Proceedings of The National Conference
On Undergraduate Research (NCUR) 2016
University of North Carolina-Asheville
Asheville, North Carolina
April 7-9, 2016

Comparison of Parametric and Nonparametric Tests for Differences in Distribution

Davis Jett, Jessica Speer
Department of Mathematics & Statistics
The University of North Carolina at Asheville
Asheville, NC 28804

Faculty Advisor: Dr. Jimin Lee

Abstract

When the population is approximately normally distributed, the two-sample t-test is appropriate to conduct a hypothesis test for the difference between two means. However, when the population is not normally distributed, the two-sample t-test has low efficiency. The Wilcoxon-Mann-Whitney U two-sample test or the Kruskal-Wallis test can be considered. These nonparametric tests are used to test for a difference in two or more samples that are drawn from the same distribution. The Kruskal-Wallis test assumes homoscedasticity across samples. For the absence of homoscedasticity (heteroscedasticity), Mood's median test is used to test if the medians of two or more populations are statistically identical or not. Using R, a well-known statistical software, simulations are conducted to explore the skewed distribution such as the Weibull and Chi-Squared distributions. The Weibull distribution has two parameters, the shape parameter (k) and the scale parameter (λ). The goal of this project to compare the efficiencies among the several different tests mentioned above. To measure the efficiency, the power – the probability of rejecting the null hypothesis when the null hypothesis is false – is calculated and compared for each test. Within this study, a real data analysis is conducted using income data which is known as typical skewed right data.

Keywords: Weibull distribution, Chi-Squared distribution, Kruskal-Wallis H-test

1. Introduction

Any introductory statistics course will more than likely only work with problems that assume the given dataset follows a normal distribution and its parametric assumptions. While this is beneficial in mastering the basics of probability and inference, real-world data will rarely follow these assumptions. Different shapes of distributions call for different types of tests. In order to accurately depict what is happening with real-world data, nonparametric distributions – right skewed in this analysis – and assumptions will be tested. The power of each parametric and nonparametric test, with respect to a particular distribution, is assessed using R statistical software.

1.1 Parametric Vs. Nonparametric

Parametric distributions have to meet certain assumptions in order to be deemed parametric. A non-parametric test does not depend on the distribution of the population. The parametric assumptions include being randomly drawn from a normally distributed population consisting of independent observations - save paired values - and of values on an interval or ratio measurement scale. Furthermore, the distributions must have respective populations of approximately equal variances, must be adequately large, and must approximately resemble a normal distribution.[1] There are many more parametric distributions than just the normal distribution. There are statistical procedures and tests specific to parametric distributions. If data does not meet these assumptions, changing the nature of the study or using mathematical transformations of the data can possibly make a once unusable dataset for parametric testing appropriate.[1]

When a distribution does not meet one or all of the parametric assumptions and there is no way to modify the data to meet the parametric assumptions, the data is nonparametric.[1] Similarly to parametric distributions, there are statistical procedures and tests specific to nonparametric distributions.

2. Models

The parametric tests, nonparametric tests, and inferential outcomes are described in this section.

2.1 Two-Sample T-Test

When testing for a difference in means between two independent samples, the parametric two-sample t-test is appropriate to use if both samples follow a normal distribution. The null and alternative hypothesis for a two-sample t-test correspond with the following equations (1) and (2), respectively[1].

$$\mathbf{H_0: \mu_1 = \mu_2} \tag{1}$$
$$\mathbf{H_a: \mu_1 \neq \mu_2} \tag{2}$$

If the resulting t-score and p-value does not prove significant, or above $\alpha = 0.05$, then we accept the null hypothesis. If the resulting t-score and p-value does prove significant, or below $\alpha = 0.05$, then we reject the null hypothesis and accept the alternative. This would be evidence of a significant difference in means of the two groups.[1]

2.2 Mann-Whitney U-test

The Mann-Whitney U-test is the nonparametric equivalent of the two-sample t-test. It is also used to compare means of two independent, or unrelated, samples for significant differences.[2] To compute the U-test, data is ranked ordered and combined into a single dataset. This combination is used to determine if the rank ordering is random or clustered.[2] If the data points of the sample are clustered, then there is evidence of a significant different between the means of the samples. Conversely, randomly distributed rank ordered data would be evidence that there is no significant difference between the means of the samples.[2] We are going to omit this test from the real data analysis because the results were very similar to those of the Kruskal-Wallis H-test.

2.3 Kruskal-Wallis H-test

The Kruskal-Wallis H-test is a nonparametric procedure for comparing two or more samples that are independent. The parametric equivalent of the Kruskal-Wallis H-test would be the one-way analysis of variance, or ANOVA.[2] When the Kruskal-Wallis H-test produces significant results, then at least one of the tested samples is significantly different from the others[2]. However, this test does not identify where the difference occurs or exactly how many differences occur. Sample contrasts, or post hoc tests, can be used to identify the particular differences between sample pairs. Tukey-Kramer confidence intervals can also be used to identify how many significant differences occur and between which sample pairs.[2]

2.4 Mood's median test

In the simulations, Mood's median test performed poorly – thus, it is not included in our real data analysis.

2.5 Power

Statistical power is the probability that a test correctly rejects the null hypothesis when the alternative hypothesis is true. When conducting a hypothesis test, there are four outcomes that are illustrated in Figure 1.[4]

	Do not reject H_0	Reject H_0
H_0 is true	A	B
H_0 is false	C	D

Figure 1. Results B and C are Type I and Type II errors, respectively. Results A and D represent correct decisions. Power is the probability that result D will occur.[4]

3. Simulation

The statistical program R was used to create the following simulations (normal and right-tailed) in order to assess the behavior of both parametric and nonparametric tests on different distributions.

3.1 Normal Distribution

For the simulation on the normal distribution, tests were conducted between equations (3) and (4).

$$\mathbf{H_0: \mu = 75} \qquad \textbf{(3)}$$
$$\mathbf{H_a: \mu > 75} \qquad \textbf{(4)}$$

A sample size of 100 and a significance level of 0.05 was used. The power was computed for each of the tests as μ was incremented according to the sample size (100). The mean was initialized to 75 and was incremented to 78. A plot of the power was created for each test as seen in Figure 2.

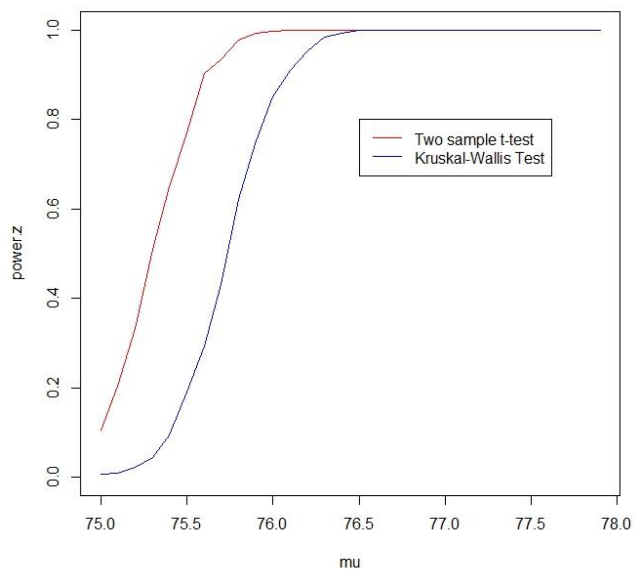

Figure 2. Plot of power for two-sample t-test and Kruskall Wallis H-Test on normal distribution

When applied to the normal distribution, the parametric two-sample test outperformed the Kruskal Wallis test with respect to power.

3.2 Chi-Squared Distribution

The simulation for the Chi-Squared distribution was similar to that of the normal distribution. Tests were conducted between equations (3) and (4), as seen above, using a sample size of 100 and a significance level of 0.05.
Then, the power was computed for each of the tests as μ was incremented from 75 to 78 and a plot of the power was created for each as seen in Figure 3.

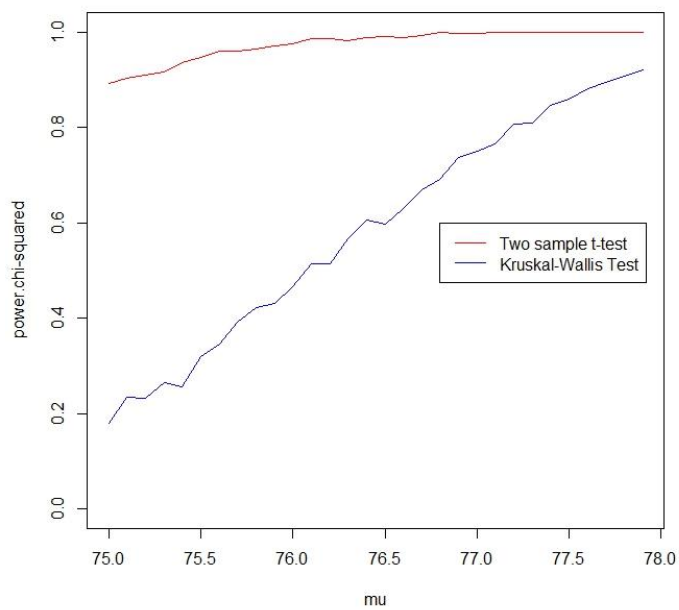

Figure 3. Plot of power for two-sample t-test and Kruskal-Wallis H-Test on Chi-Squared distribution

When applied to the Chi-Squared distribution, the nonparametric Kruskal-Wallis H-test outperformed the two-sample t-test with respect to power. The two-sample t-test immediately starts rejecting the null when μ is incremented, where the Kruskal-Wallis test increases power as μ tends away from its hypothesized value. The two-sample t-test is experiencing a validity problem. It is immediately rejecting the null as soon as the mean differs, even by 0.05, from the original value. The t-test's probability of a Type I error is vastly increased due to the central limit theorem effect. The only way that a t-test's power would not be inflated by Type I error is if the distribution followed normal assumptions. Since we are testing with the Chi-Squared distribution, we see the validity problem illustrated in the figure above.

3.3 Weibull Distribution

The Weibull distribution is dependent on two parameters; λ, the scale parameter and *k*, the shape parameter. For the simulation on the Weibull distribution a test was conducted between equations (5) and (6) using a sample size of 100, a significance level of 0.05, and a shape parameter *k* of 1.5 to achieve a right tailed distribution.

$$\mathbf{H_0: \lambda = 1} \quad \mathbf{(5)}$$
$$\mathbf{H_a: \lambda > 1} \quad \mathbf{(6)}$$

The power was computed for each of the tests as λ was incremented from 1 to 2.5 by increments of 0.05. Finally, a plot of the power was created for each test as seen in Figure 4.

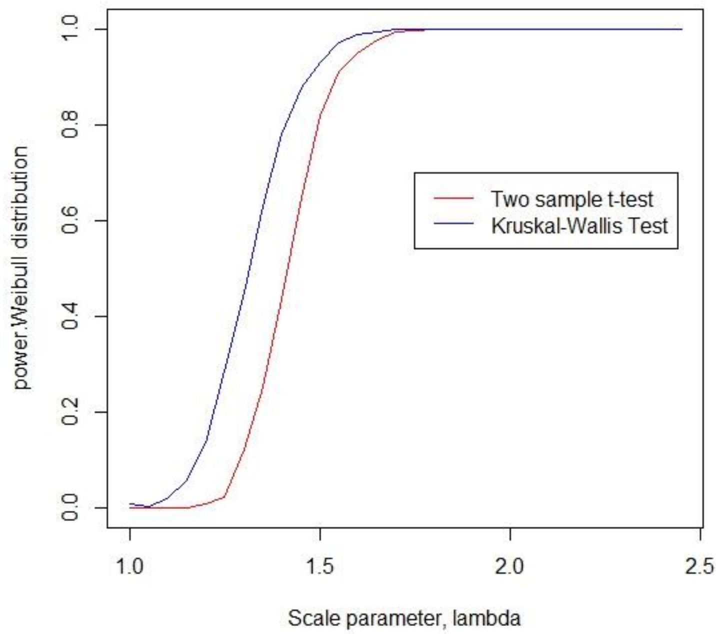

Figure 4. Plot of power for two-sample t-test and Kruskal-Wallis H-Test on Weibull distribution

When applied to the Weibull distribution, similarly to the Chi-Squared distribution, the nonparametric Kruskal-Wallis test outperformed the parametric test with respect to power in comparison to the two-sample t-test.

4. Real Data Analysis

A right skewed distribution can typically be seen when looking at salaries within academia. To examine the parametric and nonparametric tests on real-world right skewed data, an analysis was conducted to determine whether or not a difference exists in salary between male and female faculty. Data were derived from the Charlotte News and Observer for professor, assistant professor, and lecturer total salaries at University of North Carolina at Chapel Hill for the 2014-2015 academic year, including state and non-state compensation.[5] A departmental subset was also tested that included faculty from the departments of Biology, Chemistry, and Law. A two-sample t-test and Kruskal-Wallis H-Test were both used to test our hypothesis regarding faculty salary and to examine the behavior of the two tests in regard to the non-normal distribution.

4.1 Descriptive Statistics

The following table outlines the descriptive statistics for the groups tested:

Table 1. Descriptive Statistics : Sample size, means, and standard deviations for the groups of interest.

	Sample size	Mean	Standard deviation
Overall			
Male	1103	$140,486.80	$71,167.21
Female	903	$112,283	$55,116.97
Total	2006	$140,486.80	$65,934.68
Lecturers			
Male	68	$62,604.57	$24,926.72
Female	95	$58,931.75	$24,458.20
Total	163	$60,463.97	$24,645.09
Assistant Professors			
Male	228	$105,146.10	$39,560.49
Female	298	$94,369.94	$34,042.52
Total	526	$99,040.97	$36,889.89
Departmental Subset			
Male	117	$139,251.30	$54,975.41
Female	57	$122,194.90	$53,011.26
Total	174	$133,663.90	$54,777.93

4.2 Results

The following table displays the results of the parametric and nonparametric tests applied to the groups of interest:

Table 2. t-statistic, p-value for t-test, chi-squared statistic, and p-value for chi-squared statistic, followed by the 95% confidence intervals for the difference in means

	Two sample t-test		Kruskal-Wallis H-test	
	t-statistic	p-value	χ^2	p-value
Overall	9.9991	$< 2.2e\text{-}16$	113.85	$< 2.2e\text{-}16$
Lecturers	0.9348	0.3514	1.0712	0.3007
Assistant Professor	3.2862	0.001095	11.015	0.0009036
Departmental Sub	1.9678	0.0515	4.0942	0.04303

We are 95% confident that the true difference in means of the overall group is between $22,672.12 and $33,735.53.
We are 95% confident that the true difference in means of the lecturer group is between -$4,092.98 and $11,438.63.
We are 95% confident that the true difference in means of the assistant professors group is between $4,331.60 and $17,220.65.
We are 95% confident that the true difference in means of the departmental subset group is between -$113.34 and $34,226.23.

The results of the analysis show that a statistically significant difference exists in salary between males and females in academia among the overall group (professors, assistant professors, and lecturers) as well as among the assistant professors subgroup. A significant difference was not observed among the lecturer subgroup. The nonparametric Kruskal-Wallis H-Test rejected the null hypothesis of the departmental subset when the two-sample t-test failed to reject, illustrating how statistical outcome can be affected by distribution.

It is important to note that this analysis was conducted on the general population and did not compare faculty with equal qualifications and/or years of experience. Statistically, a significant difference exists, though this analysis cannot determine the cause. For instance, while some of the impact could stem from a gender wage gap, the impact could also partially stem from other factors such as departmental wage disparities (i.e. male dominated fields that earn more than female dominated fields).

5. Conclusions

5.1 Conclusions From Simulations

Using R to examine the power of certain parametric and nonparametric tests on normal and non-normal distributions, simulations were created for normal, Chi-Squared, and Weibull distributions to compare the results of a parametric two-sample t-test and those of the nonparametric Kruskal-Wallis H-Test.

The nonparametric tests outperformed the parametric tests with respect to power when tested on nonparametric distributions. For the Chi-Squared distribution, the parametric tests suggests Type I errors, while on the Weibull distribution it suggests Type II errors. While these results are representative of these simulations, variances in the simulation (such as a change in sample size) may significantly affect the outcome.

5.2 Conclusions From Real Data Analysis

The results of the analysis indicate that salary differences do exist between males and females among the overall faculty and assistant professor cohort. A significant difference was not observed among the lecturer subgroup.

The Kruskal-Wallis H-Test had a lower p-value and higher power when the null was actually false in the case of the assistant professor cohort. Additionally, the Kruskal-Wallis H-test rejected the null that differences exist in the departmental subset when the two-sample t-test did not, demonstrating discrepancies that can occur when testing on non-normal distributions.

6. Acknowledgements

The authors wish to extend their appreciation to Dr. Jimin Lee for her guidance, and Dr. Steve Patch for his assistance in reviewing the study.

7. References

1. Joan Welkowitz, Robert B. Ewen, and Jacob Cohen, Introductory Statistics for the Behavioral Sciences. (New York: Academic Press, 1971) 167:168.
2. Gregory W. Corder and Dale I. Foreman, Nonparametric Statistics: A Step-by-Step Approach. (Hoboken, New Jersey : John Wiley & Sons, 2014): 69:71, 117:119.
3. A. M. Mood, “On the Asymptotic Efficiency of Certain Nonparametric Two-Sample Tests,” The Annals of Mathematical Statistics 25, (Sep., 1954): 516, http://0-www.jstor.org.wncln.wncln.org/stable/2236833
4. Kevin R. Murphy and Brett Myors. Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests, (Mahwah, N.J.: Lawrence Erlbaum Associates, Inc, 1998.) eBook Collection (EBSCOhost), 5.
5. Charlotte News and Observer, “Look up salaries of UNC system workers”, newsobserver.com, http://www.newsobserver.com/news/databases/public-salaries/
6. Paul D. Umbach, "Gender Equity in the Academic Labor Market: An Analysis of Academic Disciplines." Research in Higher Education 48, no. 2 (2007): 169.
7. Corinne A. Moss-Racusin and others, “Science Faculty’s Subtle Gender Biases Favor Male Students”, Proceedings of the National Academy of Sciences 109, no.41, (Oct., 2012): 16474:16479.