1-1-2013

# Comparative Power Of The Anova, Randomization Anova, And Kruskal-Wallis Test

Jamie Gleason
*Wayne State University,*

# COMPARATIVE POWER OF THE ANOVA, APPROXIMATE RANDOMIZATION ANOVA, AND KRUSKAL-WALLIS TEST

by

**JAMIE H. GLEASON**

**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**

2013

MAJOR: EVALUATION AND RESEARCH

Approved by:

_____

Advisor Date

_____

_____

_____

## DEDICATION

I dedicate this to my wonderful wife, Amanda, who somehow agreed it would be a great idea to move 2,400 miles across the country so her fiancé could begin this journey, and then proceeded to marry him. Not to be outdone, though, are the Buggels, Caoimhe and Tobaira, who provided me with the daily passion and reminder to finish what I started a lifetime ago. I will forever associate their love of Mary Poppins with my desire for peace and quiet. In the end, the bulk chapter 4 and 5 were written while humming to myself about a spoonful of sugar. Now that I have completed my project, there's time to go fly a kite.

## ACKNOWLEDGMENTS

First and foremost, I need to acknowledge my mother, Judi, for supporting me throughout my educational career and always finding a way to help me to do the things I needed, or sometimes just wanted, to do. Also to my stepfather, Ed. I truly appreciate how difficult it was to meet me as a teenager, but you have always filled the role you stepped into all of those years ago.

Academically, I would like to acknowledge my committee, and thank them for making the process as painless as possible. Dr. Sawilowsky always assured me that the main goal of enduring this process is to get my degree, something that he empowered me to do through the entire dissertation process. Dr. Fahoome almost single-handedly taught me everything that I learned throughout the program. Every semester for nearly three years we were reunited… hockey season or no hockey season. Dr. Bridge and, the most recent addition, Dr. Castronova, for being willing to be active participants in my dissertation endeavor, and for the latter, after the ball had already begun rolling.

Finally, I could have never done this without the support of my family. Thank you. Now it’s my turn to give back to the Buggeldome.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1 COMPARATIVE POWER OF THE ANOVA, APPROXIMATE RANDOMIZATION ANOVA, AND KRUSKAL-WALLIS TEST

## Introduction

The Kruskal-Wallis Test is a nonparametric alternative to the one-way ANOVA when assessing for shift in location. Analysis of variance (ANOVA) has been stated to be robust to departures from population normality (Glass, Peckham & Sanders, 1972). By definition, the Kruskal-Wallis Test is robust to this violation, because it does not operate under the assumption of normality. Under conditions of substantial non-normality, the permutation ANOVA has been proposed as an alternative to ANOVA to rehabilitate its robustness properties (Potvin & Roff, 1993) and has been asserted by Hunter & May (1993) to be superior in power to the nonparametric alternative. Hunter & May (1993) suggested that degrading the data to ranks, as nonparametric tests do, can produce a more powerful test only in some situations. The basis for that assertion appears to lie in the literature exploring the comparative efficacy of parametric tests to nonparametric counterparts in the context of the independent samples t test and the Wilcoxon-Mann-Whitney test, which are two-sample tests of location (van den Brink & van den Brink, 1989). Note that Sawilowsky (1993) contested the notion that converting to ranks makes nonparametric tests less powerful, and stated that for treatment alternatives of shift in location it actually results in tests more powerful than parametric and permutation counterparts.

Many researchers have demonstrated that under conditions of normality, power advantages of parametric tests such as Student's t and ANOVA are small when compared with their nonparametric counterparts, Wilcoxon-Mann-Whitney test and Kruskal-Wallis test, respectively (Blair & Higgins, 1985; Sawilowsky, 1990;

Zimmerman & Zumbo, 1990a). More specifically, Blair (1981) reports that the asymptotic relative efficiency of the Wilcoxon test relative to the t test is .955 when the assumption of normality is perfectly met. Later, Blair and Higgins (1985) examined the t test and Wilcoxon test under 10 different population shapes, determining that not only was the Wilcoxon test more often the most powerful, but in those situations it was vastly more powerful. Zimmerman and Zumbo (1990) assert that the power advantages that exist for nonparametric tests, such as the Wilcoxon test, are due to the elimination of the outlier influence by process of ranking. Indeed, the researchers state that when non-normal distributions have a restricted range of scores, as is the case in uniform distributions, the t test outperforms the Wilcoxon, a claim supported by the finding of Blair and Higgins (1985).

The propensity of researchers to prefer the use of parametric statistics have led many to propose transforming non-normal data in an attempt to satisfy the underlying parametric assumptions (e.g, Zimmerman & Zumbo, 1990b; Andrews, Gnanadesikan & Warner, 1971). However, others have shown that data transformation for certain designs can be dramatically non-robust and in many cases can have poor power properties (Sawilowsky, Blair & Higgins, 1989). This controversy highlights the need for researchers to better understand statistical procedures available to them given an unknown or non-normal population distribution.

**Problem**

Despite intricacies in the properties of parametric and nonparametric statistics, it is largely reported that parametric tests are generally more powerful than

nonparametric tests in all aspects, whether or not empirical support is provided for the claims (Blair & Higgins, 1985). Less common are the assertions that this aforementioned superiority of parametric tests lie on the foundation that the underlying assumptions such as distribution normality, must be perfectly met for the argument to hold true (e.g., Sharp, 1979).

Micceri (1989) noted when using the Kolmogorov-Smirnov test of normality on a sample of 440 distributions from published research, 100% of the distributions were significantly non-normal at the .01 alpha level. If this occurrence holds true, there are several factors to consider when determining the appropriate analysis for a given study. Additionally, given that a large portion of educational and behavioral science data being used to make instructional and policy decisions could be considered non-normal, one must decide the value placed on statistics that would be used in decision making.

If population normality is a condition under which parametric statistics are to be used, it would follow that one wishing to use these statistics must first analyze data to determine their sample distribution. Ryan (1959) discussed the issue of experiment-wise error, the likelihood that any one analysis within a given experiment will produce a Type I error. Each analysis performed on any given data set will increase the likelihood of a Type I error occurrence. Given this increased error rate, this presents the question of whether it is worth the sacrifice to test for normality before conducting a priori analyses given that other nonparametric alternatives exist that do not rely on the underlying assumption of normality, and subsequently, do not require tests of normality.

Type I error and test robustness also have other implications in educational practice. If parametric tests require larger and equal sample sizes to remedy their sometimes excessive Type I error under non-normality, this can have a major impact on public education research practices. Having statistical options that can maintain robustness to departure from normality even in light of smaller samples, can make educational research more obtainable in the face of limited budgets.

**Purpose of the Study**

The bulk of prior comparison research has explored the comparative power of two-sample tests of location. The statistics underlying these tests expand to the tests for three or more groups, in this case the ANOVA (F-ratio), the approximate randomization ANOVA, and Kruskal-Wallis tests. Additionally, it has been suggested that under non-normal conditions, the approximate randomization ANOVA will rehabilitate the statistical power of the ANOVA, making it a better and more accurate alternative to the Kruskal-Wallis. The current study will evaluate the comparative statistical power of the three tests mentioned above - the one-way ANOVA, the approximate randomization ANOVA, and the Kruskal-Wallis test – under differing sample size and distribution. To conduct an approximate randomization ANOVA, a researcher must have the access and ability to implement a Monte Carlo simulation relevant to their data structure. Additionally, given the superior robust qualities of the Kruskal-Wallis test to the ANOVA under conditions of non-normality, its comparability to the ANOVA under conditions of normality, and the simplicity with which the Kruskal-Wallis test is performed, establishing that the Kruskal-Wallis test was at least as powerful as the two previously stated alternatives under varying conditions would

have dramatic implications for researchers both in the field and in educational institutions. Therefore, the research questions are:

(a) What is the difference in statistical power of the three tests when all parametric assumptions are met?

(b) What is the difference in statistical power of the three tests when the assumption of normality is not met?

**Assumptions and Limitations**

Critical values will be obtained for the multiple iterations of the proposed statistical tests. The accuracy of these values will be determined by the number of iterations performed on the tests, and these could vary slightly from study to study. Additionally, the distributions created for the purposes of the study will be artificial in nature and may not be an accurate representation of a real-world distribution, but rather than idealized variation of real-world distributions.

In implementing artificial effect sizes, it should be noted that these effect sizes will be identical across groups, creating an ideal situation for parametric analysis. Additionally, group sizes in the samples will always be equal, another contributor to an essentially ideal condition for statistical testing.

**Definition of Terms**

Critical Value: The critical value(s) for a hypothesis test is a threshold to which the value of the test statistic in a sample is compared to determine whether or not the null hypothesis is rejected.

Data distribution: A display of scores in which the frequency of each score is readily apparent. It has two characteristics, central tendency and variability. The name of the distribution relies heavily on the central tendency.

Degrees of Freedom (*df*): The degrees of freedom of an estimate is equal to the number of independent scores that go into the estimate minus the number of parameters estimated as intermediate steps in the estimation of the parameter itself.

Monte Carlo Estimation: Computer intensive method used to test the hypothesis that the data are a random sample from a specified population. It allows for a substantial number of theoretical simulations.

Non-normality: Used to describe values of which the frequency distribution is different from the normal probability distribution.

Nonparametric Statistics: Statistical techniques designed to be used when the data being analyzed depart from the distributions that can be analyzed with parametric statistics. In practice, this most often means data measured on a nominal or an ordinal scale.

Outlier: An observation (or subset of observations), in a set of data which appears to be inconsistent with the remainder of that set of data

Parametric Tests: Statistical procedures, based on population parameters, for testing hypotheses or estimating parameters. A parametric statistical test depends on a number of assumptions about the population from which the samples used in the test are drawn.

Robustness: Insensitivity to departures from assumptions surrounding an underlying probabilistic model.

Type I Error: Rejecting the null hypothesis (*Ho*) when in fact it is true.

Type II Error: Failing to reject the null hypothesis (*Ho*) when in fact it is false.

Violation of Assumptions: Statistical hypothesis tests generally make assumptions about the population(s) from which the data were sampled. Many normal-theory-based tests such as the *t* test and ANOVA assume that the data are sampled from one or more normal distributions. If test assumptions are violated, the test results may not be valid.

# CHAPTER 2 LITERATURE REVIEW

## Hypothesis Testing

When conducting an experiment, the researcher is not generally interested only in those individuals participating in the different treatment conditions, but rather, is attempting to make inferences about the population from which the samples come. Experiments are conducted with the participation of a sample group, and the obtained statistics provide estimates of designated parameters for different treatment populations (Keppel & Wickens, 2004). In conducting experiments, researchers generally assert that a treatment may have some defined treatment result, called a hypothesis. In doing so, two mutually exclusive hypotheses are generated regarding the treatment parameters. In hypothesis testing, these two terms are identified as the null hypothesis ($H_0$) and alternative hypothesis ($H_1$). The null hypothesis assumes no difference exists among the treatment group(s) and control group, and in situations where the null hypothesis can be rejected, the alternative hypothesis confirms the presence of a difference between groups, presumed to be due to the treatment imposed.

Parametric hypotheses suggest that either the averages of the groups are equal (e.g., $H_0$: $\mu_1=\mu_2=\mu_3$) or not equal (e.g., $H_1$: $\mu_1\neq\mu_2\neq\mu_3$). The null hypothesis for a traditional randomization test is that, “the measurement for each person or other unit that is randomly assigned will be the same under one assignment to treatments as any alternative assignment that could have resulted from the random assignment procedure”. (Edgington, 1995, p. 2). That is to say that when the randomization null hypothesis is true, random assignment of scores to different groups randomly divides

measurements among the groups. A nonparametric null hypothesis, in the case of the Kruskal-Wallis, for example, states that there is no difference between the populations being compared (Neave & Worthington, 1988). Note that in the nonparametric null alternative there is no mention of a mean or average score.

## Parametric Tests

The ANOVA relies on a group mean for hypothesis testing. Each individual score within a group is compared to the group mean, and the difference is then squared. Because of the overall importance of the mean in parametric tests, they can be susceptible to outliers. If a sample contains multiple outliers, the results of the test can be suspect. There is a plethora of support for the use of ANOVA and other parametric tests, however, there are certain expectations that comes with the use of these tests. The parametric t and ANOVA tests rely on underlying assumptions. Most notably, the assumptions state that scores should be independent of each other, meaning no score should be impacted by another's. Additionally, it is assumed that variances are equal across groups and the population distributions from which samples are drawn are normally distributed (Hunter & May, 1993).

Research has indicated that parametric tests can maintain their power properties in light of encountering some of the aforementioned violations, as long as they are not severe or are few in number (Zimmerman, 1987; Sawilowsky & Blair, 1992). The consequences of failing to meet underlying assumptions in the use of the *F*-ratio was explored by Glass, Peckham, and Sanders (1972), who reported that the *F*-ratio is robust to departures from normality. In fact, it is a relatively understood

premise that concerning Type I error, the t test is robust to non-normal distributions as long as sample sizes are equal or closely so, sample sizes approach 30 or more, and the tests are two-tailed rather than one-tailed (Sawilowsky & Blair, 1992). The issue of being robust to departures from normality, however, should not be confused with being the best statistic for the situation in question. Scheffé (1959) warned that though the F-statistic may maintain acceptable degree of power under certain non-normal situations, that should not be taken to mean that it is broadly the best statistic in relation to other available statistics given certain populations. Additionally, reported results can be confusing because of the use of inordinately small sample sizes. Boneau (1962) found the power of the t test to modestly surpass that of the Wilcoxon test for certain non-normal distributions with sample sizes approaching 5. Conversely, in exploring similar comparisons with non-normal distributions with sample sizes $n1=n2=20$ and $n1=20$ and $n2=40$, Neave and Granger (1968) reported that the Wilcoxon was superior to the t statistic, receiving a power advantage as large as .12. These conflicting results illustrate the warning posed by Scheffe (1956) regarding the selection of test statistics.

Micceri (1989) performed an evaluation of 440 educational, social and behavioral research studies and found that despite the high prevalence of use of parametric statistics, normal populations essentially do not arise in the research. Roughly 3% of the studies examined approached a normal distribution while approximately 31% exhibited extreme tail weights. He reported in his findings that this exemplifies, “the need for careful data scrutiny prior to analysis, for purposes of both selecting statistics and interpreting results” (p. 161.).

Sawilowsky and Blair (1992) noted that real and existing distributions are generally of sufficient non-normality to bring about non-robust Type I error in the t test under certain circumstances. In situations where sample sizes are equal or nearly equal, sample sizes approach 25, and the tests are two-tailed, the t test demonstrated to be reasonable robust under non-normal conditions. Glass et al. (1972) suggested there is no need to abandon the t test in the face of non-normal data. Others have provided support for the use of alternative methods in the face of non-normal samples (e.g., Scheffe, 1959; Blair, 1981; Sawilowsky & Blair, 1992).

## Permutation and Randomization Tests

In an endorsement of the applicability of permutation tests, Good (1994) stated:

> Permutation tests can be applied to continuous, ordered and categorical data, and to values that are normal, almost normal, and non-normally distributed. For almost every parametric and nonparametric test, one may obtain a distribution-free permutation counterpart. The resulting permutation test is usually as powerful as more or powerful than alternative approaches. And permutation methods can sometimes be made to work when other statistical methods fail. (p. 1)

Permutation tests can take several forms. Exact permutation tests compile all possible combinations of available data for the chosen test statistic. They are called exact because the relevant properties are specifically determined, that is an exact level of significance is determined by a significance test (Walsh, 1968). The moment approximation test uses the continuous probability density function based on the

exact lower moments of the test statistic fitted to the discrete permutation distribution. Finally, the approximate randomization test focuses on a random subset of all possible permutations (Mielke & Berry, 2001). In situations where the number of permutations may be overwhelming due to a large sample size, an approximate randomization test can be a viable alternative. Several researchers suggest that permutation and randomization tests help to rehabilitate the power of parametric tests under conditions of non-normality (Potvin & Roff, 1993; Edgington, 1995). And still others offer permutation tests as preferred alternatives to rank-based tests, citing that rank tests are less powerful than randomization tests on scores (May, Masson, & Hunter, 1989).

Unlike parametric tests, permutation tests are considered to be distribution free (Bradley, 1968), and therefore are not bound by one of the major assumptions of the parametric tests, which is that the sample is drawn from a normal population (Hunter & May, 2003). Additionally, Noreen (1989) noted that random selection is not necessary for producing internally valid results, however, lack of randomization is a barrier to making inferences to a population.

There are some assumptions underlying permutation tests, however, that are important to consider. All observations are assumed to be independent of each other (Good, 1994), exchangeability of sample data under conditions of the null hypothesis (Good, 2002), continuity of distribution (Edgington, 1995), and homogeneity of variance (Boik, 1987). Importantly, it has been asserted that permutation methods have superior power to nonparametric tests due to their use of actual data rather

than ranks (Ludbrook & Dudley, 1998), although no compelling evidence has been offered to support this assertion.

The permutation model was first introduced by Fisher (1935), and with the continuing growth of computer technology, the procedure became more feasible to conduct. In a permutation test, data are shuffled to create all possible arrangements of data values (May & Hunter, 1993). Therefore, p-values are derived from a redistribution of the existing data. For approximate randomization tests, the precision with which the p-values can be derived depends largely on the number of iterations, or re-shufflings, created with the permutation process. This method differs from the permutation method because it does not create all necessary combinations of the data, but rather a number of iterations established by the researcher, and this number can vary depending on sample size and computing power, to mention but a few factors.

Permutation statistics offer a couple of advantages over parametric methods. Researchers do not need to refer to a table of critical values for a given test as the permutation test provides critical values based on the data available (Edgington, 1995). Under normality, the permutation tests are almost as powerful as the t test (Good, 1994) and have been stated by some researchers (Rao & Sen, 2002) to be more robust than parametric tests, though others have demonstrated that when samples contain similar means and unequal variances, permutation tests do not always maintain robustness (Boik, 1987; Manly, 1995). Another advantage of permutation tests is their ability to deal with outliers by likely detecting the difference in means with outliers (Edgington, 1995). Though permutation tests have been

referred to as nonparametric because of their assumptions, Hayes (1996) disagrees with the notion that permutation tests are nonparametric. In analyzing the relationship between the t test and permutation t, he found that in most cases, the two tests yielded nearly identical results. In comparing the two tests under conditions of heteroscedasticity, non-independence, and non-normal distributions, the permutation test exhibited error rates comparable to the t test.

## Nonparametric Tests

There are at least three types of nonparametric tests: categorical, sign, and rank tests (Sawilowsky, 1990). A test can be considered to be nonparametric when it can maintain satisfactory Type I error properties when assumptions such as normality do not hold true, as they make no assumptions about population parameters (Sawilowsky & Fahoome, 2003). Because of this, nonparametric statistics are good alternatives to parametric statistics under non-normal conditions (Lehmann, 1975). Although nonparametric tests are robust to departures from normality, they do still operate under the assumptions of independence of observations, random data selection, and a continuous distribution of data (Kerlinger & Lee, 2000).

As well as being robust to non-normality, nonparametric tests have been shown to be more powerful in testing shift in location under many non-normal situations (Blair & Higgins, 1985). A trend in research involving nonparametric statistics is that in situations where nonparametric tests are less powerful than parametric tests (e.g., normality), that power gap is small, whereas the power

advantage of nonparametric tests under conditions of non-normality can be dramatic (Sawilowsky, Blair, & Higgins, 1989; Blair & Higgins, 1985).

The Kruskal-Wallis test (Kruskal & Wallis, 1952) is identified as a nonparametric test due to the fact it does not make the assumption of a normal distribution. The Kruskal-Wallis test was derived from the *F*-test, the most notable difference being that it replaces the actual observations with ranks. Each score in the sample is assigned a rank which replaces the raw value, and that rank is used in the analysis. As ANOVA is a k-sample extension of the t test, the Kruskal-Wallis test is a k-sample extension of the Mann-Whitney U test. It assumes that sampling is random and that these samplings are from a continuous distribution (Feir-Walsh & Toothaker, 1974). It has been asserted that when sampling from a normal distribution, the Kruskal-Wallis test has power almost equal to the *F*-test and is much more reliable in the presence of outliers (Neave & Worthington, 1988).

The loudest detractors from nonparametric tests would state that because they use ranks rather than the actual data, power is lost. Indeed, many researchers (e.g., Lehman 1986; Adams & Anthony, 1996) have purported that it is for this reason that permutation tests are more powerful than other nonparametric tests. As permutation tests preserve raw values, some researchers have the opinion that permutation tests are superior to rank tests (Ludbrook & Dudley, 1998). However, others make the claim that ranking scores has no impact on the data, but rather removes some of the noise (Blair & Higgins, 2000; Sawilowsky, 1993). Still others have stated that not only does ranking not create a loss in power, but the power may actually increase (Langbehn, Berger, Higgins, Blair, & Mallows, 2000).

## Underlying Assumptions

All three previously mentioned statistical tests operate under the assumption of independence of scores. That is, all scores are independent of other scores and are in no way affected by other scores. Given that all data are drawn from a random number generator, this assumption holds true in this study. The ANOVA is a parametric test and as such, assumes the sample data to be normally distributed. The randomization ANOVA and Kruskal-Wallis test make no such assumption. Other assumptions shared by both the ANOVA and the Permutation ANOVA is homogeneity of variance and the use of at least interval data, again, assumptions not made by the Kruskal-Wallis test, as data are ranked prior to analysis.

# CHAPTER 3 METHODOLOGY

## Overview

The purpose of the study is to compare the Type I error and comparative statistical power of three statistical methods for assessing a difference in means across K > 2 groups. The three hypothesis tests include (1) the classical parametric one-way ANOVA, (2) a distribution-free approximate randomization one-way ANOVA, and (3) the nonparametric Kruskal-Wallis.

## Design

Cohen (1988) suggested parameters for identifying small, medium, and large effect sizes in the one-way ANOVA layout. A small effect size was defined as f = .1σ, medium = .25σ, and large = .4σ. In keeping with the recommendations of Sawilowsky (2009) in expanding magnitudes from two sample layout, a very large effect size will be defined as f = .6σ, and a huge effect size will be defined as f = 1.0σ, where σ refers to the standard deviation of the distribution selected.

## Distributions

Data will be drawn from three theoretic distributions. Data will be sampled from a normal distribution ($\mu = 0, \sigma = 1$) to demonstrate the veracity of the Monte Carlo study. A uniform distribution and chi-square distribution (*df* = 2; also known as an exponential distribution with shape parameter =2) will be used to test conditions under which the distribution assumption does not hold. (Note that homoscedasticity will be maintained.)

The normal (Gaussian) distribution was identified due to being the ideal condition under which the ANOVA is the uniformly most powerful and unbiased

(UMPU) test. Additionally, two other distributions, uniform and chi-square, were chosen as a comparison for conditions which violate the normality assumption. The descriptions of the distributions are as follows (Sawilowsky & Fahoome, 2003):

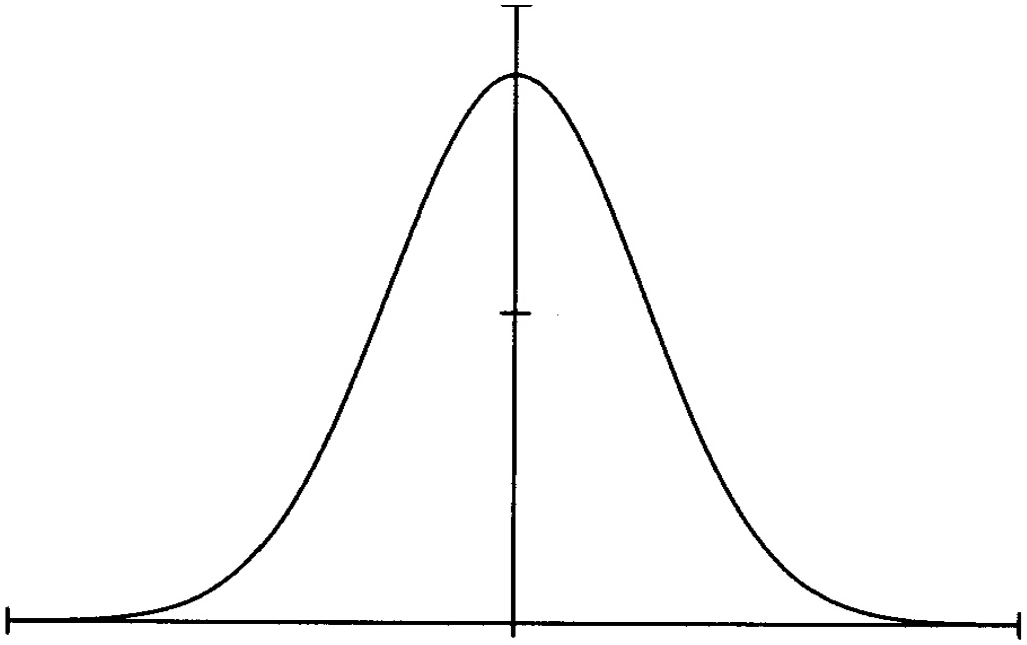

*Figure 1*. Gaussian (Normal) Distribution, Sawilowsky & Fahoome (2003).

1. Normal Distribution: This “bell shaped” curve has symmetric light tails and contains an equal distribution of scores. The mean and median = 0, and the standard deviation = 1. The probability density function of U is as follows:

$$P_U(u) = (\sqrt{2\pi})^{-1} \exp(-\tfrac{1}{2} u^2)$$

Despite being the underlying assumption for parametric tests, Micceri (1989) noted that 15% of the psychometric, achievement, criterion/mastery, and gain score studies only qualified as near-Gaussian. Most support for the use of this distribution is

derived from the central limit theorems, which postulate that “the distribution of standardized sums of random variables tends to a unit normal distribution as the number of the variables in the sum increases” (Johnson & Kotz, 1970, p.45). Johnson and Kotz (1970) also state that the normal distribution can be used to approximate to other distributions.

*Figure 2*. Uniform Distribution, Sawilowsky & Fahoome (2003).

2. Uniform Distribution: This distribution, similar to the normal distribution, is symmetric with light tails. The probability distribution function of a uniform distribution is as follows:

$$p_Y(y) = (\beta - \alpha)^{-1} \qquad (\alpha \le y \le \beta)$$

A uniform distribution is often used to represent rounding off errors when forming numbers to a set number of decimal places (Johnson & Kotz, 1970). In conditions where there is a preferences for discrete objects in which each choice is equally

likely, the likely outcomes are represented by a uniform distribution. Micceri (1989) noted that in his exploration of 440 social science and educational studies, approximately 3% of said studies conformed to a uniform distribution. However, in the engineering field in which machined parts are produced within a particular range of acceptability, these parts are often produced with variation represented by a uniform distribution (Mendenhall & Sincich, 1995).

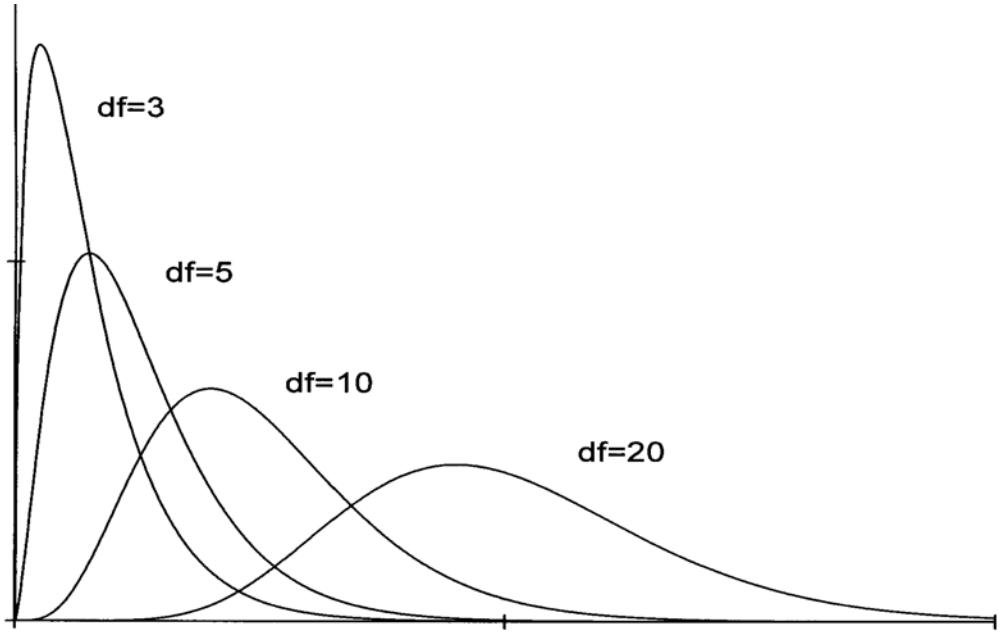


*Figure 3*. Chi-Square Distribution, Sawilowsky & Fahoome (2003).

3. Chi-Square Distribution: Also referred to as exponential when containing 2 degrees of freedom, this distribution represents the comparison between expected and actual outcomes. The probability density function is as follows:

$$p_X(x) = \sigma^{-1} \exp[-(x-\theta)/\sigma] \quad (x > \theta; \sigma > 0)$$

These distributions are relatively common when modeling the wait times between some unknown event and events recurring at random time intervals (Johnson & Kotz, 1970). For example, this can include time between arrivals at a service counter, time between earthquakes, or the length of time a machine will operate before breaking down. This can be prevalent in the mechanical and engineering field due to its relevance in electrical and mechanical lifespans.

**Sample Sizes and Effect Sizes**

Differing sample sizes will be invoked, as well as differing patterns of simulated treatment effects (tr), as noted in Table 1 below. Conditions 1, 4, 7, and 12 present the null condition. The remaining conditions present a systematic pattern of the number of non-null groups:

Table 1

*Sample Size and Treatment Conditions*

| | Sample Size | | | | |
|---|---|---|---|---|---|
| Condition | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 |
| 1 (null) | 10 | 10 | 10 | - | - |
| 2 | 10 | 10 | 10(tr) | - | - |
| 3 | 10 | 10(tr) | 10(tr) | - | - |
| 4 (null) | 30 | 30 | 30 | - | - |
| 5 | 30 | 30 | 30(tr) | - | - |
| 6 | 30 | 30(tr) | 30(tr) | - | - |
| 7 (null) | 10 | 10 | 10 | 10 | 10 |
| 8 | 10 | 10 | 10 | 10 | 10(tr) |
| 9 | 10 | 10 | 10 | 10(tr) | 10(tr) |
| 10 | 10 | 10 | 10(tr) | 10(tr) | 10(tr) |
| 11 | 10 | 10(tr) | 10(tr) | 10(tr) | 10(tr) |
| 12 (null) | 30 | 30 | 30 | 30 | 30 |
| 13 | 30 | 30 | 30 | 30 | 30(tr) |
| 14 | 30 | 30 | 30 | 30(tr) | 30(tr) |
| 15 | 30 | 30 | 30(tr) | 30(tr) | 30(tr) |
| 16 | 30 | 30(tr) | 30(tr) | 30(tr) | 30(tr) |

*Note.* The notation (tr) refers to a group that is receiving the designated treatment effect.

**Fortran Programming**

The Monte Carlo study will be performed using the Fortran programming language and the IMSL subroutine library. Pseudo-random number generators will be invoked to obtain random variates from the normal distribution, and random deviates from the non-normal distributions. Nominal alpha will be set at $\alpha$ = 0.05 and $\alpha$ = 0.01.

Each experiment will be repeated 20,000 times for each distribution under each sample size condition, treatment alternative, and alpha level. Within each iteration, however, the approximate randomization test will be conducted based on 5,000

permutations. The approximate randomization test will be performed with random permutations, which is a procedure also known as a Monte Carlo version of the test.

Each statistical test will be conducted on each sample condition under the null condition, that is, in the absence of treatment, before treatments are added. Then, power comparisons of the ANOVA, approximate randomization ANOVA, and the Kruskal-Wallis test or groups of $k$=3 and $k$=5 will be made by introducing treatment effects modeled as a shift in location parameter. For non-null conditions, a constant will be added to each treatment group in graduated increments, until all but one group has received a treatment. Treatments across groups within any analysis will be of equal magnitude and each treatment group will receive all effect sizes. Type I error and power will be identified as the rate of rejection of the null hypothesis under all treatment and distribution conditions.

The Monte Carlo study will be performed using Absoft version 11.1 compiler and written in Fortran 77 language. The program will utilize the International Mathematics and Statistics Library (IMSL) to compute the tests of significance performed. For the theoretical distributions, the program utilizes separate random number generators for the normal distribution (RNNOR), chi-squared distribution (RNCHI), and the uniform distribution (RNUN). Analyses will be performed using a Toshiba Satellite A505 computer with an Intel Core2 Duo$^{TM}$ processor (2.20 GHz x 2) and 3.87 GB of usable RAM. The computer utilizes the Windows 7 Home Premium edition with Service Pack 1.

## Presentation of Results

Results will be reported using tables of rejection rates to depict the Type I error rates under the truth of the null hypothesis, as well as the relative power of each statistical method under each sample size and distribution condition. For each experimental condition, a graph will illustrate the power curve of the three statistics being explored.

# CHAPTER 4 RESULTS

## Overview

A Monte Carlo simulation was performed to examine the Type I error rates and power properties of the ANOVA, approximate randomization ANOVA, and the Kruskal-Wallis test for data sampled from three theoretical distributions, two sample sizes ($n_i$ = 10 and n*i* = 30), and number of groups of K = 3 and K = 5. Results provide further support for previously reported findings on the t and F statistic, as well as provide new information regarding the relative powers of the three K ≥ 3 tests, when treatments were modeled as a shift of location parameter. The three theoretical distributions explored were: the normal (Gaussian) distribution, the uniform distribution, and the chi-square (df=2) distribution.

## Type I Error

To determine the Type I error properties of the statistical tests under the differing distribution and sample condition, a Monte Carlo analysis was written to tally the number of null rejections in the absence of treatment effect.

Table 2

*Rejections Under Null Condition for Normal Distribution (Type I Error)*

| | ANOVA | | Approximate Randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|
| Sample Size: | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| $n_1=n_2=n_3=10$ | 0.04825 | 0.00975 | 0.04880 | 0.01000 | 0.04860 | 0.01000 |
| $n_1=n_2=n_3=30$ | 0.04920 | 0.00915 | 0.05050 | 0.00910 | 0.04895 | 0.00975 |
| $n_1=n_2=n_3=n_4=n_5=10$ | 0.04930 | 0.01040 | 0.04930 | 0.01060 | 0.04995 | 0.01080 |
| $n_1=n_2=n_3=n_4=n_5=30$ | 0.04940 | 0.01040 | 0.04940 | 0.01070 | 0.04855 | 0.00985 |

Table 3

*Rejections Under Null Condition for Uniform Distribution (Type I Error)*

| | ANOVA | | Approximate Randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|
| Sample Size: | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| $n_1=n_2=n_3=10$ | 0.05120 | 0.01105 | 0.04915 | 0.01020 | 0.04870 | 0.01005 |
| $n_1=n_2=n_3=30$ | 0.04955 | 0.00970 | 0.04955 | 0.00985 | 0.04865 | 0.00965 |
| $n_1=n_2=n_3=n_4=n_5=10$ | 0.05080 | 0.01110 | 0.04910 | 0.01015 | 0.05040 | 0.01065 |
| $n_1=n_2=n_3=n_4=n_5=30$ | 0.04895 | 0.01000 | 0.04910 | 0.01015 | 0.04870 | 0.00995 |

Table 4

*Rejections Under Null Condition for Chi-Square (df=2) Distribution (Type I Error)*

| | ANOVA | | Approximate Randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|
| Sample Size: | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| $n_1=n_2=n_3=10$ | 0.04240 | 0.00740 | 0.05030 | 0.00985 | 0.04845 | 0.01015 |
| $n_1=n_2=n_3=30$ | 0.04460 | 0.00780 | 0.04845 | 0.00960 | 0.04890 | 0.00960 |
| $n_1=n_2=n_3=n_4=n_5=10$ | 0.04455 | 0.00890 | 0.05110 | 0.01045 | 0.04985 | 0.01075 |
| $n_1=n_2=n_3=n_4=n_5=30$ | 0.04700 | 0.01020 | 0.05010 | 0.01080 | 0.04845 | 0.00990 |

For data obtained from the normal distribution, Type I error rates were relatively consistent across samples and statistical tests. Most notably, no test was clearly superior in error rates across sample conditions. Also notable was that error rates on the whole for the normal distribution ranged from 4.825% to 5.050% across all tests at the α = .05 level and 0.910% to 1.080% at the α = .01 level. The rates fell well within range of those previously reported in the literature.

For data sampled from the uniform distribution, Type I error rates were again consistent across tests and sample conditions, though the ANOVA never demonstrated a superior error rate under any condition or alpha level. The error rates on the whole for the uniform distribution ranged from 4.865% to 5.120% across all tests at the $\alpha = .05$ level and 0.965% to 1.110% at the $\alpha = .01$ level.

For data obtained from the chi-square (df=2) distribution, the ANOVA consistently demonstrated error rates outside of conservative parameters under all but one sample condition. The error rates of the other tests remained at or around their designated level, with the rates for the approximate randomization ANOVA and Kruskal-Wallis ranging from 4.845% to 5.110% at the $\alpha = .05$ level and 0.960% and 1.080% at the $\alpha = .01$ level.

## Comparative Power Analysis

After assessing Type I error rates for each sample and distribution condition, equal treatments (tr) ranging from $0.1\sigma$ to $1.0\sigma$ were imposed on a progressive number of groups within each sample, to the maximum of k-1 groups per condition [e.g., $n_1=n_2=n_3=n_4=n_5(tr)=30$; $n_1=n_2=n_3=n_4(tr)=n_5(tr)=30$; $n_1=n_2=n_3(tr)=n_4(tr)=n_5(tr)=30$; $n_1=n_2(tr)=n_3(tr)=n_4(tr)=n_5(tr)=30$]. The results for the treatment conditions were organized into 12 tables by sample size and power curves are illustrated in a graph for each experimental condition. Note that in discussing the results of the power analysis, this researcher focuses primarily on the $\alpha = .05$ level due to its prevalence in research, though results for $\alpha = .01$ are included in all tables and follow the same trends. Following are the results of the study, organized by distribution.

## Normal Distribution

### Sample $n_1=n_2=n_3=10$

The first sample explored was $n_1=n_2=n_3=10$, in which one group received treatment. For the one treatment condition, the results of the three tests under the condition of normality were consistent across effect sizes. The results of the two treatment condition were essentially identical to the one treatment condition.

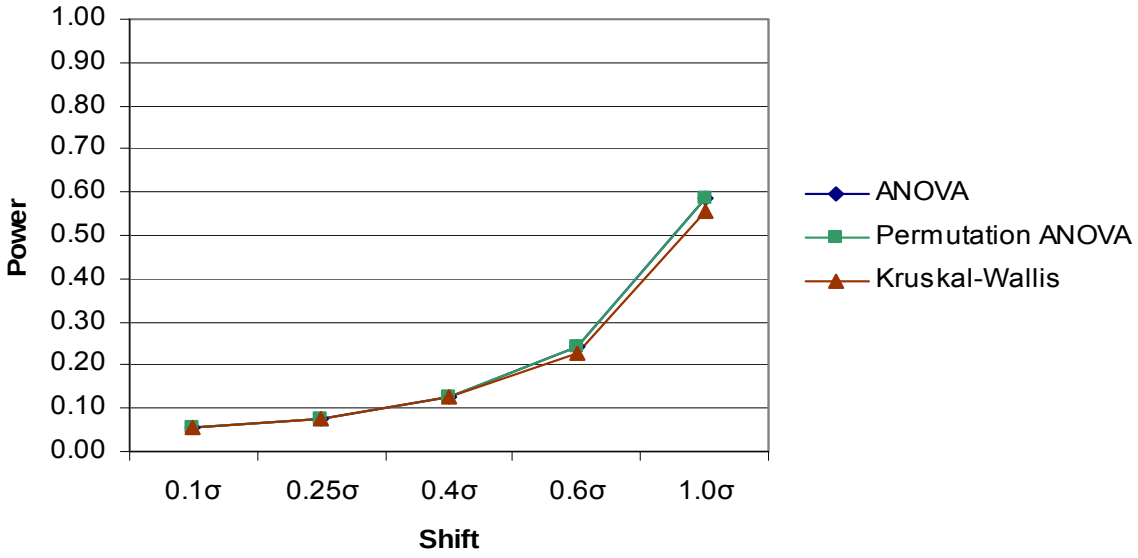


*Figure 4.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2=n_3(tr)=10$.

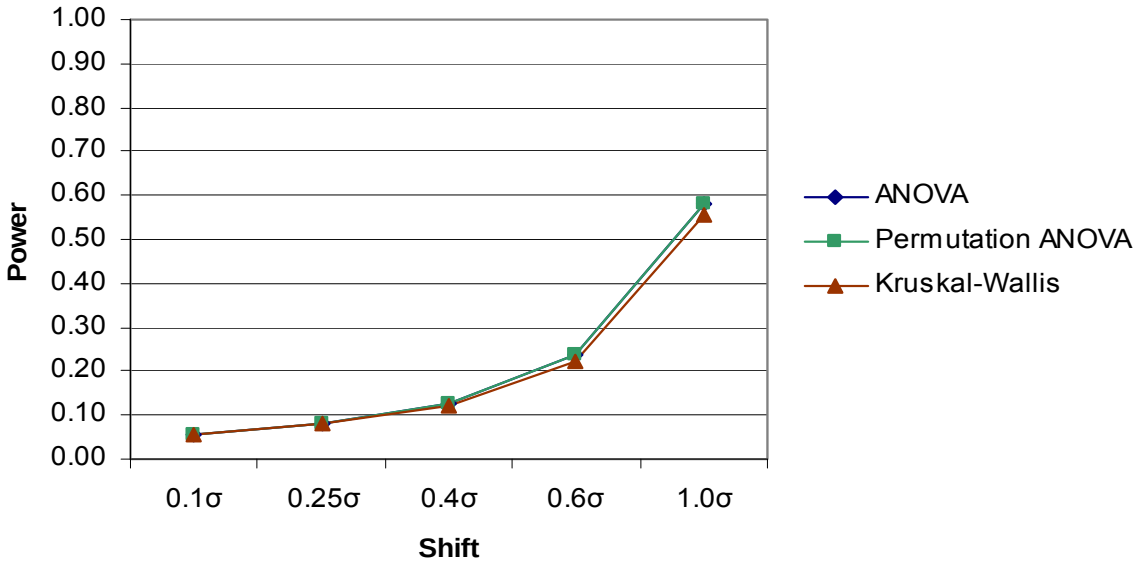


*Figure 5.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2(tr)=n_3(tr)=10$.

In the one treatment condition, the ANOVA demonstrated power of .0540 at the 0.1σ shift, increasing to .1268 at 0.4σ and .5854 at the 1.0σ shift, with the approximate randomization ANOVA nearly equal at every effect size. The Kruskal-Wallis demonstrated power of .0545 at 0.1σ, rising to .1248 at 0.4σ and .5572 at 1.0σ. The largest power discrepancy across tests was at the 1.0σ effect size, where the ANOVA and approximate randomization ANOVA achieved a power of .5854, and the Kruskal-Wallis .5572. With the exception of the 0.1σ shift, the Kruskal-Wallis trailed both the ANOVA and the approximate randomization ANOVA in power at every degree of shift. The results for the two treatment group condition were nearly identical.

**Sample $n_1=n_2=n_3=30$**

For both the one treatment and two treatment group conditions of the $n_1=n_2=n_3=30$ sample, the results of the three tests under conditions of normality were again consistent across effect sizes and the power curves were nearly identical for both treatment conditions.

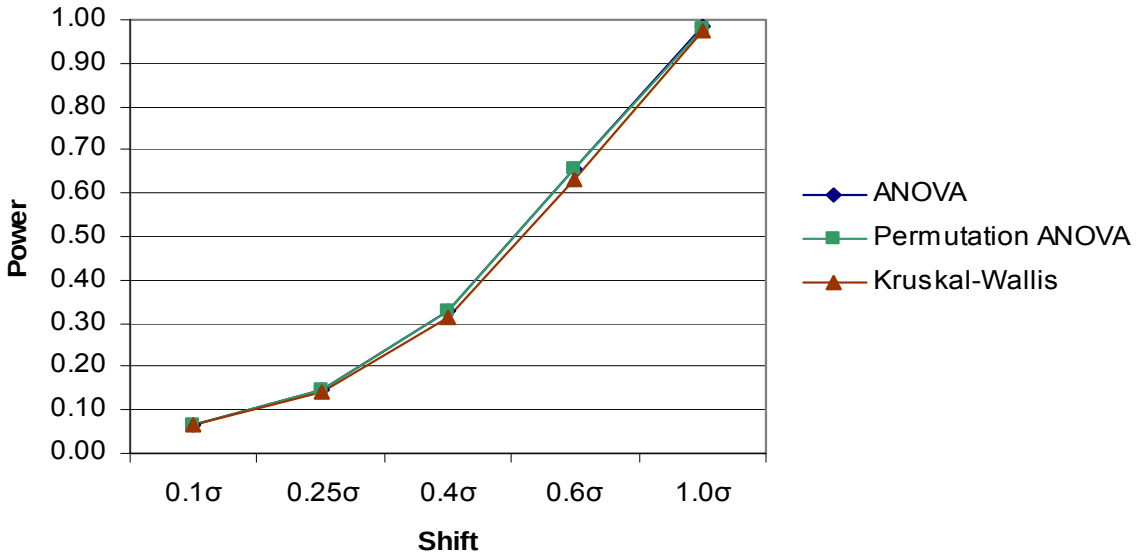


*Figure 6.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2=n_3(tr)=30$.

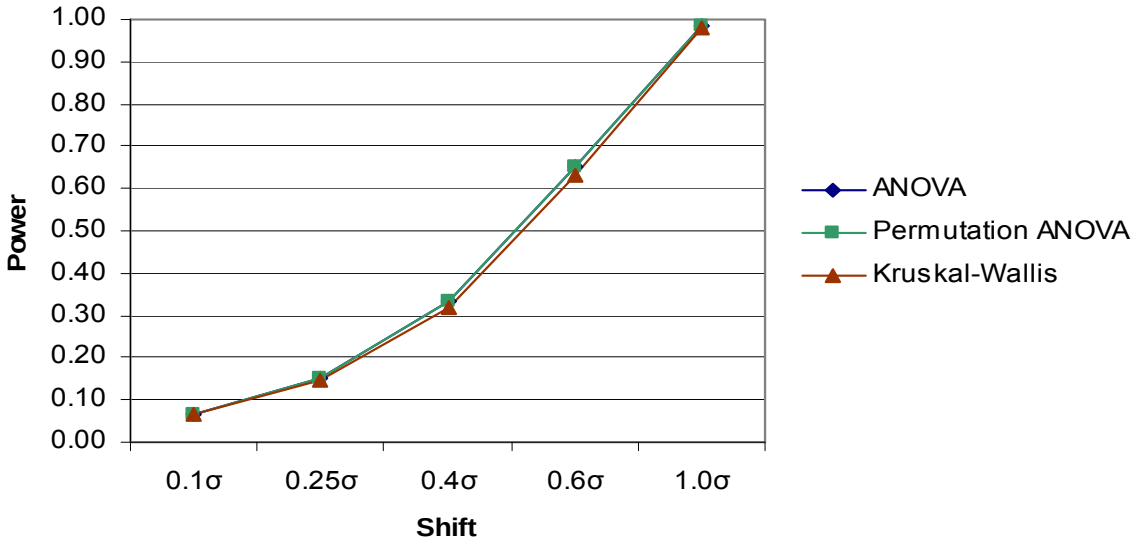


*Figure 7.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2(tr)=n_3(tr)=30$.

In the one treatment group condition, the ANOVA demonstrated power of .0632 at 0.1σ, to .3302 at 0.4σ, and increasing to .9826 at 1.0σ with the approximate randomization ANOVA nearly equal, and with one exception marginally more powerful, at every effect size. The Kruskal-Wallis demonstrated power of .0634 at 0.1σ, to .3138 at 0.4σ, and .9768 at 1.0σ. The largest power discrepancy across tests was at the 0.6σ shift, at which the ANOVA achieved .6552, the approximate randomization ANOVA .6560, and the Kruskal-Wallis .6326. The two treatment group conditions exhibited nearly identical results and trends to the one treatment condition.

**Sample $n_1=n_2=n_3=n_4=n_5=10$**

The results of the three tests under the condition of normality for the one treatment group were consistent across effect sizes.

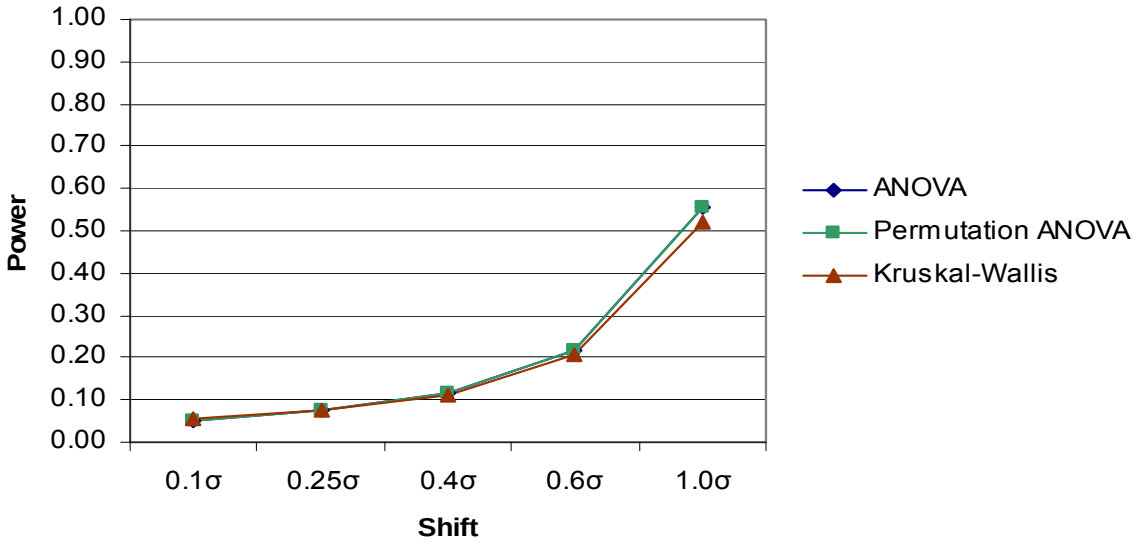


*Figure 8.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2=n_3=n_4=n_5(tr)=10$.

The ANOVA demonstrated power of .0521 at 0.1σ, to .1182 at 0.4σ, and increased steadily to .5557 at 1.0σ, with the approximate randomization ANOVA nearly equal at every effect size. The Kruskal-Wallis demonstrated power of .0534 at 0.1σ, to .1125 at 0.4σ, and rose to .5207 at 1.0σ. The largest power discrepancy across tests was at the 1.0σ effect size, where the ANOVA and approximate randomization ANOVA achieved a power of about .55, and the Kruskal-Wallis .5207. With the exception of the 0.1σ shift, the Kruskal-Wallis trailed both the ANOVA and the approximate randomization ANOVA in power.

For the two and three treatment group conditions, the relationship of the three tests with each other under all distributions very strongly resembled that of the one treatment group condition, with the difference being the pace at which the power levels increased with each shift.

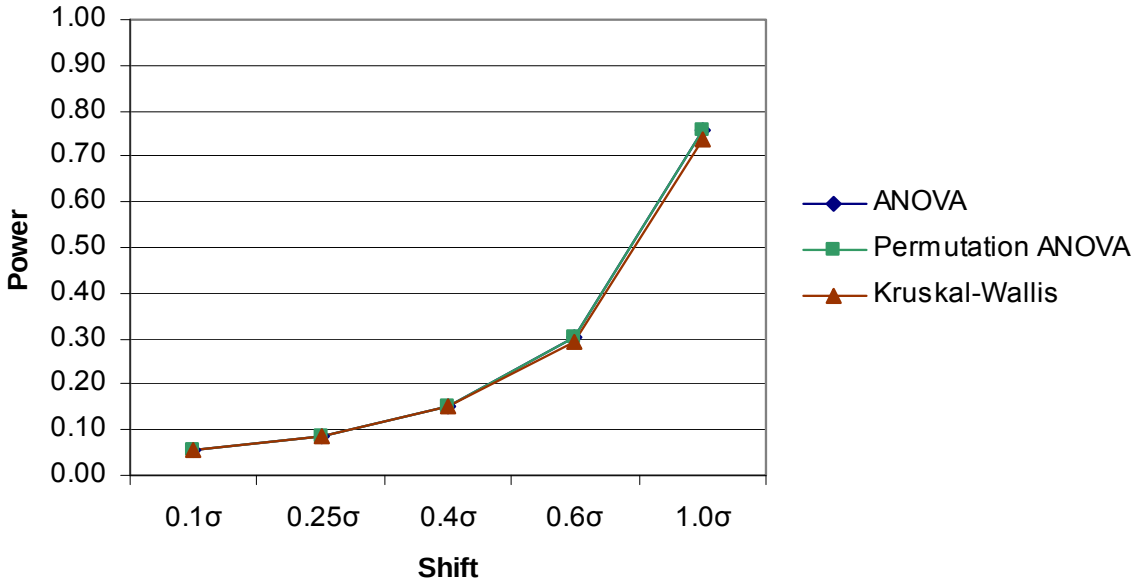


*Figure 9.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2=n_3=n_4(tr)=n_5(tr)=10$.

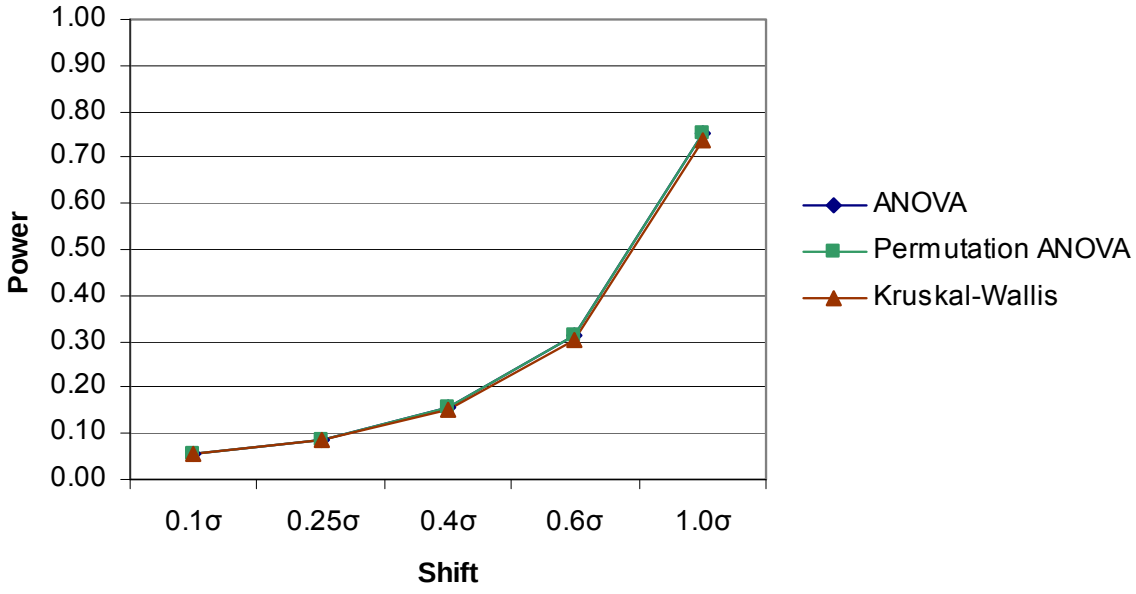


*Figure 10.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2=n_3(tr)=n_4(tr)=n_5(tr)=10$.

No test demonstrated more than a 2.0% power advantage at any shift under the normal distribution, with the ANOVA and approximate randomization ANOVA trading off very slight power advantage at alternating shifts. Under normality, all three tests gained power incrementally at roughly the same pace, with power of approximately .05 at 0.1σ, .15 at 0.4σ, and .75 at 1.0σ.

For the four treatment groups condition, all patterns under the normal distribution and shift sizes remained essentially the same as the one treatment group condition,

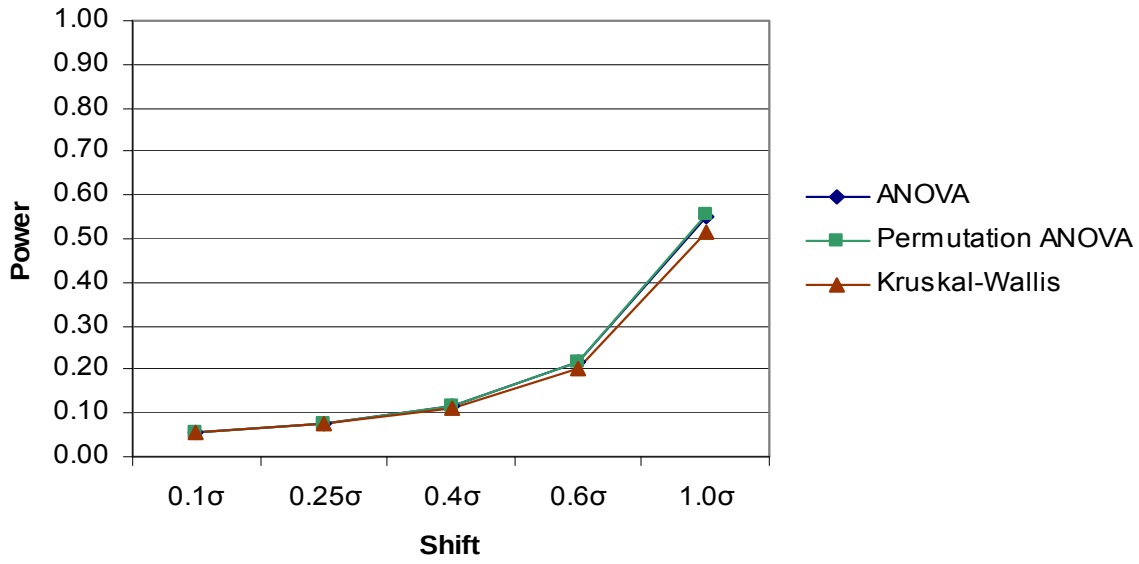


*Figure 11.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2(tr)=n_3(tr)=n_4(tr)=n_5(tr)=10$.

**Sample $n_1=n_2=n_3=n_4=n_5=30$**

The next subset of sample conditions involved exploring the effect of differing treatment effect sizes on five groups of n=30, in which one, two, three, and four

groups receive equal treatment. For the one treatment group, the power results of the three tests under conditions of normality were very similar across effect sizes.

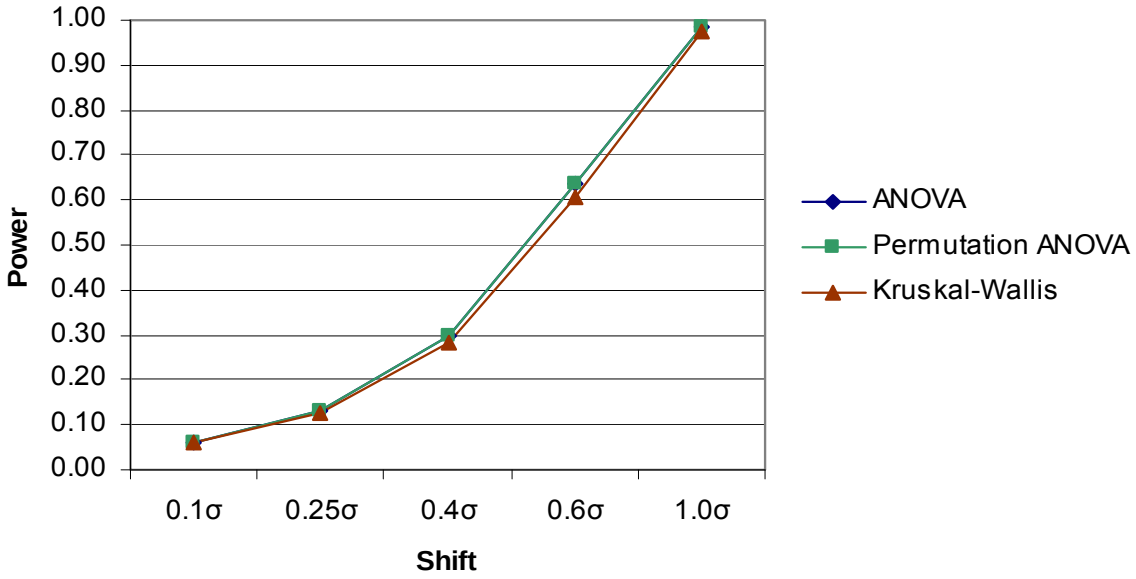


*Figure 12.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2=n_3=n_4=n_5(tr)=30$.

The ANOVA demonstrated power of .0601 at 0.1σ, to .2986 at 0.4σ, and increased to .9826 at 1.0σ. The power curve of the approximate randomization ANOVA followed that of the ANOVA almost exactly, never departing more than .0008 in power. The Kruskal-Wallis demonstrated power of .0600 at 0.1σ, to .2813 at 0.4σ, and rose to .9749 at 1.0σ. The largest power discrepancy across tests was at the 0.6σ effect size, where the ANOVA and approximate randomization ANOVA achieved a power of about .64 and the Kruskal-Wallis .61, a difference of approximately 3%. Unlike the n=10 conditions, the Kruskal-Wallis trailed both the ANOVA and the approximate randomization ANOVA in power at every effect size, though the discrepancy was usually modest.

As was the case with the $n_1=n_2=n_3=n_4=n_5=10$ subset, the four treatment groups condition demonstrated very similar patterns to the one treatment group condition under the normal distribution.

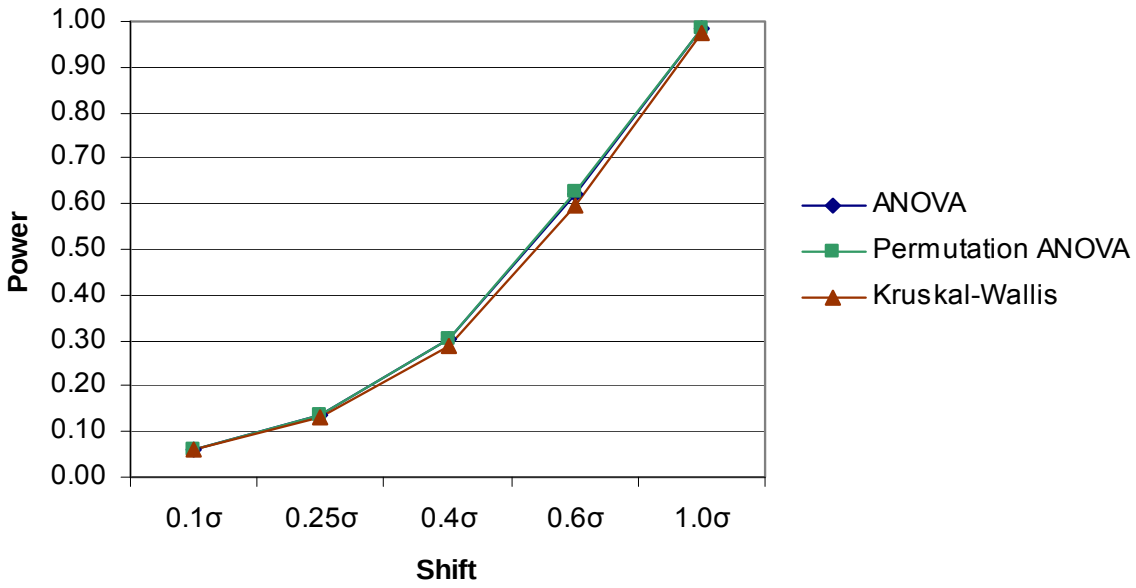


*Figure 13.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2(tr)=n_3(tr)=n_4(tr)=n_5(tr)=30$.

For the two and three treatment group conditions, the relationship of the three tests with each other under the normal distribution very strongly resembled that of the one treatment group condition. At no point was there a power discrepancy reaching 1% for any corresponding test or effect size condition between the two and three treatment group condition (the only exception being a power increase of roughly 2% for the Kruskal-Wallis at 0.25σ).

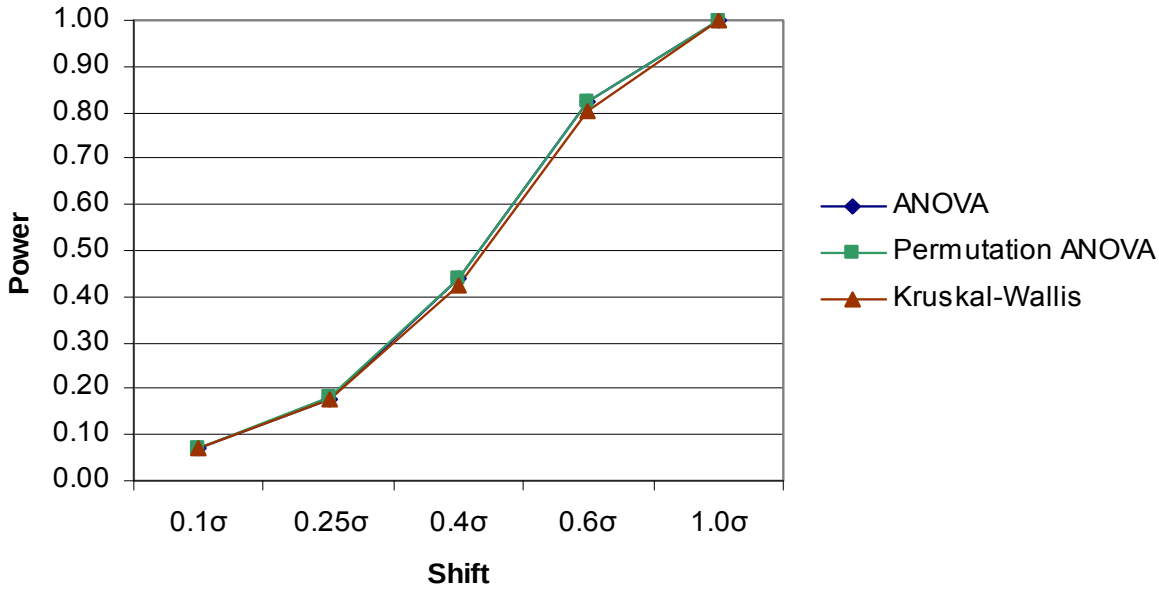


*Figure 14.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2=n_3=n_4(tr)=n_5(tr)=30$.

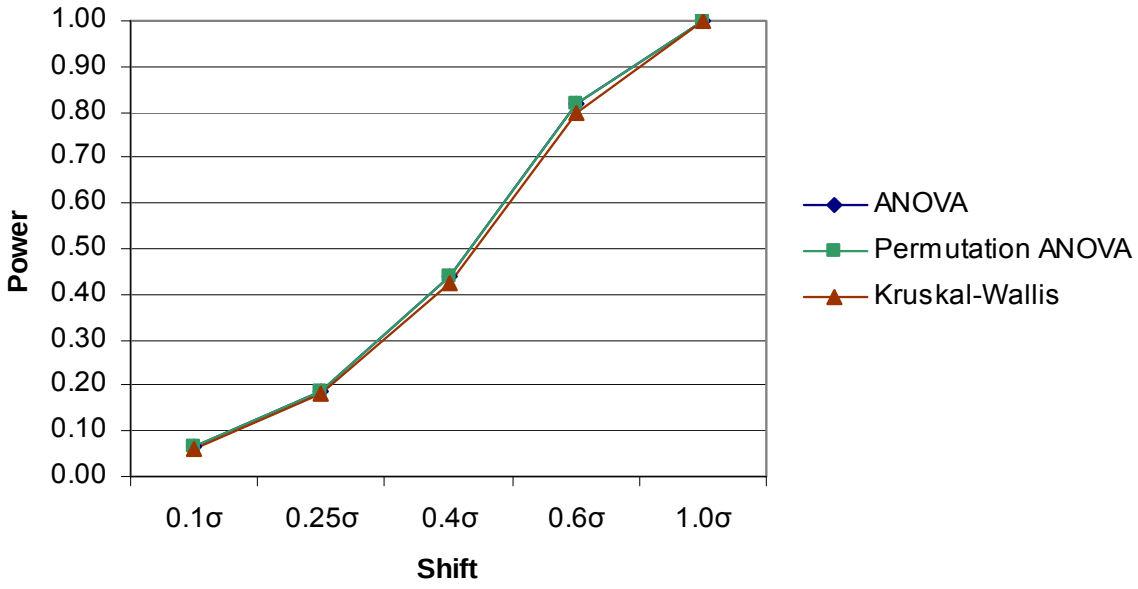


*Figure 15.* Shift vs. Power in the normal distribution for sample condition $n_1=n_2=n_3(tr)=n_4(tr)=n_5(tr)=30$.

With the exception of a nearly 2% power discrepancy between the Kruskal-Wallis (lower) and the other tests at 0.6σ for both the two and three treatment group conditions, no test demonstrated more than a 2% or higher power advantage at any shift under the normal distribution. The ANOVA and approximate randomization ANOVA traded off very slight power advantages or ties at alternating shifts. Under normality, all three tests gained power incrementally at roughly the same pace, with power of approximately .07 at 0.1σ, .43 at 0.4σ, and .99 at 1.0σ.

**Uniform Distribution**

**Sample $n_1=n_2=n_3=10$**

The first sample explored was $n_1=n_2=n_3=10$, in which one group received treatment. The same trends in power differential existed under the uniform distribution as the normal distribution, with the ANOVA and approximate randomization ANOVA outperforming the Kruskal-Wallis at every treatment level.

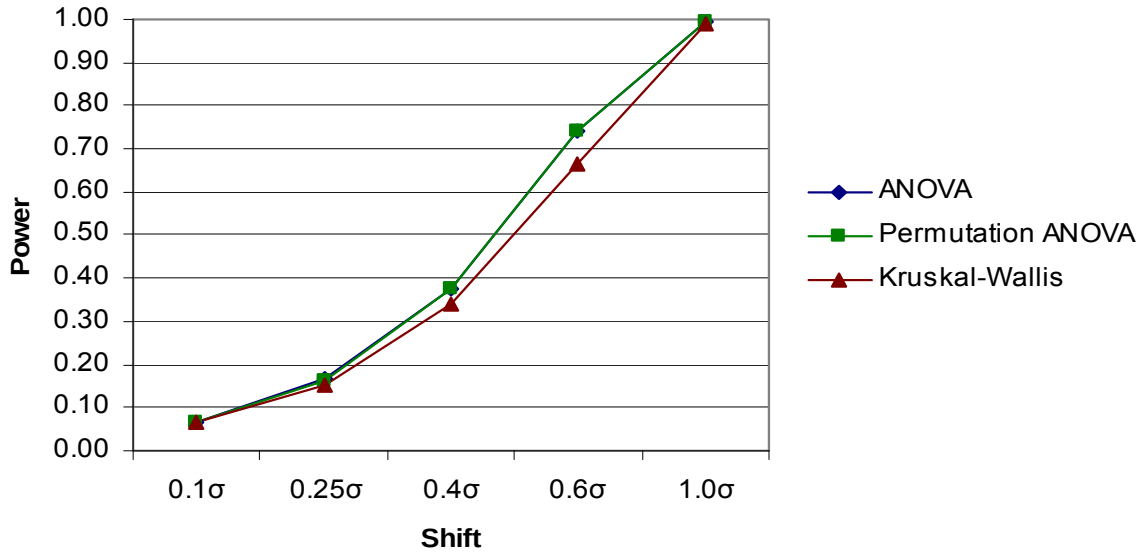


*Figure 16.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2=n_3(tr)=10$.

The ANOVA demonstrated the highest power on all but the 1.0σ shift, in which the approximate randomization ANOVA was nearly identical in power. The ANOVA ranged from a power of .0683 at 0.1σ to .37758 at 0.4σ and .9968 at 1.0σ with the approximate randomization ANOVA showing very similar, though mostly lower, levels. The Kruskal-Wallis ranged from .0657 at 0.1σ to .3402 at 0.4σ and.9877 at 1.0σ. The most notable outcome under the uniform distribution was that the Kruskal-Wallis was outperformed by the ANOVA by nearly 8% at the 0.6σ shift before demonstrating more similar power results at 1.0σ.

The two treatment group condition exhibited nearly identical results and trends to the one treatment group condition.

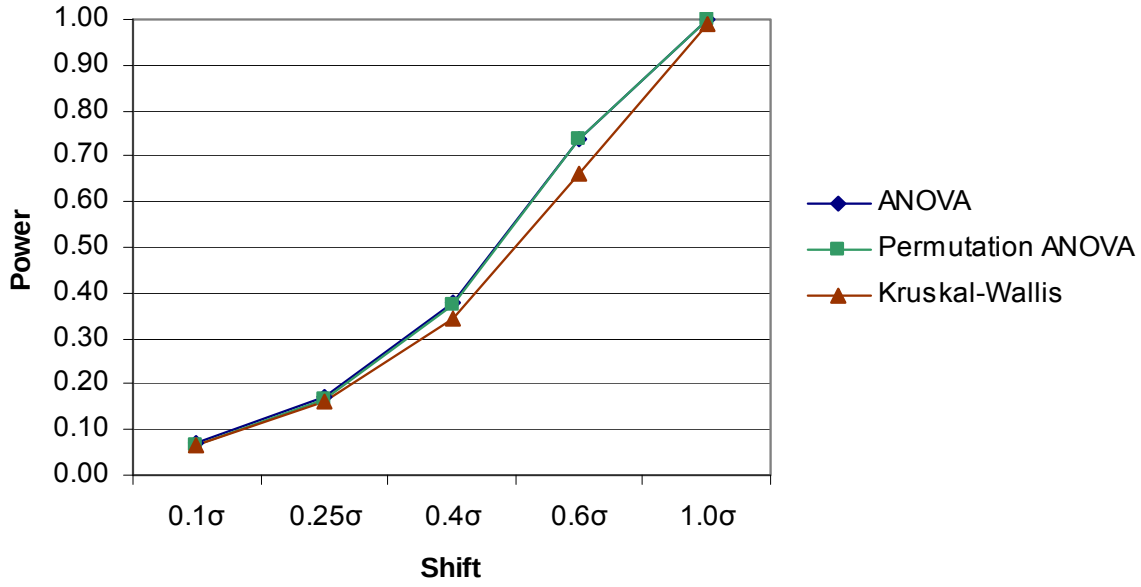


*Figure 17.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2(tr)=n_3(tr)=10$.

Power analysis results yielded almost identical rates for each shift size. Also similar was the pattern of the power drop-off for the Kruskal-Wallis from the other tests under the uniform distribution as the effect size increased, approaching a differential of 2.5% from the ANOVA and approximate randomization ANOVA.

**Sample $n_1 = n_2 = n_3 = 30$**

In the one treatment group condition, the power of the Kruskal-Wallis quickly lagged behind that of the ANOVA and approximate randomization ANOVA, reaching a drop-off of 4% at the 0.25σ shift and 5% at the 0.4σ shift.

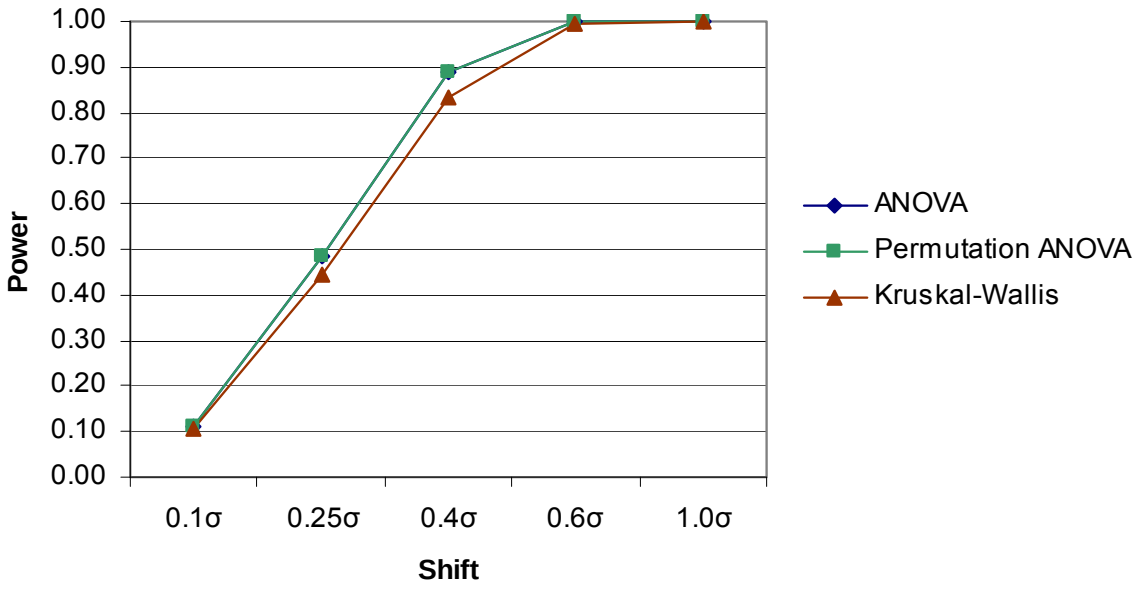


*Figure 18.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2=n_3(tr)=30$.

All tests surpassed a power of .99 at the 0.6σ shift and reached a power of 1.0 at the 1.0σ shift. The ANOVA demonstrated power of .1103 at 0.1σ, increasing to .8872 at 0.4σ with the approximate randomization ANOVA following a nearly identical

pattern. The Kruskal-Wallis demonstrated power of .1063 at 0.1σ, rising to .8344 at 0.4σ.

The two treatment group condition exhibited nearly identical results and trends to the one treatment condition.

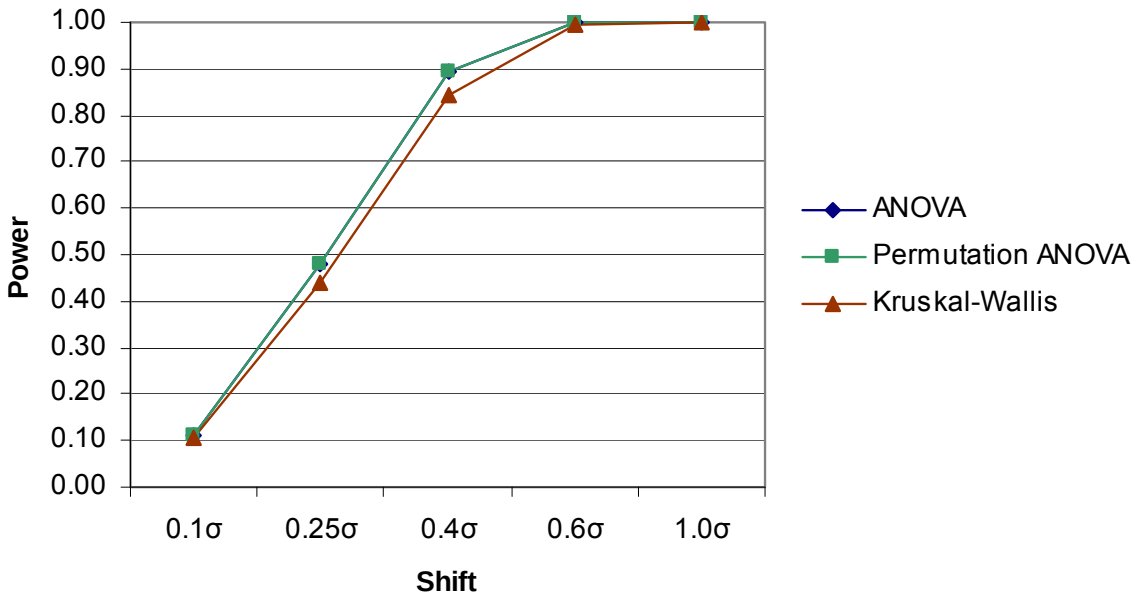


*Figure 19.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2(tr)=n_3(tr)=30$.

**Sample Condition $n_1 = n_2 = n_3 = n_4 = n_5 = 10$**

The next sample explored was $n_1=n_2=n_3=n_4=n_5=10$ in which one group received treatment of each shift size. The same trends in power differential existed under the uniform distribution as with the normal distribution, with the ANOVA and approximate randomization ANOVA outperforming the Kruskal-Wallis at every treatment level.

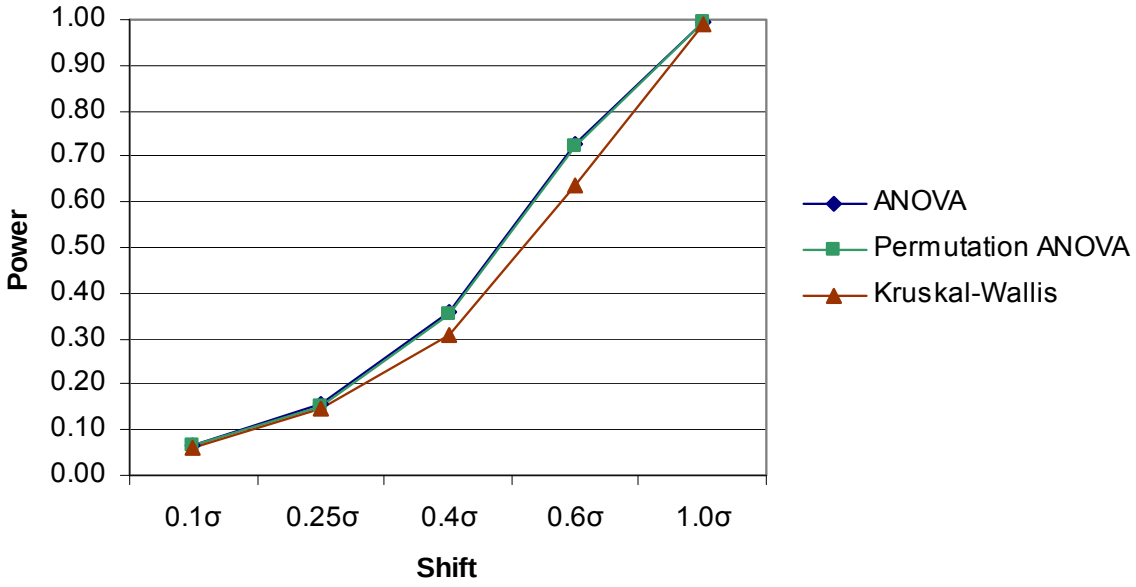


*Figure 20.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2=n_3=n_4=n_5(tr)=10$.

The ANOVA demonstrated the highest power, albeit very modest, on all degrees of shift. The ANOVA ranged from a power of .0650 at 0.1σ and .3562 at 0.4σ, to .9973 at 1.0σ with the approximate randomization ANOVA showing very similar, though mostly lower, levels. The Kruskal-Wallis ranged from .0626 at 0.1σ and .3060 at 0.4σ, to .9885 at 1.0σ. The Kruskal-Wallis was outperformed by the ANOVA by nearly 9% at the 0.6σ shift before demonstrating more similar power results at 1.0σ.

For the two and three treatment group conditions, the relationship of the three tests with each other very strongly resembled that of the one treatment group condition, with the difference being the pace at which the power levels increased with each shift. The ANOVA showed a modest power advantage over the approximate

randomization ANOVA, and both tests demonstrated up to 6% higher power than the Kruskal-Wallis.

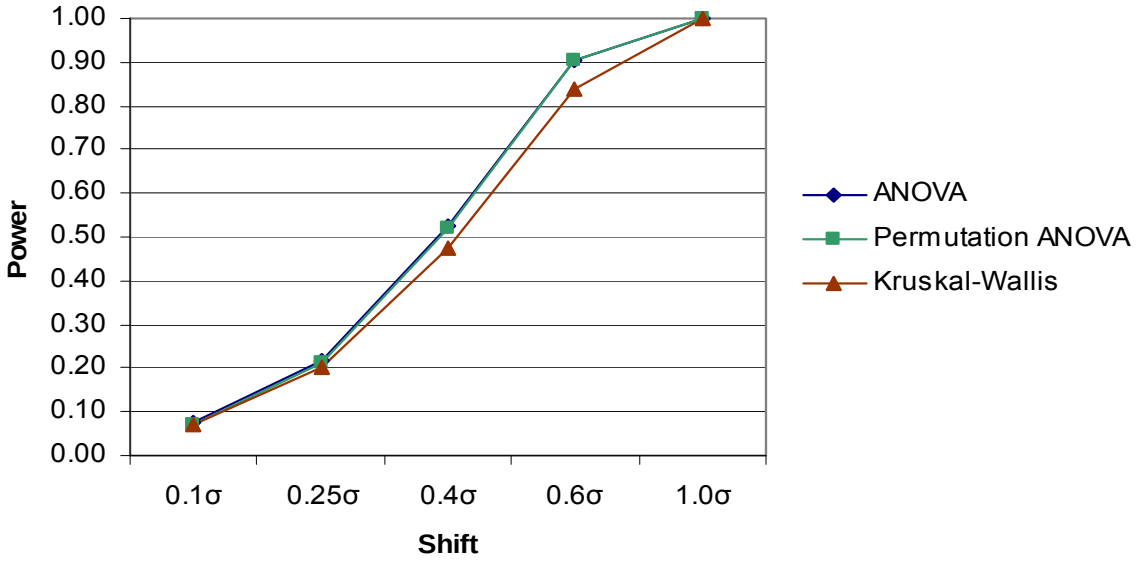


*Figure 21.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2=n_3=n_4(tr)=n_5(tr)=10$.

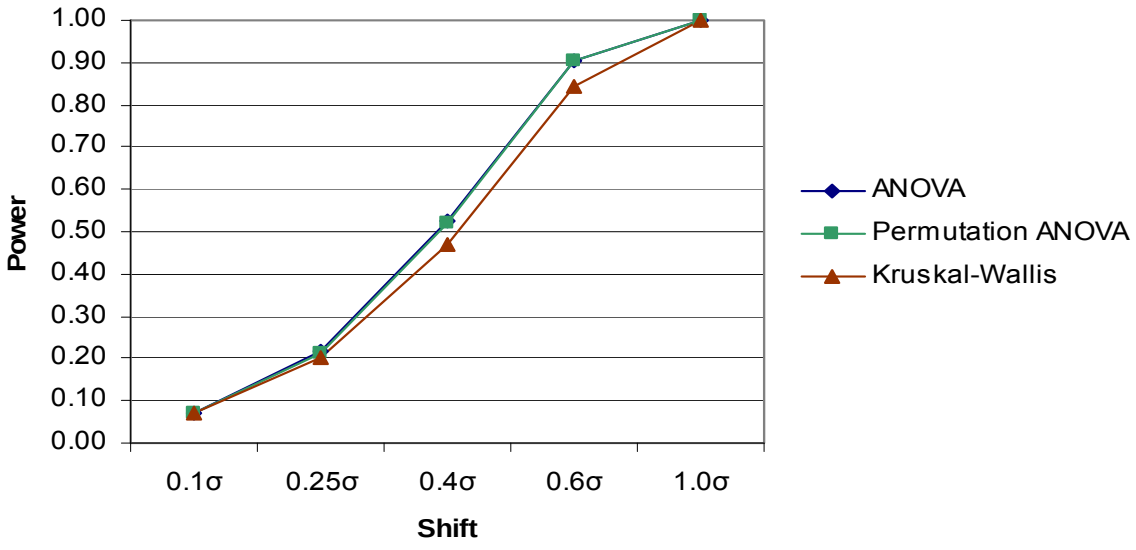


*Figure 22.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2=n_3(tr)=n_4(tr)=n_5(tr)=10$.

For the four treatment group condition, all patterns remained essentially the same as the one treatment group condition.

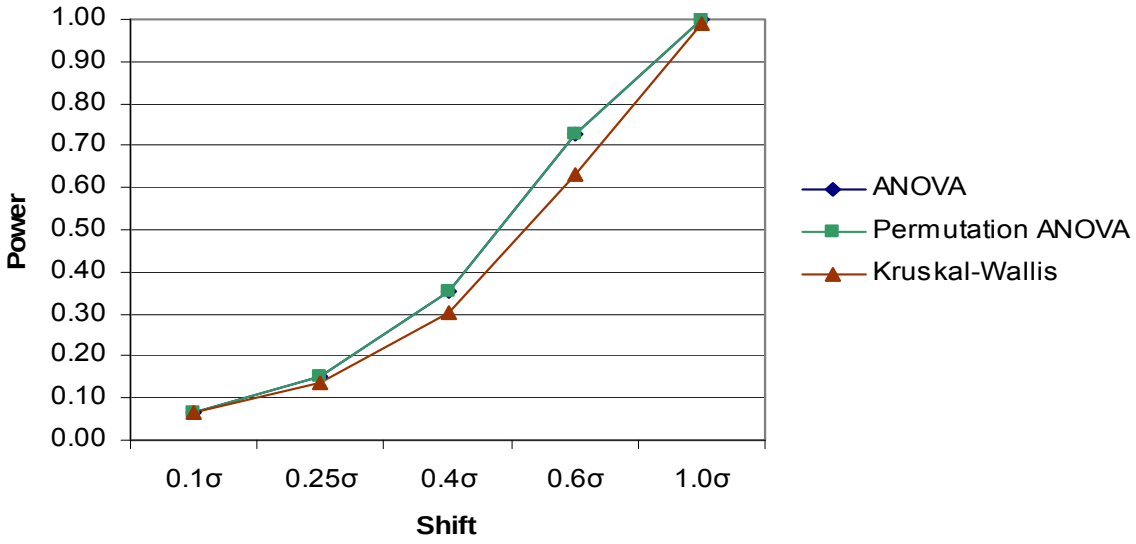


*Figure 23.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2(tr)=n_3(tr)=n_4(tr)=n_5(tr)=10$.

**Sample $n_1 = n_2 = n_3 = n_4 = n_5 = 30$**

The power differential trends that existed between treatment group conditions under the normal distribution remained under the uniform distribution, with the ANOVA and approximate randomization ANOVA outperforming the Kruskal-Wallis at all but the 1.0σ treatment effect.

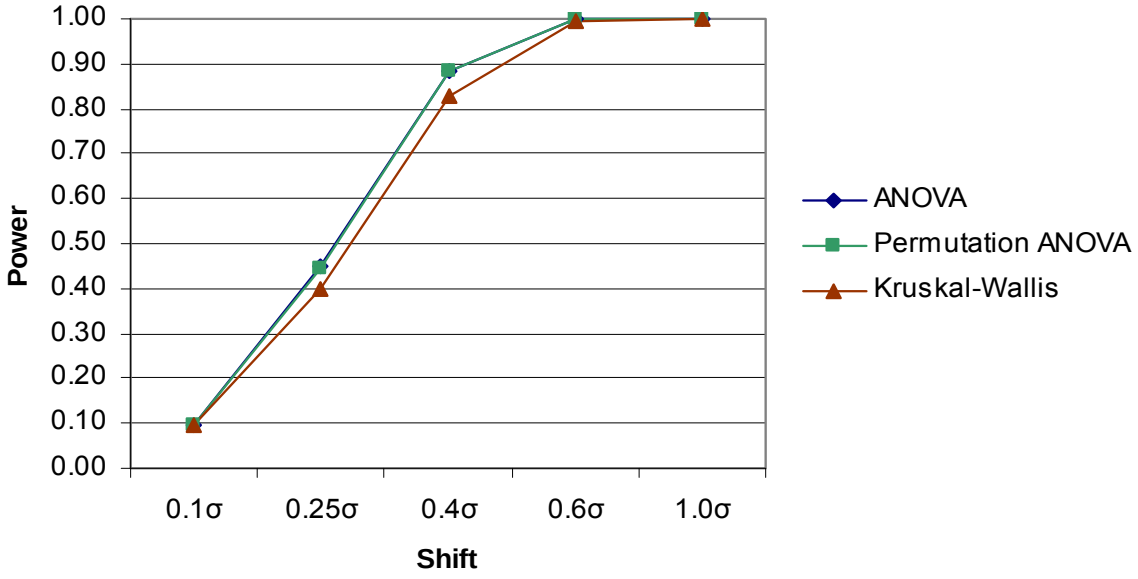


*Figure 24.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2=n_3=n_4=n_5(tr)=30$.

The ANOVA demonstrated the highest power or power equal to the approximate randomization ANOVA (and Kruskal-Wallis at 1.0σ) on all degrees of shift. The ANOVA ranged from a power of .0984 at 0.1σ and .8862 at 0.4σ, to 1.0000 at 1.0σ with the approximate randomization ANOVA showing very similar levels. The Kruskal-Wallis ranged from .0947 at 0.1σ and .8261 at 0.4σ, to 1.0000 at 1.0σ. The Kruskal-Wallis was outperformed by the ANOVA by just over 6% at the 0.4σ shift before demonstrating more similar power results at 0.6σ and 1.0σ. The four treatment group condition revealed nearly identical curves.

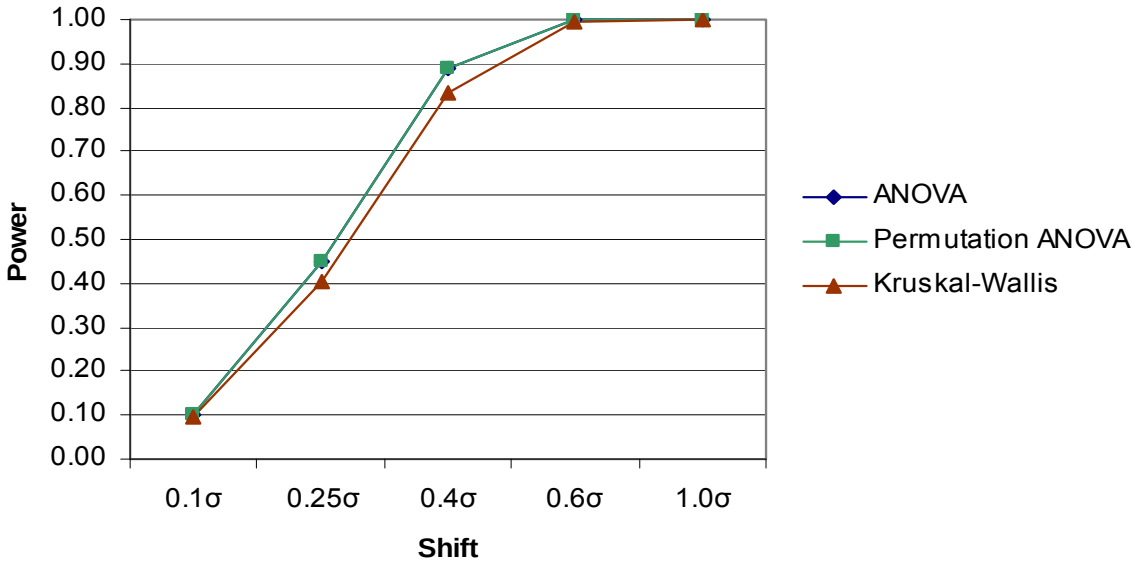


*Figure 25.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2(tr)=n_3(tr)=n_4(tr)=n_5(tr)=30$.

For the two- and three treatment group conditions, the patterns also remained constant, with the ANOVA and approximate randomization ANOVA demonstrating nearly identical power curves, and both tests demonstrating up to a 5% higher power than the Kruskal-Wallis (0.25σ).

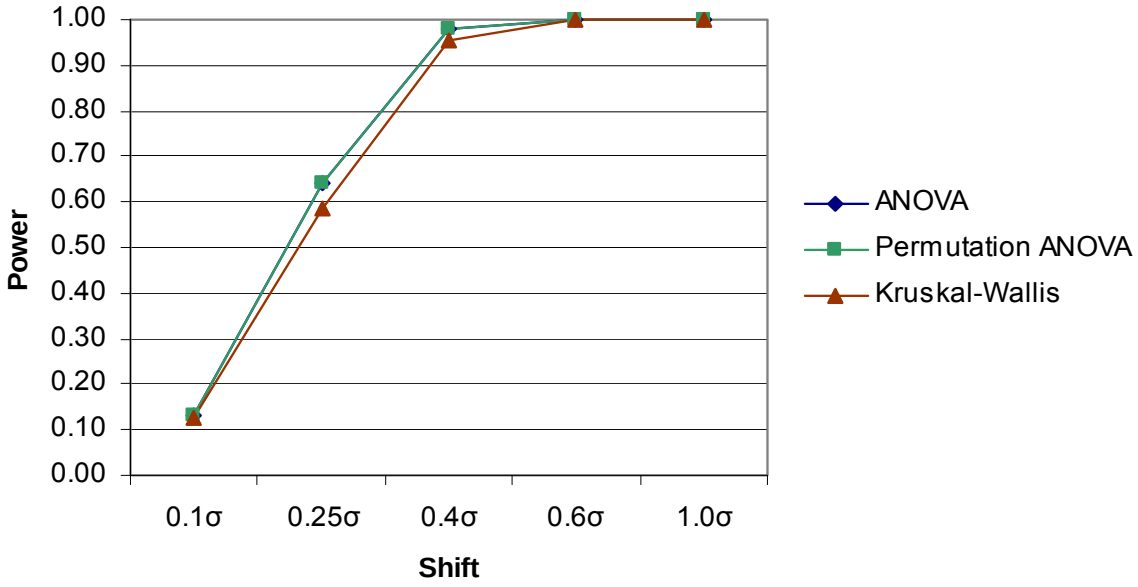


*Figure 26.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2=n_3=n_4(tr)=n_5(tr)=30$.

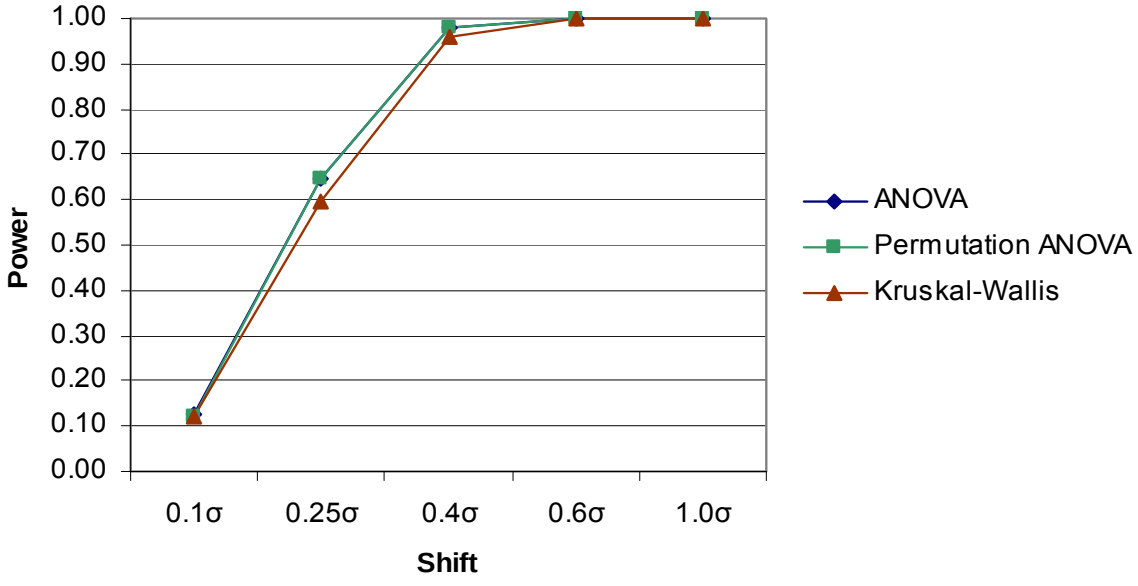


*Figure 27.* Shift vs. Power in the uniform distribution for sample condition $n_1=n_2=n_3(tr)=n_4(tr)=n_5(tr)=30$.

**Chi-Square (df=2) Distribution**

The first sample explored was $n_1=n_2=n_3=10$, in which one group received treatment. For every effect size, the ANOVA was outperformed by the approximate randomization ANOVA, which in turn, was heavily outperformed by the Kruskal-Wallis.

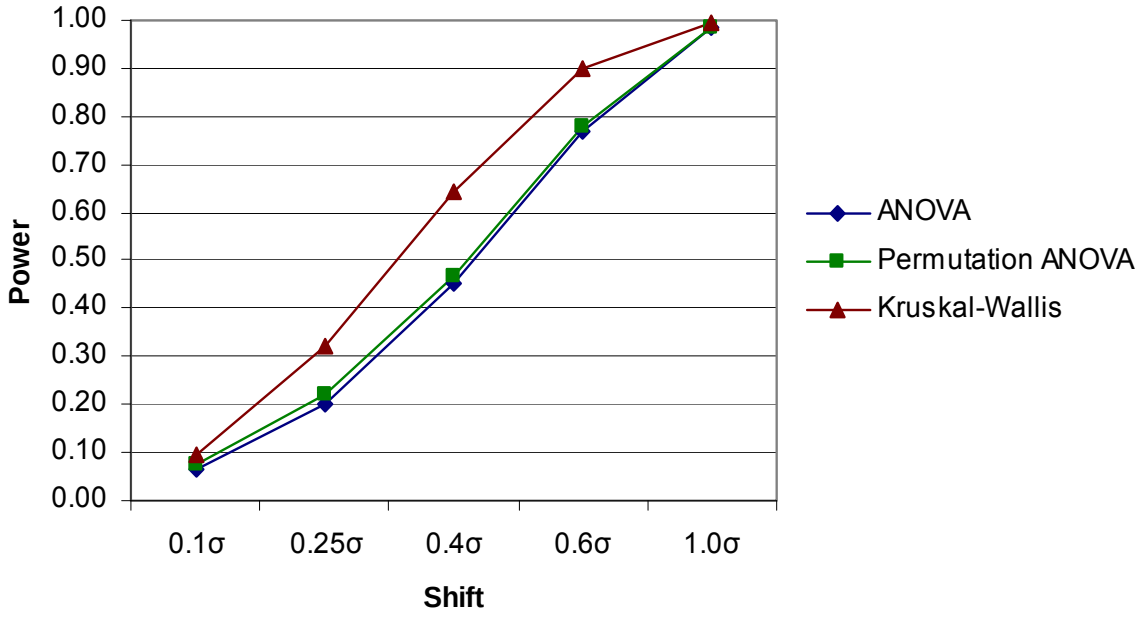


*Figure 28.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2=n_3(tr)=10$.

The ANOVA exhibited a power of .0642 at 0.1σ, .4512 at 0.4σ, and .9859 at 1.0σ, not reaching a power above .50 until the 0.6σ shift (.7700). At every treatment level, the approximate randomization ANOVA outperformed the ANOVA by 1-2%. Starting at the 0.25σ shift, a stark separation developed between the Kruskal-Wallis and its counterparts, and this separation increased dramatically through the 0.4σ shift until converging again around .99 at 1.0σ. The largest power differential for the

Kruskal-Wallis was a nearly 18% power advantage over both of the other tests at the 0.4σ shift.

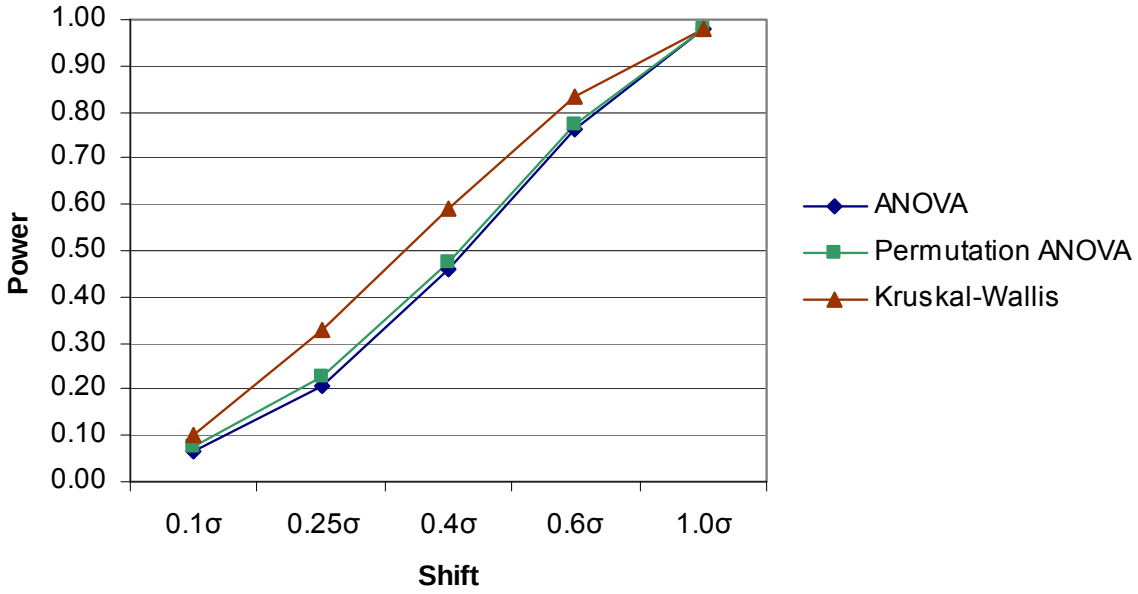


*Figure 29.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2(tr)=n_3(tr)=10$.

For the two treatment group condition, power analysis results yielded similar trends for each shift size. The only notable exception is that at the 0.4σ and 0.6σ shifts, the power of the Kruskal-Wallis was 5-8% lower than it was at the same effect size for the one treatment group condition and was 1.5% lower at the 1.0σ shift.

**Sample $n_1 = n_2 = n_3 = 30$**

For the chi-square (df=2) distribution in the one treatment group condition, the Kruskal-Wallis (.8303 at 0.25σ) quickly showed a significant power advantage over both the ANOVA and approximate randomization ANOVA, which registered power results of .5042 and .5132, respectively.

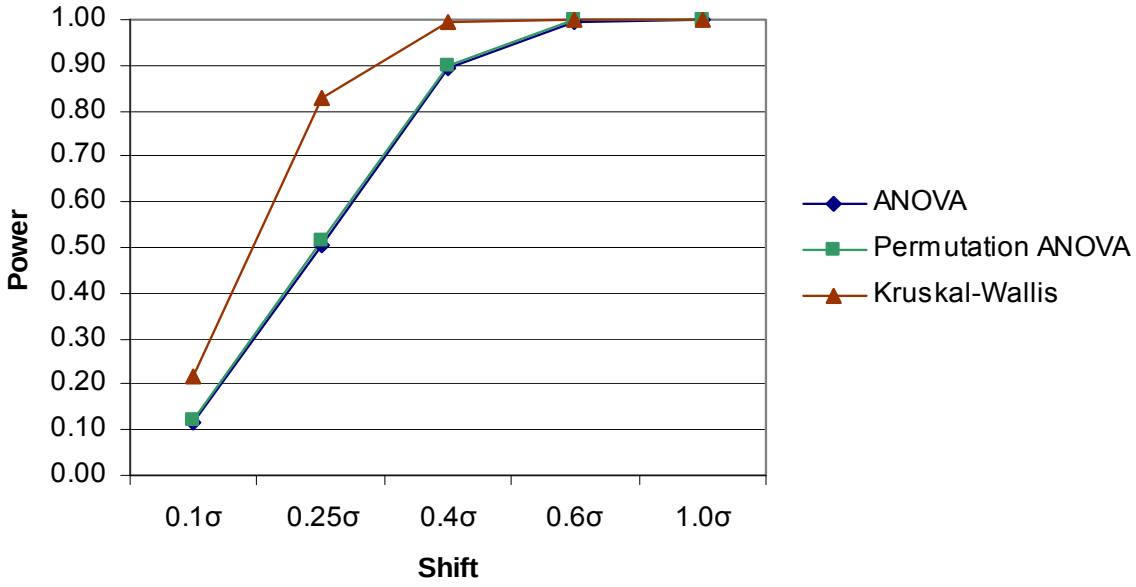


*Figure 30.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2=n_3(tr)=30$.

At the 0.4σ shift, the ANOVA demonstrated a power of .8943, the approximate randomization ANOVA showed a power of .8985, and the Kruskal-Wallis, .9937. All tests achieved a power of 1.0 at the 1.0σ shift. The results of the two treatment group condition were nearly identical, except the power of the Kruskal-Wallis was 5% lower at 0.25σ than the same effect size for the one treatment group condition.

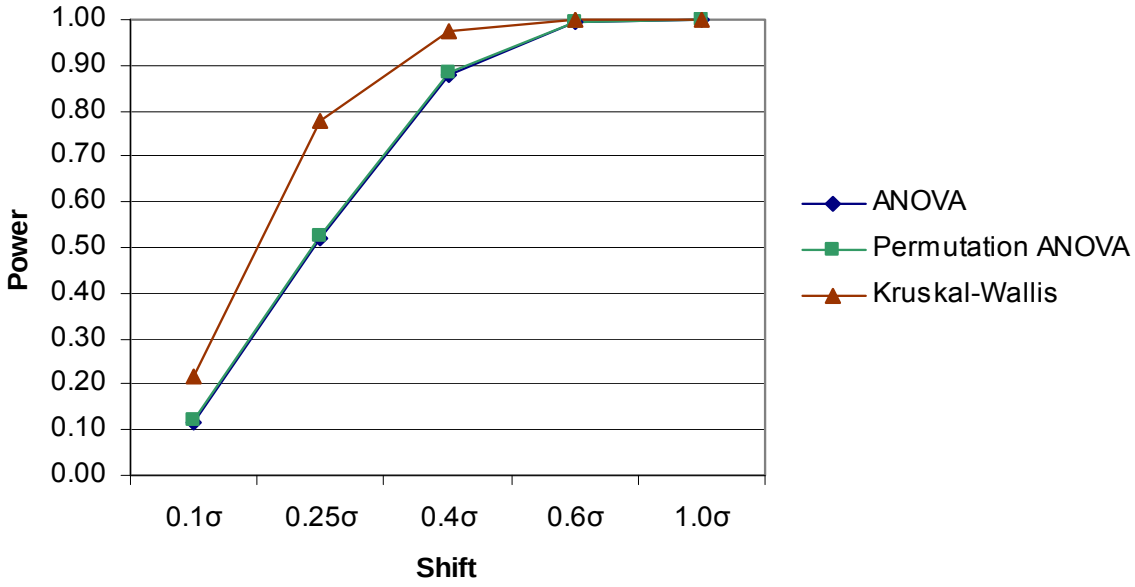


*Figure 31.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2(tr)=n_3(tr)=30$.

**Sample $n_1 = n_2 = n_3 = n_4 = n_5 = 10$**

The next sample explored was $n_1=n_2=n_3=n_4=n_5=10$ in which one group received treatment of each shift size. As with the k=3 condition, the ANOVA was outperformed by the approximate randomization ANOVA, and the Kruskal-Wallis exhibited significantly higher power than both other tests.

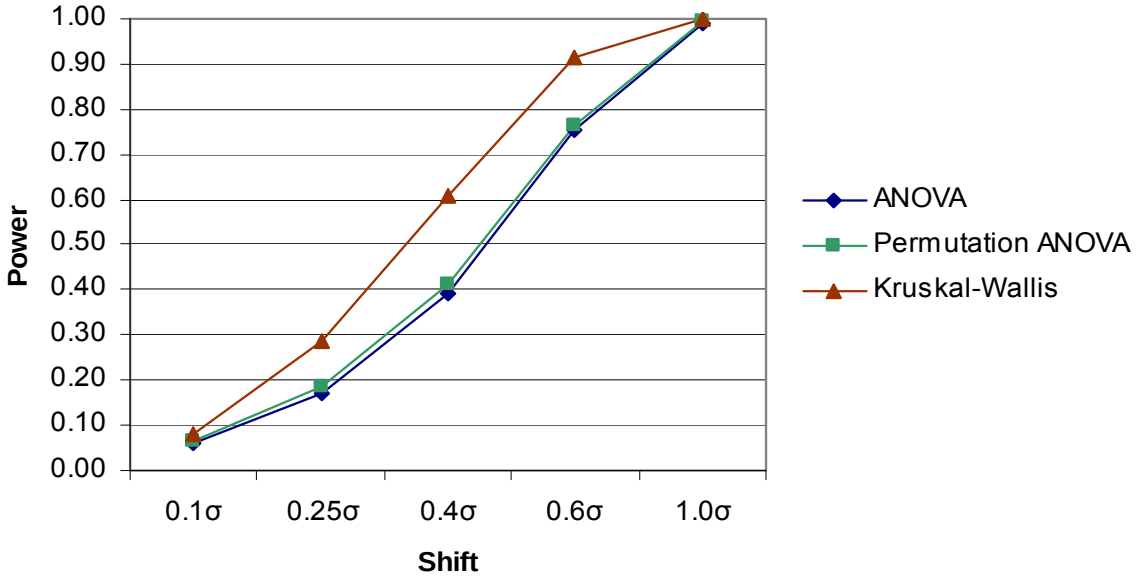


*Figure 32.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2=n_3=n_4=n_5(tr)=10$.

The ANOVA exhibited a power of .0581 at 0.1σ, to .3932 at 0.4σ, and .9924 at 1.0σ, not reaching a power above .50 until the 0.6σ shift (.7532). The Kruskal-Wallis outperformed both other tests at every level, with the largest gap of over 21% demonstrated at the 0.4σ. At this same shift, the Kruskal-Wallis also demonstrated a 19% power advantage over the approximate randomization ANOVA. The approximate randomization ANOVA was consistently more powerful than the ANOVA, though the advantage was modest overall. These patterns also remained for the two- and three treatment group conditions. At the 0.4σ shift, the Kruskal-Wallis exhibited a 19%-21% power advantage over either of the ANOVA tests, with that gap closing incrementally as shift approached 1.0σ.

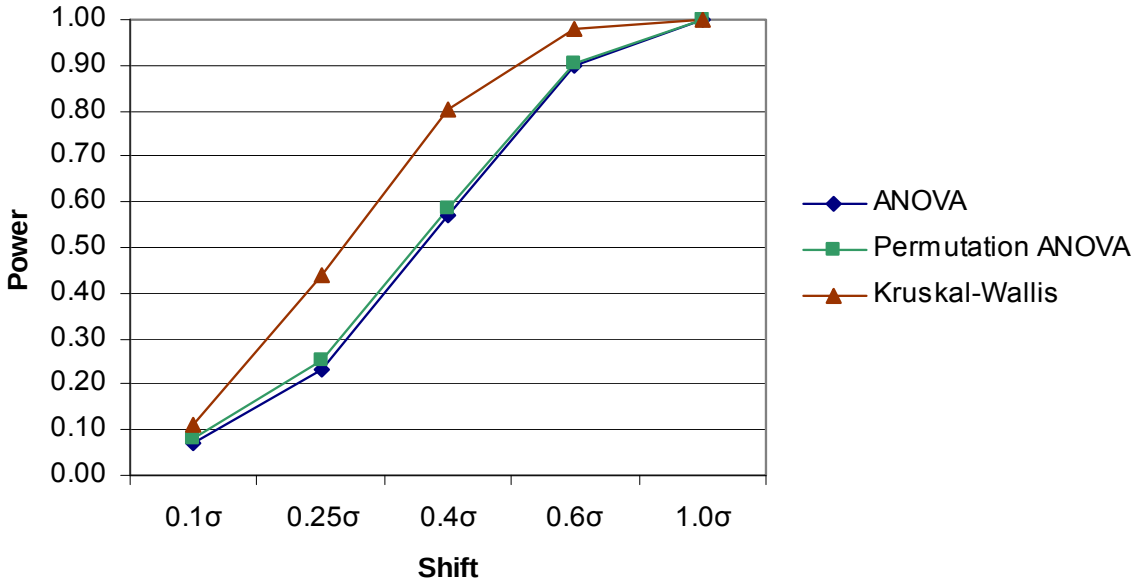


*Figure 33.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2=n_3=n_4(tr)=n_5(tr)=10$.

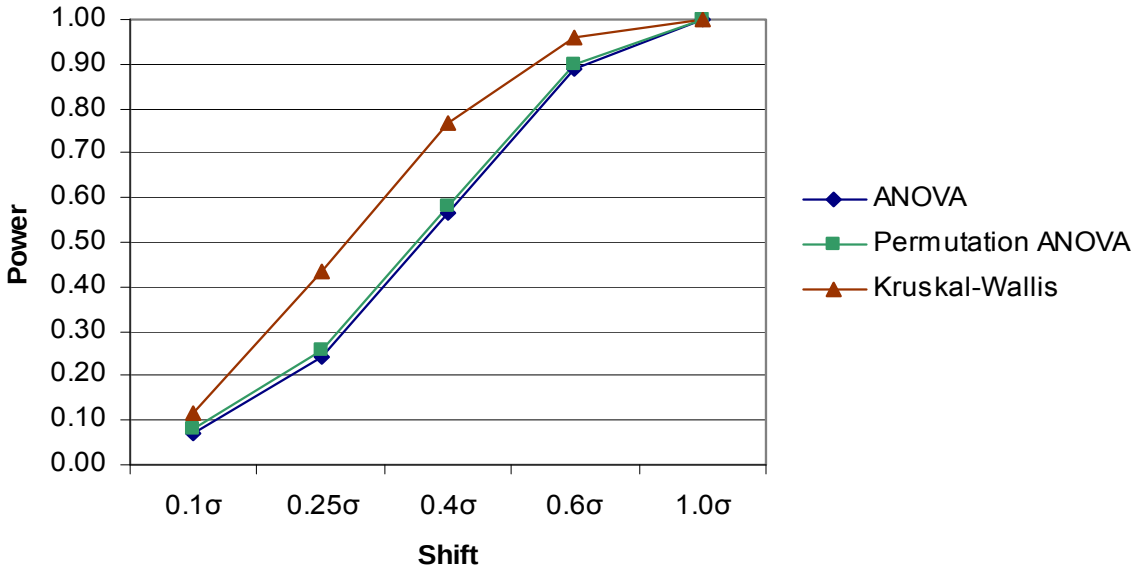


*Figure 34.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2=n_3(tr)=n_4(tr)=n_5(tr)=10$.

For the four treatment groups condition, all patterns remained essentially the same as the one treatment group condition, but there was one difference in relation to the chi-square (df=2) distribution. The Kruskal-Wallis demonstrated a significant power loss when compared with its power curve with one treatment group condition. It demonstrated a decrease in power at 0.4σ (.0584) and 0.6σ (.1142) shifts. There was also a decrease in power at the 1.0σ shift, though that drop was much smaller in magnitude (.0231).

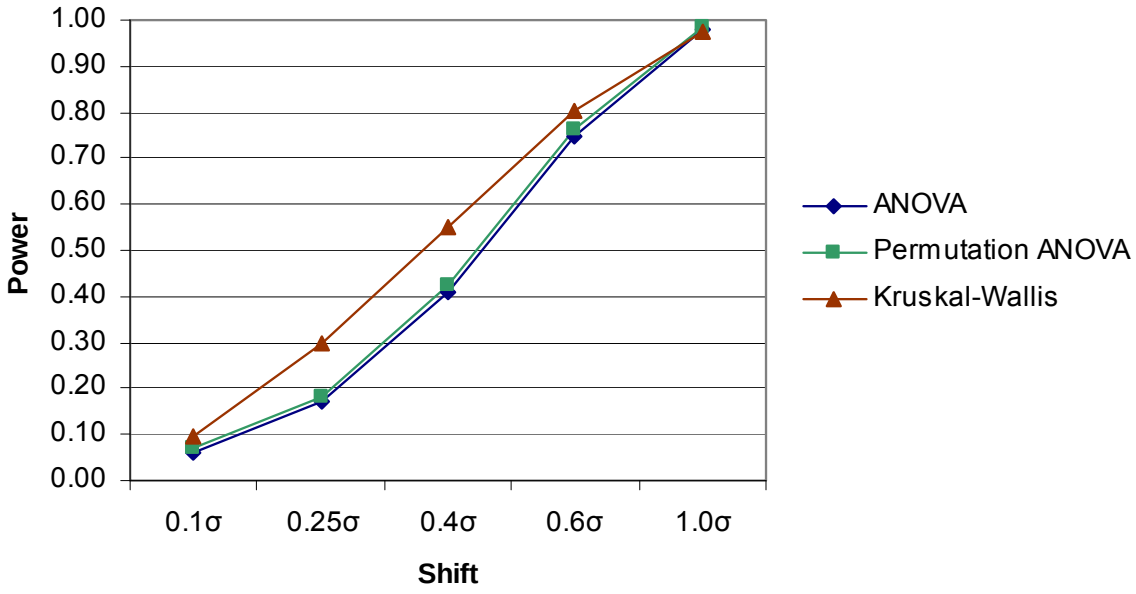


*Figure 35.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2(tr)=n_3(tr)=n_4(tr)=n_5(tr)=10$.

**Sample $n_1 = n_2 = n_3 = n_4 = n_5 = 30$**

For the one treatment group condition under the chi-square (df = 2) distribution, the Kruskal-Wallis exhibited higher power than both the ANOVA and

approximate randomization ANOVA, with the power advantage reaching 36% at 0.25σ.

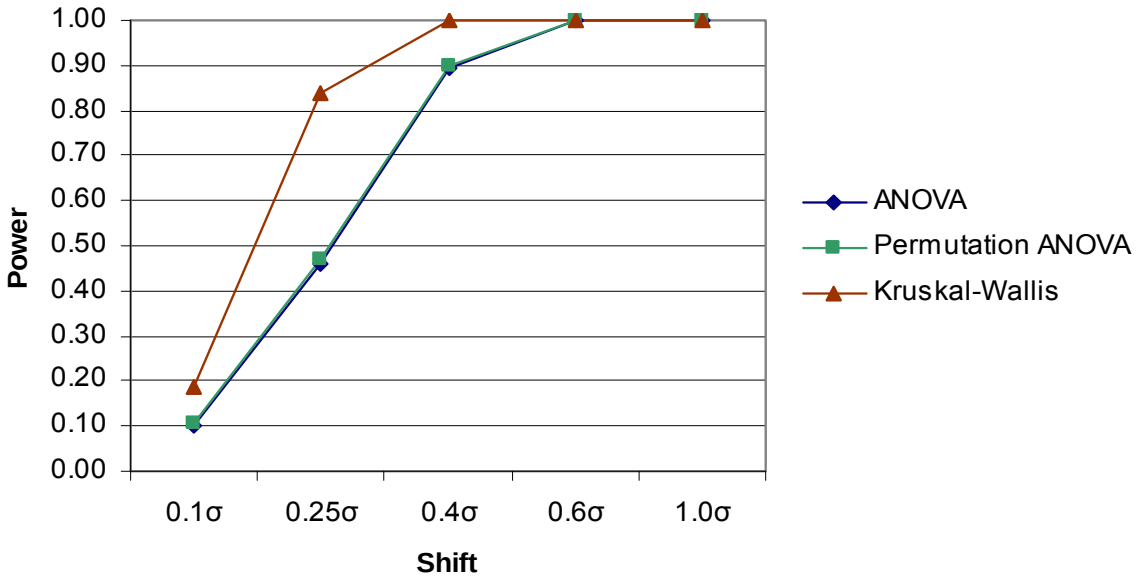


*Figure 36.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2=n_3=n_4=n_5(tr)=30$.

The ANOVA exhibited a power of .1016 at 0.1σ, to .8964 at 0.4σ, a very similar power curve to the approximate randomization ANOVA, though the approximate randomization ANOVA had a very slight edge in power until 1.0σ. The Kruskal-Wallis demonstrated power of .1857 at 0.1σ and by 0.25σ had reached a power of .8372. As with other distributions, there were many similarities between the one- and four treatment group conditions for the chi-square (df=2) distribution. One noteworthy detail was the power drop-off for the Kruskal-Wallis at 0.25σ (.0874) and 0.4σ (.0272) shifts of the four treatment group condition when compared to its power properties under the one treatment group condition. The Kruskal-Wallis also benefited in this comparison from a slight increase in power at 0.1σ.

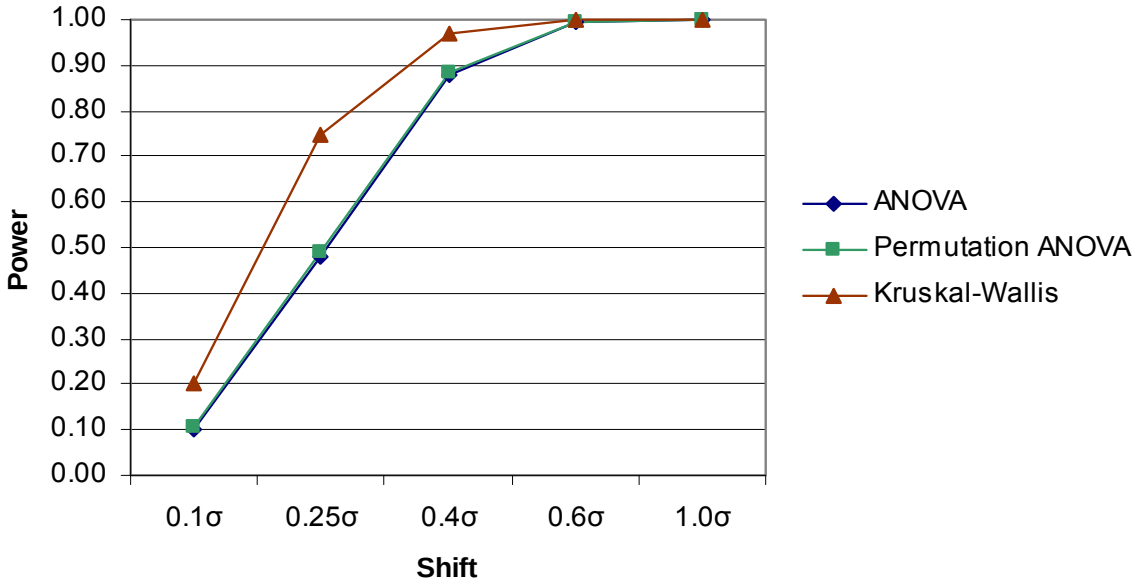


*Figure 37.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2(tr)=n_3(tr)=n_4(tr)=n_5(tr)=30$.

Congruent patterns also held for the two treatment and three treatment groups. At the 0.25σ shift, the Kruskal-Wallis exhibited a 26%-28% power advantage over either of the other tests, with that gap closing as shift approached 0.4σ.

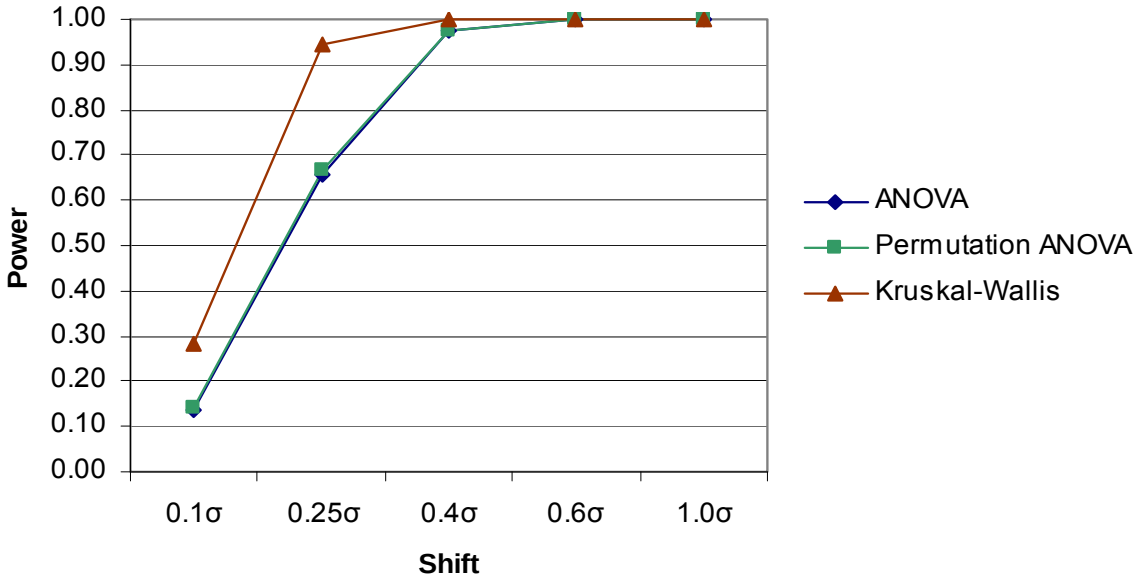


*Figure 38.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2=n_3=n_4(tr)=n_5(tr)=30$.

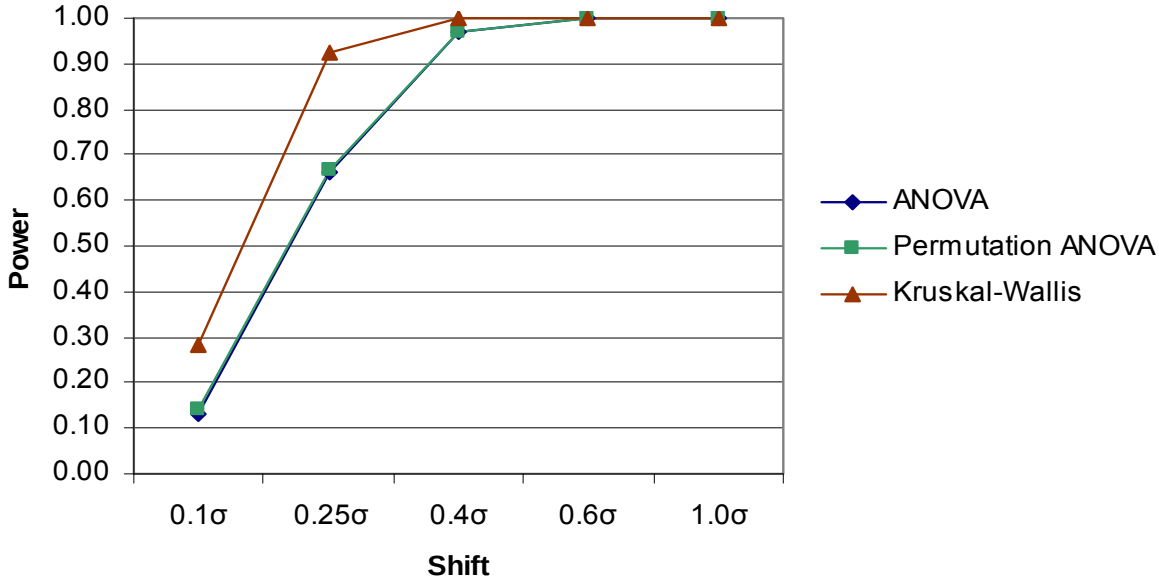


*Figure 39.* Shift vs. Power in the chi-square (df=2) distribution for sample condition $n_1=n_2=n_3(tr)=n_4(tr)=n_5(tr)=30$.

Table 5

*Rejections of the null under treatment condition for $n_1=n_2=n_3(tr)=10$*

| Effect Size | Distribution | ANOVA | | Approximate randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|---|
| | | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| 0.1σ | Normal | 0.05405 | 0.01115 | 0.05410 | 0.01140 | 0.05450 | 0.01150 |
| | Uniform | 0.06835 | 0.01690 | 0.06665 | 0.01530 | 0.06570 | 0.01490 |
| | Chi-Square | 0.06425 | 0.01355 | 0.07365 | 0.01775 | 0.09435 | 0.02410 |
| 0.25σ | Normal | 0.07410 | 0.01690 | 0.07460 | 0.01785 | 0.07385 | 0.01715 |
| | Uniform | 0.16550 | 0.05200 | 0.16200 | 0.04815 | 0.15395 | 0.04680 |
| | Chi-Square | 0.20020 | 0.06700 | 0.21885 | 0.08150 | 0.32100 | 0.12330 |
| 0.4σ | Normal | 0.12685 | 0.03615 | 0.12705 | 0.03665 | 0.12480 | 0.03600 |
| | Uniform | 0.37750 | 0.16345 | 0.37340 | 0.15445 | 0.34025 | 0.14160 |
| | Chi-Square | 0.45120 | 0.22260 | 0.46975 | 0.24575 | 0.64375 | 0.36460 |
| 0.6σ | Normal | 0.24315 | 0.08930 | 0.24440 | 0.09050 | 0.22975 | 0.08260 |
| | Uniform | 0.74345 | 0.47780 | 0.74045 | 0.46700 | 0.66750 | 0.40570 |
| | Chi-Square | 0.77000 | 0.54325 | 0.78110 | 0.56495 | 0.90160 | 0.71055 |
| 1.0σ | Normal | 0.58545 | 0.31715 | 0.58540 | 0.31735 | 0.55720 | 0.30050 |
| | Uniform | 0.99685 | 0.97640 | 0.99690 | 0.97680 | 0.98775 | 0.94050 |
| | Chi-Square | 0.98590 | 0.94020 | 0.98725 | 0.94425 | 0.99690 | 0.97565 |

Table 6

*Rejections of the null under treatment condition for $n_1=n_2(tr)=n_3(tr)=10$*

| Effect Size | Distribution | ANOVA | | Approximate randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|---|
| | | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| 0.1σ | Normal | 0.05335 | 0.01180 | 0.05390 | 0.01200 | 0.05335 | 0.01215 |
| | Uniform | 0.06840 | 0.01630 | 0.06560 | 0.01495 | 0.06525 | 0.01505 |
| | Chi-Square | 0.06630 | 0.01415 | 0.07555 | 0.01800 | 0.10115 | 0.02885 |
| 0.25σ | Normal | 0.08095 | 0.01995 | 0.08205 | 0.02025 | 0.07915 | 0.02065 |
| | Uniform | 0.17080 | 0.05685 | 0.16765 | 0.05255 | 0.16050 | 0.05135 |
| | Chi-Square | 0.20945 | 0.06985 | 0.22695 | 0.08470 | 0.32655 | 0.15300 |
| 0.4σ | Normal | 0.12565 | 0.03650 | 0.12640 | 0.03765 | 0.11990 | 0.03660 |
| | Uniform | 0.37770 | 0.16125 | 0.37280 | 0.15255 | 0.34140 | 0.13860 |
| | Chi-Square | 0.45820 | 0.23330 | 0.47695 | 0.26280 | 0.58965 | 0.37020 |
| 0.6σ | Normal | 0.23760 | 0.08685 | 0.23765 | 0.08800 | 0.22415 | 0.08495 |
| | Uniform | 0.73880 | 0.46790 | 0.73550 | 0.45490 | 0.66315 | 0.39645 |
| | Chi-Square | 0.76425 | 0.55720 | 0.77510 | 0.58840 | 0.83520 | 0.65805 |
| 1.0σ | Normal | 0.58215 | 0.31760 | 0.58275 | 0.31795 | 0.55615 | 0.29990 |
| | Uniform | 0.99775 | 0.97450 | 0.99760 | 0.97385 | 0.98800 | 0.93650 |
| | Chi-Square | 0.97925 | 0.93335 | 0.98040 | 0.94015 | 0.98055 | 0.93275 |

Table 7

*Rejections of the null under treatment condition for $n_1=n_2=n_3(tr)=30$*

| Effect Size | Distribution | ANOVA | | Approximate randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|---|
| | | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| 0.1σ | Normal | 0.06325 | 0.01430 | 0.06385 | 0.01425 | 0.06340 | 0.01405 |
| | Uniform | 0.11035 | 0.03045 | 0.10975 | 0.02955 | 0.10635 | 0.02760 |
| | Chi-Square | 0.11510 | 0.03240 | 0.11970 | 0.03680 | 0.21865 | 0.07800 |
| 0.25σ | Normal | 0.14850 | 0.04445 | 0.14885 | 0.04565 | 0.14240 | 0.04405 |
| | Uniform | 0.48475 | 0.24570 | 0.48380 | 0.24245 | 0.44405 | 0.21570 |
| | Chi-Square | 0.50425 | 0.27090 | 0.51320 | 0.28760 | 0.83030 | 0.60845 |
| 0.4σ | Normal | 0.33020 | 0.14215 | 0.33070 | 0.14495 | 0.31385 | 0.13395 |
| | Uniform | 0.88725 | 0.71755 | 0.88665 | 0.71455 | 0.83445 | 0.63680 |
| | Chi-Square | 0.89435 | 0.73550 | 0.89855 | 0.75110 | 0.99370 | 0.96545 |
| 0.6σ | Normal | 0.65525 | 0.40575 | 0.65600 | 0.40840 | 0.63265 | 0.38630 |
| | Uniform | 0.99910 | 0.99155 | 0.99910 | 0.99130 | 0.99465 | 0.97180 |
| | Chi-Square | 0.99720 | 0.98320 | 0.99750 | 0.98560 | 1.00000 | 0.99990 |
| 1.0σ | Normal | 0.98260 | 0.92745 | 0.98230 | 0.92845 | 0.97680 | 0.91535 |
| | Uniform | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| | Chi-Square | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |

Table 8

*Rejections of the null under treatment condition for $n_1=n_2(tr)=n_3(tr)=30$*

| Effect Size | Distribution | ANOVA | | Approximate randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|---|
| | | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| 0.1σ | Normal | 0.06410 | 0.01380 | 0.06455 | 0.01375 | 0.06425 | 0.01300 |
| | Uniform | 0.11170 | 0.03020 | 0.11050 | 0.02980 | 0.10800 | 0.02955 |
| | Chi-Square | 0.11540 | 0.03120 | 0.12065 | 0.03510 | 0.21505 | 0.08360 |
| 0.25σ | Normal | 0.15025 | 0.04725 | 0.15080 | 0.04750 | 0.14420 | 0.04410 |
| | Uniform | 0.47870 | 0.24760 | 0.47810 | 0.24425 | 0.43755 | 0.21965 |
| | Chi-Square | 0.51805 | 0.28625 | 0.52735 | 0.30375 | 0.77980 | 0.58230 |
| 0.4σ | Normal | 0.33085 | 0.14330 | 0.33095 | 0.14475 | 0.31620 | 0.13670 |
| | Uniform | 0.89445 | 0.72420 | 0.89350 | 0.72215 | 0.84165 | 0.64470 |
| | Chi-Square | 0.87980 | 0.73160 | 0.88440 | 0.74635 | 0.97415 | 0.91685 |
| 0.6σ | Normal | 0.65050 | 0.40330 | 0.65150 | 0.40495 | 0.63010 | 0.38275 |
| | Uniform | 0.99910 | 0.99225 | 0.99915 | 0.99180 | 0.99490 | 0.97000 |
| | Chi-Square | 0.99415 | 0.97820 | 0.99495 | 0.98045 | 0.99960 | 0.99730 |
| 1.0σ | Normal | 0.98425 | 0.92750 | 0.98450 | 0.92735 | 0.97830 | 0.91280 |
| | Uniform | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| | Chi-Square | 1.00000 | 0.99995 | 0.99995 | 0.99995 | 1.00000 | 0.99995 |

Table 9

*Rejections of the null under treatment condition for $n_1=n_2=n_3=n_4=n_5(tr)=10$*

| Effect Size | Distribution | ANOVA | | Approximate randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|---|
| | | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| 0.1σ | Normal | 0.05210 | 0.00960 | 0.05235 | 0.00990 | 0.05345 | 0.01070 |
| | Uniform | 0.06500 | 0.01460 | 0.06370 | 0.01350 | 0.06260 | 0.01430 |
| | Chi-Square | 0.05810 | 0.01325 | 0.06480 | 0.01510 | 0.08285 | 0.02030 |
| 0.25σ | Normal | 0.07540 | 0.01695 | 0.07645 | 0.01715 | 0.07580 | 0.01800 |
| | Uniform | 0.15640 | 0.04790 | 0.15370 | 0.04570 | 0.14395 | 0.04280 |
| | Chi-Square | 0.17055 | 0.05735 | 0.18355 | 0.06350 | 0.28420 | 0.09800 |
| 0.4σ | Normal | 0.11825 | 0.03375 | 0.11815 | 0.03350 | 0.11250 | 0.03105 |
| | Uniform | 0.35625 | 0.15480 | 0.35215 | 0.15045 | 0.30605 | 0.12070 |
| | Chi-Square | 0.39325 | 0.18720 | 0.41030 | 0.19920 | 0.60775 | 0.30545 |
| 0.6σ | Normal | 0.21750 | 0.07745 | 0.21790 | 0.07860 | 0.20840 | 0.07295 |
| | Uniform | 0.72700 | 0.47070 | 0.72440 | 0.46525 | 0.63580 | 0.36115 |
| | Chi-Square | 0.75320 | 0.52185 | 0.76395 | 0.53695 | 0.91625 | 0.70205 |
| 1.0σ | Normal | 0.55570 | 0.29890 | 0.55660 | 0.29970 | 0.52070 | 0.26725 |
| | Uniform | 0.99730 | 0.98070 | 0.99715 | 0.97965 | 0.98855 | 0.93565 |
| | Chi-Square | 0.99240 | 0.95910 | 0.99320 | 0.96250 | 0.99955 | 0.99040 |

Table 10

*Rejections of the null under treatment condition for $n_1=n_2=n_3=n_4(tr)=n_5(tr)=10$*

| Effect Size | Distribution | ANOVA | | Approximate randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|---|
| | | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| 0.1σ | Normal | 0.05335 | 0.01155 | 0.05325 | 0.01200 | 0.05410 | 0.01155 |
| | Uniform | 0.07330 | 0.01655 | 0.07105 | 0.01570 | 0.06885 | 0.01530 |
| | Chi-Square | 0.07070 | 0.01645 | 0.07880 | 0.01965 | 0.11220 | 0.03225 |
| 0.25σ | Normal | 0.08430 | 0.01965 | 0.08450 | 0.02040 | 0.08365 | 0.02095 |
| | Uniform | 0.21735 | 0.07480 | 0.21350 | 0.07160 | 0.20265 | 0.06750 |
| | Chi-Square | 0.23355 | 0.08565 | 0.25025 | 0.09295 | 0.43975 | 0.20960 |
| 0.4σ | Normal | 0.15280 | 0.04590 | 0.15350 | 0.04680 | 0.15040 | 0.04635 |
| | Uniform | 0.52550 | 0.27445 | 0.52180 | 0.26785 | 0.47225 | 0.23900 |
| | Chi-Square | 0.56880 | 0.32480 | 0.58670 | 0.33970 | 0.80490 | 0.58250 |
| 0.6σ | Normal | 0.30185 | 0.12325 | 0.30240 | 0.12515 | 0.29115 | 0.11860 |
| | Uniform | 0.90445 | 0.71990 | 0.90295 | 0.71495 | 0.84090 | 0.62795 |
| | Chi-Square | 0.89960 | 0.75155 | 0.90605 | 0.76415 | 0.98155 | 0.91815 |
| 1.0σ | Normal | 0.75565 | 0.51135 | 0.75695 | 0.51150 | 0.73855 | 0.49615 |
| | Uniform | 1.00000 | 0.99985 | 1.00000 | 0.99985 | 0.99985 | 0.99735 |
| | Chi-Square | 0.99905 | 0.99245 | 0.99915 | 0.99330 | 0.99985 | 0.99925 |

Table 11

*Rejections of the null under treatment condition for $n_1=n_2=n_3(tr)=n_4(tr)=n_5(tr)=10$*

| Effect Size | Distribution | ANOVA | | Approximate randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|---|
| | | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| 0.1σ | Normal | 0.05470 | 0.01160 | 0.05445 | 0.01205 | 0.05540 | 0.01150 |
| | Uniform | 0.07145 | 0.01750 | 0.06995 | 0.01655 | 0.06995 | 0.01650 |
| | Chi-Square | 0.07125 | 0.01660 | 0.07905 | 0.01905 | 0.11485 | 0.03215 |
| 0.25σ | Normal | 0.08685 | 0.02135 | 0.08765 | 0.02195 | 0.08390 | 0.02100 |
| | Uniform | 0.21660 | 0.07490 | 0.21370 | 0.07165 | 0.20100 | 0.06845 |
| | Chi-Square | 0.24440 | 0.08985 | 0.25945 | 0.09825 | 0.43450 | 0.22330 |
| 0.4σ | Normal | 0.15430 | 0.04805 | 0.15420 | 0.04840 | 0.15035 | 0.04695 |
| | Uniform | 0.52460 | 0.27340 | 0.52035 | 0.26585 | 0.46865 | 0.23530 |
| | Chi-Square | 0.56705 | 0.32865 | 0.58180 | 0.34865 | 0.77000 | 0.56515 |
| 0.6σ | Normal | 0.31345 | 0.12830 | 0.31340 | 0.12905 | 0.30295 | 0.12500 |
| | Uniform | 0.90380 | 0.73175 | 0.90295 | 0.72635 | 0.84475 | 0.64200 |
| | Chi-Square | 0.89015 | 0.74660 | 0.89655 | 0.75920 | 0.96005 | 0.88195 |
| 1.0σ | Normal | 0.75170 | 0.51120 | 0.75305 | 0.51260 | 0.73630 | 0.49670 |
| | Uniform | 1.00000 | 0.99975 | 1.00000 | 0.99970 | 0.99970 | 0.99735 |
| | Chi-Square | 0.99810 | 0.99175 | 0.99835 | 0.99305 | 0.99940 | 0.99670 |

Table 12

*Rejections of the null under treatment condition for $n_1=n_2(tr)=n_3(tr)=n_4(tr)=n_5(tr)=10$*

| Effect Size | Distribution | ANOVA α = .05 | ANOVA α = .01 | Approximate randomization ANOVA α = .05 | Approximate randomization ANOVA α = .01 | Kruskal-Wallis α = .05 | Kruskal-Wallis α = .01 |
|---|---|---|---|---|---|---|---|
| 0.1σ | Normal | 0.05420 | 0.01120 | 0.05405 | 0.01130 | 0.05405 | 0.01165 |
| | Uniform | 0.06760 | 0.01475 | 0.06575 | 0.01380 | 0.06450 | 0.01445 |
| | Chi-Square | 0.06220 | 0.01335 | 0.07010 | 0.01535 | 0.09490 | 0.02530 |
| 0.25σ | Normal | 0.07365 | 0.01790 | 0.07425 | 0.01855 | 0.07345 | 0.01790 |
| | Uniform | 0.15350 | 0.04895 | 0.15000 | 0.04660 | 0.13885 | 0.04245 |
| | Chi-Square | 0.17125 | 0.05290 | 0.18395 | 0.05930 | 0.29850 | 0.13505 |
| 0.4σ | Normal | 0.11515 | 0.03110 | 0.11510 | 0.03110 | 0.11065 | 0.03010 |
| | Uniform | 0.35490 | 0.15350 | 0.35225 | 0.14850 | 0.30475 | 0.11990 |
| | Chi-Square | 0.40880 | 0.19240 | 0.42560 | 0.20785 | 0.54935 | 0.33520 |
| 0.6σ | Normal | 0.21510 | 0.07475 | 0.21570 | 0.07630 | 0.20160 | 0.06940 |
| | Uniform | 0.72955 | 0.46820 | 0.72770 | 0.46155 | 0.63140 | 0.35340 |
| | Chi-Square | 0.75000 | 0.54510 | 0.76205 | 0.56800 | 0.80200 | 0.61930 |
| 1.0σ | Normal | 0.55280 | 0.30025 | 0.55395 | 0.30255 | 0.51650 | 0.26840 |
| | Uniform | 0.99750 | 0.98000 | 0.99750 | 0.97925 | 0.98920 | 0.93300 |
| | Chi-Square | 0.98155 | 0.94565 | 0.98255 | 0.95090 | 0.97640 | 0.91895 |

Table 13

*Rejections of the null under treatment condition for $n_1=n_2=n_3=n_4=n_5(tr)=30$*

| Effect Size | Distribution | ANOVA | | Approximate randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|---|
| | | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| 0.1σ | Normal | 0.06010 | 0.01360 | 0.06090 | 0.01355 | 0.06005 | 0.01235 |
| | Uniform | 0.09840 | 0.02540 | 0.09735 | 0.02540 | 0.09470 | 0.02455 |
| | Chi-Square | 0.10160 | 0.02685 | 0.10540 | 0.02925 | 0.18570 | 0.05705 |
| 0.25σ | Normal | 0.13165 | 0.03815 | 0.13240 | 0.03860 | 0.12625 | 0.03610 |
| | Uniform | 0.44830 | 0.22125 | 0.44655 | 0.22020 | 0.40100 | 0.18760 |
| | Chi-Square | 0.46105 | 0.23570 | 0.47075 | 0.24620 | 0.83720 | 0.59705 |
| 0.4σ | Normal | 0.29865 | 0.12165 | 0.29940 | 0.12355 | 0.28135 | 0.11355 |
| | Uniform | 0.88625 | 0.71420 | 0.88570 | 0.71345 | 0.82610 | 0.61740 |
| | Chi-Square | 0.89640 | 0.73440 | 0.90085 | 0.74490 | 0.99770 | 0.98165 |
| 0.6σ | Normal | 0.63460 | 0.38500 | 0.63500 | 0.38585 | 0.60825 | 0.35665 |
| | Uniform | 0.99895 | 0.99395 | 0.99895 | 0.99435 | 0.99615 | 0.97640 |
| | Chi-Square | 0.99860 | 0.99005 | 0.99870 | 0.99110 | 1.00000 | 1.00000 |
| 1.0σ | Normal | 0.98260 | 0.93340 | 0.98280 | 0.93310 | 0.97495 | 0.91275 |
| | Uniform | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| | Chi-Square | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |

Table 14

*Rejections under treatment condition for n1=n2=n3=n4(tr)=n5(tr)=30*

| Effect Size | Distribution | ANOVA | | Approximate randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|---|
| | | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| 0.1σ | Normal | 0.07240 | 0.01605 | 0.07270 | 0.01650 | 0.07025 | 0.01760 |
| | Uniform | 0.13015 | 0.04055 | 0.13035 | 0.04060 | 0.12685 | 0.03920 |
| | Chi-Square | 0.13435 | 0.04280 | 0.14005 | 0.04545 | 0.28465 | 0.11650 |
| 0.25σ | Normal | 0.17830 | 0.06110 | 0.17940 | 0.06250 | 0.17545 | 0.05905 |
| | Uniform | 0.63960 | 0.38920 | 0.63955 | 0.38865 | 0.58740 | 0.34205 |
| | Chi-Square | 0.65810 | 0.41945 | 0.66635 | 0.43185 | 0.94580 | 0.83680 |
| 0.4σ | Normal | 0.44160 | 0.21705 | 0.44160 | 0.21855 | 0.42235 | 0.20555 |
| | Uniform | 0.97795 | 0.91600 | 0.97740 | 0.91460 | 0.95350 | 0.86070 |
| | Chi-Square | 0.97460 | 0.91365 | 0.97625 | 0.91790 | 0.99990 | 0.99850 |
| 0.6σ | Normal | 0.82330 | 0.61870 | 0.82520 | 0.61980 | 0.80130 | 0.59435 |
| | Uniform | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 0.99995 | 0.99915 |
| | Chi-Square | 0.99995 | 0.99960 | 0.99995 | 0.99970 | 1.00000 | 1.00000 |
| 1.0σ | Normal | 0.99925 | 0.99435 | 0.99935 | 0.99445 | 0.99900 | 0.99215 |
| | Uniform | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| | Chi-Square | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |

Table 15

*Rejections of the null under treatment condition for $n_1=n_2=n_3(tr)=n_4(tr)=n_5(tr)=30$*

| Effect Size | Distribution | ANOVA | | Approximate randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|---|
| | | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| 0.1σ | Normal | 0.06510 | 0.01480 | 0.06530 | 0.01465 | 0.06190 | 0.01525 |
| | Uniform | 0.12455 | 0.03595 | 0.12310 | 0.03550 | 0.12045 | 0.03490 |
| | Chi-Square | 0.13340 | 0.03845 | 0.13905 | 0.04110 | 0.28095 | 0.11570 |
| 0.25σ | Normal | 0.18660 | 0.06120 | 0.18645 | 0.06190 | 0.17955 | 0.05965 |
| | Uniform | 0.64640 | 0.39840 | 0.64450 | 0.39625 | 0.59405 | 0.35050 |
| | Chi-Square | 0.66025 | 0.42410 | 0.66745 | 0.43645 | 0.92395 | 0.80495 |
| 0.4σ | Normal | 0.43835 | 0.21945 | 0.43905 | 0.22165 | 0.42285 | 0.20650 |
| | Uniform | 0.97945 | 0.92230 | 0.97960 | 0.92110 | 0.95725 | 0.86595 |
| | Chi-Square | 0.97020 | 0.90735 | 0.97140 | 0.91185 | 0.99905 | 0.99385 |
| 0.6σ | Normal | 0.81940 | 0.61670 | 0.81910 | 0.61840 | 0.79815 | 0.58910 |
| | Uniform | 1.00000 | 0.99995 | 1.00000 | 0.99995 | 0.99995 | 0.99905 |
| | Chi-Square | 0.99995 | 0.99895 | 0.99995 | 0.99910 | 1.00000 | 1.00000 |
| 1.0σ | Normal | 0.99940 | 0.99530 | 0.99935 | 0.99530 | 0.99925 | 0.99360 |
| | Uniform | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| | Chi-Square | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |

Table 16

*Rejections of the null under treatment condition for $n_1=n_2(tr)=n_3(tr)=n_4(tr)=n_5(tr)=30$*

| Effect Size | Distribution | ANOVA | | Approximate randomization ANOVA | | Kruskal-Wallis | |
|---|---|---|---|---|---|---|---|
| | | α = .05 | α = .01 | α = .05 | α = .01 | α = .05 | α = .01 |
| 0.1σ | Normal | 0.06125 | 0.01415 | 0.06115 | 0.01475 | 0.06110 | 0.01455 |
| | Uniform | 0.10050 | 0.02650 | 0.09885 | 0.02640 | 0.09750 | 0.02535 |
| | Chi-Square | 0.10210 | 0.02650 | 0.10670 | 0.02885 | 0.20065 | 0.07555 |
| 0.25σ | Normal | 0.13545 | 0.03835 | 0.13490 | 0.03895 | 0.13050 | 0.03660 |
| | Uniform | 0.44920 | 0.22605 | 0.44790 | 0.22530 | 0.40380 | 0.19175 |
| | Chi-Square | 0.47970 | 0.25455 | 0.48770 | 0.26355 | 0.74980 | 0.54960 |
| 0.4σ | Normal | 0.30325 | 0.12360 | 0.30335 | 0.12570 | 0.28760 | 0.11605 |
| | Uniform | 0.89020 | 0.72260 | 0.88945 | 0.72015 | 0.83190 | 0.62620 |
| | Chi-Square | 0.87910 | 0.72990 | 0.88290 | 0.73875 | 0.97050 | 0.91425 |
| 0.6σ | Normal | 0.62370 | 0.37955 | 0.62445 | 0.38115 | 0.59400 | 0.35355 |
| | Uniform | 0.99905 | 0.99430 | 0.99895 | 0.99400 | 0.99605 | 0.97515 |
| | Chi-Square | 0.99495 | 0.98350 | 0.99525 | 0.98415 | 0.99950 | 0.99645 |
| 1.0σ | Normal | 0.98355 | 0.93620 | 0.98380 | 0.93740 | 0.97725 | 0.91735 |
| | Uniform | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| | Chi-Square | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |

# CHAPTER 5 DISCUSSION

## Overview

The first section of the discussion contains the findings from the Type I error portion of the study. The power results will then be explored for the different statistical tests performed. Implications of the current study will then be discussed.

## Type I Error

Unlike other studies exploring Type I error properties of parametric and nonparametric tests (e.g., Tomarken & Serlin, 1986), the Type I error analysis for this study were performed within the framework of the experiment rather than as a separate component. Therefore, the error rates are based on 20,000 repetitions of each test, a number of iterations that is smaller than many others in published research (e.g., Weber & Sawilowsky, 2009). Similar patterns were observed at $\alpha$=.01 than were observed at $\alpha$=.05, with the latter to be discussed in more detail.

For data sampled from the normal distribution, the ANOVA demonstrated the smallest error rate under the $n_1=n_2=n_3=10$ sample condition, and the error rates as a whole were at or around the 5% level for all tests. The error rates were within the expected range according to Bradley's (1978) conservative criteria of robustness, where he asserted that a range of .9$\alpha$ to 1.1$\alpha$ constitutes a conservative limit for robustness, while a range of .5$\alpha$ to 1.5$\alpha$ is to be considered the most liberal limit.

Data sampled from the uniform distribution showed more diverse rejection rates than under the normal condition, but an expected pattern emerged. The ANOVA demonstrated a decrease in rejection rate from 5.12% on the smallest sample to 4.90% on the largest sample, with a gradual decrease at each increase in

sample. All three tests demonstrated robustness well within the conservative boundaries.

For data sampled from the chi-square distribution (df=2), a difference did emerge between tests, and similar to other findings (e.g., Tomarken & Serlin, 1986), the ANOVA demonstrated rejection rates beyond the conservative limits of robustness at all but the largest sample condition. Both the approximate randomization ANOVA and the Kruskal-Wallis maintained their robustness under the chi-square (df=2) distribution.

None of the error rates were surprising considering the past research on the matter. Sawilowsky & Blair (1992) reported that the t statistic was robust to departures from normality in situations such as equal sample sizes and samples approaching 30 or more, both conditions that appeared to have a rehabilitating effect on the ANOVA's rejection rates under the chi-square (df=2) distribution. That the other tests maintained robustness under the three distributions examined was expected considering their nonparametric nature.

## Comparative Statistical Power

### Small Sample Conditions

Regardless of the number of groups explored, results from the samples consisting of n=10 were very congruent. Not surprisingly, under the normal distribution, the ANOVA and approximate randomization ANOVA demonstrated nearly identical power. The Kruskal-Wallis trailed in power minimally throughout the range of shifts in location (see Figure 4). Research has indicated that under conditions of normality and equal sample size, the ANOVA accomplishes superior

power to nonparametric alternatives (Zimmerman & Zumbo, 1990a), and for the current n=10 conditions, that finding was replicated. Also confirmed in these findings was this comparative superiority is minimal in comparison with nonparametric alternatives (Blair & Higgins, 1985; Sawilowsky, 1990).

In exploring the results obtained for data sampled from the uniform distribution, the power curve for the ANOVA and approximate randomization ANOVA were also nearly identical, exhibiting modestly higher power (most notably at the 0.4σ and the 0.6σ magnitudes of shift in location) than the Kruskal-Wallis. This power advantage closed at the 1.0σ effect size. The Kruskal-Wallis test, however, demonstrated a dramatic power advantage over the ANOVA tests for data sampled from the chi-square (df=2) distribution, with these advantages reaching as much as 18-28%. The approximate randomization ANOVA did appear to rehabilitate the power loss of the ANOVA under the chi-square (df=2) condition, although slightly (2% at most). These power curves were also consistent at the α=.01 level, with the same patterns emerging in test comparisons.

**Large Sample Conditions**

Regardless of the number of groups in the sample, the results from the n=30 samples were also highly congruent. In fact, the patterns demonstrated by the larger samples follow those of the smaller group samples, with the exception of having steeper curves. Under all distributions, the same trends were revealed, with some power advantages slightly increasing or decreasing. Specifically, the Kruskal-Wallis exhibited a power advantage as large as 36% over the ANOVA and approximate randomization ANOVA for data obtained under the chi-square (df=2) distribution (see

Figure 30). These findings are an extension of the assertion of Neave and Granger (1968) that with larger group sizes, the Wilcoxon demonstrated larger power advantages over the t test in non-normal distributions. These power curves were also consistent at the α=.01 level, with the same patterns emerging in test comparisons.

Throughout all treatment group conditions of the n=30 samples, the power curves of all of the tests remained essentially identical under normality. The uniform distribution revealed a slight power advantage for the ANOVA and approximate randomization ANOVA when compared to the Kruskal-Wallis (see Figure 37).

### Multiple Treatment Groups

Although not an intentional goal of the current study, the impact of the number of groups receiving treatment within a sample became of particular interest. Results revealed that the effect of adding multiple treatment groups was modest. Most notably, as the number of treatment groups rose, particularly in the $n_1 = n_2 = n_3 = n_4 = n_5 = 10$ condition, the power advantage the Kruskal-Wallis demonstrated under the chi-square (df=2) decreased modestly. Conversely, as the number of treatment groups increased, the power advantage of the ANOVA and approximate randomization ANOVA over the Kruskal-Wallis under the uniform distribution also increased slightly. No other noticeable patterns were revealed by multiple treatment groups in the small samples.

## Implications

Under conditions of normality, the ANOVA is uniformly most powerful and unbiased, making it the staple analysis tool under those conditions. However, Micceri (1989) noted in his exploration of research literature, that only roughly 3% of

published research had data that approximated a normal distribution. Nonparametric tests such as the Kruskal-Wallis, by virtue of not operating under the assumption of normality, are alternatives to their parametric counterparts when normality does not hold.

Unfortunately, according to Hunter and May (1993) and others, it was posited that degrading data to ranks, as nonparametric tests do, removes valuable information and makes these tests less powerful in some situations. Moreover, as the advent of high speed personal computers has proliferated, the randomization ANOVA has been suggested as an alternative which will rehabilitate ANOVA's lack of robustness with respect to Type I error for departures from population normality, and provide an increase in statistical power under non-normal conditions (e.g., Hunter & May, 1993; Potvin & Roff, 1993), It was similarly presumed that the increased power would be superior to nonparametric rank tests due to their use of actual data rather than ranks (Ludbrook & Dudley, 1998). Assuming this is all true, in situations where the ANOVA is not most powerful, it was contended that the randomization ANOVA would prove to be the most powerful of the three options.

However, Weber and Sawilowsky (2009) previously found with regard to the two independent samples t test, its permutation analog, and the nonparametric Wilcoxon Rank-Sum test (also known as the Mann-Whitney U test), that while the permutation technique is successful in rehabilitating robustness properties, the resulting comparative statistical power generally follows the power spectrum of the parametric test. However, the rank based nonparametric test, which is by definition robust, is actually far superior in terms of its comparative power, specifically for

treatments modeled as a shift in location parameter, The results of this study demonstrated that the Weber and Sawilowsky (2009) results generalize from the k=2 independent groups to the k > 2 independent groups layouts.

## Conclusion

The purpose of this study was to explore the assertion that randomization ANOVA is better than the Kruskal-Wallis as an alternative to ANOVA under conditions on non-normality, and additionally, to explore the relationship of the three tests under conditions of normality and non-normality. In instances where it was more powerful than the ANOVA, the power curve of the approximate randomization ANOVA was only slightly better, and often impossible to distinguish. Most importantly, in situations where the ANOVA could not be said to be most powerful than the Kruskal-Wallis, nor could the approximate randomization ANOVA. In the case of the chi-square (df=2) distribution, when the Kruskal-Wallis was demonstrated to be most powerful, the approximate randomization ANOVA did perform with more power than the ANOVA. To state more clearly, the approximate randomization ANOVA appears to be more powerful than the Kruskal-Wallis only under the same conditions in which ANOVA is more powerful. Within the parameters of the current study, a modest advantage was present for the ANOVA in the uniform distribution, while the Kruskal-Wallis demonstrated a stark power advantage under the chi-square (df=2) distribution. This finding extends to K>3 tests for shifts in location the claims made by Blair (1981) and Blair & Higgins (1985) regarding two-sample tests. That is when normality conditions are perfectly met, the nonparametric test is only slightly

less powerful than the parametric test, and when those conditions are not met, the nonparametric test is often vastly more powerful.

This study utilized theoretical distributions which many researchers would suggest are not representative or productive in the experimental context. Though this could be true, what they do provide is a reference to which others can test the assumptions and results presented herein. Furthermore, they provide an understood standard by which ideas such as a "normal and non-normal" context can be examined. The question left for researchers to consider in selection of statistical tests is how close is close enough? That is to say, knowing that under perfectly met normality conditions the Kruskal-Wallis suffers a .01-.02 power disadvantage compared to the ANOVA and approximate randomization ANOVA, is that enough of a disadvantage to surrender power advantages of what was demonstrated in this study to be as large as .36? Knowing that perfect normality is rarely, if ever, achieved, this study provides more evidence that there is quite literally little to lose in using nonparametric statistics when exploring shifts in location. It should also be noted that in this study, group sizes and treatment effect sizes were always held equal, two more factors that placed the ANOVA in the best position to demonstrate superior statistical power.

In the future, researchers may want to explore opportunities to test the significance level of a noted power difference. Perhaps there is a way to test whether the difference between two power slopes is significant, or if there is not, this may be something for future research to explore. At the time of this study, the researcher was unaware of any such tools.

## REFERENCES

Adams, D. C. & Anthony, C. D. (1996). Using randomization techniques to analyse behavioural data. *Animal Behavior, 54*(4), 733-738.

Andrews, D. F., Gnanadesikan, R., & Warner, J. L. (1971). Transformations of multivariate data. *Biometrics, 27*, 825-840.

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. (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.

Boik, R. J. (1987). The Fisher-Pitman permutation test: A non-robust alternative to the normal theory F test when variances are heterogeneous. *British Journal of Mathematical and Statistical Psychology, 40*(1), 26-42.

Boneau, C. A. (1962). A comparison of the power of the U and t tests. *Psychological Review, 69,* 246-256.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Edgington, E. S. (1995). *Randomization Tests*. (3rd ed). New York, NY: Marcel Dekker.

Feir-Walsh, B. J. & Toothaker, L. E. (1974). An empirical comparison of the ANOVA f-test, normal scores test, and kruskal-wallis test under violation of assumptions. *Educational and Psychological Measurement, 34*, 789-799.

Fisher, R. A. (1935). *The design and analysis of experiments*. Edinburgh: Oliver and Boyd.

Good, P. (1994). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. New York, NY: Springer-Verlag.

Good, P. (2002). Extentions of the concept of exchangeability and their applications. *Journal of Modern Applied Statistical Methods*, *1*(2), 243-247.

Glass, G., Peckham, P., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research, 42*, 237-288.

Hayes, A. F. (1996). Permutation test is not distribution-free: Testing $H_0$: ρ=0. *Psychological Methods, 1*(2), 184-198.

Higgins, J. J. & Blair, R. C. (2000, February). Letter to the Editor. *The American Statistician*, *54*, 86.

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

Hunter, M. A. & May, R. B. (2003). Statistical testing and null distributions: What to do when samples are not random. *Canadian Journal of Experimental Psychology, 57*(3), 176-188.

Johnson, N.L., & Kotz, S. (1970). *Distributions in statistics: Continuous univariate distributions.* New York, NY: Wiley Interscience.

Keppel, G. & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, HJ: Pearson Prentice Hall.

Kerlinger, F. N. & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Belmont, CA: Cengage Learning.

Kruskal, W. H. & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47,* 583-621.

Langbehn, D. R., Berger, V. W., Higgins, J. J., Blair, R. C., & Mallows, C. L. (2000). Letters to the editor. *The American Statistician, 54*, 85-88.

Lehmann, E. L. (1975). *Nonparametrics: Statistical methods based on ranks.* San Francisco, CA: McGraw-Hill International, New York-Dusseldorf.

Ludbrook, J. & Dudley, H. (1998). Why permutation tests are superior to t and F tests in biomedical research. *The American Statistician*, *52*(2), 127-133.

Manly, B. F. (1995). Randomization tests to compare means with unequal variation. *Sankhya: The Indian Journal of Statistics, 57*(B), 200-222.

May, R. B. & Hunter, M. A. (1993). Some advantages of permutation tests. *Canadian Psychology, 34*(4), 401-407.

May, R. B., Masson, E.J., & Hunter, M. A. (1989). Randomization tests: Viable alternatives to normal curve tests. *Behavior Research Methods, Instruments, & Computers, 21*(4), 482-483.

Mendenhall, W. & Sincich, T. (1995). *Statistics for engineering and the sciences.* Englewood Cliffs, NJ: Prentice-Hall.

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

Mielke, P. W. & Berry, K. J. (2001). *Permutation methods: A distance function approach.* New York, NY: Springer.

Neave, H. R. & Granger, C. W. J. (1968). A monte carlo study comparing various two-sample tests for differences in mean. *Techometrics, 10*, 509-522.

Neave, H. R. & Worthington, P. L. (1988). *Distribution-free tests.* London: Unwin Hyman.

Noreen, E. (1989). *Computer-intensive methods for testing hypotheses.* New York: Wiley.

Potvin, C. & Roff, D. (1993). Distribution-free and robust statistical methods: Viable alternatives to parametric statistics? *Ecology, 74*(6), 1617-1628.

Rao, C. R. & Sen, P. K. (2002). Permutation scores tests for homogeneity of angular and compositional gaussian distributions. *Journal of Nonparametric Statistics*, *14*(4), 421-433.

Ryan, T. A. (1959). Multiple comparisons in psychological research. *Psychological Bulletin, 56*(1), 26-47.

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

Sawilowsky, S. S. (1993). Comments on using alternatives to normal theory statistics in social and behavioral science. *Canadian Psychology*, *34*, 398-406.

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 597-599.

Sawilowsky, S. 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*(3), 352-360.

Sawilowsky, S.S. & Fahoome, G.C. (2003). *Statistics via monte carlo simulation with Fortran*. Rochester Hills, MI: JMASM.

Sawilowsky, S. S., Blair, R. C. & Higgins, J. J. (1989). An investigation of the type I error and power properties of the rank transformation procedure in factorial ANOVA. *Journal of Educational Statistics, 14*(3), 255-267.

Scheffé, H. (1959). *The analysis of variance.* New York: Wiley.

Sharp, V. (1979). *Statistics for the social sciences.* Boston: Little, Brown.

Tomarken, A. J. & Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin, 99*(1), 90-99.

Van den Brink, W.P., & van den Brink, S.G.L. (1989). A comparison of the power of the t test, wilcoxon's test, and the approximate permutation test for the two-sample location problem. *British Journal of Mathematical and Statistical Psychology, 42*, 183-189.

Walsh, E. O. (1968). *An introduction to biochemistry.* London, England: English Universities.

Weber, M., & Sawilowsky, S. (2009). Comparative Power of the independent t, permutation t, and Wilcoxon tests. *Journal of Modern Applied Statistical Methods, 8*(1), 10-15.

Zimmerman, D. W. (1987). Comparative power of Student t test and Mann Whitney U test for unequal sample sizes and variances. *Journal of Experimental Education, 55*, 171-174.

Zimmerman, D. W. & Zumbo, B. D. (1990a). Effect of outliers on the relative power of parametric and nonparametric statistical tests. *Perceptual and Motor Skills, 71*, 339-349.

Zimmerman, D. W. & Zumbo, B.D. (1990b). The relative power of the Wilcoxon-Mann-Whitney test and Student t test under simple bounded transformations. *The Journal of General Psychology, 117*(4), 425-436.

## ABSTRACT

**COMPARATIVE POWER OF THE ANOVA, APPROXIMATE RANDOMIZATION ANOVA, AND KRUSKAL-WALLIS TEST**

by

**JAMIE H. GLEASON**

**May 2013**

**Advisor:** Dr. Shlomo Sawilowsky

**Major:** Education, Evaluation & Research

**Degree:** Doctor of Philosophy

The t test has been suggested to be robust to departures from normality as long as group sizes are equal and samples approach 30 or more. The F statistic has also been proposed to have the same robust qualities as the t, though researchers have suggested that because a test is robust to departures from normality, that does not necessarily make it the best test for every situation. With the increase in computing capabilities, the permutation ANOVA has been explored as an alternative to the ANOVA under non-normal conditions to rehabilitate the loss of statistical power. Since the permutation ANOVA does not operate under the assumption of normality and uses actual scores, many researchers suggest that the permutation ANOVA is superior to rank tests such as the Kruskal-Wallis because ranking data disposes of valuable information. To compare the power of the ANOVA, approximate randomization ANOVA, and the Kruskal-Wallis test, the researcher performed a Monte Carlo analysis on group sizes of n=10 to n=30 and groups of k=3 and k=5 using Fortran program language and the IMSL subroutine library. In 12 different treatment conditions, the researcher implemented equal treatment effect sizes of

small (0.1σ) to huge (1.0σ) on each treatment group in graduated increments, until all but one group had received a treatment. Data were drawn from three theoretical distributions: the normal (Gaussian) distribution, the uniform distribution, and the chi-square (df=2) distribution. Results indicated that regardless of the number of treatment groups, the ANOVA and approximate randomization ANOVA exhibited almost equal power under every distribution and effect size. The power of the Kruskal-Wallis was slightly less than the ANOVA and approximate randomization ANOVA under the normal and uniform condition, and was significantly more powerful under the chi-square (df=2) distribution. The sample size and treatment effect had little to do with the relationship between the performances of the three tests but did affect the rate of power increase and maximum power achieved. Implications of the findings as well the contribution to existing literature is discussed.

# AUTOBIOGRAPHICAL STATEMENT

## Jamie H. Gleason

## Education:

**Wayne State University** — Detroit, MI
Ph.D. Evaluation and Research — May 2013

**Lehigh University** — Bethlehem, PA
Master of Education, Human Development (School Psychology) — May 2003

**Wayne State University** — Detroit, MI
Bachelor of Arts, Psychology — May 2000

## Publications:

Gleason, J. H., Alexander, A. M., & Somers, C. L. (2000). Later adolescents' reactions to three types of childhood teasing: Relations of self-esteem and body image. *Social Behavior and Personality: An International Journal*, 28(5), 471-479.

Somers, C. L., Gleason, J. H., Johnson, S. A., & Fahlman, M. M. (2001). Adolescents' and teachers perception of a teen pregnancy prevention program. *American Secondary Education,* 29*,* 51-66.

Somers, C. L. & Gleason, J. H. (2001). Does the source of sex education predict adolescents' sexual knowledge, attitudes, and behaviors? *Education*, 121(4), 674-682.

## Presentations:

Gleason, J. H., & Somers, C. L. (April, 2000). Does source of sex education predict adolescents' sexual knowledge, attitudes, and behaviors? Poster accepted for presentation at the American Educational Research Association (AERA) Annual Convention, New Orleans, LA.

Somers, C. L., Gleason, J. H., & Fahlman, M. M. (April, 2000). Adolescents' and teachers' perceptions of the 'Baby Think it Over' teen pregnancy prevention program. Paper presentation submitted to the American Educational Research Association (AERA) Annual Convention, New Orleans, LA.