

Power Efficiency of Sign Test and Wilcoxon Signed Rank Test Relative to T-Test

Usman, M.

Department of Statistics, Federal Polytechnic Bali, PMB 05 Bali, Nigeria.

Abstract

Many statistical tests require that your data follow a normal distribution. Sometimes this is not the case. In some instances it is possible to transform the data to make them follow a normal distribution; in others this is not possible or the sample size might be so small that it is difficult to ascertain whether or not the data a normally distributed. In such cases it is necessary to use a statistical test that does not require the data to follow a particular distribution. In the present study, we compared the power efficiency of one sample t-test, sign test and Wilcoxon signed rank test for small samples from normal, uniform, beta and exponential distributions for which the type I error, power of the test and power efficiency were computed. The comparison indicates that, the Wilcoxon signed rank test and sign test are as good as the t-test in a situation where the underlying distribution of the population is uncertain. For increasing sample size and increasing value of population mean, the non parametric tests achieved almost the same power with the t-test in all the selected distributions.

Keyword: One Sample, T-test, Sign test, Wilcoxon signed rank test, Type I error, Power of the test and Power efficiency

1. Introduction

Many of the hypothesis tests require normal distributed populations or some tests require that population variances be equal. What if, for a given test, such requirements cannot be met? For these cases, statisticians have developed hypothesis tests that are “distribution free.” Such tests are called nonparametric tests. Nonparametric tests were created to overcome this difficulty (Rosier 2004). Non-parametric statistical test is a test whose model does not specify conditions about the parameter of the population from which the sample was drawn. Certain assumptions are associated with most non-parametric statistical tests. That is, i) the observations are independent and ii) the variable under study has underlying continuity but these assumptions are fewer and weaker than those associated with parametric tests. The focal point of parametric is some population parameter for which the sampling statistics follows a known distribution, with measurements being made at the interval or ratio scale. When one or more of these requirements or assumptions are not satisfied, then non-parametric methods can be used, which focuses particularly on the fact that the distribution of the sampling statistics is not known (Kazmier 1996). In a non-parametric tests very few assumptions are made about the distribution underlying the data and, in particular, it is not assumed to be a normal distribution. Some statisticians prefer to use the term distribution-free rather than non-parametric to describe these tests (Clarke & Cooke 1998). There is a wide range of methods that can be used in different circumstances, but some of the more commonly used are the non-parametric alternative to the t-test. Non-parametric statistical tests are concerned with the application of statistical data in nominal or ordinal scale to problems in pure science, medicine, social science engineering, health science, agricultural science and other related fields.

Most of the present analysis carried out by non science and science oriented researchers are based on parametric test, and it is often reasonable to assume that observations come from a particular family of distributions. Moreover, experience backed by theory, suggest that for measurements, inferences based on the assumption that observations form a random sample from some normal distribution may not be misleading even if the normality assumption is incorrect, but this is not always true. (Sprent 1992).

Even when a parametric test does not depend too critically on an assumption that a sample comes from a distribution in particular family; if there is doubt, then a nonparametric test needing weaker assumption is preferable. Nonparametric methods are often the only ones available for data simply order, ranks or count in various categories. Weaker assumptions do not mean nonparametric methods are assumption free. In most statistical problems what can be deduced depends upon what assumption can validly be made. Nonparametric tests often are used in conjunction with small samples, because for such samples the central limit theorem cannot be invoked. Nonparametric tests can be directed toward hypothesis concerning the form, dispersion or location (median) of the population. In the majority of the applications, the hypothesis are concerned with the value of a median, the difference between medians or the differences among several medians. This contrasts with the parametric procedures that are focused principally on population means. If normal model cannot be assumed for the data then the tests of hypothesis on means are not applicable. Nonparametric tests were created to over come this difficulty. Nonparametric tests are often (but not always) based on the use of ranks; such as Wilcoxon rank test, Sign test Wilcoxon rank sum test, Kruskal wallis test, Kolmogorov test, etc.

The objective of this paper, therefore, is to compare the one sample t-test, sign test and Wilcoxon signed rank rest with respect to how they perform for different distributions in terms of type I error and power of the test, as well as to compare the power-efficiency of sign test and Wilcoxon signed rank test relative to the one sample t-test, using small samples sizes drawn from normal, uniform, beta and exponential distributions.

2. Materials and Methods

The materials used for the analysis were generated data using simulation procedures from the required distributions. Because, it is very difficult to get data that follows these distribution pattern, even if there is, but it is very difficult to get the required number of replicates for the sample sizes of interest. The nonparametric and parametric methods of analysis are applied, using t-test, sign test and Wilcoxon signed rank test to compare the performance of each test on the generated data from the normal, uniform, exponential, and beta distributions based on the underlying criteria for assessment.

2.1 Data Simulation Procedures

The SPSS package was used for designing simulation procedure in generating the data utilized for the analysis. Random samples were simulated from Normal, Uniform, Beta and Exponential distribution respectively for sample size 6, 10, 15, 20, and 25 which were considered as small sample sizes. Each test procedures were applied on the data sets at varying sample sizes and their Type I error and Power of the tests were studied in each situation. The process was repeated 500 times for each sample size considered and average of the results were displayed in Appendix I

2.2 Criteria for Assessment

A null hypothesis can be frequently tested by different statistical tests. It is, therefore necessary to employ some objective criterion for choosing among them. Generally, a statistical test is a good one if

- it has a probability of rejecting H_0 when H_0 is true close to α level and
- a large probability of rejecting H_0 when H_0 is false (power). (Lovrich 2002)

2.3 One Sample t-Test

In one sample case a common parametric techniques is to apply a t-test to the difference between the observed (sample) mean X and the expected (population) mean μ. The t-test strictly assumes that the observations in the sample have come from a normally distributed population. The t-test also requires the observation be measured at least in an interval scale (see Siegel and Castellan 1988).

When testing H_0: $\mu = \mu_0$ against H_0: $\mu \neq \mu_o$, large values of $|T|$ indicate significance and it is calculated using

$$T = \frac{\bar{x} - \mu_0}{S_x / \sqrt{n}} \qquad (1)$$

We reject H_0 if $|T| \geq t_{\alpha/2}$, that is critical value from tables, otherwise we do not reject H_0. When parametric methods are not applicable, an appropriate non-parametric procedure can be use, which is sign test or Wilcoxon signed rank test.

2.4 One Sample Sign Test

It is called 'sign test' because it allocates a sign, either positive(+) or negative(-), to each observation according to whether it is greater or less than some hypothesized value, and considers whether it is substantially different from what we would expect by chance. If any observations are exactly equal to the hypothesized value they are ignored and dropped from the sample size (Whitley & Ball 2002).

If the null hypothesis is true we would expect approximately equal number of positive(+) and negative(-), if either positive or negative signs predominate, there is evidence that the null hypothesis is false. As a test statistic we can use number of positive or negative signs, whichever is the smallest. Exact p values for the sign test are based on the binomial distribution, and many statistical packages provide these directly.

2.5 Wilcoxon Signed Rank Test

The sign test allocates a sign to each observation according to whether it lies above or below some hypothesized value, and does not take the magnitude of the observation into account. An alternative that does account for the magnitude of the observation is the Wilcoxon signed rank test. It ranks the absolute deviations of hypothesized value from individual observations (1 for the smallest and n for the largest) and attaches a negative sign to ranks corresponding to values below hypothesized value. It denote S_+ to be the sum of positive ranks and S_- the sum of negative ranks.

$$S_+ = \sum_i^n \Psi_i r|Z_i| \quad \text{where } \Psi_i = \begin{cases} 1 & \text{if } Z_i > 0 \\ 0 & \text{if } Z < 0 \end{cases} \tag{2}$$

or

$$S_- = \sum_i^n \Psi_i r|Z_i| \quad \text{where } \Psi_i = \begin{cases} 0 & \text{if } Z_i > 0 \\ 1 & \text{if } Z_i < 0 \end{cases} \tag{3}$$

$$i = 1 \ldots n$$

and $r|Z_i|$ is the rank of absolute value of Z_i's, where $Z_i = x_i - \mu_0$ and x_i represent the individual observations. When testing H_0 against $H_1 : \mu \neq \mu_0$, we reject H_0 if $S_+ \geq \frac{n(n+1)}{2} - t_{\alpha/2}$. For exact P value, that is $\Pr(S \leq S_+) = P$, it rejects H_0 if $P \leq \alpha$.

3. Data Analysis

Based on the assumption of the foregoing tests, a fixed significance level of 5% was selected for

$H_0 : \mu = \mu_0$ that is, the mean was the true mean of the sample used

against $H_0 : \mu \neq \mu_0$ the mean was not the true mean of the sample used (4)

where μ_0 represent the value of population mean for each of the distribution at four given levels. The test was carried out on 500 samples generated for each μ_0 and each distribution. The number of times H_0 (type I error) incorrectly rejected out of 500 was counted. These were recorded as probabilities for each of the three statistical tests, under the normal, uniform, beta and exponential distributions. Similar procedure was applied for the power of test, except that the μ_0 represent the smallest value of the population mean for each type of the distributions. For example, for the normal distribution, $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ for different values of generated μ_1; for the uniform distribution $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$; for beta distribution, $H_0 : \mu = 0.2$ against $H_1 : \mu \neq 0.2$; for the exponential distribution, $H_0 : \mu = 0.5$ against $H_1 : \mu \neq 0.5$, all for different values of generated μ_1. The test was carried out on the 500 samples generated for each μ_1 and each distribution. The number of times H_0 (power of the test) was rejected out of 500 was also counted. These were also recorded as probabilities for each of the three statistical tests under each of the given distributions. Averages of the type I error and power of the test for the levels of the population means were calculated and recorded for each of the sample size under the four given distributions for the three statistical tests.

4. Results and Discussion

The Tables 1a – 4b presented at the appendix shows the results of analyses using the one sample t-test, Wilcoxon signed rank test and sign test indicating how the three tests performed based on the type I error and power of the tests both being computed at the 5% level of significance in two tailed test. The average of each value (average) of the type I error and the power of the test were calculated and recorded under each statistical test for easy comparison.

For normal distribution, the one sample t-test has the closest value to alpha (α) level shown in Table 1a, and it is therefore considered the best statistical test, followed by Wilcoxon signed rank test at the 5% level of significance. From Table 1b, the one sample t-test has the largest power of 0.9467, and it is therefore consider as the most powerful test, followed by Wilcoxon signed rank test for the normal distribution at the 5% level of significance, based on the applied criteria for assessment reported by (Lovrich, 2002).

For uniform distribution, the one sample t-test has the average type I error that is closest to the alpha (α) level given in Table 2a. It is therefore considered the best statistical test in this study, followed by Wilcoxon signed rank test at the 5% level of significance. For Table 2b, the one sample Wilcoxon signed rank test and the sign test have the largest powers of 1, and were therefore considered as the most powerful tests for the uniform distribution at the 5% level of significance, based on the applied criteria for assessment reported by (Lovrich, 2002).

For beta distribution, the one sample Wilcoxon signed rank test has the average type I error that is closest to alpha (α) level shown in Table 3a, and it is therefore considered the best statistical test, followed by the one sample t-test at the 5% level of significance. For Table 3b, the one sample t-test has the largest power of 0.9496, and it is therefore considered as the most powerful test, followed by Wilcoxon signed rank test for the beta distribution at the 5% level of significance, based on the applied criteria for assessment reported by (Lovrich, 2002).

For exponential distribution, among the two probabilities, the one closest to α level is the one sample t-test value as given in Table 4a, and it is therefore considered the best statistical test, followed by Wilcoxon signed rank test at the 5% level of significance. For Table 4b, the one sample t-test has the largest power of 0.6947, and it is therefore considered as the most powerful test, followed by Wilcoxon signed rank test for the exponential distribution at the 5% level of significance, based on the applied criteria for assessment reported by (Lovrich, 2002).

The summary of the results obtained are presented in Tables 1 and 2. The average of the averages for the type I error and the power of the test from each of the tables were taken at the conventional 5% level of significance for the three statistical tests under the four selected distributions.

Table 1: Summary Table for Type I Error α=0.05

Distributions	*t-Test*	*Wilcoxon Signed Rank Test*	*Sign Test*
Normal	0.0409	0.0385	0.0308
Uniform	0.0547	0.0427	0.0381
Beta	0.0595	0.0563	0.0606
Exponential	0.0918	0.0997	0.1346
Averages	0.0617	0.0593	0.0660

Table 1 indicates that the average type I error of the Wilcoxon signed rank test is closest to the given α level followed by the t-test. Going by the objective criterion, it is the best among the three statistical tests, (without consideration to the type of distribution) for one independent sample, but statistically there is no significant differences between the two probabilities of type I error for the t-test and sign test as well, when rounded to one decimal place, if they are compared based on the simulated data from the four selected distributions, normal, uniform, beta and exponential, using small sample sizes at the 5% levels of significance. Therefore the t-test and sign test were also considered as good statistical tests as can be seen in Table 1.

Table 2: Summary table for the power of the test α=0.05

Distributions	*t-Test*	*Wilcoxon Signed Rank Test*	*Sign Test*
Normal	0.9467	0.9281	0.8844
Uniform	0.9967	1	1
Beta	0.9496	0.9253	0.8699
Exponential	0.6947	0.6729	0.5001
Averages	0.8969	0.8816	0.8136

Table 2 indicates that the one sample t-test is the test that has the largest average power among the three statistical tests followed by the Wilcoxon signed rank test used in analyzing the data simulated from the normal, uniform, beta and exponential distributions. This shows that if the three tests were compared based on the average power of the test computed (without consideration to the type of distribution) under these distributions, the t-test is the most powerful test.

The concept of power-efficiency is concerned with the amount of increase in sample size which is necessary to make test B as powerful as test A. If test A is the most powerful known test of its type, (when used with data which meet its conditions), and if test B is another test or the same research design which is just as powerful with N_b cases as is test A with N_a cases, then,

$$\text{Power-efficiency of test B} = (100)\frac{N_a}{N_b} \text{ percent, where } N \text{ is sample size} \quad (5)$$

Therefore, the results obtained for the power of the test were used to compute the power efficiency of the Wilcoxon signed rank test and the sign test relative to t-test. The method of linear interpolation with respect to the slope of the power curves at $\mu = 1$ and 0.4 was used to define some of the power efficiencies. The results obtained are given in Table 3

Table 3: Power Efficiencies of Sign test and Wilcoxon Signed Rank test Relative to t-Test

Tests	Distributions			
	Normal	*Uniform*	*Beta*	*Exponential*
Sign Test	0.6667	1	0.526	0.61
Wilcoxon Signed Rank Test	0.9375	1	0.910	0.75

From the table above, it can be observed that the Wilcoxon signed rank test was more efficient in the three distributions. The two tests have the same power efficiency in the uniform distribution.

5. Conclusion

The results obtained from this analysis for both type I error and power of the test, indicate that, the use of Wilcoxon signed rank test as a statistical tool in carrying out any hypothesis test for a fixed 5% significance level under the distributions concerned would not cause any drastic loss of power or efficiency, because there is no significant difference between the results from type I error for t-test i.e. 0.0617 and the result from the Wilcoxon signed rank test i.e 0.0593. But the difference between the power of the test result of t-test and sign test that is 0.8969 and 0.8136 respectively is negligible, therefore it can also be a good alternative to t-test in a situation where there is uncertainty in the underlying distribution of the population The computed power of the test shows an increase in power, for increasing sample size and increasing levels of population means as it can be seen in the tables 1b, 2b, 3b and 4b in all the four selected distribution for the three statistical tests.

In view of the result obtained for the power-efficiency of Wilcoxon signed rank test relative to t-test as shown in the table 6 above, indicates that, a sample size of 100 is needed for use of Wilcoxon signed rank test, compared with sample sizes of 94, 100, 91 and 75 in normal, uniform, beta and exponential distributions for use of the t-test to obtain the same result. Similarly a sample size of 100 is needed for use of the sign test compared with sample sizes of 67, 100, 53 and 61, in normal, uniform, beta and exponential distributions respectively for use of t-test to obtain the same result. Hence the low values of the power efficiency of the sign test relative to t-test in almost all the given distributions indicate the use of the Wilcoxon signed rank test rather than the sign test, when both are available and applicable. And it is only slightly less powerful than the one sample t-test, as can be observed from the result of the analysis.

References

Hodges J.I. and Lehmann E.I. (1956) The Efficiency of some Non-parametric Competitors of the t Test Annals of Mathematical Statistics Vol. 24, 467 – 493

Kazmier, L. J. (1996). *Schaum's Outline of Business Statistics* 3rd Edition Mc Graw-Hill New York p 372.

Lovrich M. (2002) Article on Non-parametric Methods Week 12b. University of South Australia. www.roma.unise.edu.au/10920/2002/lecture – notes wk12b

Rosier, S.(2004), "Statistics: The Sign Test", Mathematics Learning Support center. P 1

Siegel S. and Castellan N.J. (1988) Non-parametric Statistics for the Behavioural Science 2nd Edition New York: McGraw. Hill.

Sprent, P. (1993). *Applied Non-parametric Statistical Methods*. 2nd Edition. Chapman and Hall

Whitley, E. and Ball, J. (2002). Statistical Review 6: Non-parametric Methods. Biomed Central Ltd London. http://ccforum.com/content/6/6/509

Appendix I

Normal Distribution

Table 1a: Type I error α = 0.05

S/N	*Sample Size*	*T Test*	*Wilcoxon Test*	*Sign Test*
1	6	0.0445	0.0315	0.031
2	10	0.044	0.0465	0.0235
3	15	0.0435	0.041	0.0315
4	20	0.037	0.037	0.0305
5	25	0.0355	0.0365	0.0375
Averages		0.0409	0.0385	0.0308

Table 1b: Power of the test α = 0.05

S/N	*Sample Size*	*T Test*	*Wilcoxon Test*	*Sign Test*
1	6	0.6255	0.564	0.5665
2	10	0.712	0.7075	0.6375
3	15	0.747	0.743	0.7015
4	20	0.757	0.7565	0.7385
5	25	0.7575	0.76	0.7535
Averages		0.7198	0.7062	0.6795

Uniform Distribution

Table 2a: Type I error α = 0.05

S/N	*Sample Size*	*T Test*	*Wilcoxon Test*	*Sign Test*
1	6	0.0585	0.0285	0.0275
2	10	0.0555	0.0435	0.025
3	15	0.055	0.05	0.0425
4	20	0.055	0.0495	0.05
5	25	0.0495	0.042	0.0455
Averages		0.0547	0.0427	0.0381

Table 2b: Power of the test α = 0.05

S/N	*Sample Size*	*T Test*	*Wilcoxon Test*	*Sign Test*
1	6	0.7535	0.7585	0.7585
2	10	0.7625	0.7575	0.759
3	15	0.768	0.7615	0.7615
4	20	0.767	0.765	0.7615
5	25	0.762	0.761	0.759
Averages		0.7626	0.7607	0.7599

Beta Distribution

Table 3a: Type I error $\alpha = 0.05$

S/N	*Sample Size*	*T Test*	*Wilcoxon Test*	*Sign Test*
1	6	0.0685	0.04	0.0375
2	10	0.057	0.0605	0.0315
3	15	0.0533	0.055	0.0555
4	20	0.0635	0.0585	0.0805
5	25	0.055	0.0675	0.098
Averages		0.0595	0.0563	0.0606

Table 3b: Power of the test $\alpha = 0.05$

S/N	*Sample Size*	*T Test*	*Wilcoxon Test*	*Sign Test*
1	6	0.634	0.547	0.543
2	10	0.719	0.714	0.609
3	15	0.7585	0.758	0.6975
4	20	0.7685	0.771	0.743
5	25	0.763	0.7675	0.76
Averages		0.7286	0.7115	0.6705

Exponential Distribution

Table 4a: Type I error $\alpha = 0.05$

S/N	*Sample Size*	*T Test*	*Wilcoxon Test*	*Sign Test*
1	6	0.1085	0.0635	0.0645
2	10	0.0975	0.082	0.0625
3	15	0.0915	0.1105	0.1345
4	20	0.078	0.1045	0.178
5	25	0.0835	0.138	0.2335
Averages		0.0918	0.0997	0.1346

Table 4b: Power of the test $\alpha = 0.05$

S/N	*Sample Size*	*T Test*	*Wilcoxon Test*	*Sign Test*
1	6	0.239	0.217	0.217
2	10	0.4795	0.4925	0.3065
3	15	0.6175	0.588	0.443
4	20	0.669	0.629	0.5195
5	25	0.7175	0.722	0.5575
Averages		0.5445	0.5297	0.4087