A MONTE CARLO INVESTIGATION INTO THE PROPERTIES OF A PROPOSED ROBUST ONE-SAMPLE TEST OF LOCATION

by

Bercedis Peterson

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1418

October 1982

A MONTE CARLO INVESTIGATION INTO THE PROPERTIES OF A PROPOSED ROBUST ONE-SAMPLE TEST OF LOCATION

by

Bercedis Peterson

A paper submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Biostatistics.

Chapel Hill

1982

Approved by:

Ronald W. Helms
Adviser

Dana Quade
*Reader

M. J. Symons
Reader

*Adviser for Master's Paper

INTRODUCTION

It has been suggested that a robust test for the location of the median of a symmetric distribution can be obtained by incorporating a preliminary test of normality. If normality is accepted, a one-sample t test is then performed. On the other hand, if normality is rejected, the Wilcoxon signed-rank test may be performed. There are, however, some problems concerning this proposed procedure. One, when the null hypothesis is true, the resulting conditional distributions of the two test statistics cannot be presumed to be the t distribution in the one case or the Wilcoxon distribution in the other. This is true in the first case because the unconditional t distribution results from calculating t values on all samples from the normal distribution, even those samples that may not pass a normality test. Likewise, the unconditional Wilcoxon distribution results from Wilcoxon statistics calculated on the whole range of possible samples from symmetric distributions, even normal samples. On the other hand, the process of filtering the samples through a normality test may cause the resulting distribution of t statistics to show more resemblance to the true t distribution than to an unconditional distribution of t statistics. A second problem with this procedure is that when the null hypothesis is false, the powers of the conditional t and Wilcoxon statistics cannot be presumed equal to the powers of their respective unconditional statistics, and thus the overall power of the proposed procedure is unknown.

This paper examines the two conditional distributions, as well as the power of the proposed procedure, via a Monte Carlo simulation. Samples were generated from each of three different distributions, including the standard normal, and the t test and Wilcoxon test were performed, conditional upon the results of a test of normality. When the null hypothesis was true, the conditional t and Wilcoxon distributions were compared to the corresponding unconditional distributions resulting from performing the t and Wilcoxon tests on all samples unconditionally.

When samples were not from the standard normal, it was suspected that the conditional t distribution under the null hypothesis might show more resemblance to the true t distribution than to its corresponding unconditional t distribution derived from non-normal data. Therefore, for non-normal data, the conditional t distribution was also compared to the unconditional t distribution based on normal data.

When the null hypothesis for the median was false, the power of the proposed procedure was compared to the powers obtained when doing the t test and the Wilcoxon test unconditionally. Support for the proposed procedure would be obtained if it could be shown that it had at least as much power as the better of the two unconditional tests. Of course, the procedure could have more power than the unconditional tests only when used on data from non-normal distributions. That is, it is known that the t test is always the most powerful test when used on samples from the normal distribution. Thus, for normal data the t test has at least as much power as any procedure which uses the t test on some samples and the Wilcoxon test on the others.

To study the proposed procedure, samples of size 10 and 50 were generated from three distributions: the standard normal, Cauchy, and Laplace. These three distributions, being symmetric, are all valid for the Wilcoxon test. The t test, being ideal for use only with samples from the normal distribution, has been shown by Hodges and Lehmann (1956) to be inferior to the Wilcoxon test when used on samples from the Laplace distribution and, even more so, the Cauchy distribution. Specifically, these authors have found that the asymptotic relative efficiency of the Wilcoxon test with respect to the t test is $3/\pi$ for normal data, $3/2$ for Laplace data, infinity for Cauchy data, and at least 0.864 for data from any symmetric distribution.

For each sample generated from these three distributions, two different tests of normality at alpha levels of .20 and .10 were used. First, the kurtosis of the sample was calculated, and a p-value for this kurtosis was obtained from tables of kurtosis of samples of size 10 and 50 from the normal distribution. This test was used since literature has shown that population kurtosis is a good indicator of the adequacy of the t test (Scheffé, 1959). Second, the Lilliefors (1967) version of the Kolmogorov-Smirnov one-sample test was used. The Kolmogorov-Smirnov test is a standard goodness-of-fit test and unlike the chi-squared test, another standard test, does not require decisions concerning the number and location of intervals into which to categorize the data. The Lilliefors version of the Kolmogorov-Smirnov test uses the sample mean and standard deviation as estimators of μ and σ.

LITERATURE REVIEW

In the only article found in the literature relating to the present topic, Easterling and Anderson (1978) speculated that a distribution of t test statistics based on samples passing a preliminary normality test would be more t-like than a distribution of t statistics based on samples failing a normality test. To investigate this conjecture, they generated samples from a number of symmetric and asymmetric distributions until 1000 samples had failed the normality test and 1000 samples had passed. Using the chi-squared goodness of fit test, they then compared the resulting conditional t distributions to theoretical t distributions. Their results offered very little support for the use of a preliminary test of normality. That is, in general, preliminary testing did not result in a more t-like distribution for the samples passing the goodness of fit test than for samples failing the test.

The investigation of these two authors differed considerably from the present study. First, Easterling and Anderson took a larger number of samples and used different generating distributions and normality tests. More importantly, however, they did not consider the use of a nonparametric test on the samples failing the normality test, and they did not examine the power of using a preliminary test of normality. Whereas their main objective was to compare the true t distribution to the two conditional t distributions which resulted from the normality test, the main interest of this paper is to determine whether the use of a preliminary test of normality preserves the desired alpha level under the true null hypothesis and maintains power under the false null hypothesis. It seems rather unusual that Easterling and Anderson would want to compare the distribution of t statistics from samples that had failed a normality test to the true t distribution, since presumably the t test would never be used upon such samples.

METHOD

General Method

One thousand samples were generated for each combination of generating distribution and sample size. Thus, six generating programs were executed in total. To make the null hypothesis false, a constant was added to each observation; this constant, called the effect size, was thus the difference between the true median of the population and the hypothesized median. The t test and Wilcoxon test were then performed on each sample both with and without the added effect size. The distribution of all 1000 t or Wilcoxon test statistics for a given effect size constitutes an unconditional distribution. The Kolmogorov-Smirnov statistic and kurtosis were also calculated on each sample. Conditional distributions were obtained by selecting samples according to the results of the two normality tests using alpha levels of both .10 and .20. Thus, for each generating distribution and sample size, 16 conditional distributions were generated: 16 = (4 normality tests) × (2 effect sizes) × (2 location tests).

Each conditional and unconditional t and Wilcoxon distribution was examined by calculating the percentage of values falling outside critical values from the theoretical t or Wilcoxon distribution, respectively. These critical values correspond to two-sided alpha levels of .02, .05, .10, and .20. For effect sizes greater than zero, however, the interest was not in the power exhibited by the conditional t and Wilcoxon distributions separately, but in the power of the proposed procedure as a whole. To calculate this power, the weighted average of the two conditional powers was computed, weighting by the numbers of observations on each distribution.

When the null hypothesis was true, two statistical tests were made on the observed percentages falling outside the specified critical values. First, all conditional alphas were compared to their respective unconditional alphas. This test determined whether using a normality test to choose a test of location changed the distribution of the location test from what it would have been without the normality test. Second, alphas from the conditional t distributions were compared to alphas from the unconditional t distributions based on normal data. This test determined whether using a normality test to choose normal-appearing samples caused the resulting distribution of the t test statistic to become more similar to the true t distribution. These two statistical tests involved calculating a z-score as:

$$Z = \frac{|p_c - p_{\mu c}| - \frac{1}{2n}}{\sqrt{\frac{p_c(1-p_c)}{n}}}$$

where p_c = the conditional alpha and $p_{\mu c}$ = the unconditional alpha (Fleiss, 1981, p. 13). Critical values of 1.960 and 2.575 were used to test the hypothesis that $p_c = p_{uc}$. In other words, this statistical test treated $p_{\mu c}$ as a population value against which a sample value, p_c, was tested. This seemed reasonable, since the conditional distribution was a subset of the unconditional distribution.

When the null hypothesis was false, powers of the proposed procedure were compared to powers of both the unconditional t test and the unconditional Wilcoxon test. The z-score above was used to make these two comparisons.

Details

Random numbers from all three distributions were obtained by generating uniform random numbers (0,1) from IMSL's subroutine GGUBS and converting these uniform numbers into standard normal, Cauchy, and Laplace random variables. Since the same starting seed for GGUBS was used when generating samples of a given size from each of the three distributions, samples from all three distributions were, in a sense, related. Different seeds, however, were used when generating samples of size 10 and 50.

Uniform numbers were converted to standard normal variables using the IMSL subroutine MDNRIS. Laplace variables were generated from uniform numbers using the algorithm:

$$\text{IF } u < .5 \text{ THEN } x = \ln(2u)$$
$$\text{ELSE } x = -\ln(2-2u) \quad .$$

Cauchy variables were generated from uniform numbers by using the equation:

$$x = \tan(\pi * (u-.5)) \quad .$$

The Lilliefors version of the Kolmogorov-Smirnov test was performed on each sample by standardizing all random variables ($z = (x-\bar{x})/sd$) and then using IMSL's subroutine NKS1 to test whether these standardized variables were from the standard normal distribution. The statistic $\max(D^+, D^-)$ from NKS1 was compared to Lilliefors (1967) critical values. These critical values are given in Table 1. It is worth noting that, although not used in this paper, Stephens (1974) has provided a smoothing of the Lilliefors critical values to get a slightly different rejection region.

Kurtosis can be calculated in at least a couple of ways. The method discussed by Fisher (1970) and implemented by SAS uses unbiased estimators of the central moments. The method of this paper, however, was to calculate the well-known moment ratio:

$$b_2 = \frac{\sum_i (X_i - \bar{X})^4 / n}{\{\sum_i (X_i - \bar{X})^2 / n\}^2} \quad .$$

Percentage points of the distribution of b_2 for samples of various sizes are given in a table in Pearson and Hartley (1976), based upon work by Pearson (1963, 1965). However, since this table was incomplete for the needs of the present paper, simulated probability points of b_2 given by D'Agostino and Tietjen (1971) were used (see Table 1). Where the percentages and sample sizes of the latter paper overlapped those of Pearson and Hartley, critical values were either identical or within .01 of each other.

To make the null hypothesis false, effect sizes of .75 and .30 for samples of size 10 and 50, respectively, were used. For these effect sizes the two-sided, one-sample t test at alpha of .05 has a power of approximately

.56 for $n = 10$ and .55 for $n = 50$ when used on samples from a normal distribution with variance of 1. These powers were found using the Department of Biostatistics' interactive program SAMSIZ.

All critical values for the t test were found using the Department of Biostatistics' interactive program CFTDIS. Critical values for the Wilcoxon signed-rank test were found in McCornack (1965) and Wilcoxon, Katti, and Wilcox (1970). Of course, since the Wilcoxon distribution is discrete, the desired alpha levels could not be obtained exactly. Therefore, critical values were chosen so that the true alpha levels came as close as possible to the desired alpha levels without exceeding them.

Tests of the Random Number Generator

To check that the random numbers had the desired distributional properties, a number of simple tests of single proportions were performed on the unconditional t and Wilcoxon distributions. Since the normal, Laplace, and Cauchy distributions were all simulated from the same set of uniform random numbers, only samples from one of the distributions, the normal, were examined. The normal distribution was chosen, since nominal alphas and powers of the t and Wilcoxon tests for normal data were readily available. Powers of the t test were obtained from the Department of Biostatistics' interactive program SAMSIZ. Powers of the Wilcoxon test for samples of size 10 were obtained from an article by Klotz (1963). Powers of the Wilcoxon test for samples of size 50 were obtained by noting that the asymptotic efficiency of the Wilcoxon test compared to the t test is $3/\pi = .955$; this implies that the power of the Wilcoxon test for samples of size 50 is approximately the same as the power of the t test for samples of size 48.

RESULTS

Results of tests which examined the accuracy of the random number generator are given in Table 2. Under the true null hypothesis, three percentages for samples of size 10 were significantly different ($.05 > P > .01$) from their expected values. No such significant differences, however, existed for samples of size 50. Furthermore, none of the powers at either sample size was significantly different from its hypothesized value. These results implied that the random numbers had the desired distributional properties.

Percentage points of the unconditional and conditional distributions under the true null hypothesis are given in Tables 3 through 8. One table is presented for each combination of generating distribution and sample size. Tables 9 through 14 give powers of the proposed procedure, where power was calculated as the weighted average of the power of the conditional t test and the power of the conditional Wilcoxon test. Again, one table is presented for each generating distribution and sample size. For purposes of comparison, beneath Tables 9 through 14 are given powers for the unconditional t and Wilcoxon tests. In all tables, entries are superscripted to indicate the results of statistical tests which compared the entries to their respective unconditional proportions. However, since only eight Cauchy samples of size 50 passed either normality test, there were not enough conditional t test statistics in this case to warrant comparisons to the unconditional t distributions.

Before examining these tables further, notice that under the true null hypothesis the distribution of the unconditional Wilcoxon statistic for a given sample size was identical for all three generating distributions. This result is not as amazing as it seems. In generating the random variables, every uniform number was transformed into a standard normal, a Cauchy, and a Laplace random variable. The method used resulted in the relationship between the uniform numbers and their transformations being monotonically increasing with the uniform number .5 being transformed into the number 0 for all three distributions. Therefore, the Wilcoxon statistic, based on ranks, was identical for normal, Cauchy, and Laplace samples generated from the same uniform numbers.

Tables 3 through 8 indicate that only the conditional t distribution based on Cauchy samples of size 10 showed any major deviation from the unconditional distribution. Furthermore, none of these tables showed any serious deviation of the conditional t distribution from its corresponding unconditional t distribution based on normal data. As mentioned above, however, these two types of comparisons could not be made for Cauchy samples of size 50. In the case of the Cauchy samples of size 10, these results implied that the filtering of the data through a normality test caused the resulting conditional t distribution to become more like the true t distribution. The filtering process did not seem to have this same effect for Laplace data, since for such data the conditional t distribution was not significantly different from its own unconditional t distribution or from the t distribution based on normal data. In the case of normal samples, the conditional t distribution was found to be similar to the unconditional t distribution, indicating that the conditional distribution resembled the true t distribution. This conclusion could be drawn, since in the case of samples from the normal distribution, the unconditional t distribution actually is the true t distribution.

As indicated in Tables 9 through 14, the power of the conditional procedure was significantly different from the unconditional Wilcoxon test only for normal samples of size 10 at alpha levels of .10 and .20. On the other hand, the power of the conditional procedure was significantly different from the unconditional t test procedure at all four alpha levles for Cauchy samples of both sizes and for Laplace samples of size 50 ($P < .00001$). A couple of powers for Laplace samples of size 10 were also significantly different from the unconditional t test procedure.

That the conditional procedure was more powerful than the unconditional Wilcoxon test when used on normal samples of size 10 was not a surprise. Since for normal samples the t test is more powerful than the Wilcoxon test, it made sense that a mixture of t tests and Wilcoxon tests would be more powerful than just Wilcoxon tests unconditionally. For normal samples of size 50, however, the proposed procedure was not significantly different from either of the unconditional tests. This result was obtained even though statistical theory shows that for normal samples the asymptotic efficiency of the Wilcoxon test relative to the t test is $3/\pi = .955$. Examination of the elements in Table 12 shows that the powers of both the unconditional t test and the proposed procedure were always numerically larger than the powers of the unconditional Wilcoxon test; these differences, however, were in all cases nonsignificant.

For Laplace samples of size 10 a couple of powers of the proposed procedure at the .20 alpha level were significantly larger than powers of the unconditional t test. All other powers of the proposed procedure were numerically larger than the unconditional t test powers, although none of these differences was significant. No systematic differences between the proposed procedure and the unconditional Wilcoxon test were obvious. For Laplace samples of size 50 the differences hinted at for samples of size 10 were all quite obviously significant ($P < .00001$). That is, the proposed procedure was shown to be significantly more powerful than the unconditional t test. As in the case of samples of size 10, however, no difference between the proposed procedure and the Wilcoxon test was found.

All the powers of the proposed procedure based on Cauchy samples of both sizes were significantly larger than powers based on the unconditional t test ($P < .00001$). None of these powers was significantly different from the Wilcoxon powers. Since for the larger Cauchy samples the proposed procedure almost always allowed the normality test to be rejected and the Wilcoxon test to be performed, the proposed procedure was essentially synonymous with the Wilcoxon test. Thus, for Cauchy samples of size 50, these results verified what was known from theory: that the power of the t test is very low when samples are taken from the Cauchy distribution.

CONCLUSIONS

When the null hypothesis was true, this paper compared all conditional t distributions to unconditional t distributions based on normal data and found no extensive significant differences. Furthermore, only for Cauchy samples of size 10 was the conditional t distribution different from its own unconditional t distribution. These results implied that for Cauchy samples of this size the conditional t distribution is more like the true t distribution than is the t distribution resulting from unconditional application of the t test. Although this comparison could not be made for Cauchy samples of size 50 due to the paucity of samples passing the normality test, it was suspected that similar results might have been obtained. These results did not confirm those of Easterling and Anderson (1978) who found that conditional t distributions based on samples passing a normality test were, in general, not more t-like than conditional t distributions based on samples failing a normality test. Of course, this seeming inconsistency could be explained by noting that the present paper never examined t distributions based on samples failing a normality test; moreover, the present study used different generating distributions than did Easterling and Anderson.

That the preliminary test of normality allowed the conditional t distributions to maintain the alpha levels of the true t distribution was support for the proposed procedure. Further support lay in the fact that all conditional Wilcoxon distributions resembled their respective unconditional Wilcoxon distributions. That is, the conditional Wilcoxon distributions also preserved the desired alpha levels. The weighted average of the empirical alphas of the conditional t and Wilcoxon distributions are given for each generating distribution and sample size at the bottom of Tables 3 through 8.

Additional support for the use of a preliminary test of normality came from an examination of the power of this procedure. For normal samples of size 10 the power of the proposed procedure was often significantly different from the power of the Wilcoxon test but never significantly different from the power of the t test. Although no significant differences were found anywhere for normal samples of size 50, the power of the proposed procedure was always numerically larger than the power of the Wilcoxon test. On the other hand, for Cauchy samples of both sizes and for Laplace samples of size 50 the power of the proposed procedure was always significantly different from the power of the t test but never significantly different from the power of the Wilcoxon. For Laplace samples of size 10 the proposed procedure always had greater power than the t test, although only two such differences were significant; the proposed procedure was not significantly different from the Wilcoxon test. Now, given that statistical theory has shown the t test better for normal data and the Wilcoxon test better for Laplace and Cauchy data, the results from this paper indicated that the power of the proposed procedure was not significantly different from the better of the two tests. The results also showed that the power of the proposed procedure was, in general, significantly better than the power of the less preferred test.

The results in this paper can also be used to help judge the relative performance of the Kolmogorov-Smirnov test and kurtosis as preliminary tests of normality. Additionally, the choice of either .10 or .20 as the alpha level for these tests can be assessed. The preferred test would presumably do a better job of allowing the robust procedure proposed in this paper to preserve the desired alpha level under the true null hypothesis and obtain the greater power under the false null hypothesis. Examination of the weighted proportions in Tables 3 through 8 and the powers in Tables 9 through 14 indicated only a few small, unsystematic differences between the normality tests, and even these differences could not be separated from the variability inherent in the simulation process. Therefore, this study did not seem to recommend one preliminary normality test or alpha level over another.

Among the limitations of this study must be included the fact that only three distributions were studied. The normal distribution was chosen because the t test is perfectly suited for normal data; the Cauchy and Laplace distributions were chosen because the Wilcoxon test is entirely appropriate for data from these two distributions, although the t test is not. What happens to the proposed procedure, however, when samples are from distributions which are appropriate for neither of these location tests? For example, how would the proposed procedure perform if the data were from a uniform distribution? It could also be asked if there exist any distributions for which the proposed procedure would be better than either the t test or the Wilcoxon test. These questions suggest possibilities for further research.

REFERENCES

D'Agostino, R.B., and Tietjen, G.L. (1971), Simulation probability points of b_2 for small samples, Biometrika, 58, 669-672.

Easterling, R.G., and Anderson, H.E. (1978), The effect of preliminary normality goodness of fit tests on subsequent inference, Journal of Statistical Computation and Simulation, 8, 1-12.

Fisher, R.A. (1970), Statistical Methods for Research Workers, 14th ed., New York: Hafner Press.

Fleiss, J.L. (1981), Statistical Methods for Rates and Proportions, 2nd ed., New York: John Wiley & Sons.

Hodges, J.L., and Lehmann, E.L. (1956), The efficiency of some nonparametric competitors of the t test, Annals of Mathematical Statistics, 27, 324-334.

Lilliefors, H.W. (1967), On the Kolmogorov-Smirnov test for normality with mean and variance unknown, Journal of the American Statistical Association, 62, 399-402.

Klotz, J. (1963), Small sample power and efficiency for the one sample Wilcoxon and normal scores tests, Annals of Mathematical Statistics, 34, 624-632.

McCornack, R.L. (1965), Extended tables of the Wilcoxon matched pair signed rank statistic, Journal of the American Statistical Association, 60, 864-871.

Pearson, E.S. (1963), Some problems arising in approximating to probability distributions, using moments, Biometrika, 50, 95-111.

Pearson, E.S. (1965), Tables of percentage points of $\sqrt{b_1}$ and b_2 in normal samples; a rounding off, Biometrika, 52, 282-285.

Scheffé, H. (1959), The Analysis of Variance, New York: John Wiley & Sons, Inc.

Shapiro, S.S., Wilk, M.B., and Chen, H.J. (1968), A comparative study of various tests for normality, Journal of the American Statistical Association, 63, 1342-1372.

Stephens, M.A. (1974), EDF statistics for goodness of fit and some comparisons, Journal of the American Statistical Association, 69, 730-737.

Wilcoxon, F., Katti, S.K., and Wilcox, R.A. (1970), Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test, in Selected Tables in Mathematical Statistics, Volume 1, Chicago: Markham Publishing Company.

TABLE 1

Two-sided Critical Values

Test	Alpha	Critical Values for N = 10	Critical Values for N = 50
Kolmogorov-Smirnov	.20	.215	$.1041 = .736/\sqrt{50}$
Test	.10	.239	$.1138 = .805/\sqrt{50}$
Kurtosis	.20	1.68, 3.53	2.25, 3.62
	.10	1.56, 3.95	2.15, 3.99
T Test	.10	1.38303	1.29907
	.10	1.83311	1.67655
	.05	2.26216	2.00958
	.02	2.82144	2.40489
Wilcoxon Signed Rank	.20	14 (α = .1934)	503 (α = .1978)
Test	.10	10 (α = .0840)	466 (α = .0990)
	.05	8 (α = .0488)	434 (α = .0494)
	.02	5 (α = .0196)	397 (α = .0196)

TABLE 2

Check on the normal random number generator:
Comparison of the observed t and Wilcoxon signed-rank distributions with their expected distributions

Sample Size	Location Test	Effect Size	True Percentage	Observed Percentage	Check if Significantly Different (P < .05)
10	t test	0	.02	.015	
			.05	.039	
			.10	.079	✓
			.20	.173	✓
10	t test	.75	.38	.369	
			.56	.574	
			.71	.719	
			.84	.849	
50	t test	0	.02	.021	
			.05	.047	
			.10	.096	
			.20	.184	
50	t test	.30	.40	.401	
			.55	.548	
			.67	.684	
			.79	.801	
10	Wilcoxon	0	.0195	.015	
			.0488	.043	
			.084	.073	
			.1934	.165	✓
10	Wilcoxon	.75	.3617	.362	
			.5437	.558	
			.7013	.719	
			.8153	.820	
50	Wilcoxon	0	.0196	.019	
			.0494	.039	
			.099	.093	
			.20	.178	
50	Wilcoxon	.30	.38	.385	
			.53	.533	
			.66	.655	
			.78	.793	

TABLE 3

Proportion of rejections under the null hypothesis among samples of size 10 from the standard normal distribution

Test	N	Nominal value of α: .02	.05	.10	.20
Unconditional t	1000	.015	.039	.079	.173
Unconditional Wilcoxon	1000	.015	.043	.073	.165
(exact nominal α)		(.0195)	(.0488)	(.0840)	(.1934)
Conditional t (given passing the test of normality)					
KS with α = .10	920	.0163	.0391	.0826	.1783
KS with α = .20	805	.0149	.0373	.0808	.1752
b_2 with α = .10	898	.0156	.0390	.0780	.1693
b_2 with α = .20	816	.0159	.0392	.0772	.1667
Conditional Wilcoxon (given failing the test of normality)					
KS with α = .10	80	0	.0250	.0500	.1500
KS with α = .20	195	.0205	.0410	.0821	.1949
b_2 with α = .10	102	.0196	.0392	.0882	.2647[1]
b_2 with α = .20	184	.0163	.0435	.0870	.2446[1]
Proposed test					
KS with α = .10	1000	.015	.038	.080	.176
KS with α = .20	1000	.016	.038	.081	.179
b_2 with α = .10	1000	.016	.039	.079	.179
b_2 with α = .20	1000	.016	.040	.079	.181

[1]Proportion was significantly different from unconditional proportion (.05 > P > .01).

TABLE 4

Proportion of rejections under the null hypothesis among samples of size 10 from the Laplace distribution*

Test	N	Nominal value of α: .02	.05	.10	.20
Unconditional t	1000	.010	.037	.074	.175
Unconditional Wilcoxon	1000	.015	.043	.073	.165
(exact nominal α)		(.0195)	(.0488)	(.0840)	(.1934)
Conditional t (given passing the test of normality)					
KS with α = .10	778	.0129	.0411	.0823	.1877
KS with α = .20	650	.0154	.0385	.0785	.1862
b_2 with α = .10	797	.0100	.0389	.0765	.1819
b_2 with α = .20	689	.0116	.0406	.0769	.1872
Conditional Wilcoxon (given failing the test of normality)					
KS with α = .10	222	.0180	.0360	.0676	.1441
KS with α = .20	350	.0143	.0400	.0771	.1629
b_2 with α = .10	203	.0197	.0443	.0788	.1872
b_2 with α = .20	311	.0129	.0450	.0836	.1833
Proposed test					
KS with α = .10	1000	.014	.040	.079	.178
KS with α = .20	1000	.015	.039	.078	.178
b_2 with α = .10	1000	.012	.040	.077	.183
b_2 with α = .20	1000	.012	.042	.079	.186

*No conditional proportions were significantly different from their respective unconditional proportions or from unconditional proportions based on normal data (P > .05).

TABLE 5

Proportion of rejections under the null hypothesis among samples of size 10 from the Cauchy distribution

Test	N	Nominal value of α .02	.05	.10	.20
Unconditional t	1000	.005	.017	.053	.142
Unconditional Wilcoxon	1000	.015	.043	.073	.165
(exact nominal α)		(.0195)	(.0488)	(.0840)	(.1934)
Conditional t (given passing the test of normality)					
KS with α = .10	328	.0152	.0335	.0915[1]	.1829[1]
KS with α = .20	240	.0208	.0417[1]	.0197[1]	.1833
b_2 with α = .10	400	.0125	.0375[1]	.1025[11]	.2200[11,2]
b_2 with α = .20	316	.0127	.0411[1]	.1044[11]	.2247[11,2]
Conditional Wilcoxon (given failing the test of normality)					
KS with α = .10	672	.0134	.0417	.0774	.1563
KS with α = .20	760	.0132	.0395	.0737	.1592
b_2 with α = .10	600	.0117	.0317	.0650	.1483
b_2 with α = .20	684	.0132	.0366	.0687	.1550
Proposed test					
KS with α = .10	1000	.014	.039	.082	.165
KS with α = .20	1000	.015	.040	.078	.165
b_2 with α = .10	1000	.012	.034	.080	.177
b_2 with α = .20	1000	.013	.038	.080	.177

[1] Proportion was significantly different from unconditional proportion (.05 > P > .01).

[11] Proportion was significantly different from unconditional proportion (P < .01).

[2] Proportion was significantly different from unconditional proportion based on normal data (.05 > P > .01).

TABLE 6

Proportion of rejections under the null hypothesis among samples of size 50 from the standard normal distribution

Test	N	.02	.05	.10	.20
		Nominal value of α			
Unconditional t	1000	.021	.047	.096	.184
Unconditional Wilcoxon	1000	.019	.039	.093	.178
(exact nominal α)		(.0196)	(.0494)	(.0990)	(.1976)
Conditional t (given passing the test of normality)					
KS with α = .10	904	.0210	.0487	.0974	.1858
KS with α = .20	813	.0221	.0529	.0984	.1943
b_2 with α = .10	902	.0222	.0477	.0976	.1840
b_2 with α = .20	790	.0228	.0468	.0987	.1823
Conditional Wilcoxon (given failing the test of normality)					
KS with α = .10	96	.0104	.0208	.0938	.1875
KS with α = .20	187	.0054[1]	.0214	.0642	.1765
b_2 with α = .10	98	.0204	.0306	.0714	.2041
b_2 with α = .20	210	.0286	.0333	.0905	.2095
Proposed test					
KS with α = .10	1000	.020	.046	.097	.186
KS with α = .20	1000	.019	.047	.092	.191
b_2 with α = .10	1000	.022	.046	.095	.186
b_2 with α = .20	1000	.024	.044	.097	.188

[1]Proportion was significantly different from unconditional proportion (.05 > P > .01).

TABLE 7

Proportion of rejections under the null hypothesis among samples of size 50 from the Laplace distribution

Test	N	Nominal value of α .02	.05	.10	.20
Unconditional t	1000	.018	.052	.094	.191
Unconditional Wilcoxon	1000	.019	.039	.093	.178
(exact nominal α)		(.0196)	(.0494)	(.0990)	(.1976)
Conditional t (given passing the test of normality)					
KS with α = .10	458	.0153	.0524	.0961	.1769
KS with α = .20	311	.0161	.0515	.0965	.1704
b_2 with α = .10	395	.0177	.0631	.1136	.2096
b_2 with α = .20	268	.0187	.0709	.1381[1,2]	.2388[2]
Conditional Wilcoxon (given failing the test of normality)					
KS with α = .10	542	.0221	.0388	.0904	.1864
KS with α = .20	689	.0189	.0377	.0871	.1785
b_2 with α = .10	604	.0215	.0431	.0878	.1772
b_2 with α = .20	732	.0191	.0410	.0820	.1639
Proposed test					
KS with α = .10	1000	.019	.045	.093	.182
KS with α = .20	1000	.018	.042	.090	.176
b_2 with α = .10	1000	.020	.051	.098	.190
b_2 with α = .20	1000	.019	.049	.097	.184

[1]Proportion was significantly different from unconditional proportion (.05 > P > .01).

[2]Proportion was significantly different from unconditional proportion based on normal data (.05 > P > .01).

TABLE 8

Proportion of rejections under the null hypothesis among samples of size 50 from the Cauchy distribution*

Test	N	Nominal value of α .02	.05	.10	.20
Unconditional t	1000	.002	.019	.062	.183
Unconditional Wilcoxon	1000	.019	.039	.093	.178
(exact nominal α)		(.0196)	(.0494)	(.0990)	(.1976)
Conditional t (given passing the test of normality)					
KS with α = .10	6	0	0	0	.1667
KS with α = .20	5	0	0	0	0
b_2 with α = .10	3	0	0	0	0
b_2 with α = .20	2	0	0	0	0
Conditional Wilcoxon (given failing the test of normality)					
KS with α = .10	994	.0191	.0392	.0936	.1781
KS with α = .20	995	.0191	.0392	.0935	.1789
b_2 with α = .10	997	.0191	.0391	.0933	.1785
b_2 with α = .20	998	.0190	.0391	.0932	.1784
Proposed test					
KS with α = .10	1000	.019	.039	.093	.177
KS with α = .20	1000	.019	.039	.093	.178
b_2 with α = .10	1000	.019	.039	.093	.178
b_2 with α = .20	1000	.019	.039	.093	.178

*No conditional Wilcoxon proportions were significantly different from unconditional proportions (P .05). The conditional t distribution could not be compared to unconditional t distributions due to the small sample size.

TABLE 9

Power of the proposed procedure for samples of size 10 from the normal distribution

Goodness of Fit Test	Nominal value of α: .02	.05	.10	.20
KS with α = .10	.371	.576	.718[ww]	.851[ww]
KS with α = .20	.372	.575	.716[ww]	.846[w]
b_2 with α = .10	.366	.571	.715[ww]	.855[ww]
b_2 with α = .20	.359	.567	.710[w]	.850[w]
Unconditional t test	.369	.574	.720	.850
Unconditional Wilcoxon test	.362	.558	.678	.821

[w]Power was significantly different from power of the unconditional Wilcoxon test (.05 > P > .01).

[ww]Power was significantly different from power of the unconditional Wilcoxon test (P < .01).

TABLE 10

Power of the proposed procedure for samples of size 10 from the Laplace distribution

Goodness of Fit Test	Nominal value of α: .02	.05	.10	.20
KS with $\alpha = .10$	.254	.392	.540	.701
KS with $\alpha = .20$	.256	.404	.547	.706
b_2 with $\alpha = .10$	.251	.391	.554	.714[t]
b_2 with $\alpha = .20$	.253	.391	.557	.720[tt]
Unconditional t test	.231	.380	.532	.683
Unconditional Wilcoxon test	.252	.400	.524	.712

[t]Power was significantly different from power of the unconditional t test (.05 > P > .01).

[tt]Power was significantly different from power of the unconditional t test (P < .01).

TABLE 11

Power of the proposed procedure for samples of size 10 from the Cauchy distribution[tt]

	Nominal value of α			
Goodness of Fit Test	.02	.05	.10	.20
KS with α = .10	.100	.182	.290	.446
KS with α = .20	.107	.186	.290	.447
b_2 with α = .10	.096	.174	.281	.456
b_2 with α = .20	.098	.179	.287	.455
Unconditional t test	.049	.107	.186	.309
Unconditional Wilcoxon test	.112	.190	.285	.449

[tt]All powers were significantly different from powers of the unconditional t test (P < .01).

TABLE 12

Power of the proposed procedure for samples of size 50 from the normal distribution*

Goodness of Fit Test	Nominal value of α .02	.05	.10	.20
KS with $\alpha = .10$	.405	.546	.682	.801
KS with $\alpha = .20$	.398	.544	.674	.803
b_2 with $\alpha = .10$	.400	.551	.682	.802
b_2 with $\alpha = .20$	.405	.553	.679	.804
Unconditional t test	.401	.548	.684	.801
Unconditional Wilcoxon test	.385	.533	.655	.793

*No powers were significantly different from powers of the unconditional t test or the unconditional Wilcoxon test ($P > .05$).

TABLE 13

Power of the proposed procedure for samples of size 50 from the Laplace distribution[tt]

Goodness of Fit Test	Nominal value of α .02	.05	.10	.20
KS with α = .10	.266	.403	.535	.689
KS with α = .20	.279	.418	.545	.691
b_2 with α = .10	.283	.409	.540	.683
b_2 with α = .20	.288	.423	.549	.690
Unconditional t test	.205	.329	.447	.604
Unconditional Wilcoxon test	.284	.432	.558	.696

[tt]All powers were significantly different from powers of the unconditional t test ($P < .01$).

TABLE 14

Power of the proposed procedure for samples of size 50 from the Cauchy distribution[tt]

Goodness of Fit Test	Nominal value of α .02	.05	.10	.20
KS with α = .10	.122	.209	.307	.463
KS with α = .20	.122	.209	.307	.463
b_2 with α = .10	.122	.210	.308	.463
b_2 with α = .20	.122	.210	.308	.463
Unconditional t test	.013	.041	.078	.221
Unconditional Wilcoxon test	.122	.210	.308	.463

[tt]All powers were significantly different from powers of the unconditional t test ($P < .01$).