Replications and Extensions of the Results from *Computer-Intensive Methods for Tests About the Mean of an Asymmetrical Distribution* by Sutton (1993)

November 28, 2005

Victoria Plamadeala
vplamade@gmu.edu
Fall 2005

Abstract

This paper replicates the evaluation of the Johnson's modified t-test vs. the Student t-test about the mean of positevely skewed distributions reported in Sutton (1993). The original evaluation compares the probability of Type I error (accuracy) and the power (efficiency) of these two tests estimated in 40,000 Monte Carlo trials using small samples coming from positively skewed distributions. For upper-tailed tests, the Johnson's test is an absolute winner. For lower-tailed tests, although the Johnson's procedure is more accurate than the Student's, its estimated Type I error is still above the nominal level. Therefore, for lower-tailed tests, an alternative testing procedure based on the bootstrap-t methodology is applied to the same data and compared to the Johnson's procedure in terms of the same criteria, the estimated probability of Type I error and power.

1 Introduction

One-sided tests about the mean of a distribution are commonly needed in applied work. The most popular parametric procedure to serve this purpose is the Student t-test, which many use as second nature. Any parametric procedure, including the Student t-test, relies on several assumptions about the distribution underlying the observed data. A violation of the distribution assumptions by the test data may lead to inflated Type I error and decreased power. In the case of the Student t-test, Johnson (1978) pointed to the effect of the symmetry violation and put forward a modified version of the t-test as a solution. Sutton (1993) compared the performance of the Student t-test and the Johnson's procedure on skewed data in an extensive Monte Carlo study. This paper replicates Sutton's reported comparison.

The first part relates three alternative procedures to test the mean of a skewed distribution, including the Student's and Johnson's procedures. The second part is a summary of Sutton (1993), followed by the replication, analysis and comparison of the results in that study. The last part is an application of the bootstrap-t methodology (Efron and Tibshirani 1993) in testing the mean of skewed distributions and a comparison of the bootstrap-t methodology with the Johnson's and Student's procedures.

2 Testing procedures for the mean when the normality assumptions are violated

One of the assumptions under the one-sided Student t-test about the mean is that the parent distribution underlying the observed data is normally distributed, which implies that the distribution is symmetric. In presence of skewness, a violation of the above-mentioned assumption, there are at least three options for a test about the mean.

1. Apply the Student t-test statistic, $t = \sqrt{n}(\bar{x} - \mu_x)/S$, and compare the t statistic to the quantiles of the T distribution with $n-1$ d.f., ignoring the skewness.

2. Modify the Student t-test statistic, the way it was suggested by Johnson (1978), so that it accounts for the skewness in the data

$$t_1 = t + \tfrac{\hat{\mu}_3}{6S^2\sqrt{n}} + \tfrac{\hat{\mu}_3}{3S^5}\sqrt{n}(\bar{x} - \mu_x)^2,$$

where $\hat{\mu}_3$ is the estimated sample skewness, and compare t_1 to the quantiles of the T distribution with $n-1$ d.f.

3. Use a bootstrap-t testing procedure which estimates t, $t = \sqrt{n}(\bar{x} - \mu_x)/S$, in the skewed data, and compare t to the distribution of

$$\hat{t} = (\bar{x}^{*j} - \bar{x})/\hat{se}(\bar{x}^{*j}).$$

> The distribution of $\hat{t}$ is estimated via bootstrap samples from the observed data. The t statistic is compared to the quantiles of the EPDF of $\hat{t}$. [1]

Those who apply the first testing procedure rely on the fact that "[for] one-sided tests about the mean of a skewed distribution, the t test is asymptotically robust for validity." [3] This is not so, however, in many practical situations and, hence, alternatives to the t-test have been searched.

The second testing procedure modifies the t-test, yet it keeps the same reference distribution as in the first case. It has showed an improvement over the t-test in many cases.

The third testing procedure keeps the same statistic, however estimates its reference distribution. The bootstrap-t method was applied in this report as an alternative test about the mean when the t-test assumptions of normality are not met.

2.1 Summary of Sutton (1993)

Sutton (1993) evaluated the Johnson's modified t-test vs. the Student t-test in a Monte Carlo study. His study also contains the evaluation of a technique based on bootstrap resampling known as the *normal approximation method*, which was propounded as yet another testing procedure for the mean of a skewed distribution. This report replicates the evaluation of the Johnson's and Student procedures only. Below is a summary of this evaluation.

2.1.1 The Monte Carlo experiment in the evaluation

In order to compare the performance of the Student's t-test with the Johnson's procedure (*two factors* in the experiment), the author considered samples coming from 19 positively skewed distributions with different values of skewness, γ_3, at 10 different sample sizes. (This is a total of 190 combinations of distribution and sample size. The distributions and sample sizes constitute *the levels* in the experiment.) The performance of the tests was evaluated in terms of the Type I error and power - *the response* of the experiment.

For each of the 190 cases mentioned above 40,000 random samples were generated. In each case the Student's and the Johnson's modified t-test were conducted about the sample mean being equal to the true population mean μ_x in 40,000 trails. (The motivation for this Monte Carlo sample size was: For a test with Type I error of 0.01, 38,032 trails are required to obtain an estimate of the Type I error rate that is within 10 percent of the true value with a probability of 0.95.[3]) Both lower- and the upper- tailed tests at $\alpha = 0.1$, 0.05 and 0.01 were conducted. The estimated probability of Type I error was the proportion of rejections in 40,000 trials, since these were false rejections. The power was estimated as follows. For an upper-tailed test of the Student's and Johnson's procedures, the sample mean was compared to $\mu_x - k\sigma_x/\sqrt{n}$ ($\mu_x + k\sigma_x/\sqrt{n}$ for a lower-tailed test), where $k = 1$, 1.5, 2 and 2.5 and $\sigma_x/\sqrt{n}$ is the standard deviation of the true population mean. The proportion of rejections in 40,000 trials was the estimated power of the test, since these were correct rejections. Finally, the testing procedures were compared in base of the estimated Type I error probability and power.

An example

This is an example of how Monte Carlo was used in the original study as well as in the replications of this report. Take the exponential distribution with $\lambda = 1$, for which the skewness is $\gamma_3 = 2.0$, and the sample size 10. This is one of the 190 cases in the study. The true population mean is $\mu_x = \frac{1}{\lambda} = 1$, and the true population standard deviation is $\sigma_x = \sqrt{\frac{1}{\lambda^2}} = 1$. Generate 40,000 samples of size 10 from this distribution. Conduct the following hypothesis test on all 40,000 samples using both the Student's and the Johnson's procedures:

- the lower-tailed test is $H_0 : \mu_0 = 1$ vs. $H_a : \mu_0 < 1$;
- the upper-tailed test is $H_0 : \mu_0 = 1$ vs. $H_a : \mu_0 > 1$.

Estimate the proportion of rejections at $\alpha = 0.05$ (then 0.01 and 0.01) in 40,000 trials for each of the four cases (Student's lower-tailed test, Student's upper-tailed test, Johnson's lower-tailed test and Johnson's upper-tailed test). These will be the estimated probabilities of Type I error under each testing case. Compare the Johnson's and Student's procedures based on the estimated Type I error probabilities. Lower probability of Type I error is better, and lower than the nominal level is best.

To obtain the power estimates under each procedure, conduct the following hypothesis tests on the same data, i.e. 40,000 times again using both the Student's and the Johnson's procedures:

- the lower-tailed test is $H_0 : \mu_0 = 1$ vs. $H_a : \mu_0 < 1 + k\frac{1}{\sqrt{n}}$;

- the upper-tailed test is $H_0 : \mu_0 = 1$ vs. $H_a : \mu_0 > 1 - k\frac{1}{\sqrt{n}}$.

Use different values of k from $\{1, 1.5, 2, 2.5\}$, which specifies that the mean under the alternative is k standard deviations away from the true mean. The estimated power is the proportion of rejections at $\alpha = 0.05$ (then 0.01 and 0.01) in 40,000 trials, since these will be correct rejections. A testing procedure with high power should be able to reject often even for $k = 1$. Compare the Johnson's and Student's procedures based on the power estimates. Larger power is better.

2.1.2 The results of the evaluation

Although the study was extensive, the results actually reported in the paper were narrowed down to only five distributions and five sample sizes in case of the Type I error probability, and only three distributions and one sample size in case of the power comparisons.

The Type I error probabilities were reported for the following distributions: Weibull ($a = 1, b = 2$)($\gamma_3 = 0.631$), the Square Root Normal Distribution (SRN)($\gamma_3 = 0.83$), Chi-squared with 3 d.f. ($\gamma_3 = 1.633$), Lognormal ($\gamma_3 = 2.89$) and ($\gamma_3 = 6.18$). The results were reported for the following sample sizes: 20, 40, 80, 160 and 320. This is a total of 15 distribution and sample size combinations. For lower-tailed rejections regions the Johnson's procedure had lower estimated Type I errors both at $\alpha = 0.01$ and 0.05 in all 15 combinations. However for upper-tailed rejections regions, the Johnson's procedure had slightly larger Type I error probabilities for all 15 cases, both at $\alpha = 0.01$ and 0.05. Nevertheless, all were well under the nominal levels.

The power comparisons were reported only for the upper-tailed rejections regions for Weibull ($a = 1, b = 2$)($\gamma_3 = 0.631$), Chi-squared with 3 d.f. ($\gamma_3 = 1.633$) and Lognormal ($\gamma_3 = 6.18$). The results were reported only for $n = 20$, at $k = 1$, 1.5, 2 and 2.5 . In all cases the Johnson's modified t-test had higher power, both at $\alpha = 0.01$ and 0.05 .

The paper concludes that in presence of skewness and small sample sizes (the latter being often the fate in applied work) the Student t-test has higher Type I error rates than the Johnson's procedure. It is also less powerful than the Johnson's test. The modified t-test's improvement over the Student's test is not without shortcomings. For very small sample sizes and considerable skewness the Johnson's modified t-test can also be inaccurate, i.e. have higher Type I error rates than the nominal levels.

3 Replication and Comparison of Results

The pseudo-random variates from various distributions in this replication were generated using the respective random number generating functions in **R**. Variates from the SRN distribution were generated using variates from a normal distribution, as described below. The seed in **R**, 100, was controlled at the beginning of every generation of 40,000 samples of a certain size and from a certain distribution.

A few comments ought to be made about how variates from the SRN ($\gamma_3 = 0.83$) were generated. In general, random variables following an SRN distribution are obtained by squaring normally distributed random variables. Sutton (1993) mentions that the respective SRN distribution has $\mu_x = 1.0$, with σ_x chosen such that the skewness coefficient is 0.83. Given these, the problem was to find μ_y and σ_y of the corresponding normal distribution. For this purpose, define $X = Y^2$, where $X \sim SRN(\mu_x,\ \sigma_x)$, $Y \sim N(\mu_y,\ \sigma_y)$ and $\gamma_3 = \frac{E([X-1]^3)}{\sigma_x^3}$. The following is true:

$$\mu_x = E(X) = E(Y^2) = \sigma_y^2 + \mu_y^2 = 1.$$

It follows that the parameters μ_y and σ_y are the solutions to this system:

$$\begin{cases} \sigma_y^2 + \mu_y^2 = 1 \\ \frac{E([X-1]^3)}{\sigma_x^3} = 0.83 \end{cases} \Leftrightarrow \begin{cases} \sigma_y^2 + \mu_y^2 = 1 \\ \frac{E(X^3)-3E(X^2)+3E(X)-1}{(E[(X-1)^2])^{3/2}} = 0.83 \end{cases} \Leftrightarrow$$

$$\begin{cases} \sigma_y^2 + \mu_y^2 = 1 \\ \frac{E(Y^6)-3E(Y^4)+3E(Y^2)-1}{[E(Y^4)-1]^{3/2}} = 0.83 \end{cases} \Leftrightarrow$$

$$\begin{cases} \sigma_y^2 + \mu_y^2 = 1 \\ \frac{(\mu_y^6+15\sigma_y^2\mu_y^4+45\sigma_y^4\mu_y^2+15\sigma_y^6)-3(\mu_y^4+6\mu_y^2\sigma_y^2+3\sigma_y^4)+3-1}{[\mu_y^4+6\mu_y^2\sigma_y^2+3\sigma_y^4-1]^{3/2}} = 0.83 \end{cases} \Leftrightarrow \begin{cases} \mu_y \approx |0.96145| \\ \sigma_y \approx 0.27466 \end{cases}.$$

Therefore, by squaring the pseudo-random variates from $N(0.96145,\ 0.27466)$ the outcome is pseudo-random variates following an SRN distribution with $\mu_x = 1$, $\sigma_x = 0.54$ and $\gamma_3 = 0.83$. The tables that follow contain the replicated results for the reported counterparts in Sutton (1993), plus results for the exponential distribution with $\lambda = 1$. A graphical summary of the results for the Lognormal distribution ($\gamma_3 = 6.18$) is in the appendix. An analysis of the replications is at the end of this section.

Table 1. Est. Type I error in $H_0: \mu_0 = \mu_x$ vs. $H_a: \mu_0 < \mu_x$

Distribution (skewness)	$n = 10$	$n = 20$	$n = 40$	$n = 80$	$n = 160$	$n = 320$
			$\alpha = 0.01$			
Exponential	0.023	0.015	0.012	0.011	0.010	0.010
($\gamma_3 = 2.0$)	0.063	0.047	0.035	0.026	0.020	0.017
Weibull	0.015	0.012	0.010	0.010	0.010	0.010
($\gamma_3 = 0.63$)	0.021	0.019	0.017	0.014	0.013	0.012
SRN	0.017	0.012	0.011	0.011	0.011	0.010
($\gamma_3 = 0.83$)	0.025	0.021	0.018	0.017	0.014	0.013
Chi-squared	0.024	0.016	0.013	0.011	0.010	0.010
($\gamma_3 = 1.63$)	0.051	0.039	0.029	0.023	0.018	0.015
Lognormal	0.033	0.023	0.017	0.016	0.013	0.012
($\gamma_3 = 2.89$)	0.066	0.053	0.042	0.034	0.026	0.021
Lognormal	0.041	0.029	0.023	0.021	0.017	0.015
($\gamma_3 = 6.18$)	0.115	0.094	0.073	0.055	0.042	0.032

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 2. Est. Type I error in $H_0: \mu_0 = \mu_x$ vs. $H_a: \mu_0 < \mu_x$

Distribution (skewness)	$n = 10$	$n = 20$	$n = 40$	$n = 80$	$n = 160$	$n = 320$
			$\alpha = 0.05$			
Exponential	0.066	0.058	0.055	0.052	0.049	0.050
($\gamma_3 = 2.0$)	0.134	0.108	0.091	0.078	0.068	0.063
Weibull	0.054	0.052	0.052	0.052	0.050	0.050
($\gamma_3 = 0.63$)	0.074	0.068	0.063	0.060	0.056	0.053
SRN	0.059	0.053	0.052	0.052	0.052	0.052
($\gamma_3 = 0.83$)	0.080	0.071	0.065	0.062	0.061	0.057
Chi-squared	0.069	0.060	0.056	0.052	0.050	0.051
($\gamma_3 = 1.63$)	0.118	0.098	0.085	0.073	0.065	0.060
Lognormal	0.087	0.075	0.068	0.061	0.058	0.055
($\gamma_3 = 2.89$)	0.144	0.122	0.103	0.090	0.079	0.071
Lognormal	0.101	0.090	0.081	0.074	0.067	0.061
($\gamma_3 = 6.18$)	0.206	0.173	0.144	0.121	0.102	0.089

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 3. Est. Type I error in $H_0 : \mu_0 = \mu_x$ vs. $H_a : \mu_0 > \mu_x$

Distribution (skewness)	$n = 10$	$n = 20$	$n = 40$	$n = 80$	$n = 160$	$n = 320$
			$\alpha = 0.01$			
Exponential	0.003	0.005	0.007	0.009	0.010	0.010
($\gamma_3 = 2.0$)	0.001	0.001	0.002	0.003	0.004	0.006
Weibull	0.007	0.007	0.008	0.010	0.010	0.010
($\gamma_3 = 0.63$)	0.005	0.005	0.006	0.008	0.008	0.009
SRN	0.006	0.008	0.008	0.009	0.010	0.009
($\gamma_3 = 0.83$)	0.004	0.005	0.005	0.006	0.007	0.007
Chi-squared	0.003	0.005	0.009	0.010	0.010	0.009
($\gamma_3 = 1.63$)	0.001	0.002	0.003	0.004	0.005	0.006
Lognormal	0.002	0.005	0.007	0.008	0.010	0.010
($\gamma_3 = 2.89$)	0.001	0.001	0.001	0.002	0.003	0.004
Lognormal	0.003	0.002	0.004	0.006	0.008	0.009
($\gamma_3 = 6.18$)	0.000	0.000	0.000	0.001	0.001	0.002

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 4. Est. Type I error in $H_0 : \mu_0 = \mu_x$ vs. $H_a : \mu_0 > \mu_x$

Distribution (skewness)	$n = 10$	$n = 20$	$n = 40$	$n = 80$	$n = 160$	$n = 320$
			$\alpha = 0.05$			
Exponential	0.029	0.038	0.045	0.049	0.050	0.052
($\gamma_3 = 2.0$)	0.014	0.019	0.025	0.033	0.037	0.041
Weibull	0.038	0.045	0.046	0.048	0.051	0.049
($\gamma_3 = 0.63$)	0.033	0.038	0.038	0.043	0.046	0.045
SRN	0.040	0.044	0.047	0.049	0.049	0.049
($\gamma_3 = 0.83$)	0.031	0.035	0.037	0.040	0.042	0.045
Chi-squared	0.031	0.040	0.046	0.050	0.051	0.050
($\gamma_3 = 1.63$)	0.018	0.023	0.030	0.035	0.038	0.040
Lognormal	0.029	0.039	0.043	0.046	0.050	0.049
($\gamma_3 = 2.89$)	0.012	0.016	0.021	0.025	0.031	0.035
Lognormal	0.024	0.033	0.037	0.043	0.046	0.048
($\gamma_3 = 6.18$)	0.006	0.008	0.012	0.016	0.022	0.026

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 5. Est. Power in $H_0 : \mu_0 = \mu_x + k\frac{\sigma_x}{\sqrt{n}}$ **vs.** $H_a : \mu_0 < \mu_x + k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.01, n = 10$				$\alpha = 0.01, n = 20$			
Exponential	0.073	0.107	0.142	0.176	0.063	0.104	0.153	0.202
($\gamma_3 = 2.0$)	0.210	0.310	0.421	0.531	0.188	0.299	0.424	0.551
Weibull	0.073	0.130	0.205	0.293	0.074	0.145	0.247	0.374
($\gamma_3 = 0.63$)	0.109	0.196	0.312	0.448	0.113	0.212	0.352	0.509
SRN	0.080	0.139	0.214	0.297	0.078	0.149	0.246	0.362
($\gamma_3 = 0.83$)	0.122	0.216	0.336	0.468	0.121	0.226	0.362	0.515
Chi-squared	0.081	0.124	0.171	0.217	0.072	0.123	0.185	0.256
($\gamma_3 = 1.63$)	0.183	0.283	0.392	0.504	0.169	0.277	0.405	0.537
Lognormal	0.110	0.156	0.197	0.234	0.096	0.148	0.204	0.255
($\gamma_3 = 2.89$)	0.243	0.359	0.475	0.581	0.222	0.341	0.469	0.591
Lognormal	0.110	0.134	0.152	0.161	0.091	0.120	0.141	0.152
($\gamma_3 = 6.18$)	0.358	0.476	0.582	0.667	0.317	0.444	0.564	0.663

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 6. Est. Power in $H_0 : \mu_0 = \mu_x + k\frac{\sigma_x}{\sqrt{n}}$ **vs.** $H_a : \mu_0 < \mu_x + k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.01, n = 40$				$\alpha = 0.01, n = 80$			
Exponential	0.068	0.122	0.194	0.279	0.073	0.144	0.246	0.369
($\gamma_3 = 2.0$)	0.164	0.278	0.417	0.559	0.146	0.262	0.407	0.565
Weibull	0.079	0.165	0.292	0.452	0.085	0.177	0.322	0.495
($\gamma_3 = 0.63$)	0.110	0.220	0.368	0.540	0.108	0.217	0.375	0.553
SRN	0.081	0.164	0.284	0.431	0.084	0.176	0.314	0.484
($\gamma_3 = 0.83$)	0.119	0.229	0.375	0.541	0.112	0.225	0.383	0.559
Chi-squared	0.074	0.138	0.227	0.334	0.077	0.155	0.265	0.404
($\gamma_3 = 1.63$)	0.152	0.268	0.405	0.555	0.137	0.250	0.397	0.559
Lognormal	0.091	0.154	0.231	0.311	0.090	0.164	0.263	0.370
($\gamma_3 = 2.89$)	0.195	0.317	0.454	0.591	0.169	0.293	0.439	0.585
Lognormal	0.092	0.134	0.175	0.205	0.097	0.158	0.227	0.287
($\gamma_3 = 6.18$)	0.275	0.407	0.536	0.651	0.238	0.370	0.507	0.632

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 7. Est. Power in $H_0 : \mu_0 = \mu_x + k\frac{\sigma_x}{\sqrt{n}}$ vs. $H_a : \mu_0 < \mu_x + k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.01, n = 160$				$\alpha = 0.01, n = 320$			
Exponential	0.077	0.161	0.282	0.430	0.081	0.172	0.309	0.476
($\gamma_3 = 2.0$)	0.130	0.248	0.397	0.565	0.119	0.235	0.393	0.568
Weibull	0.086	0.187	0.337	0.520	0.088	0.193	0.348	0.536
($\gamma_3 = 0.63$)	0.102	0.215	0.376	0.561	0.100	0.213	0.375	0.564
SRN	0.095	0.215	0.398	0.609	0.095	0.212	0.392	0.597
($\gamma_3 = 0.83$)	0.076	0.181	0.352	0.561	0.080	0.186	0.358	0.563
Chi-squared	0.080	0.168	0.299	0.458	0.082	0.176	0.317	0.489
($\gamma_3 = 1.63$)	0.123	0.238	0.393	0.563	0.112	0.229	0.385	0.566
Lognormal	0.087	0.170	0.282	0.420	0.088	0.179	0.303	0.455
($\gamma_3 = 2.89$)	0.150	0.269	0.424	0.578	0.135	0.254	0.407	0.575
Lognormal	0.097	0.170	0.261	0.356	0.096	0.179	0.287	0.404
($\gamma_3 = 6.18$)	0.199	0.334	0.478	0.613	0.176	0.302	0.451	0.598

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 8. Est. Power in $H_0 : \mu_0 = \mu_x + k\frac{\sigma_x}{\sqrt{n}}$ vs. $H_a : \mu_0 < \mu_x + k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.05, n = 10$				$\alpha = 0.05, n = 20$			
Exponential	0.183	0.250	0.311	0.358	0.197	0.292	0.387	0.469
($\gamma_3 = 2.0$)	0.367	0.501	0.622	0.726	0.343	0.488	0.626	0.744
Weibull	0.208	0.330	0.462	0.590	0.227	0.376	0.539	0.695
($\gamma_3 = 0.63$)	0.274	0.425	0.585	0.729	0.277	0.439	0.612	0.760
SRN	0.218	0.334	0.453	0.565	0.230	0.369	0.526	0.670
($\gamma_3 = 0.83$)	0.292	0.440	0.592	0.726	0.286	0.446	0.614	0.757
Chi-squared	0.202	0.282	0.357	0.424	0.212	0.321	0.434	0.540
($\gamma_3 = 1.63$)	0.341	0.476	0.607	0.719	0.325	0.473	0.618	0.742
Lognormal	0.237	0.310	0.371	0.412	0.241	0.338	0.429	0.500
($\gamma_3 = 2.89$)	0.410	0.543	0.660	0.754	0.376	0.524	0.657	0.760
Lognormal	0.219	0.257	0.272	0.276	0.231	0.284	0.319	0.330
($\gamma_3 = 6.18$)	0.507	0.629	0.722	0.795	0.462	0.598	0.707	0.789

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 9. Est. Power in $H_0 : \mu_0 = \mu_x + k\frac{\sigma_x}{\sqrt{n}}$ vs. $H_a : \mu_0 < \mu_x + k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.05, n = 40$				$\alpha = 0.05, n = 80$			
Exponential	0.214	0.343	0.473	0.600	0.230	0.373	0.533	0.681
($\gamma_3 = 2.0$)	0.319	0.476	0.628	0.759	0.305	0.467	0.633	0.770
Weibull	0.243	0.404	0.584	0.743	0.247	0.418	0.602	0.765
($\gamma_3 = 0.63$)	0.278	0.448	0.626	0.777	0.271	0.446	0.633	0.787
SRN	0.241	0.394	0.568	0.725	0.265	0.463	0.673	0.842
($\gamma_3 = 0.83$)	0.283	0.448	0.624	0.775	0.237	0.425	0.638	0.817
Chi-squared	0.229	0.361	0.508	0.642	0.234	0.384	0.549	0.701
($\gamma_3 = 1.63$)	0.312	0.468	0.624	0.757	0.294	0.460	0.629	0.768
Lognormal	0.247	0.365	0.486	0.589	0.248	0.385	0.527	0.654
($\gamma_3 = 2.89$)	0.350	0.505	0.650	0.766	0.330	0.490	0.643	0.770
Lognormal	0.251	0.339	0.407	0.452	0.261	0.375	0.480	0.565
($\gamma_3 = 6.18$)	0.423	0.569	0.693	0.785	0.390	0.542	0.676	0.780

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 10. Est. Power in $H_0 : \mu_0 = \mu_x + k\frac{\sigma_x}{\sqrt{n}}$ vs. $H_a : \mu_0 < \mu_x + k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.05, n = 160$				$\alpha = 0.05, n = 320$			
Exponential	0.238	0.391	0.565	0.722	0.244	0.409	0.588	0.746
($\gamma_3 = 2.0$)	0.292	0.457	0.633	0.779	0.283	0.456	0.635	0.783
Weibull	0.251	0.424	0.615	0.777	0.253	0.431	0.622	0.784
($\gamma_3 = 0.63$)	0.270	0.445	0.635	0.795	0.266	0.445	0.636	0.795
SRN	0.266	0.457	0.662	0.830	0.263	0.453	0.655	0.823
($\gamma_3 = 0.83$)	0.244	0.431	0.639	0.813	0.246	0.434	0.637	0.810
Chi-squared	0.241	0.401	0.576	0.731	0.245	0.409	0.595	0.756
($\gamma_3 = 1.63$)	0.287	0.456	0.630	0.777	0.276	0.449	0.634	0.786
Lognormal	0.245	0.398	0.552	0.693	0.248	0.404	0.573	0.723
($\gamma_3 = 2.89$)	0.309	0.478	0.637	0.774	0.296	0.465	0.638	0.781
Lognormal	0.258	0.391	0.520	0.632	0.257	0.398	0.542	0.671
($\gamma_3 = 6.18$)	0.360	0.517	0.661	0.776	0.333	0.496	0.651	0.779

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 11. Est. Power in $H_0 : \mu_0 = \mu_x - k\frac{\sigma_x}{\sqrt{n}}$ vs. $H_a : \mu_0 > \mu_x - k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.01, n = 10$				$\alpha = 0.01, n = 20$			
Exponential	0.041	0.147	0.430	0.834	0.075	0.221	0.492	0.806
($\gamma_3 = 2.0$)	0.015	0.054	0.176	0.437	0.024	0.084	0.240	0.516
Weibull	0.053	0.126	0.256	0.450	0.075	0.178	0.356	0.577
($\gamma_3 = 0.63$)	0.041	0.100	0.212	0.382	0.055	0.136	0.284	0.485
SRN	0.057	0.138	0.285	0.496	0.077	0.191	0.380	0.614
($\gamma_3 = 0.83$)	0.038	0.100	0.215	0.396	0.051	0.132	0.283	0.495
Chi-squared	0.044	0.143	0.368	0.709	0.076	0.213	0.456	0.736
($\gamma_3 = 1.63$)	0.020	0.065	0.181	0.406	0.030	0.097	0.252	0.499
Lognormal	0.052	0.188	0.513	0.878	0.082	0.253	0.559	0.859
($\gamma_3 = 2.89$)	0.016	0.070	0.233	0.537	0.024	0.092	0.279	0.584
Lognormal	0.058	0.360	0.905	0.981	0.091	0.367	0.823	0.990
($\gamma_3 = 6.18$)	0.012	0.105	0.472	0.777	0.016	0.101	0.411	0.796

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 12. Est. Power in $H_0 : \mu_0 = \mu_x - k\frac{\sigma_x}{\sqrt{n}}$ vs. $H_a : \mu_0 > \mu_x - k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.01, n = 40$				$\alpha = 0.01, n = 80$			
Exponential	0.094	0.241	0.488	0.758	0.099	0.241	0.465	0.714
($\gamma_3 = 2.0$)	0.037	0.115	0.287	0.551	0.050	0.140	0.316	0.563
Weibull	0.085	0.202	0.382	0.597	0.091	0.209	0.390	0.602
($\gamma_3 = 0.63$)	0.064	0.161	0.321	0.523	0.073	0.175	0.341	0.549
SRN	0.089	0.212	0.403	0.630	0.094	0.214	0.401	0.621
($\gamma_3 = 0.83$)	0.061	0.155	0.319	0.538	0.069	0.170	0.339	0.554
Chi-squared	0.093	0.234	0.460	0.712	0.102	0.238	0.448	0.686
($\gamma_3 = 1.63$)	0.045	0.128	0.295	0.540	0.058	0.152	0.325	0.561
Lognormal	0.096	0.265	0.536	0.808	0.103	0.260	0.502	0.757
($\gamma_3 = 2.89$)	0.031	0.113	0.301	0.589	0.042	0.131	0.320	0.583
Lognormal	0.103	0.340	0.713	0.957	0.110	0.315	0.626	0.890
($\gamma_3 = 6.18$)	0.021	0.105	0.366	0.739	0.027	0.115	0.344	0.677

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 13. Est. Power in $H_0 : \mu_0 = \mu_x - k\frac{\sigma_x}{\sqrt{n}}$ vs. $H_a : \mu_0 > \mu_x - k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.01, n = 160$				$\alpha = 0.01, n = 320$			
Exponential	0.102	0.235	0.446	0.681	0.101	0.230	0.425	0.648
($\gamma_3 = 2.0$)	0.060	0.157	0.331	0.568	0.070	0.173	0.348	0.566
Weibull	0.093	0.211	0.391	0.599	0.096	0.211	0.386	0.592
($\gamma_3 = 0.63$)	0.080	0.186	0.354	0.561	0.085	0.193	0.360	0.564
SRN	0.095	0.215	0.398	0.609	0.095	0.212	0.392	0.597
($\gamma_3 = 0.83$)	0.076	0.181	0.352	0.561	0.080	0.186	0.358	0.563
Chi-squared	0.102	0.234	0.434	0.657	0.099	0.225	0.416	0.637
($\gamma_3 = 1.63$)	0.066	0.169	0.342	0.566	0.073	0.179	0.350	0.570
Lognormal	0.105	0.251	0.474	0.712	0.104	0.241	0.449	0.677
($\gamma_3 = 2.89$)	0.052	0.149	0.333	0.578	0.062	0.162	0.341	0.573
Lognormal	0.112	0.292	0.564	0.822	0.110	0.273	0.517	0.764
($\gamma_3 = 6.18$)	0.036	0.128	0.336	0.631	0.044	0.142	0.338	0.604

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 14. Est. Power in $H_0 : \mu_0 = \mu_x - k\frac{\sigma_x}{\sqrt{n}}$ vs. $H_a : \mu_0 > \mu_x - k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.05, n = 10$				$\alpha = 0.05, n = 20$			
Exponential	0.242	0.527	0.847	0.989	0.275	0.528	0.800	0.955
($\gamma_3 = 2.0$)	0.138	0.343	0.655	0.916	0.173	0.381	0.664	0.892
Weibull	0.216	0.399	0.625	0.827	0.251	0.447	0.663	0.842
($\gamma_3 = 0.63$)	0.193	0.361	0.573	0.780	0.218	0.400	0.613	0.801
SRN	0.227	0.421	0.657	0.853	0.260	0.466	0.687	0.861
($\gamma_3 = 0.83$)	0.191	0.367	0.589	0.801	0.217	0.405	0.627	0.819
Chi-squared	0.241	0.491	0.782	0.959	0.272	0.508	0.758	0.929
($\gamma_3 = 1.63$)	0.154	0.349	0.624	0.875	0.186	0.387	0.645	0.864
Lognormal	0.268	0.578	0.886	0.993	0.296	0.571	0.841	0.973
($\gamma_3 = 2.89$)	0.151	0.388	0.724	0.940	0.173	0.409	0.711	0.924
Lognormal	0.345	0.824	0.998	0.999	0.345	0.723	0.971	1.000
($\gamma_3 = 6.18$)	0.156	0.547	0.913	0.974	0.165	0.490	0.866	0.983

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 15. Est. Power in $H_0 : \mu_0 = \mu_x - k\frac{\sigma_x}{\sqrt{n}}$ vs. $H_a : \mu_0 > \mu_x - k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.05, n = 40$				$\alpha = 0.05, n = 80$			
Exponential	0.282	0.518	0.761	0.923	0.279	0.500	0.730	0.893
($\gamma_3 = 2.0$)	0.198	0.402	0.660	0.868	0.215	0.415	0.655	0.848
Weibull	0.260	0.452	0.662	0.838	0.263	0.455	0.662	0.830
($\gamma_3 = 0.63$)	0.232	0.416	0.625	0.809	0.241	0.427	0.634	0.811
SRN	0.266	0.467	0.684	0.852	0.265	0.463	0.673	0.842
($\gamma_3 = 0.83$)	0.228	0.420	0.638	0.820	0.237	0.425	0.638	0.817
Chi-squared	0.280	0.501	0.732	0.900	0.281	0.490	0.713	0.877
($\gamma_3 = 1.63$)	0.208	0.407	0.647	0.846	0.226	0.419	0.649	0.836
Lognormal	0.296	0.549	0.794	0.944	0.291	0.521	0.757	0.914
($\gamma_3 = 2.89$)	0.193	0.417	0.688	0.895	0.208	0.419	0.672	0.869
Lognormal	0.332	0.649	0.909	0.994	0.319	0.593	0.848	0.970
($\gamma_3 = 6.18$)	0.176	0.460	0.795	0.969	0.189	0.439	0.738	0.933

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

Table 16. Est. Power in $H_0 : \mu_0 = \mu_x - k\frac{\sigma_x}{\sqrt{n}}$ vs. $H_a : \mu_0 > \mu_x - k\frac{\sigma_x}{\sqrt{n}}$

Distribution (skewness)	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
	$\alpha = 0.05, n = 160$				$\alpha = 0.05, n = 320$			
Exponential	0.274	0.488	0.707	0.869	0.275	0.473	0.685	0.853
($\gamma_3 = 2.0$)	0.226	0.425	0.652	0.835	0.236	0.429	0.644	0.825
Weibull	0.262	0.453	0.658	0.825	0.264	0.452	0.653	0.818
($\gamma_3 = 0.63$)	0.247	0.432	0.638	0.810	0.253	0.438	0.639	0.808
SRN	0.266	0.457	0.662	0.830	0.263	0.453	0.655	0.823
($\gamma_3 = 0.83$)	0.244	0.431	0.639	0.813	0.246	0.434	0.637	0.810
Chi-squared	0.277	0.480	0.693	0.859	0.271	0.470	0.680	0.846
($\gamma_3 = 1.63$)	0.236	0.429	0.645	0.827	0.241	0.432	0.645	0.823
Lognormal	0.286	0.502	0.726	0.887	0.279	0.487	0.704	0.865
($\gamma_3 = 2.89$)	0.220	0.424	0.658	0.849	0.229	0.428	0.650	0.833
Lognormal	0.310	0.555	0.800	0.943	0.296	0.528	0.758	0.914
($\gamma_3 = 6.18$)	0.203	0.429	0.701	0.898	0.213	0.428	0.678	0.871

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Johnson's test – the top number, for Student's test – the bottom number.

For lower-tailed tests the Johnson's procedure had consistently lower Type I error rates than the Student's t-test, for all combinations of sample size and skewness coefficient (Tables 1 and 2, Figure 1 for the Lognormal). Although Johnson's Type I error rates were not below the nominal levels, Student's were even higher. This fact rendered Student's higher power estimates for lower-tailed tests irrelevant, since this is a test which rejects a lot (an anti-conservative test), when the alternative is both false and correct (Tables 5, 6, 7, 8, 9 and 10, Figure 2). Hence, for lower-tailed tests about the mean of positively skewed distributions the Johnson's procedure is still a better choice than the Student's test, yet with pitfalls.

The picture was different, however, for upper-tailed tests. Both the Johnson's and Student's procedures had Type I error rates below the nominal levels for all combinations of sample size and skewness coefficient. Also, the Johnson's rates were slightly higher than the Student's, but this was immaterial since these were below the nominal levels (Tables 3 and 4, Figure 3). Nevertheless, the Johnson's procedure was more powerful than the Student's t-test (Tables 11, 12, 13, 14, 15 and 16, Figure 4). This leads to conclude that for upper-tailed test about the mean of positively skewed distributions the Johnson's procedure is superior to the Student's t-test.

Both the upper- and lower-tailed test findings were consistent with the results in Sutton (1993).

4 Bootstrap-t method vs. Student's and Johnson's tests

As it was shown in lower-tailed test about the mean of a positively skewed distribution, the Johnson's procedure had Type I error rates above the nominal levels. In this situation, it is of interest to study the testing properties (the Type I error and power) of the bootstrap-t method as an alternative to the Johnson procedure for lower-tailed tests.[1] What follows is an evaluation of the bootstrap-t using Monte Carlo for a sample of size 320.

[1] Efron and Tibshirani talk about using bootstrap-t for building confidence intervals only. Applying bootstrap-t as a testing procedure follows from the relationship between the confidence interval and the hypothesis test.

4.1 Description of the Monte Carlo Experiment

It was mentioned in the second section that the bootstrap-t method approximates the distribution of the studentized statistic via the estimated distribution of $\hat{t}_j = (\bar{x}^{*j} - \bar{x})/\hat{se}(\bar{x}^{*j})$. In order to study the Type I error of the bootstrap t-method the following steps were made:

1. A sample of size 320 from a skewed distributions with mean μ_x and standard deviation σ_x is drawn; $t = \sqrt{n}(\bar{x} - \mu_x)/S$ is calculated for this sample.

2. 1000 bootstrap samples are drawn from this sample.

3. The statistic $\hat{t}_j = (\bar{x}^{*j} - \bar{x})/\hat{se}(\bar{x}^{*j})$ is calculated for all 1000 bootstrap samples, where $\hat{se}(\bar{x}^{*j})$ is the standard error of $\bar{x}^*$ in the bootstrap sample j and is estimated as

$$\hat{se}(\bar{x}^{*j}) = \{\sum_{i=1}^{n}(x_i^{*j} - \bar{x}^{*j})^2/n^2\}^{1/2}.$$

   These statistics are ranked to form the EPDF of $t = \sqrt{n}(\bar{x} - \mu_x)/S$ estimated in the original sample.

4. Compare t to the 5^{th} and 1^{st} percentiles of the EPDF mentioned in (3). Reject at $\alpha = 0.05$ level of significance if t is smaller than the 5^{th} percentile. Reject at $\alpha = 0.01$ level of significance if t is smaller than the 1^{st} percentile.

5. Repeat steps one, two, three and four $40,000$ times. At each level of significance separately, the proportion of rejections in $40,000$ replications will be the estimated Type I error of the bootstrap-t method.

The power is estimated as follows.

1. The same sample as in part (1) above is taken ($n = 320$ and same skewed distributions with mean μ_x and standard deviation σ_x);
   $t = \sqrt{n}(\bar{x} - \mu_x - k\sigma_x/\sqrt{n})/S$ is calculated for this sample, where $k = 1$.

2. Same 1000 bootstrap samples are dawn from this sample as in part (2).

3. The statistic $\hat{t}_j = (\bar{x}^{*j} - \bar{x})/\hat{se}(\bar{x}^{*j})$ is calculated for all 1000 bootstrap samples, where $\hat{se}(\bar{x}^{*j})$ is the standard error of $\bar{x}^*$ in the bootstrap sample j and is estimated as

$$\hat{se}(\bar{x}^{*j}) = \{\sum_{i=1}^{n}(x_i^{*j} - \bar{x}^{*j})^2/n^2\}^{1/2}.$$

These statistics are ranked to form the EPDF of $t = \sqrt{n}(\bar{x} - \mu_x - k\sigma_x/\sqrt{n})/S$ estimated in the original sample.

4. Compare t to the 5^{th} and 1^{st} percentiles of the EPDF mentioned in (3). Reject at $\alpha = 0.05$ level of significance if t is smaller than the 5^{th} percentile. Reject at $\alpha = 0.01$ level of significance if t is smaller than the 1^{st} percentile.

5. Repeat steps one, two, three and four $40,000$ times. At each level of significance separately, the proportion of rejections in $40,000$ replications will be the estimated power of the bootstrap-t method.

6. Repeat steps 1 through 5 for $k = 1.5$ and $k = 2$.

The estimated Type I error and power from applying the bootstrap-t testing procedure on samples of size 320 coming from various skewed distributions is given in Tables 17 and 18. These are in comparison with the estimates from the Johnson's and Student's procedures.

Table 17. Est. Type I error in
$H_0 : \mu_0 = \mu_x$ vs. $H_a : \mu_0 < \mu_x$, n=320

Distribution (skewness)	$\alpha = 0.01$	$\alpha = 0.05$	Distribution (skewness)	$\alpha = 0.01$	$\alpha = 0.05$
Exponential	0.011	0.051	Chi-squared	0.011	0.052
($\gamma_3 = 2.0$)	0.010	0.050	($\gamma_3 = 1.63$)	0.010	0.051
	0.017	0.063		0.015	0.060
Weibull	0.011	0.050	Lognormal	0.014	0.056
($\gamma_3 = 0.63$)	0.010	0.050	($\gamma_3 = 2.89$)	0.012	0.055
	0.012	0.053		0.021	0.071
SRN	0.011	0.053	Lognormal	0.018	0.063
($\gamma_3 = 0.83$)	0.010	0.052	($\gamma_3 = 6.18$)	0.015	0.061
	0.013	0.057		0.032	0.089

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Bootstrap-t – top number, Johnson's test– 2^{nd} number, Student's test – 3^{rd} number.

Table 18. Est. Power in $H_0 : \mu_0 = \mu_x + k\frac{\sigma_x}{\sqrt{n}}$ vs. $H_a : \bar{x} < \mu_x + k\frac{\sigma_x}{\sqrt{n}}$, $n = 320$

Distribution	$k = 1$		$k = 1.5$		$k = 2$	
(skewness)	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
Exponential	0.087	0.248	0.183	0.412	0.323	0.592
($\gamma_3 = 2.0$)	0.081	0.244	0.172	0.409	0.309	0.588
	0.119	0.283	0.235	0.456	0.393	0.635
Weibull	0.093	0.255	0.201	0.432	0.357	0.622
($\gamma_3 = 0.63$)	0.088	0.253	0.193	0.431	0.348	0.622
	0.100	0.266	0.213	0.445	0.375	0.636
SRN	0.096	0.259	0.202	0.434	0.359	0.622
($\gamma_3 = 0.83$)	0.090	0.258	0.194	0.433	0.349	0.622
	0.107	0.273	0.221	0.452	0.384	0.640
Chi-squared	0.089	0.247	0.185	0.412	0.330	0.599
($\gamma_3 = 1.63$)	0.082	0.245	0.176	0.409	0.317	0.595
	0.112	0.276	0.229	0.449	0.385	0.634
Lognormal	0.096	0.252	0.191	0.409	0.322	0.577
($\gamma_3 = 2.89$)	0.088	0.248	0.179	0.404	0.303	0.573
	0.135	0.296	0.254	0.465	0.407	0.638
Lognormal	0.110	0.263	0.202	0.409	0.322	0.556
($\gamma_3 = 6.18$)	0.096	0.257	0.179	0.398	0.287	0.542
	0.176	0.333	0.302	0.496	0.451	0.651

NOTE: The entries are the proportions of rejections in 40,000 Monte Carlo trials, for Bootstrap-t – top number, Johnson's test – 2^{nd} number, for Student's test – 3^{rd} number.

In terms of the Type I error estimates, the bootstrap-t method was not an improvement over the Johnson's procedure. Both had almost identical estimated Type I errors (Table 17, Figure 5 for the Lognormal). In terms of the power estimates, the bootstrap-t had slightly higher power estimates than the Johnson test, but not considerably higher (Table 18, Figure 5 for the Lognormal). At least for a sample of size 320, the bootstrap-t was not an improvement over the Johnson's modified t-test for lower-tailed tests about the mean of positively skewed distributions.

5 Conclusion and Further Research

The Monte Carlo replications in this study were consistent with the results in the original paper. It was possible to confirm that for upper-tailed tests about the mean of positively skewed distributions, the Johnson's modified procedure should be preferred over the Student's t-test: it is accurate and more powerful. Even for lower-tailed tests, the Johnson's test is still superior to the Student's test, although not without shortcomings. For lower-tailed tests the Johnson's test can be inaccurate (i.e. have inflated Type I error) as well, but not as inaccurate as the Student's test.

Since the Johnson's modified procedure had unsatisfactory properties in lower-tailed tests, the bootstrap-t method was applied. The latter, however, was not an improvement over the Johnson's procedure. Hence, estimating the distribution of the studentized statistic via bootstrap did not help mitigate the inflated Type I error of lower-tailed tests about the mean of positively skewed distributions.

At least for upper-tailed tests about the mean of positively skewed distributions distributions, the Johnson's procedure should be certainly applied. Lower-tailed tests do not appear to be approached optimally via the Student's test, the Johnson's test or bootstrap-t. Perhaps understanding the nature of skewed distributions better and studying the similarities between these tests may reveal common weaknesses that could lead to a better testing procedure for the mean of skewed distributions.

References

[1] Efron, Bradley, and Robert J. Tibshirani (1993), *Introduction to Bootstrap*, Chapman and Hall/CRC, New York.

[2] Gentle, James E. (2002), *Elements of Computational Statistics*, Springer-Verlag, New York.

[3] Sutton, Clifton D. (1993), Computer-Intensive Methods for Tests About the Mean of an Asymmetrical Distribution, *Journal of American Statistical Association* **88**, 802-810.

Appendix

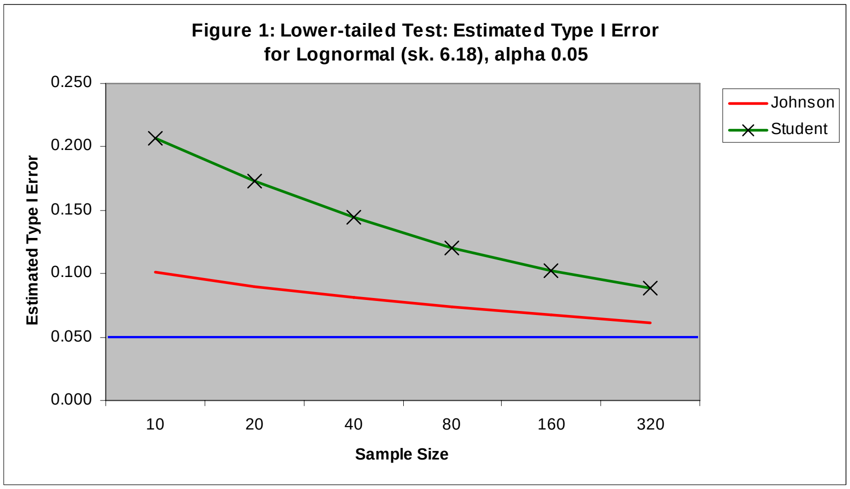

Figure 1: Lower-tailed Test: Estimated Type I Error
for Lognormal (sk. 6.18), alpha 0.05
Estimated Type I Error
0.250
0.200
0.150
0.100
0.050
0.000
10
20
40
80
160
320
Sample Size
Johnson
Student

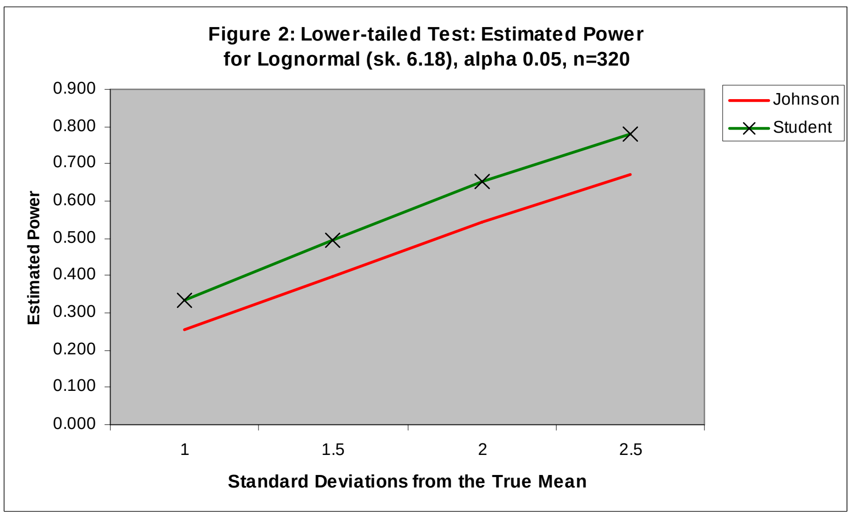

Figure 2: Lower-tailed Test: Estimated Power
for Lognormal (sk. 6.18), alpha 0.05, n=320
Estimated Power
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
1
1.5
2
2.5
Standard Deviations from the True Mean
Johnson
Student

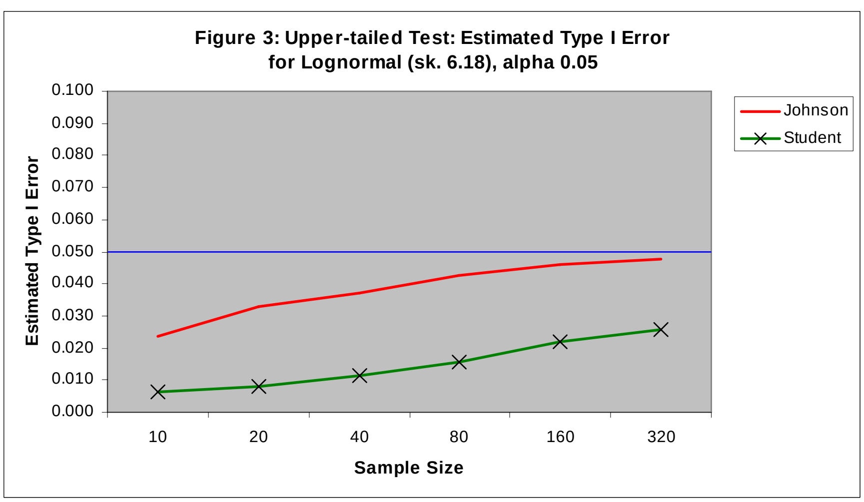

Figure 3: Upper-tailed Test: Estimated Type I Error
for Lognormal (sk. 6.18), alpha 0.05
Estimated Type I Error
0.100
0.090
0.080
0.070
0.060
0.050
0.040
0.030
0.020
0.010
0.000
10
20
40
80
160
320
Sample Size
Johnson
Student

Figure 4: Upper-tailed Test: Estimated Power for Lognormal (sk. 6.18), alpha 0.05, n=320

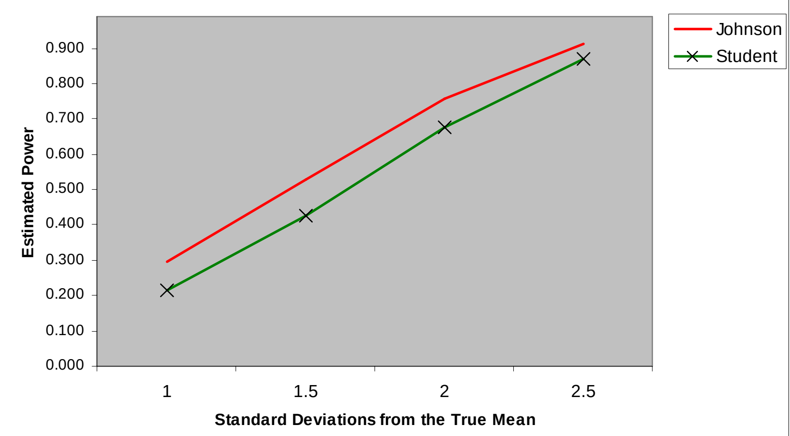

Figure 5: Lower-tailed Test: Estimated Type I Error for Lognormal (sk. 6.18), n=320, alpha 0.01 and 0.05

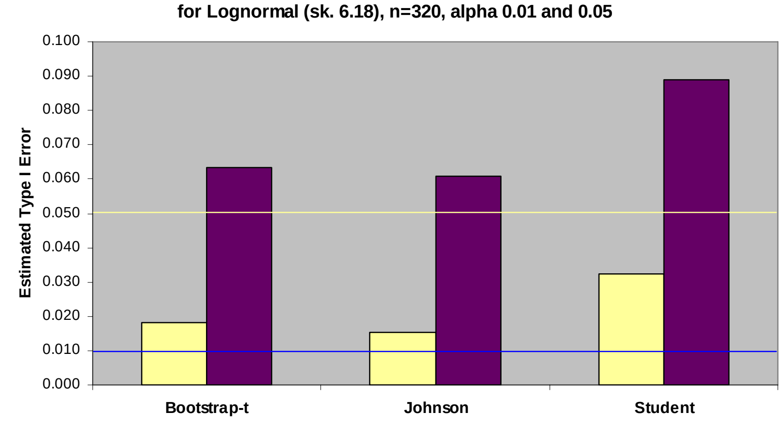

Figure 6: Lower-tailed Test: Estimated Power for Lognormal (sk. 6.18), alpha = 0.05, n=320