



Research Article

Robustness and Asymptotic Theory of Tests of Location under Violation of Normality Assumption

Ibrahim Ali¹, Nicholas P. Dibal², Yusuf A. Mohammed²

¹Department of Mathematics and Computer Science, Borno State University, Njimtilo, Maiduguri, Nigeria,

²Department of Mathematical Sciences, University of Maiduguri, Borno State Nigeria

*Corresponding author's Email: ibrahimali@bosu.edu.ng, doi.org/10.55639/607.262524

ARTICLE INFO:

ABSTRACT

Keywords:

Robustness,
Asymptotic,
Normality violation,
Monte Carlo
Simulation

The distribution of data can be normal or followed a certain distribution. Therefore, data collected requires checking normality assumption before applied to any hypothesis testing. Hence, this work investigates the robustness and asymptotic properties of one sample test statistics when data follows a particular distribution in order to identify the best one to test hypothesis about one population parameter on one sample tests statistics. The one sample test statistics considered are t, z, Sign and Wilcoxon Sign Rank test. A simulation replicated 1000 times was conducted at three sample sizes which are 10, 20 and 40 from uniform, exponential, and gamma distributions. Data analysis revealed that the z-test is the most accurate for all the sample sizes for a data from uniform distribution followed by t-test at small and moderate sample sizes while Wilcoxon sign rank-test is the most robust test to gamma and exponential distributions at all the sample sizes.

Corresponding author: Ibrahim Ali, **Email:** ibrahimali@bosu.edu.ng

Department of Mathematics and Computer Science, Borno State University, Borno State, Nigeria

INTRODUCTION

The distribution of student's t statistic for non-normal (followed another distribution) universe has been the subject of numerous studies from 1930's to the present (Cicchitelli, 1989; Usman and Oyejola, 2013; Akeyede *et al.*, 2014 and Thomas, 2016). Meanwhile, the student's t statistics is not robust to outlier. Robust statistics therefore, addresses the problem of data that deviates from normality assumptions or contain outliers (Zhao *et al.*, 2019). Statistical test or models are based on a set of assumptions; the most important includes the distribution of key variables, for examples, the normal distribution of data, model specification, model linearity or nonlinearity. Some of these assumptions are critical to the estimation process: if they are violated, the estimates become unreliable.

When there are outliers or more broadly when there are deviations from the distributional assumptions, robust statistics should function nearly as well as they should (Hampel, *et al.*, 2011). Assuming that observations are regularly distributed is standard practice in statistics. The inference fails if this assumption is broken because it forms the basis of the entire statistical system. As a result, before doing any statistical data analysis, it is imperative to verify or evaluate the validity of this assumption. The significance of testing assumptions before use is underappreciated by many scholars. One such assumption is seen in the t-test (normality, homoscedasticity and continuity of the data set). If any of these presumptions are broken, the test may be insufficient and a more reliable test may be performed to provide reliable results. The most commonly used statistical methods are correlation, regression and experimental design (Verbunck *et al.*, 2018). But all of them are based on one basic assumption, that the observation follows normal (Gaussian) distribution (Das and Imon, 2016). So, it is assumed that the population from where the sample is drawn is normally distributed. For this reason, the inferential

methods require checking the normality assumption.

The t, z, sign and Wilcoxon sign rank tests are statistical tests used to determine the significance difference between two groups of data or between one groups with a standard value. They are essential tools in statistical analysis and can help researchers determine the significance differences between groups in their data or between one group and a standard value. Hence, these tests were chosen because this research article wants to provide a suitable test of one sample test of location to the selected distributions at small, moderate and large sample sizes.

Many research contributions have been made, such as the analysis of the two-sample t-robustness test and the relationship between the population distribution's shape and the likelihood ratio test robustness (Bowden *et al.*, 2017 and 2019; Hartwig *et al.*, 2017; Qin and Priebe 2017; Choi *et al.*, 2018; Anton *et al.*, 2019 and Zhao *et al.*, 2019 and 2020). The creation of novel techniques has been a major focus of current robustness research. In addition, Slob and Burgess (2020) compared the robustness of mendelian randomization method (MR) and contamination mixture method using type I error criterion. Their analysis recommends investigators to perform variety of robust methods.

Bradley make a comprehensive series of studies of robustness to truly non-normal distributions (1977, 1978, 1980a, 1980b, 1980c). In his research, he examined the performance of the one-sample and independent samples t-tests. Bradley contrasted the results of the t-test when samples were taken from an exponential distribution and results when samples were taken from a bell-shaped (basically normal) distribution. He notes that the amount of alpha, the position of the rejection region, sample size, and the makeup of the population from which the sample was selected are all crucial considerations in one-sample t-test research (Bradley, 1978).

Bradley came to the conclusion that an alpha value of 0.05 is the alpha value that contributes the most resilience. Even though Bradley's set of studies is the most in-depth investigation of the robustness to non-normal distribution, but he hasn't looked at the non-parametric tests on other distributions and also concentrated on large sample size only.

Latest research by Sheng and Hyunseng (2021) expanded the Anderson Rubin test (AR), conditional likelihood ratio test (CLR), and Kleibergen test (K) of the exposure impact in the two sample summary data mendelian randomization (MR) when the effect is violated with the exposure. By fixing the parameters for the simulation, they compared how well these tests statistically performed on the simulated data from normal, binomial, and uniform distribution. Their investigation was 1000 times replicated, and the findings suggest using AR to evaluate faulty instruments. Burges *et al.* (2020); Qi and Chatterjee, 2020; and Slob and Burgess, 2020 using an ordinary least square regression. Tchetgen *et al.* (2017), Verbunck *et al.* (2018), Zhao *et al.* (2018), Burges *et al.* (2020) and Zhao *et al.* (2018). Abdullah and Omar (2017) demonstrate how incorrect model selection influences the significance level in hypothesis

$$\mu = E(x) = \frac{1}{\theta}$$

$$\theta = \frac{1}{\mu} = \frac{1}{0.1} = 10$$

The data were replicated 1000 times for each sample size, and the 1000 iterations computed throughout the investigation were used to count the type I errors that were made.

The following R codes is therefore used for the simulation

$$x_i = \text{rexp}(n, 10), i = 1, \dots, 1000. \text{ which form 1000 iteration}$$

```
>x_i=rexp(10,10)
```

```
>x_i=rexp(20,10)
```

```
>x_i=rexp(40,10)
```

Data were simulated to test for hypothesis ($H_0: \mu = 0.4$) from gamma distribution, the parameters $\alpha = 0.16$ and $\beta = 2.5$ were estimated using method of moment as follows:

$$\text{Mean} = \alpha\beta = 0.4 \tag{2}$$

$$\text{Variance} = \alpha\beta^2 = 1 \tag{3}$$

Solving the two equations simultaneously, we have $\alpha = 0.16$ and $\beta = 2.5$ that were used for the simulation from the gamma distribution.

testing and come to the conclusion that this is true even in the asymptotic scenario.

This research article provides a suitable test among t, z, sign and Wilcoxon sign rank tests to the selected distributions which are exponential, gamma and uniform distributions at small, moderate and large sample sizes, which show when the various selection procedures are robust as the sample sizes increase.

METHOD AND MATERIAL

Source of Data

The data used in this article for analysis were obtained by simulation from three distributions which are exponential, gamma, and uniform distributions. The mean of each distribution serves as the foundation for all speculated values. To estimate the parameter fixed for each distribution's simulation, the mean (hypothesized value) is employed. Simulated data with sample sizes of 10, 20, and 40, which represent small, moderate, and large sample sizes, respectively, were used to investigate the relative performance of the three tests.

Simulation Procedure

Data would be simulated to test for hypothesis ($H_0: \mu = 0.1$) from exponential distribution, the parameter = 10 which is estimated using method of moment is as follows:

(1)

The following R codes is therefore used for the simulation

```

xi = rgamma(n, 0.16, 2.5), i = 1, ..., 1000. which form 1000 iteration
>xi=rgamma(10,0.16,2.5)
>xi=rgamma(20,0.16,2.5)
>xi=rgamma(40,0.16,2.5)
    
```

Data would be simulated to test for hypothesis ($H_0: \mu = 0$) from uniform distribution, the parameters $\alpha = -2$ and $\beta = 2$ were estimated using method of moment as follows:

$$mean = E(x) = \frac{\alpha + \beta}{2} = 0 \tag{4}$$

$$variance = V(x) = \frac{(\alpha - \beta)^2}{12} = 1 \tag{5}$$

Solving the two equations simultaneously the parameter values would be $\alpha = -2$ and $\beta = 2$

The following R codes are therefore used for the simulation

```

xi = runif(n, -2, 2), i = 1, ..., 1000. which form 1000 iteration
>xi=runif(10,-2,2)
>xi=runif(20,-2,2)
>xi=runif(40,-2,2)
    
```

The parameter values are obtained by initial guess in most researches, but in this article we have tried to obtain the parameter values using method of moment to at least reduced biasedness. A more detailed r code of the analysis can be found in appendix.

RESULTS

Using t, z, Sign, and Wilcoxon sign rank tests, the analysis of simulated data from an exponential distribution at sample sizes of 10, 20, and 40

representing small, moderate, and large sample sizes, respectively was done to ascertain their relative performance. The table below demonstrates this:

Table 1: Type I Error of the Four Tests under Exponential Distribution

Sample Size	t-test	z-test	Sign-test	Wilcoxon Sign Rank-test
10	1000	1000	1000	0
20	1000	1000	1000	0
40	1000	1000	1000	0
Total	3000	3000	3000	0

Table 1 lists the total number of type I errors made by the four tests that were used to examine the data obtained from an exponential distribution using various sample sizes. The significance value is set as 5% (alpha level), while the p-value is generated automatically by the r package during the analysis and if p-value is less than or equal to the alpha value, a type I error is

committed (i.e., wrong rejection of the null hypothesis). The Wilcoxon sign rank-test performs the best across all sample sizes, recording zero type I errors out of a total of 1000 replications, in contrast to other tests that incorrectly rejected the true null hypothesis 1,000 times out of a total of 1000 replications. The Wilcoxon sign rank-test often performs the best

for the total sample sizes. The graphs below gave

the clear picture of the above table.

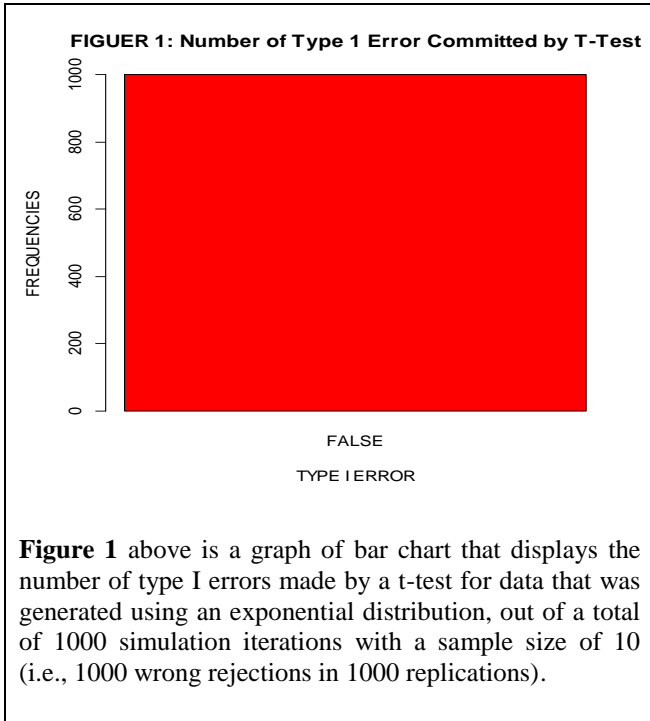


Figure 1 above is a graph of bar chart that displays the number of type I errors made by a t-test for data that was generated using an exponential distribution, out of a total of 1000 simulation iterations with a sample size of 10 (i.e., 1000 wrong rejections in 1000 replications).

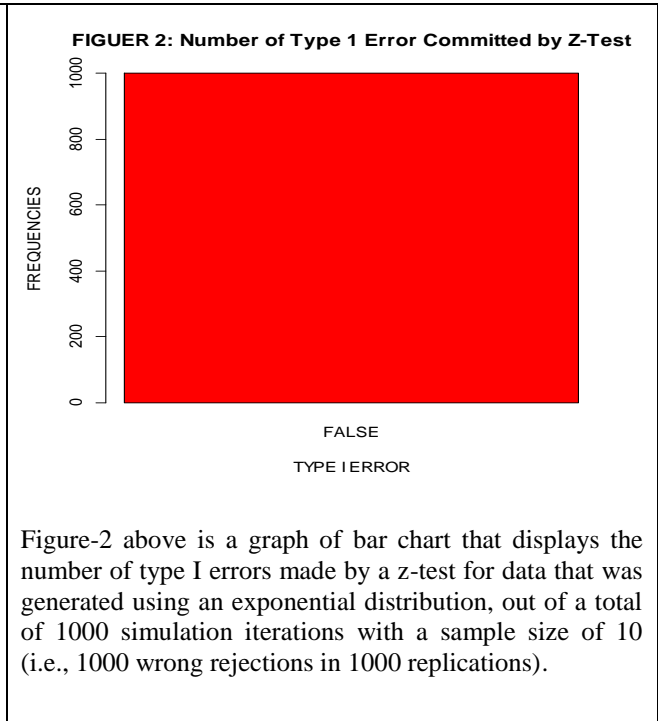


Figure-2 above is a graph of bar chart that displays the number of type I errors made by a z-test for data that was generated using an exponential distribution, out of a total of 1000 simulation iterations with a sample size of 10 (i.e., 1000 wrong rejections in 1000 replications).

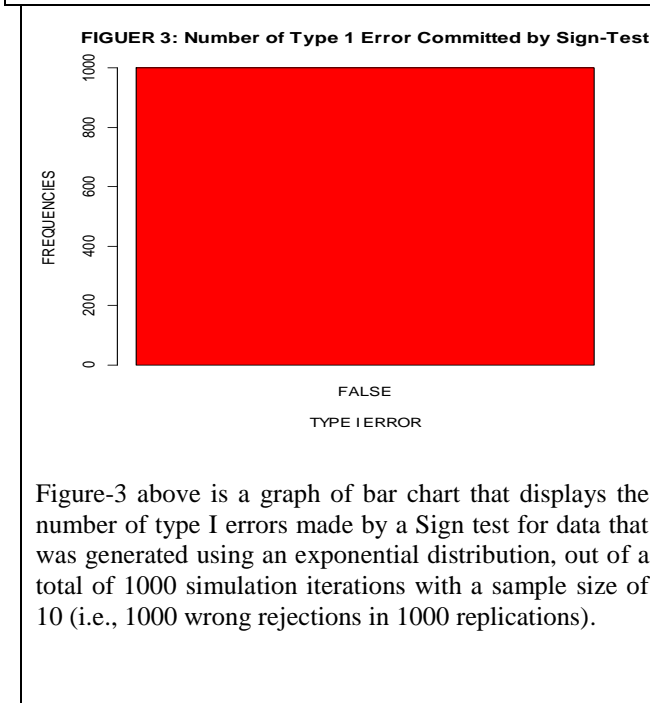


Figure-3 above is a graph of bar chart that displays the number of type I errors made by a Sign test for data that was generated using an exponential distribution, out of a total of 1000 simulation iterations with a sample size of 10 (i.e., 1000 wrong rejections in 1000 replications).

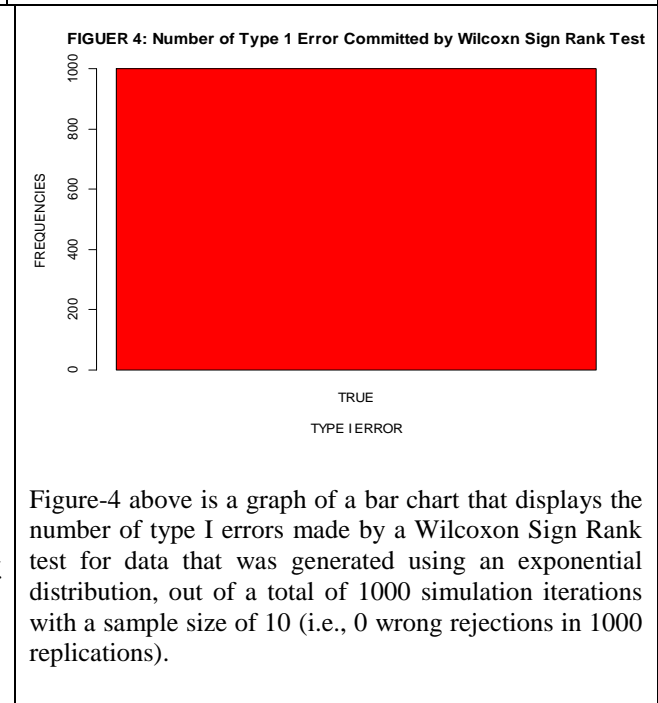
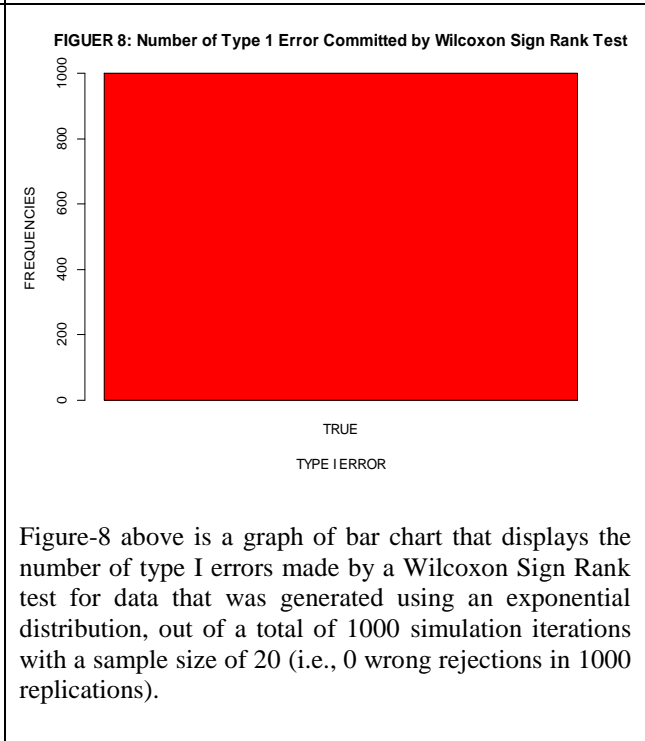
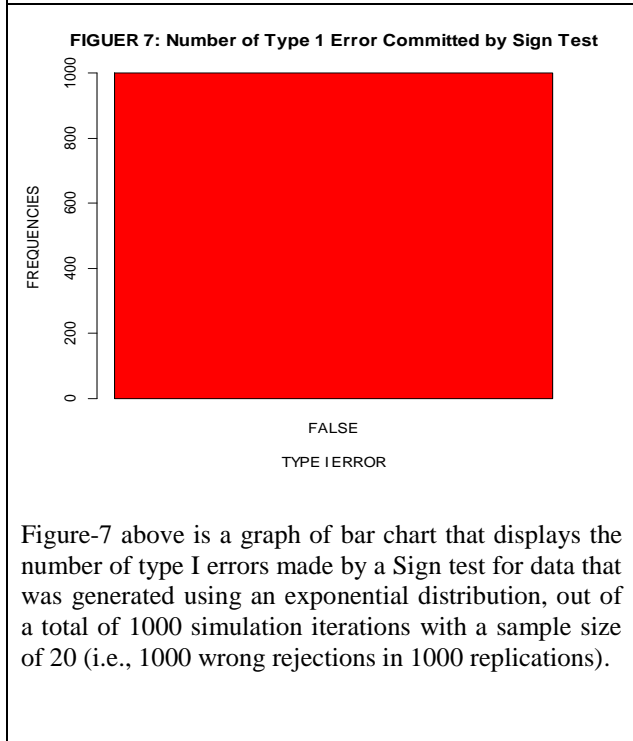
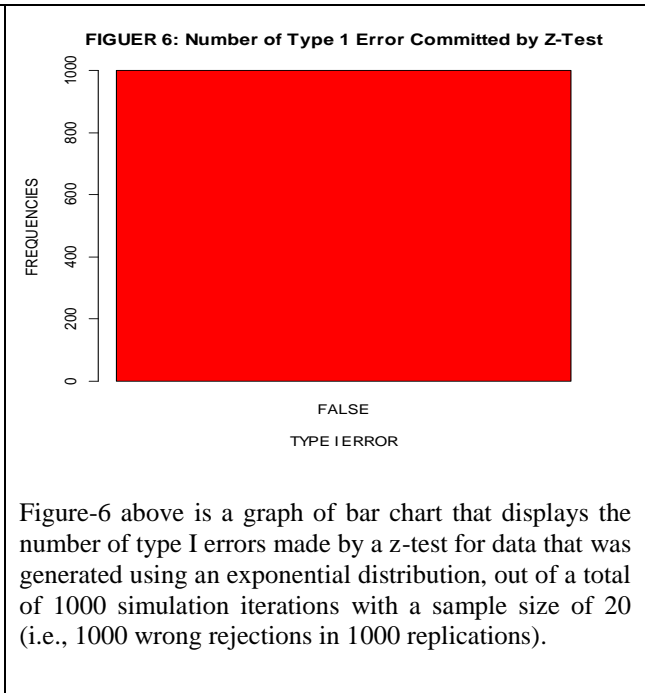
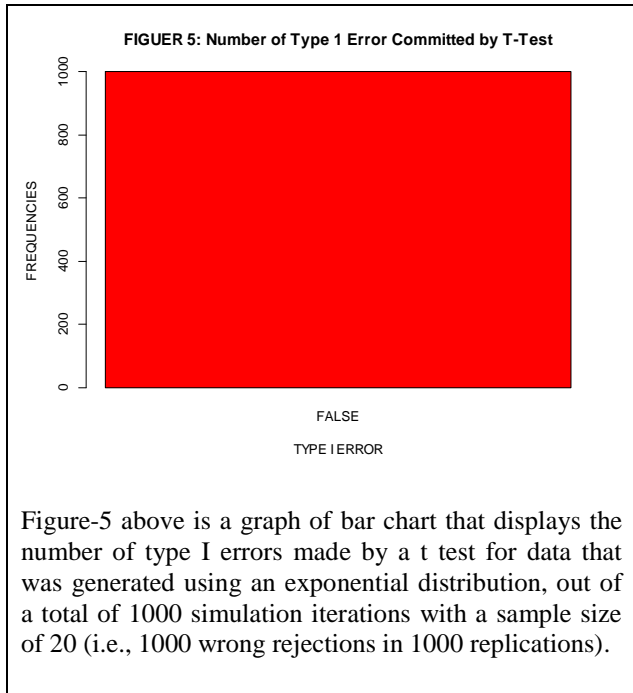


Figure-4 above is a graph of a bar chart that displays the number of type I errors made by a Wilcoxon Sign Rank test for data that was generated using an exponential distribution, out of a total of 1000 simulation iterations with a sample size of 10 (i.e., 0 wrong rejections in 1000 replications).



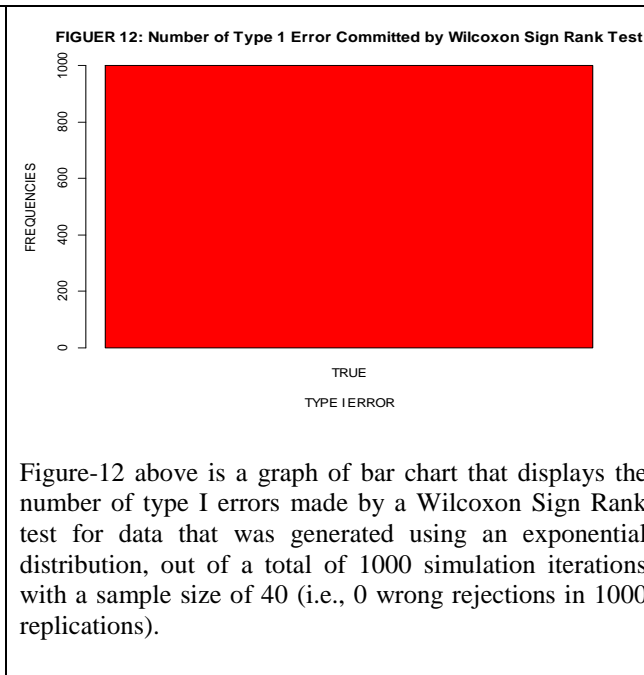
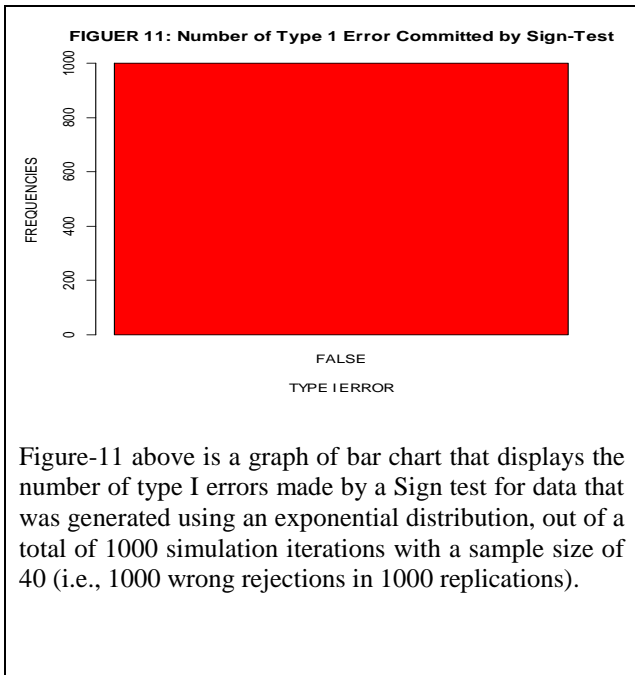
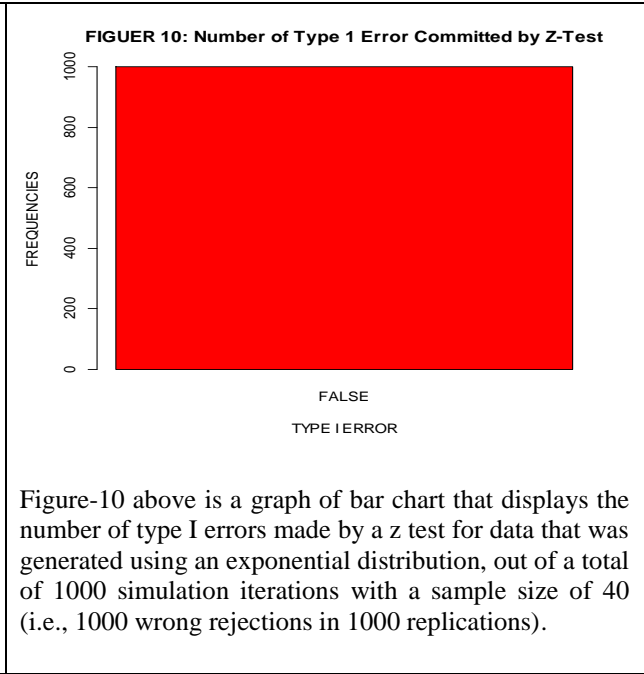
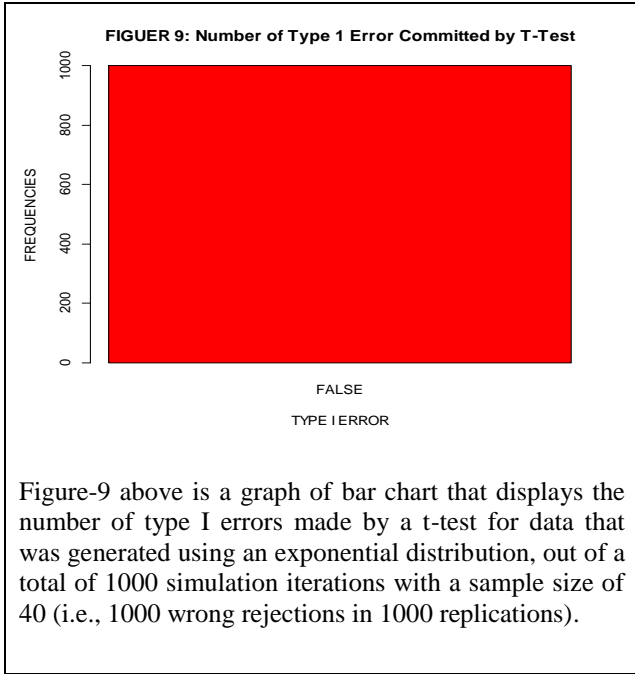
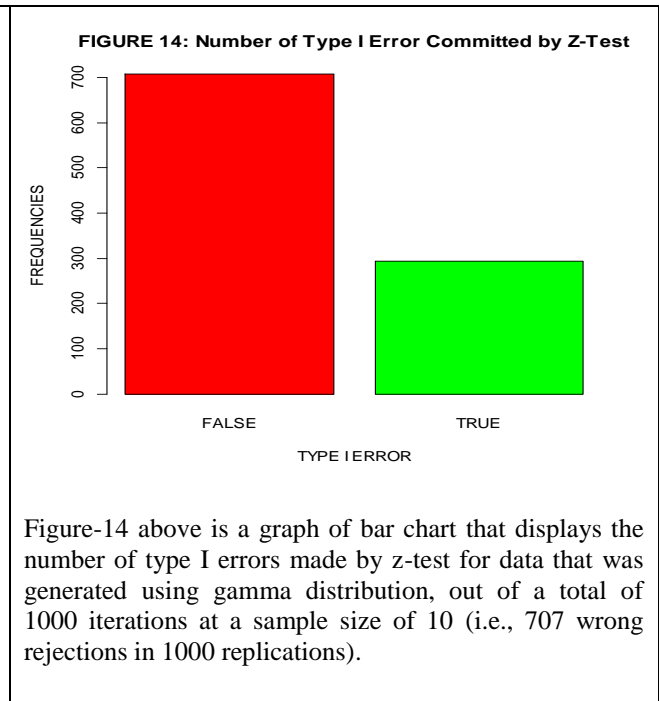
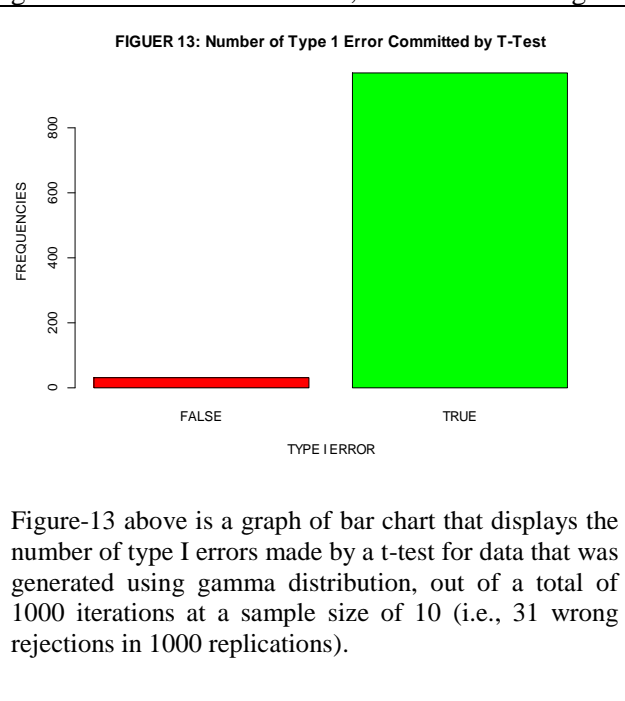


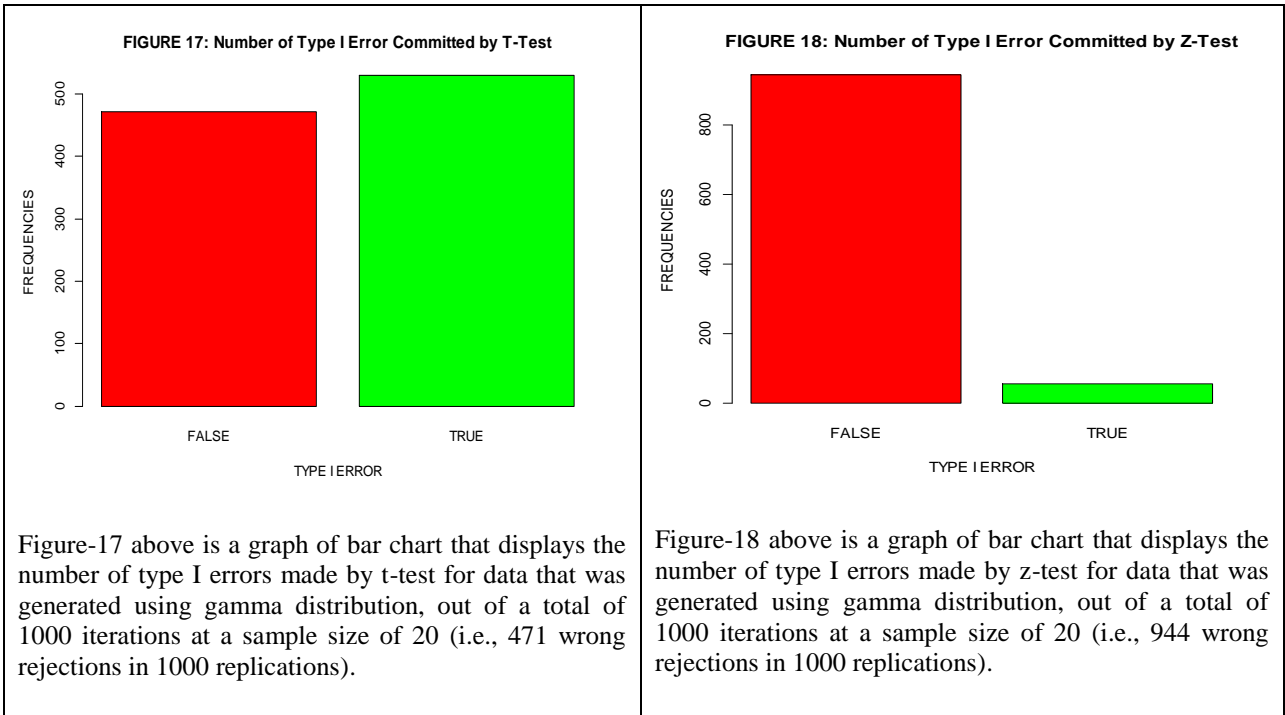
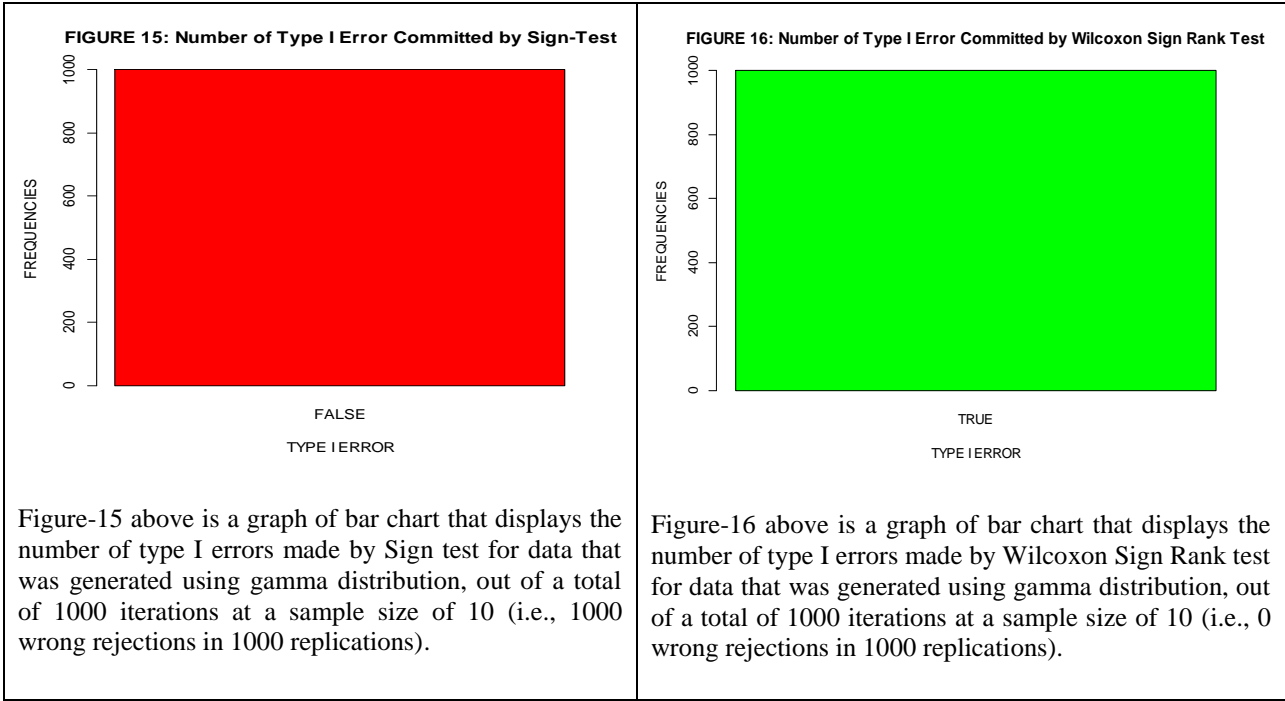
Table 2: Type I Error of the Four Tests under Gamma Distribution

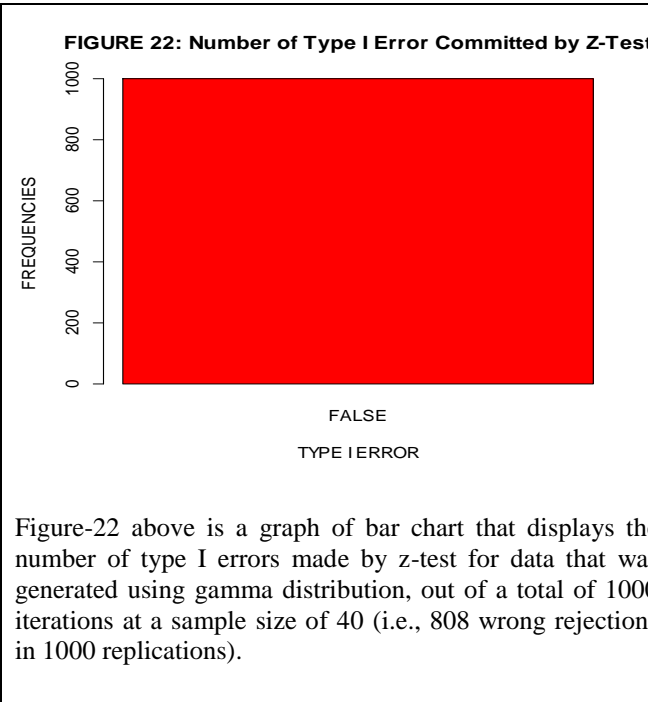
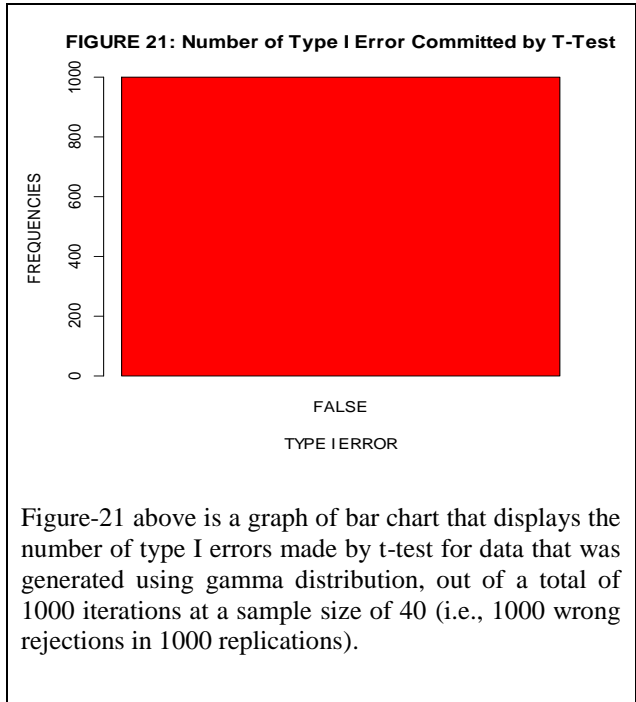
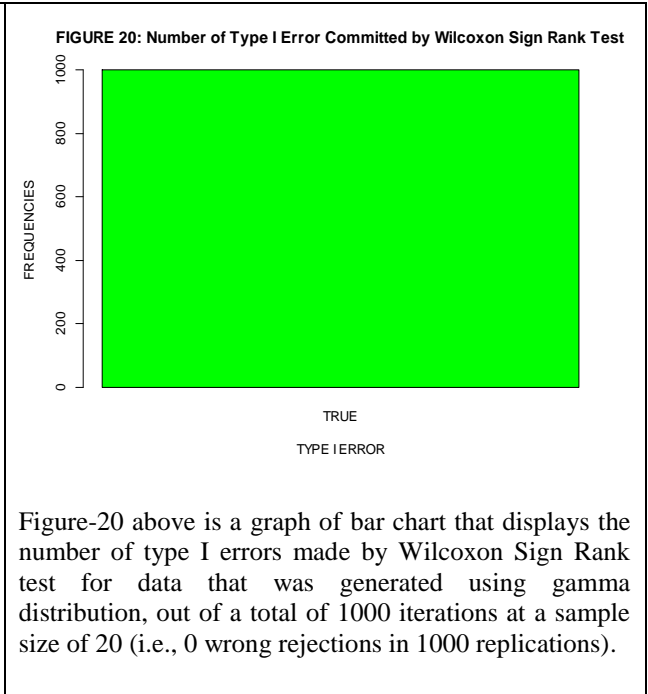
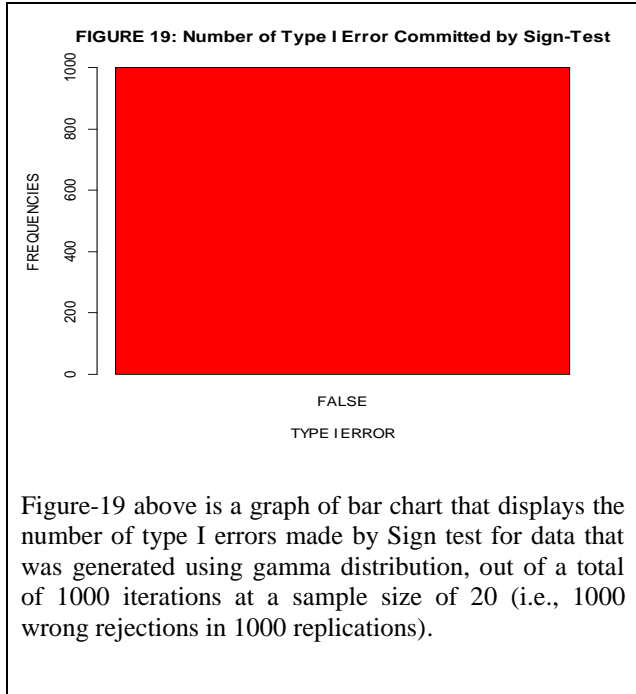
Sample Size	t-test	z-test	Sign-test	Wilcoxon Sign Rank-test
10	31	707	1000	0
20	471	944	1000	0
40	1000	808	1000	0
Total	1502	2459	3000	0

Table-2 list the total number of type I errors made by the four tests that were performed to examine the data obtained from the gamma distribution using various sample sizes. The Wilcoxon Sign Rank-test, which does not make type I error at all sample sizes, is the best. This indicates that the Wilcoxon sign rank test has the ability to identify a true null hypothesis for data from a gamma distribution at small, moderate and large

sample sizes that make it better than its counterpart. The t-test comes in second with 31 rejections out of 1000 iterations at small sample sizes and 471 rejections out of 1000 iterations at intermediate sample sizes. For the combined sample sizes, the Wilcoxon Sign Rank-test performs the best overall. This can be shown in the following graphs below:







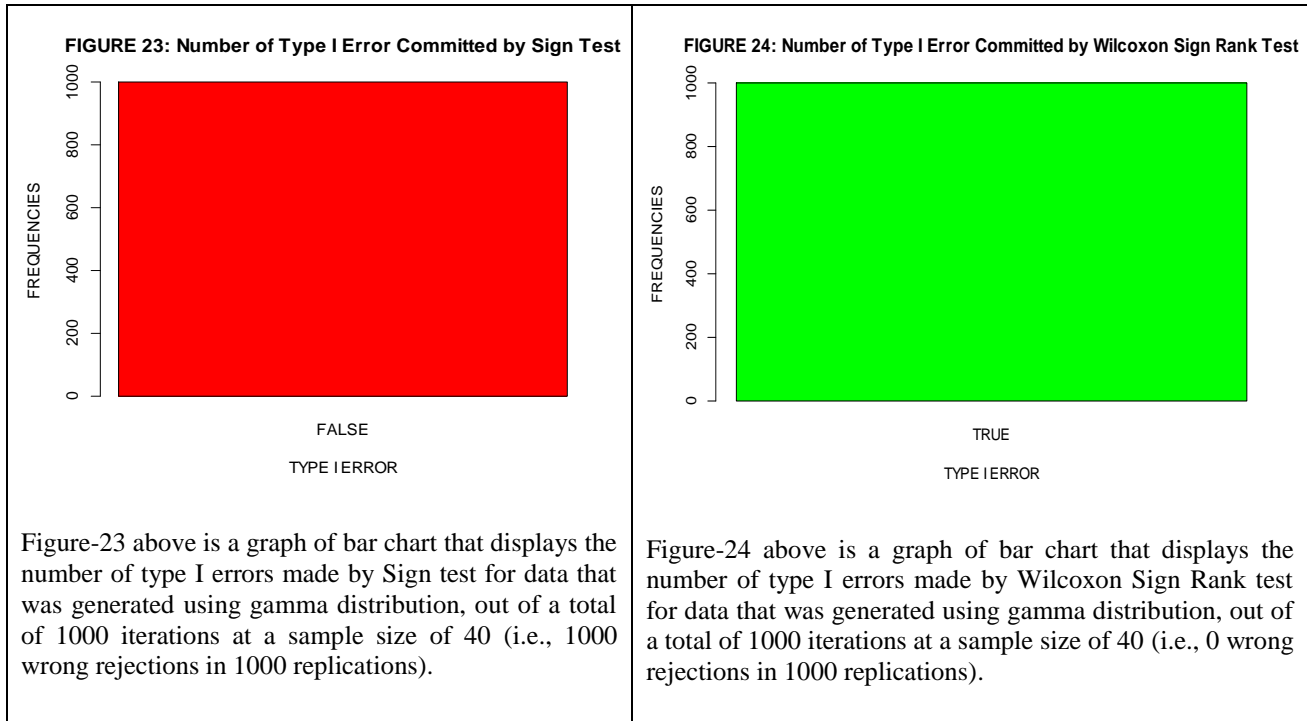
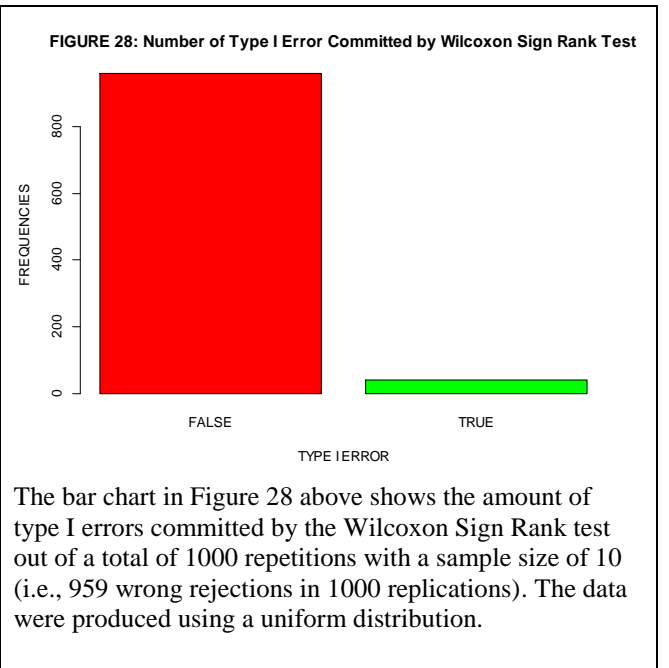
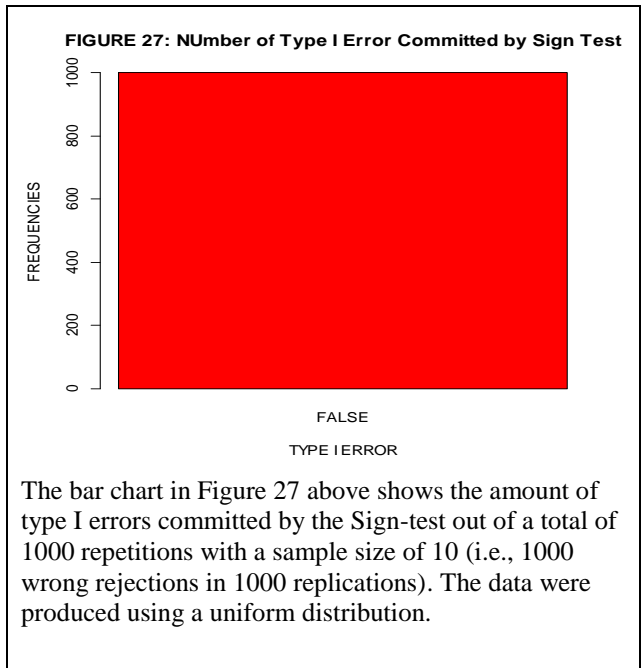
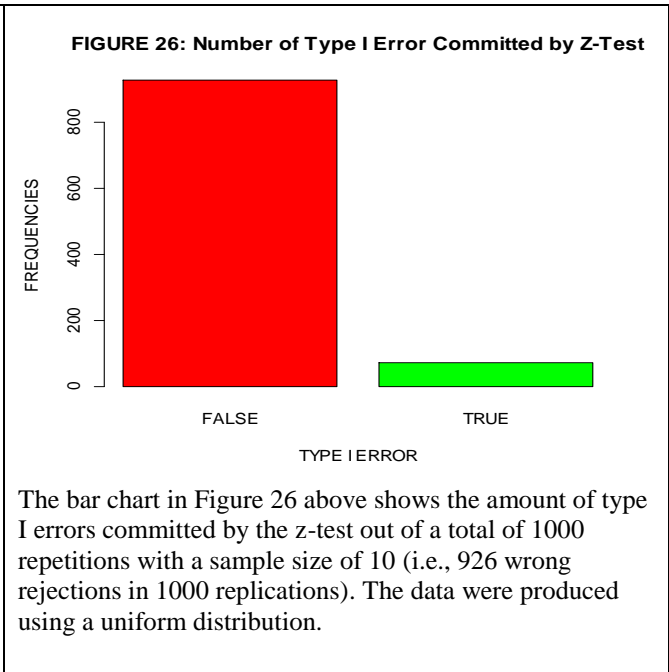
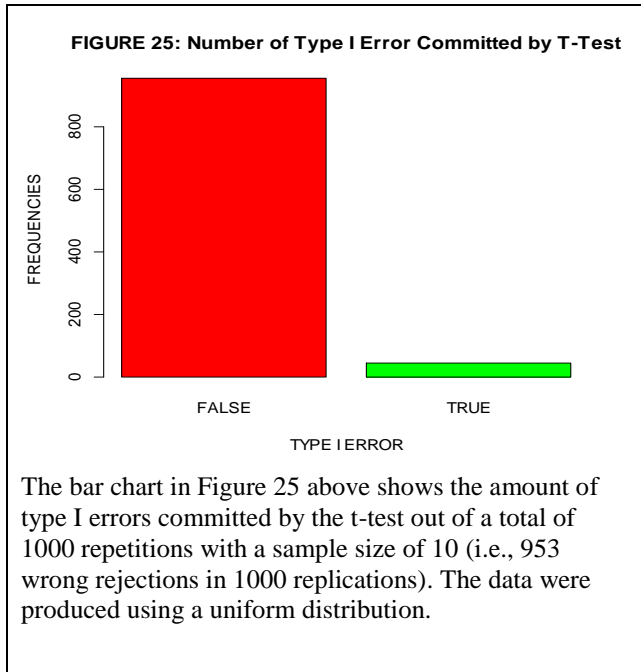


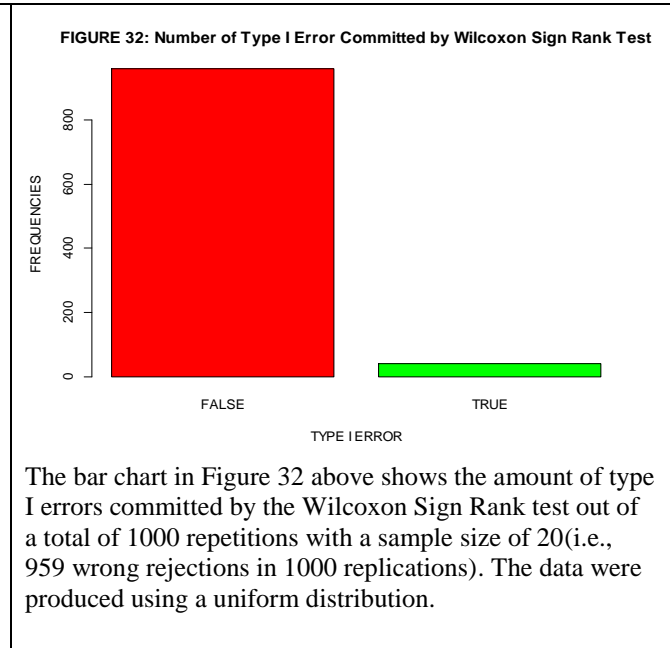
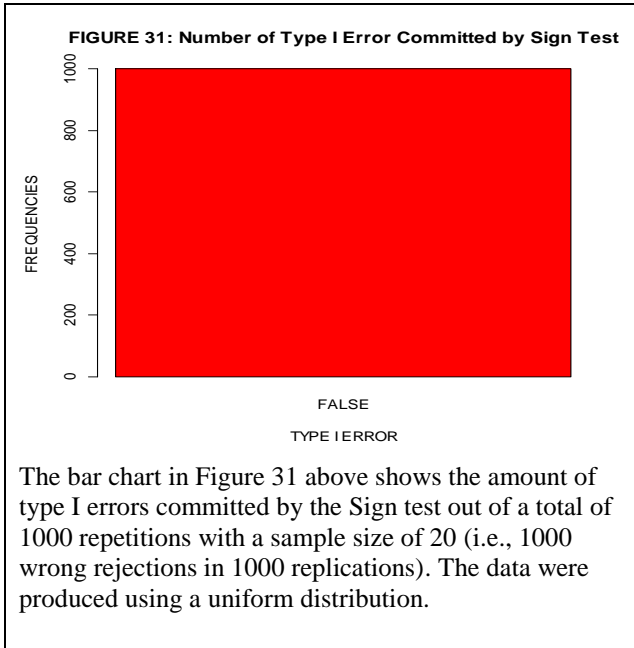
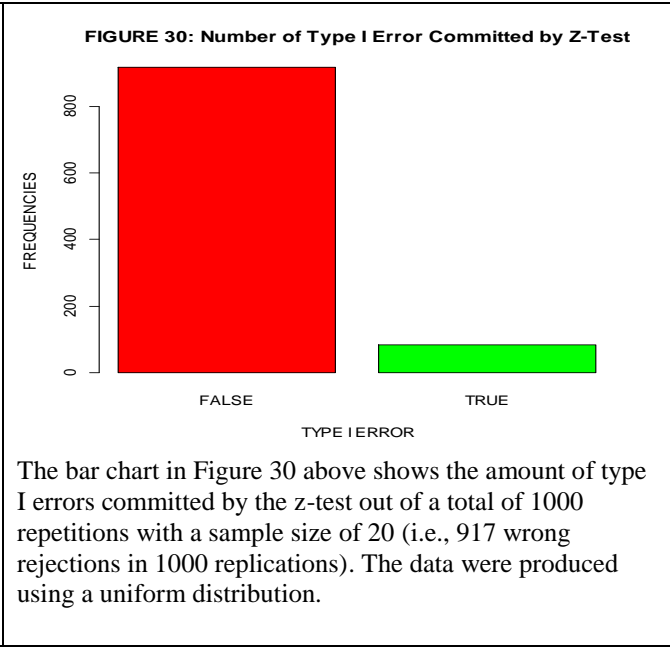
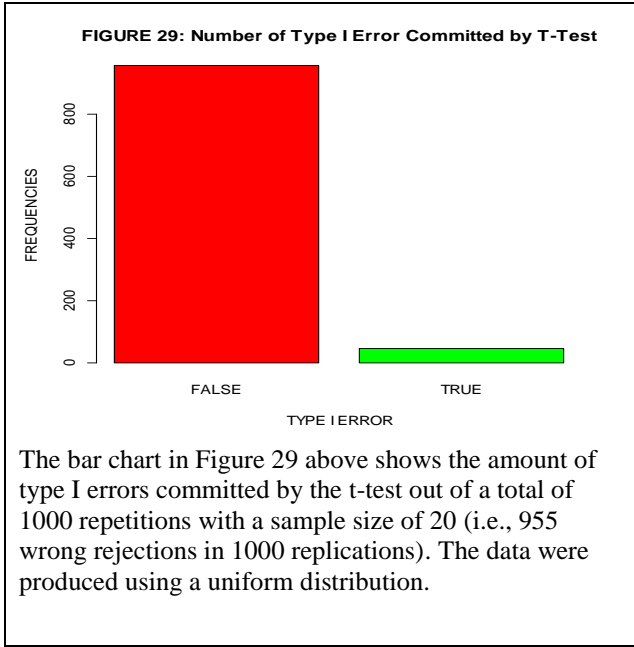
Table 3: Type I Error of the Four Tests under Uniform Distribution

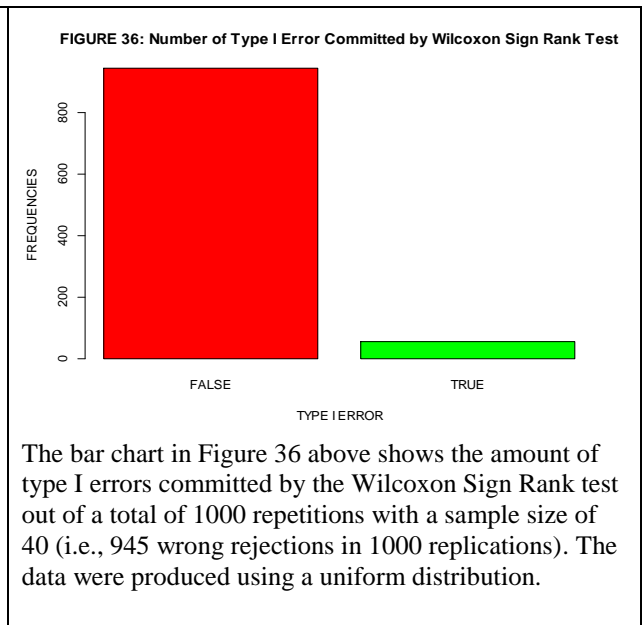
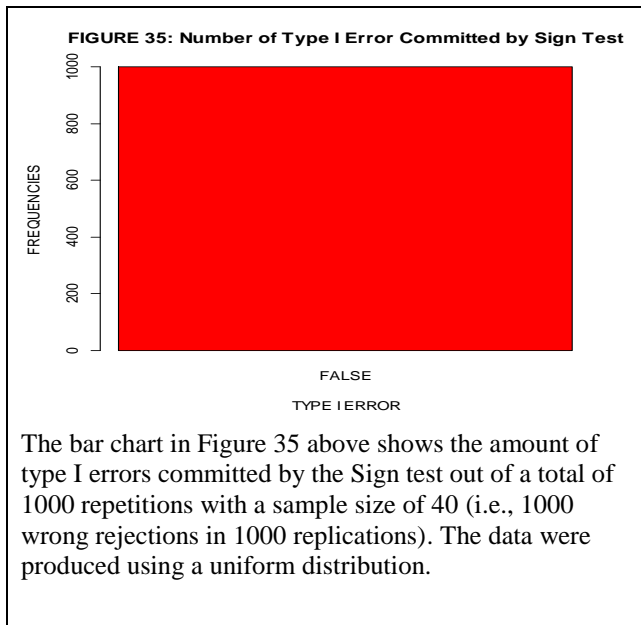
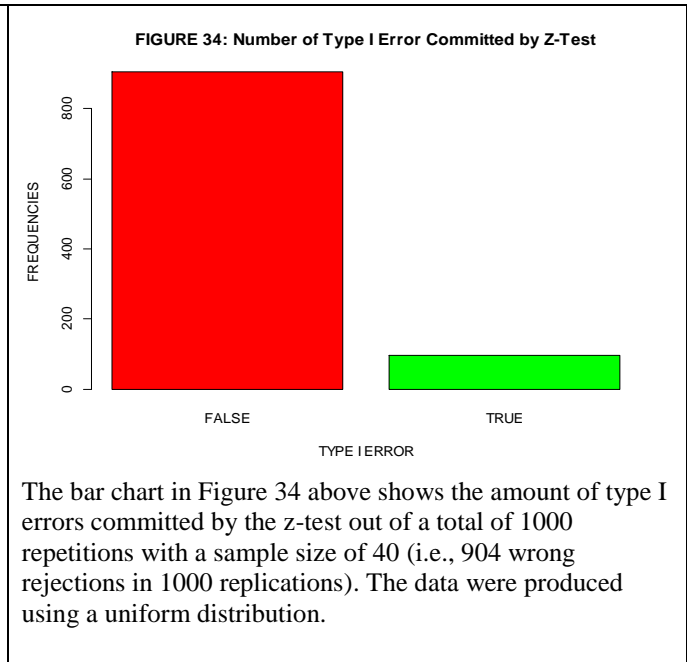
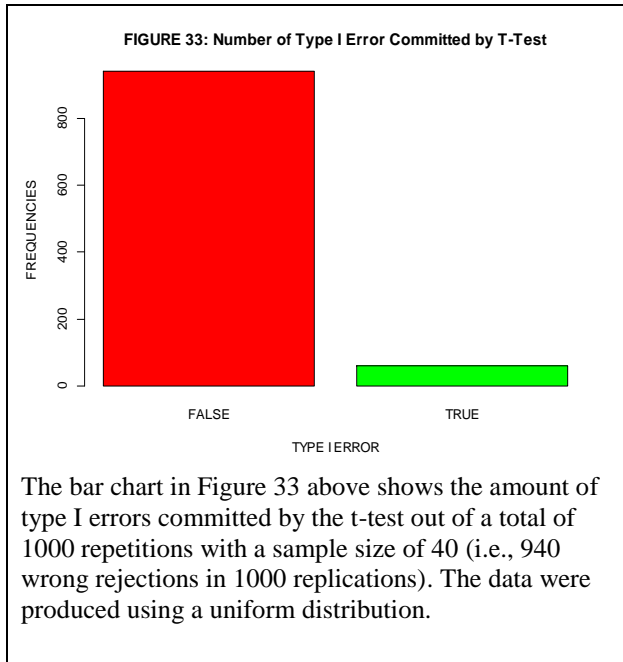
Sample Size	t-test	z-test	Sign-test	Wilcoxon Sign Rank-test
10	953	926	1000	959
20	955	917	1000	959
40	940	904	1000	945
Total	2848	2747	3000	2863

Table-3 Specify how many type I errors were made by the four tests that were used to examine the data obtained from a uniform distribution using various sample sizes. Z-test performs the best across all sample sizes, with 926, 917, and

904 rejections per 1000 iterations, respectively. T-test is next, with 953, 955, and 940 rejections per 1000 iterations, respectively. Z-tests work well for combined sample sizes in general. This is shown in the diagrams below:







SUMMARY OF FINDINGS

This study concentrated on t, z, Sign, and Wilcoxon sign rank tests when data followed another distribution as sample size increased asymptotically. The exponential, gamma, and uniform distributions (these distributions were

chosen because they are commonly right skewed distributions, and most of the real life data are positively skewed if non normal) were used to simulate data at three sample sizes (10, 20 and 40), which correspond to small, moderate, and large sample sizes, respectively. The table and the graphs demonstrated that for data from gamma

and exponential distributions, the Wilcoxon sign rank test works better at all sample sizes at 5% level of significance (i.e., it has not rejected a true null hypothesis during the analysis), in the 1000 iterations and the 1000 iterations is set in the R code algorithm. In other words, a type I error is made if p-value is less than or equal to the alpha level, and t, z and sign tests wrongly rejected the true null hypothesis while in fact it is true more than the Wilcoxon sign rank test. Based on the analysis, Wilcoxon sign rank test can be applied when a data collected followed exponential and gamma distribution at small, moderate and large sample sizes. Similarly, the z-test is the most reliable test for data from uniform distributions at all sample sizes with a lowest type I error committed compared to t, sign and Wilcoxon sign rank tests.

CONCLUSION

The investigation showed that the Wilcoxon sign rank test is the best since it has the lowest type I error for data from gamma and exponential at all sample sizes as shown in the tables as well as the bar charts (i.e., its ability of identifying the true null hypothesis than the rest of the tests). Which shows that t, z and sign tests are not robust to a data from exponential and gamma distributions for their inability to accept a true hypothesized value (i.e., their inability to identify a true null hypothesis). And if a statistical test cannot accept a true null hypothesis in a particular analysis, then by implication it cannot be reliable. Therefore, the Wilcoxon sign rank test can be

useful to researchers that performed experiment on one sample data especially those in medicine (in testifying their claims o the effectiveness of their drugs when produced) or those in manufacturing process (e.g., light bulbs, fans, etc., to verify the claims of the life span of their products) when it follows gamma and exponential at small, moderate and large sample sizes. The experiment also showed that, for all sample sizes, the z-test is the most reliable test for data from uniform distribution. Based on the results of the experiment and the findings of this study particularly the bar charts which shows the clear picture of the analysis, it is obvious that a robust test for a data from exponential, gamma and uniform distributions at small, moderate and large sample sizes were obtained. Therefore, it is recommended to apply the Wilcoxon sign rank test for data from the gamma and exponential distributions at small, moderate and large sample sizes, whereas the z-test is advised for data from the uniform distribution at small, moderate and large sample sizes. This research article is limited on the robustness of one sample test of location to a data from three distributions only, and 5% significance level and type I error are used in assessing their performance. Finally, a further research can be conducted at 1% or 10% and using a type II error of a test as a criterion for assessment to see the behaviors of the selected tests.

REFERENCES

- Abdullah, Yalcinkaya and Omar, Altindag (2017). Investigation for the robustness of significance level when the normality assumption in hypothesis test is violated. *Gazi university journal of science*. 30(2):247-259
- Akeyede, I., Usman, M. and Chiawa M. A (2014). On Consistency and Limitation of Paired t-test, and Wilcoxon Signed Rank test. *IOSR Journal of Mathematics, IOSR Journal International Organisation of Scientific Research*. 10(1): 1-6.
- Anton, A., Yichen, Q., and Carey, E. P. (2019). Robust Hypothesis Testing for Location Parameters using Lq-likelihood Ratio-Type Test in Python. Arxiv: 1911.11922 [Stat.Me].

- Bowden, j., Del Greco M, F., Minelli, C., Davey Smith, G., Shhehan N. and Thompson, J. (2017). A Framework for the Investigation of Pleiotropy in Two Sample Summary Data Mendelian Randomization. *Statistics in Medicine*. 36(11): 1783-1802.
- Bowden, j., Del Greco M, F., Minelli, C., Zhao, Q., Lawlor, D. A., Sheehan, N. A., Thompson, J. and Davey Smith, G. (2019). Improving the Accuracy of Two Sample Summary Data Mendelian Randomization: Moving Beyond the NOME Assumption. *International Journal of Epidemiology*. 48(3): 728-742.
- Burgess, S., Foley, C. N., Allara, E., Staley, J. R. and Howson, J. M. (2020). A robust and efficient Method for Mendelian Randomization with Hundreds of Genetic Variants. *Nature Communications*, 11(376).
- Burgess, Stephen and Eric Slop, A. W. (2020). A Comparison of Robust Mendelian Randomization Methods Using Summary Data. *International Genetic Epidemiology Society*. DOI: 10.1002/gepi. 2295.
- Bradley, J.V. (1977). A common situation conducive to bizarre distribution shapes. *The American Statistician*, 31, 147-150.
- Bradley, J.V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.
- Bradley, J.V. (1980a). Nonrobustness in one-sample Z and t tests: A large-scale sampling study. *Bulletin of the Psychonomic Society*, 15(1), 29-32.
- Bradley, J.V. (1980b). Nonrobustness in Z, t, and F tests at large sample sizes. *Bulletin of the Psychonomic Society*, 16, 333-336.
- Bradley, J.V. (1980c). Nonrobustness in classical tests on means and variances: A largescale sampling study. *Bulletin of the Psychonomic Society*, 15, 275-278.
- Cicchitelli, G. (1989), *On The Robustness of The One Sample t Test*. Gordon and Breach Science publishers, Inc. Printed in Great Britain
- Choi, J., Gu, J. and Shen, S. (2018). Weak Instruments Robust Inference for Two Sample Instrumental Variables Regression. *Journal of Applied Econometrics*. 33(1): 109-125.
- Hampel, F. R., Rousseeuw, P. and Stahel, H. (2011). Robust Statistics: The Approach Based on Influence Functions. *Wiley series in probability and statistics*. 9781118150689.
- Hartwig, F. P., Davey Smith, G. and Bowden, J. (2017). Robust Inference in Summary Data Mendelian Randomization via the Zero Modal Pleiotropy Assumption. *International Journal of Epidemiology*.
- Keya, Rani Das and A. H. M. Rahmatullah, Imon (2016). A Brief Review of Tests for Normality. *American Journal of Theoretical and Applied Statistics*. 5, No. 1, 5-12. doi:10.11648/j.ajtas.20160501.12
- Qi, G. and Chatterjee, N. (2020). Mendelian Randomization Analysis Using Mixture Models for Robust And Efficient Estimation of Causal Effects. *Nature commincations*. 10(1941).
- Qin, Y. and Priebe, C. E. (2017). Robust Hypothesis Testing via Lq-Likelihood. *Statistica Sinica*. 27, 1793-1813.
- Tchetgen, E. J., Sun, B. and Walter, S. (2017). The Genius Approach to Robust Mendelian Randomization Influence. Arxiv, 1709.07779 [Stat.Me].
- Thomas, Plümper and Eric Neumayer (2016). Department of Socioeconomics, Vienna University of Economics and Business, Department of Geography & Environment, London School of Economics and Political Science (LSE). *Robustness Tests and Statistical Inference*, L.S.E Research Online.
- Usman, M. and Oyejola, B. A. (2013). Models for Count Data in the Presence of Outliers in Presence of Excess zeros. *Mathematical theory and modeling*. (7): 37-42.
- Verbunck, M., Chen, C. Y., Neale, B. and Do, R. (2018). Detecrion of Widespread Horizontal Pleiotropy in Causal Relationships Inferred from Mendelian Randomizaytion Between Complex Traits and Diseases. *Nature Genetics*. 50(5), 693-698.

- Sheng, Wang and Hyunseung, Kang (2021). Weak Instrument Robust Test in Two Sample Summary Data Mendelian Randomization. *A Journal of the International Biometric Society*. Wiley, 1909.0650v3
- Zhao, Q., Wang, J., Bowden, J. and Small, D. S. (2018). Statistical Inference in Two Sample Summary Data Mendelian Randomization using Robust Adjusted Profile Score. Arxiv, 1801.09652 [Stat.Ap].
- Zhao, Q., Wang, J., Bowden, J. and Small, D. S. (2019). Two Sample Instrumental Variable Analysis using Heterogeneous Samples. *Statistical Science*. 34(2): 317-333.
- Zhao, Q., Wang, J., Hamani, G., Bowden, J. and Small, D. S. (2020). Statistical Inference in Two Sample Summary Data Mendelian Randomization using Robust Adjusted Profile Score. *Annals of statistics*. 48(3): 1742-1769.