1-1-2012

Robustness And Power Of The Kornbrot Rank Difference, Signed Ranks, And Dependent Samples T-Test

Norman Haidous
Wayne State University,

ROBUSTNESS AND POWER OF THE KORNBROT RANK DIFFERENCE, SIGNED RANKS, AND DEPENDENT SAMPLES T-TEST

by

NORMAN N. HAIDOUS

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2012

MAJOR: EVALUATION AND RESEARCH

Approved by:

____
Advisor Date

____

____

____

____

DEDICATION

To my dear wife Dr. Salima: Our first meeting in front of Pulitzer Fountain in New York City will always be a special moment for me because it was the beginning of my great love affair with you. Every day I look at you I am reminded of the excitement I felt deep in my heart, and how truly fortunate I am to have you in my life. I had hoped to meet a companion with intelligence, faith, and beauty - a person who is accomplished in who she is and could carry herself with confidence and with compassion for other people. Thank you for being so much more than I could have ever envisioned, and for taking chances and always teaching me to look forward. You are the brightest light in my life and you provide me with the direction of my life, and the warmth and support to be a better person. I am so blessed to have you in my life and to receive your love and kindness.

To my parents Abdul and Balassem Haidous: Thank you for imparting the thirst for knowledge, and the determination and confidence that I can achieve any aspiration. You taught me to look at the world with my eyes wide open and to take in everything it has to offer, to never judge, and to always look for the good. My life is better for it. Your expectations and your continued support of my educational pursuits support one of your greatest lessons in my life, that knowledge is indeed power. Your wisdom, guidance, and values will always be my guiding light on this journey. Thank you for giving me everything I have ever needed, your love and blessings. I am proud to be your son.

To the remarkable four - my sisters Sarah, Randa, Lena and Dania. Thank you for your love, support, and mostly importantly, your honesty. I am lucky to have four sisters, each possessing qualities that have contributed greatly to my life. To Sarah, the trailblazer of our group, her strength lies not only in what she does but in who she is, a remarkable person. She has always challenged my way of thinking and always had the confidence in me to make the right decisions. She is the best older sister a brother could have and I will always seek her guidance. To Randa, who has taught me to never be upset (even if she was the one throwing me under the bus), and to always appreciate humor in a difficult circumstance. She has taught me that with every challenge comes ease, never to worry. To Lena, who has shown me how to manage the multiple tasks of my life, and inspired me with her love of learning and knowledge. She has always provided her support and a high level of expectation in *whatever* I set out to do. To my youngest sister Dr. Dania, who never waivers in her love and support of her family. She has taught me strength, loyalty, and integrity, the foundations instilled in us by our grandparents. She is not only an inspiration to me, but to everyone she meets. I have learned from her to always improve myself and to look for the next challenge, chop, chop. She has taught me to take time to listen to others and always be a friend. She has shown me by example how to conduct my life by the highest of standards.

To my nieces and nephews, who at certain times in my life provided me with the love and affection I needed. Thank you to my sisters for teaching them to love their favorite Khalo.

ACKNOWLEDGMENTS

It is with great pleasure that I respectfully acknowledge the support of all those who helped me prepare and complete this dissertation study. These individuals assisted with many aspects of this study and provided their knowledge and guidance through the entire process.

Wayne State University is a university of excellence and great education, and I was very honored and fortunate to work under the mentorship of an expert in the field of Statistics, Dr. Shlomo Sawilowsky. I am very grateful for his dedication to this thesis project, and I sincerely thank him for the numerous hours of time he spent in helping me with my FORTRAN program. His knowledge and expertise has motivated and inspired me to hold myself to the highest of academic standards.

This dissertation would not have been possible without the help of my committee. Each committee member not only contributed their time and input, they shared in the interest of the final results. I would like to thank Dr. Gail Fahoome, who was so generous with her time and instruction, and who I had the pleasure of getting to know as a student in many of her classes. Dr. Fahoome provided support and encouragement at every stage of my dissertation process, and challenged me to explore many aspects of my research. She is someone who I not only respect as an instructor but as a person. Dr. Bob Pettapiece, who graciously joined my committee when a former committee member retired, provided invaluable support to this project. I became acquainted with Dr. Pettapiece as an advisor to my sister in her doctoral program, and I regard him highly for his wonderful mentorship. Dr. Pettepiece provided critical comments and direction to my paper. I extend much appreciation to my cognate advisor from the School of Business Administration, Dr. Scott Julian, who advised me regarding the application of my research to the business world. Thank you for your support and interest in my research.

I would like to thank Dr. Julie Smith who spent countless hours helping me prepare for my exams, and never asked for anything in return. She is a kind-hearted and generous person who I cannot thank enough. I wish her and Mark always the very best.

Thank you to Zahi for sacrificing time away with your wife so that she could provide me critical feedback. Thank you to Sayed and Elisabeth for the words of encouragement and support along the way, especially for the jumper cables. My parents for always reminding me how to get to Carnegie Hall, practice, practice, practice.

I would like to thank my family at the Ford Motor Company for supporting me on this journey. From the day I joined the company I never felt like anything less than being part of the Ford family. It is a company that is dedicated to the educational pursuits of their employees and their families. Ford challenges us to Go Further in our work and all aspects of our life. Thank you.

Thank you to my wife Salima and my sister Dania for keeping me focused.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Previous research has mistakenly described nonparametric, or distribution free tests, to require assumptions that are "fewer and weaker than those associated with parametric tests" (Siegel & Castellan, 1988, p. 34). It is incorrect to assume that if the underlying assumptions are easier to meet, or are weaker, that the hypotheses are of less importance or that nonparametric tests are in any way less powerful than their parametric counterparts (Sawilowsky, 1990: Blair & Higgins, 1985: Siegel & Castellan, 1988). Nonparametric tests involve few if any assumptions about the shapes of the underlying population distributions. For example, the Wilcoxon Signed-Ranks test is the nonparametric alternative to the parametric paired sample t-test when the assumption of normality is not met. Under the t-test the null hypothesis is the mean difference is zero; under the Wilcoxon Signed-Ranks test $f(x) = g(x)$, which includes all parameters. The Wilcoxon Signed-Ranks test does test the differences in means better than the t-test when population normality has not been met. Thus, nonparametric tests should be used in place of parametric tests when the assumption of population normality has been violated.

A nonparametric procedure tests hypotheses without any appeal to population parameters (Sawilowsky & Fahoome, 2003). According to Walsh (1968), a test is regarded as nonparametric when its Type I error properties are satisfied when the assumption of normality does not hold. According to Sawilowsky (1990), there are three types of nonparametric tests that are used for categorical data, signed data, and ranked data. The Chi-square test of independence, Sign test, and the Wilcoxon rank-sum tests, respectively are examples (Hinkle et al., 2003; Neave & Worthington, 1988; Wilcox, 1996).

In addition to nonparametric tests being robust with regard to Type I error, the Wilcoxon rank sum test, for example, can be much more powerful than the parametric t-test when the population's distribution is non-normal (Blair & Higgins, 1985; Hodges & Lehmann, 1956; Sawilowsky, 1990; Sawilowsky & Fahoome, 2003; Zimmerman & Zumbo, 1989). In fact, it can be up to four times more powerful than the t-test when the data are gathered from an exponential distribution (Sawilowsky & Blair, 1992).

The Wilcoxon Signed-Ranks test is often used in the field of education when investigating the effects of an intervention on a controlled and experimental group; by comparing the pretest and posttest of the two groups (Christie & Enz, 1992; Mahar & Strobert, 2010; & Mcintosh et al., 1993). For example, there have been studies that have compared the effects of tutoring students to determine if there was a statistically significant difference between the controlled and experimental groups on a posttest comparison.

Within behavior sciences and education, small sample sizes are not uncommon. Nonparametric statistics allow for research to be conducted when the assumptions of normality are not met and sample sizes are small. Specifically, within the field of education, the possibility of conducting research at a reduced or small sample size may be more cost-efficient and still provide reliable results. This allows for research within education that otherwise may not have been possible due to accessibility to large samples due to cost.

Nonparametric Rank Tests

There are several methods to compute the correlation between ranking lists (Tate and Clelland, 1957; Mosteller and Rourke, 1973; Daniel, 1978; Gibbons, 1985; Kornbrot, 1990). Most of these nonparametric statistical tests are designed to illustrate the correlation of different datasets in fields such as psychology, ecology, or material science.

A nonparametric test that developed to assess the differences in means is the Wilcoxon Signed-Ranks test (Daniel 1978; Gibbons 1985). It tests f(x) = g(x), and is comparatively more powerful than the t-test when testing for a difference in means when only the assumption of population normality has been violated. The Wilcoxon Signed-Ranks test does not require that the population is normally distributed. However, the test assumes that the population distribution is symmetric. The procedure for the Wilcoxon Signed-Ranks test is shown below:

(1) Compute differences for each of *n* pairs.

(2) Drop zeros.

(3) Order the *absolute differences* from smallest to largest.

(4) Assign ranks 1, ⋯ *n*, with average rank for ties.

(5) *S* = sum of the ranks for the pairs where the difference is positive.

(6) Compare *S* to $N\left(\frac{n(n+1)}{4}, \sqrt{\frac{n(n+1)(2n+1)}{24}}\right)$

A distribution free inferential procedure for comparing paired observations, called the ‘rank difference test’, is based on ranks, and may be suitable as a robust alternative to the related f-test in all situations where the Wilcoxon Signed-Ranks test is applicable (Kornbrot, 1990). According to Kornbrot (1990) it may be applied to ordinal data or operational measures which do not meet the assumptions of the Wilcoxon. The procedure for the rank difference test is shown below (Kornbrot, 1990):

(0) Assume there are n pairs of ordinal observations p_a [*x(i), x(j)*] with values *x(i*); *i*=1,2,…2*n*, and pair indices , *a; a*=1,2,…,*n*.

(1) Rank all the 2*n* observations, so that *r(i)* is the rank of observation *x(i)*

(2) Then perform the Wilcoxon test in the normal way on the *r(i)* rather than the *x(i)*.

For each pair of measures, define a rank difference measure t_a (*i, j*)=*r*(*i*)-*r(j)*; The statistic, D, is obtained by finding the rank, r_a *(i, j)* of $|t_a(i, j)|$ for each pair of observations, and calculating:

R_- = sum of all ranks corresponding to negative t_a.

R_+ = sum of all ranks corresponding to positive t_a.

(3) D is then smaller of R_+ and R_-.

(4) Tabulated values of the D statistic, or a normal approximation corrected for continuity may be used to calculate the probability that any given value of D would have occurred under the null hypothesis of no difference between the treatments.

The assumption of rankability of the differences is normally not met with operational measures. Kornbrot (1990) discussed this in detail with reference to operational measures of times, rates, and counts. These are common operational measures in psychology and education for example, time as an index of information processing, and counts are often used to determine errors on tasks. Kornbrot (199) presented an alternative statistical test to the Wilcoxon Signed-Ranks test called the rank difference test. This procedure is applicable when operational measures do not meet the assumptions underlying use of the Wilcoxon Signed-Ranks test, in particular if there is doubt about rankability of the difference scores. Both exact sampling distributions and large sample approximations for the sample statistic D are given in Kornbrot (1990).

According to Kornbrot (1990) the rank difference test, can be used exactly as one would use the Wilcoxon's Signed-Ranks, an alternative that is useful when one has ordinal data. Although it may be argued that it is less powerful and less efficient than the Wilcoxon Rank Sum.

Wilcoxon's Signed-Ranks test generally is used in two situations: (1) with continuous data that are, or may be, distributed non-normally, and (2) with ordinal data. However, as Kornbrot pointed out, “a procedure is meaningful for ordinal data if it gives the same result when applied to the original data, or any strictly monotone transformation of the data” (p. 244).

According to Kornbrot (1990), although the rank difference statistic D can take on half-integer, as well as integer, values, it has certain continuity advantages over the Wilcoxon for small samples.

Power

The power of a test is the probability that the test statistic will lead to rejection of the null hypothesis (H_o). Alternatively, the power of a test is 1 minus the probability of a type II error. Power is the probability of the correct decision, which implies Ho is rejected, while H_o is false.

Power depends on the following four factors (Gibbon, 1971):

1.) The degree of falseness of H_o, that is, the amount of discrepancy between the assertion as stated in H_o and the true condition.

2.) The size of the test, α, which is also called the significance level or the probability of a Type I error.

3.) The number of random observations involved in the test statistic, i.e., the sample size.

4.) The underlying population distribution that generates the random process.

The power function of a test is the power when all except one of these variable are held constant, generally item 1.

Purpose of the Study

There are situations when researchers want to compare two related samples, and the assumptions underlying the use of the applicable t-test are not met (Kornbrot, 1990). Generally, the Wilcoxon Signed-Ranks test is used in these situations.

According to Kornbrot (1990) “for non-normally distributed internal data the Wilcoxon Signed-Ranks test is a useful and legitimate procedure, although other, perhaps less familiar, robust procedures might have substantial advantages for small samples” (p. 242). Kornbrot (1990) argued that the Wilcoxon Signed-Ranks test is not meaningful in these situations because the procedure entails the subtraction of ordinal scale values. The 'rank difference test' was proposed as a more suitable alternative to the Wilcoxon Signed-Ranks test, when dealing with ordinal data and small samples.

The purpose of this study is to compare the power of the Wilcoxon Signed-Ranks test in comparison to the rank difference test when the assumption of normality is not met, the data are ordinal, and the sample size is small. The study will also investigate Kornbrot’s (1990) claim that the rank difference test should be used over the Wilcoxon Signed-Ranks tests "in all paired comparison designs where the data are not both of interval scale type and of known distribution time" (p. 258). To test Kornbrot's claim of superiority of the rank difference test, a Monte Carlo simulation using Compaq 6.6c Fortran 77 will be used.

Monte Carlo studies have documented that several nonparametric rank tests are better suited than parametric tests under numerous nonnormal distributions (Sawilowsky, 1990). Sawilowsky further pointed out that "many variables encountered in education and psychology that are treated as interval in scale may be better justified as ordinal in scale. To the extent that these variables are indeed ordinal, the loss-of-information issue vanishes" (p. 95).

A Monte Carlo simulation will be used to investigate the robustness of the statistics by repeatedly sampling from a selection of distributions and applying the statistics to the resulting samples. The samples will be changed to reflect different effect combinations for the main and interaction effects. The Monte Carlo simulation will allow for several parameters of the Wilcoxon Signed-Ranks test and the 'rank difference test,' to be assessed and controlled, with regards to shift in location.

Assumptions and Limitations

The Monte Carlo study involves random sampling and will depend on the pseudo-random number generator in the sampling process. Compaq 6.6c Fortran 77 will be used within this study. Five distributions will be compiled, and although they are extensive in their representation, they do not represent all possible distributions that could occur, as there are an infinite number. Also, the sample sizes do not represent every possible configuration that might be of importance in applied research. Although, it should be noted that the sample size augmentations and the range of the distributional types are understood to be adequate for the purpose of outlining the comparative accuracy and reliability of the ranking methods in true settings.

The study is limited to three configurations of sample sizes, and it should be understood that there are still many pairs of equal sample sizes that could be simulated. A higher statistical power may ensure the chance of generalizing the findings of the hypothesis test to setting with larger populations. There will be 100,000 replications within the Monte Carlo study.

Definition of Terms

Alpha level. The alpha level is often referred to as the level of significance and refers to “the probability of making a Type I error if Ho is rejected” (Hinkle, et al., 1998, p. 618).

Asymptotic Relative Efficiency (ARE). Is a measure of performance on large sample data from a normally distributed population (Kornbrot, 1990).

Independent Sample t Test. The independent samples t-test can be defined as “a test statistic for determining the significance of a difference between means" (for a two sample case) (Runyon & Haber, 1991, p. 337).

Interval Data. Is continuous data where differences are interpretable, but where there is no natural zero**.**

Monte Carlo. Monte Carlo studies are computer simulations that allow for the measuring of mathematical properties of statistical tests (Harwell,1990).

Normality. An underlying assumption for parametric test.

Ordinal Data. Are categorical data where there is a logical ordering to the categories.

Power. Power is "the probability of rejecting the null hypothesis when it is false" (Hinkle, et al., 1998, p. 620).

Rank Difference test. A distribution free inferential procedure for comparing paired observations, based on ranks (Kornbrot, 1990).

Robustness. Hunter and May (1993) defined robustness of a statistical test as "the extent that violating its assumptions does not appreciable affect the probability of its Type I error" (p.386). Sawilowsky (1990) added that robustness also pertains to Type II error.

t- test. A statistical procedure frequently employed by researchers to analyze the results of two group studies (Keppel, 1991). The formula for the t-test is

$$t = \frac{\overline{Y}_{A_1} - \overline{Y}_{A_2}}{\hat{\sigma}_{\overline{Y}_{A_1} - \overline{Y}_{A_2}}}$$

where $\overline{Y}_{A_1}$ and $\overline{Y}_{A_2}$ = the two means being compared

$\hat{\sigma}_{\overline{Y}_{A_1} - \overline{Y}_{A_2}}$ = the standard error of the difference between the two means (p. 121)

Type I error. A false positive error made in interpreting the results of a particular test. The null hypothesis is rejected, when it should have been accepted. This leads the researcher to arrive at the conclusion that a treatment is effective, when that is not the case.

Wilcoxon Signed-Ranks test. A nonparametric alternative to the paired Student's t-test for the case of two related samples or repeated measurements on a single sample.

CHAPTER 2

LITERATURE REVIEW

The creation and progress of ranking methods stems from two related factors: the psychological attempt to measure mental phenomena and the statistical attempt to calculate the area under the unit normal distribution. Knowledge, intellectual ability, and, personality are psychological constructs that can only be measured indirectly, not by direct observation (Dunn-Rankin, 1983). The scales that explain them are hierarchical; they result in higher or lower scores—but these scores do not reflect exact quantities.

Normal and Nonnormal Data

The t-test is one of the most regularly and widely used statistical procedures in most fields of research. Although, since most data distributions do not meet the assumption of normality, this is grounds for concern. Sawilowsky & Blair (1992) noted that the *t*-test is robust to type I error under the following circumstances:

1. Sample sizes must be equal or nearly so.
2. Sample sizes must be fairly large (25 to 30 according to Boneau, 1960)
3. Tests should be one-tailed instead of two-tailed.

According to Nunnally (1978), “test scores are seldom normally distributed” (p.160). In a study conducted by Micceri (1989), he examined at 440 large-sample real data sets from text publishers, authors of research articles, and testing results from school districts. Micceri found that none of the 440 data sets met the assumptions of normality under a Kolmogorov-Smirnov goodness-of-fit test. Although, it was noted that 19of the 440 (4.3%) real data sets approximated normal distribution, the study showed that “extremes of asymmetry and lumpiness are more the rule than the exception” (p. 161). Micceri (1989) concluded:

> The implications these findings have for normality-assuming statistics are unclear. Prior robustness studies have generally limited themselves either to computational evaluation of asymptotic theory or to Monte Carlo investigations of interesting mathematical functions (p. 163).

The data from this study emphasized the prevalence of nonnormality in real social and behavioral science data sets.

Bradley (1977) provided a rationale for adopting a statistical approach that answers to the fundamental nonnormality of most real data:

> One often hears the objection that if a distribution has a bizarre shape one should simply find and control the variable responsible for it. This outlook is appropriate enough to the area of quality control, but it is inappropriate to the behavioral sciences, and perhaps other areas, where the experimenter, even if he knew about the culprit variable and its influence upon population shape, is generally not interested in eliminating an assignable cause, but rather in coping with (i.e., drawing inferences about) a population in which it’s free to vary (p.149).

The frequency of nonnormal distributions in education, psychology, and related disciplines requires an exploration of transformation procedures.

Nonparametric Tests

As a result of the distributions discovered in the Micceri (1989) study being very unlikely, Sawilowsky (1990) concluded:

> The selection of non-parametric tests is obvious to the researcher who is convinced of the non-robustness of the t-test and ANOVA. Furthermore, the researcher who insists on the robustness of the t-test an ANOVA should be aware that non-parametric tests are also robust, and they are often more powerful under non-normality (p. 98).

Nonparametric and distribution-free statistics tend to be thought of as one of the same, but they do not have the same meaning. Nonparametric tests make no assumption regarding the parameter in a statistical density function, whereas a distribution free test makes no assumptions about the form of the sampled population (Bradley, 1968). Conover (1980) explained that the lack of consistency regarding the definition of the word nonparametric needs to be addressed by defining parametric and nonparametric as statistical hypotheses, and should not be applied to statistics, tests, or types of inference. Noether (1967) further clarified this issue:

> Originally the term non-parametric seems to be due to Wolfowitz (1942), who suggested it in order to indicate that the underlying population could not be completely specified in terms of a finite number of parameters. In a sense the word non-parametric is misleading. “Non-parametric” methods may be used to find confidence intervals for “parameters” like the median of the distribution…For practical purposes it does not seem really important to associate a precise meaning with the word non-parametric. What is important is to dissociate from the term an implication of inferiority and distrust (p. 2).

There is plenty of evidence that suggests that researchers are aware of the limitations of the independent samples t-test when the assumptions are violated. For example, between the periods of 1990 to 1995, more than 850 publications from a variety of disciplines reported using the Wilcoxon Rank Sum test (Pett, 1997, p. 170). Although, it is unlikely that all of these employed ranked data, would suggest that the researchers had chosen to convert interval level data to ranks to avert the need for the assumption of normality.

The Wilcoxon test is possibly considered the most powerful nonparametric test (Runyon & Haber, 1991). According to Bradley (1968), “In comparison to other distribution free statistics, the Wilcoxon test typically either ranks first or, when the set of tests being compared includes the optimum test for the conditions of comparison, ranks a close second” (p. 109-110). According to Kerlinger and Lee (2000), “Nonparametric methods are virtually inexhaustible” (p. 242); “given the relatively simple principles involved and the various properties of data that can be exploited: range, periodicity, distributions, and rank" (p. 425).

Rank Methods

Ranking is the method of positioning objects on an ordinal scale in relation to others. Much of the motivation for the development of the subject has in fact stemmed from the social sciences and the statistical need to calculate the area under the unit normal distribution. Knowledge, intellectual capability, and character qualities are constructs that can only be measured directly and not indirectly (Dunn-Rankin, 1983).

Ranking methods deal with statistics (rank-order statistics) constructed from the ranks, generally in random samples of observations. An ordinal scale of measurement suffices for the calculation of such statistics, but ranking methods are also frequently used even when meaningfully numerical measurements are accessible. In cases when such measurements are available the ordered observations can be denoted by x(1) ≤ x(2) ≤... ≤ $x_{(n)}$, where $x_{(i)}$(i = 1, 2, . . ., *n*) is the ith order *statistic*. Only the rank i of $x_{(i)}$ occurs in rank-order statistics; the use of $x_{(i)}$ itself leads to order statistics (Friedman, 1937).

When numerical measurements are replaced by their ranks, the issue of how much information is thereby lost. In most scenarios, the loss in efficiency or power of the standard rank tests in comparison to the best corresponding parametric tests is generally very small (Friedman, 1937). The loss is based on several constructs: the type of test used, the nature of the alternative to the null hypothesis under test, and the sample size. Because the best parametric test in a given situation depends on the often uncertain form of the underlying distribution, the performance of the two tests should also be compared for other distributional forms which may reasonably occur. Whereas the significance level of the rank test remains quite unchanged, that of the parametric test may be seriously be disturb. Even when the parametric test is not too sensitive to this, it can become inferior in power to the rank test (Friedman, 1937; Sawilowsky & Blair, 1992). Often ranking methods are much less affected by unauthentic or wild observations.

Wilcoxon Signed-Ranks Test

The Wilcoxon Signed-Ranks test does not require that the population is normally distributed. However, the test assumes that the population distribution is symmetric. The procedure for the Wilcoxon Signed-Ranks test is shown below:

(1) Compute differences for each of *n* pairs.

(2) Drop zeros.

(3) Order the *absolute differences* from smallest to largest.

(4) Assign ranks 1, ⋯ *n*, with average rank for ties.

(5) *S* = sum of the ranks for the pairs where the difference is positive.

(6) Compare *S* to $N\left(\frac{n(n+1)}{4}, \sqrt{\frac{n(n+1)(2n+1)}{24}}\right)$

The Wilcoxon Signed-Ranks test and the paired sample t-tests are often thought of as adversaries in research. Blair and Higgins (1985) pointed out that "their power/efficiency properties have not been carefully and extensively compared" (p. 120). In order to compare the efficiencies of two statistical tests, statisticians often use relative efficiency (RE) and asymptotic relative efficiency (ARE). According to Blair and Higgins, RE is defined by b/a, "where *a* is the number of observations required by Tests A to equal the power of Test b, which is based on *b* observations" (p. 120). ARE can be defined as the limiting value of b/a as *a* can vary to allow Test A the same power as Test B, while *b* approaches infinity and the treatment effect approaches 0. The BRE can also be used to compare efficiencies of two tests; it can be found by "comparing the exponential rate of convergence to zero of the Type I error rate while keeping the Type II error rate fixed" (p. 120).

According to Elshoff and Elashoff (1978), Lehmann (1975), and Siddiqui and Raghunandanan (1967), ARE's computed for the Wilcoxon Signed-Ranks test in comparison to the t-tests, there was a minimal advantage for the t-test under the normal curve. However, under the Cauchy distribution there was a large advantage for the Wilcoxon Signed-Ranks test in comparison with the t-test. In a study conducted by Klotz (1965), the BRE's were calculated for the Wilcoxon Signed-Ranks test in comparison to the t-test, the BRE was shown to be as high as 0.981, but approached 0 as the mean of the difference score population approached infinity, showing the treatment effect plays a pivotal role in the efficiency of the two tests. According to Blair and Higgins, based on the stated research regarding ARE, RE, and BRE, "when the normality assumption is relaxed the WSR test can attain truly large (in theory, infinitely large) advantages over the t-test" (p. 121).

Blair and Higgins (1985) used a Monte Carlo simulation to investigate the relative power of the Wilcoxon Signed-Ranks test and t-tests in multiple ways. Within the study, the power of the two tests were tested under 10 different population shares, 3 sample sizes, and 4 significance levels. Within the study, the algorithm utilized to construct the data sets is shown:

$x_{i1} = t_i + e_{i1}(i = 1,2,.....,n),$ where x_{i1} is the score of the ith subject at the first testing period, and t_i is the true component of the ith subject's score, and e_{i1} is the random error component of the ith subject's score at the first testing period. (p. 122)

The posttest score within the study is shown as:

$x_{i2} = c + t_i + e_{i2}(i = 1,2,.....,n),$ where x_{i2} is the score of the ith subject at the second testing period, t_i is the true component of the ith subject's score, e_{i2} is the random error component of the ith subject's score at the second testing period, and *c* is a number common to all subject's posttest scores and represents the treatment effect (p. 122).

Lastly, the difference school for which statistical tests for treatment effects is shown as:

$x_{i2} - x_{i1} = c + e_{i2} + e_{i1}(i = 1,2,.....,n)$ (p.122).

Within the study, for every value of *c,* 5,000 samples were created. For each sample, the Wilcoxon and t statistics were calculated. The following tables have been taken directly from Blair and Higgins, with each of their results being discussed.

Table 1.
Maximum Power Advantages Gained by the t and Wilcoxon Statistics Under Ten Population Shapes, $n = 10$
Blair and Higgins (1985)

	One-tailed significance levels			
Population/statistic	.050	.025	.010	.005
Normal				
T	.006	.019	.028	.047
WSR	.006	.000	.001	.001
Uniform				
T	.020	.031	.033	.052
WSR	.007	.000	.000	.000
Double exponential				
T	.000	.008	.020	.030
WSR	.024	.009	.007	.013
Truncated normal				
T	.003	.012	.029	.037
WSR	.016	.005	.004	.003
Exponential				
T	.000	.015	.029	.035
WSR	.045	.021	.024	.030
Mixed normal				
T	.000	.074	.042	.033
WSR	.309	.186	.206	.186
Mixed exponential				
T	.000	.105	.071	.032
WSR	.383	.232	.239	.241
Lognormal				
T	.000	.021	.020	.011
WSR	.104	.058	.063	.072
Chi-square				
T	.000	.026	.032	.025
WSR	.086	.049	.050	.057
Cauchy				
T	.000	.000	.000	.000
WSR	.221	.110	.109	.113

Table 2.
Maximum Power Advantages Gained by the t and Wilcoxon Statistics Under Ten Population Shapes, $n = 25$
Blair and Higgins (1985)

	One-tailed significance levels			
Population/statistic	.050	.025	.010	.005
Normal				
T	.015	.014	.014	.017
WSR	.000	.004	.000	.000
Uniform				
T	.037	.033	.038	.048
WSR	.000	.001	.001	.000
Double exponential				
T	.000	.000	.000	.000
WSR	.038	.047	.043	.041
Truncated normal				
T	.001	.001	.000	.003
WSR	.014	.018	.011	.004
Exponential				
T	.000	.000	.000	.000
WSR	.087	.100	.099	.093
Mixed normal				
T	.000	.000	.000	.000
WSR	.526	.580	.637	.639
Mixed exponential				
T	.000	.000	.000	.000
WSR	.657	.681	.705	.706
Lognormal				
T	.000	.000	.000	.000
WSR	.221	.233	.234	.227
Chi-square				
t	.000	.000	.000	.000
WSR	.178	.176	.172	.165
Cauchy				
t	.000	.000	.000	.000
WSR	.477	.499	.483	.427

Table 3.
Maximum Power Advantages Gained by the t and Wilcoxon Statistics Under Ten Population Shapes, $n = 50$
Blair and Higgins (1985)

	One-tailed significance levels			
Population/statistic	.050	.025	.010	.005
Normal				
t	.025	.015	.024	.021
WSR	.000	.001	.002	.002
Uniform				
t	.042	.040	.053	.059
WSR	.000	.000	.000	.000
Double exponential				
t	.000	.000	.000	.000
WSR	.046	.056	.059	.055
Truncated normal				
t	.004	.001	.002	.000
WSR	.016	.016	.011	.016
Exponential				
t	.000	.000	.000	.000
WSR	.115	.129	.111	.150
Mixed normal				
t	.000	.000	.000	.000
WSR	.683	.747	.768	.760
Mixed exponential				
t	.000	.000	.000	.000
WSR	.786	.822	.817	.895
Lognormal				
t	.000	.000	.000	.000
WSR	.298	.324	.351	.356
Chi-square				
t	.000	.000	.000	.000
WSR	.238	.262	.268	.270
Cauchy				
t	.000	.000	.000	.000
WSR	.659	.701	.688	.670

The results, shown in the Tables 1-3, show under normal distribution, show a slight power advantage for the t-test under all 3 sample sizes. Under uniform distribution, the t-test again showed a power advantage under all samples over the

Wilcoxon tests, but it should be noted, that the advantages were minimal. Under the double exponential distribution, the Wilcoxon Signed-Ranks test showed a slight advantage over the t-test, when $n = 10$. When $n = 25$ and $n = 50$, there was a clear benefit in using the Wilcoxon Signed-Ranks test over the t-test. Results for the truncated normal distribution showed the Wilcoxon Signed-Ranks test had a small advantage over the t-test, when $n = 10$ (for smaller values of c / σ), although it showed no significant advantage of one test over the other under truncated normal distribution. Under the exponential distribution, mixed normal distribution, mixed exponential distribution, lognormal distribution, and chi square distribution, for larger samples (25 and 50), the Wilcoxon Signed-Ranks test was the more powerful statistic in a convincing manner. Under the Cauchy Distribution, the Wilcoxon Signed-Ranks test had clear advantages at all 3 sample sizes.

According to Blair and Higgins (1985), there were three main points:

> (1) First, the objection might be raised that data, gathered in the course of social and behavioral science inquiries are rarely sufficiently nonnormal to warrant concern…Indeed, it was concern brought about by the appearance of nonnormal data that motivated the early robustness studies (p. 127).
> (2) [S]ome authors seemed to advocate the exclusive use of parametric procedures even in situations where it is known that the assumption of population normality has been seriously violated. Readers should clearly understand that the results of this study do not support a mirror-image argument to that put forth by such writers (p. 127).
> (3) [A]lthough this study dealt solely with the paired samples t-test and Wilcoxon's Signed ranks test, it should be viewed as part of a small but growing body of evidence that is seriously challenging traditional views of nonparametric statistics (p. 127-128).

Rank Difference Test

Kornbrot (1990) presented the rank difference test, a new distribution free inferential procedure for comparing paired observations. Kornbrot stated that the new procedure is a robust alternative to the t-test in all instances where the Wilcoxon Signed-Ranks test is useable. Kornbrot's argument though, is that the rank difference test has advantages over the Wilcoxon Signed-Ranks test for small samples, with similar performance for large samples.

Kornbrot (1990) argued that even though the Wilcoxon is appropriate in instances of non-normally distributed interval measures and “is known to have the maximum possible asymptotic relative efficiency (ARE) for using the formation available in the sample ranks” (p. 242), this does not imply it is the most optimal procedure for that particular data set. If dealing with interval data, Kornbrot opined there is no reason to limit analyses to information within the ranks. Secondly, when using the ARE measure, one is dealing with a measure of performance of large sample data that comes from a normally distributed population. The reasoning for using the Wilcoxon is a result of the data not being normally distributed and or the sample size is not big enough to presume the central limit theorem. Kornbrot focused on the possibility of using less familiar, robust methods to the Wilcoxon Signed-Ranks test when dealing with small samples.

In order to use the Wilcoxon Signed-Ranks test on ordinal data, the two original samples must be rankable, in addition for the difference between scores to be rankable. Kornbrot argued that

> [T]he condition to apply the Wilcoxon on ordinal data, i.e. that ‘nevertheless’ the differences are rankable, almost never occur in practice. Furthermore, the intuition that measures such as times, rates and counts, which are interval in the physical domain, are also interval as operational measures is overoptimistic, at best (p. 243).

Thus, Kornbrot (1990) argued the need for an inferential test of differences between related samples on ordinal measures.

According to Kornbrot, a statistical procedure for ordinal data are meaningful, if the same results can be found when applied on the original data. Kornbrot provided an example in Table 1 and Table 2, which demonstrates that if TIME is used as a measure of performance for the Wilcoxon test a far different value is obtained if RATE = 60/TIME is used.

Table 4. Example of Wilcoxon Signed-Ranks test applied to time/problem as measure of 'alertness' (Kornbrot, 1990)

Subject	Placebo	Drug	Plac-Drug = *d*	Signed *d* rank	Positive *d* rank	Negative *d* rank
1	4.6	2.9	1.7	6	6	
2	4.3	2.8	1.5	5	5	
3	6.7	12.0	-5.3	12		-12
4	5.8	3.8	2.1	7	7	
5	5.0	5.9	-0.9	3		-3
6	4.2	6.5	-2.3	8		-8
7	6.0	3.3	2.7	9	9	
8	2.0	2.3	-0.3	1		-1
9	2.6	2.1	0.5	2	2	
10	10.0	14.3	-4.3	11		-11
11	3.4	2.4	1.0	4	4	
12	7.1	14.0	-6.8	13		-13
13	8.6	4.9	3.7	10	10	
		mean	-.53	number	7	6
		s. e.	0.90	rank sum	43.0	48.0
		related *t*	-0.6	*W*	43.0	
	one-tailed probability:		*P*=0.289		*p*=0.446	

Table 5. Wilcoxon Signed-Ranks test applied to rate of problem solving as measure of 'alertness' (same data as in Table 4 with transformation: Rate - 60/time) (Kornbrot, 1990)

Subject	Placebo	Drug	Plac-Drug = *d*	Signed *d* rank	Positive *d* rank	Negative *d* rank
1	13	20.5	7.5	12	12	
2	13.8	21.1	7.3	11	11	
3	9.0	5.0	-4.0	4		-4
4	10.3	16.0	5.7	8	8	
5	12.0	10.1	-1.9	2		-2
6	14.2	9.2	-5.0	6		-6
7	10.0	18.0	8.0	13	13	
8	30.0	26.5	-3.5	3		-3
9	23.0	29.0	6.0	9	9	
10	6.0	4.2	-1.8	1		-1
11	17.7	24.6	6.9	10	10	
12	8.4	4.3	-4.1	5		-5
13	7.0	12.2	5.2	7	7	
		Mean	2.02	Number	7	6
		s. e.	1.48	rank sum	70.0	21.0
		related *t*	1.4	*W*	21.0	
	one-tailed probability:		*p*=0.100		*p*=0.047	

Thus, Kornbrot concluded in this situation that the Wilcoxon Signed-Ranks test is 'non-meaningful,' because even though the raw scores are ordinal, the differences obtained are not ordinal.

Kornbrot maintained that the Wilcoxon test can only be meaningful if:

$$t_{-} = t_{0} = t_{+}$$

This can only happen when dealing with interval scaled data, according to Kornbrot.

Kornbrot presented the rank difference procedure, shown below:

Rank Difference Procedure

(0) Assume there are n pairs of ordinal observations p_a [*x(i), x(j)*] with values *x(i)*; *i*=1,2,…2*n*, and pair indices , *a; a*=1,2,…,*n*.

(1) Rank all the 2*n* observations, so that *r(i)* is the rank of observation *x(i)*

(2) Then perform the Wilcoxon test in the normal way on the *r(i)* rather than the *x(i)*.

For each pair of measures, define a rank difference measure t_a *(i, j)*=*r*(*i*)-*r(j)*; The statistic, D, is obtained by finding the rank, r_a *(i, j)* of $|t_a(i, j)|$ for each pair of observations, and calculating:

R_{-} = sum of all ranks corresponding to negative t_a.

R_{+} = sum of all ranks corresponding to positive t_a.

(3) D is then smaller of R_{+} and R_{-}.

(4) Tabulated values of the D statistic, or a normal approximation corrected for continuity may be used to calculate the probability that any given value of D would have occurred under the null hypothesis of no difference between the treatments.

Table 6. Rank difference test applied to data in Table 4 (Kornbrot, 1990)

Subject	Placebo	Drug	Plac-Drug = *d*	Signed *d* rank	Positive *d* rank	Negative *d* rank
1	14	20	-6	9.0		-9.0
2	15	21	-6	9.0		-9.0
3	7	3	4	5.5	5.5	
4	11	17	-6	9.0		-9.0
5	12	10	2	1.5	1.5	
6	16	8	8	11.5	11.5	
7	9	19	-10	13.0		-13.0
8	26	24	2	1.5	1.5	
9	22	25	-3	3.5		-3.5
10	4	1	3	3.5	3.5	
11	18	23	-5	7.0		-7.0
12	6	2	4	5.5	5.5	
13	5	13	-8	11.5		-11.5
		Mean	-1.62	Number	6	7
		s. e.	1.57	rank sum	29.0	62.0
		related *t*	-1.0	*D*	29.0	
	one-tailed probability:		*p*=0.162	*p*=0.133	from Table A2 based on simulations	
				p=0.131	from Continuity adjusted normal	

The results in Table 6 show the procedure applied to the same data of Table 4. The table shows that the probability of D, under the null hypothesis is 0.133 (according to Table A2, based on 400,000 simulations in Kornbrot, 1990). As a result of normal approximation with corrections for continuity, z = 1.120, and p = 0.131 one-tailed under the null hypothesis. Thus, according to Kornbrot, the meaningfulness of the rank difference procedure “is guaranteed by the fact that it will give the same result for a set of numeric data, and any monotone transformation of that data” (p. 248).

When comparing the power of the rank difference procedure and the Wilcoxon Signed-Ranks test, asymptotically, for large samples, inferences based on both tests will have the same power. Within classical data,

> From this small and heterogeneous selection of studies it appears that using the rank difference procedure would have been similar in power to using the Wilcoxon, and substantially more powerful than using the sign test (Kornbrot, 1990, p. 253, within Table 7, p. 254).

Similar results were found regarding simulated data, where the rank difference procedure was at least as powerful as the Wilcoxon and significantly more powerful than the sign test.

Similarly, in an experiment conduct by Goldstein (1996), it was found that the Wilcoxon Signed-Ranks tests failed to give the same result when applied to the original data. Goldstein found by using the following syntax:

kornbrot *varname* = *exp [*if *exp] [*in *range]* (p. 29)

For both rank difference procedure and the Wilcoxon Signed-Ranks test, both used the same exact syntax, but the results of the p value for the Wilcoxon Signed-Ranks test did not match up, whereas they did for the rank difference procedure**.**

Kornbrot's recommendation stated that the rank difference procedure should be utilized over the Wilcoxon Signed-Ranks test in all paired comparison designs, in situations where the data are not of both interval scale type and of known distribution type. It should be noted that Kornbrot stated that the rank difference procedure *may* perform better than the Wilcoxon for small samples. Thus, the focus of the dissertation is to provide support for or an argument against this claim.

Monte Carlo Methods

According to Sawilowsky and Fahoome (2003), “Monte Carlo refers to repeated sampling from a probability distribution to determine the long run average of some parameter or characteristic” (p. 46). It is important to distinguish a Monte Carlo method from simulations and Monte Carlo simulations. Simulations are fictitious depictions of reality, whereas a Monte Carlo method is a repetitive process which in this case can be used to solve a mathematical or statistical problem. Combining the two, a Monte Carlo simulation is “is the use of a computer program to simulate some aspect of reality, and making determinations of the nature of reality or change in reality through the repeated sampling via Monte Carlo methods” (p. 46).

Pseudo-random sampling and the repeated use of Monte Carlo methods can be used to estimate a long run average (Sawilowsky & Fahoome, 2003). An example of a Monte Carlo simulation is discussed below:

> For example, a uniform random number generator can be used to simulate the tossing of a coin. Repeating the process many, many times is an application of Monte Carlo methods to that simulation to determine the long run average of what happens when a coin is tossed (p. 145).

Monte Carlo Simulations can be used to evaluate the power and robustness in respect to Type I error of a statistical test under specified conditions. Monte Carlo simulations can be conducted by using computer compilers such as Fortran or other programming languages and platforms Sawilowsky and Fahoome (2003) “found that Fortran is the shortest path to obtaining successful and useful results” (p. 46). Although the learning curve is longer than for R or SAS, the execution times are vastly superior for mathematical applications.

Type I errors, generally referred to as false positives, can have a huge impact on society, more so than Type II errors (Sawilowsky & Fahoome, 2003). Documenting Type I errors using a Monte Carlo study can be accomplished by the following example and explanation by Sawilowsky and Fahoome:

> Suppose a program has been written that draws a sample of scores from a normal distribution using normb1.f90. The population mean is known to be zero. The Z test is conducted on the difference between the sample and hypothesized population mean. If nominal alpha was set to $\alpha = 0.05$, then 1 out of every 20 repetitions of the experiment will produce a statistically significant result, even though the scores are just random numbers and there is no treatment effect. In other words, the null hypothesis is guaranteed to be true because the means of the samples randomly drawn from the population should be approximately equal, with any differences being statistically trivial (p. 173).

The purpose of Monte Carlo simulations is to allow researchers the understanding of what tests work best under certain scenarios and to reduce the occurrence of Type I error.

CHAPTER 3

METHODOLOGY

The purpose of the study is to compare the comparative power and robustness of the Wilcoxon Signed-Ranks test, 'rank difference test', and the t-test. The tests will be compared by using various distributions and sample sizes for a variety of real, nonnormal data sets and pseudo-random number generator data sets. A Monte Carlo simulation will be used to address the following research questions:

1. Do the three tests control Type I error?
2. If Type I error is not controlled, when are the tests liberal or conservative?
3. Can a recommendation be made for the best test for the management of non-normally distributed data?
4. Are there circumstances that control which of the three tests are most suitable under conditions when normality is violated?

Monte Carlo Design

According to Harwell (1990), a Monte Carlo simulation should be conducted in the following manner:

> In the typical MC study of a given statistical test the following process is repeated for a large number of samples: data are simulated which reflect a specific relationship among variables... The values of the statistical test provide information on its properties (e.g., the proportion of the “significant” values on the test). If the underlying assumption of the test were satisfied, exact statistical theory would guarantee that the test would have a specified type I error rate and would permit the probability of rejecting a false statistical hypothesis to be computed. Monte Carlo studies permit these characteristics to be examined when underlying assumptions are violated (p.4).

This study will use Monte Carlo Simulation techniques, using Compaq 6.6c Fortran 77. The program was written and compiled will be used to compare the Type I error rate, power, and robustness of the three statistical tests under the conditions of normal and non-normally distributed data sets, specific to shift in location. The data will be generated using a pseudo-random number generator via Compaq 6.6c Fortran 77; subroutines will draw from Fortran programs developed in the past from other Monte Carlo studies (Sawilowsky & Fahoome, 2003).

The Monte Carlo techniques will allow for the investigation of the comparative power and robustness of the three tests. The tests will be compared using various sample sizes and distributions. The parameters of the study are discussed below.

Sample Size and Nominal Alpha

"One of the primary motivations for utilizing tests of equivalence is that as sample sizes increases, the probability of finding even trivial mean differences statistically significant becomes larger" (Cribbie, Gruman & Arpin-Cribbie, p.5, 2004). The Monte Carlo simulation will be conducted on three equal samples of size combinations (n1, n2) = (10,10), (15,15), (20,20). Sample sizes were selected based on their representation of real world data sets often used in the behavioral and social sciences and to remain consistent with Kornbrot's (1990) study. Nominal alpha levels will be set at .01 and .05. One hundred thousand repetitions will be run in order to compare the accuracy of the two ranking methods.

Study Design

The study design will allow for the comparison of the comparative power and robustness of the Wilcoxon-Signed Ranks-Test, 'rank difference test', and the t-test. Utilizing five distributions and three sample sizes, constants will be added to investigate treatments effects particular to shift in location parameters, with nominal alpha levels of 0.05 and 0.01 being investigated.

For the purpose of investigating Type I error, 100,000 repetitions per condition will be simulated. The simulated Type I error rates will be computed and summarized for each condition and then compared to percentage of rejection rates.

Type I Error

The robustness of each statistical test with respect to Type I error will be addressed via the simulation. The Bradley (1978) liberal criterion test will be used to assess the robustness with respect to Type I error. According to Bradley's (1978) liberal criterion of robustness, a test can be considered robust if its empirical rate of Type I error, α, is within the interval $0.5\alpha < \alpha < 1.5\alpha$. Thus, if the nominal level is $\alpha = 0.05$, the empirical Type I error rate should be within the interval $.025 < \alpha < .075$. Similarly, if the nominal level is $\alpha = 0.01$, the empirical Type I error rate should be within interval $.005 < \alpha < .015$. Type I error rates above the upper robustness limit are considered liberal. Given the null hypothesis and alternative hypothesis for equivalence:

H_0: The difference between means falls above or below the limits of the equivalency interval. The means are found to be non-equivalent.

H_1 : The mean difference falls within the limits of the interval. The means are found to be equivalent.

Based on the ideology above, liberal results imply the means are equivalent, more often than at the desired Type I error, where conservative results imply the means are equivalent, less often than at the desired Type I error.

Sampling Distributions

Micceri (1989) investigated 440 univariate distributions from research within the field of education and psychology. These distributions consisted of 231 standardized achievement and ability test scores, 125 distributions of personality test scores, 49 distributors of gain scores, and 35 distributions of criterion-referenced and mastery test scores. Micceri's research discovered all of the distributions to be nonnormal at the alpha level of 0.01, using the one-sample Kolmogorov-Smirnov test.

Within this study, five population sampling distributions have been selected: 1. Gaussian (Normal) Distribution, 2. Chi-square, df = 1, 3. Chi-square, df = 2, 4. Chi-square, df = 8, 5. t distribution, df = 3. All distributions and their data sets will be created by using a pseudo-random number generator.

1. Gaussian (Normal) Distribution: This is a bell shaped distribution that has equally weighted tails and distributions of scores (excerpted from Sawilowsky & Fahoome, 2003, p. 343).

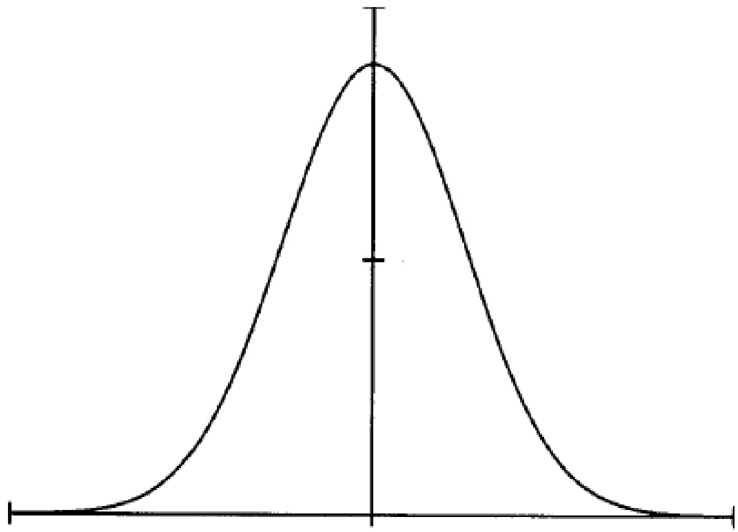

Figure 1. Gaussian (Normal) Distribution (Sawilowsky & Fahoome, 2003)

	General Form	Unit Distribution
Range:	$-\infty < x < \infty$	$-\infty < x < \infty$
μ :	μ	0
Median:	Md or μ_{Md}	0
σ^2 :	σ^2	1
γ_1 :	0	0
γ_2 :	3	3

2. Chi Square (df=1, df=2, df=8) (excerpted from Sawilowsky & Fahoome, 2003, p. 344).

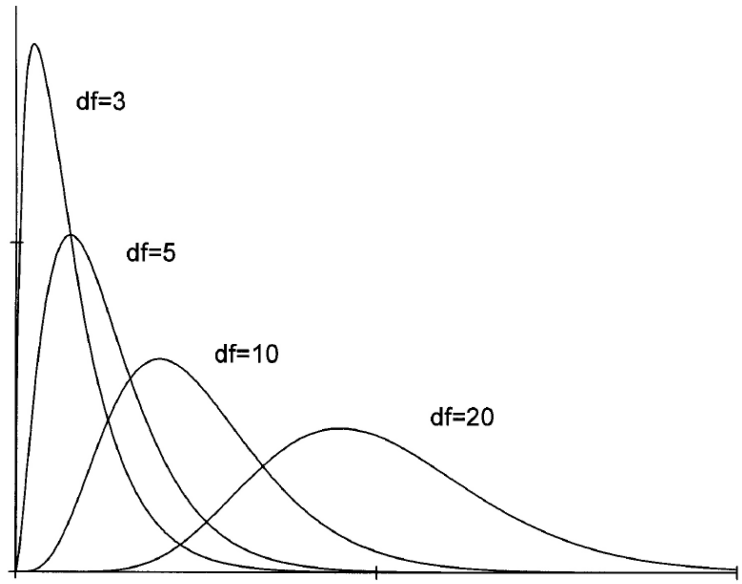

Figure 2. χ^2 (Sawilowsky & Fahoome, 2003)

	General Form	v = 1
Range:	$0 \le x \le \infty$	$0 \le x \le \infty$
μ :	v	1
Median:	$v - \frac{2}{3}$	$\frac{1}{3} or .333$
σ^2 :	2v	2
γ_1 :	$\sqrt{\frac{8}{v}}$	$\sqrt{8} or 2.828$
γ_2 :	$3 + \frac{12}{v}$	15

3. t distribution (df=3) (excerpted from Sawilowsky & Fahoome, 2003 p. 345).

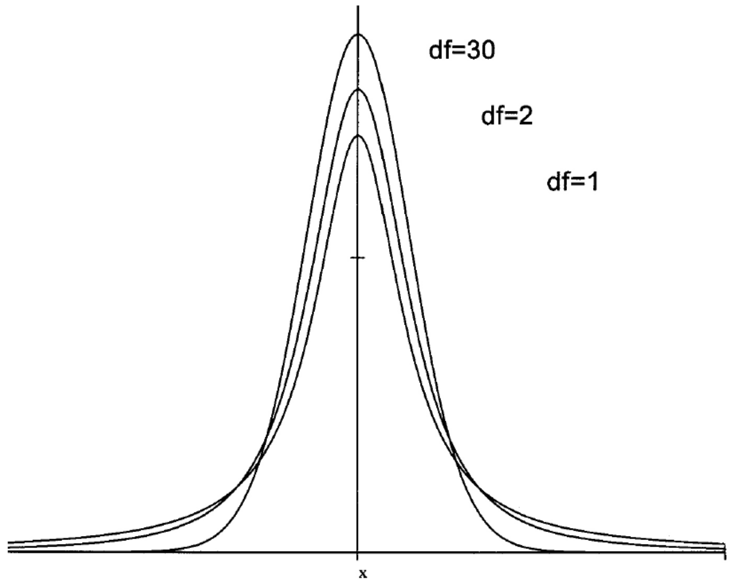

Figure 3. t (Sawilowsky & Fahoome, 2003)

	General Form	$v = 5$
Range:	$-\infty \leq x \leq \infty$	$-\infty \leq x \leq \infty$
μ:	0, $v > 1$	0
Median:	0	0
σ^2:	$\frac{v}{v-2}, v > 2$	$\frac{5}{3} or 1.667$
γ_1:	0, $v > 3$	0
γ_2:	$\frac{3(v-2)}{v-4}, v > 4$	9

Correlation

In determining the correlation between the pretest and posttest, algorithms presented by Headrick and Sawilowsky (2000) for creating correlated univariate and multivariate data for normal and nonnormal distributions will be used. This procedure is based on solving constants of the Fleishman (1978) power method. The algorithms will be used to populate matrices of correlated data for different types of distributions, presented in the table below (Normal, Chi-square df = 1, Chi-square df = 2, Chi-square df = 8, t (df = 3):

Table 7. Solutions To The Fleishman Equation For Selected Distributions.

Distribution	γ_1	γ_2	a	b	d
Normal	0	0	0	1	0
Chi-square (df =1)	$\sqrt{8}$	12	-.5207	.6146	.02007
Chi-square (df =2) (Exponential)	2	6	.3137	.8263	.02271
Chi-square (df =8)	1	1.5	-.1632	.9531	.0060
t (df = 3)	0	17	0	.3938	.1713

By doing this, it creates X and Y variables from the distributions and in turn controls skew and kurtosis.

The Headrick and Sawilowsky (2000) method for creating algorithms for correlated univariate and multivariate data are presented in Sawilowsky and Fahoome (2003). The first step in the procedure is to solve the Fleishman equation for the constants a, b, -a, and d. After the constants are identified, r can be found using the formula:

$$r_{xy} = r^2(b^2 + 6bd + 9d^2 + 2a^2r^2 + 6d^2r^4)$$ (p.300)

where $r_{xy} = 0$, .70, .80, and .90. This equation will be solved using a TI-83 calculator graphing functionality. It should be noted that the correlation of 0 will only be computed for the normal distribution.

Next, the r values obtained will be used to create intermediate standard normal variates, as shown in Table 8. Using three standard normal z scores (z_1, z_2, z_3) from normb1.f90 in Rangen 2.0, the intermediate standard normal variates will be computed using the formula below (Sawilowsky and Fahoome, 2003):

$$x_i = rz_1 + (\sqrt{1-r^2})z^2$$ (p. 301)

$$y_i = rz_1 + (\sqrt{1-r^2})z^3$$ (p. 302)

After the x_i and y_i are found, they are inputted into the Fleishman equations, along with the constants $a, b, -a,$ and d .

$$X = a + bx_i + (-a)x_i^2 + dx_i^2 \text{ (p. 302)}$$

$$Y = a + by_i + (-a)y_i^2 + dy_i^2 \text{ (p. 302)}$$

Table 8. Intermediate r Values for Various Distributions at Correlations 0.70, 0.80, and 0.90

Distribution	Intermediate *r* Values at Correlations:		
	0.70	0.80	0.90
Chi-square (df = 1)	.88909	.92960	.96633
Chi-square (df = 2)	.85998	.91319	.95973
Chi-square (df = 8)	.84466	.90058	.95271
Normal	.83666	.89443	.94868
t (df = 3)	.86665	.91814	.96118

*From Smith (2009), p.612

Although the Headrick and Sawilowsky (1999) has more mathematical computations than the Fleishman (1978) procedure, its ability to control both skew (γ_1) and kurtosis (γ_2) allows the generation of multivariate correlated data, which is a major advantage.

Location Effect Size

A shift in location will be introduced to the scores in the intervention group, represented by, by adding a percentage of the mean which corresponds to the preferred effected size. The effect size for shift in location will be conducted by using Cohen's description of common effect sizes, which will include small (0.2), medium (0.5), large (0.8), very large (1.2), and huge (2.0) for alpha levels 0.05 and 0.01. For every instance, an effect size will be added to the location of the intervention sample.

CHAPTER 4

RESULTS

The tables presented below are the results of a Monte Carlo study that was conducted with one hundred thousand iterations via Compaq 6.6c Fortran 77. Forty-eight tables are presented comparing the power of the t-test, Wilcoxon Signed-Ranks test, and the rank difference test. The simulations were conducted for the samples sizes (10, 10), (15, 15), and (20, 20); and for correlation values of .7, .8, and .9. Within each of these simulations, effect sizes of 0, .2, .5, .8, 1.2, and 2.0 were simulated. All tests were conducted at the 0.05 and 0.01 nominal alpha levels. The correlation value of 0 was also studied for the normal distribution, as discussed in Chapter 5.

Tables 9 through 24 are the results of the simulation ran for the Chi-squared df = 1, Chi-squared df = 2, Chi-squared df = 8, Normal, and, t (df = 3) distributions with a sample size of (10,10). All distributions were ran at correlations of .7,.8, and .9 and at 0 for the normal distribution, as discussed in Chapter 5. The simulation was comparing the relative power of the t-test, Wilcoxon Signed-Ranks test, and the rank difference test.

Table 9. (n1=n2)= (10,10), Chi-squared df = 1, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0492	0.0474	0.0474
	.01	0.0097	0.0055	0.0093
ES (d) = .2	.05	0.1797	0.1732	0.1732
	.01	0.0528	0.0323	0.0493
ES (d) = .5	.05	0.7298	0.7109	0.7110
	.01	0.4176	0.2950	0.3883
ES (d) = .8	.05	0.9825	0.9778	0.9778
	.01	0.8793	0.7545	0.8451
ES (d) = 1.2	.05	1.0000	0.9999	0.9999
	.01	0.9983	0.9862	0.9960
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 10. (n1=n2)= (10,10), Chi-squared df = 1, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0492	0.0480	0.0480
	.01	0.0097	0.0055	0.0095
ES (d) = .2	.05	0.2730	0.2642	0.2643
	.01	0.0938	0.0588	0.0874
ES (d) = .5	.05	0.9186	0.9056	0.9056
	.01	0.6917	0.5397	0.6514
ES (d) = .8	.05	0.9995	0.9992	0.9992
	.01	0.9878	0.9481	0.9778
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	0.9997	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 11. (n1=n2)= (10,10), Chi-squared df = 1, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0496	0.0484	0.0484
	.01	0.0099	0.0057	0.0097
ES (d) = .2	.05	0.5378	0.5179	0.5179
	.01	0.2479	0.1654	0.2322
ES (d) = .5	.05	0.9988	0.9979	0.9979
	.01	0.9784	0.9247	0.9650
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	0.9998	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 12. (n1=n2)= (10,10), Chi-squared df = 2, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0496	0.0479	0.0479
	.01	0.0099	0.0059	0.0097
ES (d) = .2	.05	0.1597	0.1541	0.1541
	.01	0.0459	0.0276	0.0429
ES (d) = .5	.05	0.6629	0.6429	0.6429
	.01	0.3491	0.2418	0.3254
ES (d) = .8	.05	0.9662	0.9579	0.9579
	.01	0.8130	0.6739	0.7760
ES (d) = 1.2	.05	0.9999	0.9997	0.9997
	.01	0.9945	0.9690	0.9885
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 13. (n1=n2)= (10,10), Chi-squared df = 2, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0502	0.0485	0.0485
	.01	0.0100	0.0054	0.0095
ES (d) = .2	.05	0.2361	0.2284	0.2285
	.01	0.0769	0.0478	0.0720
ES (d) = .5	.05	0.8715	0.8551	0.8551
	.01	0.6028	0.4543	0.5643
ES (d) = .8	.05	0.9984	0.9974	0.9974
	.01	0.9707	0.9068	0.9540
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	0.9989	0.9998
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 14. (n1=n2)= (10,10), Chi-squared df = 2, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0503	0.0495	0.0496
	.01	0.0097	0.0053	0.0093
ES (d) = .2	.05	0.4723	0.4569	0.4570
	.01	0.2042	0.1337	0.1901
ES (d) = .5	.05	0.9966	0.9946	0.9946
	.01	0.9547	0.8728	0.9333
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	0.9991	0.9998
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 15. (n1=n2)= (10,10), Chi-squared df = 8, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0507	0.0487	0.0487
	.01	0.0102	0.0060	0.0100
ES (d) = .2	.05	0.1522	0.1463	0.1463
	.01	0.0435	0.0262	0.0412
ES (d) = .5	.05	0.6328	0.6121	0.6122
	.01	0.3216	0.2210	0.2994
ES (d) = .8	.05	0.9552	0.9460	0.9461
	.01	0.7803	0.6337	0.7407
ES (d) = 1.2	.05	0.9997	0.9993	0.9993
	.01	0.9906	0.9572	0.9829
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 16. (n1=n2)= (10,10), Chi-squared df = 8, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0509	0.0493	0.0493
	.01	0.0098	0.0057	0.0096
ES (d) = .2	.05	0.2210	0.2133	0.2134
	.01	0.0703	0.0440	0.0668
ES (d) = .5	.05	0.8359	0.8199	0.8199
	.01	0.5482	0.4051	0.5119
ES (d) = .8	.05	0.9965	0.9947	0.9947
	.01	0.9528	0.8712	0.9319
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	0.9999	0.9974	0.9995
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 17. (n1=n2)= (10,10), Chi-squared df = 8, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0504	0.0487	0.0487
	.01	0.0095	0.0056	0.0096
ES (d) = .2	.05	0.4242	0.4087	0.4087
	.01	0.1719	0.1113	0.1601
ES (d) = .5	.05	0.9918	0.9889	0.9889
	.01	0.9224	0.8172	0.8946
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	0.9998	0.9972	0.9995
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 18. (n1=n2)= (10,10), Normal, $r_{xy} = 0$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0484	0.0474	0.0474
	.01	0.0099	0.0059	0.0096
ES (d) = .2	.05	0.0680	0.0660	0.0661
	.01	0.0160	0.0093	0.0152
ES (d) = .5	.05	0.1704	0.1648	0.1648
	.01	0.0492	0.0298	0.0467
ES (d) = .8	.05	0.3570	0.3449	0.3449
	.01	0.1368	0.0878	0.1276
ES (d) = 1.2	.05	0.6678	0.6489	0.6489
	.01	0.3559	0.2447	0.3302
ES (d) = 2.0	.05	0.9774	0.9706	0.9706
	.01	0.8515	0.7182	0.8167

Table 19. (n1=n2)= (10,10), Normal, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0496	0.0482	0.0482
	.01	0.0097	0.0058	0.0096
ES (d) = .2	.05	0.1134	0.1094	0.1094
	.01	0.0292	0.0177	0.0281
ES (d) = .5	.05	0.4457	0.4298	0.4298
	.01	0.1846	0.1201	0.1717
ES (d) = .8	.05	0.8256	0.8081	0.8081
	.01	0.5355	0.3930	0.5012
ES (d) = 1.2	.05	0.9907	0.9876	0.9876
	.01	0.9155	0.8077	0.8873
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	0.9999	0.9982	0.9996

Table 20. (n1=n2)= (10,10), Normal, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0487	0.0474	0.0474
	.01	0.0097	0.0055	0.0094
ES (d) = .2	.05	0.1467	0.1430	0.1431
	.01	0.0415	0.0250	0.0391
ES (d) = .5	.05	0.6072	0.5880	0.5881
	.01	0.3020	0.2032	0.2787
ES (d) = .8	.05	0.9436	0.9322	0.9322
	.01	0.7486	0.6003	0.7111
ES (d) = 1.2	.05	0.9995	0.9991	0.9991
	.01	0.9869	0.9451	0.9767
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 21. (n1=n2)= (10,10), Normal, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0499	0.0488	0.0489
	.01	0.0100	0.0060	0.0097
ES (d) = .2	.05	0.2431	0.2350	0.2351
	.01	0.0799	0.0497	0.0742
ES (d) = .5	.05	0.8809	0.8655	0.8655
	.01	0.6217	0.4712	0.5835
ES (d) = .8	.05	0.9987	0.9979	0.9979
	.01	0.9756	0.9159	0.9616
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	0.9991	0.9999
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 22. (n1=n2)= (10,10), t (df = 3), $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0494	0.0483	0.0483
	.01	0.0100	0.0058	0.0098
ES (d) = .2	.05	0.1645	0.1580	0.1580
	.01	0.0475	0.0284	0.0450
ES (d) = .5	.05	0.6789	0.6605	0.6606
	.01	0.3631	0.2525	0.3377
ES (d) = .8	.05	0.9706	0.9632	0.9632
	.01	0.8275	0.6894	0.7906
ES (d) = 1.2	.05	0.9999	0.9998	0.9998
	.01	0.9955	0.9738	0.9907
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 23. (n1=n2)= (10,10), t (df = 3), $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0488	0.0480	0.0480
	.01	0.0091	0.0055	0.0091
ES (d) = .2	.05	0.2474	0.2393	0.2394
	.01	0.0812	0.0511	0.0769
ES (d) = .5	.05	0.8844	0.8694	0.8695
	.01	0.6267	0.4770	0.5887
ES (d) = .8	.05	0.9990	0.9983	0.9983
	.01	0.9777	0.9216	0.9637
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	0.9992	0.9999
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 24. (n1=n2)= (10,10), t (df = 3), $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0502	0.0488	0.0488
	.01	0.0102	0.0062	0.0102
ES (d) = .2	.05	0.4856	0.4692	0.4692
	.01	0.2124	0.1396	0.1977
ES (d) = .5	.05	0.9972	0.9954	0.9954
	.01	0.9599	0.8835	0.9401
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	0.9992	0.9999
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Tables 25 through 40 include the results of the simulation ran for the Chi-squared df = 1, Chi-squared df = 2, Chi-squared df = 8, Normal, and, t (df = 3) distributions with a sample size of (15,15). All distributions were ran with correlations of .7,.8, and .9. Again, the simulation was comparing the relative power of the t-test, Wilcoxon Signed-Ranks test, and the rank difference test.

Table 25. (n1=n2) = (15,15), Chi-squared df = 1, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0493	0.0471	0.0471
	.01	0.0097	0.0065	0.0080
ES (d) = .2	.05	0.2575	0.2456	0.2458
	.01	0.0929	0.0687	0.0798
ES (d) = .5	.05	0.9078	0.8922	0.8923
	.01	0.7063	0.6214	0.6576
ES (d) = .8	.05	0.9996	0.9990	0.9990
	.01	0.9912	0.9803	0.9851
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 26. (n1=n2) = (15,15), Chi-squared df = 1, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0499	0.0479	0.0480
	.01	0.0096	0.0063	0.0080
ES (d) = .2	.05	0.4039	0.3839	0.3842
	.01	0.1727	0.1307	0.1495
ES (d) = .5	.05	0.9900	0.9868	0.9869
	.01	0.9312	0.8864	0.9062
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	0.9997	0.9998
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 27. (n1=n2) = (15,15), Chi-squared df = 1, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0495	0.0474	0.0474
	.01	0.0108	0.0072	0.0089
ES (d) = .2	.05	0.7442	0.7209	0.7212
	.01	0.4636	0.3827	0.4178
ES (d) = .5	.05	1.0000	1.0000	1.0000
	.01	0.9998	0.9991	0.9995
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 28. (n1=n2) = (15,15), Chi-squared df = 2, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0503	0.0473	0.0474
	.01	0.0100	0.0065	0.0082
ES (d) = .2	.05	0.2312	0.2191	0.2193
	.01	0.0799	0.0580	0.0679
ES (d) = .5	.05	0.8591	0.8413	0.8414
	.01	0.6209	0.5340	0.5720
ES (d) = .8	.05	0.9982	0.9971	0.9971
	.01	0.9780	0.9568	0.9661
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	0.9999	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 29. (n1=n2) = (15,15), Chi-squared df = 2, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0496	0.0473	0.0474
	.01	0.0095	0.0064	0.0079
ES (d) = .2	.05	0.3571	0.3388	0.3390
	.01	0.1437	0.1068	0.1241
ES (d) = .5	.05	0.9775	0.9712	0.9712
	.01	0.8793	0.8188	0.8447
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	0.9996	0.9985	0.9990
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 30. (n1=n2) = (15,15), Chi-squared df = 2, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0494	0.0473	0.0474
	.01	0.0098	0.0067	0.0084
ES (d) = .2	.05	0.6762	0.6528	0.6533
	.01	0.3911	0.3181	0.3506
ES (d) = .5	.05	1.0000	0.9999	0.9999
	.01	0.9990	0.9970	0.9979
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 31. (n1=n2) = (15,15), Chi-squared df = 8, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = .0	.05	0.0494	0.0472	0.0473
	.01	0.0098	0.0066	0.0084
ES (d) = .2	.05	0.2189	0.2065	0.2066
	.01	0.0735	0.0534	0.0629
ES (d) = .5	.05	0.8365	0.8153	0.8154
	.01	0.5839	0.4976	0.5346
ES (d) = .8	.05	0.9967	0.9953	0.9953
	.01	0.9678	0.9407	0.9523
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	0.9998	0.9999
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 32. (n1=n2) = (15,15), Chi-squared df = 8, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0488	0.0465	0.0465
	.01	0.0099	0.0066	0.0082
ES (d) = .2	.05	0.3250	0.3085	0.3087
	.01	0.1245	0.0924	0.1074
ES (d) = .5	.05	0.9639	0.9557	0.9558
	.01	0.8376	0.7692	0.7988
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	0.9990	0.9967	0.9976
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 33. (n1=n2) = (15,15), Chi-squared df = 8, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0501	0.0479	0.0480
	.01	0.0099	0.0067	0.0086
ES (d) = .2	.05	0.6153	0.5923	0.5926
	.01	0.3314	0.2650	0.2948
ES (d) = .5	.05	0.9998	0.9997	0.9997
	.01	0.9965	0.9910	0.9935
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 34. (n1=n2) = (15,15), Normal, $r_{xy} = 0$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0495	0.0480	0.0480
	.01	0.0097	0.0064	0.0079
ES (d) = .2	.05	0.0808	0.0763	0.0763
	.01	0.0200	0.0136	0.0165
ES (d) = .5	.05	0.2482	0.2351	0.2352
	.01	0.0881	0.0641	0.0750
ES (d) = .8	.05	0.5299	0.5076	0.5077
	.01	0.2608	0.2033	0.2291
ES (d) = 1.2	.05	0.8636	0.8455	0.8455
	.01	0.6265	0.5396	0.5772
ES (d) = 2.0	.05	0.9991	0.9986	0.9986
	.01	0.9864	0.9721	0.9788

Table 35. (n1=n2) = (15,15), Normal, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0494	0.0468	0.0468
	.01	0.0095	0.0064	0.0080
ES (d) = .2	.05	0.1551	0.1456	0.1457
	.01	0.0460	0.0327	0.0390
ES (d) = .5	.05	0.6438	0.6182	0.6183
	.01	0.3566	0.2843	0.3156
ES (d) = .8	.05	0.9605	0.9508	0.9508
	.01	0.8264	0.7547	0.7852
ES (d) = 1.2	.05	0.9998	0.9997	0.9997
	.01	0.9961	0.9903	0.9930
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 36. (n1=n2) = (15,15), Normal, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0491	0.0469	0.0469
	.01	0.0095	0.0063	0.0079
ES (d) = .2	.05	0.2067	0.1957	0.1958
	.01	0.0691	0.0501	0.0587
ES (d) = .5	.05	0.8112	0.7896	0.7897
	.01	0.5477	0.4611	0.4990
ES (d) = .8	.05	0.9951	0.9932	0.9932
	.01	0.9576	0.9250	0.9391
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	0.9999	0.9996	0.9997
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 37. (n1=n2) = (15,15), Normal, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0497	0.0479	0.0480
	.01	0.0103	0.0070	0.0086
ES (d) = .2	.05	0.3648	0.3462	0.3466
	.01	0.1502	0.1126	0.1304
ES (d) = .5	.05	0.9805	0.9741	0.9741
	.01	0.8893	0.8324	0.8565
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	0.9998	0.9990	0.9994
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 38. (n1=n2) = (15,15), t (df = 3), $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0501	0.0475	0.0475
	.01	0.0095	0.0061	0.0076
ES (d) = .2	.05	0.2365	0.2250	0.2251
	.01	0.0815	0.0597	0.0701
ES (d) = .5	.05	0.8705	0.8520	0.8521
	.01	0.6397	0.5529	0.5904
ES (d) = .8	.05	0.9986	0.9978	0.9978
	.01	0.9816	0.9630	0.9713
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 39. (n1=n2) = (15,15), t (df = 3), $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0495	0.0476	0.0477
	.01	0.0095	0.0061	0.0078
ES (d) = .2	.05	0.3714	0.3528	0.3531
	.01	0.1522	0.1156	0.1324
ES (d) = .5	.05	0.9818	0.9763	0.9764
	.01	0.8953	0.8398	0.8638
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	0.9998	0.9992	0.9995
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 40. (n1=n2) = (15,15), t (df = 3), $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0505	0.0483	0.0484
	.01	0.0103	0.0069	0.0087
ES (d) = .2	.05	0.6915	0.6673	0.6678
	.01	0.4059	0.3309	0.3640
ES (d) = .5	.05	1.0000	1.0000	1.0000
	.01	0.9992	0.9976	0.9984
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Tables 41 through 56 include the results of the simulation being ran for the Chi-squared df = 1, Chi-squared df = 2, Chi-squared df = 8, Normal, and, t (df = 3) distributions with a sample size of (20,20). All distributions were ran with correlations of .7,.8, and .9. Again, the simulation was comparing the relative power of the t-test, Wilcoxon Signed-Ranks test, and the rank difference test.

Table 41. (n1=n2)= (20,20), Chi-squared df = 1, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0501	0.0490	0.0490
	.01	0.0100	0.0083	0.0094
ES (d) = .2	.05	0.3448	0.3281	0.3281
	.01	0.1416	0.1216	0.1319
ES (d) = .5	.05	0.9721	0.9652	0.9652
	.01	0.8732	0.8409	0.8535
ES (d) = .8	.05	1.0000	0.9999	0.9999
	.01	0.9996	0.9991	0.9993
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 42. (n1=n2)= (20,20), Chi-squared df = 1, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0493	0.0480	0.0480
	.01	0.0100	0.0081	0.0095
ES (d) = .2	.05	0.5259	0.5035	0.5035
	.01	0.2655	0.2328	0.2486
ES (d) = .5	.05	0.9989	0.9983	0.9983
	.01	0.9882	0.9818	0.9842
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 43. (n1=n2)= (20,20), Chi-squared df = 1, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0500	0.0484	0.0484
	.01	0.0100	0.0084	0.0096
ES (d) = .2	.05	0.8655	0.8485	0.8485
	.01	0.6472	0.5992	0.6193
ES (d) = .5	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 44. (n1=n2)= (20,20), Chi-squared df = 2, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0497	0.0481	0.0481
	.01	0.0101	0.0083	0.0094
ES (d) = .2	.05	0.2990	0.2851	0.2851
	.01	0.1159	0.0985	0.1072
ES (d) = .5	.05	0.9477	0.9368	0.9368
	.01	0.8057	0.7659	0.7824
ES (d) = .8	.05	0.9999	0.9999	0.9999
	.01	0.9983	0.9968	0.9973
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 45. (n1=n2)= (20,20), Chi-squared df = 2, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0500	0.0488	0.0489
	.01	0.0101	0.0082	0.0095
ES (d) = .2	.05	0.4640	0.4443	0.4443
	.01	0.2187	0.1891	0.2032
ES (d) = .5	.05	0.9963	0.9949	0.9949
	.01	0.9712	0.9590	0.9634
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 46. (n1=n2)= (20,20), Chi-squared df = 2, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0498	0.0489	0.0489
	.01	0.0098	0.0081	0.0093
ES (d) = .2	.05	0.8120	0.7913	0.7913
	.01	0.5636	0.5158	0.5361
ES (d) = .5	.05	1.0000	1.0000	1.0000
	.01	1.0000	0.9999	0.9999
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 47. (n1=n2)= (20,20), Chi-squared df = 8, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0494	0.0474	0.0474
	.01	0.0095	0.0077	0.0088
ES (d) = .2	.05	0.2834	0.2701	0.2701
	.01	0.1070	0.0914	0.0997
ES (d) = .5	.05	0.9327	0.9206	0.9206
	.01	0.7723	0.7318	0.7486
ES (d) = .8	.05	0.9998	0.9996	0.9996
	.01	0.9968	0.9944	0.9953
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 48. (n1=n2)= (20,20), Chi-squared df = 8, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0508	0.0493	0.0493
	.01	0.0100	0.0086	0.0097
ES (d) = .2	.05	0.4247	0.4066	0.4066
	.01	0.1919	0.1682	0.1802
ES (d) = .5	.05	0.9930	0.9906	0.9906
	.01	0.9536	0.9354	0.9422
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 49. (n1=n2)= (20,20), Chi-squared df = 8, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0496	0.0486	0.0486
	.01	0.0099	0.0083	0.0094
ES (d) = .2	.05	0.7547	0.7323	0.7323
	.01	0.4907	0.4443	0.4642
ES (d) = .5	.05	1.0000	1.0000	1.0000
	.01	0.9999	0.9998	0.9998
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 50. (n1=n2)= (20,20), Normal, $r_{xy} = 0$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0497	0.0485	0.0485
	.01	0.0094	0.0076	0.0087
ES (d) = .2	.05	0.0911	0.0870	0.0870
	.01	0.0234	0.0190	0.0215
ES (d) = .5	.05	0.3225	0.3059	0.3059
	.01	0.1288	0.1114	0.1205
ES (d) = .8	.05	0.6700	0.6464	0.6464
	.01	0.3968	0.3571	0.3747
ES (d) = 1.2	.05	0.9491	0.9389	0.9389
	.01	0.8117	0.7724	0.7880
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	0.9994	0.9986	0.9989

Table 51. (n1=n2)= (20,20), Normal, $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0503	0.0490	0.0490
	.01	0.0101	0.0084	0.0095
ES (d) = .2	.05	0.1954	0.1861	0.1860
	.01	0.0663	0.0564	0.0617
ES (d) = .5	.05	0.7807	0.7609	0.7609
	.01	0.5245	0.4760	0.4971
ES (d) = .8	.05	0.9920	0.9892	0.9892
	.01	0.9472	0.9287	0.9360
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	0.9999	0.9998	0.9998
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 52. (n1=n2)= (20,20), Normal, $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0498	0.0484	0.0485
	.01	0.0108	0.0084	0.0097
ES (d) = .2	.05	0.2686	0.2555	0.2555
	.01	0.0990	0.0851	0.0919
ES (d) = .5	.05	0.9176	0.9045	0.9045
	.01	0.7408	0.6966	0.7146
ES (d) = .8	.05	0.9996	0.9993	0.9993
	.01	0.9945	0.9910	0.9922
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 53. (n1=n2)= (20,20), Normal, $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0497	0.0492	0.0492
	.01	0.0099	0.0085	0.0098
ES (d) = .2	.05	0.4755	0.4554	0.4554
	.01	0.2268	0.1980	0.2117
ES (d) = .5	.05	0.9975	0.9964	0.9964
	.01	0.9758	0.9641	0.9684
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 54. (n1=n2)= (20,20), t (df = 3), $r_{xy} = .7$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0502	0.0488	0.0488
	.01	0.0100	0.0082	0.0094
ES (d) = .2	.05	0.3105	0.2968	0.2968
	.01	0.1220	0.1043	0.1136
ES (d) = .5	.05	0.9531	0.9437	0.9437
	.01	0.8200	0.7815	0.7968
ES (d) = .8	.05	0.9999	0.9998	0.9998
	.01	0.9986	0.9974	0.9979
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 55. (n1=n2)= (20,20), t (df = 3), $r_{xy} = .8$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0500	0.0489	0.0489
	.01	0.0101	0.0086	0.0097
ES (d) = .2	.05	0.4808	0.4594	0.4594
	.01	0.2314	0.2026	0.2167
ES (d) = .5	.05	0.9977	0.9964	0.9964
	.01	0.9782	0.9672	0.9713
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Table 56. (n1=n2)= (20,20), t (df = 3), $r_{xy} = .9$

Effect Size	α level	t	Wilcoxon Signed-Ranks Test	Rank Difference Test
ES (d) = 0	.05	0.0504	0.0491	0.0491
	.01	9.7999	0.0080	0.0092
ES (d) = .2	.05	0.8247	0.8044	0.8044
	.01	0.5818	0.5344	0.5552
ES (d) = .5	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = .8	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 1.2	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000
ES (d) = 2.0	.05	1.0000	1.0000	1.0000
	.01	1.0000	1.0000	1.0000

Within Tables 9 through 56, the critical values for the t-test, Wilcoxon Signed-Ranks test, and the rank difference test are shown at varying effect sizes and alpha levels of .05 and .01. The results show that under all distributions, sample sizes, and correlations, the t-test had larger critical values than the Wilcoxon Signed-Ranks test and the rank difference test, which makes the t-test appear to have power advantages over both tests, even under non-normal distributions. The explanation for why this occurred under the non-normal distributions is discussed in Chapter 5.

The focus of the dissertation was to investigate Kornbrot's claim of superiority of the rank difference test over the Wilcoxon Signed-Ranks test. In comparing the Wilcoxon Signed-Ranks test and the rank difference test in tables 9 through 55, the data shows that there is relatively no difference in power between the two. The critical values of both tests in many cases are identical, or are fairly close in value. Although, it should be noted, that the critical values for the rank difference test were larger than the Wilcoxon Signed-Ranks test when the effect size was small and the alpha level was .01, thus appearing that the rank difference test was outperforming the Wilcoxon Signed-Ranks test. Again, the reasoning for this occurrence is discussed in Chapter 5.

CHAPTER 5

DISCUSSION

The purpose of this study was to compare the power of the Wilcoxon Signed-Ranks test in comparison to the rank difference test when the assumption of normality is not met, the data are ordinal, and the sample size is small. Furthermore, the intention of the study was to test Kornbrot's claim that the rank difference test was superior to the Wilcoxon Signed-Ranks test. Kornbrot (1990) argued that the rank difference test was better suited than the Wilcoxon Signed-Ranks test when dealing with paired comparison designs and the data are not of the interval scale type. The results do not support Kornbrot's claim, in actuality the critical values of the Wilcoxon Signed Ranks-test and the rank difference test were fairly similar. There was not a statistically significant difference between the two tests and the differences that were exhibited will be explained.

In order to test Kornbrot's claim, a Monte Carlo study was conducted. The study was conducted on three equal samples of size combinations (n_1, n_2) = (10,10), (15,15), (20,20), with nominal alpha levels of 0.01 and 0.05. There were one hundred thousand repetitions per experiment. A Compaq 6.6c Fortran 77 program was written to compare the Type I error rate and power of the t-test, Wilcoxon Signed-Ranks test, and the rank difference test under the conditions of normal and non-normally distributed data sets, including the normal, chi-square, df=1, chi-square, df=2, chi-square, df=8, and the t, df=3 distribution. The data were generated using a pseudo-random number generator via IMSL subroutines.

In order to alter the correlation between the pretest and posttest, algorithms presented by Headrick and Sawilowsky (2000) for creating correlated univariate and multivariate data for normal and nonnormal distributions were used. The algorithms were used to fill the matrices of correlated data for different types of distributions, (Normal, Chi-square df = 1, Chi-square df = 2, Chi-square df = 8, t (df = 3)).

A shift in location was introduced to the scores in the intervention group, represented by, by adding a percentage of the mean, which corresponds to the preferred effected size. The effect size for shift in location was conducted by using Cohen's description of common effect sizes, which will included small (0.2), medium (0.5), large (0.8), very large (1.2), and huge (2.0) for alpha levels 0.05 and 0.01. In each situation, an effect size was added to the location of the intervention sample.

Explanation of Results

As expected, when the correlation value was equal to 0 for the normal distribution, all three tests demonstrated low power. Of course, when setting the correlation to 0, independent samples tests should be used, but this was done only to confirm the veracity of the Fortran coding, and hence, was only necessary for the initial run on the data drawn from the normal distribution.

In all five distributions (Gaussian (normal) distribution, chi-square, df = 1, chi-square, df = 2, chi-square, df =8, and t distribution, df = 3), when the correlation increased from .7 to .8 to .9 and the effect size remained stable, the tests' rejection rates increased, as expected. The values converged to 1 as the correlation increased.

The uniformly superior results from the t-test were not as expected, but can be explained. The t-test should not be more powerful than the Wilcoxon Signed-Ranks test under the non-normal distributions that are not symmetric with light tails. However, the greatly superior results of the nonparametric tests for shift in location treatment alternatives for such nonnormally distributed data (see Sawilowsky and Blair, 1992) occur in the two independent samples layout. These results, however, do not generalize to the paired samples case (Sawilowsky, 1990).

The reason is because the t-test is not operating on data obtained from the chi squared distribution per se. Instead, it is applied to the difference distribution of the posttest score (which is distributed chi squared) and the pretest score (which is distributed chi squared). The differences, however, are a normalizing procedure, and the resulting differences are shaped more like the uniform distribution.

The uniform distribution is symmetric and has light tails. Hence, the tremendous power advantage of the nonparametric alternatives to the parametric test in the two independent samples layout vanished. Of course, the nonparametric tests should nevertheless demonstrate some minimal power advantages over the dependent sample's t-test, and yet those results were not obtained.

This can be explained, however, by the fact that the t-test is being conducted at precisely the 0.05 and 0.01 alpha levels, whereas both of the non-parametric tests are conducted at reduced alpha levels. As explained by Gibbons & Chakraborti (1991), a nonparametric test is based on the sampling distribution of discrete variables, thus constraining the possible significance levels. For example, for a two tailed test (alpha equals .05), the critical value for sample size 5 is zero as is the critical value for sample size 6. Hence, the Wilcoxon Signed-Ranks test and the rank difference test were starting off at a disadvantage. Having differing significance levels heavily limits the ability to compare the power functions of the tests.

Gibbons & Chakraborti (1991) recommended, therefore, to set the alpha level of the t-test to match the limitations of the nonparametric tests to obtain a fair comparison. Sawilowsky (personal communications), however, argued against this approach, because in practice, if a worker has selected an 0.05 or 0.01, that standard should not be modified by this limitation of the statistical tests. This suggestion was reasonable during the time period when statistical tests were conducted via obtaining critical values from tabled valued. However, today, this debate should be revisited in consideration of the ability of statistical software to compute tests at any given nominal alpha level.

Kornbrot's claim that the rank difference test is superior to the Wilcoxon Signed-Ranks test could not be supported through the results. The results for the two tests did not show that the rank difference test is more powerful than the Wilcoxon Signed-Ranks test. In fact, they are rather fairly equivalent. Although, it is difficult to compare the two tests as well, because their nominal alpha levels were also not always equal. During the simulation, the alpha levels could have been set at .046 for the Wilcoxon Signed-Ranks test and at .048 for the rank difference test. Again, this would cause the Wilcoxon Signed-Ranks test to be at a disadvantage, as a result in the differing alpha levels (Gibbons & Chakraborti, 1991). This makes it difficult to directly compare the tests. The critical values for the two tests were almost identical, even though the numbers within the tables are slightly different. There is no evidence within the results that Kornbrot's rank difference test is more powerful than the Wilcoxon Signed-Ranks test.

Recommendations

The fact that the Wilcoxon Signed-Ranks test and the rank difference test cannot take on alpha levels of .05 is a legitimate drawback to the two tests. However, for future research, it is recommended that the alpha value for the t-test be set to the same level of the competitor. For example, the alternative test can be conducted at .046, the t-test should also be set to nominal alpha at .046 so that direct comparisons can be made between the tests. This can be done by merely changing the critical values of the t-test. It is noted that .046 is not a typical alpha level, i.e. not used in textbook or research, but this will allow for direct comparison between the tests.

It is further recommended that the Wilcoxon Signed-Ranks test and the rank difference test be compared further under the same as well as different distributions, with more than one hundred thousand simulations, as presented within this study. This will help further validate the claims within this research that there is no difference in power between the Wilcoxon Signed-Ranks test and the rank difference test.

Lastly, it is recommended that college introductory statistics courses discuss the use of nonparametric statistics when the data does not meet the assumption of normality. These courses are usually limited to the parametric statistics, such as the t-test and analysis of variance. It is important for future researchers to understand under what conditions it is best to use parametric and nonparametric statistics.

REFERENCES

Blair, R. C. (1981). A reaction to ‘Consequences of failure to meet Assumptions underlying the fixed effects analysis of variance and covariance.’ *Review of Educational Research*, 51(4), 499-507.

Blair, R. C., & Higgins, J. J. (1981). A note on the asymptotic relative efficiency of the Wilcoxon rank-sum test relative to the independent means t-test under mixtures of two normal distributions. *British Journal of Mathematical and Statistical Psychology*, 34, 124-128.

Blair, R. C., & Higgins, J. J. (1985). Comparison of the power of the Paired Samples t test to that of Wilcoxon’s Signed-Ranks test under various population shapes. *Psychological Bulletin,* 97 (1), 119-128.

Boneau, C. A. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, 57, 49-64.

Bradley, J.V. (1968). *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice Hall.

Bradley, J.V. (1978). Robustness? *British Journal of Mathematical & Statistical Psychology*, 31, 144-152.

Bradley, J. V. (1977). A common situation conducive to bizarre distribution shapes. *The American Statistician, 31,* 147-150.

Christie, J.F. & Enz, B. (1992). The effects of literacy play interventions on preschoolers' play patters and literacy development. *Early Education & Development,* 3 (3), 205-220.

Conover, W. J. (1980). *Practical Nonparametric Statistics*. New York: John Wiley & Sons.

Crichton N.J. (1998). Statistical considerations in design and analysis. In: Research and development in clinical nursing practice. (Roe B. & Webb C., eds). Whurr, London. p. 209.

Dunn-Rankin, P. (1983). *Scaling Methods*. Hillsdale: Lawrence Erlbaum Associates.

Elashoff, J., & Elashoff,. (1978). Ef*fects of errors in statistical assumptions. In W. Kruskal & J. Tanur* (Eds.), International encyclopedia of statistics (pp. 229-250). New York: Free Press.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43, 521-532.

Friedman, Milton 1937 The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association* 32, 675-701.

Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin*, *87*, 564-567.

Gibbon, J. (1971). Scalar timing and semi-Markov chains. in free-operant avoidance. *Journal of Mathematical Psychology, 8,* 109–138.

Gibbon, J. & Chakraborti, S. (1991). Comparisons of the Mann-Whitney, Student's t, and Alternate t Tests for means of normal distribution. *The Journal of Experimental Education* 59(3), 258-267

Goldstein, R. (1996). *Stata Technical Bulletin*. South Salem, New York: StataCorp.

Hair J.F., Anderson, R.E., Tatham R.L., & Black W.C. (1998). *Multivariate Data Analysis* (5th edn.). Upper Saddle River, NJ: Prentice-Hall.

Harwell, M.R. (1990). Summarizing Monte Carlo results in methodological research. Paper presented at the annual meeting of the American Educational Association, Boston.

Headrick, T. C., & Sawilowsky, S. S. (2000). Weighted simplex procedures for determining boundary points and constants for the univariate and multivariate power methods. *Journal of Educational and Behavioral Statistics*, 25, 417-436.

Headrick, T. C., & Sawilowsky, S. S. (1999). Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method. *Psychometrika*, 64, 25-35.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1998). *Applied Statistics for the Behavioral Sciences (4th ed.).* New York: Houghton Mifflin Company.

Hunter M.A. & May, R.B. (1993). Some myths concerning parametric and nonparametric tests. *Canadian Psychological, 34*(4), 384-389.

Keppel, Geoffrey (1991). *Design and Analysis: A researcher's handbook* (3rd edn.). Englewood Cliffs, NJ: Prentice-Hall.

Kerlinger, F.N. & Lee, H.B. (2000). *Foundations of Behavioral Research (4th ed.)*. New York: Harcourt College Publishers.

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

Kornbrot, D. E. (1990). The rank difference test: A new and meaningful alternative to the Wilcoxon signed ranks test for ordinal data. *British Journal of Mathematical and Statistical Psychology*, 43, 241-264.

Lehmann, E.L. (1975). *Nonparametrics, statistical methods based on ranks*. San Francisco: Holden Day.

Mahar, J. & Strobert, B. (2010). The use of 360-Degree Feedback compared to traditional evaluation feedback for the professional growth of teachers in K-12 education. *Planning and Changing,* Fall 2010, 4, 147-158.

Mcintosh, R., Vaughn, S., Schumm, J.S., Haager, D., Lee, O. (1993). Observations of students with learning disabilities in general education classrooms. *Exceptional Children,* 60 (3), 249-261.

Micceri, T. (1989). The unicorn, the normal curve and other improbable creatures. *Psychological Bulletin, 105*(1), 156-166.

Nanna, M. J., & Sawilowsky, S. S. (1998). Analysis of Likert scale data in disability and medical rehabilitation research. *Psychological Methods*, *3*, 55–67.

Noether, G.E. (1967). *Elements of Nonparametric Statistics.* New York: Wiley.

Nunnally, J. C. (1978). *Psychometric Theory.* New York: McGraw-Hill.

Pett, M.A. (1997). Nonparametric statistics for health care research: Statistics for small samples and unusual distributions. Thousand Oaks, CA: SAGE Publications.

Runyon, R.P. & Haber, A. (1991). *Fundamentals of Behavioral Statistics (7th ed.).* New York: McGraw Hill.

Sawilowsky, S.S. (1990). Nonparametric test of interaction in experimental design. *Review of Educational Research, 60*(1), 91-126.

Sawilowsky, S., Blair, R. C., & Micceri, T. (1990). A PC FORTRAN subroutine library of psychology and education data sets. *Psychometrika*, 55, 729.

Sawilowsky, S. & Blair, R. C. (1992). A more realistic look at the robustness and Type II error properties of the t-test to departures from population normality. *Psychological Bulletin*, 111(2), 352-360.

Sawilowsky, S. & Fahoome, G. (2003). *Statistics Through Monte Carlo Simulation with Fortran*. Oak Park: JMASM.

Siddiqui, M., & Raghunandanan, K. (1967). Asymptotically robust estimators of location. *Journal of the American Statistical Association*, 62, 950-953.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.) New York: McGraw-Hill.

Smith, J. (2009). Intermediate r values for use in the Fleishman power method. *Journal of Modern Applied Statistical Methods,* 8(2), 610-612.

Vogt, W. P. (2005). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences* (3rd ed.). Thousand Oaks, CA: Sage Publications.

ABSTRACT

ROBUSTNESS AND POWER OF THE KORNBROT RANK DIFFERENCE, SIGNED RANKS, AND DEPENDENT SAMPLES T-TEST

by

NORMAN N. HAIDOUS

December 2012

Advisor: Dr. Shlomo S. Sawilowsky

Major: Evaluation and Research - Statistics

Degree: Doctor of Philosophy

The purpose of the study was to compare the power and accuracy of the Wilcoxon Signed-Ranks test in comparison to the rank difference test when the assumption of normality is not met, the data are ordinal, and the sample size is small. The study also investigated Kornbrot's (1990) claim that the rank difference test should be used over the Wilcoxon Signed-Ranks tests "in all paired comparison designs where the data are not both of internal scale type and of known distribution time" (p. 258).

The study design allowed for the comparison of the comparative power and robustness of the Wilcoxon-Signed Ranks-Test, rank difference test, and the t-test. Utilizing five distributions (Normal, chi-square, df = 1, chi-square, df = 2, chi-square, df = 8, and t distribution, df = 3) and three sample sizes (10, 10), (15, 15), and (20, 20). All distributions and their data sets were created by using a pseudo-random number generator, via Monte Carlo simulation, using Fortran 77.

The results showed the t-test had power advantages over the Wilcoxon Signed-Ranks test and the rank differences, but with explanation. There was no evidence within the results to confirm that the rank difference test had substantial power advantages over Wilcoxon Signed-Ranks test.

AUTOBIOGRAPHICAL STATEMENT

Norman N. Haidous
ac6702@wayne.edu

Education

Doctor of Philosophy – Educational Evaluation and Research
Wayne State University, College of Education, Detroit, Michigan
Cognate: Business December 2012

Master of Science – Mechanical Engineering
Wayne State University, College of Engineering, Detroit, Michigan
December 1999

Bachelor of Science – Industrial Engineering
Purdue University, College of Engineering, West Lafayette, Indiana
May 1996

Professional Experience

Ford Motor Company, Dearborn Michigan
Product Development Design & Release Engineer

Wayne State University, Detroit Michigan
Adjunct Faculty, School of Interdisciplinary Studies / Liberal Arts

Certification

Certified Six-Sigma Black Belt
Certified CCE CAD Capable Engineering Program

Honors/Awards

United States of America Granted Patent
Molded Assembly Having a Reduced Tendency to Squeak and a Method of Manufacturing the Assembly
Patent ID: 811404411US

2009 Society of Plastics Engineers Finalist
"Most Innovative Use of Plastics"
Wheat-Filled PP for Quarter-Trim Bin 2010 MY Ford Flex CUV

NSF 1999 Outstanding Engineering Senior Design Project to Aid Persons with Disabilities
Communication Device Mount – Wayne State University

1st Place – Industrial Engineering Technical Essay, presented by Compaq

1st Place – Industrial Engineering Senior Design of an Information System

General Military Courses - Certificate of Completion