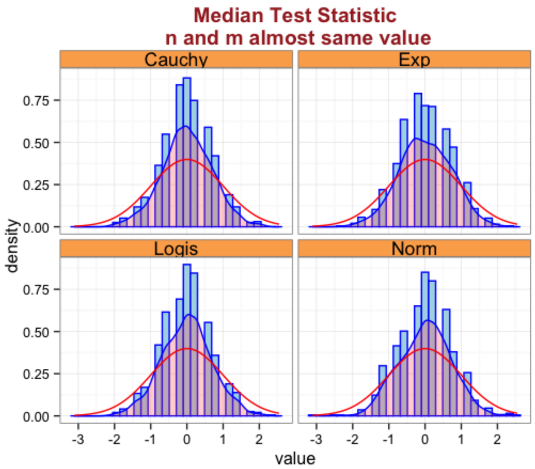
Table 3: Table shows the p-value for different normality test for Wilcoxon Rank Sum Test Statistic Value

Since all the p-values for all the distributions are greater than level of significance α (here 0.05), so we will accept the null hypothesis which is the data is from standard normal distribution.

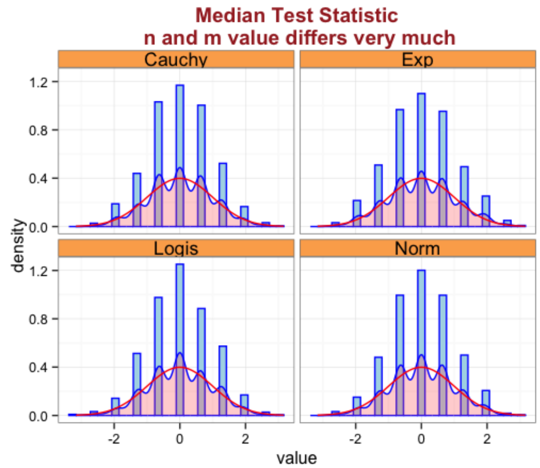
8.4 Median Test Statistic

As wilcoxon rank sum test here also we are interested to see what is the effect on the test statistic only for large n, for both m and simultaneously large. The test procedure is done under null, so we draw sample from a distribution with same location parameter.

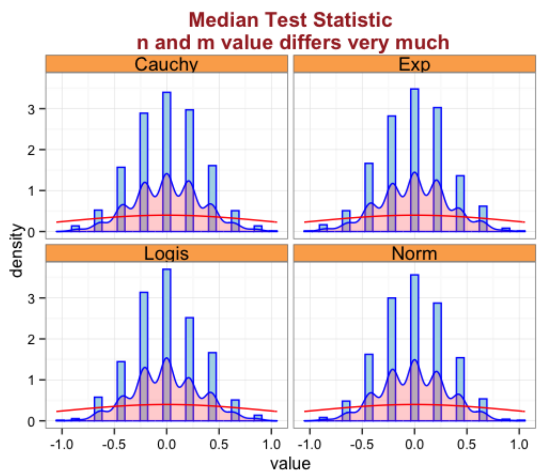
At first let us look at the test statistic when both m and n are sufficently large. From the figure 8a, it can be easily concluded that for both m and n large values the value of the median test statistic follow normality. Coming to the second plot (figure 8c & 8b), when the value of m and n differs significantly or if one of the sample is very large then it will not be normal at all. So for the first case it follows normality but when the sample sizes differs too much it doesn't. More significatly when the value of $m < 0.1 \times N$ then the approximation of the binomial is very good but when the value of $m > 0.1 \times N$ then the binomial approximation is not so good. That is why when the value of $m < 0.1 \times N$ is well approximated by $Bin(m, p)$ where $p \approx \frac{1}{2}$. But $Bin(10, p)$ is not at all a large sample in this case so the result does not show that it is normal. Here we generated data for $m = 50$ & $n = 50$ for the first case, $m = 10$ & $n = 90$ for the second one and for the last one $m = 90$ & $n = 10$.



(a) n=50 and m=50



(b) n=90 and m=10



(c) n=10 and m=90

Figure 8: Validity of Normality assumption on Median Test Statistic

For the shake of more clarity we uses two test as told earlier and the p-value of the test statistic is given below in the table 4.

n and m value	**Distribution**	**Shapiro-Wilk Test p-value**	**Kolmogorov-Smirnov Test p-value**
$n = 50$ and $m = 50$	Normal	0.2148	0.3927
	Cauchy	0.09738	0.1777
	Logistic	0.2814	0.5441
	Exponential	0.1019	0.27
$n = 90$ and $m = 10$	Normal	0.04043	0.5441
	Cauchy	0.005615	0.1122
	Logistic	0.001302	0.3927
	Exponential	0.002941	0.01195
$n = 90$ and $m = 10$	Normal	0.03505	0.01195
	Cauchy	0.006156	0.02222
	Logistic	0.04702	0.01195
	Exponential	0.009024	0.01195

Table 4: Table shows the p-value for different normality test for Median Test Statistic Value

So from the given table 4 it is clarified that when the sample size equal due to the large sample property it follows normal at the level of significance 0.05 but when n and m differs significantly then the normality deviates. In the figure 8b as $m < 0.1 \times N(= m + n)$ it deviates from the normality and it aproaches to $Bin(m, \frac{1}{2})$. Here as $m = 10$ so it does not follow the large sample, if $m \approx to40$ then it again goes to $Bin(40, \frac{1}{2}$ and the value of m also quite large to approximate the normality. To check this let us generate the sample from $n = 410$ & $m = 40$. The figure 9 and the table 5 shows that it is normal.

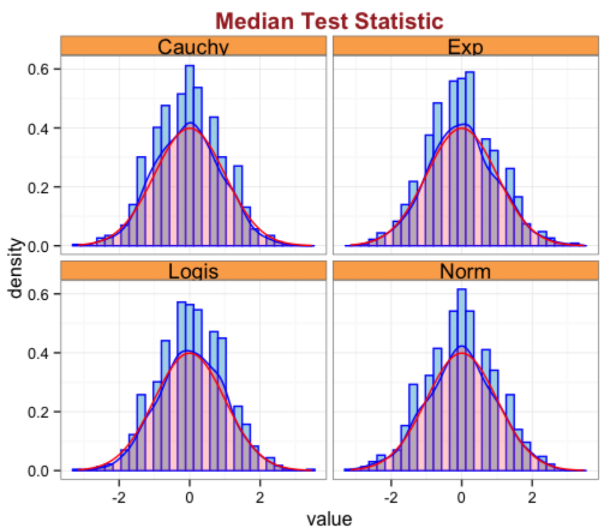


Figure 9: Validity of Normality assumption of Median Test Statistics when n=410 and m=40

Distribution	Shapiro-Wilk Test p-value	Kolmogorov-Smirnov Test p-value
Normal	0.3575	0.8643
Cauchy	0.4162	0.3927
Logistic	0.28064	0.3927
Exponential	0.5303	0.7112

Table 5: Table shows the p-value for different normality test for median test statistics for n=410 & m=40

But for the third graph 8c, it does not follow the normality and as well as it is not following Binomial distribution with given mean mp and variance $mp(1-p)$ where $p \approx \frac{1}{2}$. Actually to approximate the normality of the hypergeometric distribution it first approaches to Binomial and then after large sample it goes to normal. But hypergeometric distribution will approach to binomial iff $m < 0.1 \times N$, but here it is more or less equal to N that is why it is not following the normal distribution at all for the large sample distribution.

9 Power Curve

The power of a binary hypothesis test is the probability that the test rejects the null hypothesis (H_0) when a specific alternative hypothesis (H_1) is true. The statistical power ranges from 0 to 1, and as statistical power increases, the probability of making a type II error (wrongly failing to reject the null hypothesis) decreases.

$$\begin{aligned}\text{Power of the test} &= \text{ 1 - P(Type II error)} \\ &= \text{P(reject}H_0|H_1 \text{ is true)}\end{aligned}$$

Now we will discuss about power curve. It represents the power of the test on one axis and deviation of the mean from the target on another axis. It helps to assess the suitable size of the sample. In statistics, it is regarded as power law graph. When standard deviation and level of significance are kept constant, it shows every combination of deviation for each size of sample and power. On power curve, based on the entered values, each symbol shows a computed value.

At a particular sample size and power value, the values on the graph are examined to find out the deviation of the mean from the target. A low power hypothesis test fails to find out a practically noteworthy difference. An adequate power value is usually taken to be 0.9. By increasing the size of the sample, the power of the test also rises. The curve rises monotonically and smoothly from a height equivalent to the level of significance of the test. The rise in curve continues until it reaches the maximum height of hundred percent eventually. Here we want to compare with the parametric tests with the other non-parametric tests.

9.1 One sample

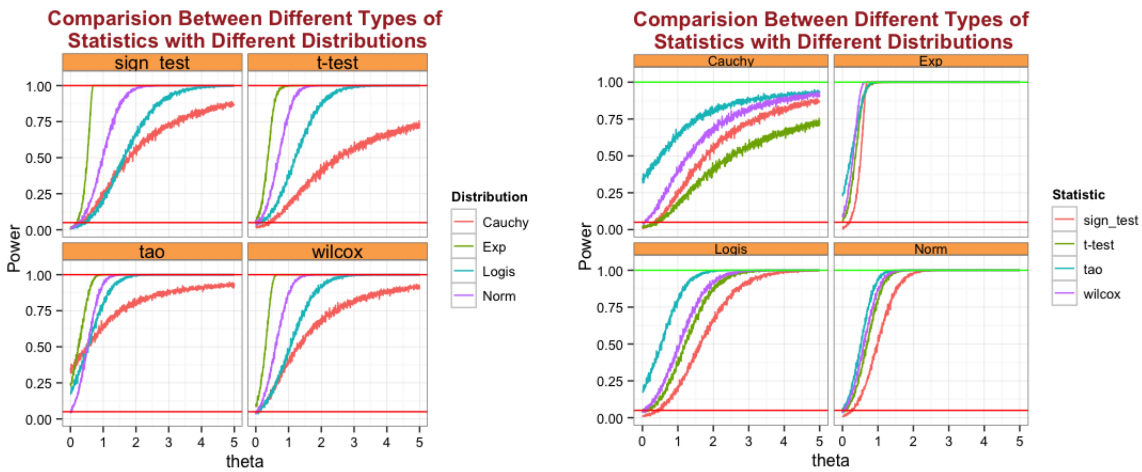


Figure 10: Power Curve for small sample

Here in Fig:10 we can notice the shape of power curves for one sample problem using different parametric and non parametric tests over four frequently used distributions.We will analyse them and will draw conclusions.

At first we notice that power curves for parametric tests are attaining the maximum power 1 for relatively lower values of the parameter and sample size than the nonparametric tests when the data is generated from the normal distribution. But when we shift our distribution normal to other then parametric test are not good at all. So we can say that parametric test is powerful when the underlying distribution is known. But when the underlying distribution is unknown then nonparametric test is better than parametric. Also when the underlying distribution changes from normal to logistic or cauchy then power curve does not even follow the level of significance i.e. 0.05 in our case. But for exponential distribution wilcox-signed rank test is not giving good result as it is assumed that wilcox test is applied on the symmetric type distribution but Exponential is right skewed distribution. So for this reason we are including the size of each nonparametric test in our project as size may not follow 0.05 for small number of observations. Before that we are interested in the large sample property of our test statistic in our project.

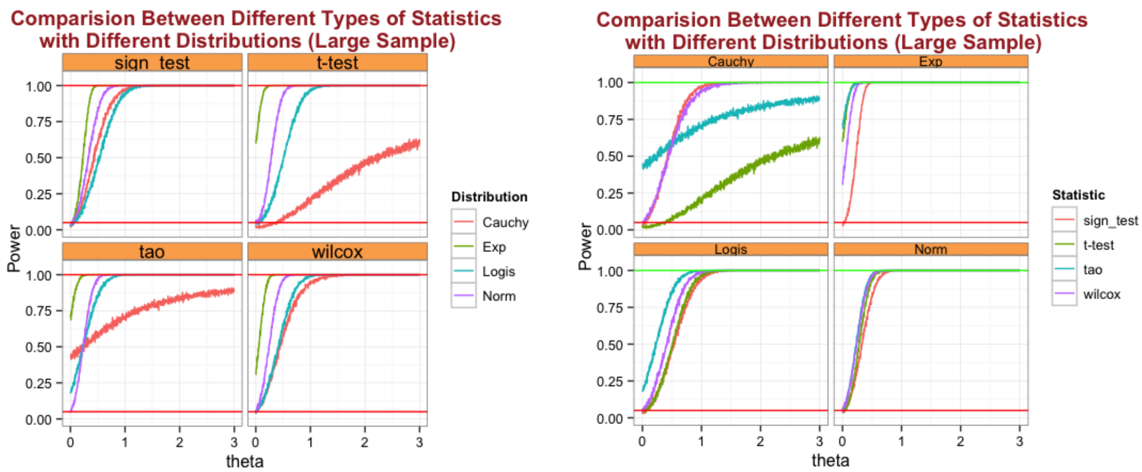


Figure 11: Power Curve for small sample

Here in Fig:11 we can notice the shape of power curves for one sample problem using different parametric and non parametric tests over four frequently used distributions.We will analyse them and will draw conclusions.

At first we notice that power curves for small sample have lesser slopes than that of large samples.Cauchy distribution doesn't have a mean,it's characteristics deviate from normal distribution.So it doesn't provide appropriate result for one sample non parametric large sample tests.

9.2 Two Sample:

Here in Fig:12 we can notice the shape of power curves for one sample problem using different parametric and non parametric tests over four frequently used distributions.We will analyse them and will draw conclusions
At first we notice that power curves for parametric tests are attaining the maximum power 1 for relatively lower values of the parameter and sample size than the nonparametric tests when the data is generated from the normal distribution. But when we shift our distribution normal to other then parametric test are not good at all. So we can say that parametric test is powerful when the underlying distribution is known. But when the underlying distribution is unknown then nonparametric test is better than parametric. Also when the underlying distribution changes from normal to logistic or cauchy then power curve does not even follow the level of significance i.e. 0.05 in our case. But for exponential distribution wilcox-signed rank test is not giving good result as it is assumed that wilcox test is applied on the symmetric type distribution but Exponential is right skewed distribution. So for this reason we are including the size of each nonparametric test in our project as size may not follow 0.05 for small number of observations. Before that we are interested in the large sample property of our test statistic in our project.

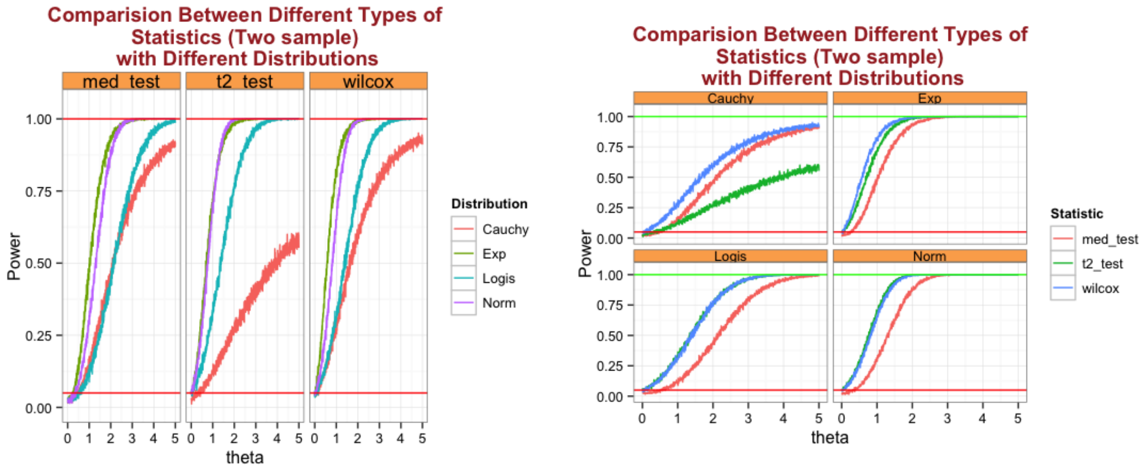


Figure 12: Power Curve for small sample

Here in Fig:14 We can notice the shape of power curves for one sample problem using different parametric and non parametric tests over four frequently used distributions.We will analyse them and will draw conclusions

At first we notice that power curves for small sample have lesser slopes than that of large samples.Parametric and nonparametric,both type of tests are giving similar power curves for every distributions but cauchy.Cauchy distribution doesn't have a mean,it's characteristics deviate from normal distribution.So it doesn't provide appropriate result for two sample non parametric large sample t2 test.

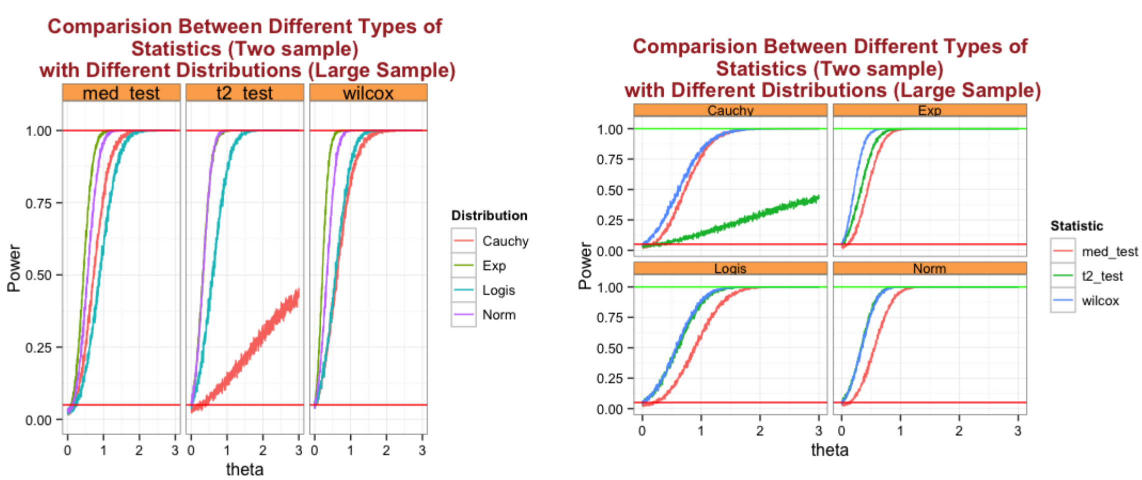


Figure 13: Power Curve for large sample

10 Size Of The Test For One Sample Location Test

In hypothesis testing, the size of a test is the (maximum) probability of committing a Type I error, that is, of incorrectly rejecting the null hypothesis when the null hypothesis is true.
For example suppose we have a $N(\mu, \sigma^2)$ population.(The variance of the distribution, denoted by σ^2, is supposed to be known). We are interested in testing $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1 \quad (\mu_1 > \mu_0)$.Now we draw a sample with independent sample observations $(x_1, x_2, x_3, \ldots, x_n)$ of size n from the population.Define

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

from the distribution and we compute the Z-statistic

$$Z = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \;\sim\; N(0, 1)$$

from the distribution and we compute the Z-statistic

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sqrt{\sigma^2}}$$

where $\bar{X}$ is the sample mean:

$$\bar{X} = \frac{1}{n} \sum X_i$$

We select a critical value z and reject the null hypothesis if $|Z_n| > z$
The size of the test is

$$\begin{aligned}
&P(\text{Rejecting the null when it is true})\\
&= P(|Z_n| > z)\\
&= 1 - F(z) + F(-z)\\
&= 2F(-z)\\
&= \alpha \quad \text{(say)}
\end{aligned}$$

Here we only concentrate ourselves into the size of one sample location problem. So

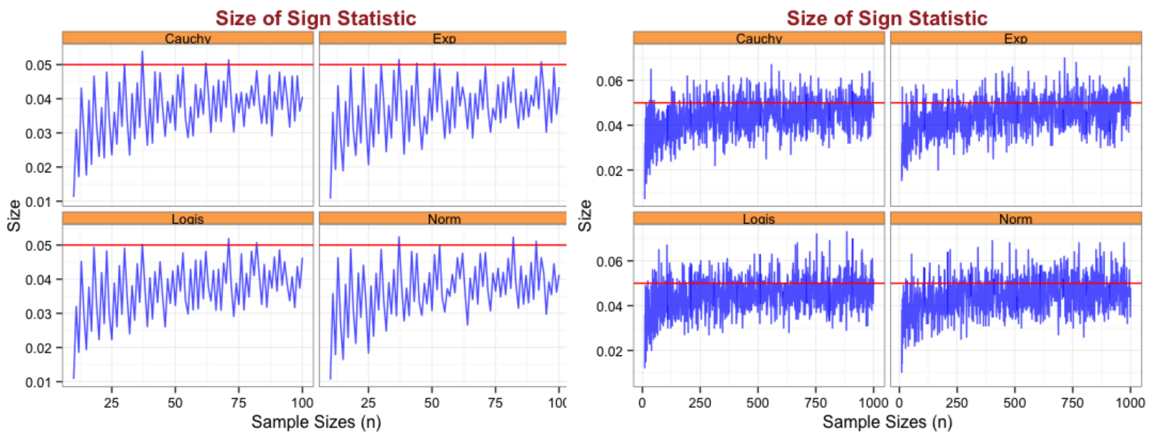


Figure 14: Power Curve for large sample

11 CODES

Here we include the necessary codes for generating the data and for the simulation study. In this project R-Software is used. The code is written in 'R-Studio'. We need few library for R to run these codes.

Library	Relevant Functions
ggplot2	For Drawing All types of graphs. Such as ggplot(), geom_line(), geom_abline(), geom_histogram() etc.
BSDA	SIGN.test
pbapply	pbsapply, pbreplicate
OneTwoSamples	mean_test1
lattice	For quick representation of graph in the inline code. Such as xyplot().
tidyr	gather
MASS	
RVAideMemoire	mood.medtest

Table 6: Table Shows the necessary Library and Corresponding Functions for the code

11.1 modified_draws.R

Here some standard distributions are defined with some modifications on location para-meter. Here we mainly uses four distributions to generate our data i.e. Cauchy (rCauchy), Exponential (rExp), Logistic (rLogis) and Normal (rNorm) distribution.

```
#modified functions to draw samples

rNorm <- function(n, location=0) {
  rnorm(n, mean = location, sd = 1)
}

rLogis<-function(n,location=0){
  rlogis(n,scale =1,location=location)
}

rCauchy=function(n,location=0){
  rcauchy(n,location = location,scale=1)
}
```

```
rExp=function(n,location=0,rate=1){
  rexp(n,rate=rate)-log(2)+location
}
```

Code 1: modified_draws.R

11.2 theme_mine.R

This code contains some basic designs of 'ggplot2' and for better representation of the graph this code is necessary.

```
theme_mine <- function (base_size = 12, base_family = "") {
  require(grid)
  theme(line = element_line(colour = "black", size = 0.5, linetype = 1, lineend =
    "butt"),
        rect = element_rect(fill = "white", colour = "black", size = 0.5, linetype
     = 1),
        text = element_text(family = base_family,face = "plain", colour = "black",
                            size = base_size, hjust = 0.5,vjust = 0.5, angle = 0,
    lineheight = 0.9),
        axis.text = element_text(size = rel(0.8),margin = margin(0.1,0.1,0.1,0.1,
    "cm")),
        axis.line = element_blank(),
        axis.text.x = element_text(vjust = 1,colour = "black",margin=margin
    (0.1,0.1,0.1,0.1,"cm") ),
        axis.text.y = element_text(hjust = 1,colour="black"),
        axis.title.x = element_text(colour = "black"),
        axis.title.y = element_text(angle = 90,colour = "black"),
        axis.ticks.length = unit(0.15, "cm"),
        #axis.ticks.margin = margin(0.1,0.1,0.1,0.1, "cm"),

        legend.background = element_rect(colour = NA),
        legend.margin = margin(0.2,0.1,0.2,0.2, "cm"),
        legend.key = element_rect(colour = "grey80"),
        legend.key.size = unit(1.2, "lines"),
        legend.key.height = NULL,
        legend.key.width = NULL,
        legend.text = element_text(size = rel(0.8)),
        legend.text.align = NULL,
        legend.title = element_text(size = rel(0.8), face = "bold", hjust = 0),
        legend.title.align = NULL,
        legend.position = "right",
        legend.direction = NULL,
        legend.justification = "center",
        legend.box = NULL,
  #grey90
        panel.background = element_rect(fill = "white", colour = NA),
        panel.border = element_rect(fill = NA, colour = "grey50"),
        panel.grid.major = element_line(colour = "grey90", size = 0.2),
        panel.grid.minor = element_line(colour = "grey98", size = 0.5),
        panel.spacing = unit(0.25, "lines"),

        strip.text = element_text(size = rel(1.1)),
        strip.background = element_rect(fill = "#fcac59", colour = "grey50"),
```

```
        strip.text.x = element_text(),
        strip.text.y = element_text(angle = -90),

        plot.background = element_rect(colour = "white"),
        plot.title = element_text(size = rel(1.2),face = "bold",colour="brown",
    margin = margin(0.25,0.25,0.25,0.25,"lines")),
        plot.margin = unit(c(1, 1, 0.8, 0.5), "lines"),complete = TRUE)
}
```

Code 2: Distribution Free and Large Sample Normality

11.3 Distribution Free & Validation Of CLT

This Code is necessary to show that the distribution of the test statistic is distribution free. Before running this code **modified_draws.R** should run.

```
setwd(dirname(rstudioapi::getSourceEditorContext()$path))
# source("sign_test.r")
# source("two_sample.R")
source("theme_mine.r")
source("modified_draws.r")




Chii=function(data,distns,seed=7){
  x=as.numeric((data[,distns]))
  shap_test=shapiro.test(x)
  set.seed(seed)
  y=rNorm(length(x))
  ks_test=ks.test(x,y)
  my_list=list("shapiro_test"=shap_test,"ks_test"=ks_test)
  return(my_list)
}







qq_one=function(n,repl=4000,distns,seed=43,param=T,encode=T){
  require(tidyr)
  sign_data=matrix(0,nrow=repl,ncol= length(distns))
  wilcox_data=matrix(0,nrow=repl,ncol= length(distns))
  if(param==T){
  param_data1=matrix(0,nrow=repl,ncol= length(distns))
  param_data2=matrix(0,nrow=repl,ncol= length(distns))
  colnames(param_data1)<-distns
```

```
  colnames(param_data2)<-distns
  }
  colnames(sign_data)<-distns
  colnames(wilcox_data)<-distns

  for(dist in distns){
   require(pbapply)
    require(BSDA)
    distn=get(paste0("r",dist))
    set.seed(seed)
    sign_data[,dist]<-pbreplicate(repl, (SIGN.test(distn(n),alternative = "g",md
   =0)$statistic-(n/2))/sqrt(n/4))
    set.seed(seed)
    wilcox_data[,dist]<-pbreplicate(repl, ((wilcox.test(distn(n),alternative = "g"
   ,mu=0)$statistic)-(n*(n+1)/4))/sqrt(n*(n+1)*(2*n+1)/24))
    if(param==T){
    set.seed(seed)
    param_data1[,dist]<-pbreplicate(repl, mean_test1(distn(n),mu=0,side = 1,sigma
   =1)$Z)
    set.seed(seed)
    param_data2[,dist]<-pbreplicate(repl, t.test(distn(n),side = "g",sigma=1)$
   statistic)
    }
  }
 if(encode==T){
   sign_data=gather(data.frame(sign_data),key="measure",value="Theoritical",distns
    )
   wilcox_data=gather(data.frame(wilcox_data),key="measure",value="Theoritical",
    distns)
   if(param==T){
     param_data1=gather(data.frame(param_data1),key="measure",value="Theoritical",
    distns)
     param_data2=gather(data.frame(param_data2),key="measure",value="Theoritical",
    distns)
     mylist=list("Sign_Data"=sign_data,"Wilcox_Data"=wilcox_data,"Normal"=param_
    data1,"t_test"=param_data2)
     return(mylist)}
   mylist=list("Sign_Data"=sign_data,"Wilcox_Data"=wilcox_data)
   return(mylist)
  }
  else{
    if(param==T){
      mylist=list("Sign_Data"=sign_data,"Wilcox_Data"=wilcox_data,"Normal"=param_
    data1,"t_test"=param_data2)
      return(mylist)}
    mylist=list("Sign_Data"=sign_data,"Wilcox_Data"=wilcox_data)
    return(mylist)
  }
}
main_data=qq_one(50,distns = c("Norm","Logis","Cauchy","Exp"),param=F)
require(ggplot2)
ggplot(main_data$Sign_Data,aes(sample=Theoritical))+geom_qq(color="blue",size=1,
    alpha=0.1)+geom_qq_line(color="red")+
  labs(title="Sign Test Statistic")+
  facet_wrap(~measure)+theme_mine() %+replace% theme(legend.position = "None")

ggplot(main_data$Wilcox_Data,aes(sample=Theoritical))+geom_qq(color="blue",size=1,
    alpha=0.1)+geom_qq_line(color="red")+
  labs(title="Wilcoxon Signed Rank\n Test Statistic")+
```

```
  facet_wrap(~measure)+theme_mine() %+replace% theme(legend.position = "None")




large_sample=qq_one(50,distns = c("Norm","Logis","Cauchy","Exp"),param=F,encode =
    F,repl=50,seed = 7)
large_sample1=qq_one(50,distns = c("Norm","Logis","Cauchy","Exp"),param=F,encode =
     T,repl=1000,seed = 7)


ggplot(data=data.frame(large_sample1$Sign_Data),aes(x=Theoritical))+
  geom_histogram(aes(y=..density..),fill="lightblue",colour="blue",bins=30)+
  geom_density(alpha=.3, fill="#FF6666",color="blue")+facet_wrap(~measure)+stat_
    function(fun=dnorm,color="red")+
  labs(title = "Sign Statistic",x="value")+theme_mine()

ggplot(data=data.frame(large_sample1$Wilcox_Data),aes(x=Theoritical))+
  geom_histogram(aes(y=..density..),fill="lightblue",colour="blue",bins=30)+
  geom_density(alpha=.3, fill="#FF6666",color="blue")+facet_wrap(~measure)+stat_
    function(fun=dnorm,color="red")+
  labs(title = "Wilcoxon Signed Rank\n Test Statistic",x="value")+theme_mine()




Chii(large_sample$Sign_Data,distns = "Cauchy")
Chii(large_sample$Sign_Data,distns = "Norm")
Chii(large_sample$Sign_Data,distns ="Logis")
Chii(large_sample$Sign_Data,distns ="Exp")


Chii(large_sample$Wilcox_Data,distns = "Cauchy")
Chii(large_sample$Wilcox_Data,distns = "Norm")
Chii(large_sample$Wilcox_Data,distns ="Logis")
Chii(large_sample$Wilcox_Data,distns ="Exp")



qq_two=function(n,m,repl=4000,distns,seed=43,param=T,encode=T){
  require(tidyr)
  med_data=matrix(0,nrow=repl,ncol= length(distns))
  wilcox_data=matrix(0,nrow=repl,ncol= length(distns))
  if(param==T){
    param_data1=matrix(0,nrow=repl,ncol= length(distns))
    colnames(param_data1)<-distns
  }
  colnames(med_data)<-distns
  colnames(wilcox_data)<-distns
  for(dist in distns){
    distn=get(paste0("r",dist))
    set.seed(seed)
    med_fun=function(){
      x=distn(n)
      y=distn(m)
      N=n+m
      combined=c(x,y)
      R=rank(combined,ties.method = "random")
      k=sum(combined<median(combined))
```

```
      p=(N-k)/N
      statistic=sum(sign(R[1:n]-((N+1)/2))+1)/2

        mn=n*p
        varian=(n*p*(1-p))#(1/(N-1))*(n/N)*(1-(n/N))*(N/(4*(N-1)))*(N-1)
        (statistic-mn)/sqrt(varian)
    }
    s=(pbreplicate(repl, med_fun(),simplify = T))
    med_data[,dist]<-s
    set.seed(seed)
    wilcox_data[,dist]<-pbreplicate(repl, ((wilcox.test(distn(n),distn(m),paired =
     F,alternative = "l")$statistic)-(m*n/2))/sqrt(n*m*(m+n+1)/12))
    if(param==T){
      set.seed(seed)
      param_data1[,dist]<-pbreplicate(repl, t.test(distn(n),distn(m),paired = F,
    alternative = "l")$statistic)

    }
  }
  if(encode==T){
    med_data=gather(data.frame(med_data),key="measure",value="Theoritical",distns)
    wilcox_data=gather(data.frame(wilcox_data),key="measure",value="Theoritical",
    distns)
    if(param==T){
      param_data1=gather(data.frame(param_data1),key="measure",value="Theoritical"
    ,distns)
      mylist=list("med_Data"=med_data,"Wilcox_Data"=wilcox_data,"t_test"=param_
    data1)
      return(mylist)}
    mylist=list("med_Data"=med_data,"Wilcox_Data"=wilcox_data)
    return(mylist)
  }
  else{
    if(param==T){
      mylist=list("med_Data"=med_data,"Wilcox_Data"=wilcox_data,"t_test"=param_
    data1)
      return(mylist)}
    mylist=list("med_Data"=med_data,"Wilcox_Data"=wilcox_data)
    return(mylist)
  }
}

main_data_two=qq_two(50,50,distns = c("Norm","Logis","Cauchy","Exp"),param=F,
    encode = T,seed = 7)
large_sample_two1=qq_two(500,500,distns = c("Norm","Logis","Cauchy","Exp"),param=F
    ,encode = T,repl=1000,seed=7)
large_sample_two=qq_two(500,500,distns = c("Norm","Logis","Cauchy","Exp"),param=F,
    encode = F,repl=50,seed=7)
large_sample_two_sep=qq_two(10,90,distns = c("Norm","Logis","Cauchy","Exp"),param=
    F,encode = F,repl=50,seed=7)
large_sample_two1_sep=qq_two(10,90,distns = c("Norm","Logis","Cauchy","Exp"),param
    =F,encode = T,repl=1000,seed=7)
large_sample_two2_sep=qq_two(90,10,distns = c("Norm","Logis","Cauchy","Exp"),param
    =F,encode = T,repl=1000,seed=7)
large_sample_two_sepp=qq_two(90,10,distns = c("Norm","Logis","Cauchy","Exp"),param
    =F,encode = F,repl=50,seed=7)
large_sample_two2_wilcox=qq_two(40,410,distns = c("Norm","Logis","Cauchy","Exp"),
    param=F,encode = T,repl=1000,seed=3)
```

```
large_sample_two_wilcox=qq_two(40,410,distns = c("Norm","Logis","Cauchy","Exp"),
    param=F,encode = F,repl=50,seed=3)

require(ggplot2)
ggplot(main_data_two$med_Data,aes(sample=Theoritical))+geom_qq(color="blue",size
    =1,alpha=0.1)+geom_qq_line(color="red")+
  labs(title="Median Test Statistic")+
  facet_wrap(~measure)+theme_mine() %+replace% theme(legend.position = "None")

ggplot(main_data_two$Wilcox_Data,aes(sample=Theoritical))+geom_qq(color="blue",
    size=1,alpha=0.1)+geom_qq_line(color="red")+
  labs(title="Wilcoxon Rank Sum Test Statistic")+
  facet_wrap(~measure)+theme_mine() %+replace% theme(legend.position = "None")



ggplot(data=data.frame(large_sample_two1$Wilcox_Data),aes(x=Theoritical))+
  geom_histogram(aes(y=..density..),fill="lightblue",colour="blue",bins=30)+
  geom_density(alpha=.3, fill="#FF6666",color="blue")+facet_wrap(~measure)+stat_
    function(fun=dnorm,color="red")+
  labs(title = "Wilcoxon Rank Sum Test Statistic\n n and m almost same value",x="
    value")+theme_mine()

ggplot(data=data.frame(large_sample_two1$med_Data),aes(x=Theoritical))+
  geom_histogram(aes(y=..density..),fill="lightblue",colour="blue",bins=30)+
  geom_density(alpha=.3, fill="#FF6666",color="blue")+facet_wrap(~measure)+stat_
    function(fun=dnorm,color="red")+
  labs(title = "Median Test Statistic\n n and m almost same value",x="value")+
    theme_mine()


ggplot(data=data.frame(large_sample_two1_sep$Wilcox_Data),aes(x=Theoritical))+
  geom_histogram(aes(y=..density..),fill="lightblue",colour="blue",bins=30)+
  geom_density(alpha=.3, fill="#FF6666",color="blue")+facet_wrap(~measure)+stat_
    function(fun=dnorm,color="red")+
  labs(title = "Wilcoxon Rank Sum Test Statistic\n n and m value differs very much
    ",x="value")+theme_mine()

ggplot(data=data.frame(large_sample_two1_sep$med_Data),aes(x=Theoritical))+
  geom_histogram(aes(y=..density..),fill="lightblue",colour="blue",bins=30)+
  geom_density(alpha=.3, fill="#FF6666",color="blue")+facet_wrap(~measure)+stat_
    function(fun=dnorm,color="red")+
  labs(title = "Median Test Statistic\n n and m value differs very much",x="value"
    )+theme_mine()

ggplot(data=data.frame(large_sample_two2_sep$med_Data),aes(x=Theoritical))+
  geom_histogram(aes(y=..density..),fill="lightblue",colour="blue",bins=30)+
  geom_density(alpha=.3, fill="#FF6666",color="blue")+facet_wrap(~measure)+stat_
    function(fun=dnorm,color="red")+
  labs(title = "Median Test Statistic\n n and m value differs very much",x="value"
    )+theme_mine()

ggplot(data=data.frame(large_sample_two2_sep$Wilcox_Data),aes(x=Theoritical))+
  geom_histogram(aes(y=..density..),fill="lightblue",colour="blue",bins=30)+
  geom_density(alpha=.3, fill="#FF6666",color="blue")+facet_wrap(~measure)+stat_
    function(fun=dnorm,color="red")+
  labs(title = "Median Test Statistic\n n and m value differs very much",x="value"
    )+theme_mine()
```

```
ggplot(data=data.frame(large_sample_two2_wilcox$med_Data),aes(x=Theoritical))+
  geom_histogram(aes(y=..density..),fill="lightblue",colour="blue",bins=30)+
  geom_density(alpha=.3, fill="#FF6666",color="blue")+facet_wrap(~measure)+stat_
    function(fun=dnorm,color="red")+
  labs(title = "Median Test Statistic",x="value")+theme_mine()









Chii(large_sample_two$med_Data,distns = "Cauchy",seed = 43)
Chii(large_sample_two$med_Data,distns = "Norm",seed = 43)
Chii(large_sample_two$med_Data,distns ="Logis",seed = 43)
Chii(large_sample_two$med_Data,distns ="Exp",seed = 43)


Chii(large_sample_two$Wilcox_Data,distns = "Cauchy")
Chii(large_sample_two$Wilcox_Data,distns = "Norm")
Chii(large_sample_two$Wilcox_Data,distns ="Logis")
Chii(large_sample_two$Wilcox_Data,distns ="Exp")



Chii(large_sample_two_sep$med_Data,distns = "Cauchy",seed = 43)
Chii(large_sample_two_sep$med_Data,distns = "Norm",seed=43)
Chii(large_sample_two_sep$med_Data,distns ="Logis",seed=43)
Chii(large_sample_two_sep$med_Data,distns ="Exp",seed=43)


Chii(large_sample_two_sep$Wilcox_Data,distns = "Cauchy")
Chii(large_sample_two_sep$Wilcox_Data,distns = "Norm")
Chii(large_sample_two_sep$Wilcox_Data,distns ="Logis")
Chii(large_sample_two_sep$Wilcox_Data,distns ="Exp")

Chii(large_sample_two_sepp$med_Data,distns = "Cauchy",seed = 43)
Chii(large_sample_two_sepp$med_Data,distns = "Norm",seed=43)
Chii(large_sample_two_sepp$med_Data,distns ="Logis",seed=43)
Chii(large_sample_two_sepp$med_Data,distns ="Exp",seed=43)

Chii(large_sample_two_sepp$Wilcox_Data,distns = "Cauchy")
Chii(large_sample_two_sepp$Wilcox_Data,distns = "Norm")
Chii(large_sample_two_sepp$Wilcox_Data,distns ="Logis")
Chii(large_sample_two_sepp$Wilcox_Data,distns ="Exp")

Chii(large_sample_two_wilcox$med_Data,distns = "Cauchy",seed = 43)
Chii(large_sample_two_wilcox$med_Data,distns = "Norm",seed=43)
Chii(large_sample_two_wilcox$med_Data,distns ="Logis",seed=43)
Chii(large_sample_two_wilcox$med_Data,distns ="Exp",seed=43)
```

Code 3: Distribution Free and Large Sample Normality

Here is the output of Shapiro-Wilk and Kolmogorov-Smirnov Test statistics for one sample location statistics.

```
> Chii(large_sample$Sign_Data,distns = "Cauchy")
$shapiro_test
```

```
  Shapiro-Wilk normality test

data:  x
W = 0.98309, p-value = 0.6874


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.1, p-value = 0.9639
alternative hypothesis: two-sided

> Chii(large_sample$Sign_Data,distns = "Norm")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.97422, p-value = 0.3405


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.2, p-value = 0.27
alternative hypothesis: two-sided

> Chii(large_sample$Sign_Data,distns ="Logis")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.98309, p-value = 0.6874


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.24, p-value = 0.1122
alternative hypothesis: two-sided

> Chii(large_sample$Sign_Data,distns ="Exp")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.9634, p-value = 0.1234


$ks_test
```

```
  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.16, p-value = 0.5441
alternative hypothesis: two-sided


> Chii(large_sample$Wilcox_Data,distns = "Cauchy")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.99191, p-value = 0.9804


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.12, p-value = 0.8643
alternative hypothesis: two-sided


> Chii(large_sample$Wilcox_Data,distns = "Norm")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.98191, p-value = 0.635


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.16, p-value = 0.5487
alternative hypothesis: two-sided


> Chii(large_sample$Wilcox_Data,distns ="Logis")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.97118, p-value = 0.2582


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.14, p-value = 0.7112
```

```
alternative hypothesis: two-sided

> Chii(large_sample$Wilcox_Data,distns ="Exp")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.99037, p-value = 0.9545


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.4, p-value = 0.0006709
alternative hypothesis: two-sided
```

Code 4: R Output For One Sample Location Test Statistic Normality Approaches

Here is the output of Shapiro-Wilk and Kolmogorov-Smirnov Test statistics for two sample location test statistics.

```
> Chii(large_sample_two$med_Data,distns = "Cauchy",seed = 43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.96094, p-value = 0.09738


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.22, p-value = 0.1777
alternative hypothesis: two-sided



> Chii(large_sample_two$med_Data,distns = "Norm",seed = 43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.9692, p-value = 0.2148


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.18, p-value = 0.3927
```

```
alternative hypothesis: two-sided



> Chii(large_sample_two$med_Data,distns ="Logis",seed = 43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.97211, p-value = 0.2814


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.16, p-value = 0.5441
alternative hypothesis: two-sided



> Chii(large_sample_two$med_Data,distns ="Exp",seed = 43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.96142, p-value = 0.1019


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.2, p-value = 0.27
alternative hypothesis: two-sided



> Chii(large_sample_two$Wilcox_Data,distns = "Cauchy")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.98971, p-value = 0.9391


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.12, p-value = 0.8643
alternative hypothesis: two-sided
```

```
> Chii(large_sample_two$Wilcox_Data,distns = "Norm")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.98067, p-value = 0.5806


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.1, p-value = 0.9639
alternative hypothesis: two-sided



> Chii(large_sample_two$Wilcox_Data,distns ="Logis")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.97968, p-value = 0.5385


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.2, p-value = 0.27
alternative hypothesis: two-sided



> Chii(large_sample_two$Wilcox_Data,distns ="Exp")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.98737, p-value = 0.8667


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.16, p-value = 0.5487
alternative hypothesis: two-sided


> Chii(large_sample_two_sep$med_Data,distns = "Cauchy",seed = 43)
```

```
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.93018, p-value = 0.005615


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.24, p-value = 0.1122
alternative hypothesis: two-sided


Warning message:
In ks.test(x, y) : cannot compute exact p-value with ties
> Chii(large_sample_two_sep$med_Data,distns = "Norm",seed=43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.95179, p-value = 0.04043


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.16, p-value = 0.5441
alternative hypothesis: two-sided


Warning message:
In ks.test(x, y) : cannot compute exact p-value with ties
> Chii(large_sample_two_sep$med_Data,distns ="Logis",seed=43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.91274, p-value = 0.001302


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.18, p-value = 0.3927
alternative hypothesis: two-sided


Warning message:
In ks.test(x, y) : cannot compute exact p-value with ties
```

```
> Chii(large_sample_two_sep$med_Data,distns ="Exp",seed=43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.92264, p-value = 0.002941


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.32, p-value = 0.01195
alternative hypothesis: two-sided


Warning message:
In ks.test(x, y) : cannot compute exact p-value with ties
>

> Chii(large_sample_two_sep$Wilcox_Data,distns = "Cauchy")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.9681, p-value = 0.1935


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.1, p-value = 0.9639
alternative hypothesis: two-sided



> Chii(large_sample_two_sep$Wilcox_Data,distns = "Norm")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.96114, p-value = 0.09927


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.2, p-value = 0.27
alternative hypothesis: two-sided
```

```
> Chii(large_sample_two_sep$Wilcox_Data,distns ="Logis")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.98876, p-value = 0.913


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.18, p-value = 0.3927
alternative hypothesis: two-sided



> Chii(large_sample_two_sep$Wilcox_Data,distns ="Exp")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.98941, p-value = 0.9314


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.1, p-value = 0.9639
alternative hypothesis: two-sided
> Chii(large_sample_two_sepp$med_Data,distns = "Cauchy",seed = 43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.93124, p-value = 0.006156


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.3, p-value = 0.02222
alternative hypothesis: two-sided



> Chii(large_sample_two_sepp$med_Data,distns = "Norm",seed=43)
$shapiro_test

  Shapiro-Wilk normality test
```

```
data:  x
W = 0.95029, p-value = 0.03505


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.32, p-value = 0.01195
alternative hypothesis: two-sided


> Chii(large_sample_two_sepp$med_Data,distns ="Logis",seed=43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.95337, p-value = 0.04702


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.32, p-value = 0.01195
alternative hypothesis: two-sided



> Chii(large_sample_two_sepp$med_Data,distns ="Exp",seed=43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.93556, p-value = 0.009024


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.32, p-value = 0.01195
alternative hypothesis: two-sided


> Chii(large_sample_two_sepp$Wilcox_Data,distns = "Cauchy")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.98636, p-value = 0.828
```

```
$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.24, p-value = 0.1122
alternative hypothesis: two-sided



> Chii(large_sample_two_sepp$Wilcox_Data,distns = "Norm")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.95772, p-value = 0.07134


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.2, p-value = 0.27
alternative hypothesis: two-sided


> Chii(large_sample_two_sepp$Wilcox_Data,distns ="Logis")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.98111, p-value = 0.5998


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.1, p-value = 0.9639
alternative hypothesis: two-sided


> Chii(large_sample_two_sepp$Wilcox_Data,distns ="Exp")
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.97216, p-value = 0.2827


$ks_test

  Two-sample Kolmogorov-Smirnov test
```

```
data:  x and y
D = 0.14, p-value = 0.7112
alternative hypothesis: two-sided


> Chii(large_sample_two_wilcox$med_Data,distns = "Cauchy",seed = 43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.97652, p-value = 0.4162


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.18, p-value = 0.3927
alternative hypothesis: two-sided


> Chii(large_sample_two_wilcox$med_Data,distns = "Norm",seed=43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.97477, p-value = 0.3575


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.12, p-value = 0.8643
alternative hypothesis: two-sided



> Chii(large_sample_two_wilcox$med_Data,distns ="Logis",seed=43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.97208, p-value = 0.2806


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.18, p-value = 0.3927
alternative hypothesis: two-sided
```

```
> Chii(large_sample_two_wilcox$med_Data,distns ="Exp",seed=43)
$shapiro_test

  Shapiro-Wilk normality test

data:  x
W = 0.97948, p-value = 0.5303


$ks_test

  Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.14, p-value = 0.7112
alternative hypothesis: two-sided
```

Code 5: R output For Two Sample Location Test Statistic Normality Approaches