A MONTE CARLO COMPUTER STUDY OF THE POWER PROPERTIES OF SIX DISTRIBUTION-FREE AND/OR NONPARAMETRIC STATISTICAL TESTS UNDER VARIOUS METHODS OF RESOLVING TIED RANKS WHEN APPLIED TO NORMAL AND NONNORMAL DATA DISTRIBUTIONS

by

BRUCE ROBERT FAY

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

2003

MAJOR: EVALUATION & RESEARCH

Approved by:

Shlomo S. Sawilowsky 10/14/03

Advisor Date

Donald R. Marcotte

Gail Fahoome

David Josiah

UMI Number: 3116494

DEDICATION

To my wife, Linda Marie Fay, who has patiently supported my interests for over 30 years.

ACKNOWLEDGMENTS

Many friends, family members and colleagues assisted in the conduct of this research and the completion of this dissertation. If I have failed to mention anyone it is an unintentional oversight. If the results are disappointing, it is my responsibility alone.

Dr. Shlomo Sawilowsky was my major advisor. He initiated and encouraged my interest in nonparametric/distribution-free tests and empirical methodologies for studying them through the use of computer simulations based on Monte Carlo techniques. He was also instrumental in my decision to pursue a Ph.D. in Education, Evaluation & Research at Wayne State University.

Special thanks go to the other members of my doctoral committee, Dr. Donald Marcotte, Dr. David Jonah and Dr. Gail Fahoome. Each contributed something special to my preparation for this work, as well as to the work itself.

Dr. Marcotte taught me the basics of applied social science statistics and data analysis. It is difficult to place a value on his practical experience and the way it informed his teaching. He was a patient but exacting teacher; patient in his explanations but exacting of rigor in my work. If that attention to detail is evident here, he is to be thanked.

Dr. Jonah helped me develop whatever depth of understanding I posses about mathematics. I lay no claim to being a mathematician, yet he was always

a patient and willing teacher. In particular he encouraged me to develop my ability to think and write about things mathematical in a mathematical way. He did this by setting excellent problems before me, allowing me to grapple with them, assisting me with well considered 'tips' when needed, and critiquing my writing in a constructive way. Thus, if my writing expresses with any clarity an underlying logical thought process, it represents his influence.

Dr. Fahoome played several roles. First and foremost, this research was an outgrowth of her dissertation. She was gracious enough to supply a copy to me and was available to discuss her work without restriction. She also supplied copies of her Fortran code, which served as a starting point for my software development efforts. She was unselfish in the extreme when it came to reviewing my code, my results, and my writing. She was also my instructor for several key courses, including Monte Carlo methods, and was a major supporter of my efforts throughout the dissertation process.

If I have had a sponsor in this endeavor it was Dr. Frank McBride. Dr. McBride was one of my first professors at Wayne State University when I initially enrolled over a decade ago. I studied philosophy of education with him, followed by independent study in philosophy of mathematics and finally a doctoral seminar in philosophy of education. He helped sharpen my ability to think critically and skeptically about the world around me. Although not directly involved in my research, he was an early, strong and consistent promoter of the idea that I should pursue a Ph.D. He knew I would earn one before I did. He was also instrumental in creating the opportunity for me to teach philosophy of education as an adjunct faculty member at Wayne State University.

Dr. Stuart Itzkowitz was another supporter who deserves special mention. As an admissions counselor in the Graduate Office of Wayne State University's College of Education, he was my first contact with the University and the College. It was in large measure because of my interactions with him that I chose to attend Wayne State University in 1990 to pursue a Master of Arts in Teaching with secondary certification in mathematics and physics. Since then I have turned to him on many occasions, often at moments of minor crisis. He was always calm, reassuring and able to resolve the problem. He checked on my progress from time to time and encouraged me to keep going.

Dr. James Gullen is an assessment consultant at Wayne RESA (the Wayne County Regional Educational Service Agency), where he and I have shared an office since July of 2000 when I began working there. Perhaps no one suffered the process of completing this research program on a daily basis more than he. His contribution was immense, the sum total of a large number of small conversations, which, unlike random errors, added up to a great deal more than zero. Having completed his Ph.D. in Education, Evaluation & Research at Wayne State University in the fall of 2000, his insights were invaluable. He also proofread various drafts of the manuscript and made many important comments thereon.

Dr. Kristine Erickson is also an assessment consultant at Wayne RESA, having joined Dr. Gullen and myself in that department during the Fall of 2001.

She completed her Ph.D. in Special Education with a cognate in Education, Evaluation & Research at Wayne State University during the winter of 2000. Her excitement, which was contagious, carried over into my project. She was my biggest fan, and I believe she was more enthusiastic about my work, at times, than I was. She was valuable beyond words in helping me navigate the doctoral process. She also willingly read and commented on the various drafts of the manuscript, where her knowledge of APA style was invaluable. Finally, it was she who encouraged me to apply for the Graduate Professional Scholarship, which greatly assisted in the completion of my program.

A note of thanks to Dr. Helen Stanks, Director of Instruction and Assessment Services at Wayne RESA, for her interest and support while completing this project. Rounding out my Wayne RESA family was research librarian Elaine Baranko. Without her assistance in locating and collecting research literature I don't know if this project would ever have been completed.

There are others to acknowledge, of course. Among classmates and colleagues special mention goes to Jack Hill for his expert consultations in matters Fortran. One of the joys of pursuing this degree was the chance to work with serious adult learners who brought a maturity and unique sense of purpose to their pursuits.

A special note of thanks is also due Wayne State University's Graduate School and College of Education for recognizing and accommodating the unique needs of adult learners and offering programs that make it possible for working adults to pursue advanced, scholarly work.

Among family members, my children, Brendan and Meghan, were good sports about having a dad who was pursing graduate studies while they were also in college. Both have proven to be serious and competent students. I am immensely proud of them, and I hope my struggles have inspired them to pursue their own educational dreams and to be lifelong learners whatever their career pursuits.

My parents, James and Eloise, were the roots of my interests. From both of them I learned the value of hard work. From my father, an engineer, came my interest in the way things work that eventually lead to my own training as an electrical engineer and eventually my interest in computers and mathematics. From my mother, a teacher of the English language, came my interest in education, an 'ear' for the language, and some sense of how to write. From both, came a love of music, art, and culture generally. They always believed that I could do whatever I set my mind too and were generous in their support of my pursuits beyond anything I can ever repay. It is of enormous importance to me that they are still alive to witness the completion of this work.

Of course, I must acknowledge my wife, Linda, who managed to meet all her personal and professional commitments while I went to class, piled books up all over the house, tinkered with my programs, and spent endless hours at the computer writing this paper. There should be a special medal for spouses of people who earn Ph.D.s.

Finally, a special acknowledgement goes to Dr. R. Clifford Blair, of the University of South Florida. Although I did not work with him directly, his influence on this study is immediate and important. Dr. Blair was a pioneer in the use of computer methods based on Monte Carlo techniques as a way to study the efficacy of statistical tests. He brought the issue of power to the forefront and was one of the first to conduct studies with realistic data representative of that encountered in the social and behavioral sciences. As Dr. Sawilowsky's major advisor, Dr. Blair's work has exerted a direct and significant influence on all of Dr. Sawilowsky's students, including me.

PREFACE

The study of nonparametric/distribution-free statistical tests using Monte Carlo techniques implemented via Fortran computer programs was a natural continuation of a long-standing interest. My Master's degree in electrical engineering, completed in May 1976, involved the design and construction of a control systems simulation program. Like the present study, this was also a Fortran program (~10,000 lines of code). Unlike the present study, it utilized methods of numerical analysis to solve sets of simultaneous differential equations representing complex physical systems, including feedback paths, and their response to various forms of input signals. During the intervening years the focus of my interests has shifted from physical to social science and the mathematics from differential equations to probability and statistics. The purpose, however, remains the same. I have long been fascinated by the potential of the computer, when combined with mathematical and scientific understanding, to model aspects of the real world that are otherwise theoretically intractable.

Bruce R. Fay

Farmington Hills, Michigan

September 2003

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 — Overview

This study examined the power properties of six nonparametric and/or distribution-free inferential statistical tests using computer-based Monte Carlo techniques. The primary focus of this study was to investigate the effect of nine different methods of resolving tied ranks prior to the computation of a test statistic and comparison to a critical value. Although the terms 'nonparametric' and 'distribution-free' are not synonymous, the term 'distribution-free' is generally used throughout this paper to refer to both types of tests.

This study was an outgrowth of Fahoome (1999), who studied the Type I error properties of large-sample approximation formulas for twenty nonparametric and/or distribution-free statistics, including the six studied here. Ties, however, were either ignored or resolved in one specific way on a test-by-test basis. Fahoome (1999) studied these statistics using the normal distribution and four of the Micceri (1986) distributions. The same pseudo-population models were used for this research.

Micceri (1986, 1989) showed that the predominant populations being studied in the social and behavioral sciences clearly did not meet the distributional assumptions of parametric tests. Bradley (1968) pointed out that distribution-free tests only fail to use sample information, relative to parametric tests, if the population assumptions are, in fact, true. Blair & Higgins (1980) restated this issue when they indicated that there was no mathematical basis for the power superiority of the Student's *t*-Test (Student, 1908) without guaranteed knowledge of the form of the population.

Cliff (1996a, 1996b) argued that much of the data, and many of the research questions, of interest in the social and behavioral sciences are fundamentally ordinal in nature and that distribution-free tests produced results that were more inferentially valid. Wilcox (1998a, 1998b) contended that while significant results obtained by traditional parametric procedures probably represented the correct detection of actual effects, many non-significant results were probably Type II errors, resulting in lost insights and missed opportunities to advance understandings in social and behavioral science.

Distribution-free tests are important in the context of social and behavioral science research because they generally make less stringent population assumptions, such as normality or equality of variance, than tests based on parametric statistics. Many distribution-free tests are based on methods involving ranks and thus apply more directly to ordinal data and are inherently more appropriate than parametric tests for ordinal research questions.

Distribution-free tests are a subset of techniques referred to as 'modern methods.' Despite the fact that many of these tests have been around for some time, they do not receive the widespread, even dominant use that they deserve (Wilcox, 1998a, 1998b). Harwell (1990) outlines many of the reasons for this lack of use, but one need only look at the history of social and behavioral science since 1800 to gain some insight into the myth of the normal distribution and the resulting edifice of statistical practice built thereon (Porter, 1986).

When the assumptions of parametric tests (e.g., Student's *t*-Test) are met, they perform in mathematically predictable ways and are the uniformly most powerful tests available. They are mathematically tractable, based on elegant and powerful mathematical theory. The merits of these tests, however, do not necessarily extend to situations where their assumptions are not met. Student's *t*-Test, for example, has robust Type I error properties under departures from normality, but is less powerful than the Wilcoxon Rank-sum (Mann-Whitney *U*) Test under these conditions (Blair & Higgins, 1980; Sawilowsky & Blair, 1992). The apparent robustness of Type I error to the violation of the normality assumption was established for both Student's *t*-Test and Fisher's *F*-Test prior to a full understanding of the importance of Type II error and power. Thus, textbook authors and research practitioners came to believe that parametric procedures could be applied with impunity under any conditions.

Tradition carries its own weight, and researchers are still not routinely trained in the use of modern methods, including distribution-free techniques. Instead, Student's *t*-Test, analysis of variance (ANOVA) and linear regression are still presented as the foundation of data analysis for experimental and quasi-experimental research in many fields (Wilcox, 1998a, 1998b; Emmons, Stallings & Layne 1990; Elmore & Woehlke, 1998). Modern methods, if they are dealt with at all, are given cursory treatment in many survey courses and textbooks with limited availability of extensive tables of critical values.

Another issue that works against the wider use of distribution-free tests is that many of them involve extensive and tedious calculations. For moderate to large samples these calculations are often only practical via computer. This tends to offset the fact that they are conceptually and computationally simpler than parametric tests and robust methods (a different subset of modern methods). Although the availability of sufficiently powerful computer hardware should no longer be an issue, many of these tests are still not included in the most popular commercial statistics packages. When they are included, their implementations are often questionable as they give inconsistent results (Bergmann, Ludbrook, & Spooren, 2000). Specialized software packages are becoming available that perform much better, but tend to be expensive.

Finally, even though the less stringent underlying assumptions of these tests are rarely met in practice, the effects of violation of assumptions on robustness of Type I error rates and power have still not been studied as extensively as they need to be using modern (empirical) investigational methods. This is especially true given the subtle nature of robustness (Bradley, 1978; Wilcox, 1998a).

Under conditions that violate their assumptions the sampling distributions of many of these tests become conditional and mathematically intractable. It is also understood that they become approximate tests, may not maintain Type I error rates as specified by nominal α, and may lose power to detect effects when those effects exist. As a result, actual performance under real-world conditions may not be well understood. Under such circumstances it is understandable that a researcher might be reluctant to use these tests.

Often overlooked is the fact that parametric tests fare no better under corresponding circumstances, and often perform worse (see for example, Blair & Higgins, 1980; van den Brink & van den Brink, 1989; Sawilowsky, 1990; Sawilowsky & Blair, 1992; Kelley, Sawilowsky, & Blair, 1994; MacDonald, 1999) Contrary or mixed results have also been reported in the literature (see for example, Srisukho & Marascuilo, 1974; Killian & Hoover, 1974; Zimmerman, 1987, 1998, 1999; Zimmerman & Zumbo, 1989, 1990a, 1990b, 1992b). Indeed, inconsistency of results is common in the literature, and problematic for researchers in this field. It is due, in part, to the manner in which empirical studies are carried out (Bradley, 1978) and to the lack of a well-defined meta-analytic approach to summarizing Monte Carlo studies (Harwell, 1992).

Of particular interest in this study was that most of the distribution-free statistics investigated utilize rank-based procedures that rely on an underlying assumption of population continuity such that samples are assumed to have no equal data values (no zero difference–scores, no tied ranks), either within groups or between groups. This assumption is almost never met in practice, however, especially with data in the social and behavioral sciences (Micceri, 1986, 1989). Sparks (1967) conducted one of the few empirical studies to have specifically examined violation of continuity and methods of resolving ties, using the Wilcoxon Rank-sum (Mann-Whitney *U*) Test and discrete approximations to the normal curve.

The practical consequence of violating the assumption of population continuity is that samples will contain equal data values resulting in zero difference–scores or tied ranks. A useful distinction can be made, however, between consequential (critical, meaningful) and inconsequential (non-critical) ties. It is possible for ties to occur in such a way that they have no effect on the calculation of the test statistic or the resulting inference, and such ties are clearly inconsequential. Ties that occur only within a group, when looking for between group effects, are often of this type. Inconsequential ties can and should be resolved by some simple procedure that maintains the integrity of the ranks, such as arbitrary assignment in sequence of the set of ranks for which the group of scores is tied.

Other ties occur in such a way that different resolutions of the tie(s) result in different values of the statistic and may result in different inferential decisions. Such ties are clearly consequential and it was the purpose of this study to examine the effect on Type I error rate and power of different methods for resolving them. This is important, in part, because these distribution-free methods are becoming more widely used as they become increasingly available in commercial software packages.

A number of methods for resolving consequential ties have been suggested in the literature, but there have been few empirical studies to determine their effect on the behavior of specific statistics. The methods fall into two basic types: A) those that calculate the test statistic in two or more ways and then combine the resulting statistics into a single statistic for testing, and B) those that alter the data (ranks) prior to calculating a single test statistic so that the data appear to meet the assumptions of the test.

This study looked at nine methods for dealing with consequential ties (zero difference–scores or tied ranks), including some of each type. The methods studied were:

3a) Resolve consequential ties in the manner least favorable to rejection of the null hypothesis (method 1) and in the manner most favorable to rejection of the null hypothesis (method 2), calculate the statistic for each of these resolutions and then calculate the mid-range (mean) value of these two statistics and use it to conduct the test.

3b) Count ties as ½ (Rosenbaum's Test and Tukey's Quick Test only).

4) Alternately resolve each set of tied-for ranks.

5) Randomly resolve each set of tied-for ranks.

6a) Assign the mid-rank of a set of tied ranks to each score without further correction.

6b) Delayed increment (Kolmogorov-Smirnov Test of General Differences only).

6c) Weighted average of all possible resolutions (Rosenbaum's Test only).

7) Drop matching tied-for ranks and reduce *N* accordingly.

8) Drop all tied-for ranks (if possible) and reduce *N* accordingly.

Methods 4), 5), 6a) and 8) were described by Bradley (1968), Gibbons & Chakraborti (1992), and others. Methods 3a) and 3b) were described by Neave & Worthington (1988). Method 3a) is related to a method described by Bradley

(1968). Method 7 was not encountered in the literature. Methods 1) and 2) were only used to investigate method 3a).

Bradley (1968) also described methods involving calculation of statistics for all possible resolutions of consequential ties, the results being used to establish probability bounds for the test or to calculate a mean probability. Although theoretically attractive, these methods are often computationally impractical and were not be part of this study other than with Rosenbaum's Test. For many tests, however, the calculation of an average statistic based on all possible resolutions of ties turns out to be equivalent to resolving each set of tied-for ranks using the mid-rank (Neave & Worthington, 1988). Bradley (1968) warned, however, that under some circumstances the use of mid-ranks might give a statistic something closer to its minimum or maximum value rather a median or mean value.

Although some of the methods for resolving ties result in a modified sample in which ties have been eliminated, others, such as mid-ranks, result in modified samples that still contain duplicate and/or non-integer ranks. These can result in values of the test statistic that are non-integral, even though this cannot happen when all assumptions of the test are met. The resulting statistics were still referred to the standard table of critical values, as the performance when used in this manner was a major point of this research.

The manner in which these methods were applied to two-independent-samples tests was fairly clear. Extension to *k*-independent-samples is discussed in Chapter 3.

Both theoretical and empirical data distributions were used as sources of samples. The normal distribution was used for reference. This distribution did not produce samples with significant numbers of duplicate data values (tied ranks) and thus served as a base line for the performance of these tests under conditions meeting their underlying continuity assumption (and those of parametric tests to which they might be compared). Of more interest, however, were the empirical distributions due to Micceri (1986) that are typical of the data found in social and behavioral science. These distributions were also discussed in Micceri (1989), Sawilowsky, Blair & Micceri (1990), and Sawilowsky & Blair (1992). They are inherently discrete, decidedly non-normal, and are referred to in the remainder of this paper as the Micceri (1986) distributions, since that is where they originally appeared.

The method used to simulate effects of known size by shifting one (or more) samples, when applied to the Micceri (1986) distributions, yielded samples with duplicate values leading to zero-difference scores or tied ranks. It was the performance of these tests with respect to the various methods of resolving ties, when used with such data, which was of primary interest in this research.

Fahoome (1999) used four of the eight Micceri (1986) distributions. However, with regard to the extreme bimodal distribution Fahoome (1999) concluded:

> ...because of the small number (6) of data points, there were an extremely large number of ties, even for relatively small sample sizes. This data is Likert-type data. The performance by most tests was extremely poor. Most of the tests had inflated Type I error rates, some as high as 0.99999. A few had very low Type I error rates. (p. 462)

In spite of this finding, the extreme bimodal distribution was retained for this study because of the widespread existence of such data.

All Type 1 error simulations were limited to equal initial per-group sample sizes in the range 3-30 to the extent that critical values were available. Power and Type III error studies were run for equal initial per-group sample sizes of 3(3)30 if critical values were available. The method of dropping ties and reducing *N* often lead to tests on unequal sample sizes for which the test statistic could not be computed. As such, it was not completely consistent with the other methods. It is so widely recommended and used, however, that it was retained in the study. This necessitated a modified approach to the Monte Carlo simulations in which untestable samples were tallied and then discarded. Samples were drawn until 10,000 testable samples were obtained, or the program reached its 10,000,000th cycle, whichever came first. A more systematic and thorough investigation of the performance of these tests under a wide range of unequal sample sizes was beyond the scope of this research.

Type I error studies and power studies were conducted for both one-sided (directional) and two-sided (omnibus) alternatives for the Kolmogorov-Smirnov Test of General Differences, Rosenbaum's Test, Tukey's Quick Test and the Wilcoxon-Mann-Whitney Test. For power studies, the one-sided test was made in the direction of the simulated effect, while significant results in the "wrong" tail constituted Type III errors. The Kruskal-Wallis Test is an inherently non-directional omnibus test, while the Terpstra-Jonckheere Test is an inherently directional test against an ordered alternative hypothesis.

1.2 — Definitions

Definitions of technical terms are included as a glossary in Appendix D.

1.3 — Research Questions

Question 1: For samples drawn from the same population, is the Type I error rate maintained between $.5\alpha$ and 1.1α for each combination of test, method, and number of groups, directionality, sample size, and distribution?

Bradley (1978) recommended conservative bounds for robust Type I error of nominal alpha $\pm$ 10% and liberal bounds of nominal alpha $\pm$ 50%. Many nonparametric and/or distribution-free tests, however, cannot achieve nominal alpha at small sample sizes. Thus, the entries in critical value tables are 'best conservative' values that may fall below the recommended 10% lower bound. As such, the main interest in the Type I error studies was the ability of each test to resist inflation of Type I error rate. Type I error rates were considered acceptable if they fell in the range of $.5\alpha$ to 1.1α.

The remaining research questions were only studied for those combinations of test, method, number of groups, directionality, sample size and distribution for which Question 1 was answered in the affirmative. The power of a test was of no interest if the Type I error rate was not robust to violations of assumptions. *A priori*, it was expected that those combinations of test conditions that produced Type I error rates well below nominal alpha would also have attenuated power.

Question 2: For samples drawn from populations differing only in location, what is the power for each combination of test, method, number of groups, directionality, sample size and distribution?

Question 3: For samples drawn from populations differing only in location, is there a preferred method of resolving tied ranks for each combination of test and distribution, irrespective of the number of groups, directionality, and sample size?

Question 4: For samples drawn from populations differing only in location, is there a preferred method of resolving tied ranks for each test, irrespective of the number of groups, directionality, sample size and distribution?

1.4 — Subjects

No living subjects, human, animal, or plant were used in this study. The Behavioral Protocol Summary Form was submitted to the Wayne State University Behavioral Investigation Committee on February 13, 2001, requesting exempt status. The Protocol was approved following expedited review (Category 5*) by the chairman for the Wayne State University Institutional Review Board (B03) for the period of March 9, 2001, through March 8, 2002. A continuation was submitted and approved for February 18, 2002 through February 17, 2003 and again for the period January 14, 2003 through January 13, 2004. Copies of the Notices of Expedited Approval are included in Appendix A.

1.5 — Assumptions and Limitations

Of the twenty statistics studied by Fahoome (1999), this research focused on six of them. These statistics are all tests of location or general difference for two or more independent-samples. Dependent-samples (repeated-measures) tests were not studied because the problem of generating correlated samples for Monte Carlo power studies is not yet fully solved (Headrick & Sawilowsky, 1999a, 1999b). Tests of pure spread are also not included here because such effects are not believed to occur widely in practice (Sawilowsky & Blair, 1992; Sawilowsky, 2002).

To the extent that the tests being studied assumed that the populations from which the samples were drawn differed only in location, that assumption was met by the method used to generate the data for the power studies. However, tests for single effects, such as pure shift, that are based on the assumption of all else equal, may not match reality very well. Bradley (1978) pointed out that a test can be robust (in some particular way) to violations of each of its assumptions, taken one at a time, and yet not be robust to violations of combinations of those assumptions, the various violations interacting to produce combinations with very particularistic consequences. While Zimmerman (1998) also addressed this issue, no attempt was made in this study to investigate the effect of distributions that differ in more than one way.

The lack of availability of extensive tables of critical values for many of the distribution-free tests continues to be a problem, especially at larger sample sizes. Although Fahoome (1999) used formulas for calculating approximate critical values for many of these tests, this study was limited to sample sizes for which tables of exact critical values were available or could be generated (Fay, 2002). This study was also limited to equal initial sample sizes. Since the method of dropping ties and reducing *N* often resulted in unequal sample sizes, this method was only studied for tests where tables of critical values for unequal sample sizes were available or could be generated.

One of the most widely suggested methods for dealing with (consequential) ties is to resolve them in all possible ways, obtaining a value of the statistic (or its associated probability) for each resolution. A mean value of the statistic (or a mean value of the probability) is then obtained and tested or probability bounds established. This method was not included in this study because of the practical difficulties involved in implementing it for even moderate sample sizes when there are numerous ties at several different values. Also, comprehensive tables of exact probabilities are even more difficult to obtain than critical value tables. This method was implemented, however, for Rosenbaum's Test as there was a practical method for doing so.

CHAPTER 2

REVIEW OF LITERATURE

2.1 — Overview

Historically, nonparametric and/or distribution-free tests predate parametric ones. Arbuthnot (1710) described the first statistical test, known today as the Sign test (as cited in Everitt, 1998; Lehmann, 1998; and others). Spurred by an interest in understanding the odds in various games of chance, mathematicians such as Fermat and Pascal laid the foundation for probability theory in the preceding century (Smith, 1959, Section IV). A fundamental principle that emerged from this period was that an event with N different (equally likely) outcomes, n ($<N$) of which were regarded as successes (and N-n as failures), had a probability of success given by the ratio n/N (Bradley, 1968).

Jacques Bernoulli (1654–1705) extended this finding into the binomial theorem, which forms the basis for the Sign test, among others (Bradley, 1968). He also contributed to the development of the law of large numbers (Boyer, 1985) and the first mathematical approach to the measurement of uncertainty (Stigler, 1986). De Moivre (1733, 1738) discovered the normal distribution as the limiting form of the binomial distribution as the number of trials became infinite (as cited in Smith, 1959, Section IV). Laplace rediscovered the normal distribution in the late 18th century followed by Gauss in the early 19th century (Smith, 1959, Section IV).

Laplace and Gauss both worked on the derivation of the normal distribution as the result of summing small, independent, elementary errors as the number of errors became infinite, in the limit (Smith, 1959, Section IV). Their derivations were special forms of the Central Limit Theorem (Bradley, 1968), the major contribution of Laplace to probability theory (Stigler, 1986). The derivation of Gauss, in particular, was based on the recognition that observational errors of measurement in astronomy should be just as likely to be in one direction (positive) as in the other (negative) and that the larger the errors were in absolute magnitude, the less common they should be. Gauss had access to empirical distributions of observational errors that appeared to be unimodal, symmetric and monotonically decreasing on either side, for which a simple equation with those properties is $\frac{dy}{dx} = -Cxy$ (Bradley, 1968). When solved for *y*, this yields $y = ke^{-\left(\frac{Cx^2}{2}\right)}$, the basic equation for a normal curve.

Thus normality originated as an idealized mathematical model for a certain type of experimental data. Although this theoretical error distribution fit astronomical observation errors quite well, no one ever claimed the fit was perfect, least of all Gauss, who, preceded by Legendre (1805), developed the theory of least squares as a way to deal with the fact that observations never exactly fit mathematical models (Smith, 1959, Section IV; Bradley, 1968). The work of Legendre, however, lacked any reference to probabilities. Between 1805 and 1812 there was a synthesis of the work of Gauss and Laplace that put the method of least squares on a solid probabilistic footing and set the stage for the development of social science statistics (Stigler, 1986).

Porter (1986) and Stigler (1986) both traced the rise of statistical thinking from 1820 to 1900. No attempt is made to recount this fascinating and complicated story here. A key figure, however, in the first two thirds of this period was Quetelet. Trained in mathematics and astronomy, he was well acquainted with the fact that errors in astronomical measurements appeared to have a Gaussian (normal) distribution, with the interpretation that the mean of the distribution corresponded to the best estimate of the true value of the observation while the variance of the distribution give some indication of the degree to which that estimate was to be trusted or accepted as true. Everitt (1998) stated that:

> ...Quetelet also built an international reputation for his work on censuses and population such as birth rates, death rates, etc. Introduced the concept of the 'average man' ('l'homme moyen) as a simple way of summarizing some characteristic of a population and devised a scheme for fitting normal distributions to grouped data that was essentially equivalent to the use of normal probability plot. ... (p. 271)

The fact that values of variates, such as height and weight in large populations, appeared to be normally distributed lead to a widespread belief that their was something fundamental about the normal distribution analogous to the relatively simple mathematical equations, such as $F = ma$, that described profound understandings in physics (Porter, 1986). Sampling distributions and inferential tests based on assumptions of normality would come later, but the Normal Mystique was already being established.

As scientists examined more and more anthropological, agricultural and even social data they increasingly saw the (Gaussian) error distribution. It

"...began to fit, and explain, almost everything and soon it was regarded as a population archetype; that which was Gaussian was considered normal and that which was non-Gaussian was regarded as abnormal" (Bradley, 1968, p. 4). The terms 'normal' and 'Gaussian' became interchangeable and it was probably unfortunate that 'normal' ultimately prevailed in common usage because of the extended connotations it eventually took on.

With wide acceptance of the belief in at least quasi-universal normality, and the mean as the 'true' value, the now classical parametric tests began to emerge by the early 20th century. From the beginning, however, there was plenty of evidence of non-Gaussian population distributions as well as scientists who were skeptical of the wide spread claims of almost universal normality.

Pearson (1895) remarked on the existence of asymmetrical frequency curves arising from both heterogeneous and homogeneous materials. With regard to asymmetry in homogeneous materials, Pearson (1895) noted that such curves arose in "many physical, economic and biological investigations" (p. 344) such as "the height of the barometer, in those for prices and for rates of interest of securities of the same class, in mortality curves, especially the percentage of deaths to cases in all kinds of fevers, in income tax and house duty returns, and in various types of anthropological measurements" (p. 344-345).

Bradley (1968) (citing Hogben, 1957, p 159-181) stated:

> ... it is a long step from astronomy to anthropology, and intuitions which prove excellent in one field may be entirely inappropriate to the other [5]. It was their less sophisticated followers, not Laplace and Gauss, who took such steps, but their prestige accompanied the misapplication of their methods. (p. 6)

Geary (1947) stated that "normality is a myth; there never was, and never will be, a normal distribution" (p. 241).

While no data, even astronomical, exactly fit a Gaussian model, data from areas such as agriculture and anthropometry often showed a good fit over the central 80–90 percent of the distribution. In light of this evidence the belief in quasi-universal normality was eventually replaced by a belief in quasi-universal quasi-normality (Bradley, 1968). One of the consequences of this historical development was that practitioners became increasingly unaware that the 'assumptions' of most statistical tests were, in fact, restrictive preconditions on their applicability, or that such restrictions existed only for reasons of theoretical (mathematical) convenience, having no necessary empirical basis.

Even with the Normal Mystique firmly in place, alternative approaches to data analysis and inferential testing were developed along with the parametric ones. Skepticism regarding the true nature of social and behavioral science data was certainly one of the reasons. Another was the complexity of the mathematical foundations of parametric tests, based as they were on asymptotic or infinite population distributions and involving calculus in their derivations. A third reason was sound scientific rationalism. "Among professional statisticians, however, there remained a lingering doubt. It does not follow logically that approximate normality and homogeneity insure approximate validity of a test which assumes exact normality and exact homogeneity" (Bradley, 1968, p. 9).

Savage (1953) placed the true beginning of distribution-free statistics in 1936 when interest developed in the "quantities and characteristics of the sample whose distributions were sensitive to the tested hypothesis and could be expressed by exact combinatorial formulas" (Bradley, 1968, p. 11). Lehmann (1998) placed the beginnings of the modern development of these tests in the papers of Hotelling & Pabst (1936), Friedman (1937), Kendall (1938), Smirnov (1939), and Wald & Wolfowitz (1940, 1943, 1944) and indicated that Scheffé (1943) contained an interesting survey of work to date.

Distribution-free tests were initially perceived as having certain weaknesses and did not gain immediate or widespread use. The new statistics tended to use medians and interquartile ranges as measures of location and spread. Although these had some obvious advantages under nonnormal (and sometimes even normal) conditions, they seemed less sophisticated and second-rate by comparison to the well-established mean and variance. The new tests were often based on ranks rather than original observations, and in that regard they seemed to discard information. Yet another drawback was that many of the tests were limited to simple situations, analogous to the comparison of two group means using the *t*-test. Meanwhile, Fisher (1935) had developed the *F*-test and firmly established the ANOVA procedure as the method of choice for the analysis of data resulting from more complex experiments.

A lack of distribution-free alternatives for more complex analyses is no longer the reality, yet these tests remain limited in the breadth of their application. Puri & Sen (1993) presented a theoretical base for nonparametric multivariate analyses. Zwick (1985) described a specific nonparametric one-way multivariate ANOVA based on the Pillai-Bartlett trace. Sawilowsky (1990) provided an extensive review of 10 nonparametric techniques that could be used to test for interactions in experimental designs, describing them as "robust, powerful, versatile, and easy to compute" (p. 91).

Harwell (1988) remarked on the use of statistical and substantive criteria for selecting between parametric and nonparametric tests. Statistical criteria relate to the ability of a test to control Type I error and power. When all underlying assumptions are met, parametric tests are the uniformly most powerful because they can utilize the additional metric information available in the data. However, a large body of literature exists consisting of both theoretical and empirical studies on the statistical robustness of a small set of parametric tests to violation of assumptions. This literature is contradictory and inconclusive in many ways, as discussed later.

Harwell (1988) asserted that the substantive (non-statistical) criteria for selecting between parametric (PAR) and nonparametric (NPAR) tests had become linked to the level of measurement, attributed to Stevens (1946), as a result of Siegel (1956). He observed that Siegel (1956) "did not attach the measurement requirement to the PAR statistical model but rather to the use of the model" (p. 37) but acknowledged that Siegel (1956) had emphasized the use of parametric tests with interval or ratio level data and nonparametric tests with nominal (categorical) or ordinal level data. Harwell (1988) observed, however, "...how religiously this prescription has been followed by researchers" (p. 37). He remarked that this practice was, unfortunately, often incorrect, the sole statistical criterion for selecting a test being the degree to which the assumptions of the test were tenable in the population.

Harwell (1988) anticipated Cliff (1989, 1993, 1996a, 1996b) when he noted that the substantive criteria for selecting a test should match the level of the data with the research question and subsequent interpretation. He remarked that "An obvious consequence of using data that convey only rank-order information in a PAR test is an increased chance of substantive misinterpretation" (p. 37). Further review of the weak measurement versus strong statistic debate is not warranted here. Stevens (1968) offered his perspective on the debate that had occurred subsequent to Stevens (1946). Sawilowsky (1993) cited Zumbo & Zimmerman (1991) as being a particularly thorough and up-to-date review of this issue. Zimmerman (1993) reiterated that the main *statistical* consideration when choosing a test of significance is how tenable the underlying population assumptions are.

Attempts were eventually made to investigate the performance of parametric tests under violations of their assumptions. Early studies used something like what we now know as Monte Carlo methods, but had significant limitations and shortcomings. The more mathematical studies used population distributions that were mathematically convenient, such as rectangular, triangular, or Pearson curves of various types. They also made additional assumptions to obtain an approximation to the true sampling distribution of the test statistic under assumption-violating conditions. When discrepancies from the normal-theory distribution of the statistic appeared, they were often attributed largely to chance or to the mathematical approximations that had been used. The empirical studies frequently used data from relatively normal populations.

The earliest studies predated computers, or at least easy and widespread access to them, so samples were drawn by hand using tables of random numbers or other (presumably) random selection processes. The number of repetitions was generally insufficient to attain a reasonable estimate of precision and samples were typically limited to smaller sizes. Results, as often as not, were due to chance effects (Bradley, 1968) and were less than convincing by modern standards. These studies also underestimated the complexity of robustness (Bradley, 1978).

Modern studies have tended to show the limitations of classical tests under violation of assumptions. Algina, Oshima, & Lin (1994), for instance, demonstrated that while Welch's (1938) Approximate Degrees of Freedom (ADPF) Test, and James's (1951, 1954) Second-order Test, control Type I error rate as well as or better than the independent samples *t*-test (when data are sampled from skewed distributions with unequal variances) none of them control it very well. They recommended that robust alternatives be considered.

Empirical studies prior to Blair & Higgins (1980) also suffered from being limited primarily to investigations of Type I error, i.e., the ability of a test to maintain the rate at which a true null hypothesis was (incorrectly) rejected at something close to nominal alpha. There is more to a statistical test, however, than robustness with respect to Type I error. To be useful, a test should also have high power to detect effects when they exist (Cohen, 1988), i.e., it should be robust with respect to Type II errors under violation of assumptions when the null hypothesis is false. It should also demonstrate relative power superiority with respect to alternative procedures. The two-independent-samples *t*-test, for example, has been shown to be quite robust with respect to both Type I and Type II error under certain violations of the normality assumption. However, Blair & Higgins (1980) stated that:

> Generally unrecognized, or at least not made apparent to the reader, is the fact that the t-test's claim to power superiority rest on certain optimal power properties that are obtained under normal theory. Thus, when the shape of the sampled population(s) is unspecified, there are no mathematical or statistical imperatives to ensure the power superiority of this statistic. (p. 157)

Another difficulty with many studies is the concept of robustness itself. Bradley (1978) made three major points concerning this topic. The first point was that there was no generally accepted (standard) quantitative definition of robustness and that studies rarely provided their own definition. (He provided several definitions that are discussed in Chapter 3.) The second point was that the Type I error properties of a test, under violation of assumptions, were very complex and determined by the particular way in which numerous factors interacted, yet studies rarely provided sufficient qualification of the exact nature and degree of violation or carefully examined all of the possible interactions. Finally, he addressed the issue of bias in the literature, particularly the mathematical treatments of Box (1954a, 1954b), Box & Anderson (1955), and Scheffé (1959) and the empirical studies of Lindquist (1953) and Boneau (1960). Box (1953) and Feir-Walsh & Toothaker (1974) also illustrated some of these biases. Bradley (1978) stated that:

> None of these authors uses a quantitative definition of robustness. Furthermore, in every case some sort of selective bias appears to be operating and that bias always seems to favour robustness. The bias then tends to be overlooked or depreciated in summarizing the actual findings and drawing generalized conclusions. And the author's overgeneralization, underqualification or use of overly exuberant language in proclaiming robustness further tends to convey the impression that robustness is a highly general phenomenon. (p. 147)

Ironically, the second point, above, was made earlier by Box & Andersen (1955). By way of more recent example, Penfield (1994) reported the results of a Monte Carlo study on the relative performance of the *t*-test, Wilcoxon-Mann-Whitney Test, van der Waerden Normal Scores Test, and the Welch-Aspin-Satterhwaite Test using two independent random samples drawn from a variety of nonnormal distributions using combinations of both equal and unequal samples sizes as well as equal and unequal variances. The Wilcoxon-Mann-Whitney test dominated when variances were equal, regardless of sample size. The *t*-test performed best for equal sample sizes with unequal variances. The Welch-Aspin-Satterthwaite test was the most powerful for unequal sample sizes combined with unequal variances.

More recently, Wilcox (1998a) offered another view of robustness as it related to "finding population *parameters*, estimators, and hypothesis-testing methods that are not drastically affected by small changes in a distribution, *F*" (p. 5). He observed that neither μ, σ, or ρ, nor their estimators, $\bar{x}$, s^2 or r, possessed qualitative, infinitesimal, or quantitative robustness. This accounts, in part, for the poor performance of parametric tests under violations of their assumptions as "arbitrarily small changes in a distribution can make σ arbitrarily large" (Wilcox, 1998a, p. 7). This results in extremely poor power for testing hypotheses using means.

It is a common recommendation and practice, when using parametric tests, to test the sample data for assumption violations prior to subjecting it to an inferential hypothesis test. Wilcox (1998a) claimed that such attempts to try and salvage methods based on means were pointless as departures from population normality small enough to be undetectable by the Kolmogorov-Smirnov Test (of general differences) cause serious problems, especially with population distributions that are near normal but heavy-tailed.

With the tendency for research designs to become ever more complex, requiring multivariate hypotheses and tests, the lack of availability of distribution-free alternatives continued to be an issue (Emmons, et al, 1990). However, if the assumptions of parametric tests are rarely met in practice, the requirements for multivariate normality, multivariate homogeneity of variance, and lack of multicollinearity, required by many of the multivariate parametric tests, must be regarded as even less tenable.

When distribution-free statistics were finally examined in comparative studies with their parametric counterparts it was usually under conditions that met all of the assumptions of the later, a practice that continues to the present in studies such as Gibbons & Chakraborti (1991). Such studies often optimized the performance of the parametric tests relative to the distribution-free ones and the distribution-free alternatives were usually found to be relatively less 'efficient', even though the difference in efficiency was often small. There were exceptions, however, such as Blair & Higgins (1980, 1985).

Bradley (1968) addressed the false perception that distribution-free tests failed to use information relative to their parametric counterparts:

> ...the utilization of the additional sample information is made possible by the additional population "information" embodied in the parametric test's assumptions. Therefore, the distribution-free test is discarding information only if the parametric test's assumptions are known to be true. (p. 12-13)

This is the more general statement of the point cited earlier from Blair & Higgins (1980) in connection with the power properties of the *t*-test.

Bradley (1968) claimed that professionals had overcome their earlier biases against distribution-free tests, yet recent reviews of statistical methods used in education, psychology, and sociology suggest otherwise (Emmons, Stallings & Layne, 1990; Elmore & Woehlke, 1998). Emmons, et al, (1990) surveyed 2,674 articles from three journals, for the years 1972 to 1987, and found that out of 41 statistical analysis techniques, ANOVA was used 37% of the time, followed by regression 13% of the time. A decrease in use was noted over that period for numerous techniques, including nonparametric ones. Elmore & Woehlke (1998) surveyed three journals from the American Educational Research Association from 1978 to 1997 and found that the seven preferred methods of data analysis, in rank order, were: ANOVA / ANCOVA (analysis of covariance), multiple regression, bivariate correlation, descriptive statistics, multivariate tests, nonparametric tests, and the *t*-test.

Cliff (1993, 1996a, 1996b) recently advanced the argument that ordinal statistics may be the most appropriate for the vast majority of social and behavioral science research and should, therefore, be more widely used. Although his argument was specifically aimed at dominance statistics (which were not part of the present study) they apply generally to modern methods, including distribution-free procedures. Wilcox (1995, 1998a, 1998b), although a proponent of robust parametric methods, advanced the argument that, while most significant research results appearing in the literature probably represent the correct detection of a true effect, many research efforts have resulted in Type II errors, with an attendant loss of knowledge.

As presented in Cliff (1993, 1996a, 1996b) the argument was in several parts. First, data in social and behavioral research, or the latent constructs that the data represent, have only ordinal justification. Second, conclusions based on ordinal data are invariant under ordinal transformations, which is not true of metric transformations. When data are not invariant under transformation, inferential conclusions may be altered. Third, the research questions of interest are themselves ordinal, such that ordinal methods answer them more directly. Fourth, ordinal methods are typically more statistically robust, when applied to real data, in the sense that results are more inferentially valid and generalizable. Fifth, ordinal methods have greater statistical resistance in the sense that they are less sensitive to being influenced by a relatively small number of observed data points, such as outliers. Collectively, these arguments speak to the issues of justification of method and the rhetoric of statistics as principled argument (Abelson, 1995).

Interestingly, Cliff (1996a) expressed reservations about some robust alternatives to his proposed procedures, specifically resampling procedures, such as randomization and bootstrap methods, as well as differential data-weighting methods, such as trimmed means or variances (as cited in Wilcox, 1996). One of his concerns with randomization approaches, such as the Wilcoxon-Mann-Whitney Test, was that they assumed populations were identical except for location, which he regarded as unrealistically narrow. In regards to bootstrap procedures, he suggested that they had not lived up to their promise, citing Wilcox (1991) with respect to inferences about Pearson *r* and citing Westfall & Young (1993) with respect to location comparisons.

Cliff (1996a) also took exception to Wilcox (1996), however, on the basis that differential data-weighting methods effectively altered the sampled population with nontrivial consequences, especially for asymmetric populations. He indicated that these alterations were unjustified because researchers had become sophisticated enough in their data collection and screening procedures that obtained data should be regarded as valid. Finally, he took exception to the complexity of the methods.

Hettmansperger (1998) echoed many of these concerns in his reply to Wilcox (1998a). Hettmansperger, McKean & Sheather (2000), however, argued in favor of robust nonparametric methods, noting in particular the need for better techniques to deal with uncontrolled quasi-experimental research designs and observational studies as well as data mining of large, unclean datasets. Zimmerman (1995) also advocated the use of robust nonparametric statistics, based on applying outlier detection and down-weighting procedures with nonparametric tests, as a way to increase their power. He noted that while the Wicoxon-Mann-Whitney Test enjoyed substantial power superiority over the *t*-test, when applied to samples from outlier-prone distributions, both tests suffered a loss of power.

Another issue of central importance to empirical studies is the methodology used to generate distributions and draw samples from them. Fleishman (1978) described four methods found in the literature for generating nonnormal distributions for Monte Carlo studies: 1) outliers, 2) extreme inverse probability functions on uniform deviates, 3) transformation of a random normal deviate by a skewing function, and 4) tabular. Fleishman (1978) developed a specific form of method 3), in the tradition of mathematical convenience, although he noted that Pearson and his students did the earliest work on nonnormality using tabular distributions. Penfield (1994) is typical of the subsequent use of the Fleishman (1978) power method. Tadikamalla (1980) offered five alternatives to Fleishman (1978) and compared their performance. Headrick & Sawilowsky (1999a, 1999b, 2000) extended the Fleishman (1978) power method to generate multivariate nonnormal data. Sawilowsky, Blair, & Micceri (1990), however, used the table method to implement the Micceri (1986) distributions.

Blair (1981) observed that "Ceiling effects, floor effects, presence of large minority groups, special scoring conventions as well as interactions between these phenomena are only some of the factors that give rise to bizarre shapes in educational data" (p. 504). Bradley (1975, 1977) reported on situations that lead to ‘bizarre’ distribution shapes such as ‘time-to-completion’ data viewed as a performance measure, noting that times tended to be somewhat normally distributed for highly error prone situations (difficult task, inept operator) but became “violently nonnormal” (Bradley, 1975, p. 321) for relatively easy tasks and/or competent operators. Pearson (1895) wrote on the problem of restriction of range for physical and biological measurements, noting that “…it *theoretically* excludes the use of the normal curve in many classes of statistics…” (p. 359) (emphasis in the original).

Micceri (1986, 1989) took on the Normal Mystique through a meta-analysis of 440 large-sample data sets from published research in social and behavioral sciences from 1982 through 1984. The picture that emerged strongly suggested that the populations under study were discrete and distinctly nonnormal. Micceri (1989) submitted that “This research was limited to measures generally avoided in the past, that is, those based on human responses to questions either testing knowledge (ability/achievement) or inventorying perceptions and opinions (psychometric)” (p. 157).

The Micceri (1986) data sets fall into four categories: 1) achievement [n=231], 2) psychometric [n=125], 3) criterion mastery [n=35], and 4) gain scores [n=49]. The sources of the data sets included: a) journal articles and researchers [265], b) national tests [30], c) statewide tests [64], d) district-wide tests [65], and e) college entrance and GRE tests [17]. None of the data passed tests of normality and very few were even reasonable approximations, with only 28.4% relatively symmetric while 30.7% were extremely asymmetric (Micceri, 1989).

Cliff (1989) asserted that the "...simplest view of testing is that it takes primary ordinal information in the form of test responses and converts it to summary ordinal information about the examinees" (p. 75). Nunnally & Bernstein (1994) remarked that "Real test scores based upon item sums are rarely normally distributed, even if the number of items is large, because items on a real test are positively correlated and not uncorrelated (independent)" (p. 168). Micceri (1989) offered his own account of factors that might contribute to a non-Gaussian distribution:

> (a) the existence of undefined subpopulations within a target population having different abilities or attitudes, (b) ceiling or floor effects, (c) variability in the difficulty of items within a measure, and (d) treatment effects that change not only the location parameter and variability but also the shape of a distribution. (p. 157)

The eight empirical distributions of Micceri (1986) included three psychometric measures and five achievement measures. The psychometric measures were: 1) discrete mass at zero with gap, 2) extreme asymmetry, and 3) extreme bimodality. The achievement measures included: 1) mass at zero, 2) extreme asymmetry, 3) multimodal and lumpy, 4) digit preference, and 5) smooth symmetric.

The assumption of normality was clearly violated in all of the Micceri (1986) distributions, which were made available as a PC Fortran subroutine module described in Sawilowsky, Blair & Micceri (1990). The distributions have subsequently been used to perform Monte Carlo studies (see for example, Sawilowsky & Markman, 1989; Sawilowsky & Brown, 1991; Sawilowsky & Blair, 1992; Sawilowsky & Hillman, 1991, 1992; Bridge & Sawilowsky, 1997; Fahoome, 1999). Using these distributions, Sawilowsky and Blair (1992) concluded that Student's *t*-Test was robust with respect to Type I error when the test was two-tailed with large samples of equal size. However, for extremely asymmetric distributions the *t*-test was non-robust with respect to Type I error.

The assumptions required for most parametric tests do not appear to be justified in the context of social and behavioral science research. Given that distribution-free tests hold out the promise of inferences that lead to conclusions that are more valid, defensible, and generalizable than those that can be arrived at using parametric tests, researchers in these fields should have a natural interest in distribution-free methods and a working knowledge of their properties and correct application. Even so, distribution-free tests are not without their mathematical assumptions (necessary and sufficient preconditions). To properly apply these tests a researcher must have knowledge of their assumptions as well as the robustness and resistance of the tests to violations of those assumptions.

The terms 'nonparametric' and 'distribution-free' are not synonymous, nor are they mutually exclusive. They stand together in contrast to the term 'parametric', which refers to tests/statistics used to make an inference about a parameter in a population (probability density function) with specific distributional requirements using data that is at least on an interval scale (Gibbons, 1985). Note, however, that the last part of this statement is a vestige of the weak-measurement versus strong-statistics debate, and is not strictly correct.

The term 'nonparametric' describes a test that does not involve a hypothesis about a parameter in a statistical density function whereas the term 'distribution-free' describes a test that makes no assumptions about the precise form of the sampled population (Gibbons, 1985). "Of the two terms, *distribution-free* comes closer to describing the qualities that make the tests desirable" (Bradley, 1968, p. 15).

The Sign test, for example, is a distribution-free test of the hypothesis that the parameter p of a binomial distribution is 0.5 whereas the Wilcoxon Signed-rank Test is a nonparametric test based on the distributional assumption of population symmetry. The Wilcoxon-Mann-Whitney Test (rank-sum) is both nonparametric and distribution-free, but presumes that the populations, under the alternative hypothesis, are identical except for location.

Although most distribution-free tests make very few assumptions, they do make some. Distribution-free tests generally share with parametric tests the fundamental assumptions that sampling and assignment are random, independent, and free of any systematic bias, as well as the distributional assumption that samples are drawn from continuous populations.

Zimmerman & Zumbo (1992a) reported that violation of random sampling produced "nonindependence of successive sample values within groups" (p. 377) resulting in higher Type I error rates, lower Type II error rates, and a spurious increase in the probability of rejecting the null hypothesis. Likewise, violation of random assignment manifested itself as "nonindependence of sample values between groups" (Zimmerman & Zumbo, 1992a, p. 378), resulting in lower Type I error rates, higher Type II error rates and a spurious decrease in the probability of rejecting the null hypothesis. Although these results were obtained with the *F*-

test across the normal and four theoretical nonnormal distributions, they suggest the difficulties that violation of such basic assumptions invites.

Other assumptions, such as symmetry or identical shapes, may also exist for specific tests. The assumptions, however, generally stop far short of a completely specified distribution, such as the requirement for normality. Further, the assumptions are often sufficient conditions for complete validity of the test rather than necessary ones, such that the test may remain valid even if the assumptions are relaxed slightly (Gibbons, 1985). Bradley (1978) should be kept in mind, however, when contemplating the meaning of 'slightly'.

The reason that only limited population assumptions are needed with distribution-free tests is that the focus of these tests is on the samples themselves. They tend to use sample-related characteristics of the observations, such as ranks, rather than population-related characteristics, such as magnitudes (Bradley, 1968).

A possible point of confusion when working with distribution-free tests is that any statistical test involves at least three different distributions: a) the population, b) the observation-characteristic actually used by the test; and c) the test statistic itself. "The distribution from which the tests are 'free' is ... the sampled population. And the freedom that they enjoy is usually relative" (Bradley, 1968, p. 15). Distribution-free tests use ordinal or categorical characteristics of the sample data rather than magnitudes, so they do not need to make assumptions about the distribution of these magnitudes (probability density function) in the population.

Distribution-free tests require knowledge of the distributions of the sample characteristic of interest and of the test statistic, but these are both generally available. When the limited assumptions are met, the distribution of the sample characteristic is both discrete and completely specifiable based on *a priori* considerations. When the sample is also finite (as it always is in practice) the distribution of the sample characteristic becomes finite as well. Under these circumstances, the distribution of the test statistic can be calculated exactly using combinatorial formulas (Bradley, 1968). Thus, the difference between parametric and distribution-free tests is not one of total dependence on, versus total independence from, distributional requirements. Rather, it is a matter of which distributions one must have information about and whether that information can reasonably be obtained or assumed.

The assumption that sampled populations are continuously distributed leads immediately to the consequence that the theoretical probability of any specific score is zero and that the probability, therefore, of equal observations is also zero, such that zero difference-scores or tied ranks are presumed not to occur (Bradley, 1968). The test statistic, however, only has its null, or tabled, distribution when: 1) the assumptions on which it is based, and 2) the null hypothesis that it tests, are both true.

Continuity is a fundamental assumption that is routinely violated in the application of almost all statistical tests to almost all real (measured) data. For tests based on ranks, continuity leads to the requirement that a total of *N* observations can be uniquely assigned ranks using consecutive, non-repeating integers from 1 to *N*. In practice this requirement can be relaxed by making a distinction between ties that are of consequence and those that are not (critical vs. non-critical).

Consequential (critical) ties are those for which differing resolutions could produce differing values of the test statistic that in turn could result in differing inferential conclusions. By contrast, ties for which all possible assignments of ranks result in the same value of the test statistic are non-consequential (non-critical) and can be arbitrarily assigned the set of consecutive ranks for which the observations are collectively tied without any adverse effect on the performance of the test (Bradley, 1968).

In practice, equal observations (zero difference-scores, tied ranks) occur regularly so that most populations, based on the samples, appear to be discrete. "Sometimes, of course, the necessary assumptions are simply violated, the test being used as an approximate test, as is almost always the case with parametric statistics" (Bradley, 1968, p. 48). More often, an attempt is made to compensate for the consequences of violating the continuity assumption by resolving zero difference-scores or tied ranks in some manner.

Gibbons & Chakraborti (1992) described two basic approaches for dealing with ties. The first involved resolving (removing) the ties in some manner prior to conducting the test. These methods effectively alter the sampled data so that the data to which the test statistic is actually applied appears to meet the underlying assumptions of the test. The second class of methods involved computing the test statistic in two or more ways, up to and including all possible resolutions of ties, and then combining the resultant values in some manner to form a single value of the statistic for testing. A third approach, available for some tests, involves a formulaic correction of the test statistic for the presence of ties.

In all three approaches the aim is to utilize existing tables of critical values and/or critical value approximation formulas, as is often the case with larger sample sizes. These techniques stand in contrast to exact (permutation) methods, methods based on bootstrap procedures, or the use of Monte Carlo simulations to compute or estimate the actual probability of the result.

Equal observations (zero difference-scores, tied ranks) arise in practice from imprecision in measurement or an inherently discrete population. The difference turns out to be immaterial in one sense, but not in another. It is immaterial in the sense that the distribution of measurements is indistinguishable regardless of the cause. Bradley (1968) wrote:

> ...if one of the methods ... is used upon the zero differences or equal observations, the numerical test outcome will be exactly the same for a given set of sample observations, irrespective of which one of the three situations obtained. And irrespective of whether H_0 is true or false, and of whether assumptions are met or failed, the distribution of the test statistic will be the same under all three situations. ... All of these considerations point to the conclusion that the discrete case can be treated in the same way as the continuous case. (p.48)

The distinction is not immaterial, however, in another sense. If the assumption of population continuity is tenable then the existence of zero difference-scores or tied ranks is an artifact resulting from a lack of precision in measurement that obscures real differences that actually exist in the population. As such, there is a 'true' value of the test statistic corresponding to one of the possible resolutions of the ambiguous scores, although it is not possible to know which resolution is the correct one (Bradley, 1968).

There is only one method that deals with this situation directly: Probability Bounds (Bradley, 1968; Gibbons & Chakraborti, 1992). Bradley (1968) refers to this as Method A, in which the test statistic is calculated twice, once resolving all consequential ties in the direction most likely to reject the null hypothesis and once in the direction least likely to do so. The probability levels associated with the test statistic T for these two resolutions are α_M and α_L respectively. The test outcome is then simply reported as $\alpha_M \leq P(T) \leq \alpha_L$. For nominal probability level α, if $\alpha \leq \alpha_M$, the null hypothesis is not rejected as both bounds lie above nominal alpha. Likewise, if $\alpha_L \leq \alpha$, then the null hypothesis is rejected as both bounds lie below nominal alpha. However, if $\alpha_M \leq \alpha \leq \alpha_L$, the test is inconclusive. The only resolution at this point would be to collect more data until both bounds fall on the same side of nominal alpha. The fact that this is not possible in many research situations is seen as a limitation of this method.

If a discrete population is the more tenable assumption, then equal observations do not necessarily require resolution as they actually occur in the population and naturally lead to zero difference-scores or tied ranks. However, for small to moderate sample sizes the observed pattern of zero difference-scores or tied ranks may not represent, even roughly, the actual pattern that exists in the population (Bradley, 1968).

Although method A is the most theoretically sound method for dealing with ties (when the continuity assumption holds) it is rarely used in practice. One reason is that it does not always result in a decision and inconclusiveness is inconvenient. Another reason may be that critical value tables are designed to allow comparison of an obtained test statistic to best conservative critical values (or estimates) at pre-specified probability (nominal alpha) levels, while method A, and several others discussed below, require the actual probability level of the obtained test statistic. This probability level is not, in general, available from tables or easily computed. Finally, without the continuity assumption, method A becomes more vague and less appropriate. Rather than α_M and α_L representing absolute bounds on *P(T)*, they provide, at best, an estimate of an interval that contains *P(T)* (Bradley, 1968).

Many other methods have been suggested for dealing with tied observations, or rather, the problem of tied ranks in order based statistics resulting from working with real (measured) data. Ties within a group are often non-critical whereas ties between groups are often critical, although this depends on the specific test. For some tests, formulaic corrections are available in conjunction with a particular method of resolving the tied ranks. In textbooks, each test is often accompanied by recommended and/or preferred methods of dealing with ties as detailed later. These methods include, but are not limited to, the following:

B – Drop tied observations prior to ranking and reducing *n* accordingly (Bradley, 1968; Gibbons & Chakraborti, 1992; Hájek, Šidák & Sen, 1999).

C – Resolve ties through random assignment (Bradley, 1968; Gibbons & Chakraborti, 1992; Hájek, Šidák & Sen, 1999).

D – Resolve ties to obtain some sort of average probability, using the mean, median, mid-range value or other weighted computation (Bradley, 1968; Gibbons & Chakraborti, 1992).

E – Alternately assign tied scores to the groups (Bradley, 1968).

F – Assign the mid-rank to all members of a set of tied values (Bradley, 1968; Gibbons & Chakraborti, 1992; Lehmann, 1998; Hájek, Šidák & Sen, 1999).

G – Resolve consequential ties in all possible ways, obtain the test statistic for each resolution, and then calculate the 'average' value of the statistic to use for the test (Neave and Worthington, 1988; Gibbons & Chakraborti, 1992; Hájek, Šidák & Sen, 1999).

H – Resolve ties in the manner least favorable to rejection of the null hypothesis (most conservative test, lowest probability of making a Type I error) (Gibbons & Chakraborti, 1992; Hájek, Šidák & Sen, 1999).

I – Resolve ties using average scores (Hájek, Šidák & Sen, 1999).

J – Resolve ties using a conditional approach in which the ties are treated as fixed (Hájek, Šidák & Sen, 1999).

Hájek, Šidák & Sen (1999) indicated that numerous papers had been published on tied observations subsequent to Hájek & Šidák (1967). Their review of that literature indicated that the methods most often presented were "a) the method of randomization, b) the method of average scores, c) the method of conditional fixing of ties" (Hájek, Šidák & Sen, 1999, p. 136). The literature they referred to, however, was mathematical statistics. Other than mid-ranks, which are a special form of the average score method, these three methods are not the ones most commonly mentioned in the applied literature.

Methods B, C, E, F, H and I are the type that alter the data prior to testing. Methods A, D, and G are the type that calculate more than one value of the test statistic. Methods D and G combine the resulting test statistics in order to arrive at a single, definitive inferential decision. Only method A leaves open the possibility that a definitive decision (reject, fail to reject) is not supported by the data and thus is inappropriate.

A selective survey of nonparametric (distribution-free) textbooks revealed that theoretically oriented mathematical statistics texts, such as Bradley (1968), Gibbons & Chakraborti (1992), and Hájek, Šidák & Sen (1999) included general treatments of the origins and consequences of ties, accompanied by a review of methods for dealing them. By contrast, applied nonparametric textbooks aimed at social and behavioral researchers, such as, Conover (1999), Deshpande, Gore & Shanubhogue (1995), Hollander & Wolfe (1999), Neave & Worthington (1988), Siegel & Castellan (1988), and Sprent & Smeeton (2001) lacked such general treatments.

Theoretical texts quickly adopted the continuity assumption and dispensed with the need for any further treatment of ties, whereas the applied texts dealt with ties on a test-by-test basis, usually by means of a single, 'preferred' method. Methods B (drop and reduce n) and F (mid-ranks) were the most frequently suggested, although Neave & Worthington (1988) often suggested method G, the use of a mean value of the statistic.

Each of these methods has potential problems. Method B, for example, results in reduced sample sizes and possible loss of power, and can result in unequal sample sizes, rendering it potentially unsuitable for repeated measures tests. It also introduces bias toward rejection of the null hypothesis (Gibbons & Chakraborti, 1992). However, when populations are assumed to be discrete, this method makes more sense as the zero difference-scores or tied ranks are real (Bradley, 1968). Also, the loss of power and bias may be minimal if the number of dropped scores is 'small' relative to the sample size, although 'small' is rarely quantified. Methods B, C, D, E, and G, however, are based on the assumption of equiprobability, which is only meaningfully under the assumption of continuity.

Method D suffers from the same basic difficulties as method A, the lack of easily available probability levels for obtained statistics. It also suffers, along with method G, from potential computational complexity depending on how the average probability is found. If a mid-range value is used, one need only calculate $(\alpha_M + \alpha_L)/2$, with α_M and α_L as described previously. If a mean or median value is desired, then the tied ranks must be resolved in all possible ways. This can become computationally daunting if multiple ties occur at multiple values. In general, if there are n_i ties at k different values, the maximum number of arrangements is $\prod_{i=1}^{k}(n_i!)$. Although only consequential ties need to be resolved, this could still become impractical for moderate to larger samples even with modern computers.

Method F often results in ranks that are not integers and still results in multiple data values that are assigned the same 'rank'. This may lead to values of a test statistic that are not integral, even though the values should be integers when the assumptions are met.

Bradley (1968) asserted that the application of methods B through E was only valid under the specific additional assumption of equiprobability. This assumption requires that when the null hypothesis is true, all possible resolutions of ambiguous scores (consequential ties) are equally likely, *a priori*, to represent the true situation. Thus zero difference-scores must be as likely to represent true positives as true negatives. The same logic applies to method G.

Given the assumption of continuity, all of the additional assumptions of methods B through E and G, including equiprobability, are met when two sufficient conditions are met. Bradley (1968) gave these as:

> (1) in the case of tests based on signs of difference-scores, *the difference-score population is symmetric about zero* when H_0 is true, or, in the case of tests based on ranks, *all of the sampled populations are identical* (and unchanging) when H_0 is true; (2) although perhaps imprecise, *measurements are unbiased*, i.e., there may be variable error but there is no constant error in measurements. (p. 50)

Method F (mid-ranks) is widely recommended and much used in practice (Gibbons & Chakraborti, 1992). It implicitly invokes the assumption of equiprobability, and, for most tests, violates other necessary assumptions as well (Bradley, 1968). Mid-ranks, by definition, occur more than once and thus do not represent a unique ordering. They are also often non-integral.

An unstated assumption with mid-ranks is that their use will yield a value of the test statistic that is approximately the average (mean or median) value of the statistic under all possible resolutions of consequential ties. In practice the resulting value of the test statistic can be much closer to the minimum or maximum value than to a middle value (Bradley, 1968). The consequence of violating these assumptions is that the distribution of the test statistic is altered to a degree that depends on the number and sizes of groups of equal observations and is conditional upon an idiosyncrasy of the sample (Bradley, 1968). Under such conditions, modern computer-based Monte Carlo methods offer one of the only practical means of studying the behavior of a test statistic.

Method G closely resembles method D, but averaging test statistics is not necessarily the same thing as averaging their probabilities. Neave & Worthington (1988) employed the arithmetic mean when finding averages, although medians and mid-range values could also be used. The mean may be justifiable if the distribution of the statistic is known to be somewhat normal, or at least symmetrical, but the median becomes equivalent under these circumstances and is generally easier to determine. Under symmetry, the mid-range value would also tend to correspond to the mean and median and is computationally much more efficient.

Equiprobability depends on population continuity and cannot be directly reconciled with population discreteness. Bradley (1968) suggested that to avoid this apparent dilemma:

> … simply omit continuity from the sufficient assumptions that guarantee equiprobability, thereby, in a sense, removing a "pure continuity" factor from the conditions which justify the use of the methods. (p. 55)

Thus, when it comes to the treatment of ties in distribution-free tests, it matters little whether discreteness is due to imprecision in measurement or is characteristic of the population, as long as all sufficient conditions are met, other than continuity (Bradley, 1968).

Bradley (1968) critiqued each of the methods (A – F). All of them suffer from the fact that the data is ambiguous as to the true situation, resulting in a loss of power. Method A makes an probability statement that is true of the entire sample, but loses power because it fails to reject the null hypothesis when $\alpha_M < P(T) < \alpha < \alpha_L$, even though rejection is appropriate. Methods B through F only estimate the exact probability level for the entire sample without providing information on the precision of estimate. Method B makes a probability statement that is true for the unambiguous portion of the sample, but loses power though reduction in sample size. Methods C through F assume that H_0 is true which tends to force the test statistic to have its null distribution, reducing power without accounting for that reduction. Bradley (1968) concluded that method A (probability bounds) was the most justifiable, followed by method B (drop and reduce *n*) and then C (random assignment), with C being used only in the case of discrete populations.

In spite of the apparent theoretical advantage for methods A, B, and C, method E (alternating assignment) has the attraction of being easy to implement (especially on a computer). Method F (use of mid-ranks) is one of the most widely suggested methods for dealing with tied ranks, along with method B.

Method G, which is similar to method D, has support in the literature and is much more practical to implement.

The effect on statistical tests under any of these methods is the same; they become approximate tests. Tables of critical values do not then apply exactly as they would when all assumptions are met and must, therefore, be applied in a conservative fashion, resulting in an additional loss of power.

Although distribution-free tests enjoy relative simplicity of derivation and ease of calculation in use, they tend to suffer from an ongoing lack of availability of extensive tables of critical values. Even when textbooks have such tables they are often abbreviated, with tables for unequal sample sizes, large sample sizes, and larger numbers of comparison groups being especially scarce. Also, the actual probability levels are rarely given.

Fay (2002) provided a review of the literature on the generation of critical values and their associated probabilities for the Kolmogorov-Smirnov Test of General Differences, Tukey's Quick Test of Location, Rosenbaum's Test, and the Wilcoxon-Mann-Whitney (rank-sum) Test. Fortran 90 programs for generating critical value tables with associated probabilities were also provided. These programs were developed as part of the present study and used to generate the corresponding critical values and associated probabilities.

The author also attempted to develop Fortran programs to generate critical values and associated probabilities for the Kruskal-Wallis Test and the Terpstra-Jonckheere Test, but without complete success. Useful references were Iman, Quade and Alexander (1975) for the Kruskal-Wallis Test and Odeh (1971) for the Terpstra-Jonckheere Test.

Many authors provide large-sample approximation formulas for use with these tests. Fahoome (1999) studied this extensively for equal sample sizes, and the reader is well advised to consult this work before relying on these formulas. Fahoome (1999) dealt with the issue of ties, however, by resolving them on a test-by-test basis using the single method that was most recommended (preferred) or most practical to implement. These results were also reported in Fahoome (2002).

The six tests investigated in this study are distribution-free and/or nonparametric tests of location or general difference. For each test the following are described: background; assumptions and hypotheses; procedure and test statistic; null distribution (absent ties); critical values, rejection region and sample size; consequences of ties and recommended methods of resolution; an example; and comments (if needed).

2.2 — Two–independent–samples Omnibus Tests (Tests of General Differences)

2.2.1 — Kolmogorov–Smirnov Test of General Differences

Background

Neave & Worthington (1988) and Conover (1999) identified this as Smirnov's (1939) application of Kolmogorov's (1933) goodness-of-fit-test. Vogt (1999) described it as a "Nonparametric test of whether two distributions differ and whether two samples may reasonably be assumed to come from the same distribution" (p. 151). Everitt (1998) described it as "A distribution free method that tests for any difference between two population probability distributions. The test is based on the absolute maximum difference between the cumulative distribution functions of the samples from each population" (p. 179). The maximum distance referred to is the vertical distance between the (graphs of the) cumulative probability distributions.

Assumptions and Hypotheses

Random and independent sampling of continuous populations is assumed with sufficient precision of measurement to avoid tied observations (Bradley, 1968; Conover, 1999). Independence of sample observations is assumed to exist both within and between groups (Hollander & Wolfe, 1999). The test can be used successfully with discrete populations, but the test becomes approximate with the tabled critical values providing best conservative estimates.

The null hypothesis for the two-sided test is that the two sampled populations have identical distributions, or H_0: [$F_A(x) = F_B(x)$ for all x from $-\infty$ to ∞] (Conover, 1999). The two-sided alternative hypothesis is simply that the two sampled populations are different in some way, or H_1: [$F_A(x) \neq F_B(x)$ for at least one value of x] (Conover, 1999).

In the case of a one-sided test the alternative hypothesis is that one population is stochastically greater than the other. For the case where the A population is suspected of having smaller values than the B population, Conover (1999) gave the null and alternative hypotheses as H_0: [$F_A(x) \leq F_B(x)$ for all x

from $-\infty$ to ∞] versus H$_1$: [$F_A(x) > F_B(x)$ for at least one value of x]. For the case where the B population is suspected of having smaller values than the A population, Conover (1999) gave the null and alternative hypotheses as H$_0$: [$F_A(x) \geq F_B(x)$ for all x from $-\infty$ to ∞] versus H$_1$: [$F_A(x) < F_B(x)$ for at least one value of x]. These hypotheses capture the idea that one of the populations is stochastically greater than the other. Neave (1981) suggested that the test only be used in the two-sided situation, the Wilcoxon–Mann–Whitney test being more powerful for the directional hypothesis.

Procedure and Test Statistic

This test can be used with unequal sample sizes. In general, assume two samples, A and B, with n_A and n_B observations respectively. Observations are combined and then ranked, keeping track of original sample membership. Conover (1999) defined the test statistic, T, in terms of two empirical distribution functions, S_A and S_B, using the supremum.

For the two-sided test, $T = \sup_x \left| S_A(x) - S_B(x) \right|$. For the one-sided test that A < B (stochastically), $T^+ = \sup_x \left[S_A(x) - S_B(x) \right]$. Finally, for the two-sided test that A > B (stochastically), $T^- = \sup_x \left[S_B(x) - S_A(x) \right]$.

The following procedure, described in Neave & Worthington (1988), is consistent with Conover (1999) above. Let there be $N = n_A + n_B$ ranked observations, each designated as an A or B. For the 'A's, maintain a count above the letter sequence, starting from zero and incremented by n_B each time an A is encountered. For the 'B's, maintain a count below the letter sequence, starting from zero and incremented by n_A each time a B is encountered. The final count for both 'A's and 'B's should be $M = n_A \times n_B$. Now compute the differences, $d_i = B_i - A_i$, by subtracting the A counts from the B counts for each letter position. Finally, find the absolute value of these differences.

For the two-sided test, take $D^* = \max|d_i|$. For a one-sided test take $D_+^* = \max|pos(d_i)|$ or $D_-^* = \max|neg(d_i)|$ depending on what is expected under H_1. Specifically, if sample A is believed to come from a population that is stochastically greater than population B, the sample A cdf should reflect this and lie to the right of the sample B cdf, such that the sample B cdf will rise more quickly than the sample A cdf. Thus, the sample B cdf would be expected to lie above (vertically) the sample A cdf and D_+^* would be the appropriate statistic. Note that $D^* = n_A n_B D$, where D is the statistic derived from a direct comparison of the sample cdf's, but is generally more convenient to use (Neave & Worthington, 1988).

A normal approximation formula is available for large sample sizes (Hollander & Wolfe, 1999). Fahoome (1999) provided an extensive study of the issue of adequate (equal) sample sizes for the use of normal approximations either in the absence of ties or with the use of mid-ranks to resolve them.

Null Distribution (absent ties)

The one-sided null distribution results from determining all possible letter sequences for a given $N = n_A + n_B$, determining the value of $D_{AB} = \max(F_A - F_B)$ (or D^*) for each arrangement, constructing a frequency table of the resulting values, and dividing by the total number of arrangements, which is given by $\binom{N}{n_A}$ (Neave & Worthington, 1988). The two-sided distribution is found by taking $D = \max[D_{AB}, D_{BA}]$. The one-sided null distribution is not symmetric and is not easily related to the two-sided null distribution, although this tends to be more of a concern with large sample sizes since the non-symmetry does not diminish and the asymptotic distribution is not normal (Neave & Worthington, 1988). This is consistent with the presentations in Bradley (1968) and Conover (1999).

Critical Values, Rejection Region and Sample Size

Critical regions are usually tabulated as $D^* \geq \textit{critical value}$, since D^* takes only integer values and is more convenient than working with D (Neave & Worthington, 1988). When the test is used with discrete populations, the tabled critical values become best conservative estimates. The smallest sample size that can be tested is 3 per group, one-sided, at $\alpha = .05$.

Siegel & Castellan (1988) contained tables for one-sided and two-sided tests of per-group sample sizes from 1 to 25 suitable for testing with unequal sample sizes. For samples exceeding 25, they provided a formula, adapted from Smirnov (1948), for determining the critical value. Fahoome (1999) studied the use of this approximation formula and it is not considered further here.

Wilcox (1997) noted that the Kolmogorov-Smirnov test has all but disappeared from applied research and introductory textbooks in social and behavioral science. He argued that it should be retained because it can have relatively high power even for shift effects and even in comparison to robust measures of location. It does suffer, however, from exact Type I error probabilities that can be unacceptably different from nominal alpha at small sample sizes. At n=m=14 the best conservative critical value for α=.05 has probability .019, the next available value having probability .12.

Consequences of Ties and Recommended Methods of Resolution

Tied observations are only of consequence if they occur in the region of maximum relevant difference. Neave & Worthington (1988) suggested two methods for dealing with consequential ties, non-consequential ties being ignored. The first method was to defer incrementing either the A or B array until the end of a sequence of ties, both arrays being incremented by the appropriate amounts at that point, and appeared to be unique to this test. The second method was to resolve consequential ties in all possible ways, calculating the statistic for each resolution, and then calculating the average (mean) value of the statistic for use in the test.

Example

Consider samples A: {1, 2, 3, 4} and B: {2, 4, 6, 8, 10} such that $n_A = 4$ and $n_B = 5$ with $N = n_A + n_B = 4 + 5 = 9$ and $n_A \times n_B = 4 \times 5 = 20$. To test H$_1$: $F_A(x) > F_B(x)$ (stochastically), combine and rank the samples to get {1, 2, 2, 3, 4, 4, 6, 8, 10}. Using the first method described above to resolve ties yields:

Count A	5		10	15		20	20	20	20
Scores	1	2	2	3	4	4	6	8	10
Group	*A*	*B*	*A*	*A*	*B*	*A*	*B*	*B*	*B*
Count B	0		4	4		8	12	16	20

with difference scores d_i:{-5, -5, -6, -11, -11, -12, -8, -4, 0}. The alternative hypothesis requires $D_-^* = \max|neg(d_i)| = \max|-12| = 12$. The critical value is 16 from Neave (1981, p.31). Since $12 < 16$ the test fails to reject the null hypothesis.

2.2.2 — Rosenbaum's Test

Background

This test first appeared in its current form in Rosenbaum (1954), which was based on Rosenbaum (1953). In both articles, Rosenbaum cites Wilks (1942) as the original source of the formulas for deriving the critical value tables. Rosenbaum (1965) reiterated this earlier work. The test is classified as a runs test. It is a quick and easy test, but is not routinely included in textbooks on nonparametric statistics. Bradley (1968) only mentioned Rosenbaum's test as a special case of a test for excludances and includances based on the multivariate hypergeometric distribution. Everitt (1998) described the test as:

> A distribution free method for the equality of the scale parameters of two populations known to have the same median. The test statistic is the total number of values in the sample from the first population that are either smaller than the smallest or larger than the largest values in the sample from the second population. (p. 289)

Everitt's (1998) description is more appropriate for Rosenbaum's (1953) test of dispersion. Neave & Worthington (1988) included this test in their chapter on tests for general differences of two independent samples, but presented the test somewhat differently as a test for general differences between two sampled populations where spread tends to increase with an increase in the mean. They noted that while Rosenbaum (1965) originally presented this as a test for differences in spread, it was more appropriate for the use that they described. It is this later use of the test that is considered here. Thus, their suggested usage is consistent with Rosenbaum (1954).

Neave & Worthington (1988) claimed that under the conditions of an increase in spread with an increase in the median, tests such as the Wilcoxon-Mann-Whitney and Tukey's quick (both described later) have almost no power because of the change in spread. Likewise, tests for spread, such as the Siegel-Tukey test (not described in this paper), have little or no power because of the change in location.

If more general differences were suspected, or needed to be protected against, the Kolvmogorov-Smirnov test (described previously) was suggested as a better choice (Neave & Worthington, 1988). Processes that are known to be exponential or Poisson in nature, where the standard deviation is related to the mean, would be excellent candidates for analysis by Rosenbaum's test. Thus, Rosenbaum's test appears to occupy a somewhat unique place among its better-known peers.

Assumptions and Hypotheses

Random and independent sampling of continuous populations is assumed with sufficiently precise measurement to avoid tied observations (Bradley, 1968). The null hypothesis is that there is no difference in the two sampled populations, or H_0: $[F_A(x) = F_B(x)]$. The alternative hypothesis can be two-sided or one-sided. The two-sided alternative hypothesis is simply that the two sampled populations are different in some way, or H_1: $[F_A(x) \neq F_B(x)]$. In the case of a one-sided test the alternative hypothesis is that one population is stochastically greater than the other, or H_1: $[F_A(x) > F_B(x)$ (stochastically)].

Procedure and Test Statistic

The following procedure was described in Neave & Worthington (1988). For the two-sided test, determine which sample has the overall greatest value and then count the number of observations in that sample that are greater than the greatest value in the other sample and let this be R, the test statistic. For the one-sided test, determine if the greatest overall value comes from the sample whose population is hypothesized under H_1 to have the greater mean. If it does, proceed as for the two-sided test. If not, set $R = 0$.

Null Distribution (absent ties)

Rosenbaum's test does not require equal sample sizes. For samples A and B of size n_A and n_B with $N = n_A + n_B$, the probability that R is formed from 'A's and takes a value of at least h is the probability of their being at least h 'A's at the right-hand end of the sequence (Neave & Worthington, 1988). This probability is given by $p_R = \dfrac{n_A!(N-h)!}{N!(n_A-h)!}$. Asymptotically, as $(n_A,\ n_B) \to \infty$ with $\dfrac{n_A}{N} \to p$, $p_R \to p^h$, and this can be generally be used to produce best conservative critical values for sample sizes not found in tables.

Critical Values, Rejection Region and Sample Size

Critical regions are of the form $R \geq$ *critical value* (Neave & Worthington, 1988). The table of critical values must be entered with n_1 as the size of the sample from which R is calculated and n_2 as the size of the other sample. The smallest-testable sample is 3 per group for a one-sided test at $\alpha = .05$.

Consequences of Ties and Recommended Methods of Resolution

Ties that occur at the maximum value are problematic but generally support a decision not to reject H_0. Ties are only consequential if they occur at the lower end of the upper extreme run used to calculate R. Consequential ties render the situation ambiguous and the test becomes approximate. Neave & Worthington (1988) recommended dealing with consequential ties by resolving them in all possible ways and calculating a mean value of R from the resulting statistics to use for the test.

Example

Consider samples A: {1, 2, 3, 4, 5} and B: {2, 4, 6, 8, 10} such that $n_A = n_B = 5$ with $N = n_A + n_B = 5 + 5 = 10$. To test H_1: [$F_A(x) > F_B(x)$ (stochastically)], check that the largest overall value comes from sample B. Since 10 > 5 continue with the procedure. Arrange the combined observations in rank order, keeping track of the original group membership.

Scores	1	2	2	3	4	4	5	6	8	10
Ranks	1	2.5	2.5	4	5.5	5.5	7	8	9	10
Group	A	A	B	A	A	B	A	B	B	B

Since the maximum value from sample A (5) is not tied with any values from sample B, none of the ties are of consequence and $R = 3$, the length of the extreme right-hand run of 'B's. The critical value at alpha = .05 is 4. Since $R < 4$, the test fails to reject H_0.

Comments

Eleven books were checked for this test (eight practical and three theoretical). It was referenced in only three of them (one practical and two theoretical). The practical text was Neave & Worthington (1988).

2.3 — Two–independent–samples Location Tests

2.3.1 — Tukey's Quick Test

Background

This test first appeared in Tukey (1959) as the Two-sample Test to Duckworth's Specifications. It is a quick test because it only requires a few of the sample observations to be ordered. It is also compact in the sense that tables of critical values are not generally needed for most applications, there being only a limited number of critical values that occur in practice. These two characteristics combine to make the test portable. Like Rosenbaum's Test, Tukey's Quick Test is based on extreme runs and is not routinely included in applied textbooks.

Assumptions and Hypotheses

Random and independent sampling of continuous populations is assumed with sufficiently precise measurement to avoid tied observations. The test is primarily a test for differences in location of the medians of the two sampled populations and is most appropriate when there is reason to believe that the sampled populations have the same spread, or better, the same shape (Neave & Worthington, 1988).

The null hypothesis is that there is no difference in the two sampled populations, H_0: $[F_A(x) = F_B(x)]$ or no difference in the medians of the populations, H_0: $[\phi_1 = \phi_2]$. The alternative hypothesis can be two-sided or one-sided. The two-sided alternative hypothesis is simply that the two sampled populations are different in some way, H_1: $[F_A(x) \neq F_B(x)]$ or have different medians, H_1: $[\phi_1 \neq \phi_2]$. In the case of a one-sided test the alternative hypothesis is that one population is stochastically greater than the other, H_1: $[F_A(x) > F_B(x)$ (stochastically)$]$, or that there is a directional difference in the medians, either H_1: $[\phi_1 < \phi_2]$ or H_1: $[\phi_1 > \phi_2]$.

Procedure and Test Statistic

The following procedure was described in Neave & Worthington (1988). It begins by arranging the sample observations in a single combined array from least to greatest, keeping track of original sample membership, say A and B, and then ranking them.

For a two-sided test, if the minimum and maximum observed values come from the same sample then the test statistic is $T_y = 0$. If the minimum and maximum observed values come from different samples, then the test statistic is the sum of the extreme runs. For example, if the minimum value comes from sample A and the maximum from sample B then count the number of 'A's from the beginning of the array until the first B is reached, say C_L, and count the number of 'B's from the end of the array back until the first A is reached, say C_U, and set $T_y = C_L + C_U$.

For a one-sided test, if the minimum and maximum observed values come from the same sample, set $T_y = 0$. If the minimum and maximum observed values come from different samples, determine if the maximum observation comes from the sample that is expected to have the greater median. If not, set $T_y = 0$. If so, calculate T_y just as for the two-sided.

Null Distribution (absent ties)

For $N = n_A + n_B$, as $(n_A,\ n_B) \to \infty$, $\frac{n_A}{N} \to p$ and $\frac{n_B}{N} \to q = 1 - p$ such that the null distribution for a one-sided test is given as $\Pr(T_y \geq h) = \frac{pq(q^h - p^h)}{q - p}$ for any $h \geq 2$, where $p \neq q$ (unequal sample sizes) (Neave & Worthington, 1988). When sample sizes are equal, $p = q = ½$ and the probability of each of the h types of letter sequences is $2^{-(h+1)}$, giving $h2^{-(h+1)}$ in total. This arises from considering the h types of letter sequences that can give rise to $T_y \geq h$ and the sum of their resulting probabilities as illustrated in Table 2-1. In the case of the two-sided test, all of these sequences may be reversed so that all of the probabilities are doubled.

Table 2-1
Null Distribution Letter Sequences for Tukey's Quick Test

Type	Letter Sequence		Asymptotic Probability
1	A B . . .	. . . (at least h–1 'B's)	pq^h
2	A A B . . .	. . . (at least h–2 'B's)	p^2q^{h-1}
3	A A A B . . .	. . . (at least h–3 'B's)	p^3q^{h-2}
.		.	.
.		.	.
.		.	.
h–2	(h–2 'A's) B . . .	. . . (at least 2 'B's)	$p^{h-2}q^3$
h–1	(h–2 'A's) B . . .	. . . (at least 1 'B')	$p^{h-1}q^2$
h	(at least h 'A's)	(at least 1 'B')	p^hq

Adapted from Neave & Worthington (1988) p. 124

<u>Critical Values, Rejection Region and Sample Size</u>

Critical regions are of the form $T_y \geq$ *critical value* and tables are available (Neave & Worthington, 1988). However, for one-sided tests with samples sizes that are 'not too small' and 'not too dissimilar', the .05 and .01 critical values are generally 6 and 9, respectively. For a two-sided test under the same conditions, the .05 and .01 critical values are generally 7 and 10, respectively. These critical values are reported to work well for ratios of sample sizes from 1 to 1.5 (Neave & Worthington, 1988).

Equal sample sizes are not required, although tables of critical values should be employed when the ratio of larger to smaller sample exceeds 1.5. The smallest sample size that can be tested is 3 per group, one-sided at α = .05.

Tukey's Quick Test is not as powerful as the Wilcoxon-Mann-Whitney Test, so the latter is preferred for smaller samples where the ease of application of the former is not needed (Neave & Worthington, 1988). It is a valid distribution-free test, however, and can be easily applied to larger samples when many other tests cannot.

Consequences of Ties and Recommended Methods of Resolution

Equal observations are only of consequence if a) they occur between samples and b) they involve the extreme runs. Neave & Worthington (1988) suggested resolving consequential ties in all possible ways and computing the average value of the test statistic, $\overline{T}_y$. Ties at the global maximum or global minimum usually result in a non-significant finding with this method since at least one, and perhaps more, of the orderings of tied values will result in $T_y = 0$ and this will tend to pull down the value of $\overline{T}_y$.

In their example for the situation involving ties at both ends, Neave & Worthington (1988, p. 125) resolved and averaged the ties at each end independently and then formed $\overline{T}_y$ by adding the two average values together. Their example gave a left-hand tie of (AAB) that resulted in runs of 'A's of 4, 5, or 6 and a right-hand tie of (AB) that resulted in runs of 'B's of 2 or 3. These were based on combinations rather than permutations, but the result turned out to be the same. The average 'A' run was 5 and the average 'B' run was 2.5, which yielded $\overline{T}_y = 7.5$. Interpreted literally, however, for each of the three 'resolutions' of the left-hand ties there are two resolutions of the right-hand tie, or 2x3=6 possible resolutions. If (L,R) represents the combinations of left and right end runs, the possible resolutions are (4,2), (4,3), (5,2), (5,3), (6,2), (6,3) with corresponding values for T_y of 6, 7, 7, 8, 8, 9. This yields $\overline{T}_y = {}^{45}\!/_{6} = 7.5$, the same as before. This result appears to hold in general, making this procedure slightly more efficient that it would be if all combinations had to be considered.

Example

Consider samples A: {1, 2, 3, 4, 5} and B: {2, 4, 6, 8, 10} such that $n_A = n_B = 5$ with $N = n_A + n_B = 5 + 5 = 10$. To test H_1: $\phi_B > \phi_A$ check that the largest overall value comes from sample B and the least overall value comes from sample A. Since 10 > 5 and 1 < 2 continue with the procedure. Arrange the combined observations in rank order, keeping track of the original group membership.

Scores	1	2	2	3	4	4	5	6	8	10
Ranks	1	2.5	2.5	4	5.5	5.5	7	8	9	10
Groups	A	A	B	A	A	B	A	B	B	B

The tie at the score of '4' is of no consequence since it is not involved in the runs at either end.

The maximum value from sample A (5) is not tied with any values from sample B, so $C_U = 3$, the length of the extreme right-hand run of 'B's. There is, however, a tie between samples at the value '2' (mid-ranks are shown). If the order of observations is left as shown above, then the left-hand run of 'A's smaller than the smallest B is 2, thus $C_L = 2$ and $T_y = C_U + C_L = 3 + 2 = 5$. If we reverse the observations at 2, then the left-hand run of 'A's is 1, such that $C_L = 1$ and $T_y = 3 + 1 = 4$. Finally, $\overline{T}_y = (5 + 4)/2 = 9/2 = 4.5$. The critical value at alpha = .05 is 6. Since $\overline{T}_y < 6$, the test fails to reject H_0.

2.3.2 — Wilcoxon–Mann–Whitney Test

Background

Wilcoxon (1945) introduced the rank-sum version of this test for equal sample sizes in the same article as the signed-rank test while Mann & Whitney (1947) independently developed the Mann–Whitney *U* test. The two versions are procedurally different but mathematically equivalent. It is often referred to in the literature as the Wilcoxon-Mann-Whitney Test (Sprent & Smeeton, 2001). The test is applied to ordinal (rank-ordered) data. Tables of critical values are more commonly available, however, for the Mann-Whitney version of the test.

In either form this is one of the better-known distribution-free tests, and is the one that corresponds most directly to Student's *t*-Test for two independent samples. It is also a powerful test, with an asymptotic relative efficiency that never falls below 0.864 with respect to the *t*-test (Lehmann, 1998, p. 80) while it is often much more powerful under conditions that violate the assumptions of the *t*-test yet respect its own assumptions (Blair & Higgins, 1980).

The Wilcoxon-Mann-Whitney Test is generally regarded as a test of whether two independent samples represent the same population versus populations that differ in location, either of their medians or with respect to the rank-ordering of their scores (Sheskin, 1997). Bergmann, Ludbrook, & Spooren (2000) described it as a test of group mean ranks or, equivalently, rank sums, for testing two different hypotheses: a) a shift in otherwise identical populations, and b) a difference in mean ranks between randomized groups. A detailed theoretical treatment of the test was given in Lehmann (1998). Kruskal (1957) detailed the history of the test from 1914 to that time.

Assumptions and Hypotheses

Under the population model, samples are assumed to be randomly and independently drawn from continuous populations with identical probability distributions other than a constant shift under H_1 (Sheskin, 1997). This is known as a translation, or location–shift model. Measurement is assumed to be sufficiently precise to avoid tied observations (Bradley, 1968). Independence is assumed both within and between samples (Conover, 1999; Hollander & Wolfe, 1999).

If F and G are the population distribution functions, the null hypothesis is H_0: $[F(x)=G(x), \forall x]$ (Hollander & Wolfe, 1999; Conover, 1999). Siegel & Castellan (1988) gave the null hypothesis as H_0: $[P[X>Y]=\frac{1}{2}]$. The location-shift model requires $G(x)=F(x-\Delta), \forall x$. The null hypothesis is then H_0: $[\Delta=0]$ (Hollander & Wolfe, 1999).

Neave & Worthington (1988) gave the null hypothesis as no difference in the medians of the populations, or H_0: $[\phi_1=\phi_2]$. With equal sample sizes this is equivalent to the hypothesis that the sum of ranks for each group is the same, or H_0: $[\sum R_1=\sum R_2]$. For unequal sample sizes this generalizes as the mean rank of the groups being equal, or H_0: $[\overline{R}_1=\overline{R}_2]$ (Sheskin, 1997). The parallel to the t-test is most evident in this form.

The test can be one-sided or two-sided. The two-sided alternative hypothesis for shift is H_1: $[\Delta \neq 0]$ (Hollander & Wolfe, 1999), H_1: $[F(x) \neq G(x)$ for some $x]$ (Conover, 1999), or H_1: $[P[X>Y] \neq \frac{1}{2}]$ (Siegel & Castellan, 1988).

The alternative hypothesis for medians is H_1: [$\phi_1 \neq \phi_2$] (Neave & Worthington, 1988). The alternative hypotheses for ranks are H_1: [$\sum R_1 \neq \sum R_2$], or H_1: [$\bar{R}_1 \neq \bar{R}_2$] (Sheskin, 1997).

For a one-sided test, the alternative hypotheses for shift are either H_1: [$\Delta < 0$], H_1: [$\Delta > 0$] (Hollander & Wolfe, 1999) or H_1: [$P[X > Y] > \frac{1}{2}$], H_1: [$P[X > Y] < \frac{1}{2}$] (Siegel & Castellan, 1988). The alternative hypotheses for medians are H_1: [$\phi_1 < \phi_2$] or H_1: [$\phi_1 > \phi_2$] (Neave & Worthington, 1988). The alternative hypotheses for ranks are H_1: [$\sum R_1 < \sum R_2$], H_1: [$\sum R_1 > \sum R_2$], H_1: [$\bar{R}_1 < \bar{R}_2$], or H_1: [$\bar{R}_1 > \bar{R}_2$] (Sheskin, 1997). Conover (1999) also gave alternative hypotheses in terms of expected values and probabilities.

Procedure and Test Statistic

Siegel & Castellan (1988) and Neave & Worthington (1988) described the Wilcoxon version of the test. Given two samples, A and B, with $N = n_A + n_B$, combine the observations in a single array, keeping track of original sample membership, and then rank them from 1 to N. Algebraic size is considered in ranking, with the lowest ranks assigned to negative values (if any) with the largest magnitudes. Compute R_A as the sum of the ranks of the observations from sample ‘A’ and R_B as the sum of the ranks of the observations from sample ‘B’. The test statistic, W, is the rank sum that would be expected to be smaller if H_1 were true.

Tables of critical values are usually given for the Mann-Whitney U test. Since they are mathematically equivalent, the results of the Wilcoxon procedure can be converted to values of U. Neave & Worthington (1988) gave the conversion for a two-sided test as: $U = \min[U_A,\ U_B]$, with $U_A = R_A - \frac{1}{2}n_A(n_A + 1)$ and $U_B = n_A n_B - U_A = R_B - \frac{1}{2}n_B(n_B + 1)$. For a one-sided test, use either U_A or U_B according to which one is expected to have the smaller value under to H_1. Converting to values of U also accounts for the effect of unequal sample sizes.

Null Distribution (absent ties)

The assumptions of the test imply that when H_0 is true, all possible sequences of observations from samples A and B are equally likely (Neave & Worthington, 1988). In any two-sample distribution-free test statistic the number of possible sequences is: $\binom{N}{n_A} = \binom{N}{n_B} = \frac{N!}{n_A!n_B!}$. To get the exact null distribution, construct the statistic for each of these ways, construct the frequency distribution of the resulting values, and divide the frequencies by the number of possible sequences (Hollander & Wolfe, 1999).

The exact null distribution is generally used up to group size 10 (Bergmann, Ludbrook, & Spooren, 2000) and depends on the number and pattern of tied values within and between groups (Lehmann, 1998). The distribution of W is symmetric about its mean value (Hollander & Wolfe, 1999) and quickly approaches the normal with increasing sample size (Neave & Worthington, 1988). A normal approximation formula is available for large samples sizes.

Hollander & Wolfe (1999) considered the exact conditional (permutation) distribution of *W* in the presence of ties when ties were resolved using mid-ranks. The program StatXact (Mehta & Patel, 1999) can generate this distribution but it is not considered further here.

Critical Values, Rejection Region and Sample Size

Critical regions are of the form $U_{min} \leq$ *critical value* and represent best conservative values (Neave & Worthington, 1988; Sheskin, 1997). The test can be applied to unequal sample sizes with appropriate critical value tables. The null distribution is symmetric and rapidly approaches normality with increasing sample size (Neave & Worthington, 1988). The method described above for generating the null distribution requires a computer for other than small sample sizes. Indeed, methods other than direct calculation may be needed as sample size increases.

Neave (1981) included critical values for all sample size combinations up to 25 per group and critical values for equal sample sizes of 26 to 50. Siegel & Castellan (1988) only provided critical value tables for all sample size combinations up to 10.

Many authors (see for example Siegel & Castellan, 1988; Hollander & Wolfe, 1999; Neave & Worthington, 1988; Spent & Smeeton, 2001), have described a large sample approximation formula, including corrections for lack of continuity and the presence of ties, but differ greatly in their recommendations as to when it is appropriate to use. Fahoome (1999) studied the issue of adequate sample sizes for use of large-sample approximation formulas and the use of such formulas is not considered further here. The smallest sample that can be tested is three per group, one-sided, at α = .05.

Consequences of Ties and Recommended Methods of Resolution

After the combined samples have been ranked, tied ranks that come from the same group are of no consequence and are resolved by arbitrarily assigning the set of tied-for ranks in sequence. Ties between groups, however, render the situation ambiguous and the test becomes approximate.

Neave & Worthington (1988) recommended resolving consequential ties in all possible ways and calculating $\overline{U}$ as the average of all of the resulting '*U*'s. This method is based on an equiprobability assumption that regards all resolutions of the tied values as equally likely (Bradley, 1968). In the case of the Wilcoxon procedure the same result can be accomplished by assigning mid-ranks. When using the Mann-Whitney procedure, the equivalent method is to assign a value of ½ to observations involved in consequential ties.

Conover (1999) described a modified form of the statistic that was approximately normally distributed and included a correction for ties. He recommended the use of this modified statistic "if there are many ties" (Conover, 1999, p. 273) without regard to sample size and without quantification of 'many'. He also devoted considerable text to formulas for calculating *p*-values.

Example

Consider samples A: {1, 2, 3, 4, 5} and B: {2, 4, 6, 8, 10} such that $n_A = n_B = 5$ with $N = n_A + n_B = 5 + 5 = 10$. To test H_1: $\phi_B > \phi_A$ arrange the combined observations in rank order, keeping track of the original group membership, and using mid-ranks to resolve consequential ties. Thus,

Value	1	2	2	3	4	4	5	6	8	10
Rank	1	2.5	2.5	4	5.5	5.5	7	8	9	10
Group	A	A	B	A	A	B	A	B	B	B

Two values are tied at '2' and two more at '4'. Both pairs of ties are consequential as they are between groups, thus mid-ranks have been used. Calculate the rank sum $R_A = 1 + 2.5 + 4 + 5.5 + 7 = 20$ as this is expected to be smaller under H_1. Compute the Mann-Whitney U equivalent of R_A as

$$U_A = R_A - \frac{1}{2} n_A (n_A + 1) = 20 - 5(5+1)/2 = 20 - 30/2 = 20 - 15 = 5.$$

The critical value at $\alpha = .05$ is 4. Since $U_A > 4$, the test fails to reject H_0.

Comments

Bergmann, Ludbrook, & Spooren (2000) compared the results of the Wilcoxon-Mann-Whitney Test for 11 PC-based statistics packages using a real data set. Results across packages were not consistent at the $\alpha = .05$ level, with probabilities ranging from 0.01473 to 0.08873. The only implementations that gave accurate results constructed the exact permutation null distribution from the actual data, although some of these failed to compensate for ties. The performance of implementations that used a large-sample approximation depended on whether corrections were used for continuity and/or ties. Most of the packages failed to document the manner in which the procedure was implemented and the authors recommended that most of them be avoided.

Fligner & Policello (1981) introduced a robust rank test for the 'Behrens-Fisher' problem. Zimmerman (1992) asserted that the Wilcoxon-Mann-Whitney Test is appropriate under violation of normality, but not violation of homogeneity of variance, consistent with the assumption of the shift-effect model investigated here. He asserted that the Fligner-Policello Test might be a more appropriate test when both assumptions were not considered tenable. Hollander & Wolfe (1999), described the Fligner-Policello Test as an extension of the Wilcoxon-Mann-Whitney Test based on the additional assumption that the two population distributions were symmetric about their respective medians, while relaxing the assumption that they have the same distributional form and, in particular, that they have the same variance. Because of the symmetry requirement, the Fligner-Policello Test is no longer strictly distribution-free. It is not considered further in this study.

Zimmerman (1995) advocated the use of robust nonparametric statistics to increase power. He applied a two-stage procedure to the *t*-test and Wilcoxon-Mann-Whitney Test that involved detection and down-weighting of outliers prior to testing.

2.4 — *k*–independent–samples Omnibus Tests — The Kruskal–Wallis Test

Background

This test was introduced in Kruskal (1952) and Kruskal & Wallis (1952). Vogt (1999) described it as "A nonparametric test of statistical significance used when testing more than two independent samples. It is an extension of the Mann-Whitney *U* test, and of the Wilcoxon [rank-sum test], to three or more independent samples. It is a nonparametric one-way ANOVA for rank order data..." (p. 151). Everitt (1998) described the test as a "...distribution free method that is the analogue of the analysis of variance of a one-way design. It tests whether the groups to be compared have the same population median..." (p. 180). The test is applied to ordinal (rank-ordered) data (Sheskin, 1997).

Power comparisons with the *F*-test are very favorable. Conover (1999) gave the following asymptotic relative efficiencies for the Kruskal-Wallis test relative to the *F*-test: 1) for distributions that differ only in their means, never less than 0.864 but as high as infinity, b) for normal populations, 0.955, c) for uniform distributions, 1.0, and d) for exponential distributions, 1.5.

Assumptions and Hypotheses

Random and independent sampling of continuous populations is assumed with sufficient precision of measurement to avoid tied observations (Bradley, 1968). Sample values are independent both with and between groups (Conover, 1999). The null hypothesis assumes further that the populations are identical. The alternative hypothesis assumes that the populations differ only in location, known as an additive or location-shift effect (Sprent & Smeeton, 2001). For k groups, the population distribution functions, $F_1, \ldots, F_k$ are assumed to have the relationship $F_j(x) = F(x - \tau_j), -\infty < j < \infty$ over all j (j = 1 to k) where F is a continuous distribution function with unknown median and τ_j is the unknown treatment effect for the jth population (Hollander & Wolfe, 1999).

Vargha & Delaney (1998) took exception to the use of the Kruskal-Wallis Test with the foregoing assumptions on the grounds that the attendant hypotheses, while mathematically correct, were too narrow to be of practical value to researchers. They claimed that the Kruskal-Wallis Test "cannot detect with consistently increasing power any alternative other than exceptions to stochastic homogeneity" (Vargha & Delaney, 1998, p.170). This, in turn, is mathematically equivalent to the "equality of expected values of the rank sample means" (Vargha & Delaney, 1998, p. 170). They argued that the requirement for identical distributions under H_0 is too strict, and that only variance homogeneity is needed. Further, they asserted that the H_1 to which the test is actually sensitive is "the tendency for observations in at least one of the populations to be larger (or smaller) than all the remaining populations together" (Vargha & Delaney, 1998,.p 186).

Hollander & Wolfe (1999) gave the null hypothesis as H_0: $[\tau_1 = \tau_2 = \cdots = \tau_k]$. Neave & Worthington (1988) and Siegal & Castellan (1988) stated the null hypothesis as no difference in the medians of the populations, or H_0: $[\phi_1 = \phi_2 = \ldots = \phi_n]$. Conover (1999) presented the null hypothesis as H_0: [All of the *k* population distribution functions are identical].

The test is two-sided with an omnibus alternative hypothesis for shift of H_1: $[\tau_1,\ldots,\tau_k \text{ not all equal}]$ (Hollander & Wolfe, 1999) or, for medians, of H_1: [not all of ϕ_1, ϕ_2, ..., ϕ_k are equal] (Neave & Worthington, 1988; Siegal & Castellan, 1988). Conover (1999) gave the alternative hypothesis as H_1: [At least one of the populations tends to yield larger observations than at least one of the other populations].

As given in Sheskin (1997), all of these hypotheses can be formulated in terms of rank-sums (for the equal sample size case) or mean ranks (for the general case) as follows: H_0: [$\sum R_1 = \sum R_2 = \ldots = \sum R_k$], H_0: [$\overline{R}_1 = \overline{R}_2 = \ldots = \overline{R}_k$], with the alternative hypothesis of H_1: [not H_0]. The alternative hypothesis is stated in this way because it only requires that some pair of groups be different, not that all groups are different, consistent with Conover (1999).

Procedure and Test Statistic

The general procedure, which does not assume equal sample sizes, is to combine the samples and rank the observations, keeping track of original group membership. For each of the k groups, let the number of observations be n_i $(i = 1, 2, \ldots, k)$ such that the total number of observations is $N = \sum_{i=1}^{k} n_i$. Calculate the rank-sum for each group as $s_i = \sum_{j=1}^{n_i} r_{ij}$, where r_{ij} is the rank assigned to the jth observation in the ith group. The sum of the mean squared ranks is calculated as $S_k = \sum_{i=1}^{k} \left(\frac{s_i^2}{n_i} \right)$. The statistic is then calculated as $H = \frac{12}{N(N+1)} S_k - 3(N+1)$. This is the common computational formulation, consistent with Sprent & Smeeton (2001), Neave & Worthington (1988), Feir-Walsh & Toothaker (1974), Siegal & Castellan (1988) and Conover (1999).

Conover (1999) defined the test statistic as $T = \frac{1}{S^2}\left(S_k - \frac{N(N+1)^2}{4}\right)$ where

S_k and N are as defined above and $S^2 = \frac{1}{N-1}\left(\sum_{\substack{all \\ ranks}} R(X_{ij})^2 - N\frac{(N+1)^2}{4}\right)$. He

noted that S^2 simplified to $N(N+1)/12$ in the absence of ties such that $T = H$ as defined above.

The definition of H was also given by Neave & Worthington (1988) and Siegal & Castellan (1988) as $H = \frac{12}{N(N+1)}\sum_{i=1}^{k} n_i\left(\overline{R}_i - \overline{R}\right)^2$, where n_i is as above, $\overline{R}_i$ is the mean rank of group i, and $\overline{R}$ is the overall mean rank of the N total observations. In this form it can be most clearly seen that the statistic is a weighted sum of squared deviations. Conover (1999), Sheskin (1997), and Siegal & Castellan (1988) discussed post-hoc procedures using pairwise comparisons, but they are not considered further here.

A normal approximation formula is available for large sample sizes (Hollander & Wolfe, 1999). Fahoome (1999) provided an extensive study of the issue of adequate (equal) sample sizes for the use of normal approximations, either in the absence of ties or with the use of mid-ranks to resolve them.

Null Distribution (absent ties)

Under the null hypothesis the k samples can all be regarded as coming from the same population. As such, all arrangements of the $N = n_1 + n_2 + \cdots + n_k$ observations are equally likely. Also, all of the arrangements of the N ranks into groups of sizes n_1, n_2, ..., n_k are equally likely, and there are $\frac{N!}{n_1!n_2!\cdots n_k!}$ such arrangements (Hollander & Wolfe, 1999; Berry & Mielke, 2000; Neave & Worthington, 1988). In theory, the null distribution is found by calculating the value of H for all of these arrangements, constructing the cumulative frequency distribution and then dividing by the total to get the required probabilities for constructing the cdf.

In practice, some other approach to generating the null distribution becomes necessary at around 1,000,000 permutations on a PC or 100,000,000 permutations on a workstation (Berry & Mielke, 2000). When the null hypothesis is true, and N is moderate to large, the test statistic has a chi-squared distribution with $k - 1$ degrees-of-freedom (Sprent & Smeeton, 2001). Siegal & Castellan (1988) and Conover (1999) recommended the use of the chi-squared approximation when $k > 4$ and $n_j > 5$ for all groups.

Hollander & Wolfe (1999) addressed the exact conditional (permutation) distribution of H using 'average' (mid) ranks to break ties. This procedure is available in StatXact (Mehta & Patel, 1999) but is not considered further here.

Critical Values, Rejection Region and Sample Size

Critical regions are of the form $H \geq \textit{critical value}$ (Neave & Worthington, 1988). Approximate critical values can be obtained from a chi-squared distribution with $k - 1$ degrees-of-freedom. The test will work with unequal sample sizes since the calculation of the statistic involves a weighted sum of squares of differences between group mean ranks and the overall mean rank, although critical value tables tend to be limited (Neave, 1981). The smallest testable sample is 3 per group at either $\alpha = .01$ or $\alpha = .05$.

Consequences of Ties and Recommended Methods of Resolution

The presence of ties violates the assumptions of the test and alters the null distribution of H. Neave & Worthington (1988) recommended the use of 'average' (mid) ranks to resolve ties. Siegal & Castellan (1988) recommended the use of mean ranks for resolving all ties, regardless of whether they were within a single group or occurred between groups.

For a 'small' (unspecified) number of ties, Neave & Worthington (1988) recommended that H be used without further correction. For a 'large' (also unspecified) number of ties they gave a correction factor as

$C = 1 - \frac{\sum(t^3 - t)}{N(N^2 - 1)} = 1 - \frac{\sum t^3 - \sum t}{N(N^2 - 1)}$ where t is the number of ties for a given tied value and the summation is taken over all sets of tied values. This is consistent with Sheskin (1997) and Siegal & Castellan (1988). The adjusted statistic is calculated as $H^* = \frac{H}{C}$, which is approximately chi-squared distributed with k–1 degrees-of-freedom.

Sprent & Smeeton (2001) gave a version of the statistic, corrected for ties using mid-ranks, based on the total sum of squared ranks, $S_r = \sum_{i=1}^{k}\sum_{j=1}^{n_i} r_{ij}^2$, and a correction factor, $C = \frac{N(N+1)^2}{4}$, as $T = \frac{(N-1)(S_k - C)}{S_r - C}$, with S_k as defined previously. Analogous to ANOVA, S_k and S_r are the "uncorrected **treatment** and

total sums of squares for ranks" (Sprent & Smeeton, 2001, p. 201) (emphasis in the original). They indicated that this corrected statistic continues to be approximately chi-squared distributed with k–1 degrees-of-freedom.

Example

Consider k = 3 groups with $N = n_A + n_B + n_C = 5 + 5 + 5 = 15$ observations: A: {2, 5, 8, 14, 17}, B: {6, 6, 9, 12, 15}, and C: {4, 7, 13, 16, 19}. To test the null hypothesis of no difference in the population medians versus the two-sided (omnibus) alternative that not all of the population medians are equal, arrange all of the observations in a single array and rank them, keeping track of the original group membership.

Scores	2	4	5	6	6	7	8	9	12	13	14	15	16	17	19
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Group	A	C	A	B	B	C	A	B	B	C	A	B	C	A	C

The two observations tied at value '6' are from the same group, thus they are not consequential for this test and have been arbitrarily assigned consecutive ranks.

From the ranked data calculate: $s_A = 1+3+7+11+14 = 36$, $s_B = 4+5+8+9+12 = 38$, and $s_C = 2+6+10+13+15 = 46$, such that

$$S_k = \frac{36^2}{5} + \frac{38^2}{5} + \frac{46^2}{5} = \frac{4856}{5} = 971.2.$$ This yields

$$H = \frac{12(971.2)}{15(15+1)} - 3(15+1) = \frac{11654.4}{240} - 3(16) = 48.56 - 48 = 0.56.$$ The critical value for this test at alpha = .05 is 5.780. Since $H < 5.780$ the test fails to reject the null hypothesis.

Comments

Berry & Mielke (2000) studied exact and resampling Monte Carlo approaches to both the Wilcoxon-Mann-Whitney Test and the Kruskal-Wallis Test. For the resampling procedures, *p* values from .001 to .10 required from 4,000 to 400,000 resamplings to obtain three places of accuracy while 40,000,000 to 4,000,000,000 resamplings were needed to obtain 5 places of accuracy.

Rust & Fligner (1984) described a modification of the Kruskal-Wallis statistic to deal with the Behrens-Fisher situation under less restrictive null and alternative hypotheses. The test remains exactly distribution-free under the original null hypothesis and is asymptotically distribution-free under the less restrictive hypothesis given that the requirement for symmetry is met (Hollander & Wolfe, 1999). Wilcox (1994) provided simulation results for the Rust-Fligner modification of the Kruskal-Wallis Test that showed very favorable performance relative to the *F*-test with population distributions that were near normal, but slightly heavy-tailed, nonnormal, and both.

Siegal & Castellan (1988) considered post-hoc multiple comparisons and comparisons of treatments versus a control. Hollander & Wolfe (1999) also discussed procedures for post-hoc multiple comparisons at length. These procedures are not considered in any further detail here.

Extensive consideration was given to the null and alternative hypotheses for which the Kruskal-Wallis test is consistent in Vargha & Delaney (1998), these being stochastic homogeneity and stochastic heterogeneity, respectively. They remarked that almost all of the prior literature on the Kruskal-Wallis Test, except for those that explicitly adopted a shift model, was inadequate and did not correctly represent the assumptions and hypotheses of the test. They also noted that the shift model, while mathematically defensible, was unrealistically restrictive.

2.5 — *k*–independent–samples Tests of Ordered Alternative Hypotheses — The Terpstra–Jonckheere Test

Background

The Terpstra-Jonckheere Test was developed independently by Terpstra (1952) and Jonckheere (1954). Like the Kruskal-Wallis Test, it is an extension of the Wilcoxon-Mann-Whitney Test on ranks for the one-way design. It differs from the Kruskal-Wallis Test in that it postulates a specific ordering of the groups under the alternative hypothesis based on prior knowledge (Hollander & Wolfe, 1999; Siegal & Castellan, 1988).

Assumptions and Hypotheses

Random and independent sampling of continuous populations is assumed with sufficient precision of measurement to avoid tied observations (Bradley, 1968). Independence of sample observations, both within and between groups, is also assumed (Hollander & Wolfe, 1999). The more general assumption is that all of the possible assignments of joint ranks are equally possible (Hollander & Wolfe, 1999). It is also assumed that the situation being tested supports an *a priori* expectation of a specific, identifiable order of the population medians. This ordering must be based on the experimental design, not on the observed data (Hollander & Wolfe, 1999; Siegal & Castellan, 1988).

The null hypothesis assumes that the populations are identical. The alternative hypothesis assumes that the populations differ only in location, known as an additive or location-shift effect. For k groups, the population distribution functions, $F_1,\ldots,F_k$ are assumed to have the relationship

$F_j(x) = F(x-\tau_j), -\infty < x < \infty$ over all j, (j = 1 to k), where F is a continuous distribution function with unknown median and τ_j is the unknown treatment effect for the jth population (Hollander & Wolfe, 1999).

Hollander & Wolfe (1999) gave the null hypothesis as H_0: $[\tau_1 = \tau_2 = \cdots = \tau_k]$ while Neave & Worthington (1988) and Siegal & Castellan (1988) stated it in terms of medians as H_0: $[\phi_1 = \phi_2 = \ldots = \phi_k]$. Sprent & Smeeton (2001) gave the null hypothesis of identical populations as H_0: $[F_1(x) = F_2(x) = \cdots = F_k(x), \forall x]$. If the k groups are numbered to correspond to the expected order, the alternative hypothesis is one-sided and given by H_1: $[\tau_1 \le \tau_2 \le \cdots \le \tau_k$, with at least one strict inequality] (Hollander & Wolfe, 1999), H_1: $[F_1(x) \le F_2(x) \le \cdots \le F_k(x)$, at least one inequality strict for some x] (Sprent & Smeeton, 2001), or H_1: $[\phi_1 \le \phi_2 \le \ldots \le \phi_k$, at least one of the inequalities is strict] (Neave & Worthington, 1988; Siegal & Castellan, 1988).

Procedure and Test Statistic

The procedure calculates the Mann-Whitney U statistic for all pairs of samples and then combines the results. If the Wilcoxon rank-sum procedure is used the resulting statistics must be converted to Mann-Whitney U statistics before being combined. For the alternative hypothesis, as stated above, the test statistic was given by Neave & Worthington (1988) as

$J = U_{21} + U_{31} + \ldots + U_{k1} + U_{32} + \ldots + U_{ij} + \ldots + U_{k(k-1)} = \sum_{j=1}^{k-1} \sum_{i=j+1}^{k} U_{ij}$, where U_{ij} represents the Mann-Whitney U statistic for each pair of samples, computed in the order dictated by H_1 to give the least value of each U_{ij}. This is consistent with Siegel & Castellan (1988) and others. To the extent that H_1 tends to be true each of the U_{ij} will tend to be small and thus their sum will tend to be small.

For k groups there will be $k(k-1)/2$ values of U. Hollander & Wolfe (1999) gave the Mann-Whitney procedure for calculating the values of U directly, including an adjustment for ties (equivalent to using mid-ranks in the Wilcoxon version of the procedure) as

$$U_{uv} = \sum_{i=1}^{n_u} \sum_{j=1}^{n_v} \phi^*\left(X_{iu}, X_{jv}\right), 1 \le u < v \le k, \text{ where } \phi^*(a,b) = \begin{cases} 1 \text{ if } a < b \\ \frac{1}{2} \text{ if } a = b \\ 0 \text{ if } a > b \end{cases}.$$

This is consistent with Siegal & Castellan (1988). The test is approximate when ties are present.

A normal approximation formula is available for large sample sizes (Hollander & Wolfe, 1999). Fahoome (1999) provided an extensive study of the issue of adequate (equal) sample sizes for the use of normal approximations either in the absence of ties or with the use of mid-ranks to resolve them. The use of such formulas is not considered further here.

Null Distribution (absent ties)

The null hypothesis implies that all samples come from identical, continuous distributions such that all possible sequences of observations are equally likely (Neave & Worthington, 1988). In theory, one obtains the null distribution by calculating *J* for all possible sequences for a given choice of *k* and n_i, constructing a frequency distribution and dividing by the total frequency (Hollander & Wolfe, 1999). This is not very practical, however, as for $N = n_1 + n_2 ... + n_k$ total observations in *k* groups, the total number of sequences is given as $\frac{N!}{n_1!n_2!...n_k!} = \frac{(\sum n_i)!}{\prod(n_i!)}$ which gets large very quickly as the number of groups and total sample size increase. Jonckheere (1954) gave a recursion formula that is the basis for the actual critical value tables (as cited in Neave & Worthington, 1988).

Critical Values, Rejection Region and Sample Size

Critical regions are of the form $J \leq \textit{critical value}$ (Neave & Worthington, 1988). The test supports unequal samples sizes and more extensive critical value tables are available as Table *R* in Neave & Worthington (1988). As the sample size increases the null distribution of *J* becomes asymptotically normal. Neave & Worthington (1988) and Siegal & Castellan (1988) gave formulas for obtaining approximate critical values.

Consequences of Ties and Recommended Methods of Resolution

The issue with regards to ties is identical to the Wilcoxon-Mann-Whitney Test and Kruskal-Wallis Test described previously and the recommendations made there apply here as well. In particular, when ties are present the test becomes approximate when referred to a standard table of critical values.

Example

Consider $k = 3$ groups with $N = n_A + n_B + n_C = 5 + 5 + 5 = 15$ observations: A: {2, 5, 8, 14, 17}, B: {6, 6, 9, 12, 15}, and C: {4, 7, 13, 16, 19}. To test the null hypothesis of no difference in the population medians versus the alternative hypothesis H$_1$: $\phi_A \leq \phi_B \leq \phi_C$ calculate the Wilcoxon rank-sum statistic for the three pairs of samples R_{BA}, R_{CA}, R_{CB}.

For R_{BA}, the data is

Scores	2	5	6	6	8	9	12	14	15	17
Ranks	1	2	3	4	5	6	7	8	9	10
Group	*A*	*A*	*B*	*B*	*A*	*B*	*B*	*A*	*B*	*A*

with $R_{BA} = 1 + 2 + 5 + 8 + 10 = 26$ and $U_{BA} = 26 - 5(5+1)/2 = 26 - 15 = 11$. The tie at '6' is within group B and so consecutive ranks were assigned. For R_{CA}, the data is

Scores	2	4	5	7	8	13	14	16	17	19
Ranks	1	2	3	4	5	6	7	8	9	10
Group	*A*	*C*	*A*	*C*	*A*	*C*	*A*	*C*	*A*	*C*

with $R_{CA} = 1 + 3 + 5 + 7 + 9 = 25$ and $U_{CA} = 25 - 5(5+1)/2 = 25 - 15 = 10$. Finally, for R_{CB}, the data is

Scores	4	6	6	7	9	12	13	15	16	17
Ranks	1	2	3	4	5	6	7	8	9	10
Group	*C*	*B*	*B*	*C*	*B*	*B*	*C*	*B*	*C*	*C*

with $R_{CB} = 2+3+5+6+8 = 24$ and $U_{CB} = 24 - 5(5+1)/2 = 24 - 15 = 9$. It follows that $W = U_{BA} + U_{CA} + U_{CB} = 11 + 10 + 9 = 30$. The critical value for this situation at α = .05 is 21. Since 30 > 21, the test fails to reject the null hypothesis.

Comments

Mielke & Berry (2000) described Fortran program RTERP for computing probabilities associated with the Terpstra-Jonckheere test. RTERP provided a probability for the value of the statistic based on a randomization routine, fully accounting for ties. They reported that for small samples the results were comparable to the exact test while for large samples the solutions were nearly exact, and that the results were uniformly better than the normal approximation. RTERP was intended to replace TJTEST (Berry & Mielke, 1997), which required that the data contain no tied values, provided an exact probability for a realized value of the statistic, and an approximate probability for a normal approximation value of the statistic.

May & Konkin (1970) proposed a modified version of the Jonckheere test. Their statistic was based on the rapid approach of the sampling distribution to the normal with increasing sample size, a *t* approximation from Jonckheere (1954), an adjustment for continuity, and the number of comparisons supporting the alternative hypothesis.

2.6 — Remarks

There are many other tests for differences among *k* independent samples (Siegal & Castellan, 1988). Some, like the median test, are known to be less powerful than tests included in this study. Others, such as a *k*-sample test of slippage (Mosteller, 1948; Mosteller & Tukey, 1950), the generalization of the Jonckheere test (Chacko, 1963; Puri, 1965), or the umbrella test (Mack & Wolfe, 1981) apply to alternative hypotheses other than a shift in location. Hollander (1963), for example, described a modification of a test proposed by Moses (1952) that was sensitive to a treatment that manifested itself as an extreme reaction in either direction. Arnold & Briley (1973) also described such a distribution-free test, which they claimed "was equivalent to a two-sample test for dispersions considered in various forms by Ansari and Bradley (1960), Siegal and Tukey (1960) and others" (p. 302). Keselman, Cribbie & Zumbo (1997) studied specialized tests for detecting location differences with skewed distributions. These specialized tests have not been reviewed here.

Monte Carlo studies form a subset of experimental mathematics (Zumbo & Zimmerman, 1993), which refers to the use of computer simulations to mimic the rules of a model via (pseudo-) random processes. They take their place between the theoretical models of mathematical statistics and the practical concerns of applied research. Their relevance lies in the attempt to bridge that gap and allow practitioners to "see statistics for what it is, a straightforward discipline designed to amplify the power of common sense in the discernment of order amid complexity" (Stevens, 1968, p. 854).

Stevens (1968) spoke eloquently to both the legitimate differences in these sometimes polar perspectives as well as the need for them to be cognizant of, and respectful towards, the concerns of the other in the name of advancing human knowledge of the real world. In describing the association between schematics (models) and empirics (data) Stevens (1968) wrote that:

> ...it is precisely by way of the proper and judicious joining of the schematic with the empirical that we achieve our beneficial and effective mappings of the universe—the schemapiric mappings known as science. The chronic danger lies in our failure to note the distinction between the map and the terrain, between the simulation and the simulated. (p. 850)

The failure to note and understand this distinction is precisely what led to, and continues to maintain, the Normal Mystique.

CHAPTER 3

METHOD

3.1 — Overview

The Type I error and power properties of six distribution-free inferential statistical tests were studied using Monte Carlo techniques implemented as Fortran computer programs. The main focus of this study was the behavior of these statistics in response to various methods of resolving tied ranks, especially when sampling from discrete, non-normal data distributions representative of the populations typically encountered in the social and behavioral sciences (Micceri, 1986).

This study was limited to *k*-independent-samples tests of location or general difference. Because of the method used to generate shifts in location, samples for power studies were guaranteed to represent populations that differed only in location. For samples derived from the discrete distributions, the method used to generate shifts in location also guaranteed that the shifts were by integral amounts. Thus it was expected that a large percentage of trials would contain at least some tied values.

A separate study was conducted to collect data on the occurrence of ties—the number of trials involving ties, the number of distinct data values involved in ties, and the number of tied observations at those data values—in order to characterize the nature and degree of 'tiedness' that occurs with these distributions across the sample sizes investigated. This is more fully discussed in section 3.7.

In the power studies the shift in location was always towards higher values such that the appropriate 1-sided test was typically in the upper-tail of the null distribution of the statistic. For 2-sided tests, a lower-tail rejection represented a Type III error (MacDonald, 1999) in which the test determined a significant difference in the 'wrong' direction. Type I errors, power and Type III errors were counted and reported, as appropriate, for each simulation run. Power, however, was determined by the appropriate rejection rate only.

3.2 — Data Distributions

The normal distribution was generated mathematically and used as a baseline for the performance of each test under conditions meeting all of its assumptions. Fortran variables of type Real (single-, double-, or quad-precision) were used with sufficient precision that most samples from the normal distribution did not contain tied values. The rare ties that did occur were resolved by whatever methods were under investigation.

The non-normal distributions due to Micceri (1986) were generated by table look up using Fortran subroutine modules based on Sawilowsky, Blair, & Micceri (1990) and Fahoome (1999). Of the eight distributions characterized in Micceri (1986), four were used in this study, following Fahoome (1999), specifically: Extreme Asymmetric (achievement) (EA); Extreme Bimodal (psychometric) (EB); Multi-modal Lumpy (achievement) (ML), and; Smooth Symmetric (SS). These data sets are included in Appendix B as the Micceri

(1986, 1989) distributions. Properties of the Micceri (1986, 1989) data sets, as reported in Sawilowsky & Blair (1992), are given in table 3.2-1. By comparison, the normal distribution has $\mu = \phi = 0$, and $\sigma = 1$.

Table 3.2-1
Properties of Selected Micceri (1986, 1989) Distributions

Name	μ	ϕ	σ	Skew	Kurtosis
EA	24.50	27.00	5.79	−1.33	4.11
EB	2.97	4.00	1.69	-0.08	1.30
ML	21.15	18.00	11.90	0.19	1.80
SS	13.19	13.00	4.91	0.01	2.66

Adapted from Sawilowsky & Blair (1992) p. 353 Table 1

3.3 — Monte Carlo Technique and Sampling

The essential method of this study was random sampling with replacement from theoretical and empirical data sets representing both continuous normal and discrete non-normal pseudo-population distributions. For Type I error studies, all samples were drawn from the same distribution and used without modification prior to resolving ties. Thus, any test producing a significant result indicated an incorrect rejection of the null hypothesis, or Type I error, whereas a failure to reject the null hypothesis was a correct decision. For those tests that could be treated as either 1-sided or 2-sided, 1-sided tests were made in both the lower and upper tails of the null distribution of the statistic at nominal α and ½ nominal α. Two-sided tests used the combined results for both tails at ½ nominal α. All tests were conducted at nominal alphas of .05 and .01.

For power studies, samples were again drawn from the same distribution. For two-sample tests, one of the samples was shifted towards higher values to simulate the effect of a difference in location, i.e., of having been drawn from an identical population with a higher median. A significant result in this case indicated a correct rejection of the null hypothesis whereas failure to reject the null hypothesis was a Type II error. In the case of tests on three or more groups, each group was shifted relative to the preceding one such that the effect size between any two adjacent samples was the same. For tests on six groups at nominal effect size multiplier 1.2, this resulted in very large differences in the medians of the populations of groups 1 and 6. For 2-sided tests, a rejection in the wrong tail represented a Type III error (MacDonald, 1999) in which the test determined a significant difference in the wrong direction.

Type I error behavior was studied first for each statistic because further investigation of the power of the test was only relevant if it maintained Type I error at something close to nominal α. Bradley (1978) suggested $.9\alpha < p < 1.1\alpha$ as a conservative range of allowed values for highly robust Type I error rates, where p is the actual probability of the resulting test statistic. These limits are nominal $\alpha \pm 10\%$, arrived at by solving $|p-\alpha| \le \frac{\alpha}{10}$. For $\alpha = .01$, this gives $.009 \le p \le .011$ or 1.0% ± 0.1%, and for $\alpha = .05$, $.045 \le p \le .055$ or 5% ± .5%. Bradley (1978) also suggested liberal bounds of $.5\alpha \le p \le 1.5\alpha$, arrived at by solving $|p-\alpha| \le \frac{\alpha}{2}$. Finally, he noted that allowing $|p-\alpha| \le \alpha$ would lead to $0 \le p \le 2\alpha$. Thus, a willingness to accept, as robust, actual p values up to twice nominal-alpha, implies a willingness to also accept any *p* value less than nominal alpha. This was not deemed to be a sufficiently strong criterion.

Bradley's (1978) robustness criteria presume that the tolerance around nominal alpha must be plus or minus the same percentage. Many distribution-free tests have statistics with discrete distributions that do not have critical values corresponding exactly to standard nominal alphas of .005, .01, .025, and .05, especially at small sample sizes. The standard practice in such cases is to use best conservative critical values, i.e., the value of the statistic with the largest possible *p* value less than or equal to nominal α. In some cases, however, this meant that the best conservative critical value for α = .05 might, in fact, have a size of .0219 even though the next available higher value of the statistic had a size of .0536. This practice intentionally errors in a conservative direction and accepts a resultant loss of power for the test.

In general, combinations of simulation parameters for each test were eliminated from further study if they demonstrated Type I error rates outside of $.5\alpha \le p \le 1.1\alpha$, especially on the higher side, when that could be determined. As long as the results were not too conservative further study seemed warranted. For α = .01 this was $.005 \le p \le .011$ while for α = .05 it was $.025 \le p \le .055$. These bounds were in the spirit of Bradley (1978), in that they were precise and explicit *a priori* statements, even though they did not correspond exactly to any of his recommendations.

The discrete nature of the distributions of the statistics for these tests meant that best-conservative critical values sometimes resulted in Type I error performance that fell outside the acceptable range on the low end, even with the Normal distribution (no ties). In other cases, the associated probabilities were not available for the critical values, so a comparison could not be made on that basis. Thus, the Type I error performance against the Normal distribution was also used as a comparative benchmark for acceptable Type I error performance with the Micceri (1986, 1989) distributions.

Power was studied for each combination of simulation parameters that demonstrated acceptable Type I error properties as just described. For each combination of test, number of groups, distribution and initial sample size, a series of progressively larger effect sizes was introduced by a shift in location. The shift in location was accomplished by adding an effect size determined as a percentage of the standard deviation of the distribution. The nominal effect sizes studied were of the form $c\sigma$, where c, the nominal effect size multiplier, equaled 0.2 (small), 0.5 (medium), 0.8 (large), and 1.2, (very large) and σ represented the standard deviation of the distribution (Cohen, 1988; Sawilowsky & Blair, 1992). The shift was accomplished by adding the quantity $c\sigma$ to each observation in the sample to be shifted. Significance tests were then performed at both nominal alpha .01 and .05, both 1-sided and 2-sided if appropriate.

In addition to being discrete, the Micceri (1986, 1989) distributions are also based on integer values. In order to generate integral shifts, the nominal effect size, $c\sigma$, was rounded to the nearest integer. The resulting actual effect size, RND($c\sigma$), was reported along with the actual effect size multiplier, c'., where

$c' = [RND(c\sigma)]/\sigma$. The actual shift and actual effect size for each nominal effect size are listed in Table 3.3-1 for each distribution.

Table 3.3-1
Actual Shifts and Effect Sizes for Nominal Effect Sizes

Name	σ	S (.2σ)	M (.5σ)	L (.8σ)	VL (1.2σ)
		Nominal Shift *Actual Shift* *Actual Effect Size (rounded)*			
EA-A	5.79	1.158 1. 0.173σ	2.895 3. 0.518σ	4.632 5. 0.864σ	6.948 7. 1.209σ
EB-P	1.69	0.338 n/a n/a	0.845 1. 0.592σ	1.352 n/a n/a	2.028 2. 1.183σ
ML-A	11.90	2.380 2. 0.168σ	5.950 6. 0.504σ	9.520 10. 0.840σ	14.280 14. 1.176σ
SS	4.91	0.982 1. 0.204σ	2.455 2. 0.407σ	3.982 4. 0.815σ	5.892 6. 1.222σ
Normal	1.00	0.200 0.200 0.200σ	0.500 0.500 0.500σ	0.800 0.800 0.800σ	1.200 1.200 1.200σ

Based on Cohen (1988) and Sawilowsky & Blair (1992)

In the case of tests on three or more groups, each group was shifted relative to the preceding one by RND($c\sigma$) such that the effect size between any two adjacent groups was the same, i.e., group one received no shift, group two was shifted by RND($c\sigma$), group three was shifted by 2xRND($c\sigma$), and so on until all groups had been treated. Generating integral shifts ensured that the scores in all samples were integers such that equal observations between samples were possible.

3.4 — Methods for Resolving Ties

Only certain types of ties are critical in terms of having a consequence on the calculated value of the test statistic and, therefore, possibly on the resulting inference. In tests of location, ties between groups are usually critical whereas ties within groups are not. Exactly what constitutes a critical tie, however, had to be determined for each test.

Non-consequential ties were resolved in a manner that did not distort the test statistic by arbitrarily assigning the set of sequential ranks for which the group of scores was tied, just as if the original values had differed slightly. This did not alter the pattern of A's, B's, C's, etc. in a combined, ranked sample, and thus could not alter the value of the statistic or resulting inference.

Resolving consequential ties is always somewhat more speculative as the original data is potentially ambiguous in some sense. The efficacy of different methods for resolving consequential ties was the main focus of this research. The following methods for resolving consequential ties were investigated:

1) Resolve all ties in the manner least favorable to rejection of H_0, with resulting statistic S_{Least}.

2) Resolve all ties in the manner most favorable to the rejection of H_0, with resulting statistic S_{Most},

3a) Calculate and test the mid-range value $S_{Mid} = (S_{Least} + S_{Most})/2$.

3b) Count ties as ½ (Rosenbaum's Test and Tukey's Quick Test only).

4) Resolve ties through alternating or rotating assignment to groups, calculate $S_{Alternating}$, and test.

5) Randomly resolve the ties, calculate S_{Random}, and test.

6a) Replace tied ranks with the mid-rank, calculate S_{Tied}, and test.

6b) Delayed increment (Kolmogorov-Smirnov Test only).

6c) Calculate the weighted average of the statistics that result from all possible resolutions of the consequential ties (Rosenbaum's Test only).

7) Drop matching ties, reduce *n*, calculate S_{Match}, and test.

8) Drop all tied observations, reduce *n*, calculate S_{Drop}, and test (may result in unequal sample sizes).

Although 11 methods are listed, methods 1 and 2 were only implemented to calculate method 3a. Discussion of each of these procedures and some examples follow.

Method 3a was used in lieu of resolving ties in all possible ways because, under the assumption of equiprobability, the number of statistics calculated is always even and precisely symmetric with the consequence that the mid-range, mean and median values are always the same. The mid-range value required only two values of the statistic and was, therefore, the most computationally efficient. All three values suffer equally from the fact that they may not be integral or match one of the actually obtained values of the statistic. Thus, none of them was preferred on that basis. They also lose validity if the equiprobability assumption does not hold. This method was investigated with all tests.

Method 3b was investigated with Rosenbaum's Test and Tukey's Quick Test only. These tests are based on runs rather than ranks and normally assign a value of '1' to observations that contribute unambiguously to the value of the test statistic, '0' otherwise. As such, they admit assignment of some value other than '1' to observations involved in consequential ties.

Method 4 was investigated with all tests. It began by drawing a random number to decide which group got the first available rank. In the case of two groups, a value less than .5000 assigned the first available rank to group A, otherwise to group B. In general, the assignment interval was $1/k$ where k was the number of groups with values involved in the tie. Whichever group got the first available rank, the assignment of ranks then alternated (rotated) between groups until all members of one group had been assigned. The remaining ranks, if any, went to the other group (which must have had more data points involved in the tie). One of the negative consequences of this method is that groups with larger numbers of observations involved in a tie typically ended up with the largest ranks in the tied-for set of ranks, perhaps resulting in inflated Type I error rates.

As an illustration of the last point, assume the situation where observations A_1, A_2, B are tied for ranks 6, 7, and 8 in a 2-independent-samples test. Now suppose the initial random number is 0.6715. Since $0.6715 > 0.5000$, the first available rank (6) would go to the first available observation from group B. Since there is only one observation from group B, ranks 7 and 8 would be assigned to observations A_1 and A_2 and the procedure would terminate.

For three or more groups the procedure was similar. When resolving ties across three groups, if the initial random number was between 0.0000 and 0.3333 the first available rank went to the first available observation from group A, greater than 0.3333 up to 0.6666, to group B and greater than 0.6666, to group C. After that, the available ranks went in order to the available observations. If the first available rank went to group C, the remaining ranks went in turn to the next available observations from groups A, B, C, A, B, C, etc. If the observations from one group were exhausted prior to the others, then an alternating assignment continued with the two remaining groups until only observations from one group remained, at which time that group received the remaining ranks.

As an example, suppose observations from three groups are tied for ranks 1 to 9 as follows: A_1, A_2, A_3, B_1, B_2, B_3, C_1, C_2, C_3. Suppose also that the initial random number is .4999. Since $.3333 < .4999 < .6666$ the first available rank (1) goes to the first available observation in the second group (B_1). Going in sequence, the observations are now ordered as B_1, C_1, A_1, B_2, C_2, A_2, B_3, C_3, A_3. Thus, the three data points from group A now have ranks 3, 6, 9, the three data points from group B have ranks 1, 4, 7, and the three data points from group C have ranks 2, 5, and 8.

As before, when the number of observations involved in a tie are not the same for each group, the group with the largest number of observations tends to get the highest rank(s). Other schemes are possible, such as an up and down pattern (A, B, C, C, B, A, etc.), but were not investigated in this study. Method 4 corresponds to method E described in Chapter 2.

Method 5 was also investigated with all tests. For each data point at which a tie occurred (for each set of tied-for ranks) the procedure drew *n* random numbers, where *n* was the number of data points involved in the tie. The *n* random numbers were then rank ordered from 1 to *n*. The set of tied-for ranks was then assigned to the equal observations in this randomly determined order.

Suppose that in a 2-independent-samples test two values from group A and one value from group B are tied for ranks 6, 7, and 8. When combining the values from both samples into a single array these values are placed into the array in group order, A's first then B's, such that the data points are in the order A_1, A_2, B. Thus, whenever there are ties between the two groups the values from group A will precede those from group B in the combined sample array. Now suppose the drawing of three random numbers produces values of 0.6025, 0.2136, and 0.5947. The rank order of these three random numbers is 3, 1, 2 and these are used to reorder the three tied values as A_2, B, A_1. Thus the two scores from group A receive ranks 6 and 8 while the score from group B receives rank 7. The method was simple to implement for two groups, and was easily extended to three or more groups. Method 5 corresponds to method C from Chapter 2.

Method 6a was the simplest to implement when it applied, which included the Wilcoxon-Mann-Whitney Test, the Kruskal-Wallis Test and the Terpstra-Jonckheere Test. Suppose that two data points from group A are tied with a data point from group B for ranks 6, 7, and 8. In the combined and ranked sample they are initially arranged as A_1, A_2, B and all three data points are assigned the mid-rank value of 7. Method 6a corresponds to method F from Chapter 2.

Method 6b was unique to the Kolmogorov-Smirnov Test as described in Chapter 2, section 2.2.1. Method 6c was implemented only for Rosenbaum's Test because it was practical to do so given that the test only looks at one end of the combined sample and some mathematical simplifications were possible. None of the methods {6a, 6b, 6c} was investigated with Tukey's Quick Test as methods 6a and 6b did not apply and method 6c proved to be too difficult to implement.

Method 7 attempted to address the problem of possible unequal sample sizes when dropping equal data values. This was accomplished by only dropping tied observations if/when they occurred in all groups. Further, only the least number of observations was dropped, i.e., if three observations in group A were tied with one observation in group B and two observations in group C, then one of the observations would be dropped from each group, since the least number of instances in any group was the one that occurred in group B. This had the obvious drawback of leaving consequential ties unresolved and thus distorting the obtained value of the statistic. Method 7 does not correspond with any of the methods described in Chapter 2.

With method 8 it was not possible to compute the statistic when too many scores were dropped or if the size of the samples became too unequal due to insufficient critical values. Thus, the number of trials in which a test could not be completed was tabulated and reported. Method 8 corresponds to method B from Chapter 2.

Methods 7 and 8 were investigated with all tests, and were run in parallel with the other methods, but often resulted in only a few testable samples out of 1,000,000 repetitions. Although this indicated that the method could not be successfully applied much of the time, it did not yield much insight into the behavior of these tests when the samples were testable. Thus, an alternative simulation design was also used in which the program cycled until it obtained 10,000 testable samples or reached 10,000,000 repetitions. Since the number of repetitions was captured, this data was also useful in establishing the relative applicability of these methods while at the same time providing more useful data than the original simulation runs relative to Type I error rates. It is the results of these alternate simulations that are reported in Chapter 4 and used as the basis for conclusions in Chapter 5.

Referring back to Chapter 2, method 3a can be seen as a combination of methods A, D and G while not exactly any of them. Method A suggested arranging ties in the manner both least and most favorable to rejection of the null hypothesis, which method 3a adopts. Method D suggested an 'average' probability calculation, whereas method G suggested an 'average' statistic, both based on all possible resolutions of ties. Method D, however, also suggested a mid-range probability calculation, which requires only the least and most favorable resolution of ties, as in method A. Method 3a adopts the mid-range calculation based on values of the statistic rather than probabilities. Method 3a also involves a single test that avoids the ambiguity of 'no decision' that can occur in method A (even though that is appropriate) and avoids the computational burden of methods A, D, or G.

Methods A and D require actual probabilities and were not proposed for this research because it appeared during the initial research phase that such probabilities would be difficult to obtain. A procedure analogous to a combination of methods A and G would be to arrange all consequential ties in the manner both most and least favorable to rejection of the null hypothesis, calculate the statistic for each case, say S_M and S_L, and then test both statistics using the standard table of critical values. If both tests failed to support H_0, it would be rejected in favor of H_1; if both tests supported H_0, it would be retained. If one test supported H_0 and the other one rejected it, no decision would be appropriate in an actual test. Although the 'no decision' situation might be worthwhile to investigate, in a Monte Carlo study one always knows the truth about the populations involved. As such, the result of any procedure was always unambiguous, both in Type I error studies and in power studies. As a practical matter, 'no decision' amounts to a failure to reject the null hypothesis. Viewed that way, the test really becomes method 1 (method H), a test of S_{Least} that is unnecessarily conservative, and was not, therefore, investigated in this study.

3.5 — Ranking Analysis of Power Results

In order to answer Research Questions 3 and 4 it was necessary to analyze the power results from a large number of simulation runs in a manner that might permit determination of the order of preference of methods across various combinations of simulation parameters for each test. Examples of the calculations performed are given in Appendix C.

The method of analysis employed for this purpose involved ranking the power results across methods for each specific combination of test, number of groups, nominal alpha level and distribution at each combination of nominal effect size multiplier and initial sample size. Mean ranks were then calculated in three ways: 1) by summing across nominal effect size multipliers at each initial sample size, 2) by summing across initial sample sizes at each nominal effect size multiplier, and 3) by summing across both nominal effect size multipliers and initial sample sizes. These mean ranks are reported in Chapter 4, section 3 after the power results for each test and were used to answer Research Question 3.

Research Question 4 requires a conclusion to be drawn about the relative behavior of the methods across distributions. The problems with this are discussed in Chapter 5. These problems notwithstanding, the results of the preceding analysis were used to determine the number of first place finishes for each test for each combination of method and distribution across nominal alpha and number of groups. If a particular method consistently had the most first place finishes for a particular test, across distributions, then it could in some sense be considered the 'best' method for that test/distribution combination. Examples of these calculations are also given in Appendix C. These are the last results reported for each test in Chapter 4, section 3.

3.6 — Critical Values and Theoretical Probabilities

Modules were developed for generating critical values and their associated probabilities for the Kolmogorov-Smirnov Test, Rosenbaum's Test, Tukey's Quick Test, and the Wilcoxon-Mann-Whitney Test. This work is described in Fay (2002) and discussed briefly in Chapters 2 and 5. For this study, best conservative critical values were used, but the ability to generate critical values and their associated probabilities opens up additional possibilities as discussed in Chapter 5.

For the Kruskal-Wallis Test and the Terpstra-Jonckheere Test, modules were created using lookup tables and populated with critical values from published sources. When available, Type I error probabilities were also included in the lookup tables. Critical values and probabilities for unequal *n* were used when available, although these tables were incomplete with respect to the sample sizes investigated. This was, undoubtedly, a limiting factor in the investigation of the drop and reduce N methods with these two tests.

3.7 — Pattern and Occurrence of Ties

A separate study was performed to determine the nature and extent of ties in samples drawn from the target populations of the main study. Results of this study are presented in Chapter 4, section 4 and discussed in Chapter 5, section 2. The methodology was to combine the samples for each group on any given trial and sort them using the same modules as the main simulations. The sorting modules kept track of group membership and maintained rank and between-group-tie status in associated arrays. The combined sample was then examined for between-group ties. Counters were maintained for each group, and for all groups combined, of both repetitions and observations by first tie, second tie, etc.

Tables 3.7-1 and 3.7-2 show the type of data that were generated. Table 3.7-1 presents data in terms of repetitions while table 3.7-2 presents data in terms of observations (data points). The structure of the tables is the same, with a table being generated for each group and for the total combined sample, both for the repetitions and the observations. Thus, for a two-group test there would be six tables: 1) group 1 repetitions, 2) group 2 repetitions, 3) all groups repetitions, 4) group 1 observations, 5) group 2 observations, and 6) all groups observations.

In all cases there are INT(N/2) rows, where N is the total (combined) sample size being tested (per-group sample size times number of groups). This is the maximum number of between-group ties that can theoretically occur, in which every observation is tied with exactly one other observation from a different group. The rows, in turn, represent the results for the first between-group tie, the second between-group tie, and so on up to the INT(N/2) between-group tie, accumulated across all repetitions.

The first column designates the group. The entries in the second column are the between-group tie occurrence, i.e., 1=1st, 2=2nd, 3=3rd, etc. The third column contains the number of repetitions (or observations) involved in the tie for that row. The forth column is the ratio of the entry in the third column (same row) to the total number of repetitions (observations). The fifth column is the reverse accumulation of the third column, while the sixth column is the ratio of the corresponding entry in the fifth column to total repetitions (observations).

Table 3.7-1
Example Repetitions Data – Two Groups, Six per group – NESM = 0.0 (no shift)
1,000,000 repetitions; 12,000,000 observations

Group	BGT	----- per BGT ----- count	ratio	----- cumulative ----- count	ratio
1	1	858,432	.8584	1,480,377	1.4804
	2	472,094	.4721	621,945	.6219
	3	132,841	.1328	149,851	.1499
	4	16,249	.0162	17,010	.0170
	5	752	.0008	761	.0008
	6	9	.0000	9	.0000
2	1	858,432	.8584	1,480,377	1.4804
	2	472,094	.4721	621,945	.6219
	3	132,841	.1328	149,851	.1499
	4	16,249	.0162	17,010	.0170
	5	752	.0008	761	.0008
	6	9	.0000	9	.0000
all	1	858,432	.8584	1,480,377	1.4804
	2	472,094	.4721	621,945	.6219
	3	132,841	.1328	149,851	.1499
	4	16,249	.0162	17,010	.0170
	5	752	.0008	761	.0008
	6	9	.0000	9	.0000

For Table 3.7-1 (repetitions) the numbers in the fifth and sixth columns do not have an obvious direct interpretation, as the reverse accumulated counts in column five can, and sometimes do, exceed the total number of repetitions. As a result, the corresponding ratios in column six can, and sometimes do, exceed 1.0. The ratio in the first row of column six, however, does give a kind of relative index of the occurrence of ties that can be compared across distributions, number of groups and per-group sample sizes. The larger this number, the more often ties are occurring in the data.

Table 3.7-2
Example Observations Data – Two Groups, Six per group
NESM = 0.0 (no shift) – 1,000,000 repetitions; 12,000,000 observations

Group	BGT	----- per BGT ----- count	ratio	----- cumulative ----- count	ratio
1	1	1,020,228	.0850	1,735,789	.1446
	2	549,569	.0458	715,561	.0596
	3	147,877	.0123	165,992	.0138
	4	17,334	.0014	18,115	.0015
	5	772	.0001	781	.0001
	6	9	.0000	9	.0000
2	1	1,020,051	.0850	1,735,177	.1446
	2	549,300	.0458	715,126	.0596
	3	147,772	.0123	165,826	.0138
	4	17,316	.0014	18,104	.0015
	5	779	.0001	788	.0001
	6	9	.0000	9	.0000
all	1	2,040,279	.1700	3,470,966	.2892
	2	1,098,869	.0916	1,430,687	.1192
	3	395,599	.0246	331,818	.0277
	4	34,650	.0029	36,219	.0030
	5	1,551	.0001	1,569	.0001
	6	18	.0000	18	.0000

For Table 3.7-2 (observations) the entry in the first row, fifth column has a direct meaning; it is the total number of observations involved in a between-group tie during that test run. Since an observation involved in a first between-group tie cannot also be involved in a subsequent (2nd, 3rd, 4th, etc.) tie, this result cannot exceed the total number of observations. The corresponding ratio in column six cannot, therefore, exceed one. Thus, this ratio gives a very direct indication of how often ties are occurring in terms of the proportion of total observations involved in between-group ties.

Although a first tie must occur before there can be a second tie, and a second tie before a third, and so on by definition, the determination of the first tie, second tie, etc. is done within each combined sample as it is drawn. If the first tie in a particular sample involves observations from groups 1 and 2, but not group 3, then the repetition counters for a first tie in groups 1, 2 and all groups get incremented, but not the counter for group 3. If a second between-group tie is then found involving observations from groups 2 and 3, the corresponding second tie counters get incremented. Thus, it is possible for a group to have been involved in more second ties than first ties, and this was, in fact, observed in the example data for each group. This cannot occur, however, for the all groups (total sample) counter.

Table 3.7-3 shows examples of repetitions and observations results for three groups of three observations each and six groups of six observations each, both at NESM 0.0 (no shift) and 1.2 (maximum shift). Data were generated for each group as well as for the total of all groups to see how ties were distributed across groups. *A priori*, it was expected that for two groups (under all conditions), and for three or more groups (with no shift), between-group ties would be equally distributed across the groups. For three or more groups, as an increasing amount of shift was introduced, one would expect the extreme groups (not shifted on the left, most shifted on the right) to have fewer observations involved in between-group ties than those groups in the middle.

Table 3.7-3
Example of comparison of ties across three or more groups
Repetitions and observations ratio for all groups
Based on 1,000,000 repetitions

Obs/grp	3				6			
NESM	0.0		1.2		0.0		1.2	
Reps | Obs Group	RC	OC	RC	OC	RC	OC	RC	OC
1	.8116	.0963	.4088	.0480	3.9955	.1292	1.3664	.0434
2	.8130	.0964	.6160	.0725	3.9959	.1292	2.2915	.0729
3	.8123	.0963	.4096	.0481	3.9964	.1292	2.5407	.0806
4					3.9961	.1292	2.5397	.0805
5					3.9943	.1291	2.2919	.0729
6					3.9956	.1292	1.3663	.0434
all	1.1779	.2890	.7068	.1686	9.1897	.7750	5.8883	.3937

RC = cumulative repetitions ratio; OC = cumulative observations ratio

3.8 — Sampling Adequacy

Prior to developing and running the main simulations, a study was also made of the adequacy of samples drawn from the Micceri (1986, 1989) distributions. The purpose of this ancillary study was to determine the smallest number of randomly drawn observations, using the MOTHER (Blair, undated) random number generator, that would produce a sample distribution that was acceptably close to the population distribution from which it came. Results are reported in Chapter 4, section 5 and discussed in Chapter 5, section 3. The methodology employed to analyze the results was as follows.

The Micceri (1986, 1989) distributions are empirically derived discrete pseudo-population distributions defined by a set of values with associated frequencies. These values and frequencies, along with cumulative frequencies and cumulative distribution functions are given in Appendix B for each distribution.

In order to sample from these distributions, they were implemented as source vectors using Fortran 90 arrays. The source vector length was the sum of the frequencies, which is the cumulative frequency at the highest value. Table B–1, for instance, shows that for the Micceri (1986) Smooth Symmetric distribution the first three values {0, 1, 2} have frequencies of {11, 21, 27} while the last value, 26, has frequency 17 and cumulative frequency 5375. Thus the source vector length is 5375. The first 11 cells in this vector contain the value 0, the next 21 the value 1 and the next 27 the value 2. The last 17 cells contain the value 26.

In order to sample from these distributions, uniform random numbers were generated on the interval (0,1), rescaled to the source vector length, converted to integers and then used as an index into the source vector. A counter vector was also created for each distribution with a length equal to the number of values in that distribution. At any given combination of simulation parameters, samples were drawn and the appropriate counters incremented. At the end of 1,000,000 trials the counts were converted to proportions of the total count (1,000,000 trials x combined sample size).

The sample proportion for each value was compared to its corresponding population proportion by calculating $\%error = (SV_i - PV_i)/PV_i$ over all values of i. This was repeated for increasing numbers of total observations with the objective of finding the number that would guarantee that the largest such error was less than or equal to 1.0%. The sample cumulative distribution function was also constructed and a similar comparison made to its population cumulative distribution function.

In addition to the percent error calculations, the Kolmogorov-Smirnov Test of General Differences for Two Independent Samples was applied to each pair of cumulative distributions (sample and population) as a 2-sided test at nominal alpha .10. A 2-sided test was used as the differences could be in either direction. Nominal alpha .10 was used to ensure a powerful test as this was more important than protecting against Type I errors.

3.9 — Computer Hardware / Software

PC compatible computer hardware was used to run all of the programs for this study. Two primary computers were used, one with an AMD Athlon processor (500 MHz, 256 Mb ram) and the other a Dell Precision Workstation 530MT with an Intel Pentium IV Xeon processor (1.8 GHz, 1 Gb ram). Both systems were running Microsoft Windows 2000 Professional with Service Pack 1 or higher. Even on the Dell workstation, some of the simulations took more than 26 hours to run.

COMPAQ Visual Fortran (CVF) Professional Version 6.5 (2000) was used to build and run the programs, including use of the IMSL (1997) libraries. Existing Fortran code, originally developed in Essential Lahey Fortran 90 (ELF90) for Fahoome (1999) was obtained from G. F. Fahoome and formed a starting point for this study. Some of this code was originally due to S. S. Sawilowsky and/or R. C. Blair, and was modified to ELF90 by G. F. Fahoome. G. F.

Fahoome was also the source of the Sawilowsky, Blair, & Micceri (1990) Fortran modules for the Micceri (1986) distributions, which served as the starting point for corresponding modules in this study.

The random number generator for this study was MOTHER (Blair, undated), which is part of version 2.1 of the BFRA module developed by R. Clifford Blair for his own personal use (undated) and obtained from G. F. Fahoome. The User's Notes indicated that:

> Version 2.X adds to 1.X the capability to call "mother" (i.e. the mother of all random number generators). Mother is reputed to have excellent properties including an extremely long period. Several included routines call mother for random numbers. Mother is actually "Luxury", a Subtract-and-borrow random number generator proposed by Marsaglia and Zaman, implemented by F. James with the name RCARRY in 1991. The code used here is excerpted (and slightly modified) from a Fortran 90 version written by Alan Miller. (p. 1)

Deng & Lin (2000) suggested that classical uniform random number generators based on linear congruential generator (LCG) techniques suffered from major defects including relatively short period length and lack of uniformity at higher dimensions. They favored the use of a new class of pseudo-random number generators known as multiple recursive generators (MRG) and matrix congruential generators (MCG). They also provided the information to build a portable, efficient version of both, known as fast multiple recursive generator (FMRG) and fast matrix congruential generator (FMCG). MOTHER (Blair, undated) appears to be an MRG. Deng and Linn (2000) also gave a set of five criteria for random number generators as "(1) period length; (2) computing efficiency; (3) portability; (4) theoretical justification on randomness; and (5) empirical performance" (p. 145). MOTHER (Blair, undated) appears to meet all of these criteria, it's only limitation being availability.

Gentle (1998) provided a thorough treatment of the subject of random number generation and its relationship to Monte Carlo studies. He also discussed simulation studies as scientific experiments, and insisted that they must meet the same high standards as any other scientific experiment, specifically: a) control, b) reproducibility, c) efficiency, and d) careful and complete documentation. The uniform random number generator directly impacts control, reproducibility, and efficiency. Portable random number generators, in particular, are essential to reproducibility. MOTHER (Blair, undated) appears to meet these standards.

3.10 — Programming Methodology

All Fortran programming was modular so that a consistent and efficient-testing protocol could be maintained. A main program was developed for each of the six tests and tested using the normal distribution. These main programs controlled the overall sequence of steps and were constructed to accomplish as much work as possible in parallel. For instance, when a sample was drawn for a particular combination of test, distribution, number or groups and sample size, all applicable methods of resolving ties were implemented on that data and the results tabulated for both 1-sided and 2-sided tests at both α = .01 and α = .05 before drawing the next sample. The main programs employed loops for distribution, sample size, number of groups, and effect size as appropriate. All modules were tested to ensure that they performed correctly before full-scale simulations were run.

Subroutine modules were used for common tasks that occurred across tests. For example, a module was created for each data distribution for use in sampling. The normal distribution was generated by mathematical formula while the Micceri (1986) distributions were generated through cumulative frequency tables (see Appendix B). The MOTHER module (Blair, undated) was used to generate uniform pseudo-random numbers on the interval (0,1). A module was created for scaling the output of the random number generator to the appropriate range of integers needed to draw samples from the various distributions by indexing their tables. Pseudo-random numbers were generated using a consistent random number generator seed of 255,255 that was reset after each simulation run of 1,000,000 trials.

Modules were developed for each of the methods of resolving tied ranks and for calculating each test statistic as well as for housekeeping chores associated with capturing the results of each simulation run, storing them and creating printable ASCII output files. Other modules were created or adapted for sorting and ranking data in linear and 2-dimensional arrays. The sorting modules used stable sorting algorithms (Knuth, 1998b). These modules kept track of the original data value, the original group membership, the assigned rank and the between-group tie status. The specific data structure created to hold these attributes consisted of four parallel vectors as illustrated in Tables 3.10-1, 3.10-2 and 3.10-3.

Table 3.10-1
Data Structure for Sampled Data Before Sorting, Ranking or Determining Between-group Ties

Value	2.1	4.7	3.5	1.9	5.6	3.8	1.1	4.7
Group	1	1	1	1	2	2	2	2
Rank	0	0	0	0	0	0	0	0
BGT	0	0	0	0	0	0	0	0

Table 3.10-2
Data Structure for Sampled Data After Sorting, Initial Ranking and Determining Between-group Ties but Prior to Resolving Between-group Ties

Value	1.1	1.9	2.1	3.5	3.8	4.7	4.7	5.6
Group	2	1	1	1	2	1	2	2
Rank	1	2	3	4	5	6	7	8
BGT	0	0	0	0	0	1	1	0

Table 3.10-3
Data Structure for Sampled Data After Resolving Ties using Mid-ranks

Value	1.1	1.9	2.1	3.5	3.8	4.7	4.7	5.6
Group	2	1	1	1	2	1	2	2
Rank	1	2	3	4	5	6.5	6.5	8
BGT	0	0	0	0	0	1	1	0

Modules were also developed to generate the critical values and associated probabilities for the Kolmogorov-Smirnov Test, Rosenbaum's Test, Tukey's Quick Test and the Wilcoxon-Mann-Whitney Test and are fully described in Fay (2002). These modules generated lookup tables that were accessed according to the sample sizes for each group and returned the best conservative critical values and associated probabilities in two vectors for nominal alpha levels, either {.005, .01, .025, .05} or {.01 and .05}.

The critical value module for the Kruskal-Wallis Test returned two critical values, without probabilities, when called with an input vector of the sample sizes for three to six groups. Table entries were based on Neave (1981) and Iman, Quade and Alexander (1975). The critical value module for the Terpstra-Jonckheere Test also returned two critical values, with associated probabilities, when called with an input vector of the sample sizes for three to six groups. Table entries were based on Neave and Worthington (1988).

All modules used dynamic array allocation programming techniques to accommodate the range of simulation parameters. Thus, for a particular statistic and type of test, the same code was used to select and process the samples regardless of the number of groups, the initial sample size, or the effect size. Appropriate arithmetic types were used to ensure adequate precision of calculated results. In particular, integer arithmetic was used whenever possible, including double or quad precision if needed.

All programs were compiled and linked to run as DOS executables for speed. They produced a minimal amount of status output to the screen using characters in a DOS box, the main results being written to ASCII text files. The screen output was sufficient, however, to track the progress of the simulation, some of which took more than 26 hours to run on the Dell Pentium IV Xeon workstation.

For each combination of test, distribution, sample size, number of groups (when appropriate) and effect size (when appropriate), random samples were drawn from the same distribution, shifts in location introduced (for power studies), tied ranks resolved using the various methods (as appropriate to the test), and the statistic(s) calculated for each method. Both the 2-sided and 1-sided inferential tests of significance were performed for each method (when appropriate) using tabled or generated critical values at nominal α of .01 and .05 (1-sided) or .005 and .025 (2-sided), with the 1-sided test in the same direction as the effect (for power studies). This process was repeated for a total of 1,000,000 trials (repetitions) with ongoing tabulation of the results.

Type I error studies counted both lower- and upper-tail rejections for 2-sided tests and combined them to get the Type I error count. For 1-sided tests only rejections in the appropriate tail were counted. For power studies both upper- and lower-tail rejections were counted. Power, however, was determined only from the results for one tail based on the direction of the shift. These counts, when converted to proportions, gave a direct indication of the power of the test. For a 2-sided test, significant results in the other tail constituted Type III errors. For both the Type I error and power studies, the Type I error rate and power were calculated at the end of the 1,000,000 trials as a ratio (proportion) of significant results to total trials and written to an ASCII file for that combination of test, sample size, distribution, and method. As mentioned previously, special simulations were designed for the 'drop and reduce N' methods that cycled up to 10,000,000 times in an attempt to get 10,000 testable samples.

Additional references and resources related to programming methodology included COMPAQ (1999), Day (1972), Etzel & Dickinson (1999), Knuth (1997, 1998a, 1998b), Kreitzberg & Shneider (1972), Law (1983), and Organick (1966). Savage (1962) was used to locate and confirm older literature.

3.11 — Statistical Tests

Specific testing protocol information is provided for each test including number of groups, sample sizes, directionality, and alpha levels tested. The maximum number of simulation runs for Type I error and power studies was planned on the basis of five distributions, ten sample sizes, four effect sizes (for power studies only) and up to four different numbers of groups (3–6) for tests involving more than two samples. One simulation run generally consisted of 1,000,000 trials for each specific combination of type (Type I error or power), test, distribution, number of groups (if appropriate), sample size and effect size (power only). During a simulation run all applicable methods of resolving ties were investigated with each set of samples and both 1- and 2-sided tests were conducted, if appropriate, at all applicable nominal alpha levels.

The number of Type I error simulation runs planned for this study was 380. The number actually conducted (not counting duplicate runs and program debugging) was 1180. The number of power simulation runs planned for this study was 1520. The number actually conducted (not counting duplicate runs) was 1840. Thus, the total number of simulation runs planned was 1900 while the number actually conducted (not counting duplicate runs) was 3020. Most of the simulations were run several times.

The programs were designed so that power results for all combinations of simulation parameters were generated whether supported by Type I error results or not. However, only power results for combinations of simulation parameters that met the robustness criteria for Type I error are reported in Chapter 4 and used as the basis for conclusions in Chapter 5.

The Kolmogorov-Smirnov Test for General Differences is a two-independent-sample test. The 2-sided test is a test against an omnibus alternative. The 1-sided test is a test against the alternative that one of the samples comes from a population that is stochastically greater than the other. A method exists for dealing with ties that appears to be unique to this test in which the incrementing of the A and B arrays is deferred until the end of a sequence of ties, both arrays being incremented by the appropriate accumulated amount at that point (Neave & Worthington, 1988). This was method 6b of the study. One-sided and 2-sided tests were conducted for both $\alpha = .01$ and $\alpha = .05$. Type I error studies: 5 distributions x 28 initial per-group sample sizes = 140 simulation runs with initial per-group sample sizes of 3(1)30. Power studies: 5 distributions x 10 sample sizes x 4 effect sizes = 200 simulation runs with initial per-group sample sizes of 3(3)30. Subtotal simulation runs, 340.

Rosenbaum's Test of Location is a two-independent-sample test of general differences. The 2-sided test is against an omnibus alternative. The 1-sided test is against the alternative that a specific sample contains the greater values. One-sided and 2-sided tests were conducted for both $\alpha = .01$ and $\alpha = .05$. Type I error studies: 5 distributions x 28 sample sizes = 140 simulation runs with initial per-group sample sizes of 3(1)30. Power studies: 5 distributions x 10 initial per-group sample sizes x 4 effect sizes = 200 simulation runs with initial per-group sample sizes of 3(3)30. Subtotal simulation runs, 340.

Tukey's Quick Test of Location to Duckworth's Specifications is a two-independent-sample test of location. One-sided and 2-sided tests were conducted for both α = .01 and α = .05. Type I error studies: 5 distributions x 28 initial per-group sample sizes = 140 simulation runs with initial per-group sample sizes of 3(1)30. Power studies: 5 distributions x 10 sample sizes x 4 effect sizes = 200 simulation runs with initial per-group sample sizes 3(3)30. Subtotal simulation runs, 340.

The Wilcoxon-Mann-Whitney Test (Wilcoxon Rank-sum Test or Mann-Whitney *U* Test) is a two-independent-sample test of location. One-sided and 2-sided tests were conducted for both α = .01 and α = .05. Type I error studies: 5 distributions x 28 sample sizes = 140 simulation runs with initial per group sample sizes of 3(1)30. Power studies: 5 distributions x 10 sample sizes x 4 effect sizes = 200 simulation runs with initial sample sizes of 3(3)30. Subtotal simulation runs, 340.

The Kruskal-Wallis Test is a *k*-independent-samples test against an omnibus alternative. Tests were conducted for both α = .01 and α = .05. Type I error studies: 5 distributions x 23 initial per-group sample sizes x 4 groupings = 460 simulation runs with initial per group sample sizes of 3(1)25. Power studies: 5 distributions x 8 sample sizes x 4 effect sizes x 4 groupings = 640 simulation runs with initial per-group sample sizes of 3(3)24. Subtotal simulation runs, 1100. Neave & Worthington (1988) described an adjusted version of the statistic based on a correction factor linked to the use of mid-ranks to resolve ties. The adjusted statistic is approximately chi-squared distributed with $k - 1$ degrees of freedom and is fully described in Chapter 2. This adjusted statistic was also calculated and reported.

The Terpstra-Jonckheere Test is a *k*-independent-samples test against an ordered alternative. Tests were conducted at both $\alpha = .01$ and $\alpha = .05$. Type I error studies: 5 distributions x 8 sample sizes x 4 groupings = 160 simulation runs with initial per-group sample sizes of 2(1)10. Power studies: 5 distributions x 5 sample sizes x 4 effect sizes x 4 groupings = 400 simulation runs with initial per-group sample sizes of 2(2)10. Subtotal simulation runs, 560.

CHAPTER 4

RESULTS

4.1 – Overview

Results are presented for all tests in the order in which they were described in Chapters 2 and 3. Table 4.1-1 shows the methods studied for each test (descriptions of each method follow the table and are repeated from Chapter 3, section 4). Type I error results are presented first for all tests in section 4.2. Power and Type III error results (significant result in the wrong tail of a 2-tailed test) are then presented in section 4.3 for all tests for those combinations of simulation parameters that demonstrated acceptable Type I error performance. Finally, section 4.4 presents results on the occurrence of ties.

Table 4.1-1
Methods of Resolving Ties by Test

Test / *Method*	*K-S*	*RB*	*TQT*	*W-M-W*	*K-W*	*T-J*
M3a	x	x	x	x	x	x
M3b		x	x			
M4	x	x	x	x	x	x
M5	x	x	x	x	x	x
M6a				x	x	x
M6b	x					
M6c		x				
M7	x	x	x	x	x	x
M8	x	x	x	x	x	x

K-S = Kolmogorov-Smirnov; *RB* = Rosenbaum; *TQT* = Tukey Quick test; *W-M-W* = Wilcoxon-Mann-Whitney; *K-W* = Kruskal-Wallis; *T-J* = Terpstra-Jonckheere

For each test, Type I error results are described in narrative and summarized in an out-of-tolerance table for a specific number of groups (2-6) for each nominal alpha level (.01 then .05). The tables are organized by method of resolving ties (rows) and distribution (columns). The rows are ordered by 1-sided and 2-sided if appropriate for the test. These are labeled as: **LT1s** – Lower Tail 1-sided; **UT1s** – Upper Tail 1-sided; **LT2s** – Lower Tail 2-sided; **UT2s** – Upper Tail 2-sided; and **BT** – Both Tails. The methods are labeled as: **M3a** – average of least and most likely to reject H_0; **M3b** – Count Ties as ½ (Rosenbaum's Test and Tukey Quick Test only); **M4** – alternating; **M5** – random; **M6a** – mid-ranks; **M6b** – delayed increment (Kolmogorov-Smirnov Test only); **M6c** – weighted average of all possible resolutions (Rosenbaum's Test only); and **M8** – drop all ties and reduce N. Method 7 (**M7**) is not represented in the tables although it was studied for every test/distribution combination. The distributions (columns) are labeled as: ***P*(T1e)** – theoretical probability of a Type I Error; **Norm** – Normal; **EA** – Micceri Extreme Asymmetric; **EB** – Micceri Extreme Bimodal; **ML** – Micceri Multi-modal Lumpy; and **SS** – Micceri Smooth Symmetric.

Entries in Type I error tables consist of a pair of numbers **(L,U)** for each combination of distribution, tail and method. The first entry in each pair (L) is the number of initial sample sizes for which the Type I error rate was less than the lower tolerance limit for the nominal alpha level and directionality. The second entry in each pair (U) is the number of initial sample sizes for which the Type I error rate was greater than the upper tolerance limit for the nominal alpha level and directionality.

The acceptable tolerance band is from 0.5α to 1.1α, inclusive, when considering results of 1-sided or omnibus tests, or both tails combined for 2-sided tests. The acceptable tolerance band is 0.25α to 0.55α, inclusive, when looking at results for either the lower tail or upper tail of a two-sided test. For tests that can be viewed as 1-sided or 2-sided, acceptable results require that the performance in each tail is very similar at approximately ½ of the overall rate. If all results are acceptable, the table entry consists of '0,0'.

Type I error results for all methods except methods 7 and 8 are based on 1,000,000 repetitions for each initial sample size for which critical values were available. Results for methods 7 and 8 are based on a maximum of 10,000 testable samples, or a maximum of 10,000,000 cycles, whichever occurred first in the simulation.

Each Type I error out-of-tolerance table is followed by a summary table for method 8. These tables present selected data on testable samples and cycles. Results are not given for method 7 because the Type I error performance was unacceptable across all testing conditions. The data includes minimum and maximum average testable sample sizes (**ATSS Min**, **ATSS Max**), the minimum and maximum number of cycles (**Cycles Min**, **Cycles Max**) and the initial sample sizes at which each of these occurred (**at ISS**). Cycle counts are rounded to the nearest 1,000 (k) up to 1,000,000 or the nearest million (M) beyond 1,000,000. Note that if the maximum number of cycles reached 10M the simulation was generally unable to obtain 10,000 testable samples. Type I error rates, Powers and Type III error rates are always presented as a proportion of the testable samples obtained.

Some tests had the theoretical probability of a Type I error, P(T1e), less than the lower limit of robustness for some or many initial sample sizes, i.e., the Terpstra-Jonckheere Test, Kolmogorov-Smirnov Test and Rosenbaum Test. P(T1e) was not available for the Kruskal-Wallis Test. In all cases where P(T1e) was available, the experimental occurrence of Type 1 errors for the Normal distribution tracked P(T1e) closely for all methods, with some departure for methods 7 and 8. This was to be expected as the normal distribution represents a population from which $P(\text{tie}) \approx 0$ and methods 7 and 8 involved separate simulations that yielded a smaller number of testable samples than the other methods.

In analyzing Type I error results in order to select combinations of distributions and methods for further study of power and Type III error a choice was made to go forward with combinations where the Type I error results were similar to or more conservative than P(T1E) and/or the results with the Normal distribution. To require strict adherence to the robust limits of $.5\alpha \leq p(\text{T1e}) \leq 1.1\alpha$ would have eliminated some tests from further study, such as the Kolmogorov-Smirnov Test, which was deemed undesirable. In general, power results were investigated if the Type I error properties were equal to or more conservative than the theoretical and/or Normal distribution results but not if they showed any tendency to be inflated.

Power and Type III error results, if appropriate, are presented in section 4.3 for all tests with acceptable Type I error results. For tests on three or more groups, all results are presented for three groups, followed by four, five and finally six groups. All table entries are rounded to the fourth decimal place.

Power and Type III error results for each test are presented in a series of tables for each nominal effect size multiplier covering both nominal alpha .01 and .05. For 2-sided tests, five tables are presented at each nominal effect size multiplier (0.2, 0.5, 0.8 and 1.2) corresponding to initial sample sizes 6, 12, 18, 24, and 30. Unlike Type I error rates, results for power and Type III error tend to behave in a more consistent and predictable manner. Thus, data for these five initial sample sizes are generally sufficient to reveal the behavior of the statistic for a specific combination of conditions.

For omnibus (Kruskal-Wallis) and ordered alternative hypothesis (Terpstra-Jonckheere) tests the concept of Type III error does not apply. Thus, results for both nominal alpha .01 and .05 are presented on a single page for each combination of nominal effect size multiplier {0.2, 0.5, 0.8, 1.2} and number of groups {3, 4, 5, 6}. Initial sample sizes 6, 12, 18 and 24 are reported for the Kruskal-Wallis Test while results for initial sample sizes 2, 4, 6, 8 and 10 are reported for the Terpstra-Jonckheere Test.

The rows of the Power and Type III error tables are organized by lower tail 1-sided (**LT1s**), upper tail 1-sided (**UT1s**), lower tail 2-sided (**LT2s**) and upper tail 2-sided (**UT2s**) when this is appropriate to the test. In all cases the lower tail results (LT1s and LT2s) represent Type III error rates and the upper tail results (UT1s and UT2s) represent the Power of the test under the specific combination of conditions (all effect shifts are towards higher values). Within each of these four directional categories are rows for each method of resolving ties. Results for method 8 (drop all ties and reduce N) are not presented because the Type I error results for this method were consistently outside the *a priori* tolerance limits. The columns represent the distributions. Results for the EB distribution are only given for nominal effect size multipliers 0.5 and 1.2 as the actual effect size is only reasonably close to the nominal effect size for these two effect size multipliers.

Three types of rank analysis tables for the power results follow power and Type III error results for each test. The first type of table gives the mean ranks of the power results for the applicable methods within each distribution and initial sample size across effect sizes. The second type of table gives the mean ranks of the power results for the applicable methods within each distribution and effect size across initial sample sizes. The third type of table gives the mean ranks of the power results for the applicable methods within each distribution across initial sample sizes and effect sizes.

Table 4.1-2
Actual Effect Size Multipliers by Nominal Effect Size Multiplier and Distribution

NESM *Distribution*	*.2*	*.5*	*.8*	*1.2*
Norm	.2	.5	.8	1.2
EA	.172 (2)	.518 (2)	.864 (1)	1.209 (2)
EB		.592 (1)		1.183 (3)
ML	.168 (3)	.504 (3)	.840 (2)	1.176 (4)
SS	.204 (1)	.407 (4)	.815 (3)	1.222 (1)

Care should be taken in comparing power results across distributions since the actual effect size multipliers for any given nominal effect size multiplier were not equal. The actual effect size multipliers and the rankings across distributions within nominal effect size multiplier are given in Table 4.1-2.

4.2 – Type I Error Results

4.2.1 – Kolmogorov-Smirnov Test of General Differences (two groups only)

Type I error results for the Kolmogorov-Smirnov Test at nominal alpha .01 are summarized in Table 4.2.1-1. The theoretical probability of a Type I error, *P*(T1e), was less than .005 for LT1s and UT1s for 14 initial samples sizes {3-7, 9, 11, 12, 15, 18, 19, 22, 23, 27} out of 28 and for BT (LT2s, UT2s) for 13 initial sample sizes {3, 4, 6, 8, 10, 11, 13, 14, 16, 17, 20, 24, 28} out of 28. *P*(T1e) never exceeded .011 due to the strictly conservative criteria used to select critical values. Table 4.2.1-2 shows testable samples results for method 8.

Type I error results for the Kolmogorov-Smirnov Test at nominal alpha .05 are summarized in Table 4.2.1-3. *P*(T1e) was less than .025 for LT1s and UT1s for six initial samples sizes {4, 6, 9, 13, 17, 22} out of 28 and for BT (LT2s, UT2s) for eight initial sample sizes {5, 7, 8, 10, 11, 14, 18, 23} out of 28. *P*(T1e) never exceeded 0.055 because of the strictly conservative criteria used to select critical values. Table 4.2.1-4 shows testable samples results for method 8.

Type I error rates for the Normal distribution at nominal alpha .01 and .05 were generally close to *P*(T1e) across methods of resolving ties as there were no ties except for the combination of method 8 with LT1s, LT2s and BT at nominal alpha .01. Otherwise, results for method 8 were similar to the other methods, but not identical due to the smaller number of repetitions involved in the method 8 simulation. These results indicate that the basic simulations worked correctly.

Table 4.2.1-1
Kolmogorov-Smirnov Test of General Differences for Two Groups
*Type I Error Out-of-Tolerance Counts for α .01, 28 Initial Sample Sizes, 3(1)30 **

			Distributions				
Tail	*Mthd*	*P*(T1e)	Norm	EA	EB	ML	SS
LT 1s	M3a	14,0	14,0	17,0	8,12	18,0	17,0
	M4	14,0	14,0	16,0	23,0	14,0	16,0
	M5	14,0	14,0	14,0	14,0	14,0	14,0
	M6b	14,0	14,0	25,0	27,0	18,0	25,0
	M8	12,0	12,0	0,25	0,25	1,21	0,23
UT 1s	M3a	14,0	14,0	17,0	8,12	18,0	17,0
	M4	14,0	14,0	16,0	23,0	15,0	17,0
	M5	14,0	14,0	14,0	14,0	14,0	14,0
	M6b	14,0	14,0	25,0	27,0	18,0	25,0
	M8	12,0	12,0	0,25	0,25	1,21	0,25
LT 2s	M3a	13,0	13,0	15,0	11,12	16,0	15,0
	M4	13,0	13,0	16,0	20,0	14,0	17,0
	M5	13,0	13,0	13,0	12,0	12,0	12,0
	M6b	13,0	14,0	25,0	27,0	17,0	24,0
	M8	11,0	0,8	0,25	0,25	0,24	0,25
UT 2s	M3a	13,0	13,0	15,0	11,12	16,0	15,0
	M4	13,0	13,0	16,0	20,0	13,0	17,0
	M5	13,0	13,0	13,0	13,0	11,0	12,0
	M6b	13,0	14,0	25,0	27,0	17,0	23,0
	M8	11,0	0,12	0,25	0,25	0,24	0,25
BT	M3a	13,0	13,0	15,0	11,12	16,0	15,0
	M4	13,0	13,0	16,0	20,0	14,0	17,0
	M5	13,0	13,0	13,0	13,0	12,0	12,0
	M6b	13,0	14,0	25,0	27,0	17,0	24,0
	M8	11,0	0,9	0,25	0,25	0,24	0,25

* Method 8 sample size range is 5(1)30.

Type I error rates for the Micceri distributions at nominal alpha .01 and .05 showed a pattern similar to *P*(T1e) across methods, but not as close as for the Normal distribution and with results somewhat to considerably more conservative. Exceptions were the combination of method 3 with the EB distribution and method 8 with all Micceri distributions, where results were inflated.

Table 4.2.1-2
Kolmogorov-Smirnov Test of General Differences for Two Groups
Method 8 – Testable Samples and Cycles for α .01
26 Initial Sample Sizes, 5(1)30

		Distributions			
Dir	*Datum*	EA	EB	ML	SS
LT 1s	ATSS Min	5.00	4.91	5.00	5.00
UT1s	at ISS	5	6	5	5
	ATSS Max	6.88	5.53	13.52	7.52
	at ISS	29	19	30	24
	Cycles Min	14k	1M	10k	11k
	at ISS	28	6	25	19
	Cycles Max	66k	10M	20k	39k
	at ISS	5	23	5	5
BT	ATSS Min	5.00	5.00	5.00	5.00
LT2s	at ISS	5	5	5	5
UT2s	ATSS Max	7.11	6.00	13.49	7.72
	at ISS	29	21	30	25
	Cycles Min	15k	1M	10k	12k
	at ISS	29	4	30	17
	Cycles Max	66k	10M	20k	40k
	at ISS	5	27	5	5

Table 4.2.1-3
Kolmogorov-Smirnov Test of General Differences for Two Groups
*Type I Error Out-of-Tolerance Counts for α .05, 28 Initial Sample Sizes, 3(1)30 **

			-------------------	Distributions		---------------------	
Tail	*Mthd*	*P*(T1e)	Norm	EA	EB	ML	SS
LT 1s	M3a	6,0	6,0	9,0	4,15	11,0	10,0
	M4	6,0	6,0	10,0	17,0	6,0	8,0
	M5	6,0	6,0	6,0	6,0	6,0	6,0
	M6b	6,0	6,0	20,0	26,0	13,0	18,0
	M8	6,0	6,0	0,27	0,27	0,21	0,24
UT 1s	M3a	6,0	6,0	9,0	4,15	11,0	10,0
	M4	6,0	6,0	8,0	17,0	6,0	8,0
	M5	6,0	6,0	6,0	6,0	6,0	6,0
	M6b	6,0	6,0	20,0	26,0	13,0	18,0
	M8	6,0	7,0	0,27	0,27	0,21	0,24
LT 2s	M3a	9,0	9,0	15,0	8,14	12,0	15,0
	M4	9,0	9,0	12,0	20,0	10,0	13,0
	M5	9,0	9,0	8,0	7,0	9,0	8,0
	M6b	9,0	9,0	23,0	27,0	15,0	24,0
	M8	9,0	10,0	0,27	0,27	0,21	0,24
UT 2s	M3a	9,0	9,0	15,0	8,14	13,0	16,0
	M4	9,0	9,0	14,0	20,0	10,0	15,0
	M5	9,0	9,0	8,0	9,0	9,0	9,0
	M6b	9,0	9,0	23,0	27,0	15,0	24,0
	M8	9,0	9,0	0,27	0,27	1,20	0,25
BT	M3a	9,0	9,0	14,0	7,14	13,0	16,0
	M4	9,0	9,0	13,0	20,0	10,0	12,0
	M5	9,0	9,0	8,0	9,0	9,0	9,0
	M6b	9,0	9,0	23,0	27,0	15,0	24,0
	M8	9,0	10,0	0,27	0,27	1,20	0,24

* Method 8 sample size range is 4(1)30 for BT (LT2s, UT2s).

Table 4.2.1-4
Kolmogorov-Smirnov Test of General Differences for Two Groups
Method 8 – Testable Samples and Cycles for α .05
27 Initial Sample Sizes, 4(1)30

		Distributions			
Dir	*Datum*	EA	EB	ML	SS
LT 1s	ATSS Min	3.00	3.00	3.00	3.00
UT1s	at ISS	3	3	3	3
	ATSS Max	6.36	4.08	13.50	7.17
	at ISS	30	14	30	23
	Cycles Min	11k	92k	10k	10k
	at ISS	24	3	17	15
	Cycles Max	21k	10M	13k	16k
	at ISS	3	18	3	3
BT	ATSS Min	4.00	4.00	4.00	4.00
LT2s	at ISS	4	4	4	4
UT2s	ATSS Max	6.59	4.92	13.49	7.33
	at ISS	30	18	30	24
	Cycles Min	12k	320k	10k	11k
	at ISS	30	4	26	17
	Cycles Max	35k	10M	16k	24k
	at ISS	4	12	4	4

4.2.2 – Rosenbaum's Test (two groups only)

Type I error results for Rosenbaum's Test at nominal alpha .01 are summarized in Table 4.2.2-1. *P*(T1e) was less than .005 for LT1s and UT1s for 11 initial samples sizes {4, 7-9, 20-26} out of 26 and for BT (LT2s, UT2s) for nine initial sample sizes {6, 7, 10-12, 27-30} out of 26. *P*(T1e) never exceeded .011 because of the strictly conservative criteria used to select critical values. Table 4.2.2-2 shows testable samples results for method 8.

Type I error results for Rosenbaum's Test at nominal alpha .05 are summarized in Table 4.2.2-3. *P*(T1e) was less than .025 for LT1s and UT1s for

11 initial sample sizes {4, 5, 16-24} out of 28 and for BT (LT2s, UT2s) for eight initial sample sizes {6, 7, 25-30} out of 27. *P*(T1e) never exceeded 0.055 because of the strictly conservative criteria used to select critical values. Table 4.2.2-4 shows testable samples results for method 8.

Type I error rates for the Normal distribution at nominal alpha .01 and .05 were generally very close to *P*(T1e) across methods of resolving ties as there were no ties. Results for method 8 were similar to the other methods but not quite as close due to the smaller number of repetitions involved in the method 8 simulation. These results indicate that the basic simulation worked correctly.

Type I error rates for the Micceri distributions at nominal alpha .01 and .05 showed a pattern similar to *P*(T1e) across methods, but not as close as for the Normal distribution and with results somewhat to considerably more conservative. Exceptions at both nominal alpha .01 and .05 were methods 3a and 3b with distribution EB, method 4 with distributions EA and EB and method 8 for all Micceri distributions, all of which showed inflated results.

Table 4.2.2-1

Rosenbaum's Test for Two Groups Type I Error Out-of-Tolerance Counts for α .01, 26 Initial Sample Sizes, 5(1)30

			Distributions				
Tail	*Mthd*	*P*(T1e)	Norm	EA	EB	ML	SS
LT 1s	M3a	11,0	11,0	18,0	19,2	11,0	17,0
	M3b	11,0	11,0	18,0	19,2	11,0	17,0
	M4	11,0	11,0	3,11	3,19	9,0	8,0
	M5	11,0	11,0	11,0	11,0	12,0	13,0
	M6c	11,0	11,0	25,0	25,0	16,0	18,0
	M8	11,0	11,0	0,25	0,25	0,20	0,25
UT 1s	M3a	11,0	12,0	18,0	19,2	16,0	17,0
	M3b	11,0	12,0	18,0	19,2	11,0	17,0
	M4	11,0	12,0	3,11	3,19	9,0	8,0
	M5	11,0	12,0	10,0	12,0	13,0	12,0
	M6c	11,0	12,0	25,0	25,0	16,0	18,0
	M8	11,0	10,0	0,25	0,25	0,19	0,25
LT 2s	M3a	9,0	9,0	13,0	23,0	16,0	12,0
	M3b	9,0	9,0	13,0	23,0	11,0	12,0
	M4	9,0	9,0	4,12	3,18	8,0	8,0
	M5	9,0	9,0	9,0	9,0	10,0	9,0
	M6c	9,0	9,0	23,0	25,0	11,0	15,0
	M8	9,0	9,0	0,25	0,25	0,25	0,25
UT 2s	M3a	9,0	9,0	13,0	24,0	11,0	13,0
	M3b	9,0	9,0	13,0	24,0	11,0	13,0
	M4	9,0	9,0	4,12	3,18	9,0	7,0
	M5	9,0	9,0	9,0	9,0	9,0	9,0
	M6c	9,0	9,0	23,0	25,0	11,0	15,0
	M8	9,0	7,0	0,25	0,25	0,25	0,25
BT	M3a	9,0	9,0	13,0	23,0	11,0	13,0
	M3b	9,0	9,0	13,0	23,0	11,0	13,0
	M4	9,0	9,0	4,12	3,18	9,0	7,0
	M5	9,0	9,0	9,0	9,0	9,0	9,0
	M6c	9,0	9,0	23,0	25,0	11,0	15,0
	M8	9,0	8,0	0,25	0,25	0,25	0,25

Table 4.2.2-2
Rosenbaum's Test for Two Groups, Method 8 – Testable Samples and Cycles for α .01, 26 Initial Sample Sizes, 5(1)30

		Distributions			
Dir	*Datum*	EA	EB	ML	SS
LT 1s	ATSS Min	5.00	4.91	5.00	5.00
UT1s	at ISS	5	6	5	5
	ATSS Max	6.88	5.51	13.48	7.53
	at ISS	29	20	30	25
	Cycles Min	14k	900k	10k	11k
	at ISS	28	6	23	18
	Cycles Max	66k	10M	20k	40k
	at ISS	5	23	5	5
BT	ATSS Min	5.00	5.00	5.00	5.00
LT2s	at ISS	5	5	5	5
UT2s	ATSS Max	7.11	5.99	13.47	7.73
	at ISS	29	21	30	25
	Cycles Min	15k	1M	10k	12k
	at ISS	29	5	29	19
	Cycles Max	66k	10M	20k	40k
	at ISS	5	19	5	5

Table 4.2.2-3
*Rosenbaum's Test for Two Groups Type I Error Out-of-Tolerance Counts for α .05, 28 Initial Sample Sizes, 3(1)30**

			Distributions				
Tail	*Mthd*	*P*(T1e)	Norm	EA	EB	ML	SS
LT 1s	M3a	11,0	11,0	19,0	14,8	17,0	18,0
	M3b	11,0	11,0	19,0	14,8	17,0	18,0
	M4	11,0	11,0	2,12	2,22	8,0	7,0
	M5	11,0	11,0	12,0	11,0	10,0	11,0
	M6c	11,0	11,0	23,0	26,0	17,0	18,0
	M8	11,0	8,0	0,27	0,27	0,18	0,23
UT 1s	M3a	11,0	12,0	19,0	14,8	17,0	18,0
	M3b	11,0	12,0	19,0	14,8	17,0	18,0
	M4	11,0	12,0	2,12	2,22	8,0	7,0
	M5	11,0	12,0	11,0	11,0	11,0	11,0
	M6c	11,0	12,0	23,0	26,0	17,0	18,0
	M8	11,0	11,0	0,27	0,27	0,18	0,23
LT 2s *	M3a	8,0	8,0	12,0	13,9	9,0	11,0
	M3b	8,0	8,0	12,0	13,9	9,0	11,0
	M4	8,0	8,0	2,16	2,20	7,0	5,0
	M5	8,0	8,0	8,0	8,0	8,0	8,0
	M6c	8,0	8,0	25,0	25,0	9,0	12,0
	M8	8,0	6,0	0,26	0,26	0,19	0,26
UT 2s *	M3a	8,0	8,0	12,0	13,9	9,0	11,0
	M3b	8,0	8,0	12,0	13,9	9,0	11,0
	M4	8,0	8,0	2,16	2,20	7,0	5,0
	M5	8,0	8,0	8,0	8,0	8,0	8,0
	M6c	8,0	8,0	25,0	25,0	9,0	12,0
	M8	8,0	7,1	0,26	0,26	0,20	0,26
BT *	M3a	8,0	8,0	12,0	13,9	9,0	11,0
	M3b	8,0	8,0	11,0	13,9	9,0	11,0
	M4	8,0	8,0	2,16	2,20	7,0	5,0
	M5	8,0	8,0	8,0	8,0	8,0	8,0
	M6c	8,0	8,0	25,0	25,0	9,0	12,0
	M8	8,0	7,0	0,26	0,26	0,20	0,26

* For LT1s and UT1s. 27 initial sample sizes 4(1)30 for BT (LT2s, UT2s).

Table 4.2.2-4
*Rosenbaum's Test for Two Groups, Method 8 – Testable Samples and Cycles for α .05, 28 Initial Sample Sizes, 3(1)30**

		Distributions			
Dir	*Datum*	EA	EB	ML	SS
LT 1s	ATSS Min	3.00	3.00	3.00	3.00
UT1s	at ISS	3	3	3	3
	ATSS Max	6.36	4.08	13.51	7.16
	at ISS	30	14	30	22
	Cycles Min	11k	92k	10k	10k
	at ISS	30	3	15	16
	Cycles Max	21k	2M	13k	17k
	at ISS	3	30	3	3
BT	ATSS Min	4.00	4.00	4.00	4.00
LT2s	at ISS	4	4	4	4
UT2s	ATSS Max	6.59	4.93	13.52	7.35
	at ISS	30	17	30	26
	Cycles Min	12k	320k	10k	11k
	at ISS	30	4	19	18
	Cycles Max	35k	5M	16k	25k
	at ISS	4	29	4	4

* For LT1s and UT1s. 27 initial sample sizes 4(1)30 for BT (LT2s, UT2s).

4.2.3 – Tukey's Quick Test (two groups only)

Type I error results for Tukey's Quick Test at nominal alpha .01 are summarized in Table 4.2.3-1 while Type I error results at nominal alpha .05 are summarized in Table 4.2.3-3. *P*(T1e) was within the robust limits for all combinations of initial samples sizes, directions and methods. Tables 4.2.3-2 and 4.2.3-4 show testable samples results for method 8 at nominal alpha .01 and .05 respectively.

Type I error rates for the Normal distribution at nominal alpha .01 and .05 were generally very close to *P*(T1e) across all methods of resolving ties as there were no ties. Results for method 8 departed slightly due to the smaller number of testable samples involved in the method 8 simulation. These results indicate that the basic simulation worked correctly.

Type I error rates for the Micceri distributions at nominal alpha .01 and .05 were very similar to P(T1e) and the Normal distribution for method 5 across all distributions and for methods 3a, 3b and 4 for distributions ML and SS with departures being slightly conservative. Results were very conservative for method 3a with distribution EA at both alpha levels, while results were slightly conservative for method 3b with distribution EA at nominal alpha .01, but inflated at nominal alpha .05. Results were somewhat to grossly inflated at both alpha levels for methods 3a and 3b with distribution EB, method 4 with distributions EA and EB and method 8 with all Micceri distributions.

Table 4.2.3-1

Tukey's Quick Test for Two Groups, Type I Error Out-of-Tolerance Counts for α .01, 26 Initial Sample Sizes, 5(1)30

			Distributions				
Tail	*Mthd*	*P*(T1e)	Norm	EA	EB	ML	SS
LT 1s	M3a	0,0	0,0	22,0	11,13	4,0	4,0
	M3b	0,0	0,0	4,0	2,11	3,0	3,0
	M4	0,0	0,0	0,13	13,2	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M8	0,0	1,0	0,26	0,26	0,26	0,26
UT 1s	M3a	0,0	0,0	22,0	11,13	3,0	4,0
	M3b	0,0	0,0	4,0	1,13	3,0	4,0
	M4	0,0	0,0	0,13	13,2	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M8	0,0	1,1	0,26	0,26	0,26	0,26
LT 2s	M3a	0,0	0,0	22,0	11,12	4,0	4,0
	M3b	0,0	0,0	5,0	3,10	4,0	4,0
	M4	0,0	0,0	0,14	14,2	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M8	0,0	2,0	0,26	0,26	0,25	0,26
UT 2s	M3a	0,0	0,0	22,0	11,12	3,0	4,0
	M3b	0,0	0,0	4,0	2,17	3,0	4,0
	M4	0,0	0,0	0,14	14,2	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M8	0,0	1,0	0,26	0,26	0,26	0,26
BT	M3a	0,0	0,0	22,0	11,12	4,0	4,0
	M3b	0,0	0,0	4,0	2,11	3,0	4,0
	M4	0,0	0,0	0,14	14,2	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M8	0,0	1,0	0,26	0,26	0,6	0,26

Table 4.2.3-2
Tukey's Quick Test for Two Groups, Method 8 – Testable Samples and Cycles for α .01, 26 Initial Sample Sizes, 5(1)30

		Distributions			
Dir	*Datum*	EA	EB	ML	SS
LT 1s	ATSS Min	5.00	4.91	5.00	5.00
UT1s	at ISS	5	6	5	5
	ATSS Max	6.88	5.51	13.48	7.53
	at ISS	29	20	30	25
	Cycles Min	14k	898k	10k	11k
	at ISS	28	6	19	18
	Cycles Max	66k	10M	20k	40k
	at ISS	5	23	5	5
BT	ATSS Min	5.00	5.00	5.00	5.00
LT2s	at ISS	5	5	5	5
UT2s	ATSS Max	7.11	5.99	13.47	7.73
	at ISS	29	21	30	25
	Cycles Min	15k	1M	10k	12k
	at ISS	29	5	29	19
	Cycles Max	66k	10M	20k	40k
	at ISS	5	19	5	5

Table 4.2.3-3

*Tukey's Quick Test for Two Groups, Type I Error Out-of-Tolerance Counts for α .05, 28 Initial Sample Sizes, 3(1)30 **

			------------------	Distributions		----------------------	
Tail	*Mthd*	*P*(T1e)	Norm	EA	EB	ML	SS
LT 1s	M3a	0,0	0,0	21,0	7,18	0,0	1,0
	M3b	0,0	0,0	1,9	3,12	0,0	1,1
	M4	0,0	0,0	0,17	9,6	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M8	0,0	0,0	0,28	0,28	0,22	0,24
UT 1s	M3a	0,0	0,0	21,0	7,18	0,0	1,0
	M3b	0,0	0,0	1,9	1,21	0,0	1,1
	M4	0,0	0,0	0,17	8,6	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M8	0,0	0,0	0,28	0,28	0,22	0,24
LT 2s *	M3a	0,0	0,0	16,0	9,17	6,0	6,0
	M3b	0,0	0,0	0,4	4,13	6,0	5,0
	M4	0,0	0,0	0,18	8,2	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M8	0,0	1,0	0,27	0,27	0,23	0,27
UT 2s *	M3a	0,0	0,0	16,0	9,17	6,0	6,0
	M3b	0,0	0,0	0,4	2,21	6,0	5,0
	M4	0,0	0,0	0,17	8,2	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M8	0,0	0,0	0,27	0,27	0,23	0,24
BT *	M3a	0,0	0,0	16,0	9,17	6,0	6,0
	M3b	0,0	0,0	0,4	3,15	6,0	5,0
	M4	0,0	0,0	0,17	8,20	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M8	0,0	0,0	0,27	0,27	0,23	0,26

* For LT1s and UT1s. 27 initial sample sizes 4(1)30 for BT (LT2s, UT2s).

Table 4.2.3-4
Tukey's Quick Test for Two Groups, Method 8 – Testable Samples and Cycles for α .05, 28 Initial Sample Sizes, 3(1)30

		Distributions			
Dir	*Datum*	EA	EB	ML	SS
LT 1s	ATSS Min	3.00	3.00	3.00	3.00
UT1s	at ISS	3	3	3	3
	ATSS Max	6.36	4.08	13.51	7.16
	at ISS	30	14	30	22
	Cycles Min	11k	92k	10k	10k
	at ISS	30	3	15	16
	Cycles Max	21k	2M	13k	17k
	at ISS	3	30	3	3
BT	ATSS Min	4.00	4.00	4.00	4.00
LT2s	at ISS	4	4	4	4
UT2s	ATSS Max	6.59	4.93	13.52	7.35
	at ISS	30	17	30	26
	Cycles Min	12k	320k	10k	11k
	at ISS	30	4	19	18
	Cycles Max	35k	5M	16k	25k
	at ISS	4	30	4	4

* For LT1s and UT1s. 27 initial sample sizes 4(1)30 for BT (LT2s, UT2s).

4.2.4 – Wilcoxon- Mann-Whitney Test (two groups only)

Type I error results for the Wilcoxon-Mann-Whitney Test at nominal alpha .01 and .05 are summarized in Tables 4.2.4-1 and 4.2.4-3 respectively. *P*(T1e) was with the robust tolerance limits for all combinations of initial sample size, direction and method. Tables 4.2.4-2 and 4.2.4-4 show method 8 testable samples results for nominal alpha .01 and .05 respectively.

Type I error rates for the Normal distribution at nominal alpha .01 and .05 were very close to *P*(T1e) across methods of resolving ties as there were no ties. Results for method 8 at nominal alpha .01 departed slightly due to the smaller number of repetitions involved in the method 8 simulation. These results indicate that the basic simulations worked correctly.

Type I error rates at both nominal alpha levels were very close to slightly conservative in comparison to the *P*(T1e) and Normal results for all of the methods except method 8 across the Micceri distributions and method 4 with distribution EA. The exceptions all had grossly inflated Type I error rates.

Table 4.2.4-1
Wilcoxon-Mann-Whitney Test for Two Groups, Type I Error Out-of-Tolerance Counts for α .01, 26 Initial Sample Sizes, 5(1)30

			Distributions				
Tail	*Mthd*	*P*(T1e)	Norm	EA	EB	ML	SS
LT 1s	M3a	0,0	0,0	0,0	3,0	0,0	0,0
	M4	0,0	0,0	0,19	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	0,0	3,0	0,0	0,0
	M8	0,0	0,0	0,26	0,26	0,26	0,26
UT 1s	M3a	0,0	0,0	0,0	3,0	0,0	0,0
	M4	0,0	0,0	0,19	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	0,0	3,0	0,0	0,0
	M8	0,0	0,2	0,26	0,26	0,26	0,26
LT 2s	M3a	0,0	0,0	0,0	5,0	0,0	0,0
	M4	0,0	0,0	0,17	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	0,0	5,0	0,0	0,0
	M8	0,0	0,0	0,26	0,26	0,26	0,26
UT 2s	M3a	0,0	0,0	0,0	5,0	0,0	0,0
	M4	0,0	0,0	0,17	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	0,0	5,0	0,0	0,0
	M8	0,0	0,2	0,26	0,26	0,25	0,26
BT	M3a	0,0	0,0	0,0	5,0	0,0	0,0
	M4	0,0	0,0	0,17	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	0,0	5,0	0,0	0,0
	M8	0,0	0,2	0,26	0,26	0,26	0,26

Table 4.2.4-2
Wilcoxon-Mann-Whitney Test for Two Groups, Method 8 – Testable Samples and Cycles for α .01, 26 Initial Sample Sizes, 5(1)30

		Distributions			
Dir	*Datum*	EA	EB	ML	SS
LT 1s	ATSS Min	5.00	4.91	5.00	5.00
UT1s	at ISS	5	6	5	5
	ATSS Max	6.88	5.51	13.48	7.53
	at ISS	29	20	30	25
	Cycles Min	14k	898k	10k	11k
	at ISS	28	6	23	18
	Cycles Max	66k	10M	20k	40k
	at ISS	5	23	5	5
BT	ATSS Min	5.00	5.00	5.00	5.00
LT2s	at ISS	5	5	5	5
UT2s	ATSS Max	7.11	5.99	13.47	7.73
	at ISS	29	21	30	25
	Cycles Min	15k	1M	10k	12k
	at ISS	29	5	29	19
	Cycles Max	66k	10M	20k	40k
	at ISS	5	19	5	5

Table 4.2.4-3
*Wilcoxon-Mann-Whitney Test for Two Groups, Type I Error Out-of-Tolerance Counts for α .05, 28 Initial Sample Sizes, 3(1)30**

			Distributions				
Tail	*Mthd*	*P*(T1e)	Norm	EA	EB	ML	SS
LT 1s	M3a	0,0	0,0	1,0	1,0	0,0	1,0
	M4	0,0	0,0	0,15	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	1,0	1,0	0,0	1,0
	M8	0,0	0,0	0,28	0,28	0,26	0,26
UT 1s	M3a	0,0	0,0	1,0	1,0	0,0	1,0
	M4	0,0	0,0	0,15	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	1,0	1,0	0,0	1,0
	M8	0,0	0,0	0,28	0,28	0,26	0,27
LT 2s	M3a	0,0	0,0	1,0	3,0	0,0	1,0
	M4	0,0	0,0	0,17	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	1,0	3,0	0,0	1,0
	M8	0,0	0,0	0,27	0,27	0,24	0,25
UT 2s	M3a	0,0	0,0	1,0	3,0	0,0	1,0
	M4	0,0	0,0	0,17	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	1,0	3,0	0,0	1,0
	M8	0,0	0,0	0,27	0,27	0,23	0,24
BT	M3a	0,0	0,0	1,0	3,0	0,0	1,0
	M4	0,0	0,0	0,17	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	1,0	3,0	0,0	1,0
	M8	0,0	0,0	0,27	0,27	0,24	0,25

* For LT1s and UT1s. 27 initial sample sizes 4(1)30 for BT (LT2s, UT2s).

Table 4.2.4-4
*Wilcoxon-Mann-Whitney Test for Two Groups, Method 8 – Testable Samples and Cycles for α .05, 28 Initial ample Sizes, 3(1)30**

		Distributions			
Dir	*Datum*	EA	EB	ML	SS
LT 1s	ATSS Min	3.00	3.00	3.00	3.00
UT1s	at ISS	3	3	3	3
	ATSS Max	11.97	6.00	17.19	13.02
	at ISS	30	30	30	30
	Cycles Min	10k	11k	10k	10k
	at ISS	20	30	10	14
	Cycles Max	21k	92k	13k	17k
	at ISS	3	3	3	3
BT	ATSS Min	4.00	4.00	4.00	4.00
LT2s	at ISS	4	4	4	4
UT2s	ATSS Max	12.00	6.34	17.27	13.04
	at ISS	30	30	30	30
	Cycles Min	10k	12k	10k	10k
	at ISS	25	30	13	22
	Cycles Max	35k	314k	16k	25k
	at ISS	4	4	4	4

* For LT1s and UT1s. 27 initial sample sizes 4(1)30 for BT (LT2s, UT2s).

4.2.5 – Kruskal-Wallis Test (three to six groups)

Type I error results for the Kruskal-Wallis Test at nominal alpha .01 are summarized in Table 4.2.5-1 for 3 to 6 groups while corresponding results for nominal alpha .05 are given in Table 4.2.5-2. Testable samples results for method 8 are given in Tables 4.2.5-3 through 4.2.5-6 for three to six groups respectively.

Tables 4.2.5-1 and 4.2.5-2 differ from the corresponding tables for the previous four tests in several ways. First, since the Kruskal-Wallis test is a test of an omnibus alternative hypothesis, 1-sided and 2-sided results have no meaning and thus are not reported. Second, the test is defined for three or more groups. Results were obtained and reported for three, four, five and six groups. Third, *P*(T1e), was not readily available for this test and thus is not reported. Fourth, a continuity corrected statistic, H_c, was available for this test for which results were calculated and reported alongside the uncorrected results (H).

Since *P*(T1e) was not readily available, the results for the Normal distribution were taken as the standard of comparison for the other test combinations. This seemed reasonable given the closeness of results between *P*(T1e) and the Normal distribution for the other tests in this study.

For three groups at nominal alpha .01, *P*(T1e) for the Normal distribution was less than .005 for initial sample size 3, while *P*(T1e) exceeded .011 for initial sample sizes 22 and 24 for methods 3 through 6 and for initial sample sizes 5, 20 and 23 for method 8.

For four groups at nominal alpha .01, *P*(T1e) for the Normal distribution remained at or above .005 for all initial sample sizes while *P*(T1e) exceeded .011 for initial sample sizes 20 and 25 for methods 3 through 6 and for initial sample size 18 for method 8.

For five groups at nominal alpha .01, *P*(T1e) for the Normal distribution remained at or above .005 for all initial sample sizes while *P*(T1e) exceeded .011 for initial sample sizes 5,19 for methods 3 through 6 and for initial sample size 15 for method 8.

Table 4.2.5-1
*Kruskal-Wallis Test for Three to Six Groups, Type I Error Out-of-Tolerance Counts for α .01, 23 Initial Sample Sizes, 3(1)25**

		Distributions**									
Groups	*Mthd*	....Norm....		......EA......		......EB......		......ML......		......SS......	
		H	Hc	H	Hc	H	Hc	H	Hc	H	Hc
3 Grps	M3a	1,2	1,2	1,11	1,14	1,19	0,23	1,3	1,3	1,7	1,10
	M4	1,2	1,2	1,18	1,21	1,0	0,14	1,3	1,4	1,7	1,9
	M5	1,2	1,2	1,3	1,6	1,1	0,23	1,3	1,3	1,6	1,5
	M6a	1,2	1,2	1,0	1,1	6,0	1,1	1,3	1,3	1,1	1,5
	M8*+	1,3	1,3	0,22	0,22	0,22	0,22	0,21	0,21	0,21	0,21
4 Grps	M3a	0,2	0,2	0,16	0,18	0,22	0,23	0,3	0,4	0,5	0,6
	M4	0,2	0,2	0,21	0,23	0,0	0,15	0,2	0,3	0,2	0,5
	M5	0,2	0,2	0,4	0,16	0,4	0,23	0,5	0,5	0,4	0,5
	M6a	0,2	0,2	0,2	0,4	8,0	0,7	0,2	0,3	0,0	0,1
	M8*+	0,1	0,1	0,22	0,22	++	++	0,22	0,22	0,22	0,22
5 Grps	M3a	0,2	0,2	0,18	0,21	0,22	0,23	0,0	0,1	0,4	0,6
	M4	0,2	0,2	0,22	0,23	0,0	0,14	0,2	0,4	0,2	0,4
	M5	0,2	0,2	0,6	0,11	0,3	0,23	0,1	0,1	0,0	0,2
	M6a	0,2	0,2	0,1	0,6	9,0	0,4	0,1	0,1	0,0	0,1
	M8*+	0,1	0,1	0,22	0,22	++	++	0,22	0,22	0,22	0,22
6 Grps	M3a	0,1	0,1	0,16	0,22	0,23	0,23	0,1	0,1	0,8	0,13
	M4	0,1	0,1	0,23	0,23	0,0	0,20	0,0	0,0	0,5	0,7
	M5	0,1	0,1	0,2	0,15	0,4	0,23	0,4	0,4	0,2	0,5
	M6a	0,1	0,1	0,0	0,1	10,0	0,3	0,1	0,2	0,2	0,3
	M8*+	0,3	0,3	0,22	0,22	++	++	0,22	0,22	0,22	0,22

** Theoretical *P*(T1e) not available.
+ See Testable Samples and Cycles Data.
++ Not enough testable samples to report results.

* For methods 3a, 4, 5 and 6. 22 initial sample sizes, 3(1)24 for method 8.

Table 4.2.5-2
*Kruskal-Wallis Test for Three to Six Groups, Type I Error Out-of-Tolerance Counts for α .05, 23 Initial Sample Sizes, 3(1)25**

Groups	*Mthd*	....Norm....		......EA......		......EB......		......ML......		......SS......	
		H	Hc	H	Hc	H	Hc	H	Hc	H	Hc
3 Grps	M3a	0,0	0,0	0,11	0,15	0,21	0,23	0,0	0,0	0,1	0,3
	M4	0,0	0,0	0,19	0,23	0,0	0,7	0,0	0,0	0,0	0,3
	M5	0,0	0,0	0,0	0,3	0,1	0,23	0,0	0,0	0,0	0,1
	M6a	0,0	0,0	0,0	0,0	0,0	0,2	0,0	0,0	0,0	0,0
	M8*+	0,0	0,0	0,22	0,22	0,22	0,22	0,21	0,22	0,22	0,22
4 Grps	M3a	0,2	0,2	0,14	0,20	0,22	0,23	0,0	0,0	0,2	0,6
	M4	0,2	0,2	0,22	0,23	0,0	0,9	0,0	0,0	0,0	0,0
	M5	0,2	0,2	0,0	0,3	0,1	0,23	0,0	0,0	0,0	0,1
	M6a	0,2	0,2	0,0	0,0	1,0	0,1	0,0	0,0	0,0	0,0
	M8*+	0,0	0,0	0,0	0,22	++	++	0,21	0,22	0,22	0,22
5 Grps	M3a	0,0	0,0	0,20	0,23	0,23	0,23	0,1	0,3	0,7	0,8
	M4	0,0	0,0	0,22	0,23	0,0	0,8	0,1	0,1	0,0	0,3
	M5	0,0	0,0	0,1	0,5	0,0	0,23	0,0	0,1	0,0	0,0
	M6a	0,0	0,0	0,0	0,1	1,0	0,0	0,0	0,0	0,0	0,0
	M8*+	0,0	0,0	0,22	0,22	++	++	0,22	0,22	0,22	0,22
6 Grps	M3a	0,0	0,0	0,17	0,22	0,23	0,23	0,2	0,2	0,11	0,13
	M4	0,0	0,0	0,23	0,23	0,0	0,7	0,0	0,1	0,1	0,4
	M5	0,0	0,0	0,1	0,9	0,1	0,23	0,0	0,1	0,0	0,1
	M6a	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0
	M8*+	0,0	0,0	0,22	0,22	++	++	0,22	0,22	0,22	0,22

Distributions** span the Norm, EA, EB, ML and SS columns.

** Theoretical *P*(T1e) not available.
\+ See Testable Samples and Cycles Data.
++ Not enough testable samples to report results.

* For methods 3a, 4, 5 and 6. 22 initial sample sizes, 3(1)24 for method 8.

For six groups at nominal alpha .01, *P*(T1e) for the Normal distribution remained at or above .005 for all initial sample sizes while *P*(T1e) exceeded .011 for initial sample sizes 15 for methods 3 through 6 and for initial sample sizes 3, 16 and 21 for method 8. All of preceding results applied equally to H and Hc as there was no difference in the statistics given the absence of ties with the Normal distribution.

For three, five and six groups at nominal alpha .05, *P*(T1e) for the Normal distribution was within the robust tolerance limits for all initial sample sizes. For four groups, *P*(T1e) for the Normal distribution remained at or above .025 for all initial sample sizes while *P*(T1e) exceeded .055 for initial sample sizes 7 and 20 for methods 3 – 6. These results applied equally to H and H_c as there was no difference in the statistics given the absence of ties with the Normal distribution.

Almost without exception, Type I error results based on H_c were inflated with respect to the results based on H. As such, test combinations were selected for further study based on the Type I error results obtained using the H statistic.

At nominal alpha .01, results for the Micceri distributions varied across method and number of groups in a way that is difficult to generalize. Results were grossly inflated for method 8 across distributions and number of groups. For distribution EB with four to six groups there were insufficient testable samples to even report results. Results were similar to slightly conservative for method 6a across distributions except for distribution EB, which had very conservative results. Method 5 was similar to somewhat inflated across distributions and number of groups. Methods 3a and 4 were similar to somewhat inflated for distributions ML and SS but extremely inflated for distributions EA and EB.

Table 4.2.5-3

Kruskal-Wallis Test for Three Groups, Method 8 – Testable Samples and Cycles for α .01 and .05, 22 Initial Sample Sizes, 3(1)24

		-------------	Distributions	-------------	
α	*Datum*	EA	EB	ML	SS
.01	ATSS Min	3.00	3.00	3.00	3.00
	at ISS	3	3	3	3
	ATSS Max	3.87	3.43	5364	4.04
	at ISS	21	15	17	13
	Cycles Min*	20k	812k	10k	15k
	at ISS	18	3	7	9
	Cycles Max*	93k	10M*	27k	49k
	at ISS	3	4	20	3
	*Testable Samples		72		
.05	ATSS Min	2.53	2.52	2.70	2.58
	at ISS	3	3	3	3
	ATSS Max	3.54	2.76	5.63	3.78
	at ISS	23	11	19	12
	Cycles Min*	14k	2.3M	10k	12k
	at ISS	19	3	7	8
	Cycles Max*	28k	10M*	27k	18k
	at ISS	3	7	22	24
	*Testable Samples		7721		

* When Cycles = 10M, testable samples < 10k.

The situation was somewhat better for nominal alpha .05. Methods 5 and 6a were similar to slightly inflated across all distributions and numbers of groups. Method 3a was similar to slightly inflated across numbers of groups for distribution ML as well as for distribution SS at three and four groups. It was quite inflated, however, for distribution SS with five and six groups. Methods 3a and 4 were grossly inflated for distributions EA and EB across numbers of groups. Method 8 was again grossly inflated across all Micceri distributions and number of groups, again with insufficient testable samples for distribution EB with four to six groups.

Table 4.2.5-4
Kruskal-Wallis Test for Four Groups, Method 8 – Testable Samples and Cycles for α .01 and .05, 22 Initial Sample Sizes, 3(1)24

		--------------	*Distributions*		-------------
α	*Datum*	EA	EB	ML	SS
.01	ATSS Min	2.36	2.25	2.59	2.42
	at ISS	3	3	3	3
	ATSS Max	2.68	2.50	3.26	2.74
	at ISS	15	9	11	8
	Cycles Min*	30k	10M*	11k	20k
	at ISS	10	3	4	6
	*Testable Samples		348		
	Cycles Max*	71k	–	80k	101k
	at ISS	3	–	15	24
.05	ATSS Min	2.12	1.94	2.45	2.20
	at ISS	3	3	3	3
	ATSS Max	2.50	2.62	3.26	2.60
	at ISS	15	11	12	8
	Cycles Min*	23k	10M*	11k	16k
	at ISS	10	3	4	5
	*Testable Samples		2		
	Cycles Max*	33k	–	81k	62k
	at ISS	3	–	14	24

* When Cycles = 10M, testable samples < 10k.

Table 4.2.5-5

Kruskal-Wallis Test for Five Groups, Method 8 – Testable Samples and Cycles for α .01 and .05, 22 Initial Sample Sizes, 3(1)24

		Distributions			
α	*Datum*	EA	EB	ML	SS
.01	ATSS Min	1.99	1.88	2.30	2.05
	at ISS	3	3	3	23
	ATSS Max	2.12	–	2.43	2.16
	at ISS	10	–	8	6
	Cycles Min*	58k	10M*	12k	30k
	at ISS	7	5	3	4
	Testable Samples		5		
	Cycles Max*	141k	–	112k	698k
	at ISS	24	–	11	24
.05	ATSS Min	1.85	1.61	2.25	1.87
	at ISS	3	3	3	23
	ATSS Max	1.99	–	2.42	2.05
	at ISS	8	–	10	6
	Cycles Min*	41k	10M*	12k	23k
	at ISS	6	3	3	4
	Testable Samples		14		
	Cycles Max*	93k	–	110k	365k
	at ISS	24	–	11	24

* When Cycles = 10M, testable samples < 10k.

Table 4.2.5-6
Kruskal-Wallis Test for Six Groups, Method 8 – Testable Samples and Cycles for α .01 and .05, 22 Initial Sample Sizes, 3(1)24

		Distributions			
α	*Datum*	EA	EB	ML	SS
.01	ATSS Min	1.80	–	1.94	1.79
	at ISS	3	–	23	21
	ATSS Max	1.85	–	2.04	1.87
	at ISS	9	–	7	5
	Cycles Min*	140k	–	22k	57k
	at ISS	5	–	3	3
	Cycles Max*	841k	–	128k	836k
	at ISS	24	–	9	24
.05	ATSS Min	1.69	–	1.88	1.65
	at ISS	3	–	24	24
	ATSS Max	1.76	–	2.03	1.79
	at ISS	10	–	8	4
	Cycles Min*	97k	–	20k	42k
	at ISS	4	–	3	3
	Cycles Max*	536k	–	126k	4.4M
	at ISS	24	–	9	24

* When Cycles = 10M, testable samples < 10k.

4.2.6 – Terpstra-Jonckheere Test (three to six groups)

Type I error results for the Terpstra-Jonckheere Test at nominal alpha .01 are summarized in Table 4.2.6-1 for three to six groups while corresponding results for nominal alpha .05 are given in Table 4.2.6-2. Testable samples results for method 8 are given in Tables 4.2.6-3 through 4.2.6-6 for three to six groups respectively.

At nominal alpha .01 *P*(T1e) was less than .005 only for initial sample size 3 for three groups and never exceeded .011 because of the strictly conservative criteria used to select critical values. For four to six groups, *P*(T1e) was within the robust tolerance limits for all initial sample sizes for which *P*(T1e) was available. *P*(T1e) was not available for five groups at initial sample sizes 8 – 10 or for six groups at initial sample sizes 7 – 10, although critical values were available and used.

At nominal alpha .05 *P*(T1e) was within the robust tolerance limits for all initial sample sizes for which *P*(T1e) was available. *P*(T1e) was not available for 5 groups at initial sample sizes 8 – 10 or for 6 groups at initial sample sizes 7 – 10, although critical values were available and used.

At nominal alpha .01 and .05 Type I error rates for the Normal distribution were very similar to *P*(T1e) for all initial sample sizes across numbers of groups. This was expected given the lack of ties with the Normal distribution. A slightly inflated departure was noted at nominal alpha .01 for method 8 with five and six groups. Some departure for method 8 was also expected due to the smaller number of repetitions involved in the method 8 simulation. These results indicate that the basic simulations worked correctly.

Type I error results for the Micceri distributions were generally very similar to *P*(T1e) and/or the Normal distribution at both nominal alpha .01 and .05 for methods 3 – 6 across number of groups with departures tending to be slightly conservative. Results for method 8 were slightly to moderately inflated for both alpha levels for all Micceri distributions across numbers of groups. There were insufficient testable samples to report results for method 8 with distribution EB for four to six groups at both alpha levels.

Table 4.2.6-1
Terpstra-Jonckheere Test for Three to Six Groups, Type I Error Out-of-Tolerance Counts for α .01, 9 Initial Sample Sizes 2(1)10, except Three Groups at 3(1)10

			Distributions				
Groups	*Mthd*	*P*(T1e)	Norm	EA	EB	ML	SS
3 Grps	M3a	1,0	1,0	1,0	2,0	1,0	1,0
	M4	1,0	1,0	0,0	1,0	1,0	1,0
	M5	1,0	1,0	1,0	1,0	1,0	1,0
	M6a	1,0	1,0	1,0	2,0	1,0	1,0
	M8+	1,0	0,0	0,7	0,8	0,5	0,6
4 Grps	M3a	0,0	0,0	1,0	2,0	1,0	1,0
	M4	0,0	0,0	0,0	1,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	1,0	2,0	1,0	1,0
	M8+	0,0	0,0	0,7	++	1,6	0,2
5 Grps*	M3a	0,0	0,0	0,0	2,0	0,0	0,0
	M4	0,0	0,0	0,0	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	0,0	2,0	0,0	0,0
	M8+	0,0	0,1	0,9	++	0,6	0,6
6 Grps*	M3a	0,0	0,0	0,0	1,0	0,0	0,0
	M4	0,0	0,0	0,4	0,1	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	0,0	1,0	0,0	0,0
	M8+	0,0	0,3	0,9	++	0,6	0,7

* *P*(T1e) not available for all critical values.
\+ See Testable Samples and Cycles Data.
++ Not enough testable samples to report results.

Table 4.2.6-2
Terpstra-Jonckheere Test for Three to Six Groups, Type I Error Out-of-Tolerance Counts for α .05, 9 Initial Sample Sizes 2(1)10

			Distributions				
Groups	*Mthd*	*P*(T1e)	Norm	EA	EB	ML	SS
3 Grps	M3a	0,0	0,0	1,0	2,0	0,0	0,0
	M4	0,0	0,0	0,0	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	1,0	2,0	0,0	0,0
	M8+	0,0	0,0	0,7	0,9	0,4	0,5
4 Grps	M3a	0,0	0,0	0,0	0,0	0,0	0,0
	M4	0,0	0,0	0,0	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	0,0	0,0	0,0	0,0
	M8+	0,0	0,0	0,9	++	0,5	0,3
5 Grps*	M3a	0,0	0,0	0,0	0,0	0,0	0,0
	M4	0,0	0,0	0,0	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	0,0	0,0	0,0	0,0
	M8+	0,0	0,0	0,9	++	0,3	0,2
6 Grps*	M3a	0,0	0,0	0,0	0,0	0,0	0,0
	M4	0,0	0,0	0,1	0,0	0,0	0,0
	M5	0,0	0,0	0,0	0,0	0,0	0,0
	M6a	0,0	0,0	0,0	0,0	0,0	0,0
	M8+	0,0	0,0	0,8	++	0,1	0,2

* *P*(T1e) not available for all critical values.
\+ See Testable Samples and Cycles Data.
++ Not enough testable samples to report results.

Table 4.2.6-3

Terpstra-Jonckheere Test for Three Groups, Method 8 – Testable Samples and Cycles for α .01 and .05, 8 Initial Sample Sizes 3(1)10 at .01, 9 at .05 2(1)10

		-------------	Distributions	-------------	
α	*Datum*	EA	EB	ML	SS
.01	ATSS Min	2.53	2.52	2.70	2.57
	at ISS	3	3	3	3
	ATSS Max	3.50	2.77	5.82	3.90
	at ISS	10	10	10	10
	Cycles Min*	20k	2M	10k	14k
	at ISS	10	3	9	7
	Cycles Max*	28k	10M	12k	18k
	at ISS	3	6	3	3
.05	ATSS Min	2.00	2.00	2.00	2.00
	at ISS	2	2	2	2
	ATSS Max	3.47	2.63	5.81	3.88
	at ISS	10	10	10	10
	Cycles Min*	19k	360k	14k	20k
	at ISS	8	2	2	2
	Cycles Max*	28k	10M	10k	14k
	at ISS	2	6	10	8

* When Cycles = 10M, testable samples < 10k.

Table 4.2.6-4
Terpstra-Jonckheere Test for Four Groups, Method 8 – Testable Samples and Cycles for α .01 and .05, 9 Initial Sample Sizes 2(1)10

		Distributions			
α	*Datum*	EA	EB	ML	SS
.01	ATSS Min	2.00	2.00	2.00	2.00
	at ISS	2	2	2	2
	ATSS Max	2.30	2.23	4.48	2.54
	at ISS	3	3	10	4
	Cycles Min*	81k	10M	20k	42k
	at ISS	2	2	2	2
	Cycles Max*	1M	(1)	1M	1M
	at ISS	10	–	10	10
.05	ATSS Min	2.00	2.00	2.00	2.00
	at ISS	2	2	2	2
	ATSS Max	2.30	2.21	4.48	2.55
	at ISS	3	3	10	4
	Cycles Min*	81k	10M	20k	43k
	at ISS	2	2	2	2
	Cycles Max*	1M	(2)	1M	1M
	at ISS	10	–	10	10

* When Cycles = 10M, testable samples < 10k.
(1) Only 5151 testable samples obtained at iss 2, 90 at iss 3, 18 at iss 4, 2 at iss 5, and none for iss 6 – 10.
(2) Only 5242 testable samples obtained at iss 2, 129 at iss 3, 14 at iss 4, 1 at iss 5, and none for iss 6 – 10.

Table 4.2.6-5

Terpstra-Jonckheere Test for Five Groups, Method 8 – Testable Samples and Cycles for α .01 and .05, 9 Initial Sample Sizes 2(1)10

		Distributions			
α	*Datum*	EA	EB	ML	SS
.01	ATSS Min	2.00	2.00	2.00	2.00
	at ISS	2	2	2	2
	ATSS Max	2.14	–	3.40	2.31
	at ISS	3	–	9	3
	Cycles Min*	320k	10M	33k	120k
	at ISS	2	2	2	2
	Cycles Max*	9M	(1)	6M	10M
	at ISS	10	–	10	10
.05	ATSS Min	2.00	2.00	2.00	2.00
	at ISS	2	2	2	2
	ATSS Max	2.14	–	3.39	2.32
	at ISS	3	–	9	3
	Cycles Min*	320k	10M	33k	120k
	at ISS	2	2	2	2
	Cycles Max*	9M	(2)	6M	10M
	at ISS	10	–	10	10

* When Cycles = 10M, testable samples < 10k.

(1) Only 18 testable samples obtained at iss 2, none for iss 3 – 10.

(2) Only 22 testable samples obtained at iss 2, none for iss 3 – 10.

Table 4.2.6-6
Terpstra-Jonckheere Test for Six Groups, Method 8 – Testable Samples and Cycles for α .01 and .05, 9 Initial Sample Sizes 2(1)10

		Distributions			
α	*Datum*	EA	EB(1)	ML	SS
.01	ATSS Min	2.00	–	2.00	2.00
	at ISS	2	–	2	2
	ATSS Max	2.03	–	2.71	2.07
	at ISS	3	–	7	3
	Cycles Min*	2M	–	61k	470k
	at ISS	2	–	2	2
	Cycles Max*	10M	–	10M	10M
	at ISS	3	–	7	4
.05	ATSS Min	2.00	–	2.00	2.00
	at ISS	2	–	2	2
	ATSS Max	2.03	–	3.88	2.07
	at ISS	3	–	10	3
	Cycles Min*	2M	–	61k	470k
	at ISS	2	–	2	2
	Cycles Max*	10M	–	10M	10M
	at ISS	3	–	7	4

* When Cycles = 10M, testable samples < 10k.
(1) No testable samples obtained at any initial sample size.

4.3 – Power and Type III Error Results

4.3.1 – Kolmogorov-Smirnov Test (two groups only)

Based on the Type I error results, power and Type III error results for the Kolmogorov-Smirnov test are only presented for the combinations of alpha level, distribution and method shown in Table 4.3.1-1.

Table 4.3.1-1
Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Method / Distribution Combinations with Acceptable Type I Error

Alpha	.01					.05				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Method										
M3a	x	x		x	x	x	x		x	x
M4	x	x		x	x	x	x		x	x
M5	x	x	x	x	x	x	x	x	x	x
M6b	x			x		x			x	

Power results for the Kolmogorov-Smirnov test at nominal alpha .01 and .05 are summarized in Table 4.3.1-2 and 4.3.1-3. These tables present the range of values obtained (minimum and maximum) for 1-tailed power, 2-tailed Type III error and 2-tailed power based on the results presented in Tables 4.3.1-4 through 4.3.1-23.

Power tended to increase monotonically across methods and distributions with increases in initial sample size and/or effect size. Likewise, Type III error tended to decrease monotonically across methods and distributions with an increase in initial sample size and/or effect size. Thus, minimum power (or maximum Type III error) usually occurred at initial sample size 6 or 12 at nominal ESM 0.2, with maximum power (or minimum Type III error) at initial sample size 30 and nominal ESM 1.2, if not sooner.

Table 4.3.1-2
Kolmogorov-Smirnov Test of General Differences for Two Groups
Range of Power and Type III Error for α .01

Min/Max	min					max				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Type										
Mthd										
Pwr 1-s										
M3a	.0030	.0019		.0017	.0007	.9351	.9999		.9061	.9346
M4	.0026	.0030		.0020	.0025	.9351	.9998		.9156	.9214
M5	.0026	.0031	.0094	.0020	.0026	.9351	.9999	.9778	.9177	.9326
M6b	.0026			.0026		.9351			.9018	
T3e 2-s										
M3a	.0000	.0000		.0000	.0000	.0011	.0028		.0014	.0007
M4	.0000	.0000		.0000	.0000	.0011	.0029		.0016	.0010
M5	.0000	.0000	.0000	.0000	.0000	.0011	.0011	.0005	.0017	.0011
M6b	.0000			.0000		.0011			.0014	
Pwr 2-s										
M3a	.0030	.0019		.0017	.0007	.8798	.9996		.8222	.8786
M4	.0026	.0030		.0020	.0025	.8798	.9991		.8372	.8597
M5	.0026	.0031	.0094	.0020	.0026	.8798	.9994	.9386	.8404	.8765
M6b	.0026			.0026		.8798			.8161	

Tables 4.3.1-4 through 4.3.1-23 give the upper and lower tail results for the Kolmogorov-Smirnov test for both 1-sided and 2-sided tests for both alpha .01 and .05. There is a table for each combination of nominal effect size multiplier {0.2, 0.5, 0.8, 1.2} and initial sample size {6, 12, 18, 24, 30}. Results are not reported for distribution EB with nominal ESM 0.2 as the actual ESM = 0.0 (no shift), which is just Type I error. Also for distribution EB, results are not reported for nominal ESM 0.8 as the actual ESM = 0.592, the same as the actual ESM for nominal ESM 0.5. Results for the normal distribution are included and are essentially identical across methods for a fixed initial sample size, effect size

and directionality. This is to be expected in the absence of ties and demonstrates that the simulations worked correctly.

Table 4.3.1-3
Kolmogorov-Smirnov Test of General Differences for Two Groups Range of Power and Type III Error for α .05

Min/Max	min					max				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Type										
Mthd										
Pwr 1-s										
M3a	.0261	.0215		.0201	.0192	.9862	.9999		.9840	.9862
M4	.0261	.0301		.0229	.0253	.9862	.9999		.9863	.9820
M5	.0261	.0309	.0830	.0230	.0260	.9862	.9999	.9985	.9869	.9853
M6b	.0261			.0201		.9862			.9828	
T3e 2-s										
M3a	.0000	.0000		.0000	.0000	.0061	.0040		.0066	.0041
M4	.0000	.0000		.0000	.0000	.0061	.0056		.0074	.0057
M5	.0000	.0000	.0001	.0000	.0000	.0061	.0056	.0030	.0075	.0059
M6b	.0000			.0000		.0061			.0066	
Pwr 2-s										
M3a	.0261	.0215		.0201	.0192	.9684	.9999		.9575	.9681
M4	.0261	.0301		.0229	.0253	.9684	.9999		.9627	.9601
M5	.0261	.0309	.0830	.0230	.0260	.9684	.9999	.9936	.9641	.9667
M6b	.0261			.0201		.9684			.9549	

A ranking analysis of the results is presented in Tables 4.3.1-24 through 4.3.1-35. These summaries were obtained by ranking the power results from Tables 4.3.1-3 through 4.3.1-22 to four decimal places (as reported), to three decimal places and to two decimal places.

Table 4.3.1-4

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0004	.0003		.0005	.0007	.0061	.0040		.0066	.0041
	M4	.0004	.0004		.0006	.0004	.0061	.0056		.0074	.0057
	M5	.0004	.0004		.0006	.0004	.0061	.0056		.0075	.0059
	M6b	.0004			.0004		.0061			.0066	
UT1s	M3a	.0030	.0019		.0017	.0007	.0261	.0215		.0201	.0192
(Pwr)	M4	.0026	.0030		.0020	.0025	.0261	.0301		.0229	.0253
	M5	.0026	.0031		.0020	.0026	.0261	.0309		.0230	.0260
	M6b	.0026			.0026		.0261			.0201	
LT2s	M3a	.0004	.0003		.0005	.0007	.0061	.0040		.0066	.0041
(T3e)	M4	.0004	.0004		.0006	.0004	.0061	.0056		.0074	.0057
	M5	.0004	.0004		.0006	.0004	.0061	.0056		.0075	.0059
	M6b	.0004			.0004		.0061			.0066	
UT2s	M3a	.0030	.0019		.0017	.0007	.0261	.0215		.0201	.0192
(Pwr)	M4	.0026	.0030		.0020	.0025	.0261	.0301		.0229	.0253
	M5	.0026	.0031		.0020	.0026	.0261	.0309		.0230	.0260
	M6b	.0026			.0026		.0261			.0201	

Table 4.3.1-5

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0011	.0028		.0014	.0007	.0197	.0147		.0226	.0143
	M4	.0011	.0011		.0016	.0010	.0197	.0181		.0255	.0179
	M5	.0011	.0011		.0017	.0011	.0197	.0191		.0258	.0195
	M6b	.0011			.0014		.0197			.0224	
UT1s	M3a	.0120	.0028		.0081	.0085	.1092	.1212		.0868	.0880
(Pwr)	M4	.0120	.0142		.0010	.0110	.1092	.1290		.0980	.1031
	M5	.0120	.0158		.0100	.0120	.1092	.1417		.0999	.1094
	M6b	.0120			.0080		.1092			.0857	
LT2s	M3a	.0011	.0028		.0014	.0007	.0052	.0038		.0064	.0036
(T3e)	M4	.0011	.0011		.0016	.0010	.0052	.0049		.0074	.0047
	M5	.0011	.0011		.0017	.0011	.0052	.0052		.0075	.0052
	M6b	.0011			.0014		.0052			.0064	
UT2s	M3a	.0120	.0028		.0081	.0085	.0408	.0425		.0298	.0308
(Pwr)	M4	.0120	.0142		.0010	.0110	.0408	.0486		.0347	.0379
	M5	.0120	.0158		.0100	.0120	.0408	.0553		.0356	.0407
	M6b	.0120			.0080		.0408			.0201	

Table 4.3.1-6

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0007	.0005		.0011	.0005	.0077	.0063		.0103	.0061
	M4	.0007	.0029		.0012	.0006	.0077	.0071		.0017	.0067
	M5	.0007	.0007		.0013	.0007	.0077	.0078		.0119	.0077
	M6b	.0007			.0010		.0077			.0101	
UT1s	M3a	.0133	.0179		.0086	.0108	.0805	.1135		.0605	.0694
(Pwr)	M4	.0133	.0028		.0101	.0119	.0805	.0983		.0689	.0729
	M5	.0133	.0196		.0106	.0136	.0805	.1154		.0708	.0805
	M6b	.0133			.0082		.0805			.0580	
LT2s	M3a	.0007	.0005		.0011	.0005	.0024	.0019		.0035	.0018
(T3e)	M4	.0007	.0029		.0012	.0006	.0024	.0022		.0040	.0020
	M5	.0007	.0007		.0013	.0007	.0024	.0024		.0041	.0023
	M6b	.0007			.0010		.0024			.0034	
UT2s	M3a	.0133	.0179		.0086	.0108	.0354	.0487		.0244	.0294
(Pwr)	M4	.0133	.0028		.0101	.0119	.0354	.0434		.0284	.0315
	M5	.0133	.0196		.0106	.0136	.0354	.0514		.0296	.0353
	M6b	.0133			.0082		.0354			.0234	

Table 4.3.1-7

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0010	.0009		.0016	.0009	.0078	.0074		.0111	.0071
	M4	.0010	.0009		.0019	.0008	.0078	.0072		.0126	.0066
	M5	.0010	.0011		.0020	.0010	.0078	.0081		.0130	.0079
	M6b	.0010			.0015		.0078			.0107	
UT1s	M3a	.0265	.0466		.0182	.0245	.1076	.1868		.0846	.1029
(Pwr)	M4	.0265	.0335		.0206	.0229	.1076	.1342		.0932	.0959
	M5	.0265	.0423		.0217	.0268	.1076	.1666		.0973	.1082
	M6b	.0265			.0166		.1076			.0784	
LT2s	M3a	.0003	.0003		.0005	.0003	.0029	.0027		.0045	.0026
(T3e)	M4	.0003	.0003		.0006	.0002	.0029	.0027		.0052	.0024
	M5	.0003	.0004		.0006	.0003	.0029	.0031		.0053	.0030
	M6b	.0003			.0005		.0029			.0044	
UT2s	M3a	.0114	.0197		.0073	.0103	.0558	.0988		.0411	.0531
(Pwr)	M4	.0114	.0142		.0084	.0096	.0058	.0704		.0459	.0492
	M5	.0114	.0182		.0089	.0113	.0558	.0882		.0482	.0566
	M6b	.0114			.0066		.0558			.0377	

Table 4.3.1-8

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0011	.0011		.0020	.0011	.0068	.0071		.0106	.0068
	M4	.0011	.0010		.0023	.0008	.0068	.0064		.0118	.0055
	M5	.0011	.0012		.0024	.0010	.0068	.0074		.0123	.0068
	M6b	.0011			.0020		.0068			.0101	
UT1s	M3a	.0388	.0848		.0280	.0392	.1262	.2575		.1023	.1272
(Pwr)	M4	.0388	.0505		.0303	.0325	.1262	.1606		.1081	.1089
	M5	.0388	.0673		.0322	.0391	.1262	.2083		.1137	.1254
	M6b	.0388			.0245		.1262			.0907	
LT2s	M3a	.0004	.0004		.0008	.0004	.0028	.0030		.0047	.0027
(T3e)	M4	.0004	.0004		.0010	.0003	.0028	.0027		.0053	.0022
	M5	.0004	.0005		.0010	.0004	.0028	.0031		.0055	.0027
	M6b	.0004			.0008		.0028			.0045	
UT2s	M3a	.0191	.0424		.0129	.0193	.0728	.1542		.0557	.0732
(Pwr)	M4	.0191	.0249		.0141	.0158	.0728	.0935		.0595	.0615
	M5	.0191	.0336		.0151	.0193	.0728	.1236		.0630	.0726
	M6b	.0191			.0111		.0728			.0488	

Table 4.3.1-9

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0001	.0000		.0002	.0001	.0017	.0010		.0025	.0017
	M4	.0001	.0001		.0002	.0001	.0017	.0013		.0027	.0024
	M5	.0001	.0001	.0002	.0002	.0001	.0017	.0013	.0030	.0028	.0026
	M6b	.0001			.0002		.0017			.0025	
UT1s	M3a	.0082	.0138		.0065	.0040	.0639	.1047		.0567	.0370
(Pwr)	M4	.0082	.0189		.0074	.0055	.0639	.1270		.0626	.0472
	M5	.0082	.0200	.0094	.0074	.0056	.0639	.1330	.0830	.0628	.0484
	M6b	.0082			.0065		.0639			.0566	
LT2s	M3a	.0001	.0000		.0002	.0001	.0017	.0010		.0025	.0017
(T3e)	M4	.0001	.0001		.0002	.0001	.0017	.0013		.0027	.0024
	M5	.0001	.0001	.0002	.0002	.0001	.0017	.0013	.0030	.0028	.0026
	M6b	.0001			.0002		.0017			.0025	
UT2s	M3a	.0082	.0138		.0065	.0040	.0639	.1047		.0567	.0370
(Pwr)	M4	.0082	.0189		.0074	.0055	.0639	.1270		.0626	.0472
	M5	.0082	.0200	.0094	.0074	.0056	.0639	.1330	.0830	.0628	.0484
	M6b	.0082			.0065		.0639			.0566	

Table 4.3.1-10
Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0001	.0001		.0004	.0001	.0036	.0025		.0069	.0045
	M4	.0001	.0001		.0004	.0002	.0036	.0034		.0078	.0059
	M5	.0001	.0001	.0005	.0004	.0003	.0036	.0033	.0092	.0079	.0065
	M6b	.0001			.0004		.0036			.0069	
UT1s	M3a	.0488	.1429		.0439	.0237	.2721	.5494		.2729	.1743
(Pwr)	M4	.0488	.1490		.0502	.0298	.2721	.5420		.2964	.1985
	M5	.0488	.1679	.0811	.0510	.0319	.2721	.5773	.4481	.2987	.2082
	M6b	.0488			.0436		.2721			.2714	
LT2s	M3a	.0001	.0001		.0004	.0001	.0008	.0005		.0018	.0009
(T3e)	M4	.0001	.0001		.0004	.0002	.0008	.0007		.0020	.0013
	M5	.0001	.0001	.0005	.0004	.0003	.0008	.0007	.0025	.0020	.0014
	M6b	.0001			.0004		.0008			.0018	
UT2s	M3a	.0488	.1429		.0439	.0237	.1293	.3184		.1240	.0726
(Pwr)	M4	.0488	.1490		.0502	.0298	.1293	.3200		.1383	.0866
	M5	.0488	.1679	.0811	.0510	.0319	.1293	.3505	.2195	.1397	.0915
	M6b	.0488			.0436		.1293			.1232	

Table 4.3.1-11

Kolmogorov-Smirnov Test of General Differences for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 0.5, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0001	.0000		.0002	.0001	.0007	.0006		.0023	.0013
	M4	.0001	.0000		.0002	.0001	.0007	.0008		.0027	.0015
	M5	.0001	.0001	.0003	.0002	.0001	.0007	.0009	.0036	.0027	.0017
	M6b	.0001			.0002		.0007			.0023	
UT1s	M3a	.0704	.3034		.0689	.0356	.2608	.6785		.2720	.1656
(Pwr)	M4	.0704	.2634		.0785	.0384	.2608	.6203		.2961	.1722
	M5	.0704	.3058	.1486	.0801	.0427	.2608	.6695	.5037	.3004	.1857
	M6b	.0704			.0673		.2608			.2672	
LT2s	M3a	.0001	.0000		.0002	.0001	.0002	.0001		.0007	.0003
(T3e)	M4	.0001	.0000		.0002	.0001	.0002	.0002		.0008	.0004
	M5	.0001	.0001	.0003	.0002	.0001	.0002	.0002	.0012	.0008	.0004
	M6b	.0001			.0002		.0007			.0007	
UT2s	M3a	.0704	.3034		.0689	.0356	.1452	.4892		.1475	.0823
(Pwr)	M4	.0704	.2634		.0785	.0384	.1452	.4341		.1642	.0871
	M5	.0704	.3058	.1486	.0801	.0427	.1452	.4857	.2994	.1669	.0956
	M6b	.0704			.0673		.1293			.1445	

Table 4.3.1-12

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0002	.0001	.0005	.0005		.0021	.0012
	M4	.0000	.0000		.0003	.0001	.0005	.0006		.0025	.0011
	M5	.0000	.0001	.0005	.0003	.0001	.0005	.0007	.0037	.0025	.0014
	M6b	.0000			.0002		.0005			.0021	
UT1s	M3a	.1438	.6135		.1561	.0835	.3642	.8740		.4020	.2510
(Pwr)	M4	.1438	.5142		.1715	.0784	.3642	.8057		.4277	.2382
	M5	.1438	.5781	.3495	.1760	.0886	.3642	.8471	.7157	.4342	.2595
	M6b	.1438			.1496		.3642			.3902	
LT2s	M3a	.0000	.0000		.0001	.0000	.0001	.0001		.0008	.0003
(T3e)	M4	.0000	.0000		.0001	.0000	.0001	.0002		.0009	.0003
	M5	.0000	.0000	.0002	.0001	.0000	.0001	.0002	.0014	.0009	.0004
	M6b	.0000			.0001		.0001			.0007	
UT2s	M3a	.0782	.4466		.0832	.0414	.2399	.7628		.2633	.1519
(Pwr)	M4	.0782	.3552		.0933	.0390	.2399	.6724		.2849	.1431
	M5	.0782	.4161	.2002	.0960	.0448	.2399	.7286	.5335	.2905	.1593
	M6b	.0782			.0792		.2399			.2538	

Table 4.3.1-13

Kolmogorov-Smirnov Test of General Differences for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 0.5, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0002	.0001	.0003	.0004		.0017	.0008
	M4	.0000	.0000		.0003	.0001	.0003	.0004		.0019	.0006
	M5	.0000	.0001	.0005	.0003	.0001	.0003	.0005	.0032	.0020	.0008
	M6b	.0000			.0002		.0003			.0016	
UT1s	M3a	.2153	.8218		.2531	.1348	.4413	.9557		.5087	.3191
(Pwr)	M4	.2153	.7127		.2694	.1164	.4413	.9062		.5286	.2852
	M5	.2153	.7715	.5516	.2764	.1331	.4413	.9326	.8455	.5377	.3151
	M6b	.2153			.2381		.4413			.4880	
LT2s	M3a	.0000	.0000		.0001	.0000	.0001	.0001		.0007	.0003
(T3e)	M4	.0000	.0000		.0001	.0000	.0001	.0001		.0008	.0002
	M5	.0000	.0000	.0002	.0001	.0000	.0001	.0001	.0013	.0008	.0003
	M6b	.0000			.0001		.0001			.0006	
UT2s	M3a	.1355	.7049		.1591	.0783	.3196	.9049		.3728	.2149
(Pwr)	M4	.1355	.5762		.1717	.0664	.3196	.8254		.3921	.1887
	M5	.1355	.6459	.3845	.1769	.0776	.3196	.8676	.7147	.4003	.2122
	M6b	.1355			.1484		.3196			.3542	

Table 4.3.1-14

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0004	.0002		.0009	.0002
	M4	.0000	.0000		.0000	.0000	.0004	.0003		.0010	.0003
	M5	.0000	.0000		.0000	.0000	.0004	.0003		.0010	.0003
	M6b	.0000			.0000		.0004			.0009	
UT1s	M3a	.0218	.0558		.0170	.0162	.1351	.2706		.1193	.1090
(Pwr)	M4	.0218	.2699		.0188	.0212	.1351	.2966		.1281	.1314
	M5	.0218	.0687		.0188	.0216	.1351	.3057		.1283	.1341
	M6b	.0218			.0170		.1351			.1193	
LT2s	M3a	.0000	.0000		.0000	.0000	.0004	.0002		.0009	.0002
(T3e)	M4	.0000	.0000		.0000	.0000	.0004	.0003		.0010	.0003
	M5	.0000	.0000		.0000	.0000	.0004	.0003		.0010	.0003
	M6b	.0000			.0000		.0004			.0009	
UT2s	M3a	.0218	.0558		.0170	.0162	.1351	.2706		.1193	.1090
(Pwr)	M4	.0218	.2699		.0188	.0212	.1351	.2966		.1281	.1314
	M5	.2018	.0687		.0188	.0216	.1351	.3057		.1283	.1341
	M6b	.0218			.0170		.1351			.1193	

Table 4.3.1-15

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0001	.0000	.0005	.0004		.0020	.0003
	M4	.0000	.0000		.0001	.0000	.0005	.0006		.0024	.0004
	M5	.0000	.0000		.0001	.0000	.0005	.0007		.0024	.0005
	M6b	.0000			.0001		.0005			.0020	
UT1s	M3a	.1443	.4585		.1354	.1172	.5080	.8635		.5523	.4558
(Pwr)	M4	.1443	.4635		.1481	.1383	.5080	.8609		.5468	.4930
	M5	.1443	.4892		.1489	.1445	.5080	.8754		.5489	.5055
	M6b	.1443			.1350		.5080			.5212	
LT2s	M3a	.0000	.0000		.0001	.0000	.0001	.0001		.0004	.0000
(T3e)	M4	.0000	.0000		.0001	.0000	.0001	.0001		.0005	.0001
	M5	.0000	.0000		.0001	.0000	.0001	.0001		.0005	.0001
	M6b	.0000			.0001		.0001			.0004	
UT2s	M3a	.1443	.4585		.1354	.1172	.3015	.6911		.2993	.2588
(Pwr)	M4	.1443	.4635		.1481	.1383	.3015	.6909		.3202	.2905
	M5	.1443	.4892		.1489	.1445	.3015	.7140		.3217	.3010
	M6b	.1443			.1350		.3015			.2985	

Table 4.3.1-16

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0001	.0001		.0005	.0000
	M4	.0000	.0000		.0000	.0000	.0001	.0001		.0005	.0000
	M5	.0000	.0000		.0000	.0000	.0001	.0001		.0006	.0001
	M6b	.0000			.0000		.0001			.0004	
UT1s	M3a	.2319	.7568		.2377	.2064	.5502	.9524		.5859	.5166
(Pwr)	M4	.2319	.7288		.2577	.2186	.5502	.9407		.6118	.5281
	M5	.2319	.7578		.2596	.2324	.5502	.9510		.6148	.5476
	M6b	.2319			.2357		.5502			.5827	
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0001		.0001	.0000
(T3e)	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
	M6b	.0000			.0000		.0001			.0001	
UT2s	M3a	.2319	.7568		.2377	.2064	.3804	.8816		.4003	.3489
(Pwr)	M4	.2319	.7288		.2577	.2186	.3804	.8609		.4253	.3618
	M5	.2319	.7587		.2596	.2324	.3804	.8806		.4282	.3798
	M6b	.2319			.2357		.3804			.3974	

Table 4.3.1-17

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0001		.0003	.0000
	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0004	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0001		.0004	.0000
	M6b	.0000			.0000		.0000			.0003	
UT1s	M3a	.4234	.9276		.4653	.4107	.7070	.9946		.7682	.6966
(Pwr)	M4	.4234	.9064		.4910	.3977	.7070	.9908		.7881	.6812
	M5	.4234	.9487		.4947	.4383	.7070	.9931		.7917	.7051
	M6b	.4234			.4594		.7070			.7627	
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0001		.0001	.0000
(T3e)	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
	M6b	.0000			.0000		.0000			.0001	
UT2s	M3a	.2882	.8432		.3152	.2766	.5691	.9832		.6247	.5565
(Pwr)	M4	.2882	.8098		.3378	.2682	.5691	.9736		.6492	.5414
	M5	.2882	.8888		.3415	.4383	.5691	.9796		.6529	.5682
	M6b	.2882			.3101		.5691			.6185	

Table 4.3.1-18

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0002	.0000
	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0003	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0003	.0000
	M6b	.0000			.0000		.0000			.0002	
UT1s	M3a	.5776	.9939		.6524	.5819	.8034	.9995		.8717	.8064
(Pwr)	M4	.5776	.9882		.6741	.5453	.8034	.9987		.8841	.7768
	M5	.5776	.9913		.6788	.5777	.8034	.9991		.8871	.8012
	M6b	.5776			.6425		.8034			.8692	
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0002	.0000
(T3e)	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
	M6b	.0000			.0000		.0000			.0001	
UT2s	M3a	.4487	.9836		.5139	.4524	.6996	.9981		.7721	.7034
(Pwr)	M4	.4487	.9709		.5372	.4184	.6996	.9958		.7932	.6688
	M5	.4487	.9777		.5420	.4496	.6996	.9970		.7971	.6985
	M6b	.4487			.5038		.6996			.7650	

Table 4.3.1-19

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0001		.0003	.0000
	M4	.0000	.0000		.0000	.0000	.0000	.0001		.0004	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0009	.0004	.0000
	M6b	.0000			.0000		.0000			.0003	
UT1s	M3a	.0646	.1300		.0364	.0507	.2883	.4529		.2064	.2462
(Pwr)	M4	.0646	.1436		.0396	.0627	.2883	.4764		.2185	.2825
	M5	.0646	.1478	.0335	.0397	.0638	.2883	.4856	.2103	.2190	.2858
	M6b	.0646			.0364		.2883			.2064	
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0001		.0003	.0000
(T3e)	M4	.0000	.0000		.0000	.0000	.0000	.0001		.0004	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0009	.0004	.0000
	M6b	.0000			.0000		.0000			.0003	
UT2s	M3a	.0646	.1300		.0364	.0507	.2883	.4529		.2064	.2462
(Pwr)	M4	.0646	.1436		.0396	.0627	.2883	.4764		.2185	.2825
	M5	.0646	.1478	.0035	.0397	.0638	.2883	.4856	.2103	.2190	.2858
	M6b	.0646			.0364		.2883			.2064	

Table 4.3.1-20
Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0001		.0004	.0000
	M4	.0000	.0000		.0000	.0000	.0000	.0001		.0005	.0001
	M5	.0000	.0000	.0001	.0000	.0000	.0000	.0002	.0020	.0005	.0000
	M6b	.0000			.0000		.0000			.0004	
UT1s	M3a	.3928	.7331		.2795	.3402	.8030	.9665		.7406	.7629
(Pwr)	M4	.3928	.7363		.2985	.3792	.8030	.9659		.7598	.7920
	M5	.3928	.7530	.1654	.2995	.3892	.8030	.9699	.8189	.7610	.8005
	M6b	.3928			.2789		.8030			.7396	
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
(T3e)	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
	M5	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0005	.0001	.0000
	M6b	.0000			.0000		.0000			.0001	
UT2s	M3a	.3928	.7331		.2795	.3402	.6153	.8907		.5100	.5616
(Pwr)	M4	.3928	.7363		.2985	.3792	.6153	.8911		.5327	.6007
	M5	.3928	.7350	.1654	.2995	.3892	.6153	.9007	.5842	.5344	.6114
	M6b	.3928			.2789		.6153			.5090	

Table 4.3.1-21

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0005	.0001	.0000
	M6b	.0000			.0000	.0000	.0000			.0001	
UT1s	M3a	.5993	.9451		.4757	.5597	.8742	.9950		.8262	.8534
(Pwr)	M4	.5993	.9375		.5001	.5782	.8742	.9937		.8426	.8614
	M5	.5993	.9458	.5881	.5027	.5953	.8742	.9949	.9152	.8447	.8717
	M6b	.5993			.4732		.8742			.8242	
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000
	M6b	.0000			.0000		.0000			.0001	
UT2s	M3a	.5993	.9451		.4757	.5597	.7580	.9816		.6664	.7261
(Pwr)	M4	.5993	.9375		.5001	.5782	.7580	.9781		.6891	.7388
	M5	.5993	.9458	.5881	.5027	.5953	.7580	.9816	.7861	.6919	.7535
	M6b	.5993			.4732		.7580			.6638	

Table 4.3.1-22

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0000	.0000
	M6b	.0000			.0000		.0000			.0000	
UT1s	M3a	.8341	.9971		.7611	.8216	.9592	.9998		.9461	.9553
(Pwr)	M4	.8341	.9957		.7802	.8138	.9592	.9997		.9527	.9511
	M5	.8341	.9966	.8867	.7833	.8303	.9592	.9998	.9880	.9540	.9575
	M6b	.8341			.7567		.9592			.9441	
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000
	M6b	.0000			.0000		.0000			.0000	
UT2s	M3a	.7215	.9903		.6125	.7040	.9122	.9992		.8749	.9051
(Pwr)	M4	.7215	.9863		.6359	.6962	.9122	.9988		.8879	.8986
	M5	.7215	.9871	.7596	.6392	.7169	.9122	.9991	.9577	.8899	.9097
	M6b	.7215			.6078		.9122			.8716	

Table 4.3.1-23

Kolmogorov-Smirnov Test of General Differences for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0000	.0000
	M6b	.0000			.0000		.0000			.0000	
UT1s	M3a	.9351	.9999		.9061	.9346	.9862	.9999		.9840	.9862
(Pwr)	M4	.9351	.9998		.9156	.9214	.9862	.9999		.9863	.9820
	M5	.9351	.9999	.9778	.9177	.9326	.9862	.9999	.9985	.9869	.9853
	M6b	.9351			.9018		.9862			.9828	
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M4	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000
	M6b	.0000			.0000		.0000			.0000	
UT2s	M3a	.8798	.9996		.8222	.8786	.9684	.9999		.9575	.9681
(Pwr)	M4	.8798	.9991		.8372	.8597	.9684	.9999		.9627	.9601
	M5	.8798	.9994	.9386	.8404	.8765	.9684	.9999	.9936	.9641	.9667
	M6b	.8798			.8161		.9684			.9549	

Table 4.3.1-24
Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .01, 1-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	2.5		3.6	3.0	3.0		3.1	3.0	2.8		2.5	2.5
	M4	1.8		1.9	2.0	1.9		2.1	1.8	1.6		2.5	1.8
	M5	1.8		1.6	1.0	1.1		2.1	1.3	1.6		2.5	1.8
	M6b			2.9				2.6				2.5	
12	M3a	3.0		2.8	3.0	3.0		3.1	3.0	2.9		3.3	2.8
	M3	2.0		2.5	2.0	2.0		2.5	2.0	2.1		2.1	1.8
	M5	1.0		1.0	1.0	1.0		1.0	1.0	1.0		1.6	1.8
	M6b			3.8				3.4				3.0	
18	M3a	2.3		3.0	3.0	2.0		3.0	3.0	1.6		3.1	2.5
	M4	2.8		2.0	2.0	3.0		2.0	2.0	3.0		1.8	2.0
	M5	1.0		1.0	1.0	1.0		1.0	1.0	1.4		1.8	1.5
	M6b			4.0				4.0				3.4	
24	M3a	1.3		3.0	2.0	1.4		3.0	2.0	1.5		3.0	2.1
	M4	3.0		2.0	3.0	3.0		2.0	3.0	2.8		2.0	2.9
	M5	1.8		1.0	1.0	1.6		1.0	1.0	1.8		1.5	1.0
	M6b			4.0				1.0				3.5	
30	M3a	1.1		3.0	1.0	1.3		3.0	1.1	1.5		2.8	1.5
	M4	3.0		2.0	3.0	2.8		2.0	3.0	2.5		1.9	3.0
	M5	1.9		1.0	2.0	2.0		1.0	1.9	2.0		1.4	1.5
	M6b			4.0				4.0				4.0	

Table 4.3.1-25

Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .01, 2-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	3.0		3.6	3.0	3.0		2.9	3.0	2.8		2.5	2.5
	M4	1.8		1.9	2.0	1.6		1.9	1.8	1.6		2.5	1.8
	M5	1.3		1.6	1.0	1.4		1.9	1.3	1.6		2.5	1.8
	M6b			2.9				3.4				2.5	
12	M3a	3.0		2.8	3.0	3.0		3.1	3.0	2.9		3.1	2.8
	M3	1.8		2.5	2.0	1.8		2.5	2.0	2.0		2.1	1.8
	M5	1.3		1.0	1.0	1.3		1.0	1.0	1.1		1.6	1.5
	M6b			3.8				3.4				3.1	
18	M3a	2.0		3.0	3.0	2.0		3.0	3.0	1.6		3.1	2.5
	M4	3.0		2.0	2.0	3.0		2.0	2.0	3.0		1.8	2.0
	M5	1.0		1.0	1.0	1.0		1.0	1.0	1.4		1.8	1.5
	M6b			4.0				4.0				3.4	
24	M3a	1.3		3.0	2.0	1.3		3.1	2.1	1.6		3.1	2.1
	M4	3.0		2.0	3.0	2.5		2.0	2.9	2.8		1.9	2.4
	M5	1.8		1.0	1.0	2.3		1.0	1.0	1.6		1.6	1.5
	M6b			4.0				3.9				3.4	
30	M3a	1.0		3.0	1.1	1.0		3.0	1.3	1.4		3.1	1.6
	M4	3.0		2.0	3.0	2.9		2.0	3.0	2.8		2.0	2.8
	M5	2.0		1.0	1.9	2.1		1.0	1.8	1.9		1.3	1.6
	M6b			4.0				4.0				3.6	

Table 4.3.1-26

Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	3.0		3.4	3.0	3.0		3.5	3.0	3.0		3.0	3.0
	M4	2.0		2.0	2.0	2.0		1.5	2.0	1.8		2.0	1.6
	M5	1.0		1.0	1.0	1.0		1.5	1.0	1.3		2.0	1.4
	M6b			3.6				3.5				3.0	
12	M3a	2.3		2.5	3.0	2.3		2.5	3.0	2.4		3.1	3.0
	M3	2.8		2.3	2.0	2.8		2.3	2.0	2.4		1.6	2.0
	M5	1.0		1.3	1.0	1.0		1.3	1.0	1.3		1.6	1.0
	M6b			4.0				4.0				3.6	
18	M3a	1.3		3.0	3.0	1.4		3.0	3.0	1.4		3.3	2.8
	M4	3.0		2.0	2.0	3.0		2.0	2.0	2.9		1.5	2.3
	M5	1.8		1.0	1.0	1.6		1.0	1.0	1.8		1.5	1.0
	M6b			4.0				4.0				3.8	
24	M3a	1.0		3.0	2.0	1.3		3.0	2.0	1.5		2.6	2.0
	M4	3.0		2.0	3.0	2.8		2.0	3.0	2.5		1.9	2.9
	M5	2.0		1.0	1.0	2.0		1.0	1.0	2.0		1.5	1.1
	M6b			4.0				4.0				4.0	
30	M3a	1.3		3.0	1.0	1.3		3.0	1.0	1.5		3.3	1.4
	M4	2.8		2.0	3.0	2.6		2.0	3.0	2.5		1.8	3.0
	M5	2.0		1.0	2.0	2.1		1.0	2.0	2.0		1.3	1.6
	M6b			4.0				4.0				3.8	

Table 4.3.1-27

Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	3.0		3.4	3.0	3.0		3.5	3.0	3.0		3.0	3.0
	M4	2.0		2.0	2.0	2.0		1.5	2.0	1.8		2.0	1.6
	M5	1.0		1.0	1.0	1.0		1.5	1.0	1.3		2.0	1.4
	M6b			3.6				3.5				3.0	
12	M3a	2.8		3.0	3.0	2.8		2.5	3.0	2.6		3.3	3.0
	M3	2.3		2.0	2.0	2.3		2.3	2.0	2.4		1.8	1.9
	M5	1.0		1.0	1.0	1.0		1.3	1.0	1.0		1.4	1.1
	M6b			4.0				4.0				3.6	
18	M3a	1.4		3.0	2.8	1.4		3.0	3.0	1.6		3.0	2.9
	M4	3.0		2.0	2.3	3.0		2.0	2.0	2.8		1.8	2.1
	M5	1.6		1.0	1.0	1.6		1.0	1.0	1.6		1.3	1.0
	M6b			4.0				4.0				4.0	
24	M3a	1.3		3.0	2.0	1.3		3.0	2.0	1.4		3.4	2.0
	M4	3.0		2.0	3.0	2.8		2.0	3.0	2.8		1.6	2.9
	M5	1.8		1.0	1.0	2.0		1.0	1.0	1.9		1.4	1.1
	M6b			4.0				4.0				3.6	
30	M3a	1.3		3.0	1.0	1.3		3.0	1.0	1.5		2.6	1.5
	M4	2.8		2.0	3.0	2.8		2.0	3.0	2.5		2.0	3.0
	M5	2.0		1.0	2.0	2.0		1.0	2.0	2.0		1.5	1.5
	M6b			4.0				4.0				3.9	

Table 4.3.1-28
Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .01, 1-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.0		3.0	2.4	2.0		2.9	2.5	1.7		2.3	2.0
	M4	2.6		2.5	2.4	2.5		2.6	2.3	2.6		2.7	2.3
	M5	1.4		1.3	1.2	1.5		1.4	1.2	1.7		2.3	1.7
	M6b			3.2				3.1				2.7	
0.5	M3a	1.6		3.1	2.4	2.0		3.0	2.4	2.0		3.1	2.3
	M4	2.6		1.9	2.4	2.6		2.1	2.3	2.5		1.9	2.2
	M5	1.8		1.1	1.2	1.4		1.3	1.3	1.5		1.5	1.5
	M6b			3.9				3.6				3.5	
0.8	M3a	2.4		3.1	2.4	2.2		3.2	2.4	2.2		3.2	2.3
	M4	2.2		1.9	2.4	2.6		1.9	2.4	2.3		1.9	2.3
	M5	1.4		1.1	1.2	1.2		1.1	1.2	1.5		1.5	1.4
	M6b			3.9				3.8				3.4	
1.2	M3a	2.1		3.1	2.4	2.3		3.1	2.4	2.3		3.1	2.5
	M4	2.6		2.0	2.4	2.4		1.9	2.4	2.2		1.7	2.3
	M5	1.3		1.0	1.2	1.3		1.1	1.2	1.5		1.7	1.2
	M6b			3.9				3.9				3.5	

Table 4.3.1-29

Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .01, 2-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.0		3.0	2.5	2.0		2.8	2.6	1.8		2.5	2.0
	M4	2.6		2.5	2.4	2.5		2.4	2.2	2.6		2.9	2.0
	M5	1.4		1.3	1.1	1.5		1.2	1.2	1.6		2.1	2.0
	M6b			3.2				3.6				2.5	
0.5	M3a	2.0		3.1	2.4	2.0		3.0	2.5	2.0		3.2	2.3
	M4	2.6		1.9	2.4	2.6		2.1	2.3	2.5		1.9	2.0
	M5	1.4		1.1	1.2	1.4		1.3	1.2	1.5		1.5	1.7
	M6b			3.9				3.6				3.4	
0.8	M3a	2.2		3.1	2.4	2.2		3.2	2.4	2.1		3.1	2.3
	M4	2.4		1.9	2.4	2.0		1.9	2.4	2.5		1.7	2.3
	M5	1.4		1.1	1.2	1.8		1.1	1.2	1.4		1.7	1.4
	M6b			3.9				3.8				3.5	
1.2	M3a	2.0		3.1	2.4	2.0		3.1	2.4	2.3		3.2	2.6
	M4	2.4		2.0	2.4	2.3		1.9	2.4	2.1		1.7	2.2
	M5	1.6		1.0	1.2	1.7		1.1	1.2	1.6		1.7	1.2
	M6b			3.9				3.9				3.4	

Table 4.3.1-30

Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.0		3.1	2.4	2.0		3.1	2.4	2.0		3.0	2.5
	M4	2.6		2.0	2.4	2.6		1.9	2.4	2.5		1.9	2.3
	M5	1.4		1.0	1.2	1.4		1.1	1.2	1.5		1.6	1.2
	M6b			3.9				3.9				3.5	
0.5	M3a	1.6		3.0	2.4	1.6		3.1	2.4	1.6		3.1	2.4
	M4	2.8		2.0	2.4	2.8		1.9	2.4	2.7		1.8	2.4
	M5	1.6		1.0	1.2	1.6		1.1	1.2	1.7		1.6	1.2
	M6b			4.0				3.9				3.5	
0.8	M3a	1.6		2.7	2.4	1.6		2.7	2.4	2.2		3.0	2.4
	M4	2.8		2.2	2.4	2.7		2.1	2.4	2.3		1.7	2.3
	M5	1.6		1.2	1.2	1.7		1.3	1.2	1.5		1.5	1.3
	M6b			3.9				3.9				3.8	
1.2	M3a	1.8		3.1	2.4	2.1		3.1	2.4	2.0		3.1	2.4
	M4	2.6		2.0	2.4	2.4		1.9	2.4	2.1		1.6	2.4
	M5	1.6		1.0	1.2	1.5		1.1	1.2	1.9		1.6	1.2
	M6b			3.9				3.9				3.7	

Table 4.3.1-31

Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.0		3.1	2.4	2.0		3.1	2.4	1.9		2.7	2.5
	M4	2.6		2.0	2.4	2.6		1.9	2.4	2.5		2.1	2.3
	M5	1.4		1.0	1.2	1.4		1.1	1.2	1.6		1.6	1.2
	M6b			3.9				3.9				3.6	
0.5	M3a	1.8		3.0	2.4	1.8		3.1	2.4	1.8		3.0	2.5
	M4	2.6		2.0	2.4	2.6		1.9	2.4	2.6		2.0	2.2
	M5	1.6		1.0	1.2	1.6		1.1	1.2	1.6		1.4	1.3
	M6b			4.0				3.9				3.6	
0.8	M3a	1.8		3.1	2.4	1.7		2.7	2.4	2.1		3.4	2.5
	M4	2.8		2.0	2.4	2.7		2.1	2.4	2.5		1.6	2.3
	M5	1.4		1.0	1.2	1.6		1.3	1.2	1.4		1.4	1.2
	M6b			3.9				3.9				3.6	
1.2	M3a	2.1		3.1	2.2	2.2		3.1	2.4	2.3		3.1	2.4
	M4	2.4		2.0	2.6	2.3		1.9	2.4	2.1		1.6	2.4
	M5	1.5		1.0	1.2	1.5		1.1	1.2	1.6		1.6	1.2
	M6b			3.9				3.9				3.7	

Table4.3.1-32

Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .01, 1-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.0		3.1	2.4	2.1		3.1	2.4	2.1		2.9	2.3
M4	2.5		2.1	2.4	2.5		2.1	2.4	2.4		2.1	2.3
M5	1.5		1.1	1.2	1.4		1.2	1.2	1.6		1.8	1.5
M6b			3.7				3.6				3.3	

Table 4.3.1-33

Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .01, 2-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.1		3.1	2.4	2.1		3.0	2.5	2.1		3.0	2.3
M4	2.5		2.1	2.4	2.4		2.1	2.3	2.4		2.1	2.1
M5	1.5		1.1	1.2	1.6		1.2	1.2	1.5		1.8	1.6
M6b			3.7				3.7				3.2	

Table 4.3.1-34

Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	1.8		3.0	2.4	1.8		3.0	2.4	2.0		3.1	2.4
M4	2.7		2.1	2.4	2.6		2.0	2.4	2.4		1.8	2.4
M5	1.6		1.1	1.2	1.6		1.2	1.2	1.7		1.6	1.2
M6b			3.9				3.9				3.6	

Table 4.3.1-35

Kolmogorov-Smirnov Test of General Differences for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	1.9		3.1	2.4	1.9		3.0	2.4	2.0		3.1	2.5
M4	2.6		2.0	2.5	2.6		2.0	2.4	2.4		1.8	2.3
M5	1.5		1.0	1.2	1.5		1.2	1.2	1.6		1.5	1.2
M6b			3.9				3.9				3.6	

Table 4.3.1-36

Kolmogorov-Smirnov Test of General Differences for Two Groups
Analysis of Mean Ranks of Power Results
Number of First Place Finishes by Distribution
Across Nominal Alpha, Direction and Number of Groups

Decimal	| ------------- 4th			------------- |	| ------------- 2^{nd}			------------- |
Method	3a	4	5	6b	3a	4	5	6b
By Initial Sample Size Across Nominal Effect Size Multiplier								
EA	10.00	0.50	9.50		9.00	1.00	10.00	
EB								
ML	0.00	0.00	20.00	0.00	0.50	3.50	15.50	0.50
SS	4.00	0.00	16.00		2.50	1.00	16.50	
By Nominal Effect Size Multiplier Across Initial Sample Size								
EA	0.00	0.00	4.00		1.50	0.00	14.50	
EB								
ML	0.00	0.00	4.00	0.00	0.50	2.50	13.00	0.00
SS	0.00	0.00	4.00		0.30	0.30	15.30	
Across Nominal Effect Size Multiplier and Initial Sample Size								
EA	0.00	0.00	4.00		0.00	0.00	4.00	
EB								
ML	0.00	0.00	4.00	0.00	0.00	0.00	4.00	0.00
SS	0.00	0.00	4.00		0.00	0.00	4.00	

Table 4.3.1-37

Kolmogorov-Smirnov Test of General Differences for Two Groups
Analysis of Mean Ranks of Power Results
Number of First Place Finishes
Across Nominal Alpha, Direction, Distributions and Number of Groups

Decimal	| ------------- 4th			------------- |	| ------------- 2^{nd}			------------- |
Method	3a	4	5	6b	3a	4	5	6b
By Initial Sample Size Across Nominal Effect Size Multiplier								
MP1i	60	60	60	20	60	60	60	20
N1Mi	14.0	0.5	45.5	0.0	12.0	5.5	42.0	0.5
PoM	0.233	0.008	0.758	0.000	0.200	0.092	0.700	0.025
PoT	0.233	0.008	0.758	0.000	0.200	0.092	0.700	0.008
By Nominal Effect Size Multiplier Across Initial Sample Size								
MP1i	48	48	48	16	48	48	48	16
N1Mi	2.0	0.0	46.0	0.0	2.33	2.83	42.83	0.0
PoM	0.042	0.000	0.958	0.000	0.049	0.059	0.892	0.000
PoT	0.042	0.000	0.958	0.000	0.049	0.059	0.892	0.000
Across Nominal Effect Size Multiplier and Initial Sample Size								
MP1i	12	12	12	4	12	12	12	4
N1Mi	0.0	0.0	12.0	0.0	0.0	0.0	12.0	0.0
PoM	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
PoT	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000

Key: MP1i: Maximum possible 1^{st} place finishes for each method
N1Mi: Actual number of 1^{st} place finishes for each method (ties count 1/n)
PoM: Proportion of Maximum Possible { N1Mi / MP1i }
PoT: Proportion of total 1^{st} place finishes { N1Mi / Σ(N1Mi) }

4.3.2 – Rosenbaum's Test (two groups only)

Based on the Type I error results, power and Type III error results for Rosenbaum's Test are only presented for the combinations of alpha level, distribution and method shown in Table 4.3.2-1.

Table 4.3.2-1
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Method / Distribution Combinations with Acceptable Type I Error

Alpha	.01					.05				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Method										
M3a	x	x		x	x	x	x		x	x
M3b	x	x		x	x	x	x		x	x
M4	x			x	x	x			x	x
M5	x	x	x	x	x	x	x	x	x	x
M6c	x			x	x	x			x	x

Power results for Rosenbaum's Test at nominal alpha .01 and .05 are summarized in Table 4.3.2-2 and 4.3.2-3. These tables present the range of values obtained (minimum and maximum) for 1-tailed power, 2-tailed Type III error and 2-tailed power based on the results presented in Tables 4.3.2-4 through 4.3.2-23.

Power tended to increase monotonically across methods and distributions with increases in initial sample size and/or effect size. Likewise, Type III error tended to decrease monotonically across methods and distributions with an increase in initial sample size and/or effect size. Thus, minimum power (or maximum Type III error) usually occurred at initial sample size 6 or 12 at nominal

ESM 0.2, with maximum power (or minimum Type III error) at initial sample size 30 and nominal ESM 1.2, if not sooner.

Table 4.3.2-2
Rosenbaum's Test for Two Groups
Range of Power and Type III Error for α .01

Min/Max	min					max				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Type										
Mthd										
Pwr 1-s										
M3a	.0152	.0147		.0111	.0017	.4762	.+		.9282	.5072
M3b	.0152	.0147		.0111	.0017	.4762	.+		.9282	.5072
M4	.0152			.0123	.0153	.4762			.9347	.5579
M5	.0152	.0147	.0514	.0123	.0153	.4762	.+	.9995	.9335	.5232
M6c	.0152			.0111	.0148	.4762			.9270	.4998
T3e 2-s										
M3a	.0000	.0000		.0000	.0000	.0013	.0003		.0010	.0010
M3b	.0000	.0000		.0000	.0000	.0013	.0003		.0010	.0010
M4	.0000			.0000	.0000	.0013			.0012	.0014
M5	.0000	.0000	.0000	.0000	.0000	.0013	.0004	.0002	.0012	.0013
M6c	.0000			.0000	.0000	.0013			.0010	.0009
Pwr 2-s										
M3a	.0026	.0019		.0017	.0017	.4216	.+		.8557	.4292
M3b	.0026	.0019		.0017	.0017	.4216	.+		.8557	.4292
M4	.0026			.0020	.0025	.4216			.8670	.4764
M5	.0026	.0031	.0094	.0020	.0026	.4216	.+	.9980	.8654	.4470
M6c	.0026			.0017	.0017	.4216			.8538	.4216

.+ = 1.0000

Tables 4.3.2-4 through 4.3.2-23 give the upper and lower tail results for Rosenbaum's Test for both 1-sided and 2-sided tests for both alpha .01 and .05. There is a table for each combination of nominal effect size multiplier {0.2, 0.5, 0.8, 1.2} and initial sample size {6, 12, 18, 24, 30}. Results are not reported for distribution EB with nominal ESM 0.2 as the actual ESM = 0.0 (no shift), which is just Type I error. Also for distribution EB, results are not reported for nominal ESM 0.8 as the actual ESM = 0.592, the same as the actual ESM for nominal ESM of 0.5. Results for Methods 3a and 3b are numerically equivalent for all combinations of alpha, distribution and direction, but both results are still reported. Results for the normal distribution are included and are essentially identical across methods for a fixed initial sample size, effect size and directionality. This is to be expected in the absence of ties and demonstrates that the simulations worked correctly.

Table 4.3.2-3
Rosenbaum's Test for Two Groups
Range of Power and Type III Error for α .05

Min/Max	min					max				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Type										
Mthd										
Pwr 1-s										
M3a	.0520	.0636		.0415	.0441	.6480	.+		.9896	.7048
M3b	.0520	.0636		.0415	.0441	.6480	.+		.9896	.7048
M4	.0520			.0453	.0541	.6480			.9908	.7482
M5	.0520	.0883	.1799	.0453	.0523	.6480	.+	.+	.9906	.7212
M6c	.0520			.0415	.0416	.6480			.9893	.7003
T3e 2-s										
M3a	.0000	.0000		.0000	.0000	.0090	.0018		.0075	.0070
M3b	.0000	.0000		.0000	.0000	.0090	.0018		.0075	.0070
M4	.0000			.0000	.0000	.0090			.0087	.0093
M5	.0000	.0000	.0000	.0000	.0000	.0090	.0027	.0015	.0086	.0089
M6c	.0000			.0000	.0000	.0090			.0074	.0064
Pwr 2-s										
M3a	.0155	.0147		.0111	.0117	.6086	.+		.9699	.6396
M3b	.0155	.0147		.0111	.0117	.6086	.+		.9699	.6396
M4	.0155			.0123	.0153	.6086			.9731	.6832
M5	.0155	.0221	.0514	.0123	.0155	.6086	.+	.+	.9725	.6579
M6c	.0155			.0111	.0115	.6086			.9693	.6346

.+ = 1.0000

A ranking analysis of the results is presented in Tables 4.3.2-24 through 4.3.2-35. These summaries were obtained by ranking the power results from Tables 4.3.2-4 through 4.3.2-23 to four decimal places (as reported), to three decimal places and to two decimal places.

Table 4.3.2-4
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0035	.0018		.0042	.0024	.0159	.0075		.0184	.0123
(T3e)	M3b	.0035	.0018		.0042	.0024	.0159	.0075		.0184	.0123
	M4	.0035			.0047	.0033	.0159			.0200	.0159
	M5	.0035	.0027		.0047	.0034	.0159	.0105		.0201	.0161
	M6c	.0035			.0042	.0023	.0159			.0184	.0121
UT1s	M3a	.0155	.0147		.0111	.0117	.0541	.0636		.0415	.0441
(Pwr)	M3b	.0155	.0147		.0111	.0117	.0541	.0636		.0415	.0441
	M4	.0155			.0123	.0153	.0541			.0453	.0541
	M5	.0155	.0221		.0123	.0155	.0541	.0883		.0453	.0539
	M6c	.0155			.0111	.0115	.0541			.0415	.0434
LT2s	M3a	.0004	.0003		.0005	.0003	.0035	.0018		.0042	.0024
(T3e)	M3b	.0004	.0003		.0005	.0003	.0035	.0018		.0042	.0024
	M4	.0004			.0006	.0004	.0035			.0047	.0033
	M5	.0004	.0004		.0006	.0004	.0035	.0027		.0047	.0034
	M6c	.0004			.0005	.0003	.0035			.0042	.0023
UT2s	M3a	.0026	.0019		.0017	.0017	.0155	.0147		.0111	.0117
(Pwr)	M3b	.0026	.0019		.0017	.0017	.0155	.0147		.0111	.0117
	M4	.0026			.0020	.0025	.0155			.0123	.0153
	M5	.0026	.0031		.0020	.0026	.0155	.0221		.0123	.0155
	M6c	.0026			.0017	.0017	.0155			.0111	.0115

Table 4.3.2-5
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0025	.0006		.0028	.0019	.0225	.0033		.0190	.0180
(T3e)	M3b	.0025	.0006		.0028	.0019	.0225	.0033		.0190	.0180
	M4	.0025			.0033	.0025	.0225			.0221	.0228
	M5	.0025	.0008		.0032	.0025	.0225	.0048		.0219	.0225
	M6c	.0025			.0028	.0017	.0225			.0188	.0171
UT1s	M3a	.0172	.0427		.0130	.0135	.0882	.2765		.0834	.0744
(Pwr)	M3b	.0172	.0427		.0130	.0135	.0882	.2765		.0834	.0744
	M4	.0172			.0151	.0173	.0882			.0945	.0898
	M5	.0172	.0578		.0150	.0169	.0882	.2926		.0939	.0873
	M6c	.0172			.0130	.0126	.0882			.0829	.0720
LT2s	M3a	.0007	.0002		.0009	.0005	.0076	.0015		.0075	.0061
(T3e)	M3b	.0007	.0002		.0009	.0005	.0076	.0015		.0075	.0061
	M4	.0007			.0011	.0007	.0076			.0087	.0079
	M5	.0007	.0003		.0011	.0007	.0076	.0020		.0086	.0078
	M6c	.0007			.0009	.0005	.0076			.0074	.0057
UT2s	M3a	.0066	.0138		.0043	.0050	.0407	.1173		.0345	.0330
(Pwr)	M3b	.0066	.0138		.0043	.0050	.0407	.1173		.0345	.0330
	M4	.0066			.0051	.0066	.0407			.0399	.0410
	M5	.0066	.0205		.0055	.0064	.0407	.1412		.0396	.0397
	M6c	.0066			.0043	.0046	.0407			.0343	.0315

Table 4.3.2-6

Rosenbaum's Test for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 0.2, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0030	.0002		.0025	.0025	.0085	.0004		.0064	.0070
(T3e)	M3b	.0030	.0002		.0025	.0025	.0085	.0004		.0064	.0070
	M4	.0030			.0031	.0033	.0085			.0078	.0092
	M5	.0030	.0003		.0030	.0032	.0085	.0007		.0076	.0089
	M6c	.0030			.0025	.0022	.0085			.0062	.0064
UT1s	M3a	.0251	.1798		.0240	.0206	.0520	.3591		.0532	.0442
(Pwr)	M3b	.0251	.1798		.0240	.0206	.0520	.3591		.0532	.0442
	M4	.0251			.0285	.0263	.0520			.0623	.0548
	M5	.0251	.1671		.0279	.0248	.0520	.3087		.0614	.0523
	M6c	.0251			.0236	.0191	.0520			.0525	.0416
LT2s	M3a	.0010	.0001		.0010	.0009	.0085	.0004		.0064	.0070
(T3e)	M3b	.0010	.0001		.0010	.0009	.0085	.0004		.0064	.0070
	M4	.0010			.0012	.0012	.0085			.0078	.0092
	M5	.0010	.0001		.0012	.0012	.0085	.0007		.0076	.0089
	M6c	.0010			.0010	.0007	.0085			.0062	.0064
UT2s	M3a	.0115	.0767		.0100	.0092	.0520	.3591		.0532	.0442
(Pwr)	M3b	.0115	.0767		.0100	.0092	.0520	.3591		.0532	.0442
	M4	.0115			.0121	.0122	.0520			.0623	.0548
	M5	.0115	.0810		.0119	.0113	.0520	.3087		.0611	.0523
	M6c	.0115			.0098	.0084	.0520			.0525	.0416

Table 4.3.2-7
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0013	.0000		.0008	.0010	.0090	.0001		.0054	.0070
(T3e)	M3b	.0013	.0000		.0008	.0010	.0090	.0001		.0054	.0070
	M4	.0013			.0011	.0014	.0090			.0067	.0093
	M5	.0013	.0000		.0010	.0013	.0090	.0003		.0066	.0088
	M6c	.0013			.0008	.0009	.0090			.0052	.0062
UT1s	M3a	.0152	.2528		.0155	.0123	.0591	.6539		.0679	.0514
(Pwr)	M3b	.0152	.2528		.0155	.0123	.0591	.6539		.0679	.0514
	M4	.0152			.0190	.0162	.0591			.0804	.0639
	M5	.0152	.1814		.0182	.0148	.0591	.4891		.0780	.0599
	M6c	.0152			.0151	.0110	.0591			.0665	.0479
LT2s	M3a	.0013	.0000		.0008	.0010	.0090	.0001		.0054	.0070
(T3e)	M3b	.0013	.0000		.0008	.0010	.0090	.0001		.0054	.0070
	M4	.0013			.0011	.0014	.0090			.0067	.0093
	M5	.0013	.0000		.0010	.0013	.0090	.0003		.0066	.0088
	M6c	.0013			.0008	.0009	.0090			.0052	.0062
UT2s	M3a	.0152	.2528		.0155	.0123	.0591	.6539		.0679	.0514
(Pwr)	M3b	.0152	.2528		.0155	.0123	.0591	.6539		.0679	.0514
	M4	.0152			.0190	.0162	.0591			.0804	.0639
	M5	.0152	.1814		.0182	.0148	.0591	.4891		.0780	.0599
	M6c	.0152			.0151	.0110	.0591			.0665	.0479

Table 4.3.2-8
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0013	.0000		.0007	.0010	.0090	.0001		.0045	.0067
(T3e)	M3b	.0013	.0000		.0007	.0010	.0090	.0001		.0045	.0067
	M4	.0013			.0009	.0014	.0090			.0059	.0090
	M5	.0013	.0000		.0009	.0013	.0090	.0001		.0057	.0085
	M6c	.0013			.0007	.0008	.0090			.0043	.0058
UT1s	M3a	.0177	.5155		.0210	.0150	.0642	.8548		.0824	.0582
(Pwr)	M3b	.0177	.5155		.0210	.0150	.0642	.8548		.0824	.0582
	M4	.0177			.0259	.0198	.0642			.0987	.0725
	M5	.0177	.3164		.0247	.0175	.0642	.6528		.0949	.0671
	M6c	.0177			.0212	.0132	.0642			.0802	.0536
LT2s	M3a	.0005	.0000		.0003	.0004	.0034	.0000		.0018	.0026
(T3e)	M3b	.0005	.0000		.0003	.0004	.0034	.0000		.0018	.0026
	M4	.0005			.0004	.0005	.0034			.0024	.0035
	M5	.0005	.0000		.0003	.0005	.0034	.0000		.0023	.0033
	M6c	.0005			.0003	.0003	.0034			.0017	.0021
UT2s	M3a	.0091	.3302		.0102	.0073	.0341	.7042		.0422	.0298
(Pwr)	M3b	.0091	.3302		.0102	.0073	.0341	.7042		.0422	.0298
	M4	.0091			.0127	.0099	.0341			.0514	.0384
	M5	.0091	.1911		.0122	.0087	.0341	.4775		.0492	.0349
	M6c	.0091			.0098	.0063	.0341			.0408	.0268

Table 4.3.2-9

Rosenbaum's Test for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 0.5, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0009	.0031		.0017	.0010	.0053	.0011		.0085	.0058
(T3e)	M3b	.0009	.0031		.0017	.0010	.0053	.0011		.0085	.0058
	M4	.0009			.0019	.0014	.0053			.0093	.0077
	M5	.0009	.0004	.0015	.0019	.0015	.0053	.0015	.0050	.0093	.0078
	M6c	.0009			.0017	.0010	.0053			.0085	.0057
UT1s	M3a	.0393	.0938		.0276	.0231	.1128	.3012		.0878	.0752
(Pwr)	M3b	.0393	.0938		.0276	.0231	.1128	.3012		.0878	.0752
	M4	.0393			.0302	.0294	.1128			.0947	.0902
	M5	.0393	.1191	.0514	.0303	.0295	.1128	.3445	.1799	.0945	.0893
	M6c	.0393			.0276	.0228	.1128			.0878	.0742
LT2s	M3a	.0001	.0000		.0002	.0001	.0009	.0003		.0017	.0010
(T3e)	M3b	.0001	.0000		.0002	.0001	.0009	.0003		.0017	.0010
	M4	.0001			.0002	.0001	.0009			.0019	.0014
	M5	.0001	.0001	.0002	.0002	.0001	.0009	.0004	.0015	.0019	.0015
	M6c	.0001			.0002	.0001	.0009			.0017	.0010
UT2s	M3a	.0082	.0138		.0065	.0040	.0393	.0938		.0276	.0231
(Pwr)	M3b	.0082	.0138		.0065	.0040	.0393	.0938		.0276	.0231
	M4	.0082			.0074	.0055	.0393			.0302	.0294
	M5	.0082	.0200	.0094	.0074	.0056	.0393	.1191	.0514	.0303	.0295
	M6c	.0082			.0065	.0040	.0393			.0276	.0228

Table 4.3.2-10
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0004	.0000		.0007	.0006	.0060	.0001		.0042	.0077
(T3e)	M3b	.0004	.0000		.0007	.0006	.0060	.0001		.0042	.0077
	M4	.0004			.0008	.0008	.0060			.0048	.0102
	M5	.0004	.0000	.0001	.0008	.0008	.0060	.0001	.0008	.0047	.0101
	M6c	.0004			.0007	.0005	.0060			.0042	.0073
UT1s	M3a	.0542	.4876		.0524	.0304	.1920	.8765		.2569	.1311
(Pwr)	M3b	.0542	.4876		.0524	.0304	.1920	.8765		.2569	.1311
	M4	.0542			.0593	.0381	.1920			.2811	.1537
	M5	.0542	.5027	.1983	.0589	.0367	.1920	.8728	.5976	.2792	.1490
	M6c	.0542			.0522	.0290	.1920			.2560	.1277
LT2s	M3a	.0001	.0000		.0002	.0001	.0016	.0000		.0018	.0021
(T3e)	M3b	.0001	.0000		.0002	.0001	.0016	.0000		.0018	.0021
	M4	.0001			.0003	.0002	.0016			.0021	.0030
	M5	.0001	.0000	.0001	.0003	.0002	.0016	.0000	.0003	.0021	.0029
	M6c	.0001			.0002	.0001	.0016			.0018	.0020
UT2s	M3a	.0251	.2704		.0185	.0130	.1063	.7102		.1252	.0656
(Pwr)	M3b	.0251	.2704		.0185	.0130	.1063	.7102		.1252	.0656
	M4	.0251			.0212	.0168	.1063			.1395	.0797
	M5	.0251	.2932	.0869	.0211	.0162	.1063	.7139	.3774	.1386	.0767
	M6c	.0251			.0184	.0122	.1063			.1248	.0633

Table 4.3.2-11
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0004	.0000		.0003	.0007	.0016	.0000		.0006	.0022
(T3e)	M3b	.0004	.0000		.0003	.0007	.0016	.0000		.0006	.0022
	M4	.0004			.0003	.0009	.0016			.0007	.0030
	M5	.0004	.0000	.0000	.0003	.0009	.0016	.0000	.0001	.0007	.0029
	M6c	.0004			.0003	.0006	.0016			.0006	.0020
UT1s	M3a	.0819	.9033		.1371	.0479	.1401	.9652		.2502	.0910
(Pwr)	M3b	.0819	.9033		.1371	.0479	.1401	.9652		.2502	.0910
	M4	.0819			.1563	.0594	.1401			.2796	.1096
	M5	.0819	.8808	.5148	.1535	.0560	.1401	.9524	.7005	.2746	.1038
	M6c	.0819			.1359	.0451	.1401			.2482	.0869
LT2s	M3a	.0001	.0000		.0001	.0002	.0016	.0000		.0006	.0022
(T3e)	M3b	.0001	.0000		.0001	.0002	.0016	.0000		.0006	.0022
	M4	.0001			.0001	.0003	.0016			.0007	.0030
	M5	.0001	.0000	.0000	.0001	.0003	.0016	.0000	.0001	.0007	.0029
	M6c	.0001			.0001	.0002	.0016			.0006	.0020
UT2s	M3a	.0456	.7876		.0685	.0242	.1401	.9652		.2502	.0910
(Pwr)	M3b	.0456	.7876		.0685	.0242	.1401	.9652		.2502	.0910
	M4	.0456			.0796	.0307	.1401			.2796	.1096
	M5	.0456	.7589	.3338	.0780	.0286	.1401	.9524	.7005	.2746	.1038
	M6c	.0456			.0678	.0224	.1401			.2482	.0869

Table 4.3.2-12
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0001	.0000		.0000	.0002	.0014	.0000		.0002	.0020
(T3e)	M3b	.0001	.0000		.0000	.0002	.0014	.0000		.0002	.0020
	M4	.0001			.0001	.0003	.0014			.0003	.0027
	M5	.0001	.0000	.0000	.0000	.0003	.0014	.0000	.0000	.0003	.0026
	M6c	.0001			.0000	.0002	.0014			.0002	.0017
UT1s	M3a	.0608	.9708		.1363	.0341	.1622	.9976		.3746	.1117
(Pwr)	M3b	.0608	.9708		.1363	.0341	.1622	.9976		.3746	.1117
	M4	.0608			.1582	.0431	.1622			.4156	.1341
	M5	.0608	.9506	.6193	.1531	.0394	.1622	.9944	.8858	.4049	.1255
	M6c	.0608			.1340	.0312	.1622			.3703	.1060
LT2s	M3a	.0001	.0000		.0000	.0002	.0014	.0000		.0002	.0020
(T3e)	M3b	.0001	.0000		.0000	.0002	.0014	.0000		.0002	.0020
	M4	.0001			.0001	.0003	.0014			.0003	.0027
	M5	.0001	.0000	.0000	.0000	.0003	.0014	.0000	.0000	.0003	.0026
	M6c	.0001			.0000	.0002	.0014			.0002	.0017
UT2s	M3a	.0608	.9708		.1363	.0341	.1622	.9976		.3746	.1117
(Pwr)	M3b	.0608	.9708		.1363	.0341	.1622	.9976		.3746	.1117
	M4	.0608			.1582	.0431	.1622			.4156	.1341
	M5	.0608	.9506	.6193	.1531	.0394	.1622	.9944	.8858	.4049	.1255
	M6c	.0608			.1340	.0312	.1622			.3703	.1060

Table 4.3.2-13
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0001	.0000		.0000	.0002	.0013	.0000		.0001	.0017
(T3e)	M3b	.0001	.0000		.0000	.0002	.0013	.0000		.0001	.0017
	M4	.0001			.0000	.0003	.0013			.0001	.0023
	M5	.0001	.0000	.0000	.0000	.0002	.0013	.0000	.0000	.0001	.0022
	M6c	.0001			.0000	.0001	.0013			.0001	.0014
UT1s	M3a	.0721	.9975		.2173	.0434	.1786	.9999		.4993	.1294
(Pwr)	M3b	.0721	.9975		.2173	.0434	.1786	.9999		.4993	.1294
	M4	.0721			.2511	.0551	.1786			.5460	.1549
	M5	.0721	.9928	.8234	.2401	.0492	.1786	.9995	.9637	.5293	.1433
	M6c	.0721			.2126	.0395	.1786			.4920	.1222
LT2s	M3a	.0000	.0000		.0000	.0001	.0004	.0000		.0000	.0005
(T3e)	M3b	.0000	.0000		.0000	.0001	.0004	.0000		.0000	.0005
	M4	.0000			.0000	.0001	.0004			.0000	.0008
	M5	.0000	.0000	.0000	.0000	.0001	.0004	.0000	.0000	.0000	.0007
	M6c	.000			.0000	.0001	.0004			.0000	.0004
UT2s	M3a	.0443	.9917		.1303	.0239	.1152	.9994		.3414	.0761
(Pwr)	M3b	.0443	.9917		.1303	.0239	.1152	.9994		.6414	.0761
	M4	.0443			.1538	.0311	.1152			.3845	.0942
	M5	.0443	.9800	.7001	.1465	.0272	.1152	.9979	.9116	.3696	.0857
	M6c	.0443			.1269	.0213	.1152			.3351	.0706

Table 4.3.2-14
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0002	.0001		.0006	.0001	.0015	.0002		.0037	.0010
(T3e)	M3b	.0002	.0001		.0006	.0001	.0015	.0002		.0037	.0010
	M4	.0002			.0007	.0002	.0015			.0041	.0015
	M5	.0002	.0001		.0007	.0002	.0015	.0003		.0042	.0015
	M6c	.0002			.0006	.0001	.0015			.0037	.0010
UT1s	M3a	.0873	.2599		.0565	.0714	.2070	.5749		.1581	.1798
(Pwr)	M3b	.0873	.2599		.0565	.0714	.2070	.5749		.1581	.1798
	M4	.0873			.0608	.0863	.2070			.1673	.2063
	M5	.0873	.2917		.0607	.0863	.2070	.6069		.1672	.2042
	M6c	.0873			.0565	.0707	.2070			.1580	.1781
LT2s	M3a	.0000	.0000		.0000	.0000	.0002	.0001		.0006	.0001
(T3e)	M3b	.0000	.0000		.0000	.0000	.0002	.0001		.0006	.0001
	M4	.0000			.0000	.0000	.0002			.0007	.0002
	M5	.0000	.0000		.0003	.0000	.0002	.0001		.0007	.0002
	M6c	.0000			.0000	.0000	.0002			.0006	.0001
UT2s	M3a	.0218	.0558		.0170	.0162	.0873	.2599		.0565	.0714
(Pwr)	M3b	.0218	.0558		.0170	.0162	.0873	.2599		.0565	.0714
	M4	.0218			.0188	.0212	.0873			.0608	.0863
	M5	.0218	.0687		.0188	.0212	.0873	.2917		.0607	.0863
	M6c	.0218			.0170	.0162	.0873			.0565	.0707

Table 4.3.2-15
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0002	.0000	.0013	.0000		.0013	.0010
(T3e)	M3b	.0000	.0000		.0002	.0000	.0013	.0000		.0013	.0010
	M4	.0000			.0002	.0001	.0013			.0015	.0013
	M5	.0000	.0000		.0003	.0001	.0013	.0000		.0014	.0014
	M6c	.0000			.0002	.0000	.0013			.0013	.0009
UT1s	M3a	.1354	.8463		.1458	.1145	.3442	.9858		.5139	.3156
(Pwr)	M3b	.1354	.8463		.1458	.1145	.3442	.9858		.5139	.3156
	M4	.1354			.1586	.1360	.3442			.5371	.3526
	M5	.1354	.8526		.1581	.1298	.3442	.9857		.5356	.3424
	M6c	.1354			.1455	.1112	.3442			.5132	.3111
LT2s	M3a	.0000	.0000		.0001	.0000	.0002	.0000		.0006	.0002
(T3e)	M3b	.0000	.0000		.0001	.0000	.0002	.0000		.0006	.0002
	M4	.0000			.0001	.0000	.0002			.0006	.0003
	M5	.0000	.0000		.0001	.0000	.0002	.0000		.0006	.0003
	M6c	.0000			.0001	.0000	.0002			.0006	.0002
UT2s	M3a	.0746	.6690		.0585	.0612	.2244	.9456		.3004	.1968
(Pwr)	M3b	.0746	.6690		.0585	.0612	.2244	.9456		.3004	.1968
	M4	.0746			.0648	.0749	.2244			.3205	.2268
	M5	.0746	.6844		.0646	.0713	.2244	.9469		.3196	.2181
	M6c	.0746			.0584	.0588	.2244			.2999	.1926

Table 4.3.2-16
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0001	.0000	.0002	.0000		.0001	.0001
(T3e)	M3b	.0000	.0000		.0001	.0000	.0002	.0000		.0001	.0001
	M4	.0000			.0001	.0001	.0002			.0001	.0002
	M5	.0000	.0000		.0001	.0001	.0002	.0000		.0001	.0002
	M6c	.0000			.0001	.0000	.0002			.0001	.0001
UT1s	M3a	.1998	.9948		.4140	.1815	.2912	.9989		.6020	.0001
(Pwr)	M3b	.1998	.9948		.4140	.1815	.2912	.9989		.6020	.2763
	M4	.1998			.4418	.2115	.2912			.6294	.3135
	M5	.1998	.9938		.4380	.1987	.2912	.9985		.6255	.2985
	M6c	.1998			.4124	.1755	.2912			.6004	.2697
LT2s	M3a	.0000	.0000		.0000	.0000	.0002	.0000		.0001	.0001
(T3e)	M3b	.0000	.0000		.0000	.0000	.0002	.0000		.0001	.0001
	M4	.0000			.0000	.0000	.0002			.0001	.0001
	M5	.0000	.0000		.0000	.0000	.0002	.0000		.0001	.0001
	M6c	.0000			.0000	.0000	.0002			.0001	.0001
UT2s	M3a	.1305			.2504	.1137	.2912	.9989		.6020	.2763
(Pwr)	M3b	.1305			.2504	.1137	.2912	.9989		.6020	.2763
	M4	.1305			.2725	.1362	.2912			.6294	.3135
	M5	.1305	.9785		.2701	.1264	.2912	.9985		.6255	.2985
	M6c	.1305			.2493	.1088	.2912			.6004	.2697

Table 4.3.2-17

Rosenbaum's Test for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 0.8, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0002	.0000		.0000	.0001
(T3e)	M3b	.0000	.0000		.0000	.0000	.0002	.0000		.0000	.0001
	M4	.0000			.0000	.0000	.0002			.0000	.0001
	M5	.0000	.0000		.0000	.0000	.0002	.0000		.0000	.0001
	M6c	.0000			.0000	.0000	.0002			.0001	.0001
UT1s	M3a	.1711	.9996		.4953	.1599	.3351	.+		.8054	.3370
(Pwr)	M3b	.1711	.9996		.4953	.1599	.3351	.+		.8054	.3370
	M4	.1711			.5269	.1896	.3351			.8259	.3792
	M5	.1711	.9993		.5204	.1734	.3351	.+		.8211	.3589
	M6c	.1711			.4922	.1526	.3351			.8030	.3291
LT2s	M3a	.0000	.0000		.0000	.0000	.0002	.0000		.0000	.0001
(T3e)	M3b	.0000	.0000		.0000	.0000	.0002	.0000		.0000	.0001
	M4	.0000			.0000	.0000	.0002			.0000	.0001
	M5	.0000	.0000		.0000	.0000	.0002	.0000		.0000	.0001
	M6c	.0000			.0000	.0000	.0002			.0000	.0001
UT2s	M3a	.1711	.9996		.4953	.1599	.3351	.+		.8054	.3370
(Pwr)	M3b	.1711	.9996		.4953	.1599	.3351	.+		.8054	.3370
	M4	.1711			.5269	.1896	.3351			.8259	.3792
	M5	.1711	.9993		.5204	.1734	.3351	.+		.8211	.3589
	M6c	.1711			.4922	.1526	.3351			.8030	.3291

.+ = 1.0000

Table 4.3.2-18

Rosenbaum's Test for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 0.8, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0001	.0000		.0000	.0001
(T3e)	M3b	.0000	.0000		.0000	.0000	.0001	.0000		.0000	.0001
	M4	.0000			.0000	.0000	.0001			.0000	.0001
	M5	.0000	.0000		.0000	.0000	.0001	.0000		.0000	.0001
	M6c	.0000			.0000	.0000	.0001			.0000	.0001
UT1s	M3a	.2019	.+		.7028	.2036	.3668	.+		.9147	.3875
(Pwr)	M3b	.2019	.+		.7028	.2036	.3668	.+		.9147	.6875
	M4	.2019			.7312	.2395	.3668			.9261	.4337
	M5	.2019	.+		.7222	.2181	.3668	.+		.9221	.4084
	M6c	.2019			.6982	.1949	.3668			.9124	.3789
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0001
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M6c	.0000			.0000	.0000	.0000			.0000	.0000
UT2s	M3a	.1443	.+		.5566	.1404	.2755	.+		.8261	.2861
(Pwr)	M3b	.1443	.+		.5566	.1404	.2755	.+		.8261	.2861
	M4	.1443			.5907	.1692	.2755			.8461	.3281
	M5	.1443	.+		.5806	.1514	.2755	.+		.8397	.3036
	M6c	.1443			.5516	.1326	.2755			.8225	.2769

.+ = 1.0000

Table 4.3.2-19
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0002	.0000	.0002	.0000		.0015	.0001
(T3e)	M3b	.0000	.0000		.0002	.0000	.0002	.0000		.0015	.0001
	M4	.0000			.0003	.0000	.0002			.0017	.0002
	M5	.0000	.0000	.0004	.0003	.0000	.0002	.0001	.0015	.0017	.0002
	M6c	.0000			.0002	.0000	.0002			.0015	.0001
UT1s	M3a	.1994	.4438		.1019	.1709	.3803	.7639		.2487	.3439
(Pwr)	M3b	.1994	.4438		.1019	.1709	.3803	.7639		.2487	.3439
	M4	.1994			.1085	.1976	.3803			.2609	.3802
	M5	.1994	.4736	.1554	.1085	.1966	.3803	.7834	.4146	.2607	.3761
	M6c	.1994			.1018	.1698	.3803			.2486	.3420
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0002	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000	.0000		.0002	.0000
	M4	.0000			.0000	.0000	.0000			.0003	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0004	.0003	.0000
	M6c	.0000			.0000	.0000	.0000			.0002	.0000
UT2s	M3a	.0646	.1300		.0364	.0507	.1994	.4438		.1019	.1709
(Pwr)	M3b	.0646	.1300		.0364	.0507	.1994	.4438		.1019	.1709
	M4	.0646			.0396	.0627	.1994			.1085	.1976
	M5	.0646	.1478	.0335	.0367	.0638	.1994	.4736	.1554	.1085	.1966
	M6c	.0646			.0364	.0507	.1994			.1018	.1698

Table 4.3.2-20
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0001	.0000		.0003	.0001
(T3e)	M3b	.0000	.0000		.0000	.0000	.0001	.0000		.0003	.0001
	M4	.0000			.0000	.0000	.0001			.0003	.0001
	M5	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0001	.0003	.0001
	M6c	.0000			.0000	.0000	.0001			.0003	.0001
UT1s	M3a	.3232	.9627		.2817	.2945	.5786	.9985		.7054	.5651
(Pwr)	M3b	.3232	.9627		.2817	.2945	.5786	.9985		.7054	.5651
	M4	.3232			.2974	.3323	.5786			.7054	.6047
	M5	.3232	.9645	.6600	.2973	.3193	.5786	.9985	.9422	.7054	.5905
	M6c	.3232			.2812	.2896	.5786			.7054	.5612
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
	M4	.0000			.0000	.0000	.0000			.0001	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0000
	M6c	.0000			.0000	.0000	.0000			.0001	.0000
UT2s	M3a	.2164	.8833		.1347	.1899	.4460	.9911		.4873	.4220
(Pwr)	M3b	.2164	.8833		.1347	.1899	.4460	.9911		.4873	.4220
	M4	.2164			.1453	.2207	.4460			.5056	.4630
	M5	.2164	.8901	.4376	.1452	.2105	.4460	.9914	.8382	.5054	.4481
	M6c	.2164			.1344	.1858	.4460			.4868	.4172

Table 4.3.2-21
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6c	.0000			.0000	.0000	.0000			.0000	.0000
UT1s	M3a	.4414	.9998		.6648	.4334	.5481	.+		.8202	.5522
(Pwr)	M3b	.4414	.9998		.6648	.4334	.5481	.+		.8202	.5522
	M4	.4414			.6828	.4775	.5481			.8326	.5954
	M5	.4414	.9997	.9619	.6815	.4549	.5481	.+	.9882	.8317	.5739
	M6c	.4414			.6636	.4269	.5481			.8194	.5469
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6c	.0000			.0000	.0000	.0000			.0000	.0000
UT2s	M3a	.3422	.9987		.4808	.3264	.5481	.+		.8202	.5522
(Pwr)	M3b	.3422	.9987		.4808	.3264	.5481	.+		.8202	.5522
	M4	.3422			.5010	.3684	.5481			.8326	.5954
	M5	.3422	.9985	.9015	.5000	.3465	.5481	.+	.9882	.8317	.5739
	M6c	.3422			.4795	.3194	.5481			.8194	.5469

.+ = 1.0000

Table 4.3.2-22
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6c	.0000			.0000	.0000	.0000			.0000	.0000
UT1s	M3a	.4216	.+		.7806	.4292	.6086	.+		.9528	.6369
(Pwr)	M3b	.4216	.+		.7806	.4292	.6086	.+		.9528	.6369
	M4	.4216			.7955	.4764	.6086			.9573	.6832
	M5	.4216	.+	.9914	.7939	.4470	.6086	.+	.9994	.9568	.6576
	M6c	.4216			.7789	.4216	.6086			.9522	.6346
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6c	.0000			.0000	.0000	.0000			.0000	.0000
UT2s	M3a	.4216	.+		.7806	.4292	.6086	.+		.9528	.6396
(Pwr)	M3b	.4216	.+		.7806	.4292	.6086	.+		.9528	.6396
	M4	.4216			.7955	.4764	.6086			.9573	.6832
	M5	.4216	.+	.9914	.7939	.4470	.6086	.+	.9994	.9568	.6579
	M6c	.4216			.7789	.4216	.6086			.9522	.6346

.+ = 1.0000

Table 4.3.2-23
Rosenbaum's Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6c	.0000			.0000	.0000	.0000			.0000	.0000
UT1s	M3a	.4762	.+		.9282	.5072	.6480	.+		.9896	.7048
(Pwr)	M3b	.4762	.+		.9282	.5072	.6480	.+		.9896	.7048
	M4	.4762			.9347	.5579	.6480			.9908	.7482
	M5	.4762	.+	.9995	.9335	.5232	.6480	.+	.+	.9906	.7212
	M6c	.4762			.9270	.4998	.6480			.9893	.7003
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6c	.0000			.0000	.0000	.0000			.0000	.0000
UT2s	M3a	.3970	.+		.8557	.4159	.5605	.+		.9699	.6048
(Pwr)	M3b	.3970	.+		.8557	.4159	.5605	.+		.9699	.6048
	M4	.3970			.8670	.4666	.5605			.9731	.6533
	M5	.3970	.+	.9980	.8654	.4306	.5605	.+	.+	.9725	.6214
	M6c	.3970			.8538	.4075	.5605			.9693	.5986

.+ = 1.0000

Table 4.3.2-24
Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .01, 1-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	2.5		3.9	3.5	2.5		4.0	3.9	2.5		3.3	4.0
	M3b	2.5		3.9	3.5	2.5		4.0	3.9	2.5		3.3	4.0
	M4			1.5	1.5			1.5	1.3			2.6	1.5
	M5	1.0		1.5	1.5	1.0		1.5	1.8	1.0		2.6	1.5
	M6c			4.3	5.0			4.0	4.3			3.3	4.0
12	M3a	2.5		3.6	3.5	2.5		3.9	3.5	2.3		4.0	4.0
	M3b	2.5		3.6	3.5	2.5		3.9	3.5	2.3		4.0	4.0
	M4			1.0	1.0			1.4	1.1			1.5	1.3
	M5	1.0		2.0	2.0	1.0		1.6	1.9	1.5		1.5	1.8
	M6c			4.8	5.0			4.3	5.0			4.0	4.0
18	M3a	1.5		3.5	3.5	1.6		3.6	3.5	1.8		4.0	3.9
	M3b	1.5		3.5	3.5	1.6		3.6	3.5	1.8		4.0	3.9
	M4			1.0	1.0			1.0	1.0			1.4	1.1
	M5	3.0		2.0	2.0	2.8		2.0	2.0	2.5		1.6	2.3
	M6c			5.0	5.0			4.8	5.0			4.0	3.9
24	M3a	1.6		3.5	3.5	1.6		3.5	3.5	1.8		3.5	3.6
	M3b	1.6		3.5	3.5	1.6		3.5	3.5	1.8		3.5	3.6
	M4			1.0	1.0			1.0	1.0			1.5	1.1
	M5	2.8		2.0	2.0	2.8		2.0	2.0	2.5		2.3	2.3
	M6c			5.0	5.0			5.0	5.0			4.3	4.4
30	M3a	1.8		3.8	3.5	1.8		3.6	3.5	1.8		3.5	3.4
	M3b	1.8		3.8	3.5	1.8		3.6	3.5	1.8		3.5	3.4
	M4			1.0	1.0			1.0	1.0			1.5	1.4
	M5	2.5		2.0	2.0	2.5		2.0	2.0	2.5		2.6	2.1
	M6c			4.5	5.0			4.8	5.0			3.9	4.8

Table 4.3.2-25

Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .01, 2-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	2.5		4.0	4.0	2.5		3.5	4.0	2.4		3.0	3.5
	M3b	2.5		4.0	4.0	2.5		3.5	4.0	2.4		3.0	3.5
	M4			1.4	1.9			2.1	1.6			3.0	2.3
	M5	1.0		1.6	1.1	1.0		2.4	1.4	1.3		3.0	2.3
	M6c			4.0	4.0			3.5	4.0			3.0	3.5
12	M3a	2.5		3.6	3.5	2.5		3.6	3.6	2.5		3.5	3.6
	M3b	2.5		3.6	3.5	2.5		3.6	3.6	2.5		3.5	3.6
	M4			1.3	1.0			1.6	1.0			2.3	1.6
	M5	1.0		1.8	2.0	1.0		1.4	2.0	1.0		2.3	1.9
	M6c			4.8	5.0			4.8	4.8			3.5	4.3
18	M3a	1.8		3.5	3.5	1.9		3.6	3.5	1.9		3.8	3.6
	M3b	1.8		3.5	3.5	1.9		3.6	3.5	1.9		3.8	3.6
	M4			1.0	1.0			1.1	1.0			1.9	1.6
	M5	2.5		2.0	2.0	2.3		1.9	2.0	2.3		1.9	2.1
	M6c			5.0	5.0			4.8	5.0			3.8	4.0
24	M3a	1.6		3.5	3.5	1.6		3.5	3.5	1.8		3.5	3.6
	M3b	1.6		3.5	3.5	1.6		3.5	3.5	1.8		3.5	3.6
	M4			1.0	1.0			1.0	1.0			1.5	1.1
	M5	2.8		2.0	2.0	2.8		2.0	2.0	2.5		2.3	2.3
	M6c			5.0	5.0			5.0	5.0			4.3	4.4
30	M3a	1.8		3.5	3.5	1.8		3.6	3.5	1.8		3.5	3.5
	M3b	1.8		3.5	3.5	1.8		3.6	3.5	1.8		3.5	3.5
	M4			1.0	1.0			1.0	1.0			1.8	1.6
	M5	2.5		2.0	2.0	2.5		2.0	2.0	2.5		2.0	2.1
	M6c			5.0	5.0			4.8	5.0			4.3	4.3

Table 4.3.2-26
Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	2.5		3.8	3.5	2.5		4.0	3.6	2.5		3.8	3.9
	M3b	2.5		3.8	3.5	2.5		4.0	3.6	2.5		3.8	3.9
	M4			1.1	1.0			1.5	1.1			1.9	1.4
	M5	1.0		1.9	2.0	1.0		1.5	1.9	1.0		1.9	1.6
	M6c			4.5	5.0			4.0	4.8			3.8	4.3
12	M3a	1.9		3.5	3.5	2.0		3.8	3.5	2.0		3.9	3.8
	M3b	1.9		3.5	3.5	2.0		3.8	3.5	2.0		3.9	3.8
	M4			1.0	1.0			1.0	1.0			1.5	1.3
	M5	2.3		2.0	2.0	2.0		2.0	2.0	2.0		1.5	1.8
	M6c			5.0	5.0			4.5	5.0			4.3	4.5
18	M3a	1.6		3.5	3.5	1.8		3.6	3.5	1.8		4.0	3.9
	M3b	1.6		3.5	3.5	1.8		3.6	3.5	1.8		4.0	3.9
	M4			1.0	1.0			1.0	1.0			1.4	1.1
	M5	2.8		2.0	2.0	2.5		2.0	2.0	2.5		1.6	1.9
	M6c			5.0	5.0			4.8	5.0			4.0	4.3
24	M3a	1.8		3.5	3.5	1.8		3.6	3.5	1.8		3.9	3.8
	M3b	1.8		3.5	3.5	1.8		3.6	3.5	1.8		3.9	3.8
	M4			1.0	1.0			1.1	1.0			1.3	1.3
	M5	2.5		2.0	2.0	2.5		1.9	2.0	2.5		1.8	1.8
	M6c			5.0	5.0			4.8	5.0			4.3	4.5
30	M3a	1.8		3.5	3.5	1.9		3.5	3.5	1.9		3.6	3.6
	M3b	1.8		3.5	3.5	1.9		3.5	3.5	1.9		3.6	3.6
	M4			1.0	1.0			1.0	1.0			1.5	1.1
	M5	2.5		2.0	2.0	2.3		2.0	2.0	2.3		2.3	1.9
	M6c			5.0	5.0			5.0	5.0			4.0	4.8

Table 4.3.2-27

Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	2.5		3.9	3.5	2.5		3.9	3.9	2.5		3.3	4.0
	M3b	2.5		3.9	3.5	2.5		3.9	3.9	2.5		3.3	4.0
	M4			1.4	1.6			1.5	1.6			2.6	1.5
	M5	1.0		1.6	1.4	1.0		1.5	1.4	1.0		2.6	1.5
	M6c			4.3	5.0			4.3	4.3			3.3	4.0
12	M3a	2.5		3.5	3.5	2.4		3.9	3.5	2.1		3.9	3.8
	M3b	2.5		3.5	3.5	2.4		3.9	3.5	2.1		3.9	3.8
	M4			1.0	1.0			1.1	1.0			1.5	1.3
	M5	1.0		2.0	2.0	1.3		1.9	2.0	1.8		1.5	1.8
	M6c			5.0	5.0			4.3	5.0			4.3	4.5
18	M3a	1.6		3.5	3.5	1.8		3.6	3.5	1.8		4.0	3.9
	M3b	1.6		3.5	3.5	1.8		3.6	3.5	1.8		4.0	3.9
	M4			1.0	1.0			1.0	1.0			1.4	1.1
	M5	2.8		2.0	2.0	2.5		2.0	2.0	2.5		1.6	1.9
	M6c			5.0	5.0			4.8	5.0			4.0	4.3
24	M3a	1.8		3.5	3.5	1.8		3.5	3.5	1.8		3.9	3.8
	M3b	1.8		3.5	3.5	1.8		3.5	3.5	1.8		3.9	3.8
	M4			1.0	1.0			1.1	1.0			1.3	1.3
	M5	2.5		2.0	2.0	2.5		1.9	2.0	2.5		1.8	1.8
	M6c			5.0	5.0			5.0	5.0			4.3	4.5
30	M3a	1.8		3.5	3.5	1.8		3.5	3.5	1.9		3.6	3.6
	M3b	1.8		3.5	3.5	1.8		3.5	3.5	1.9		3.6	3.6
	M4			1.0	1.0			1.1	1.0			1.6	1.1
	M5	2.5		2.0	2.0	2.5		1.9	2.0	2.3		2.1	2.3
	M6c			5.0	5.0			5.0	5.0			4.0	4.4

Table 4.3.2-28
Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .01, 1-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	1.9		3.9	3.5	1.9		3.9	3.6	1.9		3.5	3.5
	M3b	1.9		3.9	3.5	1.9		3.9	3.6	1.9		3.5	3.5
	M4			1.1	1.1			1.2	1.2			2.0	1.5
	M5	2.2		1.9	1.9	2.2		1.8	1.8	2.2		2.5	2.5
	M6c			4.2	5.0			4.2	4.8			3.5	4.0
0.5	M3a	1.9		3.6	3.5	1.9		3.7	3.6	1.9		3.6	4.0
	M3b	1.9		3.6	3.5	1.9		3.7	3.6	1.9		3.6	4.0
	M4			1.2	1.2			1.2	1.0			1.5	1.4
	M5	2.2		1.8	1.8	2.2		1.8	2.0	2.2		2.1	1.6
	M6c			4.8	5.0			4.6	4.8			4.2	4.0
0.8	M3a	2.0		3.6	3.5	2.0		3.7	3.6	2.1		3.7	3.8
	M3b	2.0		3.6	3.5	2.0		3.7	3.6	2.1		3.7	3.8
	M4			1.0	1.1			1.1	1.1			1.6	1.1
	M5	2.0		2.0	1.9	2.0		1.9	1.9	1.8		2.0	1.9
	M6c			4.8	5.0			4.6	4.8			4.0	4.4
1.2	M3a	2.1		3.5	3.5	2.2		3.6	3.5	2.1		3.8	3.8
	M3b	2.1		3.5	3.5	2.2		3.6	3.5	2.1		3.8	3.8
	M4			1.1	1.0			1.2	1.0			1.7	1.1
	M5	1.8		1.9	2.0	1.6		1.8	2.0	1.8		1.9	1.9
	M6c			5.0	5.0			4.8	5.0			3.8	4.4

Table 4.3.2-29
Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .01, 2-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.1		3.7	3.6	2.1		3.7	3.7	1.9		3.2	3.0
	3b	2.1		3.7	3.6	2.1		3.7	3.7	1.9		3.2	3.0
	M4			1.3	1.2			1.7	1.1			2.7	2.5
	M5	1.8		1.7	1.8	1.8		1.9	1.9	2.2		2.7	3.0
	M6c			4.6	4.8			4.0	4.6			3.2	3.5
0.5	M3a	1.9		3.6	3.6	1.9		3.4	3.6	1.9		3.5	4.0
	M3b	1.9		3.6	3.6	1.9		3.4	3.6	1.9		3.5	4.0
	M4			1.1	1.2			1.5	1.1			2.0	1.5
	M5	2.2		1.9	1.8	2.2		2.1	1.9	2.2		2.2	1.5
	M6c			4.8	4.8			4.6	4.8			3.8	4.0
0.8	M3a	2.0		3.6	3.6	2.0		3.6	3.6	2.2		3.4	3.6
	M3b	2.0		3.6	3.6	2.0		3.6	3.6	2.2		3.4	3.6
	M4			1.1	1.1			1.2	1.1			1.9	1.5
	M5	2.0		1.9	1.9	2.0		1.8	1.9	1.9		2.3	2.1
	M6c			4.8	4.8			4.8	4.8			4.0	4.2
1.2	M3a	2.1		3.6	3.6	2.0		3.6	3.6	2.2		3.7	3.7
	M3b	2.1		3.6	3.6	2.2		3.6	3.6	2.2		3.7	3.7
	M4			1.0	1.2			1.1	1.2			1.7	1.1
	M5	1.8		2.0	1.8	1.6		1.9	1.8	1.9		1.9	1.9
	M6c			4.8	4.8			4.8	4.8			4.0	4.6

Table 4.3.2-30

Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	1.9		3.6	3.5	1.9		3.8	3.6	1.9		4.0	3.9
	M3b	1.9		3.6	3.5	1.9		3.8	3.6	1.9		4.0	3.9
	M4			1.1	1.0			1.1	1.1			1.4	1.5
	M5	2.2		1.9	2.0	2.2		1.9	1.9	2.2		1.6	1.5
	M6c			4.8	5.0			4.4	4.8			4.0	4.2
0.5	M3a	1.7		3.6	3.5	1.8		3.6	3.5	1.8		3.7	3.8
	M3b	1.7		3.6	3.5	1.8		3.6	3.5	1.8		3.7	3.8
	M4			1.0	1.0			1.1	1.0			1.5	1.3
	M5	2.6		2.0	2.0	2.4		1.9	2.0	2.4		2.1	1.7
	M6c			4.8	5.0			4.8	5.0			4.0	4.4
0.8	M3a	1.9		3.5	3.5	2.1		3.6	3.5	2.1		3.9	3.6
	M3b	1.9		3.5	3.5	2.1		3.6	3.5	2.1		3.9	3.6
	M4			1.0	1.0			1.1	1.0			1.3	1.0
	M5	2.2		2.0	2.0	1.8		1.9	2.0	1.8		1.7	2.0
	M6c			5.0	5.0			4.8	5.0			4.2	4.8
1.2	M3a	2.1		3.5	3.5	2.1		3.8	3.5	2.1		3.7	3.8
	M3b	2.1		3.5	3.5	2.1		3.8	3.5	2.1		3.7	3.8
	M4			1.0	1.0			1.2	1.0			1.8	1.1
	M5	1.8		2.0	2.0	1.8		1.8	2.0	1.8		1.8	1.9
	M6c			5.0	5.0			4.4	5.0			4.0	4.4

Table 4.3.2-31
Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	1.9		3.6	3.5	1.9		3.7	3.6	1.9		3.8	3.9
	M3b	1.9		3.6	3.5	1.9		3.7	3.6	1.9		3.8	3.9
	M4			1.1	1.2			1.2	1.2			1.8	1.4
	M5	2.2		1.9	1.8	2.2		1.8	1.8	2.2		1.8	1.9
	M6c			4.8	5.0			4.6	4.8			3.8	3.9
0.5	M3a	1.9		3.6	3.5	1.9		3.7	3.6	1.9		3.7	3.8
	M3b	1.9		3.6	3.5	1.9		3.7	3.6	1.9		3.7	3.8
	M4			1.2	1.2			1.1	1.2			1.5	1.4
	M5	2.2		1.8	1.8	2.2		1.9	1.8	2.2		2.1	1.6
	M6c			4.8	5.0			4.6	4.8			4.0	4.4
0.8	M3a	2.1		3.6	3.5	2.2		3.7	3.6	2.1		3.6	3.6
	M3b	2.1		3.6	3.5	2.2		3.7	3.6	2.1		3.6	3.6
	M4			1.0	1.1			1.1	1.1			1.6	1.1
	M5	1.8		2.0	1.9	1.6		1.9	1.9	1.8		2.0	1.9
	M6c			4.8	5.0			4.6	4.8			4.2	4.8
1.2	M3a	2.2		3.5	3.5	2.1		3.6	3.5	2.1		3.8	3.9
	M3b	2.2		3.5	3.5	2.1		3.6	3.5	2.1		3.8	3.9
	M4			1.0	1.0			1.3	1.0			1.8	1.1
	M5	1.6		2.0	2.0	1.8		1.7	2.0	1.8		1.8	1.9
	M6c			5.0	5.0			4.8	5.0			3.8	4.2

Table 4.3.2-32
Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .01, 1-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.0		3.7	3.5	2.0		3.7	3.6	2.0		3.7	3.8
M3b	2.0		3.7	3.5	2.0		3.7	3.6	2.0		3.7	3.8
M4			1.1	1.1			1.2	1.1			1.7	1.3
M5	2.1		1.9	1.9	2.0		1.8	1.9	2.0		2.1	2.0
M6c			4.7	5.0			4.6	4.9			3.9	4.2

Table 4.3.2-33
Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .01, 2-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.0		3.6	3.6	2.1		3.6	3.6	2.1		3.5	3.6
M3b	2.0		3.6	3.6	2.1		3.6	3.6	2.1		3.5	3.6
M4			1.1	1.2			1.4	1.1			2.1	1.7
M5	2.0		1.9	1.8	1.9		1.9	1.9	1.9		2.3	2.1
M6c			4.8	4.8			4.6	4.8			3.8	4.1

Table 4.3.2-34
Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	1.9		3.6	3.5	2.0		3.7	3.5	2.0		3.8	3.8
M3b	1.9		3.6	3.5	2.0		3.7	3.5	2.0		3.8	3.8
M4			1.0	1.0			1.1	2.0			1.5	1.2
M5	2.2		2.0	2.0	2.1		1.9	2.0	2.1		1.8	1.8
M6c			4.9	5.0			4.6	5.0			4.1	4.5

Table 4.3.2-35
Rosenbaum's Test for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.0		3.6	3.5	2.0		3.7	3.6	2.0		3.7	3.8
M3b	2.0		3.6	3.5	2.0		3.7	3.6	2.0		3.7	3.8
M4			1.1	1.1			1.2	1.1			1.7	1.3
M5	2.0		1.9	1.9	2.0		1.8	1.9	2.0		1.9	1.8
M6c			4.9	5.0			4.7	4.9			4.0	4.3

Table 4.3.2-36

Rosenbaum's Test for Two Groups
Analysis of Mean Ranks of Power Results
Number of First Place Finishes by Distribution
Across Nominal Alpha, Direction and Number of Groups

Decimal	4th				2nd			
Method	3a*	4	5	6b	3a*	4	5	6b
By Initial Sample Size Across Nominal Effect Size Multiplier								
EA	6.50		7.00		6.33		7.33	
EB								
ML	0.00	19.50	0.50	0.00	0.20	15.20	4.20	0.20
SS	0.00	17.50	2.50	0.00	0.00	18.50	1.50	0.00
By Nominal Effect Size Multiplier Across Initial Sample Size								
EA	4.67		6.67		4.00		8.00	
EB								
ML	0.00	16.00	0.00	0.00	0.00	14.00	2.00	0.00
SS	0.00	16.00	0.00	0.00	0.00	15.00	1.00	0.00
Across Nominal Effect Size Multiplier and Initial Sample Size								
EA	1.67		0.67		1.17		1.67	
EB								
ML	0.00	4.00	0.00	0.00	0.00	4.00	0.00	0.00
SS	0.00	4.00	0.00	0.00	0.00	4.00	0.00	0.00

* Note: Results for method 3b were identical to method 3a and are not shown.

Table 4.3.2-37

Rosenbaum's Test for Two Groups
Analysis of Mean Ranks of Power Results
Across Nominal Alpha, Direction and Distributions

Decimal	4th				2nd			
Method	3a/b*	4	5	6c	3a/b*	4	5	6c
By Initial Sample Size Across Nominal Effect Size Multiplier								
MP1i	60	40	60	40	60	40	60	40
N1Mi	16.5	37.0	10.0	0.0	6.53	33.70	13.03	0.20
PoM	0.108	0.925	0.167	0.000	0.109	0.843	0.217	0.005
PoT*	0.108	0.617	0.167	0.000	0.109	0.562	0.217	0.003
By Nominal Effect Size Multiplier Across Initial Sample Size								
MP1i	48	32	48	32	48	32	48	32
N1Mi	4.67	32.00	6.67	0.00	4.0	29.0	11.0	0.0
PoM	0.097	1.000	0.139	0.000	0.083	0.906	0.229	0.000
PoT*	0.097	0.667	0.139	0.000	0.083	0.604	0.229	0.000
Across Nominal Effect Size Multiplier and Initial Sample Size								
MP1i	12	8	12	8	12	8	12	8
N1Mi	1.67	8.00	0.67	0.00	1.17	8.00	1.67	0.00
PoM	0.139	1.000	0.056	0.000	0.097	1.000	0.139	0.000
PoT*	0.139	0.667	0.056	0.000	0.097	0.667	0.139	0.000

Key: MP1i: Maximum possible 1st place finishes for each method
N1Mi: Actual number of 1st place finishes for each method (ties count 1/n)
PoM: Proportion of Maximum Possible { N1Mi / MP1i }
PoT: Proportion of total 1st place finishes { N1Mi / Σ(N1Mi) }

* Note: Results for method 3b were identical to method 3a and are not shown but are included in the total 1st place finishes Σ(N1Mi) for calculating PoT.

4.3.3 – Tukey's Quick Test (two groups only)

Based on the Type I error results, power and Type III error results for Tukey's Quick Test are only presented for the combinations of alpha level, distribution and method shown in Table 4.3.3-1.

Table 4.3.3-1
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Method / Distribution Combinations with Acceptable Type I Error

Alpha	.01					.05				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Method										
M3a	x			x	x	x			x	x
M3b	x	x		x	x	x			x	x
M4	x			x	x	x			x	x
M5	x	x	x	x	x	x	x	x	x	x

Power results for Tukey's Quick Test at nominal alpha .01 and .05 are summarized in Table 4.3.3-2 and 4.3.3-3. These tables present the range of values obtained (minimum and maximum) for 1-tailed power, 2-tailed Type III error and 2-tailed power based on the results presented in Tables 4.3.3-4 through 4.3.3-23.

Power tended to increase monotonically across methods and distributions with increases in initial sample size and/or effect size. Likewise, Type III error tended to decrease monotonically across methods and distributions with an increase in initial sample size and/or effect size. Thus, minimum power (or maximum Type III error) usually occurred at initial sample size 6 or 12 at nominal

ESM 0.2, with maximum power (or minimum Type III error) at initial sample size 30 and nominal ESM 1.2, if not sooner.

Table 4.3.3-2
Tukey's Quick Test for Two Groups
Range of Power and Type III Error for α .01

Min/Max	min					max				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Type										
Mthd										
Pwr 1-s										
M3a	.0180			.0152	.0151	.7410			.9999	.8380
M3b	.0180	.0197		.0152	.0152	.7410			.9999	.8455
M4	.0180			.0170	.0172	.7410			.9999	.8456
M5	.0180	.0240	.0819	.0171	.0178	.7410	.9291	.+	.9999	.8458
T3e 2-s										
M3a	.0000			.0000	.0000	.0018			.0019	.0014
M3b	.0000	.0000		.0000	.0000	.0018			.0019	.0014
M4	.0000			.0000	.0000	.0018			.0021	.0017
M5	.0000	.0000	.0000	.0000	.0000	.0018	.0016	.0050	.0021	.0018
Pwr 2-s										
M3a	.0098			.0075	.0081	.6893			.9998	.7853
M3b	.0098	.0101		.0075	.0081	.6893			.9998	.7914
M4	.0098			.0084	.0095	.6893			.9997	.7930
M5	.0098	.0123	.0373	.0084	.0097	.6893	.9291	.+	.9998	.7934

.+ = 1.0000

Tables 4.3.3-4 through 4.3.3-23 give the upper and lower tail results for Tukey's Quick Test for both 1-sided and 2-sided tests for both alpha .01 and .05. There is a table for each combination of nominal effect size multiplier {0.2, 0.5, 0.8, 1.2} and initial sample size {6, 12, 18, 24, 30}. Results are not reported for distribution EB with nominal ESM 0.2 as the actual ESM = 0.0 (no shift), which is just Type I error. Also for distribution EB, results are not reported for nominal

ESM 0.8 as the actual ESM = 0.592, the same as the actual ESM for nominal ESM of 0.5. Results for the normal distribution are included and are essentially identical across methods for a fixed initial sample size, effect size and directionality. This is to be expected in the absence of ties and demonstrates that the simulations worked correctly.

Table 4.3.3-3
Tukey's Quick Test for Two Groups
Range of Power and Type III Error for α .05

Min/Max	min					max				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Type / *Mthd*										
Pwr 1-s										
M3a	.0654			.0591	.0562	.8585			.+	.9461
M3b	.0654			.0592	.0564	.8585			.+	.9546
M4	.0654			.0655	.0646	.8585			.+	.9503
M5	.0654	.0871	.3121	.0659	.0652	.8585	.9291	.+	.+	.9493
T3e 2-s										
M3a	.0000			.0000	.0000	.0079			.0072	.0065
M3b	.0000			.0000	.0000	.0079			.0072	.0065
M4	.0000			.0000	.0000	.0079			.0082	.0077
M5	.0000	.0000	.0000	.0000	.0000	.0079	.0061	.0012	.0083	.0079
Pwr 2-s										
M3a	.0351			.0302	.0298	.7869			.+	.8826
M3b	.0351			.0302	.0298	.7869			.+	.8910
M4	.0351			.0328	.0345	.7869			.+	.8898
M5	.0351	.0458	.1690	.0339	.0350	.7869	.9291	.+	.+	.8896

.+ = 1.0000

A ranking analysis of the results is presented in Tables 4.3.3-24 through 4.3.3-35. These summaries were obtained by ranking the power results from

Tables 4.3.3-4 through 4.3.3-23 to four decimal places (as reported), to three decimal places and to two decimal places.

Table 4.3.3-4
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0038			.0037	.0030	.0168			.0150	.0137
	M3b	.0038	.0025		.0037	.0030	.0168			.0150	.0138
	M4	.0038			.0041	.0037	.0168			.0169	.0163
	M5	.0038	.0031		.0042	.0038	.0168	.0122		.0170	.0167
UT1s	M3a	.0187			.0152	.0156	.0654			.0591	.0562
(Pwr)	M3b	.0187	.0197		.0152	.0156	.0654			.0592	.0564
	M4	.0187			.0170	.0182	.0654			.0655	.0646
	M5	.0187	.0240		.0171	.0185	.0654	.0871		.0659	.0652
LT2s	M3a	.0018			.0019	.0014	.0079			.0072	.0065
(T3e)	M3b	.0018	.0013		.0019	.0014	.0079			.0072	.0065
	M4	.0018			.0021	.0017	.0079			.0082	.0077
	M5	.0018	.0016		.0021	.0018	.0079	.0061		.0083	.0079
UT2s	M3a	.0098			.0075	.0081	.0351			.0302	.0298
(Pwr)	M3b	.0098	.0101		.0075	.0081	.0351			.0302	.0298
	M4	.0098			.0084	.0095	.0351			.0328	.0345
	M5	.0098	.0123		.0084	.0097	.0351	.0458		.0339	.0350

Table 4.3.3-5

Tukey's Quick Test for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 0.2, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0017			.0012	.0015	.0163			.0105	.0127
	M3b	.0017	.0006		.0012	.0015	.0163			.0110	.0135
	M4	.0017			.0014	.0017	.0163			.0126	.0155
	M5	.0017	.0007		.0015	.0018	.0163	.0049		.0127	.0160
UT1s	M3a	.0180			.0184	.0151	.0913			.0999	.0789
(Pwr)	M3b	.0180	.0322		.0185	.0152	.0913			.1028	.0832
	M4	.0180			.0208	.0172	.0913			.1119	.0910
	M5	.0180	.0363		.0213	.0178	.0913	.1851		.1132	.0921
LT2s	M3a	.0008			.0006	.0007	.0078			.0054	.0063
(T3e)	M3b	.0008	.0003		.0006	.0007	.0078			.0054	.0065
	M4	.0008			.0007	.0008	.0078			.0064	.0075
	M5	.0008	.0003		.0008	.0008	.0078	.0027		.0064	.0078
UT2s	M3a	.0099			.0100	.0083	.0549			.0591	.0472
(Pwr)	M3b	.0099	.0171		.0100	.0083	.0549			.0602	.0488
	M4	.0099			.0117	.0095	.0549			.0664	.0540
	M5	.0099	.0196		.0116	.0098	.0549	.1132		.0674	.0551

Table 4.3.3-6
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0016			.0009	.0013	.0151			.0079	.0113
	M3b	.0016	.0003		.0010	.0013	.0151			.0085	.0124
	M4	.0016			.0012	.0015	.0151			.0098	.0140
	M5	.0016	.0003		.0012	.0016	.0151	.0017		.0098	.0144
UT1s	M3a	.0235			.0280	.0201	.1032			.1265	.0922
(Pwr)	M3b	.0235	.0817		.0285	.0207	.1032			.1324	.0993
	M4	.0235			.0316	.0229	.1032			.1430	.1058
	M5	.0235	.0777		.0324	.0237	.1032	.2865		.1436	.1076
LT2s	M3a	.0007			.0004	.0006	.0072			.0039	.0055
(T3e)	M3b	.0007	.0001		.0004	.0006	.0072			.0041	.0059
	M4	.0007			.0006	.0007	.0072			.0049	.0068
	M5	.0007	.0001		.0006	.0007	.0072	.0009		.0049	.0071
UT2s	M3a	.0138			.0159	.0116	.0653			.0788	.0574
(Pwr)	M3b	.0138	.0464		.0161	.0118	.0653			.0816	.0607
	M4	.0138			.0181	.0130	.0653			.0889	.0658
	M5	.0138	.0455		.0186	.0136	.0653	.1972		.0900	.0671

Table 4.3.3-7

Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0016			.0007	.0012	.0144			.0058	.0100
	M3b	.0016	.0001		.0008	.0012	.0144			.0066	.0114
	M4	.0016			.0009	.0014	.0144			.0076	.0127
	M5	.0016	.0001		.0009	.0015	.0144	.0006		.0076	.0132
UT1s	M3a	.0269			.0365	.0245	.1109			.1539	.1022
(Pwr)	M3b	.0269	.1765		.0375	.0255	.1109			.1617	.1116
	M4	.0269			.0413	.0274	.1109			.1749	.1166
	M5	.0269	.1356		.0419	.0284	.1109	.3796		.1750	.1186
LT2s	M3a	.0007			.0003	.0005	.0070			.0030	.0049
(T3e)	M3b	.0007	.0001		.0004	.0006	.0070			.0032	.0049
	M4	.0007			.0004	.0006	.0070			.0038	.0062
	M5	.0007	.0001		.0004	.0007	.0070	.0004		.0038	.0064
UT2s	M3a	.0160			.0216	.0145	.0715			.0976	.0653
(Pwr)	M3b	.0160	.1094		.0221	.0150	.0715			.1017	.0700
	M4	.0160			.0244	.0161	.0715			.1111	.0741
	M5	.0160	.0855		.0250	.0167	.0715	.2879		.1115	.0757

Table 4.3.3-8
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0015			.0005	.0010	.0137			.0042	.0092
	M3b	.0015	.0001		.0006	.0011	.0137			.0050	.0105
	M4	.0015			.0007	.0012	.0137			.0056	.0114
	M5	.0015	.0001		.0007	.0013	.0137	.0003		.0057	.0120
UT1s	M3a	.0305			.0458	.0281	.1168			.1862	.1104
(Pwr)	M3b	.0305	.3129		.0472	.0295	.1168			.1958	.1217
	M4	.0305			.0523	.0314	.1168			.2124	.1257
	M5	.0305	.2075		.0529	.0323	.1168	.4567		.2116	.1278
LT2s	M3a	.0007			.0003	.0005	.0031			.0011	.0022
(T3e)	M3b	.0007	.0000		.0003	.0005	.0031			.0012	.0024
	M4	.0007			.0003	.0005	.0031			.0015	.0026
	M5	.0007	.0000		.0003	.0006	.0031	.0001		.0014	.0028
UT2s	M3a	.0186			.0277	.0171	.0488			.0745	.0455
(Pwr)	M3b	.0186	.2188		.0284	.0177	.0488			.0772	.0485
	M4	.0186			.0315	.0188	.0488			.0853	.0510
	M5	.0186	.1413		.0321	.0195	.0488	.2880		.0855	.0524

Table 4.3.3-9
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0009			.0009	.0012	.0047			.0036	.0060
	M3b	.0009	.0004		.0009	.0012	.0047			.0037	.0060
	M4	.0009			.0010	.0015	.0047			.0041	.0073
	M5	.0009	.0005	.0076	.0010	.0015	.0047	.0018	.0019	.0042	.0075
UT1s	M3a	.0498			.0478	.0315	.1461			.1686	.1001
(Pwr)	M3b	.0498	.0871		.0478	.0315	.1461			.1687	.1005
	M4	.0498			.0521	.0361	.1461			.1819	.1130
	M5	.0498	.0963	.0819	.0522	.0365	.1461	.2539	.3121	.1823	.1139
LT2s	M3a	.0004			.0005	.0005	.0020			.0017	.0026
(T3e)	M3b	.0004	.0002		.0005	.0005	.0020			.0017	.0026
	M4	.0004			.0005	.0007	.0020			.0019	.0032
	M5	.0004	.0003	.0050	.0005	.0007	.0020	.0009	.0012	.0020	.0033
UT2s	M3a	.0279			.0242	.0170	.0867			.0921	.0566
(Pwr)	M3b	.0279	.0522		.0242	.0170	.0867			.0921	.0566
	M4	.0279			.0265	.0197	.0867			.1000	.0645
	M5	.0279	.0583	.0373	.0265	.0200	.0867	.1577	.1690	.1004	.0652

Table 4.3.3-10
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0002			.0001	.0003	.0030			.0009	.0040
	M3b	.0002	.0000		.0001	.0003	.0030			.0009	.0043
	M4	.0002			.0001	.0004	.0030			.0010	.0053
	M5	.0002	.0000	.0000	.0001	.0004	.0030	.0001	.0001	.0011	.0054
UT1s	M3a	.0691			.1281	.0395	.2302			.3849	.1576
(Pwr)	M3b	.0691	.2362		.1284	.0398	.2302			.3907	.1650
	M4	.0691			.1394	.0443	.2302			.4087	.1760
	M5	.0691	.2469	.4692	.1405	.0453	.2302	.5594	.8620	.4102	.1777
LT2s	M3a	.0001			.0000	.0001	.0012			.0005	.0018
(T3e)	M3b	.0001	.0000		.0000	.0001	.0012			.0005	.0018
	M4	.0001			.0001	.0002	.0012			.0005	.0022
	M5	.0001	.0000	.0000	.0001	.0002	.0012	.0000	.0000	.0006	.0023
UT2s	M3a	.0433			.0801	.0234	.1604			.2799	.1023
(Pwr)	M3b	.0433	.1521		.0802	.0235	.1604			.2823	.1064
	M4	.0433			.0881	.0263	.1604			.2989	.1153
	M5	.0433	.1620	.3260	.0891	.0271	.1604	.4651	.7575	.3007	.1170

Table 4.3.3-11
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0001			.0000	.0002	.0021			.0002	.0029
	M3b	.0001	.0000		.0000	.0003	.0021			.0002	.0032
	M4	.0001			.0000	.0003	.0021			.0003	.0038
	M5	.0001	.0000	.0000	.0000	.0003	.0021	.0000	.0000	.0003	.0040
UT1s	M3a	.0999			.2346	.0590	.2724			.5327	.1963
(Pwr)	M3b	.0999	.5498		.2359	.0605	.2724			.5416	.2090
	M4	.0999			.2496	.0650	.2724			.5600	.2177
	M5	.0999	.5223	.8662	.2516	.0663	.2724	.6764	.9733	.5602	.2198
LT2s	M3a	.0000			.0000	.0001	.0008			.0001	.0012
(T3e)	M3b	.0000	.0000		.0000	.0001	.0008			.0001	.0013
	M4	.0000			.0000	.0001	.0008			.0001	.0016
	M5	.0000	.0000	.0000	.0000	.0001	.0008	.0000	.0000	.0001	.0017
UT2s	M3a	.0677			.1694	.0372	.2011			.4181	.1359
(Pwr)	M3b	.0677	.4475		.1700	.0379	.2011			.4232	.1425
	M4	.0677			.1806	.0411	.2011			.4420	.1508
	M5	.0677	.4274	.7857	.1829	.0422	.2011	.6490	.9550	.4434	.1527

Table 4.3.3-12

Tukey's Quick Test for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 0.5, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0001			.0000	.0002	.0017			.0001	.0021
	M3b	.0001	.0000		.0000	.0002	.0017			.0001	.0024
	M4	.0001			.0000	.0002	.0017			.0001	.0028
	M5	.0001	.0000	.0000	.0000	.0002	.0017	.0000	.0000	.0001	.0030
UT1s	M3a	.1220			.3361	.0759	.3004			.6689	.2273
(Pwr)	M3b	.1220	.7149		.3382	.0787	.3004			.6783	.2445
	M4	.1220			.3565	.0831	.3004			.6954	.2505
	M5	.1220	.6734	.9625	.3571	.0845	.3004	.7138	.9860	.6925	.2527
LT2s	M3a	.0000			.0000	.0001	.0007			.0000	.0009
(T3e)	M3b	.0000	.0000		.0000	.0001	.0007			.0000	.0010
	M4	.0000			.0000	.0001	.0007			.0000	.0011
	M5	.0000	.0000	.0000	.0000	.0001	.0007	.0000	.0000	.0000	.0012
UT2s	M3a	.0859			.2529	.0498	.2289			.5521	.1627
(Pwr)	M3b	.0859	.6771		.2541	.0512	.2289			.5582	.1725
	M4	.0859			.2687	.0545	.2289			.5790	.1787
	M5	.0859	.6307	.9426	.2701	.0557	.2289	.7092	.9820	.5768	.1812

Table 4.3.3-13
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0001			.0000	.0001	.0014			.0000	.0017
	M3b	.0001	.0000		.0000	.0001	.0014			.0000	.0019
	M4	.0001			.0000	.0001	.0014			.0000	.0021
	M5	.0001	.0000	.0000	.0000	.0002	.0014	.0000	.0000	.0000	.0024
UT1s	M3a	.1401			.4554	.0926	.3217			.7873	.2541
(Pwr)	M3b	.1401	.7650		.4577	.0966	.3217			.7957	.2744
	M4	.1401			.4805	.1001	.3217			.8081	.2773
	M5	.1401	.7273	.9800	.4767	.1019	.3217	.7345	.9884	.8030	.2796
LT2s	M3a	.0000			.0000	.0000	.0002			.0000	.0003
(T3e)	M3b	.0000	.0000		.0000	.0001	.0002			.0000	.0003
	M4	.0000			.0000	.0001	.0002			.0000	.0003
	M5	.0000	.0000	.0000	.0000	.0001	.0002	.0000	.0000	.0000	.0004
UT2s	M3a	.1012			.3525	.0625	.1869			.5690	.1334
(Pwr)	M3b	.1012	.7586		.3538	.0647	.1869			.5727	.1405
	M4	.1012			.3741	.0674	.1869			.5955	.1449
	M5	.1012	.7164	.9732	.3727	.0691	.1869	.7322	.9844	.5904	.1469

Table 4.3.3-14
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0002			.0002	.0001	.0011			.0008	.0008
	M3b	.0002	.0001		.0002	.0001	.0011			.0008	.0008
	M4	.0002			.0003	.0002	.0011			.0010	.0011
	M5	.0002	.0001		.0003	.0002	.0011	.0003		.0010	.0012
UT1s	M3a	.1134			.1088	.0992	.2751			.3351	.2499
(Pwr)	M3b	.1134	.1936		.1088	.0992	.2751			.3352	.2508
	M4	.1134			.1163	.1106	.2751			.3522	.2727
	M5	.1134	.2063		.1166	.1114	.2751	.4333		.3527	.2738
LT2s	M3a	.0001			.0001	.0001	.0004			.0004	.0003
(T3e)	M3b	.0001	.0000		.0001	.0001	.0004			.0004	.0003
	M4	.0001			.0001	.0001	.0004			.0005	.0004
	M5	.0001	.0001		.0001	.0001	.0004	.0002		.0005	.0004
UT2s	M3a	.0684			.0572	.0588	.1799			.1997	.1608
(Pwr)	M3b	.0684	.1337		.0572	.0588	.1799			.1997	.1608
	M4	.0684			.0613	.0666	.1799			.2120	.1769
	M5	.0684	.1433		.0615	.0674	.1799	.2992		.2124	.1784

Table 4.3.3-15
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000			.0000	.0000	.0004			.0000	.0003
	M3b	.0000	.0000		.0000	.0000	.0004			.0001	.0003
	M4	.0000			.0000	.0000	.0004			.0001	.0003
	M5	.0000	.0000		.0000	.0000	.0004	.0000		.0001	.0003
UT1s	M3a	.1915			.3936	.1764	.4368			.7597	.4261
(Pwr)	M3b	.1915	.5233		.3938	.1778	.4368			.7634	.4410
	M4	.1915			.4120	.1898	.4368			.7762	.4542
	M5	.1915	.5322		.4155	.1918	.4368	.7462		.7791	.4557
LT2s	M3a	.0000			.0000	.0000	.0001			.0000	.0001
(T3e)	M3b	.0000	.0000		.0000	.0000	.0001			.0000	.0001
	M4	.0000			.0000	.0000	.0001			.0000	.0001
	M5	.0000	.0000		.0000	.0000	.0001	.0000		.0000	.0001
UT2s	M3a	.1356			.2840	.1226	.3448			.6427	.3302
(Pwr)	M3b	.1356	.3947		.2840	.1230	.3448			.6443	.3381
	M4	.1356			.3003	.1324	.3448			.6618	.3531
	M5	.1356	.4091		.3034	.1343	.3448	.7091		.6650	.3549

Table 4.3.3-16
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000			.0000	.0000	.0002			.0000	.0001
	M3b	.0000	.0000		.0000	.0000	.0002			.0000	.0001
	M4	.0000			.0000	.0000	.0002			.0000	.0002
	M5	.0000	.0000		.0000	.0000	.0002	.0000		.0000	.0002
UT1s	M3a	.2752			.6980	.2747	.5123			.9294	.5347
(Pwr)	M3b	.2752	.8000		.6987	.2796	.5123			.9330	.5567
	M4	.2752			.7098	.2902	.5123			.9354	.5624
	M5	.2752	.7780		.7160	.2924	.5123	.8003		.9368	.5634
LT2s	M3a	.0000			.0000	.0000	.0001			.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0001			.0000	.0000
	M4	.0000			.0000	.0000	.0001			.0000	.0001
	M5	.0000	.0000		.0000	.0000	.0001	.0000		.0000	.0001
UT2s	M3a	.2124			.5913	.2080	.4287			.8734	.4419
(Pwr)	M3b	.2124	.7656		.5916	.2105	.4287			.8757	.4565
	M4	.2124			.6046	.2197	.4287			.8810	.4656
	M5	.2124	.7458		.6121	.2221	.4287	.7987		.8843	.4675

Table 4.3.3-17
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000			.0000	.0000	.0001			.0000	.0000
	M3b	.0000	.0000		.0000	.0000	.0001			.0000	.0000
	M4	.0000			.0000	.0000	.0001			.0000	.0001
	M5	.0000	.0000		.0000	.0000	.0001	.0000		.0000	.0001
UT1s	M3a	.3341			.8838	.3575	.5585			.9833	.6131
(Pwr)	M3b	.3341	.8529		.8846	.3658	.5585			.9852	.6387
	M4	.3341			.8877	.3732	.5585			.9846	.6386
	M5	.3341	.8292		.8914	.3762	.5585	.8302		.9850	.6390
LT2s	M3a	.0000			.0000	.0000	.0000			.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000			.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
UT2s	M3a	.2692			.8160	.2841	.4823			.9655	.5259
(Pwr)	M3b	.2692	.8510		.8165	.2891	.4823			.9671	.5449
	M4	.2692			.8207	.2961	.4823			.9674	.5493
	M5	.2692	.8268		.8268	.2992	.4823	.8301		.9684	.5504

.+ = 1.0000

Table 4.3.3-18
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000			.0000	.0000	.0001			.0000	.0000
	M3b	.0000	.0000		.0000	.0000	.0001			.0000	.0000
	M4	.0000			.0000	.0000	.0001			.0000	.0000
	M5	.0000	.0000		.0000	.0000	.0001	.0000		.0000	.0000
UT1s	M3a	.3778			.9649	.4300	.5912			.9965	.6742
(Pwr)	M3b	.3778	.8778		.9654	.4411	.5912			.9971	.7012
	M4	.3778			.9653	.4458	.5912			.9967	.6963
	M5	.3778	.8545		.9663	.4480	.5912	.8545		.9968	.6969
LT2s	M3a	.0000			.0000	.0000	.0000			.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000			.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
UT2s	M3a	.3131			.9349	.3538	.4479			.9827	.5113
(Pwr)	M3b	.3131	.8778		.9352	.3612	.4479			.9833	.5274
	M4	.3131			.9352	.3666	.4479			.9829	.5305
	M5	.3131	.8544		.9375	.3688	.4479	.8545		.9835	.5318

.+ = 1.0000

Table 4.3.3-19
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000			.0001	.0000	.0001			.0002	.0001
	M3b	.0000	.0000		.0001	.0000	.0001			.0002	.0001
	M4	.0000			.0001	.0000	.0001			.0002	.0001
	M5	.0000	.0000	.0001	.0001	.0000	.0001	.0001	.0002	.0002	.0001
UT1s	M3a	.2616			.2044	.2378	.5005			.5221	.4759
(Pwr)	M3b	.2616	.3150		.2044	.2378	.5005			.5221	.4772
	M4	.2616			.2159	.2582	.5005			.5389	.5032
	M5	.2616	.3292	.2136	.2159	.2594	.5005	.5906	.5994	.5396	.5044
LT2s	M3a	.0000			.0000	.0000	.0000			.0001	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000			.0001	.0000
	M4	.0000			.0001	.0000	.0000			.0001	.0000
	M5	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0002	.0001	.0000
UT2s	M3a	.1741			.1144	.1546	.3718			.3436	.3454
(Pwr)	M3b	.1741	.2371		.1144	.1546	.3718			.3436	.3454
	M4	.1741			.1218	.1706	.3718			.3591	.3700
	M5	.1741	.2489	.1051	.1219	.1719	.3718	.4405	.3885	.3593	.3711

Table 4.3.3-20

Tukey's Quick Test for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 1.2, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000			.0000	.0000	.0000			.0000	.0000
	M3b	.0000	.0000		.0000	.0000	.0000			.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT1s	M3a	.4667			.7199	.4622	.7217			.9562	.7391
(Pwr)	M3b	.4667	.7412		.7200	.4649	.7217			.9571	.7545
	M4	.4667			.7347	.4815	.7217			.9603	.7611
	M5	.4667	.7425	.8936	.7374	.4834	.7217	.8417	.9936	.9607	.7613
LT2s	M3a	.0000			.0000	.0000	.0000			.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000			.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT2s	M3a	.3792			.5899	.3705	.6440			.9075	.6535
(Pwr)	M3b	.3792	.6316		.5899	.3715	.6440			.9080	.6634
	M4	.3792			.6078	.3875	.6440			.9147	.6751
	M5	.3792	.6423	.7998	.6104	.3895	.6440	.8320	.9810	.9158	.6761

Table 4.3.3-21
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000			.0000	.0000	.0000			.0000	.0000
	M3b	.0000	.0000		.0000	.0000	.0000			.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT1s	M3a	.6115			.9677	.6478	.7966			.9979	.8512
(Pwr)	M3b	.6115	.8959		.9678	.6542	.7966			.9980	.8661
	M4	.6115			.9695	.6627	.7966			.9980	.8652
	M5	.6115	.8798	.9984	.9705	.6638	.7966	.8820	.+	.9981	.8648
LT2s	M3a	.0000			.0000	.0000	.0000			.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000			.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT2s	M3a	.5382			.9351	.5659	.7439			.9943	.7933
(Pwr)	M3b	.5382	.8900		.9352	.5702	.7439			.9944	.8060
	M4	.5382			.9387	.5800	.7439			.9946	.8088
	M5	.5382	.8745	.9952	.9406	.5819	.7439	.8819	.9999	.9948	.8086

.+ = 1.0000

Table 4.3.3-22
Tukey's Quick Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000			.0000	.0000	.0000			.0000	.0000
	M3b	.0000	.0000		.0000	.0000	.0000			.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT1s	M3a	.6910			.9981	.7620	.8353			.9999	.9110
(Pwr)	M3b	.6910	.9250		.9982	.7702	.8353			.9999	.9226
	M4	.6910			.9981	.7733	.8353			.9999	.9190
	M5	.6910	.9095	.+	.9982	.7737	.8353	.9096	.+	.9999	.9181
LT2s	M3a	.0000			.0000	.0000	.0000			.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000			.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT2s	M3a	.6304			.9950	.6956	.7952			.9998	.8710
(Pwr)	M3b	.6304	.9249		.9950	.7017	.7952			.9998	.8825
	M4	.6304			.9949	.7070	.7952			.9998	.8816
	M5	.6304	.9095	.+	.9953	.7076	.7952	.9096	.+	.9998	.8808

.+ = 1.0000

Table 4.3.3-23

Tukey's Quick Test for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 1.2, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000			.0000	.0000	.0000			.0000	.0000
	M3b	.0000	.0000		.0000	.0000	.0000			.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT1s	M3a	.7410			.9999	.8380	.8585			.+	.9461
(Pwr)	M3b	.7410	.9432		.9999	.8455	.8585			.+	.9546
	M4	.7410			.9999	.8456	.8585			.+	.9503
	M5	.7410	.9291	.+	.9999	.8458	.8585	.9291	.+	.+	.9493
LT2s	M3a	.0000			.0000	.0000	.0000			.0000	.0000
(T3e)	M3b	.0000	.0000		.0000	.0000	.0000			.0000	.0000
	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT2s	M3a	.6893			.9998	.7853	.7869			.+	.8826
(Pwr)	M3b	.6893	.9432		.9998	.7914	.7869			.+	.8910
	M4	.6893			.9997	.7930	.7869			.+	.8898
	M5	.6893	.9291	.+	.9998	.7934	.7869	.9291	.+	.+	.8896

.+ = 1.0000

Table 4.3.3-24
Tukey's Quick Test for Two Groups
Mean Ranks of Power Results, α .01, 1-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		| ----------4th ----------				| ---------- 3rd ----------				| ---------- 2nd -------- |			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			3.5	3.5			3.5	3.5			3.0	3.3
	M3b	1.5		3.5	3.5	1.5		3.5	3.5	1.9		3.0	3.3
	M4			1.9	1.9			1.6	1.8			2.0	1.8
	M5	1.5		1.1	1.1	1.5		1.4	1.3	1.1		2.0	1.8
12	M3a			4.0	4.0			3.6	3.8			3.3	3.1
	M3b	1.8		3.0	3.0	1.8		3.4	3.3	1.9		3.3	3.1
	M4			2.0	2.0			1.9	2.0			2.0	2.1
	M5	1.3		1.0	1.0	1.3		1.1	1.0	1.1		1.5	1.6
18	M3a			4.0	4.0			3.9	4.0			3.1	3.4
	M3b	1.0		3.0	3.0	1.0		3.1	3.0	1.1		2.9	3.1
	M4			2.0	2.0			1.9	2.0			2.1	1.8
	M5	2.0		1.0	1.0	2.0		1.1	1.0	1.9		1.9	1.8
24	M3a			3.9	4.0			3.6	4.0			3.0	3.6
	M3b	1.0		2.6	3.0	1.0		2.9	3.0	1.0		3.0	2.3
	M4			2.4	2.0			2.1	1.9			2.0	2.3
	M5	2.0		1.1	1.0	2.0		1.4	1.1	2.0		2.0	1.9
30	M3a	1.0		3.6	4.0			3.4	4.0			3.1	3.6
	M3b			2.6	3.0	1.0		2.9	2.8	1.0		2.6	2.4
	M4	2.0		2.1	2.0			2.1	2.0			2.1	2.0
	M5			1.6	1.0	2.0		1.6	1.3	2.0		2.1	2.0

Table 4.3.3-25
Tukey's Quick Test for Two Groups
Mean Ranks of Power Results, α .01, 2-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			3.5	3.5			3.3	3.5			3.0	3.0
	M3b	2.0		3.5	3.5	2.0		3.3	3.5	1.9		3.0	3.0
	M4			1.8	2.0			1.9	1.6			2.0	2.0
	M5	1.0		1.3	1.0	1.0		1.6	1.4	1.1		2.0	2.0
12	M3a			3.6	3.9			3.5	3.8			3.3	3.3
	M3b	2.0		3.4	3.1	2.0		3.5	3.3	1.9		3.3	3.3
	M4			2.0	2.0			1.9	1.9			1.9	1.8
	M5	1.0		1.0	1.0	1.0		1.1	1.1	1.1		1.6	1.8
18	M3a			4.0	4.0			3.8	3.9			3.0	3.0
	M3b	1.0		3.0	3.0	1.1		3.3	3.1	1.1		3.0	3.0
	M4			2.0	2.0			2.0	2.0			2.1	2.0
	M5	2.0		1.0	1.0	1.9		1.0	1.0	1.9		1.9	2.0
24	M3a			3.6	4.0			3.5	3.9			2.9	3.6
	M3b	1.0		2.9	3.0	1.0		3.0	3.1	1.0		2.9	2.9
	M4			2.5	2.0			2.1	2.0			2.9	2.0
	M5	2.0		1.0	1.0	2.0		1.4	1.0	2.0		1.4	1.5
30	M3a			3.5	4.0	1.0		3.3	4.0			3.1	3.1
	M3b	1.0		2.6	3.0			3.0	3.0	1.0		2.6	2.9
	M4			2.4	2.0	2.0		2.0	1.9			2.1	2.0
	M5	2.0		1.5	1.0			1.8	1.1	2.0		2.1	2.0

Table 4.3.3-26
Tukey's Quick Test for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			3.9	4.0			3.5	3.9			3.5	3.4
	M3b			3.1	3.0			3.5	3.1			3.5	3.4
	M4			2.0	2.0			1.8	1.9			1.5	1.9
	M5			1.0	1.0			1.3	1.1			1.5	1.4
12	M3a			4.0	4.0			4.0	4.0			3.4	3.9
	M3b			3.0	3.0			3.0	3.0			3.1	3.1
	M4			2.0	2.0			2.0	1.9			1.8	1.6
	M5			1.0	1.0			1.0	1.1			1.8	1.4
18	M3a			4.0	4.0			3.6	4.0			3.4	4.0
	M3b			2.9	2.5			2.9	2.5			3.1	2.4
	M4			2.1	2.0			2.0	2.1			1.8	1.6
	M5			1.0	1.5			1.5	1.4			1.8	2.0
24	M3a			3.6	4.0			3.6	4.0			3.5	4.0
	M3b			2.4	2.3			2.6	2.3			2.5	2.5
	M4			2.1	2.3			1.8	2.0			2.3	1.8
	M5			1.9	1.5			2.0	1.8			1.8	1.8
30	M3a			3.6	4.0			3.3	4.0			3.3	3.6
	M3b			2.4	2.5			2.8	2.0			2.6	2.6
	M4			1.9	2.0			1.9	2.4			1.9	1.9
	M5			2.1	1.5			2.1	1.6			2.3	1.9

Table 4.3.3-27

Tukey's Quick Test for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			3.5	3.5			3.5	3.5			3.3	3.3
	M3b			3.5	3.5			3.5	3.5			3.3	3.3
	M4			2.0	2.0			1.6	1.8			1.8	2.3
	M5			1.0	1.0			1.4	1.3			1.8	1.3
12	M3a			4.0	4.0			3.9	4.0			3.4	3.8
	M3b			3.0	3.0			3.1	3.0			3.4	3.0
	M4			2.0	2.0			2.0	2.0			2.0	1.9
	M5			1.0	1.0			1.0	1.0			1.3	1.4
18	M3a			4.0	4.0			3.9	4.0			3.4	3.8
	M3b			3.0	3.0			3.1	3.0			2.9	3.0
	M4			2.0	1.8			1.9	1.9			1.9	1.6
	M5			1.0	1.3			1.1	1.1			1.9	1.6
24	M3a			3.6	4.0			3.6	4.0			3.1	3.8
	M3b			2.9	2.5			2.8	2.5			2.9	2.8
	M4			1.9	2.0			2.0	2.0			2.0	2.0
	M5			1.6	1.5			1.6	1.5			2.0	1.5
30	M3a			3.6	4.0			3.6	4.0			3.1	3.6
	M3b			2.6	2.5			2.6	2.5			2.9	2.3
	M4			2.1	2.0			2.1	2.1			1.9	2.3
	M5			1.6	1.5			1.6	1.4			2.1	1.9

Table 4.3.3-28
Tukey's Quick Test for Two Groups
Mean Ranks of Power Results, α .01, 1-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			3.9	3.9			3.8	3.8			2.5	2.8
	M3b	1.4		3.1	3.1	1.4		3.2	3.2	1.4		2.5	2.4
	M4			2.0	2.0			1.7	1.9			2.5	2.4
	M5	1.6		1.0	1.0	1.6		1.3	1.1	1.6		2.5	2.4
0.5	M3a			3.9	3.9			3.9	3.8			3.4	3.3
	M3b	1.0		3.1	3.1	1.0		3.1	3.2	1.4		3.2	2.9
	M4			1.8	2.0			1.7	1.9			1.7	2.1
	M5	2.0		1.2	1.0	2.0		1.3	1.1	1.6		1.7	1.7
0.8	M3a			3.9	3.9			3.6	3.9	1.4		3.6	3.8
	M3b	1.2		2.9	3.1	1.2		3.2	3.1			3.2	3.1
	M4			2.2	2.0			2.2	1.9	1.6		1.8	1.7
	M5	1.8		1.0	1.0	1.8		1.0	1.1			1.4	1.4
1.2	M3a			3.9	3.9			3.1	3.9			2.9	3.7
	M3b	1.4		3.1	3.1	1.4		3.1	2.9	1.3		2.9	2.9
	M4			2.0	2.0			2.1	2.0			2.2	1.7
	M5	1.6		1.0	1.0	1.8		1.7	1.2	1.7		2.0	1.7

Table 4.3.3-29

Tukey's Quick Test for Two Groups
Mean Ranks of Power Results, α .01, 2-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			3.8	3.8			3.3	3.6			2.6	2.8
	M3b	1.4		3.2	3.2	1.5		3.3	3.4	1.3		2.6	2.4
	M4			1.9	2.0			1.9	1.8			2.6	2.4
	M5	1.6		1.1	1.0	1.5		1.5	1.2	1.7		2.6	2.4
0.5	M3a			3.9	3.9			3.8	3.9			3.5	3.0
	M3b	1.4		3.1	3.1	1.4		3.2	3.1	1.4		3.5	3.0
	M4			1.7	2.0			1.7	1.9			1.5	2.2
	M5	1.6		1.3	1.0	1.6		1.3	1.1	1.6		1.5	1.8
0.8	M3a			3.8	3.9			3.6	3.8			3.3	3.7
	M3b	1.4		3.1	3.1	1.4		3.2	3.2	1.4		2.9	3.3
	M4			2.1	2.0			2.2	1.9			2.2	1.5
	M5	1.6		1.0	1.0	1.6		1.0	1.1	1.6		1.6	1.5
1.2	M3a			3.1	3.9			3.1	3.9			2.8	3.3
	M3b	1.4		2.9	3.1	1.4		3.1	3.1	1.4		2.8	3.3
	M4			2.0	2.0			2.1	1.9			2.5	1.7
	M5	1.6		1.0	1.0	1.6		1.7	1.1	1.6		1.9	1.7

Table 4.3.3-30
Tukey's Quick Test for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			4.0	4.0			3.9	3.9			3.7	3.7
	M3b			3.0	3.0			3.1	3.1			3.3	3.1
	M4			1.8	2.0			1.7	1.9			1.6	1.8
	M5			1.2	1.0			1.3	1.1			1.4	1.4
0.5	M3a			4.0	4.0			3.9	4.0			3.9	3.9
	M3b			3.0	3.0			3.1	3.0			3.0	3.1
	M4			1.6	2.0			1.4	2.0			1.3	1.5
	M5			1.4	1.0			1.6	1.0			1.8	1.5
0.8	M3a			4.0	4.0			3.6	4.0			3.3	3.9
	M3b			2.2	2.8			2.8	2.4			2.9	2.5
	M4			2.4	2.2			2.1	2.2			2.1	1.9
	M5			1.4	1.0			1.5	1.4			1.7	1.7
1.2	M3a			3.3	4.0			3.0	4.0			2.7	3.6
	M3b			2.8	1.8			2.8	1.8			2.7	2.5
	M4			2.3	2.0			2.3	2.1			2.3	1.8
	M5			1.6	2.2			1.9	2.1			2.3	2.1

Table 4.3.3-31

Tukey's Quick Test for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			3.9	3.9			3.9	3.9			3.4	3.0
	M3b			3.1	3.1			3.1	3.1			3.2	3.0
	M4			2.0	2.0			2.0	1.9			1.7	2.6
	M5			1.0	1.0			1.0	1.1			1.7	1.4
0.5	M3a			3.9	3.9			3.9	3.9			3.6	3.7
	M3b			3.1	3.1			3.1	3.1			3.4	3.0
	M4			1.6	2.0			1.5	1.9			1.4	2.0
	M5			1.4	1.0			1.5	1.1			1.6	1.3
0.8	M3a			3.9	3.9			3.9	3.9			3.2	3.9
	M3b			2.9	3.1			2.8	3.1			2.8	2.9
	M4			2.2	2.0			2.2	2.0			2.1	1.6
	M5			1.0	1.0			1.1	1.0			1.9	1.6
1.2	M3a			3.3	3.9			3.1	3.9			2.8	3.9
	M3b			2.9	2.3			3.1	2.3			2.8	2.5
	M4			2.2	1.8			2.0	2.0			2.4	1.8
	M5			1.6	2.0			1.8	1.8			2.0	1.8

Table 4.3.3-32

Tukey's Quick Test for Two Groups

Mean Ranks of Power Results, α .01, 1-sided, by

Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution / *Mthd*	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
M3a			3.8	3.9			3.6	3.9			3.1	3.4
M3b	1.3		3.0	3.1	1.3		3.2	3.1	1.4		3.0	2.8
M4			2.1	2.0			1.9	1.9			2.1	2.0
M5	1.8		1.2	1.0	1.8		1.3	1.1	1.6		1.9	1.8

Table 4.3.3-33

Tukey's Quick Test for Two Groups

Mean Ranks of Power Results, α .01, 2-sided, by

Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution / *Mthd*	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
M3a			3.7	3.9			3.5	3.8			3.1	3.2
M3b	1.4		3.1	3.1	1.4		3.2	3.2	1.4		3.0	3.0
M4			2.1	2.0			2.0	1.9			2.2	2.0
M5	1.6		1.2	1.0	1.6		1.4	1.1	1.6		1.8	1.9

Table 4.3.3-34
Tukey's Quick Test for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a			3.8	4.0			3.6	4.0			3.4	3.8
M3b			2.8	2.7			3.0	2.6			3.0	2.8
M4			2.0	2.1			1.9	2.1			1.8	1.8
M5			1.4	1.3			1.6	1.4			1.8	1.7

Table 4.3.3-35
Tukey's Quick Test for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a			3.8	3.9			3.7	3.9			3.3	3.6
M3b			3.0	2.9			3.0	2.9			3.1	2.9
M4			2.0	2.0			1.9	2.0			1.9	2.0
M5			1.3	1.3			1.4	1.3			1.8	1.5

Table 4.3.3-36
Tukey's Quick Test for Two Groups
Analysis of Mean Ranks of Power Results
Number of First Place Finishes by Distribution
Across Nominal Alpha, Direction and Number of Groups

Decimal	4th				2nd			
Method	3a	3b	4	5	3a	3b	4	5
By Initial Sample Size Across Nominal Effect Size Multiplier								
EA								
EB								
ML	0.00	0.00	1.00	19.00	0.00	0.00	7.50	12.50
SS	0.00	0.00	0.00	20.00	0.00	0.00	6.00	14.00
By Nominal Effect Size Multiplier Across Initial Sample Size								
EA								
EB								
ML	0.00	0.00	0.00	16.00	0.25	0.25	4.25	11.25
SS	0.00	0.00	1.00	15.00	0.00	0.67	4.67	10.67
Across Nominal Effect Size Multiplier and Initial Sample Size								
EA								
EB								
ML	0.00	0.00	0.00	4.00	0.00	0.00	0.50	3.50
SS	0.00	0.00	0.00	4.00	0.00	0.00	0.00	4.00

Table 4.3.3-37
Tukey's Quick Test for Two Groups
Analysis of Mean Ranks of Power Results
Across Nominal Alpha, Direction and Distributions

Decimal	4th				2nd			
Method	3a*	3b*	4	5	3a*	3b*	4	5
By Initial Sample Size across Nominal Effect Size Multiplier								
MP1i	40	40	40	40	40	40	40	40
N1Mi	0.0	0.0	1.0	39.0	0.0	0.0	13.5	26.5
PoM	0.000	0.000	0.025	0.975	0.000	0.000	0.338	0.663
PoT	0.000	0.000	0.025	0.975	0.000	0.000	0.338	0.663
By Nominal Effect Size Multiplier across Initial Sample Size								
MP1i	32	32	32	32	32	32	32	32
N1Mi	0.0	0.0	1.0	31.0	.25	.92	8.92	21.92
PoM	0.000	0.000	0.031	0.969	0.008	0.029	0.279	0.685
PoT	0.000	0.000	0.031	0.969	0.008	0.029	0.279	0.685
By Initial Sample Size and Nominal Effect Size Multiplier								
MP1i	8	8	8	8	8	8	8	8
N1Mi	0.0	0.0	0.0	8.0	0.0	0.0	0.5	7.5
PoM	0.000	0.000	0.000	1.000	0.000	0.000	0.063	0.938
PoT	0.000	0.000	0.000	1.000	0.000	0.000	0.063	0.938

Key: MP1i: Maximum possible 1st place finishes for each method
N1Mi: Actual number of 1st place finishes for each method (ties count 1/n)
PoM: Proportion of Maximum Possible { N1Mi / MP1i }
PoT: Proportion of total 1st place finishes { N1Mi / Σ(N1Mi) }

Note: Results for method 3a and 3b are identical, but both are shown.

4.3.4 – Wilcoxon-Mann-Whitney Test (two groups only)

Based on the Type I error results, power and Type III error results for the Wilcoxon-Mann-Whitney Test are only presented for the combinations of alpha level, distribution and method shown in Table 4.3.4-1.

Table 4.3.4-1
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Method / Distribution Combinations with Acceptable Type I Error

Alpha	.01					.05				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Method										
M3a	x	x	x	x	x	x	x	x	x	x
M4	x		x	x	x	x		x	x	x
M5	x	x	x	x	x	x	x	x	x	x
M6a	x	x	x	x	x	x	x	x	x	x

Power results for the Wilcoxon-Mann-Whitney Test at nominal alpha .01 and .05 are summarized in Table 4.3.4-2 and 4.3.4-3. These tables present the range of values obtained (minimum and maximum) for 1-tailed power, 2-tailed Type III error and 2-tailed power based on the results presented in Tables 4.3.4-4 through 4.3.4-23.

Power tended to increase monotonically across methods and distributions with increases in initial sample size and/or effect size. Likewise, Type III error tended to decrease monotonically across methods and distributions with an increase in initial sample size and/or effect size. Thus, minimum power (or maximum Type III error) usually occurred at initial sample size 6 or 12 at nominal

ESM 0.2, with maximum power (or minimum Type III error) at initial sample size 30 and nominal ESM 1.2, if not sooner.

Table 4.3.4-2
Wilcoxon-Mann-Whitney Test for Two Groups
Range of Power and Type III Error for α .01

Min/Max	min					max				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Type / *Mthd*										
Pwr 1-s										
M3a	.0165	.0174	.0532	.0134	.0140	.9821	.9935	.9353	.9566	.9828
M4	.0165		.0731	.0147	.0164	.9821		.9270	.9571	.9828
M5	.0165	.0210	.0663	.0148	.0163	.9821	.9934	.9342	.9569	.9827
M6a	.0165	.0174	.0532	.0134	.0140	.9821	.9935	.9953	.9566	.9828
T3e 2-s										
M3a	.0000	.0000	.0000	.0000	.0000	.0018	.0013	.0003	.0019	.0014
M4	.0000		.0000	.0000	.0000	.0018		.0005	.0021	.0017
M5	.0000	.0000	.0000	.0000	.0000	.0018	.0016	.0005	.0021	.0018
M6a	.0000	.0000	.0000	.0000	.0000	.0018	.0013	.0003	.0019	.0014
Pwr 2-s										
M3a	.0098	.0098	.0281	.0075	.0080	.9653	.9866	.8942	.9249	.9666
M4	.0098		.0381	.0084	.0095	.9653		.8834	.9257	.9667
M5	.0098	.0123	.0373	.0085	.0097	.9653	.9866	.8929	.9254	.9665
M6a	.0098	.0098	.0281	.0075	.0080	.9653	.9866	.8942	.9249	.9666

Tables 4.3.4-4 through 4.3.4-23 give the upper and lower tail results for the Wilcoxon-Mann-Whitney Test for both 1-sided and 2-sided tests for both alpha .01 and .05. There is a table for each combination of nominal effect size multiplier {0.2, 0.5, 0.8, 1.2} and initial sample size {6, 12, 18, 24, 30}. Results are not reported for distribution EB with nominal ESM 0.2 as the actual ESM = 0.0 (no shift), which is just Type I error. Also for distribution EB, results are not reported for nominal ESM 0.8 as the actual ESM = 0.592, the same as the actual ESM for nominal ESM 0.5. Results for the normal distribution are included and

are essentially identical across methods for a fixed initial sample size, effect size and directionality. This is to be expected in the absence of ties and demonstrates that the simulations worked correctly.

Table 4.3.4-3
Wilcoxon-Mann-Whitney Test for Two Groups
Range of Power and Type III Error for α .05

Min/Max	min					max				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Type / *Mthd*										
Pwr 1-s										
M3a	.0860	.0973	.2573	.0767	.0784	.9976	.9992	.9859	.9918	.9978
M4	.0860		.2900	.0819	.0858	.9976		.9831	.9919	.9978
M5	.0860	.1072	.2731	.0821	.0860	.9976	.9992	.9854	.9919	.9977
M6a	.0860	.0973	.2573	.0767	.0784	.9976	.9992	.9859	.9919	.9978
T3e 2-s										
M3a	.0000	.0000	.0000	.0000	.0000	.0095	.0062	.0013	.0093	.0080
M4	.0000		.0000	.0000	.0000	.0095		.0017	.0102	.0093
M5	.0000	.0000	.0000	.0000	.0000	.0095	.0073	.0015	.0103	.0094
M6a	.0000	.0000	.0000	.0000	.0000	.0095	.0062	.0013	.0093	.0080
Pwr 2-s										
M3a	.0412	.0458	.1376	.0357	.0366	.9937	.9979	.9698	.9814	.9940
M4	.0412		.1654	.0388	.0413	.9937		.9651	.9816	.9939
M5	.0412	.0521	.1532	.0389	.0413	.9937	.9979	.9691	.9815	.9940
M6a	.0412	.0458	.1376	.0357	.0366	.9937	.9979	.9698	.9814	.9940

A ranking analysis of the results is presented in Tables 4.3.4-24 through 4.3.4-35. These summaries were obtained by ranking the power results from Tables 4.3.4-3 through 4.3.4-22 to four decimal places (as reported), to three decimal places and to two decimal places.

Table 4.3.4-4
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0033	.0022		.0033	.0026	.0232	.0161		.0229	.0204
(T3e)	M4	.0033			.0037	.0032	.0232			.0248	.0231
	M5	.0033	.0027		.0037	.0032	.0232	.0185		.0249	.0231
	M6a	.0033	.0022		.0033	.0026	.0232	.0161		.0229	.0204
UT1s	M3a	.0165	.0174		.0134	.0140	.0860	.0973		.0767	.0784
(Pwr)	M4	.0165			.0147	.0164	.0860			.0819	.0858
	M5	.0165	.0210		.0148	.0163	.0860	.1072		.0821	.0860
	M6a	.0165	.0174		.0134	.0140	.0860	.0973		.0767	.0784
LT2s	M3a	.0018	.0013		.0019	.0014	.0095	.0062		.0093	.0080
(T3e)	M4	.0018			.0021	.0017	.0095			.0102	.0093
	M5	.0018	.0016		.0021	.0018	.0095	.0073		.0103	.0094
	M6a	.0018	.0013		.0019	.0014	.0095	.0062		.0093	.0080
UT2s	M3a	.0098	.0098		.0075	.0080	.0412	.0458		.0357	.0366
(Pwr)	M4	.0098			.0084	.0095	.0412			.0388	.0413
	M5	.0098	.0123		.0085	.0097	.0412	.0521		.0389	.0413
	M6a	.0098	.0098		.0075	.0080	.0412	.0458		.0357	.0366

1) For distribution EB, actual ESM = 0.0 (no shift) for Nominal ESM = 0.2. Results differ from those for the NORM distribution due to the presence of ties, but are not reported since no shift has been introduced.

2) M3a and M6a appear to yield identical results for any given combination of alpha level, distribution and direction.

Table 4.3.4-5

Wilcoxon-Mann-Whitney Test for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 0.2, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0023	.0014		.0026	.0022	.0154	.0095		.0163	.0141
(T3e)	M4	.0023			.0027	.0023	.0154			.0169	.0149
	M5	.0023	.0016		.0027	.0023	.0154	.0103		.0170	.0151
	M6a	.0023	.0014		.0026	.0022	.0154	.0095		.0163	.0141
UT1s	M3a	.0265	.0366		.0236	.0250	.1066	.1417		.0995	.1033
(Pwr)	M4	.0265			.0245	.0265	.1066			.1022	.1069
	M5	.0265	.0394		.0248	.0267	.1066	.1471		.1026	.1072
	M6a	.0265	.0366		.0236	.0250	.1066	.1417		.0995	.1033
LT2s	M3a	.0010	.0007		.0012	.0009	.0069	.0042		.0075	.0064
(T3e)	M4	.0010			.0013	.0010	.0069			.0077	.0069
	M5	.0010	.0007		.0013	.0011	.0069	.0046		.0079	.0069
	M6a	.0010	.0007		.0012	.0009	.0069	.0042		.0075	.0064
UT2s	M3a	.0140	.0193		.0121	.0129	.0605	.0820		.0549	.0579
(Pwr)	M4	.0140			.0126	.0138	.0605			.0568	.0604
	M5	.0140	.0211		.0127	.0140	.0605	.0864		.0572	.0607
	M6a	.0140	.0193		.0121	.0129	.0605	.0820		.0549	.0579

Table 4.3.4-6
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0017	.0009		.0021	.0017	.0127	.0070		.0141	.0122
(T3e)	M4	.0017			.0022	.0018	.0127			.0143	.0127
	M5	.0017	.0010		.0022	.0018	.0127	.0075		.0145	.0128
	M6a	.0017	.0009		.0021	.0017	.0127	.0070		.0141	.0112
UT1s	M3a	.0371	.0570		.0327	.0354	.1393	.1936		.1293	.1357
(Pwr)	M4	.0371			.0333	.0366	.1393			.1310	.1382
	M5	.0371	.0594		.0336	.0367	.1393	.1974		.1313	.1386
	M6a	.0371	.0570		.0327	.0354	.1393	.1936		.1293	.1357
LT2s	M3a	.0008	.0004		.0010	.0007	.0053	.0028		.0060	.0050
(T3e)	M4	.0008			.0010	.0008	.0053			.0061	.0052
	M5	.0008	.0005		.0011	.0008	.0053	.0031		.0062	.0053
	M6a	.0008	.0004		.0010	.0007	.0053	.0028		.0060	.0050
UT2s	M3a	.0211	.0337		.0184	.0204	.0787	.1149		.0717	.0758
(Pwr)	M4	.0211			.0188	.0211	.0787			.0727	.0777
	M5	.0211	.0356		.0189	.0213	.0787	.1183		.0732	.0779
	M6a	.0211	.0337		.0184	.0204	.0787	.1149		.0717	.0758

Table 4.3.4-7
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.2, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0014	.0006		.0016	.0013	.0101	.0049		.0111	.0098
(T3e)	M4	.0014			.0016	.0014	.0101			.0112	.0100
	M5	.0014	.0007		.0017	.0014	.0101	.0051		.0114	.0101
	M6a	.0014	.0006		.0016	.0013	.0101	.0049		.0111	.0098
UT1s	M3a	.0460	.0769		.0417	.0451	.1600	.2333		.1494	.1579
(Pwr)	M4	.0460			.0419	.0460	.1600			.1505	.1598
	M5	.0460	.0789		.0424	.0462	.1600	.2362		.1511	.1601
	M6a	.0460	.0769		.0417	.0451	.1600	.2333		.1494	.1579
LT2s	M3a	.0006	.0003		.0007	.0006	.0041	.0019		.0047	.0040
(T3e)	M4	.0006			.0008	.0007	.0041			.0047	.0041
	M5	.0006	.0003		.0008	.0007	.0041	.0021		.0049	.0042
	M6a	.0006	.0003		.0007	.0006	.0041	.0019		.0047	.0040
UT2s	M3a	.0271	.0471		.0240	.0264	.0944	.1462		.0869	.0928
(Pwr)	M4	.0271			.0242	.0270	.0944			.0976	.0942
	M5	.0271	.0488		.0245	.0273	.0944	.1494		.0882	.0945
	M6a	.0271	.0471		.0240	.0264	.0944	.1462		.0869	.0928

Table 4.3.4-8

Wilcoxon-Mann-Whitney Test for Two Groups

Power and Type III Error for α .01 and .05

Nominal Effect Size Multiplier 0.2, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0011	.0004		.0013	.0010	.0084	.0036		.0095	.0080
(T3e)	M4	.0011			.0013	.0011	.0084			.0094	.0082
	M5	.0011	.0005		.0013	.0011	.0084	.0038		.0096	.0083
	M6a	.0011	.0004		.0013	.0010	.0084	.0036		.0095	.0080
UT1s	M3a	.0564	.0978		.0504	.0547	.1840	.2752		.1717	.1815
(Pwr)	M4	.0564			.0505	.0555	.1840			.1723	.1829
	M5	.0564	.0998		.0509	.0556	.1840	.2775		.1729	.1831
	M6a	.0564	.0978		.0504	.0547	.1840	.2752		.1717	.1815
LT2s	M3a	.0004	.0002		.0006	.0004	.0034	.0014		.0039	.0032
(T3e)	M4	.0004			.0006	.0005	.0034			.0040	.0033
	M5	.0004	.0002		.0006	.0005	.0034	.0015		.0040	.0034
	M6a	.0004	.0002		.0006	.0004	.0034	.0014		.0039	.0032
UT2s	M3a	.0320	.0589		.0282	.0308	.1115	.1788		.1022	.1094
(Pwr)	M4	.0320			.0283	.0313	.1115			.1025	.1104
	M5	.0320	.0607		.0286	.0316	.1115	.1812		.1030	.1107
	M6a	.0320	.0589		.0282	.0308	.1115	.1788		.1022	.1094

Table 4.3.4-9
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0008	.0004	.0005	.0008	.0010	.0068	.0031	.0041	.0066	.0091
(T3e)	M4	.0008		.0008	.0009	.0012	.0068		.0057	.0072	.0104
	M5	.0008	.0005	.0007	.0009	.0013	.0068	.0035	.0051	.0072	.0105
	M6a	.0008	.0004	.0005	.0008	.0010	.0068	.0031	.0041	.0066	.0091
UT1s	M3a	.0447	.0791	.0532	.0416	.0285	.1848	.2857	.2573	.1874	.1354
(Pwr)	M4	.0447		.0731	.0452	.0327	.1848		.2900	.1970	.1461
	M5	.0447	.0874	.0663	.0450	.0327	.1848	.2993	.2731	.1970	.1464
	M6a	.0447	.0791	.0532	.0416	.0285	.1848	.2857	.2573	.1874	.1354
LT2s	M3a	.0004	.0002	.0003	.0005	.0005	.0025	.0009	.0013	.0024	.0033
(T3e)	M4	.0004		.0005	.0005	.0007	.0025		.0017	.0026	.0039
	M5	.0004	.0003	.0005	.0005	.0007	.0025	.0011	.0015	.0026	.0039
	M6a	.0004	.0002	.0003	.0005	.0005	.0025	.0009	.0013	.0024	.0033
UT2s	M3a	.0279	.0513	.0281	.0242	.0168	.1001	.1650	.1376	.1002	.0686
(Pwr)	M4	.0279		.0381	.0265	.0197	.1001		.1654	.1067	.0763
	M5	.0279	.0583	.0373	.0265	.0200	.1001	.1780	.1532	.1067	.0762
	M6a	.0279	.0513	.0281	.0242	.0168	.1001	.1650	.1376	.1002	.0686

Table 4.3.4-10
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0002	.0001	.0001	.0003	.0004	.0021	.0005	.0008	.0020	.0040
(T3e)	M4	.0002		.0002	.0003	.0005	.0021		.0012	.0021	.0043
	M5	.0002	.0001	.0001	.0003	.0005	.0021	.0006	.0010	.0021	.0044
	M6a	.0002	.0001	.0001	.0003	.0004	.0021	.0005	.0008	.0020	.0040
UT1s	M3a	.1040	.2260	.1900	.1129	.0656	.2906	.4849	.4552	.3137	.2117
(Pwr)	M4	.1040		.2211	.1166	.0684	.2906		.4750	.3195	.2172
	M5	.1040	.2326	.2015	.1166	.0688	.2906	.4906	.4623	.3198	.2173
	M6a	.1040	.2260	.1900	.1129	.0656	.2906	.4849	.4552	.3137	.2117
LT2s	M3a	.0001	.0000	.0000	.0001	.0002	.0008	.0002	.0003	.0008	.0015
(T3e)	M4	.0001		.0001	.0001	.0002	.0008		.0005	.0009	.0017
	M5	.0001	.0000	.0001	.0001	.0002	.0008	.0002	.0004	.0009	.0017
	M6a	.0001	.0000	.0000	.0001	.0002	.0008	.0002	.0003	.0008	.0015
UT2s	M3a	.0627	.1516	.1192	.0674	.0373	.1927	.3615	.3264	.2096	.1321
(Pwr)	M4	.0627		.1456	.0699	.0394	.1927		.3531	.2146	.1363
	M5	.0627	.1572	.1289	.0701	.0396	.1927	.3681	.3368	.2147	.1367
	M6a	.0627	.1516	.1192	.0674	.0373	.1927	.3615	.3264	.2096	.1321

Table 4.3.4-11

Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0001	.0000	.0000	.0001	.0002	.0010	.0002	.0002	.0009	.0024
(T3e)	M4	.0001		.0000	.0001	.0002	.0010		.0004	.0009	.0025
	M5	.0001	.0000	.0000	.0001	.0002	.0010	.0002	.0003	.0010	.0025
	M6a	.0001	.0000	.0000	.0001	.0002	.0010	.0002	.0002	.0009	.0024
UT1s	M3a	.1723	.3784	.3379	.1932	.1068	.4094	.6563	.6263	.4447	.2997
(Pwr)	M4	.1723		.3643	.1961	.1093	.4094		.6322	.4480	.3031
	M5	.1723	.3822	.3447	.1961	.1096	.4094	.6584	.6272	.4480	.3035
	M6a	.1723	.3784	.3379	.1932	.1068	.4094	.6563	.6263	.4447	.2997
LT2s	M3a	.0000	.0000	.0000	.0000	.0001	.0004	.0001	.0001	.0003	.0009
(T3e)	M4	.0000		.0000	.0000	.0001	.0004		.0001	.0003	.0009
	M5	.0000	.0000	.0000	.0000	.0001	.0004	.0001	.0001	.0003	.0009
	M6a	.0000	.0000	.0000	.0000	.0001	.0004	.0001	.0001	.0003	.0009
UT2s	M3a	.1163	.2891	.2478	.1314	.0678	.2861	.5273	.4916	.3160	.1940
(Pwr)	M4	.1163		.2772	.1337	.0696	.2861		.5079	.3194	.1973
	M5	.1163	.2928	.2555	.1336	.0700	.2861	.5304	.4952	.3192	.1975
	M6a	.1163	.2891	.2478	.1314	.0678	.2861	.5273	.4916	.3160	.1940

Table 4.3.4-12
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000	.0000	.0001	.0001	.0004	.0001	.0001	.0004	.0014
(T3e)	M4	.0000		.0000	.0000	.0001	.0004		.0001	.0004	.0014
	M5	.0000	.0000	.0000	.0000	.0001	.0004	.0001	.0001	.0004	.0015
	M6a	.0000	.0000	.0000	.0001	.0001	.0004	.0001	.0001	.0004	.0014
UT1s	M3a	.2436	.5157	.4710	.2754	.1506	.5019	.7668	.7394	.5432	.3687
(Pwr)	M4	.2436		.4896	.2778	.1526	.5019		.7373	.5457	.3709
	M5	.2436	.5179	.4741	.2774	.1527	.5019	.7678	.7373	.5454	.3712
	M6a	.2436	.5157	.4710	.2754	.1506	.5019	.7668	.7394	.5432	.3687
LT2s	M3a	.0000	.0000	.0000	.0001	.0000	.0001	.0000	.0000	.0001	.0005
(T3e)	M4	.0000		.0000	.0000	.0000	.0001		.0001	.0001	.0005
	M5	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000	.0001	.0005
	M6a	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000	.0001	.0005
UT2s	M3a	.1729	.4185	.3712	.1985	.1001	.3743	.6589	.6229	.4142	.2547
(Pwr)	M4	.1729		.3962	.2003	.1016	.3743		.6296	.4168	.2570
	M5	.1729	.4212	.3757	.2003	.1019	.3743	.6603	.6227	.4164	.2572
	M6a	.1729	.4185	.3712	.1985	.1001	.3743	.6589	.6229	.4142	.2547

Table 4.3.4-13

Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.5, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000	.0000	.0000	.0000	.0002	.0000	.0000	.0002	.0007
(T3e)	M4	.0000		.0000	.0000	.0001	.0002		.0001	.0002	.0007
	M5	.0000	.0000	.0000	.0000	.0001	.0002	.0000	.0000	.0002	.0007
	M6a	.0000	.0000	.0000	.0000	.0000	.0002	.0000	.0000	.0002	.0007
UT1s	M3a	.3158	.6338	.5893	.3582	.1951	.5865	.8475	.8229	.6322	.4370
(Pwr)	M4	.3158		.5988	.3604	.1967	.5865		.8164	.6342	.4385
	M5	.3158	.6348	.5894	.3598	.1971	.5865	.8476	.8200	.6335	.4386
	M6a	.3158	.6338	.5893	.3582	.1951	.5865	.8475	.8229	.6322	.4370
LT2s	M3a	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000	.0001	.0002
(T3e)	M4	.0000		.0000	.0000	.0000	.0001		.0000	.0001	.0002
	M5	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000	.0001	.0002
	M6a	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000	.0001	.0002
UT2s	M3a	.2266	.5301	.4797	.2625	.1299	.4584	.7613	.7271	.5059	.3145
(Pwr)	M4	.2266		.4982	.2644	.1313	.4584		.7261	.5081	.3164
	M5	.2266	.5315	.4820	.2639	.1316	.4584	.7618	.7253	.5076	.3165
	M6a	.2266	.5301	.4797	.2625	.1299	.4584	.7613	.7271	.5059	.3145

Table 4.3.4-14
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0001	.0001		.0002	.0001	.0016	.0007		.0021	.0014
(T3e)	M4	.0001			.0002	.0001	.0016			.0023	.0016
	M5	.0001	.0001		.0002	.0001	.0016	.0007		.0023	.0017
	M6a	.0001	.0001		.0002	.0001	.0016	.0007		.0021	.0014
UT1s	M3a	.1037	.1829		.0949	.0917	.3338	.4857		.3389	.3170
(Pwr)	M4	.1037			.1011	.1023	.3338			.3507	.3337
	M5	.1037	.1943		.1010	.1021	.3338	.4985		.3507	.3339
	M6a	.1037	.1829		.0949	.0917	.3338	.4857		.3389	.3170
LT2s	M3a	.0001	.0000		.0001	.0001	.0005	.0002		.0006	.0004
(T3e)	M4	.0001			.0001	.0001	.0005			.0007	.0005
	M5	.0001	.0001		.0001	.0001	.0005	.0002		.0007	.0005
	M6a	.0001	.0000		.0001	.0001	.0005	.0002		.0006	.0004
UT2s	M3a	.0684	.1330		.0572	.0585	.2046	.3276		.2041	.1889
(Pwr)	M4	.0684			.0613	.0666	.2046			.2139	.2038
	M5	.0684	.1433		.0615	.0674	.2046	.3432		.2139	.2038
	M6a	.0684	.1330		.0572	.0585	.2046	.3276		.2041	.1889

Table 4.3.4-15
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0002	.0000		.0003	.0002
(T3e)	M4	.0000			.0000	.0000	.0002			.0003	.0002
	M5	.0000	.0000		.0000	.0000	.0002	.0000		.0003	.0002
	M6a	.0000	.0000		.0000	.0000	.0002	.0000		.0003	.0002
UT1s	M3a	.2775	.4973		.2959	.2705	.5556	.7607		.5809	.5493
(Pwr)	M4	.2775			.3020	.2774	.5556			.5867	.5553
	M5	.2775	.5042		.3022	.2774	.5556	.7647		.5869	.5553
	M6a	.2775	.4973		.2959	.2705	.5556	.7607		.5809	.5493
LT2s	M3a	.0000	.0000		.0000	.0000	.0001	.0000		.0001	.0001
(T3e)	M4	.0000	.0000		.0000	.0000	.0001	.0000		.0001	.0001
	M5	.0000	.0000		.0000	.0000	.0001	.0000		.0001	.0001
	M6a	.0000	.0000		.0000	.0000	.0001	.0000		.0001	.0001
UT2s	M3a	.1920	.3879		.2032	.1849	.4274	.6529		.4515	.4194
(Pwr)	M4	.1920			.2085	.1910	.4274			.4578	.4263
	M5	.1920	.3946		.2085	.1912	.4274	.6579		.4580	.4265
	M6a	.1920	.3879		.2032	.1849	.4274	.6529		.4515	.4194

Table 4.3.4-16
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
(T3e)	M4	.0000			.0000	.0000	.0000			.0001	.0001
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0001
	M6a	.0000	.0000		.0000	.0000	.0000	.0000		.0001	.0000
UT1s	M3a	.4593	.7288		.4921	.4548	.7351	.9065		.7619	.7337
(Pwr)	M4	.4593			.4956	.4589	.7351			.7645	.7360
	M5	.4593	.7312		.4957	.4587	.7351	.9070		.7642	.7355
	M6a	.4593	.7288		.4921	.4548	.7351	.9065		.7619	.7337
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M6a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
UT2s	M3a	.3607	.6412		.3905	.3563	.6128	.8382		.6446	.6095
(Pwr)	M4	.3607			.3940	.3604	.6128			.6478	.6125
	M5	.3607	.6439		.3941	.3604	.6128	.8397		.6477	.6125
	M6a	.3607	.6412		.3905	.3563	.6128	.8382		.6446	.6095

Table 4.3.4-17

Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M6a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
UT1s	M3a	.6138	.8642		.6506	.6122	.8417	.9627		.8626	.8409
(Pwr)	M4	.6138			.6527	.6139	.8417			.8639	.8418
	M5	.6138	.8647		.6526	.6139	.8417	.9629		.8636	.8414
	M6a	.6138	.8642		.6506	.6122	.8417	.9627		.8626	.8409
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M6a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
UT2s	M3a	.5144	.8038		.5530	.5120	.7483	.9289		.7784	.7480
(Pwr)	M4	.5144			.5555	.5143	.7483			.7799	.7491
	M5	.5144	.8045		.5554	.5143	.7483	.9292		.7798	.7493
	M6a	.5144	.8038		.5530	.5120	.7483	.9289		.7784	.7480

Table 4.3.4-18

Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 0.8, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M6a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
UT1s	M3a	.7345	.9359		.7681	.7344	.9093	.9860		.9238	.9095
(Pwr)	M4	.7345			.7693	.7354	.9093			.9245	.9097
	M5	.7345	.9359		.7691	.7353	.9093	.9859		.9244	.9094
	M6a	.7345	.9359		.7681	.7344	.9093	.9860		.9238	.9095
LT2s	M3a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
(T3e)	M4	.0000			.0000	.0000	.0000			.0000	.0000
	M5	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
	M6a	.0000	.0000		.0000	.0000	.0000	.0000		.0000	.0000
UT2s	M3a	.6376	.8959		.6772	.6368	.8437	.9703		.8666	.8432
(Pwr)	M4	.6376			.6785	.6379	.8437			.8676	.8438
	M5	.6376	.8958		.6784	.6380	.8437	.9703		.8673	.8435
	M6a	.6376	.8959		.6772	.6368	.8437	.9703		.8666	.8432

Table4.3.4-19
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 6

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000	.0001	.0001	.0000	.0002	.0002	.0011	.0006	.0002
(T3e)	M4	.0000		.0001	.0001	.0000	.0002		.0015	.0007	.0002
	M5	.0000	.0000	.0001	.0001	.0000	.0002	.0002	.0014	.0007	.0002
	M6a	.0000	.0000	.0001	.0001	.0000	.0002	.0002	.0011	.0006	.0002
UT1s	M3a	.2440	.3039	.1557	.1795	.2239	.5750	.6517	.4732	.5053	.5599
(Pwr)	M4	.2440		.1707	.1891	.2427	.5750		.4827	.5177	.5775
	M5	.2440	.3171	.1719	.1890	.2423	.5750	.6619	.4823	.5175	.5773
	M6a	.2440	.3039	.1557	.1795	.2239	.5750	.6517	.4732	.5053	.5599
LT2s	M3a	.0000	.0000	.0001	.0000	.0000	.0001	.0000	.0002	.0001	.0000
(T3e)	M4	.0000		.0001	.0001	.0000	.0001		.0002	.0002	.0000
	M5	.0000	.0000	.0001	.0000	.0000	.0001	.0000	.0002	.0002	.0001
	M6a	.0000	.0000	.0001	.0000	.0000	.0001	.0000	.0002	.0001	.0000
UT2s	M3a	.1741	.2365	.0901	.1144	.1543	.4100	.4900	.3110	.3404	.3914
(Pwr)	M4	.1741		.0999	.1218	.1706	.4100		.3201	.3526	.4115
	M5	.1741	.2489	.1051	.1219	.1719	.4100	.5058	.3207	.3524	.4113
	M6a	.1741	.2365	.0901	.1144	.1543	.4100	.4900	.3110	.3404	.3914

Table 4.3.4-20

Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 12

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000	.0001	.0000	.0000	.0001	.0000	.0001	.0000	.0000
(T3e)	M4	.0000		.0001	.0000	.0000	.0001		.0001	.0000	.0000
	M5	.0000	.0000	.0001	.0000	.0000	.0001	.0000	.0001	.0000	.0000
	M6a	.0000	.0000	.0001	.0000	.0000	.0001	.0000	.0001	.0000	.0000
UT1s	M3a	.6182	.7230	.4894	.5307	.6159	.8541	.9046	.7642	.7955	.8546
(Pwr)	M4	.6182		.4915	.5378	.6223	.8541		.7647	.7998	.8573
	M5	.6182	.7278	.4964	.5375	.6220	.8541	.9065	.7683	.7994	.8569
	M6a	.6182	.7230	.4894	.5307	.6159	.8541	.9047	.7642	.7955	.8546
LT2s	M3a	.0000	.0000	.0001	.0000	.0000	.0001	.0000	.0001	.0000	.0000
(T3e)	M4	.0000		.0001	.0000	.0000	.0001		.0001	.0000	.0000
	M5	.0000	.0000	.0001	.0000	.0000	.0001	.0000	.0001	.0000	.0000
	M6a	.0000	.0000	.0001	.0000	.0000	.0001	.0000	.0001	.0000	.0000
UT2s	M3a	.5028	.6214	.3706	.4114	.4974	.7644	.8396	.6505	.6906	.7637
(Pwr)	M4	.5028		.3724	.4190	.5054	.7644		.6530	.6964	.7683
	M5	.5028	.6274	.3775	.4188	.5053	.7644	.8425	.6568	.6960	.7682
	M6a	.5028	.6214	.3706	.4114	.4974	.7644	.8396	.6505	.6906	.7637

Table 4.3.4-21
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 18

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
(T3e)	M4	.0000		.0000	.0000	.0000	.0000		.0000	.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT1s	M3a	.8454	.9110	.7279	.7689	.8469	.9613	.9803	.9069	.9286	.9626
(Pwr)	M4	.8454		.7181	.7719	.8484	.9613		.9009	.9299	.9627
	M5	.8454	.9118	.7282	.7713	.8483	.9613	.9804	.9059	.9295	.9626
	M6a	.8454	.9110	.7279	.7689	.8469	.9613	.9803	.9069	.9286	.9626
LT2s	M3a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
(T3e)	M4	.0000		.0000	.0000	.0000	.0000		.0000	.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT2s	M3a	.7746	.8636	.6365	.6832	.7762	.9218	.9576	.8388	.8700	.9229
(Pwr)	M4	.7746		.6295	.6870	.7780	.9218		.8322	.8719	.9237
	M5	.7746	.8647	.6388	.6863	.7780	.9218	.9580	.8392	.8715	.9235
	M6a	.7746	.8636	.6365	.6832	.7762	.9218	.9576	.8388	.8700	.9229

Table 4.3.4-22
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 24

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
(T3e)	M4	.0000		.0000	.0000	.0000	.0000		.0000	.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT1s	M3a	.9450	.9750	.8625	.8967	.9461	.9900	.9959	.9626	.9752	.9903
(Pwr)	M4	.9450		.8526	.8980	.9463	.9900		.9579	.9755	.9904
	M5	.9450	.9751	.8616	.8975	.9460	.9900	.9959	.9617	.9754	.9902
	M6a	.9450	.9750	.8625	.8967	.9461	.9900	.9959	.9626	.9752	.9903
LT2s	M3a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
(T3e)	M4	.0000		.0000	.0000	.0000	.0000		.0000	.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT2s	M3a	.9095	.9568	.8007	.8433	.9112	.9767	.9901	.9282	.9495	.9774
(Pwr)	M4	.9095		.7884	.8452	.9115	.9767		.9216	.9502	.9773
	M5	.9095	.9570	.8000	.8445	.9112	.9767	.9901	.9275	.9499	.9773
	M6a	.9095	.9568	.8007	.8433	.9112	.9767	.9901	.9282	.9495	.9774

Table 4.3.4-23
Wilcoxon-Mann-Whitney Test for Two Groups
Power and Type III Error for α .01 and .05
Nominal Effect Size Multiplier 1.2, Initial Sample Size 30

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Tail	*Mthd*										
LT1s	M3a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
(T3e)	M4	.0000		.0000	.0000	.0000	.0000		.0000	.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT1s	M3a	.9821	.9935	.9353	.9566	.9828	.9976	.9992	.9859	.9918	.9978
(Pwr)	M4	.9821		.9270	.9571	.9828	.9976		.9831	.9919	.9978
	M5	.9821	.9934	.9342	.9569	.9827	.9976	.9992	.9854	.9919	.9977
	M6a	.9821	.9935	.9353	.9566	.9828	.9976	.9992	.9859	.9918	.9978
LT2s	M3a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
(T3e)	M4	.0000		.0000	.0000	.0000	.0000		.0000	.0000	.0000
	M5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	M6a	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
UT2s	M3a	.9653	.9866	.8942	.9249	.9666	.9937	.9979	.9698	.9814	.9940
(Pwr)	M4	.9653		.8834	.9257	.9667	.9937		.9651	.9816	.9939
	M5	.9653	.9866	.8929	.9254	.9665	.9937	.9979	.9691	.9815	.9940
	M6a	.9653	.9866	.8942	.9249	.9666	.9937	.9979	.9698	.9814	.9940

Table 4.3.4-24
Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .01, 1-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.4	3.5	3.3	3.3
	M4		1.5	1.3	1.1		1.5	1.5	1.4		1.5	1.8	1.8
	M5	1.0	1.5	1.8	1.9	1.0	1.5	1.5	1.6	1.3	1.5	1.8	1.8
	M6a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.4	3.5	3.3	3.3
12	M3a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.1	3.0	3.0	2.8
	M4		1.5	1.6	1.6		1.5	1.5	1.6		1.8	2.0	2.3
	M5	1.0	1.5	1.4	1.4	1.0	1.5	1.5	1.4	1.8	2.3	2.0	2.3
	M6a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.1	3.0	3.0	2.8
18	M3a	2.5	3.0	3.5	3.5	2.5	2.8	3.4	3.5	2.0	2.5	3.0	2.8
	M4		2.5	1.6	1.5		2.5	1.8	1.6		2.5	2.0	2.3
	M5	1.0	1.5	1.4	1.5	1.0	2.0	1.5	1.4	2.0	2.5	2.0	2.3
	M6a	2.5	3.0	3.5	3.5	2.5	2.8	3.4	3.5	2.0	2.5	3.0	2.8
24	M3a	2.5	3.5	3.5	3.3	2.4	2.5	3.3	3.3	2.0	2.5	2.5	2.5
	M4		1.5	1.3	1.6		2.5	1.6	1.8		2.5	2.5	2.5
	M5	1.0	1.5	1.8	1.9	1.3	2.5	1.9	1.8	2.0	2.5	2.5	2.5
	M6a	2.5	3.5	3.5	3.3	2.4	2.5	3.3	3.3	2.0	2.5	2.5	2.5
30	M3a	2.1	2.5	3.5	3.1	2.1	2.3	3.3	3.3	2.0	2.3	2.5	3.0
	M4		2.5	1.3	1.8		2.5	1.8	1.8		2.3	2.5	2.0
	M5	1.8	2.5	1.8	2.0	1.8	3.0	1.8	1.8	2.0	3.3	2.5	2.0
	M6a	2.1	2.5	3.5	3.1	2.1	2.3	3.3	3.3	2.0	2.3	2.5	3.0

Table 4.3.4-25
Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .01, 2-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	2.5	3.5	3.5	3.5	2.5	3.5	3.4	3.5	2.4	3.5	3.0	3.0
	M4		1.5	1.9	2.0		1.5	2.0	1.6		1.8	2.0	2.0
	M5	2.0	1.5	1.1	1.0	1.0	1.5	1.3	1.4	1.3	1.3	2.0	2.0
	M6a	2.5	3.5	3.5	3.5	2.5	3.5	3.4	3.5	2.4	3.5	3.0	3.0
12	M3a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.3	3.3	3.0	3.0
	M4		1.5	1.6	1.8		1.5	1.5	1.6		2.0	2.0	2.0
	M5	1.0	1.5	1.4	1.3	1.0	1.5	1.5	1.4	1.5	1.5	2.0	2.0
	M6a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.3	3.3	3.0	3.0
18	M3a	2.5	3.0	3.5	3.5	2.5	3.0	3.5	3.5	2.1	2.8	2.8	2.5
	M4		2.5	1.5	1.8		2.5	1.4	1.5		2.5	2.3	2.5
	M5	1.0	1.5	1.5	1.3	1.0	1.5	1.6	1.5	1.8	2.0	2.3	2.5
	M6a	2.5	3.0	3.5	3.5	2.5	3.0	3.5	3.5	2.1	2.8	2.8	2.5
24	M3a	2.5	2.5	3.5	3.4	2.4	2.5	3.5	3.4	2.0	2.8	2.9	2.5
	M4		2.5	1.4	1.6		2.5	1.6	1.4		2.5	1.9	2.5
	M5	1.0	2.5	1.6	1.6	1.3	2.5	1.4	1.9	2.0	2.0	2.4	2.5
	M6a	2.5	2.5	3.5	3.4	2.4	2.5	3.5	3.4	2.0	2.8	2.9	2.5
30	M3a	2.1	2.5	3.5	3.3	2.3	2.5	3.3	3.3	2.0	2.8	2.8	2.5
	M4		2.5	1.3	1.8		2.5	1.6	1.9		1.8	2.3	2.5
	M5	1.8	2.5	1.8	1.8	1.5	2.5	1.9	1.6	2.0	2.8	2.3	2.5
	M6a	2.1	2.5	3.5	3.3	2.3	2.5	3.3	3.3	2.0	2.8	2.8	2.5

Table 4.3.4-26
Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.5	3.5	3.3	3.5
	M4		1.0	1.5	1.8		1.0	1.5	1.4		1.3	1.8	1.5
	M5	1.0	2.0	1.5	1.3	1.0	2.0	1.5	1.6	1.0	1.8	1.8	1.5
	M6a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.5	3.5	3.3	3.5
12	M3a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.4	3.0	3.0	3.5
	M4		1.5	1.8	1.6		1.5	1.5	1.5		2.0	2.0	1.5
	M5	1.0	1.5	1.3	1.4	1.0	1.5	1.5	1.5	1.3	2.0	2.0	1.5
	M6a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.4	3.0	3.0	3.5
18	M3a	2.5	2.5	3.5	3.4	2.3	2.5	3.5	3.3	2.1	2.3	2.8	2.8
	M4		2.5	1.4	1.5		2.5	1.4	2.0		3.3	2.3	2.3
	M5	1.0	2.5	1.6	1.8	1.5	2.5	1.6	1.5	1.8	2.3	2.3	2.3
	M6a	2.5	2.5	3.5	3.4	2.3	2.5	3.5	3.3	2.1	2.3	2.8	2.8
24	M3a	2.4	1.5	3.5	3.3	2.3	1.5	3.4	3.1	2.1	2.5	2.8	2.5
	M4		3.8	1.3	1.5		3.8	1.3	1.6		2.5	2.3	2.5
	M5	1.3	3.3	1.8	2.0	1.5	3.3	2.0	2.1	1.8	2.5	2.3	2.5
	M6a	2.4	1.5	3.5	3.3	2.3	1.5	3.4	3.1	2.1	2.5	2.8	2.5
30	M3a	2.1	1.5	3.5	2.9	2.1	1.5	3.0	2.9	2.0	2.3	2.5	2.5
	M4		4.0	1.4	1.8		4.0	2.0	1.9		3.3	2.5	2.5
	M5	1.8	3.0	1.6	2.5	1.8	3.0	2.0	2.4	2.0	2.3	2.5	2.5
	M6a	2.1	1.5	3.5	2.9	2.1	1.5	3.0	2.9	2.0	2.3	2.5	2.5

Table 4.3.4-27

Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.4	3.5	3.3	3.3
	M4		1.5	1.5	1.3		1.5	1.4	1.4		1.3	1.8	1.8
	M5	1.0	1.5	1.5	1.8	1.0	1.5	1.6	1.6	1.3	1.8	1.8	1.8
	M6a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.4	3.5	3.3	3.3
12	M3a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.4	3.3	3.3	3.3
	M4		1.5	1.8	1.8		1.5	1.5	1.9		2.0	1.8	1.8
	M5	1.0	1.5	1.3	1.3	1.0	1.5	1.5	1.1	1.3	1.5	1.8	1.8
	M6a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.4	3.3	3.3	3.3
18	M3a	2.5	3.0	3.5	3.5	2.4	2.8	3.5	3.5	2.1	2.8	2.8	2.8
	M4		2.5	1.3	1.6		2.5	1.5	1.6		2.5	2.3	2.3
	M5	1.0	1.5	1.8	1.4	1.3	2.0	1.5	1.4	1.8	2.0	2.3	2.3
	M6a	2.5	3.0	3.5	3.5	2.4	2.8	3.5	3.5	2.1	2.8	2.8	2.8
24	M3a	2.4	2.0	3.5	3.0	2.3	2.5	3.5	3.5	2.0	2.5	2.9	2.8
	M4		2.5	1.3	2.1		2.5	1.3	1.6		2.5	1.9	2.3
	M5	1.3	3.5	1.8	1.9	1.5	2.5	1.8	1.4	2.0	2.5	2.4	2.3
	M6a	2.4	2.0	3.5	3.0	2.3	2.5	3.5	3.5	2.0	2.5	2.9	2.8
30	M3a	2.3	1.5	3.5	3.1	2.3	1.5	3.5	3.3	2.0	2.5	2.5	2.8
	M4		3.5	1.3	2.3		3.5	1.5	2.0		2.5	2.5	2.3
	M5	1.5	3.5	1.8	1.5	1.5	3.5	1.5	1.5	2.0	2.5	2.5	2.3
	M6a	2.3	1.5	3.5	3.1	2.3	1.5	3.5	3.3	2.0	2.5	2.5	2.8

Table 4.3.4-28
Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .01, 1-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.5		3.5	3.5	2.5		3.2	3.5	2.0		2.5	2.9
	M4			2.0	1.8			2.0	1.5			2.5	2.1
	M5	1.0		1.0	1.2	1.0		1.6	1.5	2.0		2.5	2.1
	M6a	2.5		3.5	3.5	2.5		3.2	3.5	2.0		2.5	2.9
0.5	M3a	2.5	3.5	3.5	3.5	2.5	3.4	3.5	3.5	2.1	3.2	3.1	2.5
	M4		1.0	1.2	1.9		1.0	1.4	1.7		1.1	1.9	2.5
	M5	1.0	2.0	1.8	1.1	1.0	2.2	1.6	1.3	1.8	2.5	1.9	2.5
	M6a	2.5	3.5	3.5	3.5	2.5	3.4	3.5	3.5	2.1	3.2	3.1	2.5
0.8	M3a	2.4		3.5	3.5	2.4		3.5	3.5	2.1		2.9	3.3
	M4			1.4	1.2			1.5	1.5			2.1	1.7
	M5	1.2		1.6	1.8	1.2		1.5	1.5	1.8		2.1	1.7
	M6a	2.4		3.5	3.5	2.4		3.5	3.5	2.1		2.9	3.3
1.2	M3a	2.3	2.9	3.5	3.0	2.2	2.4	3.3	3.1	2.2	2.3	2.9	2.7
	M4		2.8	1.0	1.2		3.2	1.6	1.8		3.1	2.1	2.3
	M5	1.4	1.4	2.0	2.8	1.6	2.0	1.8	2.0	1.6	2.3	2.1	2.3
	M6a	2.3	2.9	3.5	3.0	2.2	2.4	3.3	3.1	2.2	2.3	2.9	2.7

Table 4.3.4-29
Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .01, 2-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		| ----------4th ----------				| ---------- 3rd ----------				| ---------- 2nd -------- |			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.5		3.5	3.5	2.5		3.3	3.4	2.1		2.5	2.5
	M4			2.0	2.0			2.2	1.8			2.5	2.5
	M5	1.0		1.0	1.0	1.0		1.2	1.4	1.8		2.5	2.5
	M6a	2.5		3.5	3.5	2.5		3.3	3.4	2.1		2.5	2.5
0.5	M3a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.2	3.4	2.7	2.5
	M4		1.0	1.4	2.0		1.0	1.5	1.7		1.1	2.3	2.5
	M5	1.0	2.0	1.6	1.0	1.0	2.0	1.5	1.3	1.6	2.1	2.3	2.5
	M6a	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.5	2.2	3.4	2.7	2.5
0.8	M3a	2.3		3.5	3.5	2.4		3.5	3.5	2.1		2.9	2.9
	M4			1.5	1.8			1.5	1.5			2.1	2.1
	M5	1.4		1.5	1.2	1.2		1.5	1.5	1.8		2.1	2.1
	M6a	2.3		3.5	3.5	2.4		3.5	3.5	2.1		2.9	2.9
1.2	M3a	2.4	2.5	3.5	3.2	2.3	2.5	3.4	3.3	2.2	2.6	3.4	2.9
	M4		3.2	1.2	1.3		3.2	1.3	1.4		3.1	1.4	2.1
	M5	1.2	1.8	1.8	2.3	1.4	1.8	1.9	2.0	1.6	1.7	1.8	2.1
	M6a	2.4	2.5	3.5	3.2	2.3	2.5	3.4	3.3	2.2	2.6	3.4	2.9

Table 4.3.4-30
Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.5		3.5	3.5	2.5		3.4	3.5	2.4		2.5	2.9
	M4			2.0	2.0			1.9	1.6			2.5	2.1
	M5	1.0		1.0	1.0	1.0		1.3	1.4	1.2		2.5	2.1
	M6a	2.5		3.5	3.5	2.5		3.4	3.5	2.4		2.5	2.9
0.5	M3a	2.5	2.7	3.5	3.5	2.4	2.7	3.5	3.5	2.2	2.8	3.3	2.9
	M4		2.1	1.4	2.0		2.1	1.4	1.6		1.9	1.7	2.1
	M5	1.0	2.5	1.6	1.0	1.2	2.5	1.6	1.4	1.6	2.5	1.7	2.1
	M6a	2.5	2.7	3.5	3.5	2.4	2.7	3.5	3.5	2.2	2.8	3.3	2.9
0.8	M3a	2.3		3.5	3.3	2.2		3.4	3.1	2.1		2.9	3.1
	M4			1.3	1.3			1.3	1.5			2.1	1.9
	M5	1.4		1.7	2.1	1.6		1.9	2.3	1.8		2.1	1.9
	M6a	2.3		3.5	3.3	2.2		3.4	3.1	2.1		2.9	3.1
1.2	M3a	2.3	2.3	3.5	2.9	2.2	2.3	3.2	2.9	2.2	2.6	2.7	2.9
	M4		3.0	1.1	1.2		3.0	1.5	2.0		3.0	2.3	2.1
	M5	1.4	2.4	1.9	3.0	1.6	2.4	2.1	2.2	1.6	1.8	2.3	2.1
	M6a	2.3	2.3	3.5	2.9	2.2	2.3	3.2	2.9	2.2	2.6	2.7	2.9

Table 4.3.4-31
Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.5		3.5	3.5	2.5		3.5	3.5	2.2		2.8	2.5
	M4			2.0	1.9			1.4	1.8			2.0	2.5
	M5	1.0		1.0	1.1	1.0		1.6	1.2	1.6		2.4	2.5
	M6a	2.5		3.5	3.5	2.5		3.5	3.5	2.2		2.8	2.5
0.5	M3a	2.5	2.9	3.5	3.5	2.5	3.0	3.5	3.5	2.2	3.2	2.9	3.5
	M4		1.4	1.3	1.8		1.4	1.4	1.8		1.3	2.1	1.5
	M5	1.0	2.8	1.7	1.2	1.0	2.6	1.6	1.2	1.6	2.3	2.1	1.5
	M6a	2.5	2.9	3.5	3.5	2.5	3.0	3.5	3.5	2.2	3.2	2.9	3.5
0.8	M3a	2.4		3.5	3.5	2.3		3.5	3.5	2.2		3.1	2.9
	M4			1.3	1.4			1.5	1.6			1.9	2.1
	M5	1.2		1.7	1.6	1.4		1.5	1.4	1.6		1.9	2.1
	M6a	2.4		3.5	3.5	2.3		3.5	3.5	2.2		3.1	2.9
1.2	M3a	2.3	2.5	3.5	2.8	2.2	2.5	3.5	3.3	2.1	2.6	2.9	2.9
	M4		3.2	1.0	2.1		3.2	1.4	1.6		3.0	2.1	2.1
	M5	1.4	1.8	2.0	2.3	1.6	1.8	1.6	1.8	1.8	1.8	2.1	2.1
	M6a	2.3	2.5	3.5	2.8	2.2	2.5	3.5	3.3	2.1	2.6	2.9	2.9

Table 4.3.4-32
Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .01, 1-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.5	3.2	3.5	3.4	2.4	2.9	3.4	3.4	2.1	2.8	2.9	2.9
M4		1.9	1.4	1.5		2.1	1.6	1.6		2.1	2.2	2.2
M5	1.1	1.7	1.6	1.7	1.2	2.1	1.6	1.6	1.8	2.4	2.2	2.2
M6a	2.5	3.2	3.5	3.4	2.4	2.9	3.4	3.4	2.1	2.8	2.9	2.9

Table 4.3.4-33
Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .01, 2-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.4	3.0	3.5	3.4	2.4	3.0	3.4	3.4	2.2	3.0	2.9	2.7
M4		2.1	1.5	1.8		2.1	1.6	1.6		2.1	2.1	2.3
M5	1.2	1.9	1.5	1.4	1.2	1.9	1.5	1.6	1.7	1.9	2.2	2.3
M6a	2.4	3.0	3.5	3.4	2.4	3.0	3.4	3.4	2.2	3.0	2.9	2.7

Table 4.3.4-34

Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .05, 1-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.4	2.5	3.5	3.3	2.3	2.5	3.4	3.3	2.2	2.7	2.9	3.0
M4		2.6	1.5	1.6		2.6	1.5	1.7		2.5	2.2	2.1
M5	1.2	2.5	1.6	1.8	1.4	2.5	1.7	1.8	1.6	2.2	2.2	2.1
M6a	2.4	2.5	3.5	3.3	2.3	2.5	3.4	3.3	2.2	2.7	2.9	3.0

Table 4.3.4-35

Wilcoxon-Mann-Whitney Test for Two Groups
Mean Ranks of Power Results, α .05, 2-sided, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.4	2.7	3.5	3.3	2.4	2.8	3.5	3.5	2.2	2.9	2.9	3.0
M4		2.3	1.4	1.8		2.3	1.4	1.7		2.2	2.0	2.1
M5	1.2	2.3	1.6	1.6	1.3	2.2	1.6	1.4	1.7	2.1	2.1	2.1
M6a	2.4	2.7	3.5	3.3	2.4	2.8	3.5	3.5	2.2	2.9	2.9	3.0

Table 4.3.4-36
Wilcoxon-Mann-Whitney Test for Two Groups
Analysis of Mean Ranks of Power Results
Number of First Place Finishes by Distribution
Across Nominal Alpha, Direction and Number of Groups

Decimal	4th				2nd			
Method	3a	4	5	6a	3a	4	5	6a
By Initial Sample Size Across Nominal Effect Size Multiplier								
EA	0.00		20.00	0.00	2.67		14.67	2.67
EB	0.00	8.50	11.50	0.00	2.25	6.58	8.92	2.25
ML	0.00	12.50	7.50	0.00	1.00	10.00	8.00	1.00
SS	0.00	8.50	11.50	0.00	1.50	8.50	8.50	1.50
By Nominal Effect Size Multiplier Across Initial Sample Size								
EA	0.00		16.00	0.00	0.33		15.33	0.33
EB	0.00	4.00	4.00	0.00	0.33	4.33	3.33	0.33
ML	0.00	11.50	4.50	0.00	0.75	8.25	6.25	0.75
SS	0.00	7.00	9.00	0.00	1.00	7.00	7.00	1.00
Across Nominal Effect Size Multiplier and Initial Sample Size								
EA	0.00		4.00	0.00	0.00		4.00	0.00
EB	0.33	0.50	2.83	0.33	0.00	1.00	3.00	0.00
ML	0.00	3.50	0.50	0.00	0.00	3.00	1.00	0.00
SS	0.00	2.00	2.00	0.00	0.00	2.00	2.00	0.00

Table 4.3.4-37
Wilcoxon-Mann-Whitney Test for Two Groups
Analysis of Mean Ranks of Power Results
Across Nominal Alpha, Direction and Distributions

Decimal	4th				2nd			
Method	3a	4	5	6a	3a	4	5	6a
By Initial Sample Size across Nominal Effect Size Multiplier								
MP1i	80	60	80	80	80	60	80	80
N1Mi	0.0	29.5	50.5	0.0	7.42	25.08	40.08	7.42
PoM	0.000	0.492	0.631	0.000	0.093	0.418	0.501	0.093
PoT	0.000	0.369	0.631	0.000	0.093	0.314	0.501	0.093
By Nominal Effect Size Multiplier across Initial Sample Size								
MP1i	56	40	56	56	56	40	56	56
N1Mi	0.0	22.5	33.5	0.0	2.42	19.58	31.92	2.42
PoM	0.000	0.563	0.598	0.000	0.043	0.490	0.570	0.043
PoT	0.000	0.402	0.598	0.000	0.043	0.348	0.567	0.043
By Initial Sample Size and Nominal Effect Size Multiplier								
MP1i	16	12	16	16	16	12	16	16
N1Mi	0.33	6.00	9.33	0.33	0.0	6.0	10.0	0.0
PoM	0.021	0.500	0.583	0.021	0.000	0.500	0.625	0.000
PoT	0.021	0.375	0.583	0.021	0.000	0.375	0.625	0.000

Key: MP1i: Maximum possible 1st place finishes for each method
N1Mi: Actual number of 1st place finishes for each method (ties count 1/n)
PoM: Proportion of Maximum Possible { N1Mi / MP1i }
PoT: Proportion of total 1st place finishes { N1Mi / Σ(N1Mi) }

4.3.5 – Kruskal-Wallis Test (three to six groups)

Based on the Type I error results, power results for the Kruskal-Wallis Test are only presented for the combinations of alpha level, distribution and method shown in Table 4.3.5-1. Only power results for the H statistic are reported as the Type I error results based on the continuity corrected H_c statistic were generally inflated, relative to the results based on the H statistic, and often exceeded the upper robust tolerance limit across distributions, methods and initial sample sizes.

Table 4.3.5-1
Kruskal-Wallis Test for Three to Six Groups
Power and Type III Error for α .01 and .05
Method / Distribution Combinations with Acceptable Type I Error

Alpha	.01					.05				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Method										
Three Groups										
M3a	x			x		x			x	x
M4	x		x	x		x		x	x	x
M5	x	x	x	x	x	x	x	x	x	x
M6a	x	x	x	x	x	x	x	x	x	x
Four Groups										
M3a	x			x	x	x			x	x
M4	x		x	x	x	x		x	x	x
M5	x	x	x	x	x	x	x	x	x	x
M6a	x	x	x	x	x	x	x	x	x	x
Five Groups										
M3a	x			x	x	x			x	
M4	x		x	x	x	x		x	x	x
M5	x	x	x	x	x	x	x	x	x	x
M6a	x	x	x	x	x	x	x	x	x	x
Six Groups										
M3a	x			x		x			x	
M4	x		x	x	x	x		x	x	x
M5	x	x	x	x	x	x	x	x	x	x
M6a	x	x	x	x	x	x	x	x	x	x

Power results for the Kruskal-Wallis Test at nominal alpha .01 and .05 are summarized in Table 4.3.5-2 and 4.3.5-3. These tables present the range of values obtained (minimum and maximum) based on the results presented in Tables 4.3.5-4 through 4.3.5-23.

Table 4.3.5-2
Kruskal-Wallis Test for Three to Six Groups
Range of Power for α .01

Min/Max	min (ISS = 6, NESM = 0.2)					max (ISS = 24, NESM = 1.2)				
Dist *Grps | Mthd*	Norm	EA	EB*	ML	SS	Norm	EA	EB	ML	SS
Three Groups										
M3a	.0185			.0160		.+			.+	
M4	.0185		.1151	.0155		.+		.+	.+	
M5	.0185	.0278	.1229	.0164	.0187	.+	.+	.+	.+	.+
M6a	.0185	.0257	.1137	.0158	.0172	.+	.+	.+	.+	.+
Four Groups										
M3a	.0296			.0248	.0295	.+			.+	.+
M4	.0296		.2756	.0233	.0257	.+		.+	.+	.+
M5	.0296	.0509	.3035	.0250	.0296	.+	.+	.+	.+	.+
M6a	.0296	.0488	.2930	.0245	.0284	.+	.+	.+	.+	.+
Five Groups										
M3a	.0492			.1028	.1461	.+			.+	.+
M4	.0492		.6191	.0958	.1256	.+		.+	.+	.+
M5	.0492	.0891	.6501	.0121	.1413	.+	.+	.+	.+	.+
M6a	.0492	.0867	.6434	.1018	.1396	.+	.+	.+	.+	.+
Six Groups										
M3a	.0876			.0614		.+			.+	
M4	.0876		.9617	.0597	.0815	.+		.+	.+	.+
M5	.0876	.1515	.9599	.0612	.0871	.+	.+	.+	.+	.+
M6a	.0876	.1484	.9645	.0608	.0608	.+	.+	.+	.+	.+

* ESM = 0.5, .+ = 1.0000

Table 4.3.5-3
Kruskal-Wallis Test for Three to Six Groups
Range of Power for α .05

Min/Max	min (ISS = 6, NESM = 0.2)					max (ISS = 24, NESM = 1.2)				
Dist	Norm	EA	EB*	ML	SS	Norm	EA	EB	ML	SS
Grps | Mthd										
Three Groups										
M3a	.0780			.0712	.0769	.+			.+	.+
M4	.0780		.3193	.0690	.0712	.+		.+	.+	.+
M5	.0780	.1027	.3349	.0726	.0783	.+	.+	.+	.+	.+
M6a	.0780	.0973	.3221	.0703	.0742	.+	.+	.+	.+	.+
Four Groups										
M3a	.1107			.0986	.1119	.+			.+	.+
M4	.1107		.5544	.0936	.0094	.+		.+	.+	.+
M5	.1170	.1581	.5838	.0986	.1106	.+	.+	.+	.+	.+
M6a	.1170	.1542	.5779	.0977	.1080	.+	.+	.+	.+	.+
Five Groups										
M3a	.1654			.1369		.+			.+	
M4	.1654		.8697	.1313	.1515	.+		.+	.+	.+
M5	.1654	.2395	.8835	.1367	.1646	.+	.+	.+	.+	.+
M6a	.1654	.2352	.8861	.1357	.1615	.+	.+	.+	.+	.+
Six Groups										
M3a	.2479			.1924		.+			.+	
M4	.2479		.9971	.1877	.2346	.+		.+	.+	.+
M5	.2479	.3457	.9969	.1921	.2470	.+	.+	.+	.+	.+
M6a	.2479	.3418	.9979	.1909	.2438	.+	.+	.+	.+	.+

* ESM = 0.5, .+ = 1.0000

Up to the maximum of 1.0000, power tended to increase monotonically across methods and distributions with increases in initial sample size and/or effect size. Thus, minimum power usually occurred at initial sample size 6 or 12 at nominal ESM 0.2, with maximum power at initial sample size 24 and nominal ESM 1.2, if not sooner. Power results typically reached 1.0000 prior to a nominal effect size multiplier of 1.2 and initial sample size of 24. This was because the effect shift was applied between each pair of adjacent groups. Thus for six groups at nominal effect size multiplier 0.2, the nominal effect size multiplier between groups 1 and 6 (lowest and highest scores) was actually 1.0, irrespective of the sample size. At nominal effect size multiplier 1.2, the nominal effect size multiplier between these same two groups was 6.0. Under these conditions the test was going to be very powerful even with small sample sizes.

Tables 4.3.5-4 through 4.3.5-23 give the omnibus power results for the Kruskal-Wallis Test for both alpha .01 and .05 for three to six groups. There is a table for each combination of number of groups {3, 4, 5, 6} and nominal effect size multiplier {0.2, 0.5, 0.8, 1.2}. Results are not reported for distribution EB with nominal ESM 0.2 as the actual ESM = 0.0 (no shift), which is just Type I error. Also for distribution EB, results are not reported for nominal ESM 0.8 as the actual ESM = 0.592, the same as the actual ESM for nominal ESM 0.5. Results for the normal distribution are included and are essentially identical across methods for a fixed initial sample size and effect size. This is to be expected in the absence of ties and demonstrates that the simulations worked correctly.

A ranking analysis of the results is presented in Tables 4.3.5-24 through 4.3.5-35. These summaries were obtained by ranking the power results from Tables 4.3.5-3 through 4.3.5-22 to four decimal places (as reported), to three decimal places and to two decimal places.

Table 4.3.5-4

Kruskal-Wallis Test (Three Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.0185			.0160		.0780			.0712	.0769
	M4	.0185			.0155		.0780			.0690	.0712
	M5	.0185	.0278		.0164	.0187	.0780	.1027		.0726	.0783
	M6a	.0185	.0257		.0158	.0172	.0780	.0973		.0703	.0742
12	M3a	.0336			.0286		.1179			.1072	.1215
	M4	.0336			.0262		.1179			.0998	.1053
	M5	.0336	.0598		.0287	.0334	.1179	.1791		.1069	.1184
	M6a	.0336	.0578		.0284	.0325	.1179	.1762		.1061	.1165
18	M3a	.0512			.0433		.1585			.1419	.1651
	M4	.0512			.0389		.1585			.1304	.1407
	M5	.0512	.1002		.0430	.0512	.1585	.2556		.1406	.1578
	M6a	.0512	.0985		.0429	.0504	.1585	.2536		.1402	.1565
24	M3a	.0711			.0599		.2019			.1803	.2134
	M4	.0711			.0534		.2019			.1656	.1812
	M5	.0711	.1469		.0594	.0709	.2019	.3361		.1783	.2019
	M6a	.0711	.1451		.0591	.0702	.2019	.3341		.1779	.2006

Table 4.3.5-5

Kruskal-Wallis Test (Three Groups), Power for α .01 and .05

Nominal Effect Size Multiplier 0.5

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.0891			.0817		.2541			.2493	.1735
	M4	.0891		.1151	.0797		.2541		.3193	.2435	.1590
	M5	.0891	.1889	.1229	.0833	.0537	.2541	.4096	.3349	.2519	.1752
	M6a	.0891	.1848	.1137	.0813	.0510	.2541	.4027	.3221	.2476	.1692
12	M3a	.2728			.2734		.5205			.5259	.3600
	M4	.2728		.3968	.2618		.5205		.6537	.5104	.3241
	M5	.2728	.5169	.4073	.2734	.1518	.5205	.7514	.6665	.5247	.3531
	M6a	.2728	.5158	.4026	.2726	.1501	.5205	.7515	.6655	.5243	.3514
18	M3a	.4781			.4833		.7198			.7251	.5284
	M4	.4781		.6474	.4666		.7198		.8414	.7102	.4833
	M5	.4781	.7681	.6596	.4821	.2777	.7198	.9122	.8516	.7235	.5161
	M6a	.4781	.7683	.6589	.4820	.2764	.7198	.9126	.8529	.7235	.5153
24	M3a	.6556			.6617		.8501			.8542	.6702
	M4	.6556		.8140	.6450		.8501		.9345	.8432	.6250
	M5	.6556	.9042	.8257	.6600	.4082	.8501	.9730	.9413	.8527	.6549
	M6a	.6556	.9047	.8266	.6600	.4073	.8501	.9734	.9426	.8529	.6548

Table 4.3.5-6

Kruskal-Wallis Test (Three Groups), Power for α .01 and .05

Nominal Effect Size Multiplier 0.8

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.2939			.2691		.5719			.5620	.5749
	M4	.2939			.2632		.5719			.5515	.5486
	M5	.2939	.4587		.2724	.2943	.5719	.7226		.5641	.5747
	M6a	.2939	.4543		.2683	.2880	.5719	.7185		.5597	.5691
12	M3a	.7523			.7457		.9133			.9130	.9197
	M4	.7523			.7323		.9133			.9056	.9027
	M5	.7523	.8952		.7453	.7563	.9133	.9719		.9122	.9161
	M6a	.7523	.8955		.7450	.7563	.9133	.9722		.9124	.9165
18	M3a	.9430			.9408		.9872			.9869	.9892
	M4	.9430			.9352		.9872			.9851	.9852
	M5	.9430	.9878		.9403	.9452	.9872	.9981		.9867	.9881
	M6a	.9430	.9878		.9404	.9457	.9872	.9981		.9867	.9883
24	M3a	.9897			.9893		.9985			.9985	.9989
	M4	.9897			.9879		.9985			.9982	.9983
	M5	.9897	.9990		.9891	.9907	.9985	.9999		.9984	.9987
	M6a	.9897	.9990		.9891	.9908	.9985	.9999		.9984	.9988

Table 4.3.5-7
Kruskal-Wallis Test (Three Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 1.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.7052			.6406		.9077			.8946	.9181
	M4	.7052		.5660	.6317		.9077		.8750	.8866	.9058
	M5	.7052	.7222	.6188	.6444	.7152	.9077	.9089	.8934	.8949	.9164
	M6a	.7052	.7193	.6040	.6397	.7107	.9077	.9076	.9024	.8933	.9158
12	M3a	.9918			.9901		.9991			.9990	.9994
	M4	.9918		.9904	.9884		.9991		.9993	.9988	.9991
	M5	.9918	.9906	.9897	.9899	.9934	.9991	.9988	.9991	.9990	.9993
	M6a	.9918	.9907	.9940	.9900	.9937	.9991	.9988	.9997	.9990	.9994
18	M3a	.9999			.9999		.+			.+	.+
	M4	.9999		.+	.9999		.+		.+	.+	.+
	M5	.9999	.9999	.9999	.9999	.9999	.+	.+	.+	.+	.+
	M6a	.9999	.9999	.+	.9999	.9999	.+	.+	.+	.+	.+
24	M3a	.+			.+		.+			.+	.+
	M4	.+		.+	.+		.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+

+ = 1.0000

Table 4.3.5-8

Kruskal-Wallis Test (Four Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.0296			.0248	.0295	.1107			.0986	.1119
	M4	.0296			.0233	.0257	.1107			.0936	.0994
	M5	.0296	.0509		.0250	.0296	.1107	.1581		.0986	.1106
	M6a	.0296	.0488		.0245	.0284	.1107	.1542		.0977	.1080
12	M3a	.0682			.0542	.0698	.1997			.1719	.2065
	M4	.0682			.0493	.0577	.1997			.1599	.1777
	M5	.0682	.1300		.0540	.0680	.1997	.3112		.1711	.2002
	M6a	.0682	.1278		.0537	.0667	.1997	.3084		.1703	.1979
18	M3a	.1198			.0932	.1251	.2949			.2487	.3076
	M4	.1198			.0841	.1026	.2949			.2304	.2651
	M5	.1198	.2373		.0924	.1194	.2949	.4633		.2465	.2944
	M6a	.1198	.2354		.0921	.1181	.2949	.4615		.2461	.2929
24	M3a	.1818			.1409	.1925	.3905			.3302	.4112
	M4	.1818			.1272	.1592	.3905			.3071	.3576
	M5	.1818	.3556		.1393	.1819	.3905	.5986		.3269	.3908
	M6a	.1818	.3540		.1390	.1805	.3905	.5975		.3264	.3896

Table 4.3.5-9

Kruskal-Wallis Test (Four Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.5

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.2476			.2179	.1350	.5049			.4726	.3402
	M4	.2476		.2756	.2097	.1201	.5049		.5544	.4591	.3100
	M5	.2476	.4192	.3035	.2182	.1345	.5049	.6717	.5838	.4716	.3366
	M6a	.2476	.4173	.2930	.2169	.1318	.5049	.6707	.5779	.4710	.3337
12	M3a	.6808			.6401	.4289	.8721			.8488	.6864
	M4	.6808		.7718	.6235	.3867	.8721		.9250	.8375	.6434
	M5	.6808	.8653	.7821	.6391	.4218	.8721	.9597	.9277	.8477	.6770
	M6a	.6808	.8655	.7830	.6389	.4203	.8721	.9597	.9297	.8477	.6767
18	M3a	.9088			.8843	.6977	.9761			.9676	.8786
	M4	.9088		.9582	.8743	.6528	.9761		.9918	.9637	.8491
	M5	.9088	.9821	.9577	.8833	.6869	.9761	.9968	.9912	.9671	.8702
	M6a	.9088	.9823	.9597	.8834	.6866	.9761	.9969	.9921	.9672	.8706
24	M3a	.9801			.9708	.8647	.9965			.9944	.9596
	M4	.9801		.9943	.9671	.8324	.9965		.9993	.9934	.9448
	M5	.9801	.9983	.9939	.9703	.8548	.9965	.9998	.9991	.9942	.9545
	M6a	.9801	.9983	.9945	.9704	.8552	.9965	.9998	.9993	.9943	.9548

Table 4.3.5-10
Kruskal-Wallis Test (Four Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.8

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.7285			.7480	.7401	.9122			.9349	.9216
	M4	.7285			.7338	.7136	.9122			.9279	.9065
	M5	.7285	.4587		.7469	.7361	.9122	.9506		.9341	.9187
	M6a	.7285	.4543		.7468	.7360	.9122	.9510		.9343	.9193
12	M3a	.9939			.9975	.9956	.9993			.9998	.9996
	M4	.9939			.9970	.9937	.9993			.9998	.9994
	M5	.9939	.9982		.9974	.9951	.9993	.9999		.9998	.9996
	M6a	.9939	.9983		.9975	.9953	.9993	.9999		.9998	.9996
18	M3a	.+			.+	.+	.+			.+	.+
	M4	.+			.+	.+	.+			.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+
24	M3a	.+			.+	.+	.+			.+	.+
	M4	.+			.+	.+	.+			.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+

+ = 1.0000

Table 4.3.5-11
Kruskal-Wallis Test (Four Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 1.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.9898			.9965	.9943	.9992			.+	.9997
	M4	.9898		.9971	.9956	.9925	.9992		.+	.9999	.9995
	M5	.9898	.9851	.9970	.9963	.9938	.9992	.9986	.+	.+	.9996
	M6a	.9898	.9854	.9992	.9964	.9941	.9992	.9987	.+	.+	.9997
12	M3a	.+			.+	.+	.+			.+	.+
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
18	M3a	.+			.+	.+	.+			.+	.+
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
24	M3a	.+			.+	.+	.+			.+	.+
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+

+ = 1.0000

Table 4.3.5-12
Kruskal-Wallis Test (Five Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.0492			.0380	.0494	.1654			.1369	
	M4	.0492			.0363	.0439	.1654			.1313	.1515
	M5	.0492	.0891		.0381	.0492	.1654	.2395		.1367	.1646
	M6a	.0492	.0867		.0377	.0474	.1654	.2352		.1357	.1615
12	M3a	.1416			.1028	.1461	.3349			.2690	
	M4	.1416			.0958	.1256	.3349			.2553	.3080
	M5	.1416	.2578		.1021	.1413	.3349	.4903		.2675	.3344
	M6a	.1416	.2555		.1018	.1396	.3349	.4888		.2669	.3322
18	M3a	.2678			.1931	.2797	.5064			.4113	
	M4	.2678			.1798	.2432	.5064			.3913	.4754
	M5	.2678	.4614		.1914	.2678	.5064	.6994		.4083	.5066
	M6a	.2678	.4595		.1910	.2663	.5064	.6989		.4079	.5051
24	M3a	.4067			.2994	.4258	.6525			.5415	
	M4	.4067			.2807	.3757	.6525			.5188	.6230
	M5	.4067	.6537		.2965	.4062	.6525	.8377		.5372	.6530
	M6a	.4067	.6429		.2962	.4049	.6525	.8373		.5368	.6520

Table 4.3.5-13

Kruskal-Wallis Test (Five Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.5

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.5414			.4829	.3116	.7948			.7592	
	M4	.5414		.6191	.4711	.2889	.7948		.8697	.7480	.5560
	M5	.5414	.6972	.6501	.4824	.3087	.7948	.8821	.8835	.7584	.5810
	M6a	.5414	.6968	.6434	.4817	.3059	.7948	.8821	.8861	.7579	.5789
12	M3a	.9592			.9422	.7928	.9921			.9881	
	M4	.9592		.9920	.9364	.7621	.9921		.9993	.9863	.9143
	M5	.9592	.9879	.9909	.9416	.7855	.9921	.9982	.9990	.9879	.9263
	M6a	.9592	.9880	.9928	.9416	.7855	.9921	.9982	.9994	.9879	.9265
18	M3a	.9984			.9972	.9643	.9999			.9997	
	M4	.9984		.+	.9967	.9536	.9999		.+	.9997	.9899
	M5	.9984	.9999	.9999	.9971	.9610	.9999	.+	.+	.9997	.9919
	M6a	.9984	.9999	.+	.9972	.9614	.9999	.+	.+	.9997	.9920
24	M3a	.+			.9999	.9957	.+			.+	
	M4	.+		.+	.9999	.9936	.+		.+	.+	.9991
	M5	.+	.+	.+	.9999	.9950	.+	.+	.+	.+	.9993
	M6a	.+	.+	.+	.9999	.9999	.+	.+	.+	.+	.9993

+ = 1.0000

Table 4.3.5-14

Kruskal-Wallis Test (Five Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.8

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.9744			.9932	.9809	.9970			.9998	
	M4	.9744			.9922	.9770	.9970			.9997	.9974
	M5	.9744	.9860		.9930	.9797	.9970	.9985		.9998	.9979
	M6a	.9744	.9863		.9931	.9803	.9970	.9985		.9998	.9980
12	M3a	.+			.+	.+	.+			.+	
	M4	.+			.+	.+	.+			.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+
18	M3a	.+			.+	.+	.+			.+	
	M4	.+			.+	.+	.+			.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+
24	M3a	.+			.+	.+	.+			.+	
	M4	.+			.+	.+	.+			.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+

+ = 1.0000

Table 4.3.5-15
Kruskal-Wallis Test (Five Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 1.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.+			.+	.+	.+			.+	
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
12	M3a	.+			.+	.+	.+			.+	
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
18	M3a	.+			.+	.+	.+			.+	
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
24	M3a	.+			.+	.+	.+			.+	
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+

+ = 1.0000

Table 4.3.5-16
Kruskal-Wallis Test (Six Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.0876			.0614		.2479			.1924	
	M4	.0876			.0597	.0815	.2479			.1877	.2346
	M5	.0876	.1515		.0612	.0871	.2479	.3457		.1921	.2470
	M6a	.0876	.1484		.0608	.0848	.2479	.3418		.1909	.2438
12	M3a	.2804			.1889		.5265			.4093	
	M4	.2804			.1813	.2633	.5265			.3972	.5057
	M5	.2804	.4456		.1879	.2808	.5265	.6866		.4073	.5273
	M6a	.2804	.4432		.1873	.2788	.5265	.6852		.4067	.5256
18	M3a	.5085			.3594		.7439			.6093	
	M4	.5085			.3459	.4866	.7439			.5942	.7251
	M5	.5085	.7119		.3569	.5093	.7439	.8804		.6062	.7445
	M6a	.5085	.7109		.3564	.5082	.7439	.8802		.6057	.7437
24	M3a	.7037			.5340		.8765			.7618	
	M4	.7037			.5174	.6838	.8765			.7477	.8649
	M5	.7037	.8766		.5303	.7042	.8765	.9616		.7582	.8776
	M6a	.7037	.8764		.5300	.7039	.8765	.9616		.7581	.8775

Table 4.3.5-17
Kruskal-Wallis Test (Six Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.5

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.8467			.8184		.9603			.9546	
	M4	.8467		.9617	.8107	.5728	.9603		.9971	.9509	.8120
	M5	.8467	.9061	.9599	.8175	.5867	.9603	.9771	.9969	.9540	.8233
	M6a	.8467	.9060	.9645	.8173	.5844	.9603	.9773	.9979	.9541	.8230
12	M3a	.9992			.9992		.+			.+	
	M4	.9992		.+	.9990	.9698	.+		.+	.9999	.9946
	M5	.9992	.9998	.+	.9991	.9734	.+	.+	.+	.+	.9954
	M6a	.9992	.9998	.+	.9992	.9736	.+	.+	.+	.+	.9955
18	M3a	.+			.+		.+			.+	
	M4	.+		.+	.+	.9992	.+		.+	.+	.9999
	M5	.+	.+	.+	.+	.9994	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.9994	.+	.+	.+	.+	.+
24	M3a	.+			.+	.+	.+			.+	
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+

+ = 1.0000

Table 4.3.5-18

Kruskal-Wallis Test (Six Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.8

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.9997			.+		.+			.+	
	M4	.9997			.+	.9999	.+			.+	.+
	M5	.9997	.9999		.+	.9999	.+	.+		.+	.+
	M6a	.9997	.9999		.+	.9999	.+	.+		.+	.+
12	M3a	.+			.+		.+			.+	
	M4	.+			.+	.+	.+			.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+
18	M3a	.+			.+		.+			.+	
	M4	.+			.+	.+	.+			.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+
24	M3a	.+			.+		.+			.+	
	M4	.+			.+	.+	.+			.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+

+ = 1.0000

Table 4.3.5-19

Kruskal-Wallis Test (Six Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 1.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
6	M3a	.+			.+		.+			.+	
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
12	M3a	.+			.+		.+			.+	
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
18	M3a	.+			.+		.+			.+	
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
24	M3a	.+			.+		.+			.+	
	M4	.+		.+	.+	.+	.+		.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+

+ = 1.0000

Table 4.3.5-20
Kruskal-Wallis Test
Mean Ranks of Power Results for Three Groups, α .01, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			2.0				2.1				2.3	
	M4		2.5	1.0			2.5	3.6			2.3	3.3	
	M5	1.0	1.0	1.0	1.0	1.0	1.0	1.4	1.0	1.3	1.3	2.3	1.4
	M6a	2.0	2.5	3.0	2.0	2.0	2.5	2.9	2.0	1.8	2.5	2.3	1.6
12	M3a			1.4				1.9				2.3	
	M4		2.5	4.0			2.8	4.0			2.3	3.3	
	M5	1.5	2.0	1.9	1.4	1.4	1.8	1.9	1.5	1.5	1.5	2.3	1.5
	M6a	1.5	1.5	2.8	1.6	1.6	1.5	2.3	1.5	1.5	2.3	2.3	1.5
18	M3a			1.4				1.6				2.4	
	M4		2.3	3.6			2.3	3.6			2.5	2.9	
	M5	1.5	2.0	2.4	1.4	1.4	2.0	2.4	1.4	1.5	1.8	2.4	1.5
	M6a	1.5	1.8	2.6	1.6	1.6	1.8	2.4	1.6	1.5	1.8	2.4	1.5
24	M3a			1.4				1.6				2.3	
	M4		2.5	3.6			2.5	3.6			2.5	3.3	
	M5	1.5	2.0	2.4	1.4	1.5	2.0	2.4	1.3	1.5	1.8	2.3	1.5
	M6a	1.5	1.5	2.6	1.6	1.5	1.5	2.4	1.8	1.5	1.8	2.3	1.5

Table 4.3.5-21
Kruskal-Wallis Test
Mean Ranks of Power Results for Four Groups, α .01, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			1.5	1.3			1.4	1.5			2.4	2.0
	M4		2.5	4.0	4.0		2.8	3.8	4.0		2.5	3.4	3.3
	M5	1.3	2.0	1.8	2.0	1.1	1.8	2.3	1.9	1.5	1.5	1.9	2.4
	M6a	1.8	1.5	2.8	2.8	1.9	1.5	2.6	2.6	1.5	2.0	2.4	2.4
12	M3a			1.5	1.4			1.8	1.4			2.4	1.9
	M4		2.5	3.6	3.6		2.5	3.5	3.6		2.5	2.9	3.6
	M5	1.6	2.0	2.4	2.4	1.5	2.0	2.6	2.3	1.5	1.8	2.4	2.3
	M6a	1.4	1.5	2.5	2.6	1.5	1.5	2.1	2.8	1.5	1.8	2.4	2.3
18	M3a			1.8	1.8			1.8	1.8			2.3	1.8
	M4		2.0	3.3	3.3		2.3	3.3	3.3		2.0	3.3	3.3
	M5	1.5	2.5	2.5	2.3	1.4	2.3	2.5	2.4	1.5	2.0	2.3	2.5
	M6a	1.5	1.5	2.5	2.8	1.6	1.5	2.5	2.6	1.5	2.0	2.3	2.5
24	M3a			1.8	1.8			1.8	1.8			2.4	2.1
	M4		2.0	3.3	3.3		2.3	3.3	3.3		2.0	2.9	3.3
	M5	1.4	2.5	2.5	2.3	1.4	2.3	2.5	2.4	1.4	2.0	2.4	2.5
	M6a	1.6	1.5	2.5	2.8	1.6	1.5	2.5	2.6	1.6	2.0	2.4	2.1

Table 4.3.5-22

Kruskal-Wallis Test

Mean Ranks of Power Results for Five Groups, α .01, by

Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			1.6	1.4			1.9	1.5			2.4	2.3
	M4		2.5	3.6	3.6		2.5	3.6	3.6		2.5	2.9	3.3
	M5	1.4	1.5	2.1	2.4	1.4	1.5	2.3	2.1	1.5	1.5	2.4	2.3
	M6a	1.6	2.0	2.6	2.6	1.6	2.0	2.3	2.8	1.5	2.0	2.4	2.3
12	M3a			1.8	1.8			2.0	1.8			2.5	2.0
	M4		2.0	3.3	3.3		2.0	3.3	3.3		2.0	2.5	3.3
	M5	1.5	2.5	2.4	2.4	1.4	2.5	2.4	2.4	1.5	2.0	2.5	2.4
	M6a	1.5	1.5	2.6	2.6	1.6	1.5	2.4	2.6	1.5	2.0	2.5	2.4
18	M3a			1.9	1.8			2.1	1.8			2.4	2.0
	M4		1.8	3.3	3.3		2.0	2.9	3.3		2.0	2.9	3.3
	M5	1.4	2.5	2.5	2.5	1.4	2.0	2.5	2.4	1.5	2.0	2.4	2.4
	M6a	1.6	1.8	2.4	2.5	1.6	2.0	2.5	2.6	1.5	2.0	2.4	2.4
24	M3a			2.1	2.0			2.1	2.0			2.4	2.0
	M4		2.0	2.9	3.3		2.0	2.9	3.3		2.0	2.9	3.3
	M5	1.4	2.0	2.4	2.5	1.4	2.0	2.4	2.5	1.4	2.0	2.4	2.3
	M6a	1.6	2.0	2.6	2.3	1.6	2.0	2.6	2.3	1.8	2.0	2.4	2.5

Table 4.3.5-23

Kruskal-Wallis Test

Mean Ranks of Power Results for Six Groups, α .01, by

Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			1.8				2.1				2.4	
	M4		2.0	3.3	2.5		2.0	3.3	2.5		2.0	2.9	2.4
	M5	1.3	2.5	2.3	1.5	1.4	2.5	2.1	1.5	1.5	2.0	2.4	1.5
	M6a	1.8	1.5	2.8	2.0	1.6	1.5	2.5	2.0	1.5	2.0	2.4	2.1
12	M3a			1.9				2.1				2.4	
	M4		2.0	3.3	2.5		2.0	2.9	2.5		2.0	2.9	2.3
	M5	1.4	2.0	2.5	1.8	1.4	2.0	2.5	1.8	1.4	2.0	2.4	1.9
	M6a	1.6	2.0	2.4	1.8	1.6	2.0	2.5	1.8	1.6	2.0	2.4	1.9
18	M3a			2.1				2.1				2.4	
	M4		2.0	2.9	2.5		2.0	2.9	2.3		2.0	2.9	2.3
	M5	1.4	2.0	2.4	1.6	1.4	2.0	2.4	1.8	1.5	2.0	2.4	1.9
	M6a	1.6	2.0	2.6	1.9	1.6	2.0	2.6	2.0	1.5	2.0	2.4	1.9
24	M3a			2.1				2.1				2.4	
	M4		2.0	2.9	2.3		2.0	2.9	2.3		2.0	2.9	2.3
	M5	1.4	2.0	2.4	1.8	1.5	2.0	2.5	1.9	1.5	2.0	2.4	1.9
	M6a	1.6	2.0	2.6	2.0	1.5	2.0	2.5	1.9	1.5	2.0	2.4	1.9

Table 4.3.5-24
Kruskal-Wallis Test
Mean Ranks of Power Results for Three Groups, α .05, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			2.0	1.5			1.9	1.6			2.3	2.0
	M4		3.0	4.0	4.0		3.0	4.0	4.0		2.8	3.3	3.4
	M5	1.0	1.5	1.0	1.5	1.0	1.5	1.1	1.5	1.5	1.5	2.3	2.1
	M6a	2.0	1.5	3.0	3.0	2.0	1.5	3.0	2.9	1.5	1.8	2.3	2.5
12	M3a			1.3	1.1			1.5	1.4			2.0	1.9
	M4		2.5	4.0	4.0		2.8	3.6	3.6		2.5	3.3	3.6
	M5	1.6	2.0	2.3	2.5	1.5	1.8	2.1	2.4	1.5	1.8	2.4	2.3
	M6a	1.4	1.5	2.5	2.4	1.5	1.5	2.8	2.6	1.5	1.8	2.4	2.3
18	M3a			1.4	1.4			1.8	1.4			2.0	1.8
	M4		2.5	3.6	3.6		2.5	3.6	3.6		2.5	3.3	3.3
	M5	1.5	2.0	2.4	2.4	1.5	2.0	2.1	2.3	1.4	1.8	2.4	2.5
	M6a	1.5	1.5	2.6	2.6	1.5	1.5	2.5	2.8	1.6	1.8	2.4	2.5
24	M3a			1.4	1.4			1.4	1.6			2.3	1.8
	M4		2.5	3.6	3.6		2.5	3.4	3.6		2.5	3.3	3.3
	M5	1.5	2.0	2.5	2.4	1.4	2.0	2.6	2.3	1.4	1.8	2.3	2.5
	M6a	1.5	1.5	2.5	2.6	1.6	1.5	2.6	2.5	1.6	1.8	2.3	2.5

Table 4.3.5-25
Kruskal-Wallis Test
Mean Ranks of Power Results for Four Groups, α .05, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			1.5	1.1			1.4	1.4			2.3	2.0
	M4		2.5	3.6	4.0		2.5	4.0	3.6		2.5	3.3	3.6
	M5	1.3	1.5	2.3	2.5	1.3	1.5	2.0	2.3	1.4	1.8	2.3	2.0
	M6a	1.8	2.0	2.6	2.4	1.8	2.0	2.6	2.8	1.6	1.8	2.3	2.4
12	M3a			1.8	1.6			1.8	1.6			2.3	1.8
	M4		2.5	3.3	3.6		2.5	3.3	3.6		2.0	3.3	3.3
	M5	1.4	2.0	2.4	2.1	1.4	2.0	2.4	2.3	1.5	2.0	2.3	2.5
	M6a	1.6	1.5	2.6	2.65	1.6	1.5	2.6	2.5	1.5	2.0	2.3	2.5
18	M3a			1.8	1.8			1.8	1.8			2.3	1.8
	M4		2.0	3.3	3.3		2.0	3.3	3.3		2.0	3.3	3.3
	M5	1.5	2.5	2.5	2.5	1.4	2.5	2.4	2.5	1.5	2.0	2.3	2.5
	M6a	1.5	1.5	2.5	2.5	1.6	1.5	2.6	2.5	1.5	2.0	2.3	2.5
24	M3a			1.8	1.8			2.0	1.8			2.4	1.8
	M4		1.8	3.3	3.3		2.0	3.3	3.3		2.0	2.9	3.3
	M5	1.4	2.5	2.5	2.5	1.4	2.0	2.3	2.4	1.5	2.0	2.4	2.5
	M6a	1.6	1.8	2.5	2.5	1.6	2.0	2.5	2.6	1.5	2.0	2.4	2.5

Table 4.3.5-26
Kruskal-Wallis Test
Mean Ranks of Power Results for Five Groups, α .05, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			1.6				1.9				2.3	
	M4		2.5	3.6	2.8		1.5	3.3	2.8		2.5	3.3	2.5
	M5	1.4	2.0	2.1	1.5	1.4	2.5	2.3	1.4	1.5	2.0	2.3	1.8
	M6a	1.6	1.5	2.6	1.8	1.6	2.0	2.6	1.9	1.5	1.5	2.3	1.8
12	M3a			1.8				2.0				2.4	
	M4		2.0	3.3	2.5		2.0	3.3	2.5		2.0	2.9	2.5
	M5	1.4	2.5	2.4	1.8	1.4	2.0	2.3	1.8	1.5	2.0	2.4	1.8
	M6a	1.6	1.5	2.6	1.8	1.6	2.0	2.5	1.8	1.5	2.0	2.4	1.8
18	M3a			2.1				2.1				2.4	
	M4		2.0	2.9	2.5		2.0	2.9	2.5		2.0	2.9	2.3
	M5	1.4	2.0	2.4	1.8	1.5	2.0	2.5	1.8	1.5	2.0	2.4	1.9
	M6a	1.6	2.0	2.6	1.8	1.5	2.0	2.5	1.8	1.5	2.0	2.4	1.9
24	M3a			2.1				2.1				2.4	
	M4		1.0	2.9	2.5		2.0	2.9	2.3		2.0	2.9	2.3
	M5	1.4	2.0	2.4	1.6	1.5	2.0	2.5	1.8	1.5	2.0	2.4	1.9
	M6a	1.6	2.0	2.6	1.9	1.5	2.0	2.5	2.0	1.5	2.0	2.4	1.9

Table 4.3.5-27

Kruskal-Wallis Test

Mean Ranks of Power Results for Six Groups, α .05, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
6	M3a			1.8				1.9				2.5	
	M4		2.0	3.3	2.5		2.3	3.3	2.5		2.0	2.5	2.5
	M5	1.5	2.5	2.5	1.6	1.4	2.3	2.3	1.6	1.4	2.0	2.5	1.6
	M6a	1.5	1.5	2.5	1.9	1.6	1.5	2.6	1.9	1.6	2.0	2.5	1.9
12	M3a			2.0				2.1				2.4	
	M4		2.0	3.3	2.5		2.0	2.9	2.4		2.0	2.9	2.5
	M5	1.4	2.0	2.3	1.8	1.4	2.0	2.5	1.9	1.5	2.0	2.4	1.8
	M6a	1.6	2.0	2.5	1.8	1.6	2.0	2.5	1.8	1.5	2.0	2.4	1.8
18	M3a			2.1				2.1				2.4	
	M4		2.0	2.9	2.5		2.0	2.9	2.3		2.0	2.9	2.3
	M5	1.4	2.0	2.4	1.6	1.5	2.0	2.5	1.8	1.5	2.0	2.4	1.9
	M6a	1.6	2.0	2.6	1.9	1.5	2.0	2.5	2.0	1.5	2.0	2.4	1.9
24	M3a			2.1				2.1				2.4	
	M4		2.0	2.9	2.3		2.0	2.9	2.3		2.0	2.9	2.3
	M5	1.5	2.0	2.4	1.8	1.5	2.0	2.5	1.9	1.5	2.0	2.4	1.9
	M6a	1.5	2.0	2.6	2.0	1.5	2.0	2.5	1.9	1.5	2.0	2.4	1.9

Table 4.3.5-28
Kruskal-Wallis Test
Mean Ranks of Power Results for Three Groups, α .01, by Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			1.5				1.8				2.4	
	M4			4.0				3.6				2.9	
	M5	1.0		1.5	1.0	1.0		2.1	1.1	1.5		2.4	1.5
	M6a	2.0		3.0	2.0	2.0		2.5	1.9	1.5		2.4	1.5
0.5	M3a			1.4				1.5				2.1	
	M4		2.8	4.0			2.8	4.0			2.5	3.6	
	M5	1.5	1.3	1.8	1.0	1.4	1.3	2.0	1.0	1.4	1.4	2.1	1.5
	M6a	1.5	2.0	2.9	2.0	1.6	2.0	2.5	2.0	1.6	2.1	2.1	1.5
0.8	M3a			1.3				1.8				2.3	
	M4			4.0				4.0				3.3	
	M5	1.5		2.1	1.6	1.5		1.9	1.5	1.4		2.3	1.5
	M6a	1.5		2.6	1.4	1.5		2.4	1.5	1.6		2.3	1.5
1.2	M3a			2.0				2.3				2.4	
	M4		2.1	3.3			2.3	3.3			2.3	2.9	
	M5	1.5	2.3	2.3	1.5	1.4	2.1	2.0	1.5	1.5	1.8	2.4	1.4
	M6a	1.5	1.6	2.5	1.5	1.6	1.6	2.5	1.5	1.5	2.0	2.4	1.6

Table 4.3.5-29

Kruskal-Wallis Test

Mean Ranks of Power Results for Four Groups, α .01, by Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			1.3	1.3			1.5	1.1			2.4	1.9
	M4			4.0	4.0			4.0	4.0			3.4	3.6
	M5	1.0		1.8	1.8	1.0		2.3	1.9	1.4		1.9	2.3
	M6a	2.0		3.0	3.0	2.0		2.3	3.0	1.6		2.4	2.3
0.5	M3a			1.3	1.0			1.1	1.1			2.1	1.1
	M4		2.5	4.0	4.0		2.8	4.0	4.0		2.5	3.6	4.0
	M5	1.6	2.3	2.3	2.0	1.5	2.0	2.3	2.1	1.5	1.6	2.1	2.6
	M6a	1.4	1.3	2.5	3.0	1.5	1.3	2.6	2.8	1.5	1.9	2.1	2.3
0.8	M3a			1.9	1.8			1.9	1.8			2.4	2.3
	M4			3.3	3.3			3.1	3.3			2.9	3.3
	M5	1.5		2.5	2.5	1.4		2.8	2.5	1.5		2.4	2.3
	M6a	1.5		2.4	2.5	1.6		2.3	2.5	1.5		2.4	2.3
1.2	M3a			2.1	2.1			2.1	2.4			2.5	2.5
	M4		2.0	2.9	2.9		2.1	2.6	2.9		2.0	2.5	2.5
	M5	1.6	2.3	2.6	2.6	1.5	2.1	2.6	2.4	1.5	2.0	2.5	2.5
	M6a	1.4	1.8	2.4	2.4	1.5	1.8	2.6	2.4	1.5	2.0	2.5	2.5

Table 4.3.5-30
Kruskal-Wallis Test
Mean Ranks of Power Results for Five Groups, α .01, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			1.3	1.0			1.3	1.1			2.3	1.3
	M4			4.0	4.0			4.0	4.0			3.3	4.0
	M5	1.0		1.8	2.0	1.0		2.3	1.9	1.4		2.3	2.3
	M6a	2.0		3.0	3.0	2.0		2.5	3.0	1.6		2.3	2.5
0.5	M3a			1.5	1.3			2.0	1.3			2.4	2.0
	M4		2.1	3.6	4.0		2.3	3.3	4.0		2.3	2.9	4.0
	M5	1.5	2.3	2.5	2.6	1.5	2.0	2.4	2.5	1.5	1.8	2.4	2.0
	M6a	1.5	1.6	2.4	2.1	1.5	1.8	2.4	2.3	1.5	2.0	2.4	2.0
0.8	M3a			2.1	2.1			2.4	2.1			2.5	2.5
	M4			2.9	2.9			2.9	2.9			2.5	2.5
	M5	1.6		2.6	2.6	1.5		2.4	2.5	1.5		2.5	2.5
	M6a	1.4		2.4	2.4	1.5		2.4	2.5	1.5		2.5	2.5
1.2	M3a			2.5	2.5			2.5	2.5			2.5	2.5
	M4		2.0	2.5	2.5		2.0	2.5	2.5		2.0	2.5	2.5
	M5	1.5	2.0	2.5	2.5	1.5	2.0	2.5	2.5	1.5	2.0	2.5	2.5
	M6a	1.5	2.0	2.5	2.5	1.5	2.0	2.5	2.5	1.5	2.0	2.5	2.5

Table 4.3.5-31
Kruskal-Wallis Test
Mean Ranks of Power Results for Six Groups, α .01, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			1.0				1.3				2.1	
	M4			4.0	3.0			4.0	3.0			3.6	2.9
	M5	1.0		2.0	1.0	1.1		2.3	1.1	1.4		2.1	1.4
	M6a	2.0		3.0	2.0	1.9		2.5	1.9	1.6		2.1	1.8
0.5	M3a			1.9				2.3				2.4	
	M4		2.0	3.3	2.8		2.0	2.9	2.5		2.0	2.9	2.3
	M5	1.4	2.3	2.5	1.6	1.5	2.3	2.3	1.8	1.5	2.0	2.4	1.8
	M6a	1.6	1.8	2.4	1.6	1.5	1.8	2.6	1.8	1.5	2.0	2.4	2.0
0.8	M3a			2.5				2.5				2.5	
	M4			2.5	2.0			2.5	2.0			2.5	2.0
	M5	1.5		2.5	2.0	1.5		2.5	2.0	1.5		2.5	2.0
	M6a	1.5		2.5	2.0	1.5		2.5	2.0	1.5		2.5	2.0
1.2	M3a			2.5				2.5				2.5	
	M4		2.0	2.5	2.0		2.0	2.5	2.0		2.0	2.5	2.0
	M5	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0
	M6a	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0

Table 4.3.5-32

Kruskal-Wallis Test

Mean Ranks of Power Results for Three Groups, α .05, by Effect Size, Method and Distribution across Initial Sample Size

Decimal		| ----------4th ----------				| ---------- 3rd ----------				| ---------- 2nd -------- |		
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			1.3	1.3			1.5	1.3			2.1	1.4
	M4			4.0	4.0			4.0	4.0			3.6	3.4
	M5	1.0		1.8	1.8	1.0		1.6	1.8	1.3		2.1	2.6
	M6a	2.0		3.0	3.0	2.0		2.9	3.0	1.8		2.1	2.6
0.5	M3a			1.3	1.3			1.3	1.3			1.5	1.4
	M4		3.0	4.0	4.0		3.0	4.0	4.0		2.9	4.0	4.0
	M5	1.8	1.5	2.1	1.8	1.6	1.5	2.0	1.9	1.5	1.4	2.3	2.1
	M6a	1.3	1.5	2.6	3.0	1.4	1.5	2.8	2.9	1.5	1.8	2.3	2.5
0.8	M3a			1.3	1.0			1.5	1.4			2.4	2.3
	M4			4.0	4.0			3.8	4.0			2.9	3.3
	M5	1.5		2.3	2.8	1.4		2.1	2.3	1.5		2.4	2.3
	M6a	1.5		2.5	2.3	1.6		2.6	2.4	1.5		2.4	2.3
1.2	M3a			2.3	1.9			2.3	2.1			2.5	2.4
	M4		2.3	3.3	3.3		2.4	2.9	2.9		2.3	2.5	2.9
	M5	1.4	2.3	2.0	2.5	1.4	2.1	2.3	2.5	1.5	2.0	2.5	2.4
	M6a	1.6	1.5	2.5	2.4	1.6	1.5	2.6	2.5	1.5	1.8	2.5	2.4

Table 4.3.5-33
Kruskal-Wallis Test
Mean Ranks of Power Results for Four Groups, α .05, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			1.1	1.0			1.1	1.0			2.0	1.3
	M4			4.0	4.0			4.0	4.0			4.0	4.0
	M5	1.0		1.9	2.0	1.0		1.9	2.0	1.4		2.0	2.4
	M6a	2.0		3.0	3.0	2.0		3.0	3.0	1.6		2.0	2.4
0.5	M3a			1.0	1.0			1.3	1.0			2.1	1.1
	M4		2.4	4.0	4.0		2.5	4.0	4.0		2.3	3.6	4.0
	M5	1.5	2.3	2.6	2.5	1.4	2.5	2.3	2.5	1.5	1.9	2.1	2.3
	M6a	1.5	1.4	2.4	2.5	1.6	1.5	2.5	2.5	1.5	1.9	2.1	2.6
0.8	M3a			2.1	2.0			2.1	2.0			2.5	2.4
	M4			2.9	3.3			2.9	3.3			2.5	2.9
	M5	1.6		2.6	2.5	1.5		2.5	2.4	1.5		2.5	2.4
	M6a	1.4		2.4	2.3	1.5		2.5	2.4	1.5		2.5	2.4
1.2	M3a			2.5	2.3			2.4	2.5			2.5	2.5
	M4		2.0	2.5	2.9		2.0	2.9	2.5		2.0	2.5	2.5
	M5	1.4	2.0	2.5	2.6	1.5	2.0	2.4	2.5	1.5	2.0	2.5	2.5
	M6a	1.6	2.0	2.5	2.3	1.5	2.0	2.4	2.5	1.5	2.0	2.5	2.5

Table 4.3.5-34
Kruskal-Wallis Test
Mean Ranks of Power Results for Five Groups, α .05, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		----------4th----------				---------- 3rd ----------				---------- 2nd --------			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			1.0				1.1				2.0	
	M4			4.0	3.0			4.0	3.0			4.0	3.0
	M5	1.0		2.0	1.0	1.3		2.1	1.0	1.5		2.0	1.5
	M6a	2.0		3.0	2.0	1.8		2.8	2.0	1.5		2.0	1.5
0.5	M3a			1.8				2.0				2.4	
	M4		2.3	3.3	3.0		1.8	3.3	2.8		2.3	2.9	2.5
	M5	1.5	2.3	2.4	1.6	1.5	2.3	2.4	1.8	1.5	2.0	2.4	1.8
	M6a	1.5	1.5	2.6	1.4	1.5	2.0	2.4	1.5	1.5	1.8	2.4	1.8
0.8	M3a			2.4				2.5				2.5	
	M4			2.9	2.3			2.5	2.3			2.5	2.0
	M5	1.5		2.4	2.0	1.5		2.5	1.9	1.5		2.5	2.0
	M6a	1.5		2.4	1.8	1.5		2.5	1.9	1.5		2.5	2.0
1.2	M3a			2.5				2.5				2.5	
	M4		2.0	2.5	2.0		2.0	2.5	2.0		2.0	2.5	2.0
	M5	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0
	M6a	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0

Table 4.3.5-35
Kruskal-Wallis Test
Mean Ranks of Power Results for Six Groups, α .05, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		----------4th ----------				---------- 3rd ----------				---------- 2nd --------			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a			1.0				1.1				2.1	
	M4			4.0	3.0			4.0	3.0			3.6	3.0
	M5	1.1		2.0	1.0	1.3		2.3	1.1	1.4		2.1	1.4
	M6a	1.9		3.0	2.0	1.8		2.6	1.9	1.6		2.1	1.6
0.5	M3a			2.0				2.1				2.5	
	M4		2.0	3.3	2.8		2.1	2.9	2.4		2.0	2.5	2.5
	M5	1.6	2.3	2.5	1.8	1.5	2.1	2.5	2.0	1.5	2.0	2.5	1.8
	M6a	1.4	1.8	2.3	1.5	1.5	1.8	2.5	1.6	1.5	2.0	2.5	1.8
0.8	M3a			2.5				2.5				2.5	
	M4			2.5	2.0			2.5	2.0			2.5	2.0
	M5	1.5		2.5	2.0	1.5		2.5	2.0	1.5		2.5	2.0
	M6a	1.5		2.5	2.0	1.5		2.5	2.0	1.5		2.5	2.0
1.2	M3a			2.5				2.5				2.5	
	M4		2.0	2.5	2.0		2.0	2.5	2.0		2.0	2.5	2.0
	M5	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0
	M6a	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0	1.5	2.0	2.5	2.0

Table 4.3.5-36
Kruskal-Wallis Test
Mean Ranks of Power Results for Three Groups, α .01, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution / *Mthd*	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
M3a			1.5				1.8				2.3	
M4		2.4	3.8			2.5	3.7			2.4	3.2	
M5	1.4	1.8	1.9	1.3	1.3	1.7	2.0	1.3	1.4	1.6	2.3	1.5
M6a	1.6	1.8	2.8	1.7	1.7	1.8	2.5	1.7	1.6	2.1	2.3	1.5

Table 4.3.5-37
Kruskal-Wallis Test
Mean Ranks of Power Results for Four Groups, α .01, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution / *Mthd*	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
M3a			1.6	1.5			1.7	1.6			2.3	1.9
M4		2.3	3.5	3.5		2.4	3.4	3.5		2.3	3.1	3.3
M5	1.4	2.3	2.3	2.2	1.3	2.1	2.5	2.2	1.5	1.8	2.2	2.4
M6a	1.6	1.5	2.6	2.7	1.7	1.5	2.4	2.7	1.5	1.9	2.3	2.3

Table 4.3.5-38
Kruskal-Wallis Test
Mean Ranks of Power Results for Five Groups, α .01, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	----------4th ----------				---------- 3rd ----------				---------- 2nd --------			
Distribution / *Mthd*	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
M3a			1.8	1.7			2.0	1.8			2.4	2.1
M4		2.1	3.3	3.3		2.1	3.2	3.3		2.1	2.8	3.3
M5	1.4	2.1	2.3	2.4	1.4	2.0	2.4	2.3	1.5	1.9	2.4	2.3
M6a	1.6	1.8	2.6	2.5	1.6	1.9	2.4	2.6	1.5	2.0	2.4	2.4

Table 4.3.5-39
Kruskal-Wallis Test
Mean Ranks of Power Results for Six Groups, α .01, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	----------4th ----------				---------- 3rd ----------				---------- 2nd --------			
Distribution / *Mthd*	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
M3a			2.0				2.1				2.4	
M4		2.0	3.1	2.4		2.0	3.0	2.4		2.0	2.9	2.3
M5	1.3	2.1	2.4	1.7	1.4	2.1	2.4	1.7	1.5	2.0	2.4	1.8
M6a	1.7	1.9	2.6	1.9	1.6	1.9	2.5	1.9	1.5	2.0	2.4	1.9

Table 4.3.5-40
Kruskal-Wallis Test
Mean Ranks of Power Results for Three Groups, α .05, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a			1.5	1.3			1.6	1.5			2.1	1.8
M4		2.6	3.8	3.8		2.7	3.7	3.7		2.6	3.3	3.4
M5	1.4	1.9	2.0	2.2	1.3	1.8	2.0	2.1	1.4	1.7	2.3	2.3
M6a	1.6	1.5	2.7	2.7	1.7	1.5	2.7	2.7	1.6	1.8	2.3	2.4

Table 4.3.5-41
Kruskal-Wallis Test
Mean Ranks of Power Results for Four Groups, α .05, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a			1.7	1.6			1.7	1.6			2.3	1.8
M4		2.2	3.3	3.5		2.3	3.4	3.4		2.1	3.2	3.3
M5	1.4	2.1	2.4	2.4	1.3	2.0	2.3	2.3	1.5	1.9	2.3	2.4
M6a	1.6	1.7	2.6	2.5	1.7	1.8	2.6	2.6	1.5	1.9	2.3	2.5

Table 4.3.5-42
Kruskal-Wallis Test
Mean Ranks of Power Results for Five Groups, α .05, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a			1.9				2.0				2.3	
M4		2.1	3.2	2.6		1.9	3.1	2.5		2.1	3.0	2.4
M5	1.4	2.1	2.3	1.7	1.4	2.1	2.4	1.7	1.5	2.0	2.3	1.8
M6a	1.6	1.8	2.6	1.8	1.6	2.0	2.5	1.8	1.5	1.9	2.3	1.8

Table 4.3.5-43
Kruskal-Wallis Test
Mean Ranks of Power Results for Six Groups, α .05, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a			2.0				2.1				2.4	
M4		2.0	3.1	2.4		2.1	3.0	2.3		2.0	2.8	2.4
M5	1.4	2.1	2.4	1.7	1.4	2.1	2.4	1.8	1.5	2.0	2.4	1.8
M6a	1.6	1.9	2.6	1.9	1.6	1.9	2.5	1.9	1.5	2.0	2.4	1.8

Table 4.3.5-44
Kruskal-Wallis Test
Analysis of Mean Ranks of Power Results
Number of First Place Finishes by Distribution
Across Nominal Alpha and Groups (.01 4, .01 5, .05 3, .05 4)

Decimal	4th				2nd			
Method	3a	4	5	6a	3a	4	5	6a
By Initial Sample Size Across Nominal Effect Size Multiplier								
EA			11.50	4.50			10.50	5.50
EB		3.33	2.83	9.83		2.67	8.17	5.17
ML	16.00	0.00	0.00	0.00	6.25	0.25	5.25	4.25
SS	16.00	0.00	0.00	0.00	14.33	0.00	0.83	0.83
By Nominal Effect Size Multiplier Across Initial Sample Size								
EA			8.50	7.50			10.00	6.00
EB		0.67	1.17	6.17		1.00	4.50	2.50
ML	13.50	0.50	1.50	0.50	5.17	1.50	5.17	4.17
SS	14.75	0.25	0.25	0.75	9.67	1.00	2.67	2.67
Across Nominal Effect Size Multiplier and Initial Sample Size								
EA			4.00	0.00			2.50	1.50
EB		0.00	0.00	4.00		0.00	3.50	0.50
ML	4.00	0.00	0.00	0.00	1.67	0.00	1.67	0.67
SS	4.00	0.00	0.00	0.00	4.00	0.00	0.00	0.00

Table 4.3.5-45
Kruskal-Wallis Test
Analysis of Mean Ranks of Power Results
Number of First Place Finishes by Distribution
Across Nominal Alpha and Groups (.01 6, .05 5, .05 6)

Decimal	4th				2nd			
Method	3a	4	5	6a	3a	4	5	6a
By Initial Sample Size Across Nominal Effect Size Multiplier								
EA			11.00	1.00			7.00	5.00
EB		2.67	2.67	6.67		3.67	4.67	3.67
ML	12.00	0.00	0.00	0.00	3.92	0.25	3.92	3.92
SS		0.00	10.00	2.00		0.00	7.00	5.00
By Nominal Effect Size Multiplier Across Initial Sample Size								
EA			7.50	4.50			7.00	5.00
EB		1.00	1.00	4.00		1.67	1.67	2.67
ML	7.58	1.25	1.58	1.58	3.42	1.75	3.42	3.42
SS		1.67	5.17	5.17		2.00	6.50	3.50
Across Nominal Effect Size Multiplier and Initial Sample Size								
EA			3.00	0.00			1.50	1.50
EB		0.00	0.00	3.00		0.67	0.67	1.67
ML	3.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00
SS		0.00	3.00	0.00		0.00	2.00	1.00

Table 4.3.5-46
Kruskal-Wallis Test
Analysis of Mean Ranks of Power Results
Across Nominal Alpha and Groups (.01 4, .01 5, .05 3, .05 4) and Distributions

Decimal	| ------------- 4th	------------- |			| ------------- 2nd	------------- |		
Method	3a	4	5	6a	3a	4	5	6a
By Initial Sample Size across Nominal Effect Size Multiplier								
MP1i	32	48	64	64	32	48	64	64
N1Mi	32.00	3.33	14.33	14.33	20.58	2.92	24.75	15.75
PoM	1.000	0.069	0.224	0.224	0.643	0.061	0.387	0.246
PoT	0.500	0.052	0.224	0.224	0.322	0.046	0.387	0.246
By Nominal Effect Size Multiplier across Initial Sample Size								
MP1i	56	40	56	56	56	40	56	56
N1Mi	28.25	1.42	11.42	14.92	14.83	3.50	22.33	15.33
PoM	0.883	0.035	0.204	0.266	0.464	0.088	0.399	0.274
PoT	0.504	0.025	0.204	0.266	0.265	0.063	0.399	0.274
By Initial Sample Size and Nominal Effect Size Multiplier								
MP1i	8	12	16	16	8	12	16	16
N1Mi	8.0	0.0	4.0	4.0	5.67	0.00	7.67	2.67
PoM	1.000	0.000	0.250	0.250	0.708	0.000	0.479	0.167
PoT	0.500	0.000	0.250	0.250	0.354	0.000	0.479	0.167

Key: MP1i: Maximum possible 1st place finishes for each method
N1Mi: Actual number of 1st place finishes for each method (ties count 1/n)
PoM: Proportion of Maximum Possible { N1Mi / MP1i }
PoT: Proportion of total 1st place finishes { N1Mi / Σ(N1Mi) }

Table 4.3.4-47
Kruskal-Wallis Test
Analysis of Mean Ranks of Power Results
Across Nominal Alpha and Groups (.01 6, .05 5, .05 6) and Distributions

Decimal	4th				2nd			
Method	3a	4	5	6a	3a	4	5	6a
By Initial Sample Size across Nominal Effect Size Multiplier								
MP1i	12	36	48	48	12	36	48	48
N1Mi	12.00	2.67	23.67	9.67	3.92	3.92	22.58	17.58
PoM	1.000	0.074	0.493	0.201	0.326	0.109	0.470	0.366
PoT	0.250	0.056	0.493	0.201	0.082	0.082	0.470	0.366
By Nominal Effect Size Multiplier across Initial Sample Size								
MP1i	12	30	42	42	12	30	42	42
N1Mi	7.58	3.92	15.25	15.25	3.42	5.42	18.58	14.58
PoM	0.632	0.131	0.363	0.363	0.285	0.181	0.442	0.347
PoT	0.181	0.093	0.363	0.363	0.081	0.129	0.442	0.347
By Initial Sample Size and Nominal Effect Size Multiplier								
MP1i	3	9	12	12	3	9	12	12
N1Mi	3.0	0.0	6.0	3.0	1.00	0.67	5.17	5.17
PoM	1.000	0.000	0.500	0.250	0.333	0.074	0.431	0.431
PoT	0.250	0.000	0.500	0.250	0.083	0.056	0.431	0.431

Key: MP1i: Maximum possible 1st place finishes for each method
N1Mi: Actual number of 1st place finishes for each method (ties count 1/n)
PoM: Proportion of Maximum Possible { N1Mi / MP1i }
PoT: Proportion of total 1st place finishes { N1Mi / Σ(N1Mi) }

4.3.6 – Terpstra-Jonckheere Test (three to six groups)

Based on the Type I error results, power results for the Terpstra-Jonckheere Test are only presented for the combinations of alpha level, distribution and method shown in Table 4.3.6-1.

Table 4.3.6-1
Terpstra-Jonckheere Test for Three to Six Groups
Power and Type III Error for α .01 and .05
Method / Distribution Combinations with Acceptable Type I Error

Alpha	.01					.05				
Dist	Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
Method										
Three Groups										
M3a						x	x	x	x	x
M4						x	x	x	x	x
M5						x	x	x	x	x
M6a						x	x	x	x	x
Four Groups										
M3a	x	x	x	x	x	x	x	x	x	x
M4	x	x	x	x	x	x	x	x	x	x
M5	x	x	x	x	x	x	x	x	x	x
M6a	x	x	x	x	x	x	x	x	x	x
Five Groups										
M3a	x	x	x	x	x	x	x	x	x	x
M4	x	x	x	x	x	x	x	x	x	x
M5	x	x	x	x	x	x	x	x	x	x
M6a	x	x	x	x	x	x	x	x	x	x
Six Groups										
M3a	x	x	x	x	x	x	x	x	x	x
M4	x	x	x	x	x	x	x	x	x	x
M5	x	x	x	x	x	x	x	x	x	x
M6a	x	x	x	x	x	x	x	x	x	x

Power results for the Terpstra-Jonckheere Test at nominal alpha .01 and .05 are summarized in Table 4.3.6-2 and 4.3.6-3. These tables present the range of values obtained (minimum and maximum) based on the results presented in Tables 4.3.6-4 through 4.3.6-23. The concept of Type III error does not apply to the Terpstra-Jonckheere Test as it is an inherently directional test against an ordered alternative hypothesis.

Table 4.3.6-2
Terpstra-Jonckheere Test for Three to Six Groups
Range of Power for α .01

Min/Max	min (ISS = 2, NESM = 0.2)					max (ISS = 10, NESM = 1.2)				
Dist	Norm	EA	EB*	ML	SS	Norm	EA	EB	ML	SS
Grps | Mthd										
Three Groups										
M3a										
M4										
M5										
M6a										
Four Groups										
M3a	.0178	.0203	.0905	.0136	.0130	.+	.+	.+	.+	.+
M4	.0178	.0284	.1135	.0166	.0181	.+	.+	.+	.+	.+
M5	.0178	.0284	.1104	.0163	.0179	.+	.+	.+	.+	.+
M6a	.0178	.0203	.0905	.0136	.0130	.+	.+	.+	.+	.+
Five Groups										
M3a	.0514	.0723	.3074	.0401	.0439	.+	.+	.+	.+	.+
M4	.0514	.0857	.3414	.0465	.0536	.+	.+	.+	.+	.+
M5	.0514	.0838	.3401	.0453	.0518	.+	.+	.+	.+	.+
M6a	.0514	.0723	.3074	.0401	.0439	.+	.+	.+	.+	.+
Six Groups										
M3a	.0867	.1289	.6120	.0648	.0777	.+	.+	.+	.+	.+
M4	.0867	.1483	.6534	.0740	.0921	.+	.+	.+	.+	.+
M5	.0867	.1408	.6362	.0711	.0868	.+	.+	.+	.+	.+
M6a	.0867	.1289	.6120	.0648	.0777	.+	.+	.+	.+	.+

* ESM = 0.5, .+ = 1.0000

Table 4.3.6-3
Terpstra-Jonckheere Test for Three to Six Groups
Range of Power for α .05

Min/Max	min (ISS = 2, NESM = 0.2)					max (ISS = 10, NESM = 1.2)				
Dist	Norm	EA	EB*	ML	SS	Norm	EA	EB	ML	SS
Grps | Mthd										
Three Groups										
M3a	.0640	.0652	.1654	.0544	.0510	.9996	.9996	.9999	.9996	.9998
M4	.0640	.0837	.2040	.0610	.0637	.9996	.9995	.9997	.9996	.9997
M5	.0640	.0815	.1957	.0611	.0641	.9996	.9995	.9996	.9996	.9998
M6a	.0640	.0652	.1654	.0544	.0510	.9996	.9996	.9999	.9996	.9998
Four Groups										
M3a	.1308	.1538	.3973	.1097	.1132	.+	.+	.+	.+	.+
M4	.1308	.1687	.4237	.1191	.1258	.+	.+	.+	.+	.+
M5	.1308	.1749	.4388	.1206	.1310	.+	.+	.+	.+	.+
M6a	.1308	.1538	.3973	.1097	.1132	.+	.+	.+	.+	.+
Five Groups										
M3a	.1680	.2154	.6308	.1387	.1521	.+	.+	.+	.+	.+
M4	.1680	.2319	.6358	.1506	.1659	.+	.+	.+	.+	.+
M5	.1680	.2336	.6548	.1504	.1678	.+	.+	.+	.+	.+
M6a	.1680	.2154	.6308	.1387	.1521	.+	.+	.+	.+	.+
Six Groups										
M3a	.2428	.3159	.9082	.1967	.2285	.+	.+	.+	.+	.+
M4	.2428	.3386	.9132	.2118	.2473	.+	.+	.+	.+	.+
M5	.2428	.3308	.9065	.2082	.2431	.+	.+	.+	.+	.+
M6a	.2428	.3159	.9082	.1967	.2285	.+	.+	.+	.+	.+

* ESM = 0.5, .+ = 1.0000

Up to the maximum of 1.0000, power tended to increase monotonically across methods and distributions with increases in initial sample size and/or effect size. Thus, minimum power usually occurred at initial sample size 2 at nominal ESM 0.2, with maximum power at initial sample size 10 and nominal ESM 1.2, if not sooner. Power results typically reached 1.0000 prior to a nominal effect size multiplier of 1.2 and initial sample size of 10. This was because the effect shift was applied between each pair of adjacent groups. Thus for six groups at nominal effect size multiplier 0.2, the nominal effect size multiplier between groups 1 and 6 (lowest and highest scores) was actually 1.0, irrespective of the sample size. At nominal effect size multiplier 1.2, the nominal effect size multiplier between these same two groups was 6.0. Under these conditions the test was going to be very powerful even with small sample sizes.

Tables 4.3.6-4 through 4.3.6-23 give the power results for the Terpstra-Jonckheere Test for both alpha .01 and .05. There is a table for each combination of number of groups {3, 4, 5, 6}, and nominal effect size multiplier {0.2, 0.5, 0.8, 1.2}. Each table covers all methods for initial sample sizes {2, 4, 6, 8, 10}. Results are not reported for distribution EB with nominal ESM 0.2 as the actual ESM = 0.0 (no shift), which is just Type I error. Also for distribution EB, results are not reported for nominal ESM 0.8 as the actual ESM = 0.592, the same as the actual ESM for nominal ESM 0.5. Results for the normal distribution are included and are essentially identical across methods for a fixed initial sample size and effect size. This is to be expected in the absence of ties and demonstrates that the simulations worked correctly.

A ranking analysis of the results is presented in Tables 4.3.6-24 through 4.3.6-35. These summaries were obtained by ranking the power results from Tables 4.3.6-3 through 4.3.6-22 to four decimal places (as reported), to three decimal places and to two decimal places.

Table 4.3.6-4

Terpstra-Jonckheere Test (Three Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a						.0640	.0652		.0544	.0510
	M4						.0640	.0837		.0610	.0637
	M5						.0640	.0815		.0611	.0641
	M6a						.0640	.0652		.0544	.0510
4	M3a	.0322	.0416		.0265	.0279	.1203	.1503		.1053	.1102
	M4	.0322	.0450		.0283	.0296	.1203	.1526		.1086	.1106
	M5	.0322	.0477		.0294	.0320	.1203	.1631		.1131	.1202
	M6a	.0322	.0412		.0265	.0279	.1203	.1503		.1053	.1102
6	M3a	.0418	.0628		.0359	.0390	.1556	.2114		.1395	.1488
	M4	.0418	.0628		.0361	.0377	.1556	.2037		.1382	.1421
	M5	.0418	.0679		.0382	.0421	.1556	.2199		.1448	.1553
	M6a	.0418	.0628		.0359	.0390	.1556	.2114		.1395	.1488
8	M3a	.0549	.0889		.0469	.0523	.1801	.2569		.1621	.1751
	M4	.0549	.0851		.0457	.0487	.1801	.2424		.1580	.1641
	M5	.0549	.0933		.0487	.0548	.1801	.2630		.1660	.1802
	M6a	.0549	.0889		.0469	.0523	.1801	.2569		.1621	.1751
10	M3a	.0623	.1067		.0541	.0603	.2101	.3050		.1914	.2065
	M4	.0623	.1000		.0519	.0554	.2101	.2843		.1852	.1921
	M5	.0623	.1102		.0556	.0624	.2101	.3096		.1946	.2106
	M6a	.0623	.1067		.0541	.0603	.2101	.3050		.1914	.2065

Table 4.3.6-5

Terpstra-Jonckheere Test (Three Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.5

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a						.1437	.2056	.1654	.1374	.0926
	M4						.1437	.2307	.2040	.1484	.1120
	M5						.1437	.2299	.1957	.1484	.1120
	M6a						.1437	.2056	.1654	.1374	.0926
4	M3a	.1272	.2342	.1736	.1216	.0749	.3377	.4919	.4302	.3337	.2360
	M4	.1272	.2389	.1856	.1276	.0788	.3377	.4904	.4339	.3409	.2360
	M5	.1272	.2504	.1944	.1305	.0842	.3377	.5081	.4502	.3471	.2512
	M6a	.1272	.2342	.1736	.1216	.0749	.3377	.4919	.4302	.3337	.2360
6	M3a	.2097	.3906	.3162	.2102	.1259	.4770	.6737	.6152	.4828	.3453
	M4	.2097	.3796	.3188	.2122	.1225	.4770	.6585	.6057	.4829	.3334
	M5	.2097	.4001	.3279	.2178	.1327	.4770	.6804	.6208	.4915	.3545
	M6a	.2097	.3906	.3162	.2102	.1259	.4770	.6737	.6152	.4828	.3453
8	M3a	.3054	.5413	.4593	.3121	.1869	.5826	.7861	.7344	.5913	.4283
	M4	.3054	.5191	.4553	.3099	.1774	.5826	.7668	.7220	.5871	.4103
	M5	.3054	.5468	.4662	.3179	.1924	.5826	.7892	.7356	.5967	.4345
	M6a	.3054	.5413	.4593	.3121	.1869	.5826	.7861	.7344	.5913	.4283
10	M3a	.3835	.6496	.5663	.3945	.2349	.6777	.8670	.8241	.6892	.5125
	M4	.3835	.6215	.5582	.3892	.2212	.6777	.8487	.8118	.6831	.4912
	M5	.3835	.6527	.5698	.3989	.2394	.6777	.8681	.8232	.6925	.5165
	M6a	.3835	.6496	.5663	.3945	.2349	.6777	.8670	.8241	.6892	.5125

Table 4.3.6-6
Terpstra-Jonckheere Test (Three Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.8

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a						.2659	.3585		.2484	.2322
	M4						.2659	.3832		.2621	.2658
	M5						.2659	.3835		.2623	.2656
	M6a						.2659	.3585		.2484	.2322
4	M3a	.3270	.4908		.3090	.3069	.6274	.7585		.6178	.6139
	M4	.3270	.4946		.3185	.3158	.6274	.7572		.6244	.6130
	M5	.3270	.5084		.3241	.3278	.6274	.7699		.6321	.6301
	M6a	.3270	.4908		.3090	.3069	.6274	.7585		.6178	.6139
6	M3a	.5397	.7276		.5322	.5314	.8145	.9119		.8143	.8120
	M4	.5397	.7171		.5339	.5231	.8145	.9053		.8130	.8019
	M5	.5397	.7348		.5425	.5416	.8145	.9145		.8201	.8167
	M6a	.5397	.7256		.5322	.5314	.8145	.9119		.8143	.8120
8	M3a	.7169	.8726		.7195	.7152	.9073	.9672		.9106	.9079
	M4	.7169	.8604		.7156	.7012	.9073	.9624		.9079	.8991
	M5	.7169	.8747		.7248	.7196	.9073	.9676		.9127	.9091
	M6a	.7169	.8726		.7195	.7152	.9073	.9672		.9106	.9079
10	M3a	.8252	.9399		.8288	.8261	.9576	.9894		.9598	.9588
	M4	.8252	.9307		.8239	.8125	.9576	.9869		.9878	.9533
	M5	.8252	.9405		.8315	.8278	.9576	.9895		.9606	.9591
	M6a	.8252	.9399		.8288	.8261	.9576	.9894		.9598	.9588

Table 4.3.6-7

Terpstra-Jonckheere Test (Three Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 1.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a						.4715	.4981	.3108	.3822	.4325
	M4						.4715	.5206	.3533	.3990	.4732
	M5						.4715	.5209	.3537	.3991	.4728
	M6a						.4715	.4981	.3108	.3822	.4325
4	M3a	.6741	.7125	.5345	.5877	.6607	.9034	.9089	.8682	.8758	.9036
	M4	.6741	.7149	.5270	.5988	.6701	.9034	.9077	.8502	.8784	.9031
	M5	.6741	.7267	.5711	.6049	.6801	.9034	.9144	.8741	.8840	.9096
	M6a	.6741	.7125	.5345	.5877	.6607	.9034	.9089	.8682	.8758	.9036
6	M3a	.9022	.9146	.8548	.8674	.9064	.9843	.9846	.9839	.9804	.9861
	M4	.9022	.9094	.8404	.8670	.9020	.9843	.9829	.9779	.9795	.9846
	M5	.9022	.9176	.8553	.8732	.9094	.9843	.9852	.9800	.9813	.9864
	M6a	.9022	.9146	.8548	.8674	.9064	.9843	.9846	.9839	.9804	.9861
8	M3a	.9775	.9801	.9704	.9688	.9804	.9975	.9974	.9983	.9969	.9980
	M4	.9775	.9771	.9641	.9674	.9782	.9975	.9969	.9972	.9966	.9976
	M5	.9775	.9805	.9661	.9699	.9806	.9975	.9974	.9971	.9970	.9980
	M6a	.9775	.9801	.9704	.9688	.9804	.9975	.9974	.9983	.9969	.9980
10	M3a	.9949	.9953	.9944	.9929	.9959	.9996	.9996	.9999	.9996	.9998
	M4	.9949	.9941	.9924	.9922	.9951	.9996	.9995	.9997	.9996	.9997
	M5	.9949	.9953	.9921	.9931	.9959	.9996	.9996	.9996	.9996	.9998
	M6a	.9949	.9953	.9944	.9929	.9959	.9996	.9996	.9999	.9996	.9998

Table 4.3.6-8

Terpstra-Jonckheere Test (Four Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.2

Alpha		|---------------- .01				----------------|	|---------------- .05				----------------|
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.0178	.0203		.0136	.0130	.1308	.1538		.1097	.1132
	M4	.0178	.0284		.0166	.0181	.1308	.1687		.1191	.1258
	M5	.0178	.0284		.0163	.0179	.1308	.1749		.1206	.1310
	M6a	.0178	.0203		.0136	.0130	.1308	.1538		.1097	.1132
4	M3a	.0615	.0947		.0502	.0570	.1807	.2494		.1570	.1726
	M4	.0615	.0961		.0524	.0579	.1807	.2455		.1593	.1696
	M5	.0615	.1020		.0537	.0616	.1807	.2594		.1637	.1814
	M6a	.0615	.0947		.0502	.0570	.1807	.2494		.1570	.1726
6	M3a	.0848	.1440		.0699	.0814	.2594	.3684		.2277	.2538
	M4	.0848	.1386		.0699	.0781	.2594	.3525		.2251	.2420
	M5	.0848	.1490		.0723	.0851	.2594	.3747		.2322	.2594
	M6a	.0848	.1440		.0699	.0814	.2594	.3684		.2277	.2538
8	M3a	.1187	.2078		.0977	.1152	.3208	.4595		.2833	.3173
	M4	.1187	.1953		.0955	.1082	.3208	.4354		.2772	.3007
	M5	.1187	.2121		.0998	.1184	.3208	.4632		.2865	.3211
	M6a	.1187	.2078		.0977	.1152	.3208	.4595		.2833	.3173
10	M3a	.1549	.2733		.1275	.1512	.3811	.5419		.3360	.3772
	M4	.1549	.2533		.1235	.1409	.3811	.5111		.3277	.3571
	M5	.1549	.2766		.1292	.1537	.3811	.5442		.3388	.3802
	M6a	.1549	.2733		.1275	.1512	.3811	.5419		.3360	.3772

Table 4.3.6-9

Terpstra-Jonckheere Test (Four Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.5

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.0771	.1435	.0905	.0697	.0386	.3688	.4841	.3973	.3421	.2482
	M4	.0771	.1677	.1135	.0781	.0509	.3688	.5050	.4237	.3580	.2688
	M5	.0771	.1677	.1104	.0772	.0503	.3688	.5100	.4388	.3603	.2759
	M6a	.0771	.1435	.0905	.0697	.0386	.3688	.4841	.3973	.3421	.2482
4	M3a	.3504	.5451	.4467	.3309	.2106	.6174	.7754	.7185	.5996	.4473
	M4	.3504	.5438	.4464	.3385	.2119	.6174	.7706	.7098	.6038	.4414
	M5	.3504	.5555	.4606	.3421	.2208	.6174	.7817	.7250	.6098	.4584
	M6a	.3504	.5451	.4467	.3309	.2106	.6174	.7754	.7185	.5996	.4473
6	M3a	.5439	.7604	.6825	.5261	.3448	.8120	.9247	.9013	.8015	.6423
	M4	.5439	.7477	.6741	.5261	.3352	.8120	.9175	.8944	.7995	.6270
	M5	.5439	.7638	.6850	.5322	.3511	.8120	.9257	.8994	.8051	.6468
	M6a	.5439	.7604	.6825	.5261	.3448	.8120	.9247	.9013	.8015	.6423
8	M3a	.7152	.8915	.8455	.7004	.4897	.9077	.9744	.9653	.9010	.7657
	M4	.7152	.8791	.8386	.6970	.4732	.9077	.9698	.9620	.8986	.7498
	M5	.7152	.8923	.8439	.7036	.4936	.9077	.9744	.9630	.9025	.7673
	M6a	.7152	.8915	.8455	.7004	.4897	.9077	.9744	.9653	.9010	.7657
10	M3a	.8328	.9550	.9317	.8222	.6176	.9566	.9919	.9886	.9530	.8513
	M4	.8328	.9464	.9272	.8181	.5985	.9566	.9898	.9872	.9511	.8375
	M5	.8328	.9553	.9291	.8237	.6198	.9566	.9919	.9874	.9533	.8518
	M6a	.8328	.9550	.9317	.8222	.6176	.9566	.9919	.9886	.9530	.8513

Table 4.3.6-10

Terpstra-Jonckheere Test (Four Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.8

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.2146	.3352		.1847	.1817	.6643	.7454		.6448	.6363
	M4	.2146	.3647		.1995	.2178	.6643	.7600		.6609	.6608
	M5	.2146	.3638		.1987	.2141	.6643	.7646		.6647	.6668
	M6a	.2146	.3352		.1847	.1817	.6643	.7454		.6448	.6363
4	M3a	.7745	.8751		.7832	.7734	.9324	.9672		.9467	.9346
	M4	.7745	.8728		.7871	.7731	.9324	.9654		.9466	.9315
	M5	.7745	.8801		.7927	.7812	.9324	.9685		.9496	.9364
	M6a	.7745	.8751		.7832	.7734	.9324	.9672		.9467	.9346
6	M3a	.9421	.9790		.9580	.9463	.9919	.9976		.9960	.9930
	M4	.9421	.9763		.9565	.9423	.9919	.9971		.9957	.9919
	M5	.9421	.9794		.9592	.9470	.9919	.9976		.9961	.9929
	M6a	.9421	.9790		.9580	.9463	.9919	.9976		.9960	.9930
8	M3a	.9894	.9975		.9946	.9906	.9990	.9998		.9997	.9993
	M4	.9894	.9969		.9941	.9892	.9990	.9998		.9997	.9991
	M5	.9894	.9976		.9948	.9905	.9990	.9998		.9997	.9992
	M6a	.9894	.9975		.9946	.9906	.9990	.9998		.9997	.9993
10	M3a	.9984	.9997		.9982	.9986	.9990	.9998		.9997	.9993
	M4	.9984	.9996		.9994	.9984	.9999	.+		.+	.+
	M5	.9984	.9997		.9995	.9986	.9999	.+		.+	.+
	M6a	.9984	.9997		.9995	.9986	.9990	.9998		.9997	.9993

+ = 1.0000

Table 4.3.6-11
Terpstra-Jonckheere Test (Four Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 1.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.4979	.5339	.2820	.3890	.4556	.9159	.9055	.9141	.9031	.9105
	M4	.4979	.5619	.3416	.4125	.5071	.9159	.9129	.9020	.9106	.9213
	M5	.4979	.5616	.3675	.4109	.4999	.9159	.9154	.9126	.9127	.9222
	M6a	.4979	.5339	.2820	.3890	.4556	.9159	.9055	.9141	.9031	.9105
4	M3a	.9866	.9851	.9959	.9902	.9896	.9988	.9985	.+	.9997	.9993
	M4	.9866	.9841	.9924	.9905	.9895	.9988	.9983	.9999	.9997	.9992
	M5	.9866	.9859	.9920	.9910	.9900	.9988	.9986	.9998	.9997	.9993
	M6a	.9866	.9851	.9959	.9902	.9896	.9988	.9985	.+	.9997	.9993
6	M3a	.9997	.9997	.+	.+	.9999	.+	.+	.+	.+	.+
	M4	.9997	.9996	.+	.+	.9999	.+	.+	.+	.+	.+
	M5	.9997	.9997	.+	.+	.9999	.+	.+	.+	.+	.+
	M6a	.9997	.9997	.+	.+	.9999	.+	.+	.+	.+	.+
8	M3a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M4	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
10	M3a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M4	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+

+ = 1.0000

Table 4.3.6-12
Terpstra-Jonckheere Test (Five Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.0514	.0723		.0401	.0439	.1680	.2154		.1387	.1521
	M4	.0514	.0857		.0465	.0536	.1680	.2319		.1506	.1659
	M5	.0514	.0838		.0453	.0518	.1680	.2336		.1504	.1678
	M6a	.0514	.0723		.0401	.0439	.1680	.2154		.1387	.1521
4	M3a	.1028	.1699		.0811	.0974	.3005	.4109		.2559	.2924
	M4	.1028	.1740		.0848	.1006	.3005	.4104		.2602	.2923
	M5	.1028	.1766		.0847	.1026	.3005	.4184		.2624	.3001
	M6a	.1028	.1699		.0811	.0974	.3005	.4109		.2559	.2924
6	M3a	.1816	.3010		.1451	.1776	.4218	.5703		.3632	.4171
	M4	.1816	.2959		.1467	.1762	.4218	.5594		.3634	.4103
	M5	.1816	.3060		.1478	.1817	.4218	.5735		.3674	.4212
	M6a	.1816	.3010		.1451	.1776	.4218	.5703		.3632	.4171
8	M3a	.2534	.4142		.2021	.2499	.5193	.6845		.4510	.5153
	M4	.2534	.4003		.2020	.2449	.5193	.6673		.4491	.5052
	M5	.2534	.4169		.2043	.2529	.5193	.6858		.4536	.5180
	M6a	.2534	.4142		.2021	.2499	.5193	.6845		.4510	.5153
10	M3a	.3299	.5236		.2647	.3282	.6040	.7728		.5286	.6027
	M4	.3299	.5022		.2632	.3203	.6040	.7526		.5255	.5915
	M5	.3299	.5254		.2666	.3306	.6040	.7728		.5304	.6041
	M6a	.3299	.5236		.2647	.3282	.6040	.7728		.5286	.6027

Table 4.3.6-13

Terpstra-Jonckheere Test (Five Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.5

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.2790	.4143	.3074	.2411	.1561	.5684	.6849	.6308	.5256	.3960
	M4	.2790	.4427	.3414	.2605	.1803	.5684	.7020	.6358	.5441	.4175
	M5	.2790	.4381	.3401	.2577	.1757	.5684	.7026	.6548	.5448	.4199
	M6a	.2790	.4143	.3074	.2411	.1561	.5684	.6849	.6308	.5256	.3960
4	M3a	.6564	.8144	.7862	.6235	.4332	.8875	.9505	.9609	.8760	.7351
	M4	.6564	.8143	.7835	.6303	.4385	.8875	.9493	.9567	.8773	.7330
	M5	.6564	.8183	.7878	.6313	.4418	.8875	.9514	.9572	.8791	.7397
	M6a	.6564	.8144	.7862	.6235	.4332	.8875	.9505	.9609	.8760	.7351
6	M3a	.8880	.9622	.9692	.8747	.6942	.9761	.9938	.9974	.9737	.8960
	M4	.8880	.9595	.9671	.8745	.6902	.9761	.9930	.9968	.9732	.8918
	M5	.8880	.9626	.9654	.8766	.6970	.9761	.9937	.9965	.9740	.8966
	M6a	.8880	.9622	.9692	.8747	.6942	.9761	.9938	.9974	.9737	.8960
8	M3a	.9668	.9930	.9964	.9620	.8436	.9953	.9992	.9999	.9947	.9604
	M4	.9668	.9918	.9960	.9614	.8383	.9953	.9991	.9998	.9945	.9576
	M5	.9668	.9930	.9954	.9623	.8443	.9953	.9992	.9997	.9947	.9604
	M6a	.9668	.9930	.9964	.9620	.8436	.9953	.9992	.9999	.9947	.9604
10	M3a	.9914	.9989	.9997	.9901	.9276	.9991	.+	.+	.9990	.9858
	M4	.9914	.9986	.9996	.9898	.9235	.9991	.9999	.+	.9990	.9846
	M5	.9914	.9988	.9995	.9902	.9272	.9991	.9999	.+	.9990	.9856
	M6a	.9914	.9989	.9997	.9901	.9276	.9991	.+	.+	.9990	.9858

+ = 1.0000

Table 4.3.6-14

Terpstra-Jonckheere Test (Five Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.8

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.6546	.7483		.6426	.6316	.8956	.9306		.9191	.8920
	M4	.6546	.7685		.6673	.6686	.8956	.9356		.9266	.9025
	M5	.6546	.7662		.6648	.6593	.8956	.9369		.9272	.9017
	M6a	.6546	.7483		.6426	.6316	.8956	.9306		.9191	.8920
4	M3a	.9786	.9920		.9929	.9821	.9982	.9995		.9998	.9988
	M4	.9786	.9914		.9930	.9823	.9982	.9994		.9998	.9987
	M5	.9786	.9923		.9933	.9824	.9982	.9995		.9999	.9987
	M6a	.9786	.9920		.9929	.9821	.9982	.9995		.9998	.9988
6	M3a	.9996	.9999		.+	.9997	.+	.+		.+	.+
	M4	.9996	.9999		.+	.9997	.+	.+		.+	.+
	M5	.9996	.9999		.+	.9997	.+	.+		.+	.+
	M6a	.9996	.9999		.+	.9997	.+	.+		.+	.+
8	M3a	.+	.+		.+	.+	.+	.+		.+	.+
	M4	.+	.+		.+	.+	.+	.+		.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+
10	M3a	.+	.+		.+	.+	.+	.+		.+	.+
	M4	.+	.+		.+	.+	.+	.+		.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+

+ = 1.0000

Table 4.3.6-15

Terpstra-Jonckheere Test (Five Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 1.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.9446	.9374	.9667	.9468	.9461	.9952	.9947	.+	.9984	.9964
	M4	.9446	.9450	.9677	.9546	.9570	.9952	.9953	.9996	.9987	.9972
	M5	.9446	.9444	.9555	.9530	.9529	.9952	.9954	.9993	.9986	.9969
	M6a	.9446	.9374	.9667	.9468	.9461	.9952	.9947	.+	.9984	.9964
4	M3a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M4	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
6	M3a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M4	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
8	M3a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M4	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
10	M3a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M4	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+

+ = 1.0000

Table 4.3.6-16

Terpstra-Jonckheere Test (Six Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.0867	.1289		.0648	.0777	.2428	.3159		.1967	.2285
	M4	.0867	.1483		.0740	.0921	.2428	.3386		.2118	.2473
	M5	.0867	.1408		.0711	.0868	.2428	.3308		.2082	.2431
	M6a	.0867	.1289		.0648	.0777	.2428	.3159		.1967	.2285
4	M3a	.1964	.3084		.1490	.1901	.4487	.5809		.3765	.4433
	M4	.1964	.3203		.1564	.2006	.4487	.5889		.3853	.4519
	M5	.1964	.3141		.1529	.1955	.4487	.5855		.3819	.4492
	M6a	.1964	.3084		.1490	.1901	.4487	.5809		.3765	.4433
6	M3a	.3378	.5051		.2611	.3343	.6171	.7606		.5290	.6147
	M4	.3378	.5074		.2679	.3423	.6171	.7597		.5349	.6190
	M5	.3378	.5078		.2639	.3383	.6171	.7617		.5319	.6170
	M6a	.3378	.5051		.2611	.3343	.6171	.7606		.5290	.6147
8	M3a	.4591	.6513		.3616	.4578	.7278	.8594		.6378	.7284
	M4	.4591	.6454		.3679	.4637	.7278	.8540		.6426	.7303
	M5	.4591	.6521		.3635	.4598	.7278	.8594		.6395	.7290
	M6a	.4591	.6513		.3616	.4578	.7278	.8594		.6378	.7284
10	M3a	.5707	.7632		.4580	.5697	.8163	.9220		.7307	.8167
	M4	.5707	.7518		.4637	.5737	.8163	.9123		.7346	.8177
	M5	.5707	.7632		.4594	.5707	.8163	.9216		.7316	.8167
	M6a	.5707	.7632		.4580	.5697	.8163	.9220		.7307	.8167

Table 4.3.6-17

Terpstra-Jonckheere Test (Six Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.5

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.5243	.6581	.6120	.4694	.3223	.7958	.8672	.9082	.7716	.6140
	M4	.5243	.6821	.6534	.4945	.3556	.7958	.8771	.9132	.7860	.6360
	M5	.5243	.6734	.6362	.4890	.3417	.7958	.8744	.9065	.7846	.6293
	M6a	.5243	.6581	.6120	.4694	.3223	.7958	.8672	.9082	.7716	.6140
4	M3a	.9162	.9649	.9914	.9115	.7393	.9855	.9948	.9999	.9871	.9236
	M4	.9162	.9653	.9910	.9142	.7498	.9855	.9948	.9998	.9873	.9258
	M5	.9162	.9655	.9882	.9136	.7432	.9855	.9948	.9996	.9874	.9244
	M6a	.9162	.9649	.9914	.9115	.7393	.9855	.9948	.9999	.9871	.9236
6	M3a	.9922	.9981	.+	.9930	.9320	.9993	.9999	.+	.9995	.9878
	M4	.9922	.9979	.+	.9931	.9341	.9993	.9999	.+	.9995	.9880
	M5	.9922	.9981	.+	.9931	.9323	.9993	.9999	.+	.9995	.9878
	M6a	.9922	.9981	.+	.9930	.9320	.9993	.9999	.+	.9995	.9878
8	M3a	.9993	.9999	.+	.9995	.9841	.+	.+	.+	.+	.9981
	M4	.9993	.9999	.+	.9995	.9841	.+	.+	.+	.+	.9981
	M5	.9993	.9999	.+	.9995	.9841	.+	.+	.+	.+	.9980
	M6a	.9993	.9999	.+	.9995	.9841	.+	.+	.+	.+	.9981
10	M3a	.+	.+	.+	.+	.9967	.+	.+	.+	.+	.9997
	M4	.+	.+	.+	.+	.9968	.+	.+	.+	.+	.9997
	M5	.+	.+	.+	.+	.9967	.+	.+	.+	.+	.9997
	M6a	.+	.+	.+	.+	.9967	.+	.+	.+	.+	.9997

+ = 1.0000

Table 4.3.6-18

Terpstra-Jonckheere Test (Six Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 0.8

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.9181	.9519		.9517	.9207	.9885	.9949		.9979	.9909
	M4	.9181	.9578		.9594	.9335	.9885	.9954		.9984	.9923
	M5	.9181	.9562		.9571	.9267	.9885	.9953		.9982	.9913
	M6a	.9181	.9519		.9517	.9207	.9885	.9949		.9979	.9909
4	M3a	.9998	.+		.+	.9999	.+	.+		.+	.+
	M4	.9998	.+		.+	.9999	.+	.+		.+	.+
	M5	.9998	.+		.+	.9999	.+	.+		.+	.+
	M6a	.9998	.+		.+	.9999	.+	.+		.+	.+
6	M3a	.+	.+		.+	.+	.+	.+		.+	.+
	M4	.+	.+		.+	.+	.+	.+		.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+
8	M3a	.+	.+		.+	.+	.+	.+		.+	.+
	M4	.+	.+		.+	.+	.+	.+		.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+
10	M3a	.+	.+		.+	.+	.+	.+		.+	.+
	M4	.+	.+		.+	.+	.+	.+		.+	.+
	M5	.+	.+		.+	.+	.+	.+		.+	.+
	M6a	.+	.+		.+	.+	.+	.+		.+	.+

+ = 1.0000

Table 4.3.6-19
Terpstra-Jonckheere Test (Six Groups), Power for α .01 and .05
Nominal Effect Size Multiplier 1.2

Alpha		.01					.05				
Distribution		Norm	EA	EB	ML	SS	Norm	EA	EB	ML	SS
ISS	*Mthd*										
2	M3a	.9985	.9986	.+	.9997	.9992	.+	.+	.+	.+	.+
	M4	.9985	.9989	.+	.9998	.9995	.+	.+	.+	.+	.+
	M5	.9985	.9988	.9999	.9998	.9992	.+	.+	.+	.+	.+
	M6a	.9985	.9986	.+	.9997	.9992	.+	.+	.+	.+	.+
4	M3a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M4	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
6	M3a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M4	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+					
8	M3a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M4	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
10	M3a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M4	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M5	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+
	M6a	.+	.+	.+	.+	.+	.+	.+	.+	.+	.+

+ = 1.0000

There is no table for three groups at nominal alpha .01 as there was no data for this combination of test conditions.

Table 4.3.6-20
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Four Groups, α .01, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
2	M3a	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	
	M4	1.3	1.5	1.0	1.0	1.4	1.5	1.0	1.1	1.5	1.8	1.5	
	M5	1.8	1.5	2.0	2.0	1.6	1.5	2.0	1.9	1.5	1.3	1.5	
	M6a	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	
4	M3a	2.8	2.0	3.5	3.0	2.8	2.0	3.5	3.1	2.5	2.3	3.0	
	M4	3.5	3.5	2.0	3.0	3.5	3.8	1.9	2.4	3.4	3.3	2.0	
	M5	1.0	2.5	1.0	1.0	1.0	2.3	1.1	1.4	1.6	2.3	2.0	
	M6a	2.8	2.0	3.5	3.0	2.8	2.0	3.5	3.1	2.5	2.3	3.0	
6	M3a	2.4	2.5	2.8	2.5	2.4	2.5	2.6	2.5	2.5	2.5	2.5	
	M4	4.0	3.3	3.1	3.6	3.6	3.3	3.0	3.6	3.0	3.3	2.5	
	M5	1.3	1.8	1.4	1.4	1.6	1.8	1.3	1.4	2.0	1.8	2.5	
	M6a	2.4	2.5	2.8	2.5	2.4	2.5	3.1	2.5	2.5	2.5	2.5	
8	M3a	2.5	2.0	2.5	2.3	2.3	2.0	2.4	2.4	2.3	2.0	2.5	
	M4	3.6	3.3	3.6	3.6	3.6	3.3	3.6	3.6	3.3	3.0	2.5	
	M5	1.4	2.8	1.4	1.9	1.9	2.8	1.6	1.6	2.3	3.0	2.5	
	M6a	2.5	2.0	2.5	2.3	2.3	2.0	2.4	2.4	2.3	2.0	2.5	
10	M3a	2.4	2.0	2.9	2.4	2.4	2.0	2.9	2.4	2.4	2.5	2.4	
	M4	3.6	3.3	3.4	3.6	3.3	3.3	3.4	3.6	3.3	2.5	2.9	
	M5	1.6	2.8	1.5	1.6	2.0	2.8	1.5	1.6	2.0	2.5	2.4	
	M6a	2.4	2.0	2.3	2.4	2.4	2.0	2.3	2.4	2.4	2.5	2.4	

Table 4.3.6-21
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Five Groups, α .01, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
2	M3a	3.5	3.0	3.5	3.3	3.5	3.0	3.5	3.5	3.4	2.8	3.3	3.4
	M4	1.0	1.0	1.3	1.0	1.0	1.0	1.0	1.0	1.3	1.8	1.6	1.3
	M5	2.0	3.0	1.8	2.5	2.0	3.0	2.0	2.0	2.0	2.8	1.9	2.0
	M6a	3.5	3.0	3.5	3.3	3.5	3.0	3.5	3.5	3.4	2.8	3.3	3.4
4	M3a	2.8	2.5	3.3	3.3	2.8	2.5	3.0	3.0	2.8	2.3	2.8	2.8
	M4	3.1	3.3	1.9	2.1	2.9	3.3	2.1	2.3	2.8	3.3	2.3	2.3
	M5	1.4	1.8	1.6	1.4	1.6	1.8	1.9	1.8	1.8	2.3	2.3	2.3
	M6a	2.8	2.5	3.3	3.3	2.8	2.5	3.0	3.0	2.8	2.3	2.8	2.8
6	M3a	2.5	2.0	2.8	2.5	2.5	2.0	2.9	2.5	2.6	2.5	2.6	2.6
	M4	3.3	2.8	2.8	3.3	3.3	2.8	2.5	3.3	2.6	2.5	2.6	2.6
	M5	1.8	3.3	1.8	1.8	1.8	3.3	1.8	1.8	2.1	2.5	2.1	2.1
	M6a	2.5	2.0	2.8	2.5	2.5	2.0	2.9	2.5	2.6	2.5	2.6	2.6
8	M3a	2.4	2.0	2.5	2.5	2.4	2.3	2.5	2.4	2.5	2.5	2.5	2.4
	M4	3.3	2.8	3.3	3.3	3.3	2.3	3.0	3.3	2.9	2.5	2.5	2.9
	M5	2.0	3.3	1.8	1.8	2.0	3.3	2.0	2.0	2.1	2.5	2.5	2.4
	M6a	2.4	2.0	2.5	2.5	2.4	2.3	2.5	2.4	2.5	2.5	2.5	2.4
10	M3a	2.3	2.0	2.5	2.3	2.5	2.5	2.5	2.3	2.5	2.5	2.6	2.3
	M4	3.3	2.8	3.3	3.3	2.9	2.5	2.9	3.3	2.9	2.5	2.6	3.3
	M5	2.3	3.3	1.8	2.3	2.1	2.5	2.1	2.3	2.1	2.5	2.1	2.3
	M6a	2.3	2.0	2.5	2.3	2.5	2.5	2.5	2.3	2.5	2.5	2.6	2.3

Table 4.3.6-22
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Six Groups, α .01, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
2	M3a	3.5	2.8	3.5	3.4	3.3	3.0	3.3	3.4	3.3	3.0	3.3	3.3
	M4	1.0	1.5	1.1	1.0	1.4	1.8	1.4	1.0	1.5	1.8	1.8	1.6
	M5	2.0	3.0	1.9	2.3	2.1	2.3	2.1	2.3	2.0	2.3	1.8	1.9
	M6a	3.5	2.8	3.5	3.4	3.3	3.0	3.3	3.4	3.3	3.0	3.3	3.3
4	M3a	3.0	2.0	3.0	3.0	2.9	2.3	3.0	3.0	2.9	2.5	2.6	2.9
	M4	2.0	2.8	1.8	1.8	2.3	2.3	1.9	1.8	1.9	2.5	2.1	1.9
	M5	2.0	3.3	2.3	2.3	2.0	3.3	2.1	2.3	2.4	2.5	2.6	2.4
	M6a	3.0	2.0	3.0	3.0	2.9	2.3	3.0	3.0	2.9	2.5	2.6	2.9
6	M3a	2.6	2.5	2.5	3.0	2.8	2.5	2.8	2.9	2.5	2.5	2.6	2.8
	M4	2.8	2.5	2.4	1.8	2.4	2.5	2.1	1.8	2.5	2.5	2.1	2.3
	M5	2.0	2.5	2.6	2.3	2.1	2.5	2.4	2.5	2.5	2.5	2.6	2.3
	M6a	2.6	2.5	2.5	3.0	2.8	2.5	2.8	2.9	2.5	2.5	2.6	2.8
8	M3a	2.5	2.5	2.8	2.8	2.5	2.5	2.8	2.8	2.5	2.5	2.6	2.5
	M4	2.9	2.5	2.1	2.1	2.9	2.5	2.1	2.1	2.5	2.5	2.1	2.5
	M5	2.1	2.5	2.4	2.4	2.1	2.5	2.4	2.4	2.5	2.5	2.6	2.5
	M6a	2.5	2.5	2.8	2.8	2.5	2.5	2.8	2.8	2.5	2.5	2.6	2.5
10	M3a	2.4	2.5	2.8	2.6	2.4	2.5	2.8	2.8	2.4	2.5	2.5	2.5
	M4	2.9	2.5	2.1	2.5	2.9	2.5	2.1	2.1	2.9	2.5	2.5	2.5
	M5	2.4	2.5	2.4	2.3	2.4	2.5	2.4	2.4	2.4	2.5	2.5	2.5
	M6a	2.4	2.5	2.8	2.6	2.4	2.5	2.8	2.8	2.4	2.5	2.5	2.5

Table 4.3.6-23
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Three Groups, α .05, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
2	M3a	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5
	M4	1.5	1.5	1.9	1.4	1.4	1.5	1.5	1.5	1.5	1.5	1.5	1.5
	M5	1.5	1.5	1.1	1.6	1.6	1.5	1.5	1.5	1.5	1.5	1.5	1.5
	M6a	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5
4	M3a	2.8	3.0	3.5	2.9	2.8	3.0	3.5	2.9	2.9	2.5	2.9	3.0
	M4	3.5	3.0	2.0	3.3	3.5	3.0	2.0	3.3	2.9	3.5	2.5	3.0
	M5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.4	1.5	1.8	1.0
	M6a	2.8	3.0	3.5	2.9	2.8	3.0	3.5	2.9	2.9	2.5	2.9	3.0
6	M3a	2.5	2.0	2.8	2.5	2.4	2.0	2.8	2.4	2.6	2.3	2.8	2.1
	M4	4.0	4.0	3.5	4.0	4.0	4.0	3.5	4.0	3.4	3.3	2.8	4.0
	M5	1.0	2.0	1.0	1.0	1.3	2.0	1.0	1.3	1.4	2.3	1.8	1.8
	M6a	2.5	2.0	2.8	2.5	2.4	2.0	2.8	2.4	2.6	2.3	2.8	2.1
8	M3a	2.6	2.0	2.5	2.4	2.5	2.0	2.5	2.5	2.1	2.5	2.8	2.1
	M4	3.5	3.5	4.0	4.0	3.6	3.8	3.6	3.6	3.6	3.3	2.8	3.6
	M5	1.3	2.5	1.0	1.3	1.4	2.3	1.4	1.4	2.1	1.8	1.8	2.1
	M6a	2.6	2.0	2.5	2.4	2.5	2.0	2.5	2.5	2.1	2.5	2.8	2.1
10	M3a	2.4	1.5	2.8	2.4	2.5	2.0	2.8	2.4	2.3	2.3	2.5	2.4
	M4	4.0	3.5	3.1	4.0	3.6	3.3	2.9	3.6	3.3	3.3	2.5	3.3
	M5	1.3	3.5	1.4	1.3	1.4	2.8	1.6	1.6	2.3	2.3	2.5	2.0
	M6a	2.4	1.5	2.8	2.4	2.5	2.0	2.8	2.4	2.3	2.3	2.5	2.4

Table 4.3.6-24
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Four Groups, α .05, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
2	M3a	3.5	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.4	2.8	3.5	3.5
	M4	2.0	3.0	2.0	2.0	2.0	3.0	2.0	2.0	1.9	3.0	1.5	1.8
	M5	1.0	2.0	1.0	1.0	1.0	2.0	1.0	1.0	1.4	1.5	1.5	1.3
	M6a	3.5	2.5	3.5	3.5	3.5	2.5	3.5	3.5	3.4	2.8	3.5	3.5
4	M3a	2.5	2.0	3.0	2.4	2.4	2.5	3.1	2.5	2.5	2.5	2.6	2.8
	M4	4.0	3.5	2.6	4.0	4.0	3.3	2.4	3.6	3.0	3.3	2.6	3.1
	M5	1.0	2.5	1.4	1.3	1.3	1.8	1.4	1.4	2.0	1.8	2.1	1.4
	M6a	2.5	2.0	3.0	2.4	2.4	2.5	3.1	2.5	2.5	2.5	2.6	2.8
6	M3a	2.4	2.0	2.5	2.3	2.4	2.0	2.5	2.4	2.4	2.3	2.6	2.5
	M4	3.6	3.3	3.6	3.6	3.6	3.3	3.3	3.6	3.3	3.3	2.6	3.3
	M5	1.6	2.8	1.4	1.9	1.6	2.8	1.8	1.6	2.0	2.3	2.1	1.8
	M6a	2.4	2.0	2.5	2.3	2.4	2.0	2.5	2.4	2.4	2.3	2.6	2.5
8	M3a	2.4	2.0	2.5	2.3	2.4	2.0	2.5	2.5	2.4	2.0	2.6	2.3
	M4	3.3	3.3	3.3	3.6	3.3	3.3	3.3	3.3	2.9	3.0	2.6	3.3
	M5	2.0	2.8	1.8	1.9	2.0	2.8	1.8	1.8	2.4	3.0	2.1	2.3
	M6a	2.4	2.0	2.5	2.3	2.4	2.0	2.5	2.5	2.4	2.0	2.6	2.3
10	M3a	2.6	2.0	2.8	2.8	2.4	2.0	2.4	2.5	2.4	2.5	2.4	2.3
	M4	3.3	3.3	3.0	3.0	3.3	3.0	3.3	3.3	2.9	2.5	2.9	3.3
	M5	1.8	2.8	1.5	1.5	2.0	3.0	2.0	1.8	2.4	2.5	2.4	2.3
	M6a	2.6	2.0	2.8	2.8	2.4	2.0	2.4	2.5	2.4	2.5	2.4	2.3

Table 4.3.6-25
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Five Groups, α .05, by
Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
2	M3a	3.5	2.5	3.5	3.5	3.3	2.8	3.5	3.5	3.5	3.0	3.3	3.3
	M4	2.0	2.5	1.5	1.5	2.1	2.0	1.5	1.8	1.5	2.3	1.8	1.8
	M5	1.0	2.5	1.5	1.5	1.4	2.5	1.5	1.3	1.5	1.8	1.8	1.8
	M6a	3.5	2.5	3.5	3.5	3.3	2.8	3.5	3.5	3.5	3.0	3.3	3.3
4	M3a	2.4	2.0	3.1	2.3	2.4	2.0	3.0	2.6	2.4	2.5	2.5	2.3
	M4	3.6	3.3	2.4	3.5	3.3	3.0	2.3	3.0	2.4	2.5	2.5	2.8
	M5	1.6	2.8	1.4	2.0	2.0	3.0	1.8	1.8	2.9	2.5	2.5	2.8
	M6a	2.4	2.0	3.1	2.3	2.4	2.0	3.0	2.6	2.4	2.5	2.5	2.3
6	M3a	2.3	2.0	2.8	2.5	2.4	2.5	2.5	2.5	2.4	2.5	2.6	2.3
	M4	3.3	3.3	2.8	3.3	3.3	2.5	3.0	3.3	2.9	2.5	2.6	3.3
	M5	2.3	2.8	1.8	1.8	2.0	2.5	2.0	1.8	2.4	2.5	2.1	2.3
	M6a	2.3	2.0	2.8	2.5	2.4	2.5	2.5	2.5	2.4	2.5	2.6	2.3
8	M3a	2.4	2.0	2.4	2.4	2.5	2.5	2.5	2.4	2.5	2.5	2.5	2.4
	M4	3.3	2.8	3.3	3.3	2.9	2.5	2.9	3.3	2.9	2.5	2.5	2.9
	M5	2.0	3.3	2.0	2.0	2.1	2.5	2.1	2.0	2.1	2.5	2.5	2.4
	M6a	2.4	2.0	2.4	2.4	2.5	2.5	2.5	2.4	2.5	2.5	2.5	2.4
10	M3a	2.1	2.5	2.5	2.3	2.4	2.5	2.5	2.4	2.4	2.5	2.5	2.3
	M4	3.1	2.5	2.9	3.3	2.9	2.5	2.9	3.3	2.9	2.5	2.5	3.3
	M5	2.6	2.5	2.1	2.3	2.4	2.5	2.1	2.0	2.4	2.5	2.5	2.3
	M6a	2.1	2.5	2.5	2.3	2.4	2.5	2.5	2.4	2.4	2.5	2.5	2.3

Table 4.3.6-26

Terpstra-Jonckheere Test

Mean Ranks of Power Results for Six Groups, α .05, by

Initial Sample Size, Method and Distribution across Effect Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	*Mthd*												
2	M3a	3.3	2.5	3.3	3.3	3.0	2.5	3.0	3.1	3.1	2.5	3.0	3.0
	M4	1.4	1.8	1.4	1.4	1.8	1.8	1.8	1.4	1.5	2.5	1.9	1.8
	M5	2.1	3.3	2.1	2.1	2.3	3.3	2.3	2.4	2.3	2.5	2.1	2.3
	M6a	3.3	2.5	3.3	3.3	3.0	2.5	3.0	3.1	3.1	2.5	3.0	3.0
4	M3a	2.8	2.0	3.0	3.0	2.8	2.5	2.8	2.9	2.8	2.5	2.6	2.9
	M4	2.1	2.8	2.0	1.8	2.1	2.5	2.1	1.8	2.3	2.5	2.1	1.9
	M5	2.4	3.3	2.0	2.3	2.4	2.5	2.4	2.5	2.3	2.5	2.6	2.4
	M6a	2.8	2.0	3.0	3.0	2.8	2.5	2.8	2.9	2.8	2.5	2.6	2.9
6	M3a	2.5	2.5	2.8	2.9	2.5	2.5	2.8	2.8	2.5	2.5	2.6	2.8
	M4	2.9	2.5	2.1	1.8	2.9	2.5	2.1	2.1	2.5	2.5	2.1	2.3
	M5	2.1	2.5	2.4	2.5	2.1	2.5	2.4	2.4	2.5	2.5	2.6	2.3
	M6a	2.5	2.5	2.8	2.9	2.5	2.5	2.8	2.8	2.5	2.5	2.6	2.8
8	M3a	2.4	2.5	2.8	2.6	2.4	2.5	2.8	2.8	2.4	2.5	2.5	2.5
	M4	2.9	2.5	2.1	2.0	2.9	2.5	2.1	2.1	2.9	2.5	2.5	2.5
	M5	2.4	2.5	2.4	2.8	2.4	2.5	2.4	2.4	2.4	2.5	2.5	2.5
	M6a	2.4	2.5	2.8	2.6	2.4	2.5	2.8	2.8	2.4	2.5	2.5	2.5
10	M3a	2.3	2.5	2.8	2.6	2.4	2.5	2.8	2.6	2.4	2.5	2.5	2.5
	M4	2.9	2.5	2.1	2.1	2.9	2.5	2.1	2.1	2.9	2.5	2.5	2.5
	M5	2.6	2.5	2.4	2.6	2.4	2.5	2.4	2.6	2.4	2.5	2.5	2.5
	M6a	2.3	2.5	2.8	2.6	2.4	2.5	2.8	2.6	2.4	2.5	2.5	2.5

There is no table for three groups at nominal alpha .01 as there was no data for this combination of test conditions.

Table 4.3.6-27
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Four Groups, α .01, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.9		3.0	2.9	2.9		3.0	2.9	2.9		2.6	2.6
	M4	3.1		2.8	3.0	3.1		2.8	3.1	2.8		2.6	3.0
	M5	1.1		1.2	1.2	1.1		1.2	1.1	1.4		2.2	1.8
	M6a	2.9		3.0	2.9	2.9		3.0	2.9	2.9		2.6	2.6
0.5	M3a	2.7	2.3	3.0	2.9	2.5	2.3	3.0	2.9	2.4	2.6	2.9	2.7
	M4	3.5	3.4	2.8	3.0	3.5	3.4	2.8	3.0	3.5	2.9	2.1	3.1
	M5	1.1	2.0	1.2	1.2	1.5	2.0	1.2	1.2	1.7	1.9	2.1	1.5
	M6a	2.7	2.3	3.0	2.9	2.5	2.3	3.0	2.9	2.4	2.6	2.9	2.7
0.8	M3a	2.6		3.2	2.4	2.5		3.1	2.6	2.6		2.9	2.7
	M4	3.4		2.8	3.4	3.1		2.8	3.2	2.6		2.1	2.6
	M5	1.4		1.3	1.8	1.9		1.5	1.6	2.2		2.1	2.0
	M6a	2.6		2.7	2.4	2.5		2.6	2.6	2.6		2.9	2.7
1.2	M3a	2.6	2.5	2.9	2.7	2.7	2.5	2.8	2.7	2.6	2.5	2.7	2.7
	M4	2.8	2.5	2.1	2.5	2.6	2.6	1.9	2.2	2.6	2.6	2.3	2.2
	M5	2.0	2.5	2.1	2.1	2.0	2.4	2.1	2.4	2.2	2.4	2.3	2.4
	M6a	2.6	2.5	2.9	2.7	2.7	2.5	3.2	2.7	2.6	2.5	2.7	2.7

Table 4.3.6-28

Terpstra-Jonckheere Test

Mean Ranks of Power Results for Five Groups, α .01, by Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.9		3.1	2.9	2.9		3.2	2.9	2.9		2.8	2.5
	M4	3.0		2.4	3.0	3.0		2.3	3.0	3.0		2.4	2.9
	M5	1.2		1.4	1.2	1.2		1.3	1.2	1.2		2.0	2.1
	M6a	2.9		3.1	2.9	2.9		3.2	2.9	2.9		2.8	2.5
0.5	M3a	2.4	2.1	2.9	2.7	2.7	2.4	2.9	2.6	2.8	2.6	3.0	2.9
	M4	3.4	2.8	3.0	3.0	2.9	2.5	2.5	3.0	2.4	2.6	2.2	2.5
	M5	1.8	3.0	1.2	1.6	1.7	2.7	1.7	1.8	2.0	2.2	1.8	1.7
	M6a	2.4	2.1	2.9	2.7	2.7	2.4	2.9	2.6	2.8	2.6	3.0	2.9
0.8	M3a	2.7		2.9	2.7	2.6		2.7	2.7	2.7		2.7	2.7
	M4	2.5		2.1	2.1	2.5		2.2	2.2	2.3		2.2	2.2
	M5	2.1		2.1	2.5	2.3		2.4	2.4	2.3		2.4	2.4
	M6a	2.7		2.9	2.7	2.6		2.7	2.7	2.7		2.7	2.7
1.2	M3a	2.7	2.5	2.7	2.7	2.7	2.5	2.7	2.7	2.6	2.4	2.5	2.6
	M4	2.2	2.2	2.4	2.2	2.2	2.2	2.2	2.2	2.2	2.4	2.5	2.2
	M5	2.4	2.8	2.2	2.4	2.4	2.8	2.4	2.4	2.6	2.8	2.5	2.6
	M6a	2.7	2.5	2.7	2.7	2.7	2.5	2.7	2.7	2.6	2.4	2.5	2.6

Table 4.3.6-29
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Six Groups, α .01, by
Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	3.0		3.5	3.5	3.0		3.5	3.5	2.7		3.0	3.1
	M4	2.4		1.0	1.0	2.4		1.0	1.0	2.2		1.4	1.9
	M5	1.8		2.0	2.0	1.6		2.0	2.0	2.4		2.6	1.9
	M6a	3.0		3.5	3.5	3.0		3.5	3.5	2.7		3.0	3.1
0.5	M3a	2.8	2.5	2.7	3.0	2.8	2.6	2.9	3.0	2.9	2.7	2.7	2.8
	M4	2.4	2.3	2.1	1.9	2.3	2.1	2.0	1.6	2.0	2.2	2.3	1.9
	M5	2.0	2.7	2.5	2.1	2.1	2.7	2.2	2.4	2.2	2.4	2.3	2.5
	M6a	2.8	2.5	2.7	3.0	2.8	2.6	2.9	3.0	2.9	2.7	2.7	2.8
0.8	M3a	2.7		2.7	2.7	2.7		2.7	2.7	2.7		2.7	2.7
	M4	2.2		2.2	2.2	2.2		2.2	2.2	2.3		2.3	2.3
	M5	2.4		2.4	2.4	2.4		2.4	2.4	2.3		2.3	2.3
	M6a	2.7		2.7	2.7	2.7		2.7	2.7	2.7		2.7	2.7
1.2	M3a	2.7	2.4	2.7	2.6	2.5	2.5	2.5	2.6	2.5	2.5	2.5	2.5
	M4	2.2	2.4	2.3	2.2	2.5	2.5	2.5	2.2	2.5	2.5	2.5	2.5
	M5	2.4	2.8	2.3	2.6	2.5	2.5	2.5	2.6	2.5	2.5	2.5	2.5
	M6a	2.7	2.4	2.7	2.6	2.5	2.5	2.5	2.6	2.5	2.5	2.5	2.5

Table 4.3.6-30
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Three Groups, α .05, by Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.9		2.9	2.9	2.9		2.9	2.9	2.6		2.8	2.5
	M4	3.0		3.2	3.2	3.0		3.1	3.1	3.3		2.4	3.3
	M5	1.2		1.0	1.0	1.2		1.1	1.1	1.5		2.0	1.7
	M6a	2.9		2.9	2.9	2.9		2.9	2.9	2.6		2.8	2.5
0.5	M3a	2.9	2.7	3.1	2.8	2.7	2.7	3.0	2.8	2.6	2.6	3.0	2.6
	M4	3.0	3.0	2.7	3.3	3.4	3.0	2.9	3.3	3.3	3.3	2.7	3.3
	M5	1.2	1.6	1.1	1.1	1.2	1.6	1.1	1.1	1.5	1.5	1.3	1.5
	M6a	2.9	2.7	3.1	2.8	2.7	2.7	3.0	2.8	2.6	2.6	3.0	2.6
0.8	M3a	2.7		3.1	2.7	2.7		3.1	2.6	2.7		3.0	2.7
	M4	3.6		2.8	3.4	3.6		2.5	3.5	2.7		2.2	3.0
	M5	1.0		1.0	1.2	1.0		1.3	1.3	1.9		1.8	1.6
	M6a	2.7		3.1	2.7	2.7		3.1	2.6	2.7		3.0	2.7
1.2	M3a	2.5	2.1	2.9	2.5	2.6	2.3	3.0	2.6	2.8	2.6	2.7	2.7
	M4	3.6	3.2	2.9	3.4	2.9	3.2	2.3	2.9	2.4	2.6	2.3	2.7
	M5	1.4	2.6	1.3	1.6	1.9	2.2	1.7	1.9	2.0	2.2	2.3	1.9
	M6a	2.5	2.1	2.9	2.5	2.6	2.3	3.0	2.6	2.8	2.6	2.7	2.7

Table 4.3.6-31
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Four Groups, α .05, by Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.7		2.9	2.7	2.7		2.9	2.7	2.5		2.7	2.6
	M4	3.6		3.2	3.6	3.6		3.2	3.6	3.3		2.7	3.3
	M5	1.0		1.0	1.0	1.0		1.0	1.0	1.7		1.9	1.5
	M6a	2.7		2.9	2.7	2.7		2.9	2.7	2.5		2.7	2.6
0.5	M3a	2.5	2.1	2.9	2.7	2.5	2.1	2.8	2.7	2.6	2.4	2.9	2.5
	M4	3.6	3.6	3.2	3.6	3.6	3.5	3.2	3.6	2.9	3.2	2.5	3.6
	M5	1.4	2.2	1.0	1.0	1.4	2.3	1.2	1.0	1.9	2.0	1.7	1.4
	M6a	2.5	2.1	2.9	2.7	2.5	2.1	2.8	2.7	2.6	2.4	2.9	2.5
0.8	M3a	2.8		2.9	2.5	2.6		2.8	2.6	2.7		2.7	2.8
	M4	2.8		2.8	3.1	3.0		2.5	3.0	2.3		2.3	2.5
	M5	1.6		1.4	1.9	1.8		1.9	1.8	2.3		2.3	1.9
	M6a	2.8		2.9	2.5	2.6		2.8	2.6	2.7		2.7	2.8
1.2	M3a	2.7	2.1	2.7	2.6	2.6	2.3	2.7	2.7	2.6	2.4	2.7	2.7
	M4	2.7	2.9	2.4	2.7	2.7	2.8	2.4	2.4	2.6	2.8	2.3	2.3
	M5	1.9	2.9	2.2	2.1	2.1	2.6	2.2	2.2	2.2	2.4	2.3	2.3
	M6a	2.7	2.1	2.7	2.6	2.6	2.3	2.7	2.7	2.6	2.4	2.7	2.7

Table 4.3.6-32

Terpstra-Jonckheere Test

Mean Ranks of Power Results for Five Groups, α .05, by

Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.6		3.1	2.7	2.6		3.0	2.8	2.4		2.8	2.3
	M4	3.6		2.6	3.6	3.6		2.8	3.4	3.1		2.4	3.1
	M5	1.2		1.2	1.0	1.2		1.2	1.0	2.1		2.0	2.3
	M6a	2.6		3.1	2.7	2.6		3.0	2.8	2.4		2.8	2.3
0.5	M3a	2.2	2.1	2.8	2.4	2.5	2.5	2.8	2.5	2.7	2.7	2.7	2.4
	M4	3.5	3.1	2.9	3.6	3.0	2.6	2.6	3.6	2.3	2.4	2.3	3.2
	M5	2.1	2.7	1.5	1.6	2.0	2.4	1.8	1.4	2.3	2.2	2.3	2.0
	M6a	2.2	2.1	2.8	2.4	2.5	2.5	2.8	2.5	2.7	2.7	2.7	2.4
0.8	M3a	2.6		2.8	2.5	2.7		2.7	2.7	2.7		2.7	2.7
	M4	2.7		2.5	2.4	2.4		2.3	2.3	2.3		2.3	2.3
	M5	2.1		1.9	2.6	2.2		2.3	2.3	2.3		2.3	2.3
	M6a	2.6		2.8	2.5	2.7		2.7	2.7	2.7		2.7	2.7
1.2	M3a	2.7	2.3	2.7	2.7	2.5	2.4	2.7	2.7	2.7	2.5	2.5	2.5
	M4	2.4	2.6	2.2	2.2	2.5	2.4	2.3	2.3	2.3	2.5	2.5	2.5
	M5	2.2	2.8	2.4	2.4	2.5	2.8	2.3	2.3	2.3	2.5	2.5	2.5
	M6a	2.7	2.3	2.7	2.7	2.5	2.4	2.7	2.7	2.7	2.5	2.5	2.5

Table 4.3.6-33

Terpstra-Jonckheere Test

Mean Ranks of Power Results for Six Groups, α .05, by

Effect Size, Method and Distribution across Initial Sample Size

Decimal		4th				3rd				2nd			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
NESM	*Mthd*												
0.2	M3a	2.6		3.5	3.4	2.7		3.5	3.4	2.7		2.9	3.1
	M4	2.8		1.0	1.0	2.8		1.0	1.0	2.6		1.7	1.8
	M5	2.0		2.0	2.2	1.8		2.0	2.2	2.0		2.5	2.0
	M6a	2.6		3.5	3.4	2.7		3.5	3.4	2.7		2.9	3.1
0.5	M3a	2.7	2.3	2.9	2.9	2.7	2.5	2.7	2.8	2.6	2.5	2.7	2.8
	M4	2.2	2.3	2.1	1.5	2.2	2.2	2.2	1.9	2.2	2.5	2.2	1.9
	M5	2.4	3.1	2.1	2.7	2.4	2.8	2.4	2.5	2.6	2.5	2.4	2.5
	M6a	2.7	2.3	2.9	2.9	2.7	2.5	2.7	2.8	2.6	2.5	2.7	2.8
0.8	M3a	2.7		2.7	2.7	2.5		2.5	2.6	2.7		2.5	2.5
	M4	2.2		2.2	2.2	2.5		2.5	2.2	2.3		2.5	2.5
	M5	2.4		2.4	2.4	2.5		2.5	2.6	2.3		2.5	2.5
	M6a	2.7		2.7	2.7	2.5		2.5	2.6	2.7		2.5	2.5
1.2	M3a	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
	M4	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
	M5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
	M6a	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5

Table 4.3.6-34

Terpstra-Jonckheere Test

Mean Ranks of Power Results for Three Groups, α .01, by

Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a												
M4			no									
M5		data										
M6a												

Table 4.3.6-35

Terpstra-Jonckheere Test

Mean Ranks of Power Results for Four Groups, α .01, by

Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.7	2.4	3.0	2.7	2.7	2.4	3.0	2.8	2.6	2.6	2.8	2.7
M4	3.2	3.0	2.6	3.0	3.1	3.0	2.6	2.9	2.9	2.8	2.3	2.7
M5	1.4	2.3	1.5	1.6	1.6	2.2	1.5	1.6	1.9	2.2	2.2	1.9
M6a	2.7	2.4	2.9	2.7	2.7	2.4	3.0	2.8	2.6	2.6	2.8	2.7

Table 4.3.6-36
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Five Groups, α .01, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.7	2.3	2.9	2.8	2.7	2.5	2.9	2.7	2.8	2.5	2.8	2.7
M4	2.8	2.5	2.5	2.6	2.7	2.4	2.3	2.6	2.5	2.5	2.3	2.5
M5	1.9	2.9	1.7	1.9	1.9	2.8	2.0	2.0	2.0	2.5	2.2	2.2
M6a	2.7	2.3	2.9	2.8	2.7	2.5	2.9	2.7	2.8	2.5	2.8	2.7

Table 4.3.6-37
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Six Groups, α .01, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
Mthd												
M3a	2.8	2.5	2.9	3.0	2.8	2.6	2.9	3.0	2.7	2.6	2.7	2.8
M4	2.3	2.4	1.9	1.8	2.4	2.3	1.9	1.8	2.3	2.4	2.1	2.2
M5	2.1	2.8	2.3	2.3	2.2	2.6	2.3	2.4	2.4	2.5	2.4	2.3
M6a	2.8	2.5	2.9	3.0	2.8	2.6	2.9	3.0	2.7	2.6	2.7	2.8

Table 4.3.6-38
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Three Groups, α .05, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	---4th---				---3rd---				---2nd---			
Distribution / *Mthd*	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
M3a	2.8	2.4	3.0	2.7	2.7	2.5	3.0	2.7	2.7	2.6	2.9	2.6
M4	3.3	3.1	2.9	3.3	3.2	3.1	2.7	3.2	2.9	3.0	2.4	3.1
M5	1.2	2.1	1.1	1.2	1.3	1.9	1.3	1.4	1.7	1.9	1.9	1.7
M6a	2.8	2.4	3.0	2.7	2.7	2.5	3.0	2.7	2.7	2.6	2.9	2.6

Table 4.3.6-39
Terpstra-Jonckheere Test
Mean Ranks of Power Results for Four Groups, α .05, by
Method and Distribution across Initial Sample Size and Effect Size

Decimal	---4th---				---3rd---				---2nd---			
Distribution / *Mthd*	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
M3a	2.7	2.1	2.9	2.6	2.6	2.2	2.8	2.7	2.6	2.4	2.8	2.7
M4	3.2	3.3	2.9	3.3	3.2	3.2	2.8	3.2	2.8	3.0	2.5	2.9
M5	1.5	2.6	1.4	1.5	1.6	2.5	1.6	1.5	2.0	2.2	2.1	1.8
M6a	2.7	2.1	2.9	2.6	2.6	2.2	2.8	2.7	2.6	2.4	2.8	2.7

Table 4.3.6-40

Terpstra-Jonckheere Test

Mean Ranks of Power Results for Five Groups, α .05, by

Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution / *Mthd*	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
M3a	2.5	2.2	2.9	2.6	2.6	2.5	2.8	2.7	2.6	2.6	2.7	2.5
M4	3.1	2.9	2.6	3.0	2.9	2.5	2.5	2.9	2.5	2.5	2.4	2.8
M5	1.9	2.8	1.8	1.9	2.0	2.6	1.9	1.8	2.3	2.4	2.3	2.3
M6a	2.5	2.2	2.9	2.6	2.6	2.5	2.8	2.7	2.6	2.6	2.7	2.5

Table 4.3.6-41

Terpstra-Jonckheere Test

Mean Ranks of Power Results for Six Groups, α .05, by

Method and Distribution across Initial Sample Size and Effect Size

Decimal	4th				3rd				2nd			
Distribution / *Mthd*	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
M3a	2.6	2.4	2.9	2.9	2.6	2.5	2.8	2.8	2.6	2.5	2.7	2.7
M4	2.4	2.4	2.0	1.8	2.5	2.4	2.1	1.9	2.4	2.5	2.2	2.2
M5	2.3	2.8	2.3	2.5	2.3	2.7	2.4	2.5	2.4	2.5	2.5	2.4
M6a	2.6	2.4	2.9	2.9	2.6	2.5	2.8	2.8	2.6	2.5	2.7	2.7

Table 4.3.6-42
Terpstra-Jonckheere Test
Analysis of Mean Ranks of Power Results
Number of First Place Finishes by Distribution
Across Nominal Alpha and Groups (.01 4-6, .05 3-6)

Decimal	| ------------- 4th			------------- |	| ------------- 2nd			------------- |
Method	3a	4	5	6a	3a	4	5	6a
By Initial Sample Size Across Nominal Effect Size Multiplier								
EA	2.33	6.00	24.33	2.33	4.42	7.08	19.08	4.42
EB	10.83	6.00	7.33	10.83	7.17	7.00	13.67	7.17
ML	0.00	12.00	23.00	0.00	3.17	13.00	15.67	3.17
SS	0.67	12.50	21.17	0.67	4.33	10.00	16.33	4.33
By Nominal Effect Size Multiplier Across Initial Sample Size								
EA	0.25	5.25	22.25	0.25	0.50	8.00	19.00	0.50
EB	4.17	3.17	2.50	4.17	1.67	2.33	8.33	1.67
ML	0.25	8.25	19.25	0.25	1.25	10.25	15.25	1.25
SS	0.25	11.25	16.25	0.25	1.33	9.00	16.33	1.33
Across Nominal Effect Size Multiplier and Initial Sample Size								
EA	0.00	0.00	7.00	0.00	0.00	1.50	5.50	0.00
EB	1.83	1.33	2.00	1.83	0.50	1.50	4.50	0.50
ML	0.00	2.00	5.00	0.00	0.00	2.00	5.00	0.00
SS	0.00	2.00	5.00	0.00	0.00	2.00	5.00	0.00

Table 4.3.6-43
Terpstra-Jonckheere Test
Analysis of Mean Ranks of Power Results
Across Nominal Alpha, Groups and Distributions

Decimal	| ------------- 4th			------------- |	| ------------- 2^{nd}			------------- |
Method	3a	4	5	6a	3a	4	5	6a
By Initial Sample Size across Nominal Effect Size Multiplier								
MP1i	140	140	140	140	140	140	140	140
N1Mi	13.83	36.50	75.83	13.83	19.08	37.08	64.75	19.08
PoM	0.099	0.261	0.542	0.099	0.136	0.265	0.463	0.136
PoT	0.099	0.261	0.542	0.099	0.136	0.265	0.463	0.136
By Nominal Effect Size Multiplier across Initial Sample Size								
MP1i	98	98	98	98	98	98	98	98
N1Mi	4.92	27.92	60.25	4.92	4.75	29.58	58.92	4.75
PoM	0.050	0.285	0.615	0.050	0.048	0.302	0.601	0.048
PoT	0.050	0.285	0.615	0.050	0.048	0.302	0.601	0.048
By Initial Sample Size and Nominal Effect Size Multiplier								
MP1i	28	28	28	28	28	28	28	28
N1Mi	1.83	5.33	19.00	1.83	0.5	7.0	20.0	0.5
PoM	0.065	0.190	0.679	0.065	0.018	0.250	0.714	0.018
PoT	0.065	0.190	0.679	0.065	0.018	0.250	0.714	0.018

Key: MP1i: Maximum possible 1st place finishes for each method
N1Mi: Actual number of 1st place finishes for each method (ties count 1/n)
PoM: Proportion of Maximum Possible { N1Mi / MP1i }
PoT: Proportion of total 1st place finishes { N1Mi / Σ(N1Mi) }

4.4 – Occurrence of Ties

All of the preceding results are only of interest if ties occur often and/or in large numbers when samples are drawn from the Micceri distributions. Thus, an ancillary part of this research was to study the frequency and pattern of occurrence of ties (equal observations resulting in tied ranks) in samples drawn from said distributions. This part of the study generated over a thousand pages of data, which are briefly summarized in tables 4.4-1 through 4.4-23.

Tables 4.4-1 and 4.4-2 show the type of data that was generated. Table 4.4-1 presents data in terms of repetitions while Table 4.4-2 presents data in terms of observations (data points). The structure of the tables is the same, with a table being generated for each group and for the total combined sample, both for the repetitions and the observations. Thus, for a two-group test there would be six tables (group 1 repetitions, group 2 repetitions, all groups repetitions, group 1 observations, group 2 observations, all groups observations).

In all cases there are INT(N/2) rows, where N is the total (combined) sample size being tested (per-group sample size times number of groups). This is the maximum number of between-group ties that can theoretically occur, in which every observation is tied with exactly one other observation from a different group. The rows, in turn, represent the results for the first between-group tie, the second between-group tie, and so on up to the INT(N/2) between-group tie, accumulated across all repetitions.

The first column designates the group. The entries in the second column are the between-group tie occurrence, i.e., 1=1st, 2=2nd, 3=3rd, etc. The third column contains the number of repetitions (or observations) involved in the tie for that row. The fourth column is the ratio of the entry in the third column (same row) to the total number of repetitions (observations). The fifth column is the reverse accumulation of the third column, while the sixth column is the ratio of the corresponding entry in the fifth column to total repetitions (observations).

Table 4.4-1
Micceri Smooth Symmetric (Achievement) Distribution
Repetitions Data – Two Groups, Six per group – NESM = 0.0 (no shift)
1,000,000 repetitions; 12,000,000 observations

Group	BGT	----- per BGT -----		----- cumulative -----	
		count	ratio	count	ratio
1	1	858,432	.8584	1,480,377	1.4804
	2	472,094	.4721	621,945	.6219
	3	132,841	.1328	149,851	.1499
	4	16,249	.0162	17,010	.0170
	5	752	.0008	761	.0008
	6	9	.0000	9	.0000
2	1	858,432	.8584	1,480,377	1.4804
	2	472,094	.4721	621,945	.6219
	3	132,841	.1328	149,851	.1499
	4	16,249	.0162	17,010	.0170
	5	752	.0008	761	.0008
	6	9	.0000	9	.0000
all	1	858,432	.8584	1,480,377	1.4804
	2	472,094	.4721	621,945	.6219
	3	132,841	.1328	149,851	.1499
	4	16,249	.0162	17,010	.0170
	5	752	.0008	761	.0008
	6	9	.0000	9	.0000

For Table 4.4-1 (repetitions) the numbers in the fifth and sixth columns do not have an obvious direct interpretation, as the reverse accumulated counts in column five can, and sometimes do, exceed the total number of repetitions. As a result, the corresponding ratios in column six can, and sometimes do, exceed 1.0. The ratio in the first row of column six, however, does give a kind of relative index of the occurrence of ties that can be compared across distributions, number of groups and per-group sample sizes. Simply put, the larger this number, the more often ties are occurring in the data.

For Table 4.4-2 (observations) the entry in the first row, fifth column has a direct meaning; it is the total number of observations involved in a between-group tie during that test run. Since an observation involved in a first between-group tie cannot also be involved in a subsequent (2^{nd}, 3^{rd}, 4^{th}, etc.) tie, this result cannot exceed the total number of observations tested. The corresponding ratio in column six cannot, therefore, exceed one. Thus, this ratio gives a very direct indication of how often ties are occurring in terms of the proportion of total observations involved in between-group ties.

Although a first tie must occur before there can be a second tie, and a second tie before a third, and so on by definition, the determination of the first, second, etc. tie is within each combined sample as it is drawn. If the first tie in a particular sample involves observations from groups 1 and 2, but not group 3, then the repetition counters for a first tie in groups 1, 2 and all groups get incremented, but not the counter for group 3. If a second between-group tie is then found involving observations from groups 2 and 3, the corresponding second tie counters get incremented. Thus, it is possible for a group to have been involved in more second ties than first ties, and this is, in fact, observed in the data for each group. This cannot occur, however, for the all groups (total sample) counter.

Table 4.4-2
Micceri Smooth Symmetric (Achievement) Distribution
Observations Data – Two Groups, Six per group – NESM = 0.0 (no shift)
1,000,000 repetitions; 12,000,000 observations

Group	BGT	----- per BGT ----- count	ratio	----- cumulative ----- count	ratio
1	1	1,020,228	.0850	1,735,789	.1446
	2	549,569	.0458	715,561	.0596
	3	147,877	.0123	165,992	.0138
	4	17,334	.0014	18,115	.0015
	5	772	.0001	781	.0001
	6	9	.0000	9	.0000
2	1	1,020,051	.0850	1,735,177	.1446
	2	549,300	.0458	715,126	.0596
	3	147,772	.0123	165,826	.0138
	4	17,316	.0014	18,104	.0015
	5	779	.0001	788	.0001
	6	9	.0000	9	.0000
all	1	2,040,279	.1700	3,470,966	.2892
	2	1,098,869	.0916	1,430,687	.1192
	3	395,599	.0246	331,818	.0277
	4	34,650	.0029	36,219	.0030
	5	1,551	.0001	1,569	.0001
	6	18	.0000	18	.0000

Data was generated for each group as well as for the total of all groups to see how ties were distributed across groups. *A priori*, it was expected that for two groups (under all conditions), and for three or more groups (with no shift), between-group ties would be equally distributed across the groups. This was, in fact, observed within sampling error. For three or more groups, as an increasing amount of shift was introduced, one would expect the extreme groups (not shifted on the left, most shifted on the right) to have fewer observations involved in between-group ties than those groups in the middle. This pattern was also present. Table 4.4-3 shows the repetitions and observations results for the Micceri Smooth Symmetric distribution with three groups of three observations each and six groups of six observations each, both at NESM 0.0 (no shift) and 1.2 (maximum shift).

Table 4.4-3
Micceri Smooth Symmetric (Achievement) Distribution
Comparison of ties across three or more groups
Repetitions and observations ratio for all groups
Based on 1,000,000 repetitions

Obs/grp	----------------- 3 -----------------				------------------ 6 ------------------			
NESM	0.0		1.2		0.0		1.2	
Reps | Obs Group	RC	OC	RC	OC	RC	OC	RC	OC
1	.8116	.0963	.4088	.0480	3.9955	.1292	1.3664	.0434
2	.8130	.0964	.6160	.0725	3.9959	.1292	2.2915	.0729
3	.8123	.0963	.4096	.0481	3.9964	.1292	2.5407	.0806
4					3.9961	.1292	2.5397	.0805
5					3.9943	.1291	2.2919	.0729
6					3.9956	.1292	1.3663	.0434
all	1.1779	.2890	.7068	.1686	9.1897	.7750	5.8883	.3937

RC = cumulative repetitions ratio; OC = cumulative observations ratio

In Tables 4.4-4 through 4.4-23 the data on repetitions and observations is combined. Each table contains the results for one of the four Micceri distributions (EA, EB, ML, SS) for a specific number of groups (2, 3, 4, 5, 6). For each combination, results are presented at various per-group sample sizes.

A priori, it would be expected that the number of ties would increase as the per-group sample size increased, all other factors being held constant, and this was, in fact, observed. For two groups, results are given at per-group sample sizes of 3, 6, 9, 12, and 15. For three, four, and five groups, results are given for per-group sample sizes of 3, 6, and 9, while for six groups, results for per-group sample sizes of 3 and 6 are reported. In all cases, results are given for nominal effect size multipliers of 0.0, 0.2, 0.5, 0.8, and 1.2 (except for the Extreme Bi-modal distribution, for which NESMs of .2 and .8 are not defined). The total sample size is also given and is equal to the number of groups times the per-group sample size.

Simulation results in tables 4.4-4 through 4.4-23 are based on 1,000,000 repetitions per combination. Both repetition and observation results are selectively combined in each table. The first column is the per-group sample size (**SS**). The second column is the total sample size (**TSS**). The third column is the nominal effect size multiplier (**NESM**). Columns four through ten present seven results for each combination as follows: 4) **R-1**, the ratio of repetitions with a first between-group tie to total repetitions; 5) **R-C**, the ratio of cumulative repetitions with one or more between-group ties to total repetitions; 6) **O-1**, the ratio of observations involved in a first between-group tie to the total number of observations; 7) **O-A**, the ratio of observations involved in all between-group ties to total observations; 8) **M#**, the maximum number of between-group ties observed; 9) **R-M#**, the number of repetitions in which the maximum number of between-group ties was observed, and; 10) **O-M#**, the number of observations involved in the maximum number of between-group ties. Note again that R-C can, and often does, result in ratios greater than 1.0.

The simulations were also run for the normal distribution. *A priori*, results should have been zero for all combinations. Minor discrepancies were noted, particularly at combinations of larger numbers of groups and greater per-group sample sizes. However, when ties did occur, they were generally confined to a first tie, and were very few in number. As such, they are not presented here in table form.

Table 4.4-4
Micceri Extreme Asymmetric (Achievement) Distribution
Occurrence of Between-group Ties, Two Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	6	0.0	.524	.61	.198	.23	3	1976	3952
		0.2	.460	.52	.172	.19	3	1140	2280
		0.5	.329	.36	.120	.13	3	353	706
		0.8	.238	.25	.086	.09	3	128	256
		1.2	.180	.19	.065	.07	3	41	82
6	12	0.0	.927	1.76	.200	.39	6	7	14
		0.2	.884	1.53	.188	.33	6	8	16
		0.5	.752	1.11	.154	.23	5	87	188
		0.8	.613	.81	.123	.16	5	14	29
		1.2	.499	.62	.099	.12	5	2	4
9	18	0.0	.994	2.97	.149	.50	8	8	17
		0.2	.986	2.64	.146	.43	8	2	4
		0.5	.938	2.00	.137	.31	7	21	50
		0.8	.848	1.50	.122	.22	6	96	241
		1.2	.746	1.16	.106	.17	6	25	58
12	24	0.0	.+	4.11	.111	.58	10	1	3
		0.2	.999	3.71	.111	.51	9	37	97
		0.5	.988	2.90	.110	.37	9	3	8
		0.8	.949	2.24	.106	.27	8	9	24
		1.2	.886	1.76	.099	.21	8	2	4
15	30	0.0	.+	5.15	.088	.64	11	17	45
		0.2	.+	4.69	.088	.56	11	6	21
		0.5	.998	3.77	.089	.42	10	1	2
		0.8	.985	2.97	.089	.31	10	2	6
		1.2	.952	2.36	.087	.24	9	1	2

* For R-1: .+ indicates greater than .999

Table 4.4-5
Micceri Extreme Asymmetric (Achievement) Distribution
Occurrence of Between-group Ties, Three Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	9	0.0	.894	1.46	.238	.39	4	3690	7679
		0.2	.825	1.24	.214	.32	4	1834	3769
		0.5	.653	.86	.163	.21	4	365	743
		0.8	.501	.60	.122	.15	4	81	166
		1.2	.386	.44	.093	.11	4	16	32
6	18	0.0	.+	3.59	.140	.58	8	55	120
		0.2	.998	3.27	.139	.50	8	19	40
		0.5	.975	2.51	.134	.36	8	1	2
		0.8	.911	1.88	.123	.26	8	1	2
		1.2	.818	1.42	.109	.19	7	2	4
9	27	0.0	1.0	5.44	.092	.69	11	12	32
		0.2	.+	5.17	.092	.61	11	8	17
		0.5	.+	4.28	.093	.46	10	37	91
		0.8	.992	3.35	.094	.34	10	2	5
		1.2	.965	2.61	.092	.26	9	6	13

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-6
Micceri Extreme Asymmetric (Achievement) Distribution
Occurrence of Between-group Ties, Four Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	12	0.0	.989	2.36	.202	.50	6	48	96
		0.2	.964	2.06	.192	.41	6	29	58
		0.5	.852	1.44	.161	.27	6	3	6
		0.8	.702	1.01	.130	.18	5	62	125
		1.2	.558	.72	.101	.13	5	10	20
6	24	0.0	1.0	5.17	.101	.69	11	1	2
		0.2	.+	4.97	.102	.61	11	1	3
		0.5	.999	4.03	.102	.44	10	7	16
		0.8	.985	3.04	.101	.32	10	1	2
		1.2	.942	2.27	.095	.23	9	1	2
9	36	0.0	1.0	7.34	.067	.78	14	4	14
		0.2	1.0	7.41	.067	.71	14	5	13
		0.5	.+	6.62	.069	.55	13	40	102
		0.8	.+	5.31	.071	.41	13	2	6
		1.2	.996	4.11	.072	.31	12	2	4

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-7
Micceri Extreme Asymmetric (Achievement) Distribution
Occurrence of Between-group Ties, Five Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	15	0.0	.+	3.25	.162	.58	7	145	298
		0.2	.995	2.91	.159	.48	7	84	174
		0.5	.946	2.06	.145	.31	7	1	2
		0.8	.830	1.43	.123	.21	6	31	62
		1.2	.683	.99	.100	.14	6	2	4
6	30	0.0	1.0	6.51	.079	.75	12	37	105
		0.2	1.0	6.60	.080	.68	13	2	4
		0.5	.+	5.59	.081	.50	12	15	31
		0.8	.998	4.24	.082	.36	11	12	24
		1.2	.981	3.11	.080	.25	10	4	9
9	45	0.0	1.0	8.89	.053	.83	17	1	5
		0.2	1.0	9.42	.053	.77	16	16	46
		0.5	1.0	8.93	.054	.62	17	1	2
		0.8	.+	7.28	.056	.46	16	1	2
		1.2	.+	5.61	.057	.34	14	5	14

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-8
Micceri Extreme Asymmetric (Achievement) Distribution
Occurrence of Between-group Ties, Six Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	18	0.0	.+	4.09	.133	.64	8	256	564
		0.2	.+	3.79	.133	.53	8	173	361
		0.5	.982	2.71	.126	.35	8	7	14
		0.8	.905	1.85	.113	.23	8	1	2
		1.2	.773	1.27	.094	.15	7	1	2
6	36	0.0	1.0	7.67	.066	.80	14	6	20
		0.2	1.0	8.15	.066	.73	14	30	77
		0.5	.+	7.17	.068	.54	14	9	19
		0.8	.+	5.44	.069	.38	13	5	11
		1.2	.994	3.96	.067	.27	12	3	8

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-9
Micceri Extreme Bi-modal (Psychometric) Distribution
Occurrence of Between-group Ties, Two Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	6	0.0	.892	1.28	.422	.58	3	25070	50140
		0.5	.558	.64	.237	.27	3	1626	3252
		1.2	.320	.34	.130	.14	3	222	444
6	12	0.0	.998	2.47	.369	.79	5	1436	3660
		0.5	.896	1.44	.259	.41	5	27	58
		1.2	.657	.86	.176	.23	4	82	209
9	18	0.0	.+	3.15	.353	.88	6	29	106
		0.5	.978	2.03	.240	.49	5	872	3325
		1.2	.832	1.30	.190	.29	4	1121	3722
12	24	0.0	1.0	3.55	.350	.92	6	144	722
		0.5	.996	2.46	.226	.54	5	3901	20723
		1.2	.917	1.65	.194	.33	4	4283	18136
15	30	0.0	1.0	3.83	.349	.94	6	583	4073
		0.5	.+	2.78	.218	.58	5	10042	68421
		1.2	.958	1.93	.194	.37	4	10228	52576

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-10
Micceri Extreme Bi-modal (Psychometric) Distribution
Occurrence of Between-group Ties, Three Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	9	0.0	.999	2.26	.391	.79	4	20256	45418
		0.5	.879	1.42	.260	.41	4	3358	7052
		1.2	.751	1.00	.226	.29	4	453	924
6	18	0.0	1.0	3.40	.352	.92	6	56	202
		0.5	.996	2.91	.203	.59	6	510	1500
		1.2	.975	2.06	.205	.42	6	25	59
9	27	0.0	1.0	3.91	.349	.96	6	679	4273
		0.5	.+	3.79	.176	.67	6	6649	28806
		1.2	.998	2.78	.176	.48	6	551	1805

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-11

Micceri Extreme Bi-modal (Psychometric) Distribution
Occurrence of Between-group Ties, Four Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	12	0.0	.+	2.89	.362	.88	6	2	4
		0.5	.983	2.31	.234	.53	6	36	72
		1.2	.917	1.65	.211	.37	5	503	1044
6	24	0.0	1.0	3.86	.349	.96	6	484	2485
		0.5	.+	4.17	.164	.70	7	1144	3833
		1.2	.999	3.25	.159	.51	8	4	13
9	36	0.0	1.0	4.31	.348	.98	6	3252	28589
		0.5	1.0	5.03	.139	.77	7	10352	49247
		1.2	.+	4.26	.132	.58	8	307	1107

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-12
Micceri Extreme Bi-modal (Psychometric) Distribution
Occurrence of Between-group Ties, Five Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	15	0.0	1.0	3.29	.354	.92	6	21	59
		0.5	.999	3.10	.200	.61	7	62	136
		1.2	.975	2.31	.181	.42	7	3	7
6	30	0.0	1.0	4.15	.349	.97	6	1782	12646
		0.5	1.0	5.15	.133	.76	8	1302	4508
		1.2	.+	4.44	.127	.57	9	246	637
9	45	0.0	1.0	4.46	.346	.99	6	8215	93219
		0.5	1.0	6.03	.111	.81	8	10768	51736
		1.2	1.0	5.74	.106	.64	10	115	371

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-13
Micceri Extreme Bi-modal (Psychometric) Distribution
Occurrence of Between-group Ties, Six Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	18	0.0	1.0	3.57	.351	.94	6	132	478
		0.5	.+	3.89	.168	.66	8	59	130
		1.2	.993	2.96	.154	.45	8	6	13
6	36	0.0	1.0	4.37	.348	.98	6	3998	35138
		0.5	1.0	6.12	.111	.80	9	1391	4900
		1.2	.+	5.64	.106	.61	11	57	148

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-14
Micceri Multi-modal Lumpy (Achievement) Distribution
Occurrence of Between-group Ties, Two Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1	R-C	O-1	O-C	M#	R-M#	O-M#
3	6	0.0	.226	.24	.078	.08	3	118	236
		0.2	.204	.21	.070	.07	3	91	182
		0.5	.170	.18	.058	.06	3	55	110
		0.8	.137	.14	.047	.05	3	31	62
		1.2	.115	.12	.039	.04	3	26	52
6	12	0.0	.633	.87	.115	.16	5	35	71
		0.2	.590	.78	.107	.14	5	23	46
		0.5	.514	.65	.092	.12	5	9	18
		0.8	.435	.52	.078	.09	5	8	16
		1.2	.376	.44	.067	.08	5	2	4
9	18	0.0	.890	1.77	.114	.22	7	15	30
		0.2	.858	1.60	.110	.20	7	13	26
		0.5	.794	1.35	.099	.17	7	2	4
		0.8	.710	1.09	.088	.13	6	21	43
		1.2	.640	.92	.079	.11	6	11	23
12	24	0.0	.978	2.87	.099	.28	9	14	28
		0.2	.965	2.61	.098	.26	9	6	12
		0.5	.934	2.21	.092	.21	9	2	4
		0.8	.880	1.80	.086	.17	8	6	13
		1.2	.825	1.52	.080	.15	7	29	61
15	30	0.0	.997	4.10	.085	.34	12	1	2
		0.2	.994	3.74	.084	.31	11	3	6
		0.5	.983	3.20	.081	.26	11	1	2
		0.8	.959	2.62	.078	.21	9	34	74
		1.2	.927	2.21	.074	.18	10	1	2

Table 4.4-15
Micceri Multi-modal Lumpy (Achievement) Distribution
Occurrence of Between-group Ties, Three Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	9	0.0	.543	.67	.127	.16	4	146	294
		0.2	.492	.59	.115	.14	4	99	198
		0.5	.400	.46	.093	.11	4	37	74
		0.8	.338	.38	.078	.09	4	12	24
		1.2	.276	.30	.063	.07	4	1	2
6	18	0.0	.955	2.28	.121	.28	8	3	6
		0.2	.931	2.05	.117	.25	7	46	93
		0.5	.866	1.64	.106	.20	7	9	18
		0.8	.799	1.36	.097	.16	6	123	253
		1.2	.712	1.09	.086	.13	7	1	2
9	27	0.0	.+	4.41	.090	.39	11	9	18
		0.2	.997	4.02	.089	.35	11	1	2
		0.5	.988	3.29	.085	.28	10	3	6
		0.8	.970	2.75	.083	.23	9	19	39
		1.2	.932	2.22	.078	.19	8	45	93

* For R-1: .+ indicates greater than .999

Table 4.4-16
Micceri Multi-modal Lumpy (Achievement) Distribution
Occurrence of Between-group Ties, Four Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	12	0.0	.796	1.25	.143	.22	5	212	427
		0.2	.738	1.09	.131	.19	6	1	2
		0.5	.619	.83	.108	.14	5	23	47
		0.8	.525	.66	.092	.12	5	5	10
		1.2	.413	.49	.071	.08	5	2	4
6	24	0.0	.998	4.04	.098	.39	10	5	10
		0.2	.995	3.62	.096	.34	10	2	4
		0.5	.978	2.86	.091	.27	9	15	30
		0.8	.945	2.32	.087	.21	8	34	71
		1.2	.874	1.75	.079	.16	8	4	8
9	36	0.0	1.0	7.42	.069	.52	15	1	2
		0.2	.+	6.81	.068	.46	14	14	30
		0.5	.+	5.57	.066	.36	13	8	16
		0.8	.998	4.47	.064	.30	13	2	4
		1.2	.989	3.53	.063	.22	11	6	12

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-17
Micceri Multi-modal Lumpy (Achievement) Distribution
Occurrence of Between-group Ties, Five Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	15	0.0	.933	1.96	.137	.28	7	4	8
		0.2	.889	1.69	.128	.24	7	1	2
		0.5	.783	1.27	.111	.18	6	19	38
		0.8	.665	.95	.093	.13	6	1	2
		1.2	.525	.68	.073	.09	5	26	52
6	30	0.0	.+	5.98	.081	.48	13	1	2
		0.2	.+	5.37	.079	.42	12	5	10
		0.5	.998	4.22	.075	.32	11	9	18
		0.8	.986	3.28	.073	.24	10	13	26
		1.2	.946	2.41	.069	.17	9	18	37
9	45	0.0	1.0	10.49	.055	.61	18	12	25
		0.2	1.0	9.74	.055	.55	18	3	6
		0.5	1.0	7.99	.053	.43	16	8	16
		0.8	.+	6.39	.052	.33	16	1	2
		1.2	.998	4.85	.051	.25	13	8	16

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-18

Micceri Multi-modal Lumpy (Achievement) Distribution
Occurrence of Between-group Ties, Six Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	18	0.0	.984	2.76	.123	.34	8	8	16
		0.2	.961	2.36	.117	.28	8	1	2
		0.5	.885	1.73	.105	.21	7	14	29
		0.8	.764	1.24	.089	.14	7	1	2
		1.2	.615	.86	.071	.10	6	8	16
6	36	0.0	1.0	8.01	.068	.55	15	3	6
		0.2	1.0	7.23	.066	.48	15	2	5
		0.5	.+	5.64	.063	.36	13	9	20
		0.8	.997	4.24	.062	.26	12	4	8
		1.2	.977	3.07	.059	.19	11	1	2

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-19
Micceri Smooth Symmetric Distribution
Occurrence of Between-group Ties, Two Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	6	0.0	.401	.45	.143	.16	3	859	1718
		0.2	.396	.44	.141	.16	3	795	1590
		0.5	.387	.43	.138	.15	3	761	1522
		0.8	.350	.38	.125	.14	3	553	1106
		1.2	.295	.32	.105	.11	3	295	590
6	12	0.0	.858	1.48	.170	.29	6	9	18
		0.2	.854	1.46	.169	.29	6	1	2
		0.5	.844	1.43	.167	.28	6	7	14
		0.8	.802	1.28	.157	.25	6	1	2
		1.2	.728	1.07	.141	.21	5	128	263
9	18	0.0	.984	2.78	.141	.40	8	16	32
		0.2	.983	2.76	.141	.39	8	11	22
		0.5	.980	2.69	.140	.38	8	12	24
		0.8	.966	2.43	.137	.34	8	4	8
		1.2	.932	2.06	.131	.29	8	1	2
12	24	0.0	.999	4.18	.113	.48	10	13	27
		0.2	.999	4.14	.113	.48	10	9	19
		0.5	.999	4.04	.113	.47	10	10	22
		0.8	.996	3.67	.113	.42	10	1	2
		1.2	.987	3.13	.112	.35	10	1	2
15	30	0.0	.+	5.55	.093	.56	13	1	2
		0.2	.+	5.50	.094	.55	13	1	2
		0.5	.+	3.38	.094	.54	12	8	17
		0.8	.+	4.91	.095	.49	12	2	4
		1.2	.998	4.21	.096	.41	11	4	8

* For R-1: .+ indicates greater than .999

Table 4.4-20
Micceri Smooth Symmetric Distribution
Occurrence of Between-group Ties, Three Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	9	0.0	.797	1.18	.197	.29	4	1674	3399
		0.2	.789	1.16	.195	.28	4	1660	3352
		0.5	.767	1.11	.189	.27	4	1283	2598
		0.8	.684	.92	.166	.22	4	515	1043
		1.2	.566	.71	.136	.17	4	135	273
6	18	0.0	.998	3.48	.138	.48	8	94	191
		0.2	.997	3.44	.138	.48	8	76	155
		0.5	.996	3.31	.138	.46	8	50	102
		0.8	.987	2.86	.135	.39	8	7	16
		1.2	.956	2.28	.127	.30	8	1	2
9	27	0.0	.+	5.94	.097	.62	12	3	6
		0.2	.+	5.87	.097	.61	13	1	2
		0.5	.+	5.69	.098	.59	12	2	4
		0.8	.+	5.04	.098	.50	11	12	26
		1.2	.998	4.17	.096	.40	10	37	81

* For R-1: .+ indicates greater than .999

Table 4.4-21
Micceri Smooth Symmetric Distribution
Occurrence of Between-group Ties, Four Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	12	0.0	.963	2.08	.186	.40	6	27	54
		0.2	.958	2.03	.184	.39	6	28	56
		0.5	.943	1.90	.180	.36	6	16	32
		0.8	.869	1.50	.161	.28	6	1	2
		1.2	.743	1.10	.134	.20	5	85	174
6	24	0.0	.+	5.56	.106	.62	11	11	23
		0.2	.+	5.47	.106	.60	11	12	24
		0.5	.+	5.23	.106	.57	11	1	2
		0.8	.+	4.43	.103	.46	10	20	42
		1.2	.994	3.48	.100	.35	9	46	96
9	36	0.0	1.0	8.71	.073	.75	15	8	16
		0.2	1.0	8.62	.074	.73	15	4	8
		0.5	1.0	8.38	.074	.70	15	2	4
		0.8	1.0	7.51	.074	.59	14	29	65
		1.2	.+	6.27	.072	.46	14	1	2

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-22
Micceri Smooth Symmetric Distribution
Occurrence of Between-group Ties, Five Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	15	0.0	.997	3.07	.168	.48	7	150	304
		0.2	.995	2.99	.157	.47	7	116	235
		0.5	.990	2.75	.154	.42	7	49	98
		0.8	.950	2.09	.142	.31	7	7	14
		1.2	.850	1.49	.123	.22	6	46	93
6	30	0.0	1.0	7.48	.085	.71	14	1	2
		0.2	1.0	7.37	.085	.69	13	20	41
		0.5	1.0	7.06	.085	.64	13	8	16
		0.8	.+	6.00	.083	.51	13	1	2
		1.2	.+	4.68	.080	.37	12	2	5
9	45	0.0	1.0	10.95	.059	.82	18	3	6
		0.2	1.0	10.90	.059	.81	17	45	101
		0.5	1.0	10.73	.059	.76	18	2	4
		0.8	1.0	9.92	.059	.64	18	1	2
		1.2	1.0	8.36	.058	.49	16	17	38

* For R-1: .+ indicates greater than .999; 1.0 is 100%

Table 4.4-23
Micceri Smooth Symmetric Distribution
Occurrence of Between-group Ties, Six Groups – 1,000,000 repetitions

SS	TSS	NESM	R-1*	R-C	O-1	O-C	M#	R-M#	O-M#
3	18	0.0	.+	4.10	.135	.56	9	1	2
		0.2	.+	3.97	.134	.53	9	1	2
		0.5	.999	3.61	.131	.47	9	2	4
		0.8	.981	2.68	.123	.33	8	1	2
		1.2	.912	1.89	.111	.23	7	22	44
6	36	0.0	1.0	9.19	.072	.78	15	29	59
		0.2	1.0	9.09	.072	.75	16	1	2
		0.5	1.0	8.79	.071	.69	16	1	2
		0.8	1.0	7.56	.069	.54	15	4	8
		1.2	.+	5.89	.067	.39	13	19	42

* For R-1: .+ indicates greater than .999; 1.0 is 100%

4.5 – Sampling Adequacy

Prior to conducting the main simulations an ancillary study was performed to determine the number of simulation cycles required to obtain adequate samples from the Micceri (1986, 1989) distributions using the MOTHER (Blair, undated) random number generator with an initial seed of 255,255. Jack Hill (private correspondence) investigated initial seeds with the MOTHER random number generator and validated the use of 255,255. The methodology used for the sampling adequacy study is described in Chapter 3, Section 8 and conclusions about sampling adequacy are included in Chapter 5, Section 3.

The results of this ancillary study are summarized in tables 4.5-1 through 4.5-4 based on the percent error between each obtained sample distribution and its population distribution at each value. Not all values are shown. The percent error between each obtained sample cumulative distribution and its corresponding population cumulative distribution was also investigated. The results were very similar to those shown and not given here.

In tables 4.5-1 through 4.5-4, the following apply to columns A through D:

A. Total observations (obs) in thousands = (total sample size) x (total repetitions).

B. The ratio of total observations (obs) to the source vector length (SVL), rounded to the nearest integer. The source vector length is the length of the generating vector that defines a population distribution as described in Chapter 3, section 8.

C. The number of population values at which the observed sample distribution and the population distribution differed by 1.0 % or more.

D. The maximum percent error between the obtained sample distribution and the population distribution.

Table 4.5-1
Sampling Adequacy for Micceri (1986, 1989) Extreme Asymmetric Distribution 27 Values [1,27,1] with Source Vector Length (SVL) 2,768

A Total Observations (1,000's)	B Obs/SVL ratio (approx)	C pd diff # of values with |%error| ≥ 1.0%	D pd diff maximum % error
60	22	17	6.16
180	65	18	5.61
480	173	10	-2.47
1,800	650	6	-1.48
4,800	1,734	4	-1.75
18,000	6,503	0	-0.81
48,000	17,341	0	0.33

Table 4.5-2
Sampling Adequacy for Micceri (1986, 1989) Extreme Bi-modal Distribution 6 Values [1,6,1] with Source Vector Length (SVL) 665

A Total Observations (1,000's)	B Obs/SVL ratio (approx)	C pd diff # of values with |%error| ≥ 1.0%	D pd diff maximum % error
60	90	1	-1.11
180	271	1	-1.97
480	722	1	-1.91
1,800	2,707	0	-0.48
3,600	5,414	1	-1.01
4,800	7,218	0	-0.84
18,000	27,068	0	0.43
48,000	72,180	0	0.43

Table 4.5-3
Sampling Adequacy for Micceri (1986, 1989) Multi-modal Lumpy Distribution 44 Values [1,44,1] with Source Vector Length (SVL) 467

A Total Observations (1,000's)	B Obs/SVL ratio (approx)	C pd diff # of values with |%error| ≥ 1.0%	D pd diff maximum % error
60	128	34	6.63
180	385	26	-5.30
480	1,028	17	3.58
1,800	3,854	4	1.49
4,800	10,278	0	-0.75
18,000	38,544	0	0.40
48,000	102,784	0	0.22

The total number of observations needed to make the difference between each obtained sample distribution and its population distribution less than or equal to 1.0% for all values was approximately 18,000,000 for distribution EA, 6,000,000 for distribution SS and 4,800,000 for distributions EB and ML. The complete set of data points for each population distribution were fitted with a smooth curve and plotted. These curves indicated that approximately 10,000,000 observations might be sufficient to minimize the error to 1.0% or less. Since the minimum combined sample size studied was six (two groups of three or three groups of two), 6,000,000 observations, requiring 1,000,000 repetitions, was selected as a reasonable compromise for the lower limit, even though the EA distribution might have benefited from a slightly larger choice.

Table 4.5-4
Sampling Adequacy for Micceri (1986, 1989) Smooth Symmetric Distribution 27 Values [1,27,1] with Source Vector Length (SVL) 5,375

A Total Observations (1,000's)	B Obs/SVL ratio (approx)	C pd diff # of values with |%error| ≥ 1.0%	D pd diff maximum % error
60	11	17	-9.52
180	33	15	-4.97
480	89	9	2.64
1,800	335	3	-1.21
4,800	893	1	-1.17
18,000	3,349	0	0.56
48,000	8,930	0	0.38

In addition to the 1.0 % error criteria, each distribution of sample values was tested against its corresponding population distribution using the 2-sided Kolmogorov-Smirnov Test of General Differences at nominal alpha .10. A 2-sided test was used because the distribution of sample values tended to deviate from its Micceri (1986, 1989) population distribution both positively and negatively. Nominal alpha was set at 10% to ensure that the test was powerful at detecting real differences. Even with the elevated risk of a false positive, this test never produced a significant result. Thus, the 1.0% error criteria was more stringent than that represented by the use of the 2-sided Kolmogorov-Smirnov Test at nominal alpha .10.

The sampling adequacy study also investigated the use of total numbers of observations that were integral multiples of the source vector length. The results did not differ from those obtained with total observations that were integral multiples of 10, and so are not reported here.

CHAPTER 5

DISCUSSION

5.1 — Overview

Before discussing the four research questions posed in Chapter 1, the issues of occurrence of ties and sampling adequacy are addressed. The findings with respect to the original research questions are then presented, followed by comments on the apparent effect of the discreteness of the population and critical value distributions on the results. Limitations of the study are dealt with next and possible directions for further research are suggested. Conclusions are based on the results presented in Chapter 4. Throughout this chapter, methods and distributions are abbreviated as shown in Tables 5.1-1 and 5.1-2. Table 5.1-2 also contains the parameters for each population distribution.

Table 5.1-1
Methods of Resolving Equal Data Values

Abbrev.	Name
3a	Average of Least (M1) and Most (M2) Likely to Reject H_0
3b	Count Ties as ½
4	Alternating
5	Random
6a	Mid-ranks
6b	Delayed Increment
6c	Weighted Average of All Possible Resolutions
7	Drop Matching Ties and Reduce N
8	Drop All Ties and Reduce N

Table 5.1-2
Micceri (1986, 1989) Distributions (from Table 3.2-1)

Abbrev.	Name	min	max	μ	ϕ	σ	skew	kurtosis
EA	Extreme Asymmetric	0	30	24.50	27.	5.79	-1.33	4.11
EB	Extreme Bi-modal	0	7	2.97	4.	1.69	-0.08	1.30
ML	Multi-modal Lumpy	0	43	21.15	18.	11.90	0.19	1.80
SS	Smooth Symmetric	0	26	13.19	13.	4.91	0.01	2.66
Norm	Normal	$-\infty$	$+\infty$	0.00	0.	1.00	0.00	3.00

Adapted from Sawilowsky & Blair (1992) p.353 Table 1

5.2 — Occurrence of Ties

Preliminary to the main study itself, the occurrence of between-group ties was investigated when sampling from the various distributions used in this study. Since the main purpose of this study was to determine the relative efficacy of different methods of resolving consequential ties, the whole enterprise would have been largely without relevance if ties did not occur with some relatively high frequency when samples were drawn from the Micceri (1986, 1989) distributions. Consequential ties turn out to be the ones that occur between different groups in any given sample. The results of the investigation of the occurrence of ties are described in Chapter 4, section 4, and it is clear from those results that ties occur often in samples drawn from the Micceri (1986, 1989) distributions.

Tables 4.4-4 through 4.4-8 provide a summary of the results for distribution EA for two to six groups. At nominal effect size multiplier 0.0 (no shift) the least number of repetitions involved in a first between-group tie was 52.4% for two groups of three taking in 19.8% of the observations. Overall 23% of the observations where involved in a consequential tie. These rates rose quickly to 92.7% of repetitions experiencing a first between-group tie for two groups of six and 89.4% for three groups of three. The corresponding observations were 20.0% for first tie and 39% overall for two groups of six and 23.8% for first tie and 39% overall for three groups of three. At any given number of groups and initial sample size, these rates declined steadily with increases in the nominal effect size multiplier. Even so, the smallest values obtained were for two groups of three at nominal effect size multiplier 1.2 where 18.0% of repetitions were involved in a first tie taking in 6.5% of the total observations, with 7% of the observations involved in a between-group tie.

The EA distribution has 31 distinct values. Not surprisingly, the EB distribution, with only 8 distinct values, produced much higher numbers of between-group ties. Tables 4.4-9 through 4.4-13 summarize the results for the EB distribution. For two groups of three at nominal effect size multiplier 0.0, 89.2% of repetitions had a first between-group tie involving 42.2% of the observations, with 58% of observations overall involved in a between-group tie. Indeed, at two groups of six and above, essentially every repetition involved between-group ties. At two groups of 15, 34.9% of the observations were involved in a first between-group tie with 94% of total observations involved in a between-group tie. For three or more groups, most sample sizes saw greater than 90% of the observations involved in a between-group tie.

The ML distribution has 44 distinct values. Results are given in Tables 4.4-14 through 4.4-18. For two groups of three at nominal effect size multiplier 0.0, 22.6% of repetitions had first between-group ties involving 7.8% of the observations with 8% of observations involved in between-group ties. At nominal effect size multiplier 1.2 these rates dropped to 11.5%, 3.9% and 4% respectively. For two groups of six at nominal effect size multiplier 0.0, the rate of repetitions experiencing a first tie nearly tripled to 63.3% while the observations increased to 11.5% for first ties and 16% overall, double that of two groups of three. For three groups of three, 54.3% of repetitions saw a first tie involving 12.7% of the observations with 16% of the observations involved in a between-group tie. At six groups of three almost every repetition saw a between-group tie at 98.4%. The rate of observations involved in first ties was 12.3% with 34% of observations involved in a between-group tie.

Finally, the SS distribution, with 27 distinct values, experienced ties at a rate slightly lower than the EA distribution. These results are given in Tables 4.4-19 through 4.4-23. If the rate at which between-group ties was experienced depended only on the number of distinct values in the distribution one would have expected the SS distribution to have slightly higher rates than the EA distribution. Thus, it is possible that shape, or some other factor, also contributed to the rate of occurrence and pattern of ties.

All other factors being equal, the following were both expected and generally observed:

- Samples drawn from populations with fewer defined values tended to produce more between-group ties than samples drawn from populations with more defined values.

- Smaller per-group sample sizes tended to produce fewer between-group ties than larger per-group sample sizes.
- Samples drawn from the same population tended to produce between-group ties more often than samples drawn from populations separated by large effect sizes.
- For three or more groups, the middle most group(s) tended to be involved in between-group ties more often than the extreme groups.
- Multiple ties were common, i.e., ties involving more than one occurrence of a value within a group as well as ties involving more than one value in a group.

Two conclusions that can be drawn from the occurrence of ties study are:

(1) When using this group of tests with real data, of which the Micceri (1986, 1989) distributions are representative, effective methods of dealing with consequential ties are a necessity; and

(2) It is doubtful that these tests can be successfully applied to data from discrete populations with very few values, such as Micceri's (1986, 1989) Extreme Bi-modal Psychometric (EB) distribution.

Both points are supported by the results from Chapter 4 and are discussed further in the Chapter 5, section 4.

5.3 — Sampling Adequacy

Results of a sampling adequacy study are presented in Chapter 4, section 5. The results indicated that for the random number generator and initial seed used (MOTHER with 255,255), more observations were needed to ensure a sufficiently close approximation to the population distributions than originally expected. Although it was possible to get an idea of what the results would look like in as few as 1,000 trials, and short runs of 10,000 trials were used for program debugging and tryout prior to the full simulation runs, the results from this number of repetitions at the sample sizes being investigated were not sufficient to generate defensible answers to the primary research questions.

Simulations based on 1,000,000 trials offered an acceptable balance of good distribution fit, precision and stability of results, and execution time. Given the power of modern personal computers one should expect to see Monte Carlo results based on considerably larger numbers of trials than have historically appeared in the literature.

5.4 — Research Questions

The findings with respect to the original research questions are presented. Note that methods 1, 2 and 7 were investigated for all six tests, but results were not presented in Chapter 4. Methods 1 and 2 were only investigated in order to investigate method 3a. The Type I error results for method 1 were, as expected, very conservative while the results for method 2 were equally liberal (inflated). The Type I error results for method 7 were also extremely inflated. Methods 3a, 4, 5 and 8 were the only methods applied to all six tests, although each test included at least one of the method's {3a, 3b} and most included at least one of the method's {6a, 6b, 6c}, Tukey's Quick Test being the exception.

As described in Chapter 3, methods 7 and 8 were investigated in two different ways: 1) in parallel with methods 1 – 6 (as appropriate) for each test, and 2) through independent simulations. Although the parallel results were not reported in Chapter 4, comments regarding them follow.

As described in Chapter 3, the main simulations were designed to draw 1,000,000 sets of samples for each combination of simulation parameters and test them. Because methods 7 and 8 dropped tied values they often resulted in untestable samples, either because there was no data left or because critical values were unavailable, due either to small and/or unequal sample sizes. Untestable results were counted separately from significant and non-significant results, although the specific reasons that samples were untestable was not captured. The results of the main Type I error simulations revealed a very high percentage of untestable samples and grossly inflated Type I error rates, especially with the EB distribution. This was true even though Type I error rates were based on total samples. An adjustment to use testable samples as the denominator of the Type I error rate would have inflated the results further, as seen in the separate method 7/8 studies described next.

Separate method 7/8 simulations were also run for each test in which the program was allowed to cycle up to 10,000,000 times in an attempt to get 10,000 testable samples. These simulations almost always took more than 10,000 cycles to get 10,000 testable samples and under many combinations of simulation parameters took hundreds of thousands or millions of cycles to do so. Indeed, it was not uncommon that they failed to obtain 10,000 testable samples and sometimes obtained only a few (<10) or none. Type I error results from these simulations, however, were calculated based on the number of testable samples obtained to try and capture a more accurate picture of the true Type I error properties. These are the Type I error results presented in Chapter 4, along with information about the relationship of cycles to testable samples.

Both simulation approaches calculated and reported average per-group sample sizes for methods 7 and 8. These average sample sizes revealed the behavior of these methods in yet another way. As expected, average sample size was almost always smaller than the initial sample size, indicating that ties had occurred and been dropped. Interestingly, the average sample size seemed to approach an upper limit as initial sample size increased. For some tests, this limit was approached fairly quickly and was well below the initial sample size. Thus, it appears that most of the data was being thrown away as sample size increased.

In a sense these two simulation approaches yielded the same results from slightly different perspectives. Either way, methods 7 and 8 produced a problematic number of untestable samples and grossly inflated Type I error rates. The Type I error results for method 7 were so inflated that they were not even reported in Chapter 4. The Type I error results for method 8 were also inflated, but were presented in Chapter 4 in order to support certain conclusions in this chapter.

5.4.1 – Research Question 1

For samples drawn from the same population, is the Type I error rate maintained between .5α and 1.1α for each combination of test, method, number of groups, directionality, sample size and distribution?

The short answer is no, not for most combinations of simulation parameters with the Micceri (1986, 1989) distributions, and sometimes not even with the Normal distribution. This was because the theoretical probability of a Type I error, *P*(T1e), was sometimes outside the robust tolerance limits to begin with, albeit usually in a conservative direction due to the use of best conservative critical values. Results are detailed in Chapter 4, section 2 and described following.

Methods 3a, 4, 5, 6b and 8 were investigated for the Kolmogorov-Smirnov Test of General Differences for Two Groups. At nominal alpha .01, table 4.2.1-1 shows that about half of the initial sample sizes had *P*(T1e) $< 0.5\alpha$, both 1-sided and 2-sided (upper, lower and both tails). Results for the Normal distribution closely matched these theoretical probabilities. Method 5 was consistently close to these results across all Micceri (1986, 1989) distributions. Method 4 and method 3a exhibited slightly more conservative results across directions and tails, for distributions EA, ML and SS. Results of method 3a with distribution EB were grossly inflated while results of method 4 with distribution EB were very conservative. Method 6b yielded somewhat conservative results with distribution ML and very conservative results with distributions EA, EB and SS. Results for method 8 were grossly inflated across all Micceri (1986, 1989) distributions.

For the Kolmogorov-Smirnov Test at nominal alpha .05, Table 4.2.1-3 shows that about a fourth of the initial sample sizes had *P*(T1e) $< 0.5\alpha$, 1-sided and about a third of the initial sample sizes had *P*(T1e) $< 0.5\alpha$, 2-sided (upper, lower and both tails). Again, results for the Normal distribution closely matched these theoretical probabilities. Method 5 was consistently close to these results as well across all Micceri (1986, 1989) distributions. Results for method 4 with the ML distribution were very similar to those for method 5, but somewhat more conservative for distributions EA and SS and very conservative for distribution EB. Method 3a yielded results that were even more conservative for distributions EA, ML and SS, but grossly inflated for distribution EB. Method 6b produced somewhat conservative results with distribution ML and very conservative results with distributions EA, EB and SS. Results for method 8 were grossly inflated across all Micceri (1986, 1989) distributions.

Based solely on Type I error performance, method 5 would be the method of choice for the Kolmogorov-Smirnov Test. None of the methods seemed to work well with distribution EB, and method 8 should be avoided.

For Rosenbaum's Test for Two Groups, methods 3a, 3b, 4, 5, 6c and 8 were investigated. At nominal alpha .01, Table 4.2.2-1 shows that about 40% of the initial sample sizes had $P(\text{T1e}) < 0.5\alpha$, 1-sided and about a third had $P(\text{T1e}) < 0.5\alpha$, 2-sided (upper, lower and both tails). Results for the Normal distribution closely matched these theoretical probabilities, as did the results for method 5 across all Micceri (1986, 1989) distributions. Method 4 produced slightly inflated but acceptable results with distributions ML and SS but grossly inflated results with distributions EA and EB. Results for methods 3a and 3b were similar to each other and were slightly to somewhat conservative relative to $P(\text{T1e})$ for distributions ML, EA and SS but very conservative for distribution EB 2-sided and inflated for distribution EB 1-sided. Method 6c produced very conservative results for distributions EA and EB and somewhat conservative results for distributions ML and SS. Results for method 8 were grossly inflated across all Micceri (1986, 1989) distributions.

Table 4.2.2-3 shows the results for Rosenbaum's Test at nominal alpha .05. The pattern of out-of-tolerance counts is essentially the same as for nominal alpha .01 except that the results for methods 3a and 3b with distribution EB were grossly inflated for all directionalities and tails. Based solely on the Type I error performance, method 5 would be the method of choice for Rosenbaum's Test for both nominal alpha levels. Again, none of the methods seemed to work well with distribution EB and method 8 should be avoided.

Type I error results for Tukey's Quick Test of Location for Two Groups at nominal alpha .01 are given in Table 4.2.3-1. Methods 3a, 3b, 4, 5 and 8 were examined. *P*(T1e) was within the robust tolerance limits for all initial sample sizes, directionalities and tails. Results for the Normal distribution matched the theoretical probabilities for all methods, directions and tails except method 8, which was based on a maximum of 10,000 testable samples. Results for method 5 matched *P*(T1e) across all Micceri (1986, 1989) distributions and conditions. Method 3a gave very conservative results for distribution EA, somewhat conservative results for distributions ML and SS and grossly inflated results for distribution EB. Method 3b gave somewhat conservative results for distributions EA, ML and SS and grossly inflated results for distribution EB. Results for method 4 with distributions ML and SS were within the robust tolerance limits, but results for distributions EA were grossly inflated while with distribution EB results were outside the robust tolerance limits in both directions. Results for method 8 were grossly inflated across all Micceri (1986, 1989) distributions.

Results for nominal alpha .05 did not follow the same exact pattern as for nominal alpha .01. Again, all *P*(T1e) were within the robust tolerance limits, as were all of the results for the Normal distribution and for method 5 across all Micceri (1986, 1989) distributions. Method 3a produced very conservative results with distribution EA, grossly inflated results with distribution EB, within tolerance results with distribution ML (1-sided), somewhat conservative results with distribution ML (2-sided), slightly conservative results with distribution SS (1-sided) and somewhat conservative results with distribution SS (2-sided). Method 3b produced somewhat to grossly inflated results with distributions EA and EB, within tolerance to slightly out of tolerance results with distributions ML and SS (1-sided) and somewhat conservative results with distributions ML (2-sided) and SS (2-sided). Method 8 was grossly inflated across all Micceri (1986, 1989) distributions.

Based solely on the Type I error performance, method 5 would be the method of choice for Tukey's Quick Test for both nominal alpha levels. Distribution EB continued to be problematic for most methods and method 8 did not perform acceptably with any distribution and should be avoided.

P(T1e) for the Wilcoxon-Mann-Whitney Test was within the robust tolerance limits for all initial sample sizes for all combinations of test conditions at both nominal alpha .01 and .05 as shown in Tables 4.2.4-1 and 4.2.4-3. Methods 3a, 4, 5, 6a and 8 were studied. Results for the Normal distribution were also within the robust tolerance limits except for method 8 at nominal alpha .01, which was slightly inflated. All results for method 5 were within the robust tolerance limits as were all results for distribution ML. Results for method 8 were grossly inflated across the Micceri (1986, 1989) distributions at both alpha levels, as were the results for method 4 with distribution EA. All other results at nominal alpha .01 were within the robust tolerance limits except for methods 3a and 6a with distribution EB, which were somewhat conservative. At nominal alpha .05, the same general pattern continued with results across the Micceri (1986, 1989) distributions within the robust tolerance limits or slightly to somewhat conservative (methods 3a and 6a with distribution EB 2-sided).

Based solely on Type I error results, the method of choice for the Wilcoxon-Mann-Whitney Test at both nominal alpha levels would probably be method 5, although the evidence in favor of that choice is not as overwhelming as it is for the Kolmogorov-Smirnov Test, Rosenbaum's Test and Tukey's Quick Test. Note that all of the methods (except method 8) worked well with distribution ML, which has the most values of the four Micceri (1986, 1989) distributions. Method 8, however, should not be used.

P(T1e) was not available for the Kruskal-Wallis Test (against an omnibus alternative hypothesis) so results for the Normal distribution were taken as the standard for comparison with the Micceri (1986, 1989) distributions in determining the Type I error performance of methods 3a, 4, 5, 6a and 8. Tables 4.2.5-1 and 4.2.5-2 give the results for nominal alpha .01 and .05 respectively, each for three to six groups. Results for both the H statistic and the H_c statistic (continuity correction) are reported but all subsequent results are based on the H statistic as the H_c statistic tended to produce inflated results compared to the H statistic across nominal alpha level, number of groups, distribution, nominal effect size multiplier and method.

At nominal alpha .01 and .05, results for method 3a were grossly inflated across number of groups for distributions EA and EB, close for distribution ML and somewhat to considerably inflated for distribution SS. Results for method 4 at both alpha levels were grossly inflated for distribution EA, close for distributions EB and ML, and close to somewhat inflated for distribution SS (nominal alpha .01 at 3 and 6 groups). Results for method 5 at nominal alpha .01 were slightly to somewhat inflated, but matched the Normal results very closely for nominal alpha .05. At nominal alpha .01, method 6a was generally very close to the Normal results except with distribution EB, where results were quite conservative. At nominal alpha .05, however, results for method 6a across the Micceri (1986, 1989) distributions were quite close to those for the Normal distribution, most of them being within the robust tolerance limits or slightly conservative (distribution EB with 4 and 5 groups).

Based solely on Type I error performance, method 6a would probably be the method of choice for the Kruskal-Wallis Test across nominal alpha levels, although method 5 would be an equally reasonable choice at nominal alpha .05 and not a bad choice at nominal alpha .01. Results for method 8 were grossly inflated across all test conditions and it should not be used.

Type I error results for the Terpstra-Jonckheere Test (against an ordered alternative hypothesis) are given in Tables 4.2.6-1 and 4.2.6-2 for nominal alpha .01 and .05, respectively. Each table covers methods 3a, 4, 5, 6a and 8 for three to six groups. *P*(T1e) was available for many, but not all, initial sample sizes. In all cases where *P*(T1e) was available, it was within the robust tolerance limits except for three groups at nominal alpha .01, where one value was below the lower limit. Results for the Normal distribution matched the theoretical results very closely except for method 8 at nominal alpha .01. As with all of the preceding tests, results for method 8 were generally quite inflated across the Micceri (1986, 1989) distributions at both nominal alpha levels.

Results for method 5 matched the *P*(T1e) and Normal results across all Micceri (1986, 1989) distributions at both nominal alpha levels. At nominal alpha .01, methods 3a, 4 and 6a also showed good agreement with a tendency to be slightly conservative when deviating. The exception was method 4 with distribution EA, which was somewhat inflated. At nominal alpha .05, methods 3a, 4 and 6a also matched the *P*(T1e)/Normal results across distributions and number of groups quite closely. The exceptions were methods 3a and 6a with distributions EA and EB at three groups, which were slightly conservative, and method 4 with distribution EA at six groups, which was slightly inflated.

Based solely on Type I error performance, method 5 would probably be the method of choice for the Terpstra-Jonckheere Test based on the better results at nominal alpha .01. Although the choice is less compelling at nominal alpha .05, method 5 is the only method that matched the *P*(T1e)/Normal results across all Micceri (1986, 1989) distributions and number of groups. Method 8 should be avoided.

On the basis of Type I error performance alone, method 5 appears to be the method of choice for five of the six tests irrespective of the combination of the various test parameters. Since method 5 is also a viable choice for the Kruskal-Wallis Test, it would be the overall method of choice for this set of tests, based solely on Type I error performance, if a single method had to be chosen. It is clear from this study that the often-recommended method (8) of dropping ties and reducing N is not to be recommended for any of these tests under any circumstances.

5.4.2 – Research Question 2

For samples drawn from populations differing only in location, what is the power for each combination of test, method, number of groups, directionality, sample size and distribution?

This question should also include nominal alpha level and nominal effect size multiplier. Results are presented in Chapter 4, section 3, including Type III error results where appropriate. Results are only reported for those combinations of test parameters that exhibited acceptable Type I error performance.

5.4.3 – Research Question 3

For samples drawn from populations differing only in location, is there a preferred method of resolving tied ranks for each combination of test and distribution, irrespective of the number of groups, directionality and sample size?

This research question asks for results across number of groups, directionality and sample size and should have included nominal alpha and nominal effect size multiplier. Chapter 4, section 3 reports the rankings of the power results across methods by distribution for each combination of test, nominal alpha level, number of groups and directionality in three ways: 1) across nominal effect size multiplier at each initial sample size; 2) across initial sample size at each nominal effect size multiplier, and 3) across nominal effect size multiplier and initial sample size. The question then is, given the apparent superiority of method 5 with respect to Type I error performance, did the power results also support that choice?

Table 4.3.1-36 shows the results by distribution and method for the Kolmogorov-Smirnov Test of General Differences for Two Groups. The picture presented there suggests that method 5 is the most powerful with each of the three distributions investigated (EA, ML and SS) except when looking at initial sample size across nominal effect size multiplier for distribution EA, where method 3a performed equally well with method 5.

Table 4.3.2-36 shows the results by distribution and method for Rosenbaum's Test of Location for Two Groups. For distribution EA, only methods 3a, 3b and 5 were investigated. Results for methods 3a and 3b were identical, so they were tied for rank 1 twice as often as the number of first place finishes shown, i.e., if either method was eliminated the number of first place finishes for the other would double. Thus, they appear to have a slight edge over method 5 for distribution EA. It is clear, however, that for distributions ML and

SS, method 4 was the most powerful.

Table 4.3.3-36 shows the results by distribution and method for Tukey's Quick Test of Location for Two Groups. Methods 3a, 3b, 4 and 5 were investigated for distributions ML and SS. The picture presented there suggests that method 5 was consistently the most powerful.

Table 4.3.4-36 shows the results by distribution and method for the Wilcoxon-Mann-Whitney Test for Two Groups. Methods 3a, 4, 5 and 6a were investigated for all four Micceri distributions except for method 4 with distribution EA. The picture presented there suggests that method 5 was the most powerful with distributions EA and EB and equal to slightly better than method 4 for distribution SS. For distribution ML, however, method 4 was generally the most powerful method.

Tables 4.3.5-44 and 4.3.5-45 show the results by distribution and method for the Kruskal-Wallis Test. Table 4.3.5-44 presents results for nominal alpha and number of groups (.01, 4), (.01, 5), (.05, 3) and (.05, 4) as the combination of methods and distributions was the same for these conditions. Likewise, Table 4.3.5-45 presents results for nominal alpha and number of groups (.01, 6), (.05, 5) and (.05, 6). The same picture is presented in both tables. For distribution EA, method 5 is generally more powerful than method 6a. For distribution EB, method 6a seems to have the edge when the rankings are done on power to four decimals, but this advantage disappears somewhat when the ranking is done on power to two decimals, tending to favor method 5 instead. In those cases where method 3a was investigated with distributions ML and SS it was generally superior. In the absence of method 3a, however, method 5 regained the advantage or shared it with method 6a.

Table 4.3.6-42 shows the results by distribution and method for the Terpstra-Jonckheere Test. The table shows the number of first place finishes, based on mean ranks, for each combination of methods 3a, 4, 5 and 6 with distributions EA, EB, ML and SS across nominal alphas and number of groups. The picture presented in Table 4.3.6-42 shows that method 5 is the most powerful with distributions EA, ML and SS whether the rankings are based on power to four decimals or two decimals. At four decimals, the four methods appear to be somewhat similar in performance with distribution EB, but method 5 emerges as the method of choice when the rankings are based on power to two decimals.

When considering each combination of test and distribution across all other simulation parameters, a single method was not uniformly most powerful. Method 5, however, was generally the most powerful method except for some idiosyncratic departures.

5.4.4 – Research Question 4

For samples drawn from populations differing only in location, is there a preferred method of resolving tied ranks for each test, irrespective of the number of groups, directionality, sample size and distribution?

This research question should also include nominal alpha and nominal effect size multiplier. While the Type I error results and the discussion under Research Question 3 suggest that method 5 may be the preferred method for most tests, this research question asks for results across distributions. This is difficult due to: 1) the lack of equivalence of the actual effect size multipliers, and; 2) the fact that power results were not obtained for all combinations of test, distribution and method. Table 5.4-1 summarizes the issue.

Table 5.4-1
Actual Effect Size Multipliers by Distribution and Nominal Effect Size Multiplier and Ranks Across Distributions (within Nominal Effect Size Multiplier)

Distribution	*Nominal Effect Size Multiplier* .2	.5	.8	1.2	
Actual Effect Size Multipliers					
Norm	.2	.5	.8	1.2	
EA	.173	.518	.864	1.209	
EB	.n/a	.592	.n/a	1.183	
ML	.168	.504	.840	1.176	
SS	.204	.407	.815	1.222	
Ranks with Normal					*mean rank*
Norm	2	4	4	3	3.25
EA	3	2	1	2	2.00
EB		1		4	2.50
ML	4	3	2	5	3.50
SS	1	5	3	1	2.50
Ranks w/o Normal					*mean rank*
EA	2	2	1	2	1.75
EB		1		3	2.00
ML	3	3	2	4	3.00
SS	1	4	3	1	2.25

The first part of the table restates the actual effect size multipliers for each distribution for each nominal effect size multiplier. The middle part of the table gives the ranking of these actual effect size multipliers across distributions at each nominal effect size multiplier, including the Normal distribution. The last part of the table repeats the ranking, without the Normal distribution. The last part of the table must be kept clearly in mind when comparing power results across the Micceri (1986, 1989) distributions as these ranking define the expected order of results if all methods performed equally well across distributions at a given nominal effect size multiplier.

Based on the mean rank of power results across test conditions, the following results were obtained with respect to methods of choice across distributions:

- For the Kolmogorov-Smirnov Test of General Differences, the last part of Table 4.3.1-37 clearly indicates that method 5 is preferred.
- For Rosenbaum's Test, the last part of Table 4.3.2-37 clearly indicates that method 4 is preferred.
- For Tukey's Quick Test, the last part of Table 4.3.3-37 clearly indicates that method 5 is preferred.
- For the Wilcoxon-Mann-Whitney Test, the last part of Table 4.3.4-37 clearly indicates that method 5 is preferred, with method 4 as a viable second choice.
- For the Kruskal-Wallis Test, the last part of Tables 4.3.5-46 and 4.3.5-47 clearly indicate that method 5 is preferred, with method 6a as a viable second choice and method 3a acceptable under certain conditions.
- For the Terpstra-Jonckheere Test, the last part of Table 4.3.6-43 clearly indicates that method 5 is preferred.

Thus it appears that there is a preferred method of resolving tied observations for each test, and that method is generally method 5 (random) with the exception of Rosenbaum's Test, where method 4 appears to be favored.

5.4.5 – Additional Research Question

Is there a preferred method of resolving ties across all tests and conditions in this study?

On the basis of Type I error performance alone, method 5 would be the method of choice across all tests under all combinations of study parameters if a single method had to be selected. Further, the answers to research questions 2, 3 and 4 tend to support the use of method 5 with most of these tests under most combinations of simulation parameters. Thus, it is reasonable to conclude that if one had to choose a single method for resolving equal data values (tied ranks), method 5 (random resolution) would be the method of choice. It's also clear from this study, however, that the often-recommended method of dropping ties and reducing N, method 8, is not to be recommended for any of these tests under any circumstances.

5.5 — Discreteness

It is intuitively obvious that, all other things being equal, the more distinct values a population distribution has, the less likely ties are to occur in random samples, either within or between groups. The results of this study suggest that all of the methods studied work better with many-valued discrete distributions, such as the ML distribution, than they do with distributions that have very few distinct values, such as the EB distribution. The only empirical study known to the author to explicitly investigate the effect of the discreteness of a distribution on the performance of one of the tests in this study is Sparks (1967), who studied the Wilcoxon-Mann-Whitney Test using successively coarser approximations to the Normal distribution. The results of the current study support Sparks (1967) findings, i.e., that the Type I error performance of these tests with populations whose distributions violate the assumption of continuity gets progressively worse as the number of categories (distinct values) used to approximate a smooth distribution gets smaller. Thus, it appears that discreteness of population distributions is an important factor in the performance of these tests vis-à-vis consequential ties and ways of dealing with them.

Discreteness is also an issue with some of these tests in that the critical values are themselves from a discrete distribution. Indeed, the very conservative Type I error performance of some of the tests, such as the Kolmogorov-Smirnov Test, were directly attributable to the best-conservative critical values at the chosen nominal alpha levels. This is discussed more fully in section 5.7.

5.6 — Additional Limitations Of The Study

This research originated with Fahoome (1999). The decision to do power studies forced a narrowing of the focus to tests of location involving independent samples as the amount of work involved in doing power studies on multiple methods of resolving ties using five distributions at four different effect sizes, with up to four different numbers of groups, was considerable. The tests that were eliminated included single-sample tests of location as well as tests of association, correlation, spread and dependent-samples tests. The tests for spread were eliminated because they are not of much practical interest while the dependent-samples tests were eliminated because the problem of generating correlated data for such tests is not yet fully worked out. Thus, of the 21 statistics studied by Fahoome (1999), only six were retained for this research.

Sawilowsky (2002), in writing about the 'Behrens-Fisher problem', argued that the problem was of no practical importance as treatment effects in social and behavioral science rarely produce changes in spread without an accompanying change in location (see also Sawilowsky & Blair, 1992). He asserted that researchers should be concerned with tests for shifts in location accompanied by changes in spread. Such effects are not represented in the present research, although Rosenbaum's Test of Location is particularly appropriate for situations in which an increase in the median is accompanied by an increase in spread. Thus, tests for pure changes in scale were not included in this study. The median test was also excluded as Freidlin & Gastwirth (2000) showed that it lacked power in comparison to the Wilcoxon-Mann-Whitney Test.

One implication of Sawilowsky (2002) was that changes in location without accompanying changes in spread seemed equally unlikely. To the extent that this is true, the pure shift methodology of the present study was limiting. Bradley (1978) made a strong case that tests for single effects (such as the shift in a population mean) that were based on 'all else equal' assumptions, such as homogeneity of variance across groups, may not match reality very well, the real interest being in the complex interactions of various factors. Sawilowsky & Blair (1992) also established the importance of comparative power studies between tests. The current research, although complex, does not directly address these concerns.

This research was limited by lack of another kind of comparability as well. When dealing with discrete distributions that are integer valued, such as the Micceri (1986, 1989) distributions, it makes sense to deal with integer amounts of shift so that the shifted population continues to reflect the same properties as the un-shifted population, especially as regards being integer valued. In this research, however, it was absolutely necessary to shift the Micceri (1986, 1989) distributions by integer amounts in order to ensure the occurrence of tied values. As a consequence, at each nominal effect size multiplier {0.2, 0.5, 0.8, 1.2}, the actual effect size multiplier for each Micceri (1986, 1989) distribution turned out to be different from the nominal value and from the values for the other distributions. These characteristics are summarized in Table 5.4-1. The lack of effect size multiplier equality made it difficult to draw a solid conclusion about the methods across distributions as called for in Research Question 4.

Bradley's (1968) methods A and D call for the use of actual probability values. Method A obtains probability bounds, α_M and α_L, by resolving tied ranks in the manner most likely to lead to rejection of the null hypothesis and then in the manner least likely to lead to rejection of the null hypothesis. Method D resolves ties in all possible ways, obtaining the probability of the resulting statistic for each resolution and then constructing the cumulative probability for use in the next step, where the mean, median, or mid-range probability were suggested as the probability of the test. However, to calculate the mid-range probability, (α_M +

α_L) / 2, only two resolutions are needed, the most and least favorable, as described above. These methods, based on combining probabilities, were not investigated directly, in part because it appeared at the start of this research that obtaining theoretical probabilities for the critical values for most of the tests would be very difficult. Instead, ties were resolved in the manner least and most likely to lead to rejection of the null hypothesis and the mid-range value of the statistic was computed as the mean of these two values and used for the test. Averaging the values of the statistics in this way is almost certainly not equivalent to Bradley's (1968) methods.

As discussed previously in section 5.5, Micceri's (1986, 1989) extreme bi-modal distribution with its very short scale was problematic. This is supposed to represent Likert-scale data such as encountered with surveys. In practice, however, it doesn't make any sense to take a survey with Likert-scale items that range from 0 to 7 and say that the expected effect of a treatment will be to shift the responses to range from 3 to 10. Indeed, one would expect the same survey to be re-administered after treatment such that the items would still range from 0 to 7. If the intended effect of the treatment were to shift scores towards higher values, one would expect a piling up of scores at the high end, with a left skew to the distribution of scores.

The same issue applies to any distribution that represents scores on an instrument with a fixed range, whether 0 to 100 for a typical classroom assessment or 400 to 1600 on the composite S.A.T. (Scholastic Aptitude Test, Educational Testing Service). Many other effects can be hypothesized. Pure shift effects do not necessarily represent reality well with short scale, discrete distributions. They also suffer from the problem that substantial effect size multipliers serve to reduce the overlap between the two population distributions, thus reducing the opportunity to get tied observations. The limitations imposed by using the Micceri (1986, 1989) distributions in a pure shift model are balanced by the fact that they resulted from a meta-analysis of empirical data.

In an effort to reveal differences in the rankings of the power results that were of practical significance, the results were rounded to three and two decimal places prior to ranking. However, since rounding has the effect of creating fewer categories, it can certainly increase the apparent difference between adjacent values as much as it can reduce it. For example, consider results to four decimal places of {.5010, .5049, .5051, .5092} with ranks {4, 3, 2, 1}. When rounded to two decimal places and then ranked one gets {.50, .50, .51, .51} with ranks {3.5, 3.5, 1.5, 1.5} although clearly a more reasonable ranking would be {4, 2.5, 2.5, 1} given the closeness of agreement of the middle two values and their distance from the other two. This is in fact what happens when they are rounded to three decimal places, but other examples could easily be given for which two decimal places gives a clearer picture than three. Thus, a better way than rounding might be used to determine closeness of agreement prior to ranking the results.

Ancillary to the main study, critical values and associated probabilities were generated for the Kolmogorov-Smirnov Test, Rosenbaum's Test, Tukey's Quick Test and the Wilcoxon-Mann-Whitney Test as described in Fay (2002). In this work, best conservative values were generated, checked against published tables (Neave, 1981; Neave & Worthington, 1988) and found to be in extremely high agreement. The major advantages of generated critical values included: 1) convenience, 2) availability of values for larger sample sizes, 3) availability of values for a much greater combination of unequal sample sizes and 4) availability of the mathematical probability of a Type 1 error for each pair of initial sample sizes. In particular, the ability to generate critical values for a greater range of unequal sample sizes expanded the testability of the 'drop ties and reduce N' methods.

The availability of *P*(T1e) revealed that in some cases the best conservative critical values were extremely conservative, with probabilities that fell below the lower limit of robustness chosen for this study, while the next highest value had a probability that was greater than nominal alpha but below the upper robust limit, and therefore much closer to nominal alpha. Indeed, there were many cases in which the best conservative critical value had a probability that was within the robust limits but further away from nominal alpha than the next available one whose probability just slightly exceeded nominal alpha and was well within the upper robust limit. Thus, what was presumed to be generous limits on Type I error robustness of 0.5α to 1.1α turned out not to be for some tests, even in the absence of ties.

An unsuccessful attempt was made to generate critical values with associated probabilities for the Kruskal-Wallis and Terpstra-Jonckheere Tests so existing tabled critical values were used instead. This was due primarily to published articles that were not quite detailed enough relative to the author's mathematical knowledge. Best conservative critical values were available for equal sample sizes from 2 to 25 for 3, 4, 5 and 6 groups except for 3 groups at sample size 2 but Type I error probabilities were not available. Critical values were also available for some unequal sample size combinations (Neave, 1981).

Best conservative critical values for the Terpstra-Jonckheere Test were available for equal sample sizes from 2 to 10 for 3, 4, 5 and 6 groups and Type I error probabilities were available for some sample size combinations. Critical values were also available for some unequal sample size combinations (Neave & Worthington, 1988). In general, the lack of available critical values for unequal sample sizes probably had an adverse effect on the performance of the “drop and reduce N” methods for these two tests, limiting the reduced samples that could be tested to those for which critical values were available.

5.7 — Topics For Further Research

Given the ability to generate critical values and their associated probabilities, Bradley's (1968) methods based on probabilities should be investigated. In addition, the relative performance of averaging probabilities versus averaging statistics should be explored.

Further research into population distributions in the social and behavioral sciences, in particular into the effects that actually occur in different research settings, is needed. For one, Micceri's (1986, 1989) empirical distributions could be extended to provide more realistic sampling for power studies. The smooth symmetric distribution currently ranges from 0 to 26. Instead of being shifted to range from 4 to 30 to simulate a pure shift in location, it could be modified in several ways. One is to shift the median up while reducing the standard deviation in order to maintain the shape and still stay within the same range of values. Another would be to shift the median up and change the shape to be more left skewed.

Another direction of inquiry would be to develop a set of distributions like the EA, ML and SS, that all have the same standard deviations, this being the key issue in obtaining equal effect sizes. Starting from the Micceri (1986) distributions, the granularity and range would undoubtedly need to be adjusted along with the specific frequencies. Yet another possibility would be to fit mathematically continuous curves to the Micceri (1986,1989) distributions to establish ideal behavior (no ties) as a basis for comparison with the results obtained from the Micceri (1986, 1989) models.

Many other possibilities can be hypothesized. The problem is that any attempt to modify the Micceri (1986, 1989) distributions through other than empirical means immediately takes the researcher out of the realm of working with empirically derived pseudo-population models and into the make-believe world of invented models. This leaves open the possibility of creating population models, intentionally or otherwise, that could lead to, support or favor particular outcomes. Perhaps what is needed is replication of Micceri's (1986) method with an extension to pre- and post-treatment models that reflect what really happens as a result of various treatments or interventions in various settings.

Although not employed directly in this research, the ability to generate critical values allows them to be obtained for any alpha level and according to selection rules other than best conservative. For instance, a best fit criteria could be defined as 'closest to nominal α but not greater than 1.1α' as discussed in Fay (2002). Note that this best fit criterion does NOT produce the same critical values as a best conservative rule with a slightly higher nominal alpha.

By way of illustration, consider nominal alpha .05 and two potential critical values with probabilities .0495 and .0510. Under the best fit rule given above the value with probability .0495 would be selected as the critical value, just as in the best conservative case, whereas a best conservative rule with nominal alpha .055 would select the value with probability .0510 even though it is further from the desired nominal level. Now consider potential critical values with probabilities of .0249 and .0526. At nominal alpha .05, a best conservative rule would result in the choice of the value with probability .0249 whereas the best fit rule would select the value with probability .0526. In this case the later value is clearly a better choice in that it is much closer to nominal alpha and within the upper limit of robustness. It appears that some tests with discrete valued distributions of their statistics might benefit in terms of their size through the use of critical value selection criteria other than best conservative.

While the range of initial sample sizes for which critical values could be generated for each test was often greater than that available from existing tables, it was not unlimited. Some of the generating programs inevitably ran into computer word length limitations and stopped producing accurate critical values due to overflow. Such breakdowns in the procedures generally occurred beyond the maximum sample sizes planned for this study, and so did not pose a direct limitation. The most direct solution to this problem is probably a larger word length computer. In some cases, more sophisticated algorithms might extend the range of values. Recursive relations, when they exist, are especially useful. Most of these tests have large-sample approximation formulas (Fahoome, 1999). Given that these formulas have been investigated carefully, the lack of available critical values at larger sample sizes should not be a serious limitation.

As discussed in Bradley (1978), a test that is robust with respect to Type I error under a violation of only one of its assumptions may not be robust under simultaneous violation of two or more assumptions. Given that real populations almost certainly violate more than one of the limited assumptions made by nonparametric / distribution-free tests, additional studies are needed.

Indeed, robustness itself remains an issue. Serlin (2000) identified the two types of errors that can be made in Monte Carlo studies as: "(a) concluding that a statistical procedure is robust when it is not or (b) concluding that it is not when it is." (p. 230). He also noted that "In previous attempts to apply standard statistical design principles to Monte Carlo studies, the less severe of these errors has been wrongly designated the Type I error." (p. 230). He stated that "...concluding that a methodology is robust when it is not is the more serious of the errors. The null hypothesis in a Monte Carlo study should state that the methodology under examination is not robust, and the alternative hypothesis should state that the methodology is robust." (p. 232). Serlin (2000) described the use of a range null hypothesis to test for robustness, and gave methods to calculate the power of such a test as well as the number of iterations required to detect robustness given specific criteria for a desired Type I error rate and level of power. One of the conclusions of his approach is that the number of iterations in a Monte Carlo study can become too large and overpower the test of significance. Serlin's (2000) recommendations regarding the optimal number of iterations suggest much smaller numbers than used in the present study. Although robustness was not being determined by means of significance tests in the present study, this finding seems to be at odds with the number of iterations needed to obtain acceptable fit of the sample distribution to the population distribution.

Although some work has been done in the generation of correlated data for dependent-samples tests (Headrick & Sawilowsky, 1999b) additional research is needed. Tests of location for repeated-measures designs are widely used and therefore of great practical importance. The present research needs to be extended to available distribution-free and/or nonparametric dependent-samples tests using realistic and representative data.

Finally, an entirely different approach to dealing with equal data values (tied ranks) in nonparametric statistical tests may be found in Rayner and Best (2001). They stated that "The prime objective of this book is to provide a *unification* and *extension* of some popular nonparametric statistical tests by linking them to tests based on models for data that can be presented in contingency tables." (p. vii) (emphasis in original). Their approach works with nominal and ordinal data, and encompasses such well known nonparametric procedures as the "sign, median, Wilcoxon, Kruskal-Wallis, Page, Friedman, Durbin, Cochran, Spearman and Kendall's tau tests." (p. vii). Ties are a concern in all of these tests, some of which were part of the present study. However, they go on to say:

> For almost all of these tests, and for some other important tests also, we are able to present the data in contingency tables. We can then calculate a Pearson-type χ^2 statistic, and its *components*. In the case of univariate data, the initial tests based on these components detect mean differences between treatments, and in the case of bivariate data, they detect correlations. The later components provide tests that detect variance, skewness and higher moment differences between treatments with univariate data, and higher bivariate moment differences with bivariate data. This approach provides a *unification* of much popular nonparametric statistical inference, and makes the traditional, most commonly performed nonparametric analyses much *more complete and informative*. In addition, the contingency table approach means tied data are easily handled, and almost exact Monte Carlo p-values can be obtained. Modern statistical packages such as StatXact (1995) calculate p-values this way. (p. vii)

Thus, they deal with ties by modeling them using scoring categories for data presented in contingency tables and partitioning a Person-type χ^2 statistic. They conclude by noting that "...many common nonparametric tests can be obtained as low order components of Pearson's χ^2_P statistic. ... The subsequent higher order components give potentially very useful extensions that define new nonparametric tests." (Rayner & Best, 2001, p. 196).

5.8 — Final Remarks

This study examined various methods of resolving equal data values (tied ranks) in a set of nonparametric and/or distribution-free statistical tests of location or general difference for *k* independent samples using Monte Carlo simulations with normal and discrete, nonnormal data. These tests were all based on the assumption of continuity in the underlying population. As such, the presence of ties, which occurred frequently with the discrete, non-normal populations, was problematic. The pattern and occurrence of ties was studied separately and quantified.

Of the methods investigated for resolving ties, random resolution seemed to work best for the majority of combinations of simulation parameters. The often-recommended method of dropping tied values and reducing the sample size performed very poorly across all combinations of simulation parameters and should not be used. All of these tests also performed poorly with the Micceri Extreme Bi-modal distribution and probably should not be used with discrete population distributions that contain relatively few distinct values.

Also investigated were the number of observations that needed to be drawn from the Micceri distributions in order to obtain a good fit between the sample distribution and the population distribution. A total of 6,000,000 observations seemed to be a reasonable lower limit. This translated into 1,000,000 iterations for two groups with initial samples sizes of three each.

Finally, methods were implemented for generating critical values and associated probabilities for four of the tests. For some tests the probabilities associated with the best conservative critical values were so conservative that alternative strategies, such as best fit, should be investigated. The ability to generate critical values and associated probabilities makes this possible.

APPENDIX A

HIC Approval

A Behavioral Protocol Summary Form for this study was submitted to the Wayne State University Human Investigation Committee on February 13, 2001. Notice Of Expedited Approval was received for the period March 9, 2001 through March 8, 2002 and a copy is included. Notice Of Expedited Continuation Approval was received for the periods February 18, 2002 through February 17, 2003 and January 14, 2003 through January 13, 2004, copies of which are also included.

WAYNE STATE UNIVERSITY

HUMAN INVESTIGATION COMMITTEE
4201 St. Antoine Blvd. , UHC 6-G
Detroit, Michigan 48201
Phone: (313) 577-1628
Fax: (313) 993-7122
HIC Website: www.orsps.wayne.edu

NOTICE OF EXPEDITED APPROVAL

TO: **Bruce R. Fay**
Evaluation & Research
3W, EDU

FROM: Peter A. Lichtenberg, Ph.D. Peter A Lichtenberg
Chairman, Behavioral Institutional Review Board (B03)

DATE: March 9, 2001

RE: Protocol # 03-33-01(B03)-ER "A Monte Carlo Computer Simulation Study of the Power Properties of 14 Nonparametric/Distribution-free Statistics under Six Different Methods of Resolving Tied Ranks when Applied to Normal and Non-normal Data Distributions" No funding requested

The above-referenced Protocol was **APPROVED** following Expedited Review (Category 5*) by the Chairman for the Wayne State University Institutional Review Board (B03) for the period of **March 9, 2001 through March 8, 2002.**

EXPIRATION DATE: March 8, 2002

This approval does not replace any departmental or other approvals that may be required.

Federal regulations require that all research be reviewed at least annually. **It is the Principal Investigator's responsibility to obtain review and continued approval before the expiration date.** You may not continue any research activity beyond the expiration date without HIC approval.

- If you wish to have your protocol approved for continuation after the above approval period, please submit a completed Continuation Form at least six weeks before the expiration date. It may take up to six weeks from the time of submission to the time of approval to process your continuation request.
  Failure to receive approval for continuation before the expiration date will result in the automatic suspension of the approval of this protocol on the expiration date. Information collected following suspension is unapproved research and can never be reported or published as research data.
- If you do not wish continued approval, please submit a completed Closure Form when the study is terminated.

All changes or amendments to your protocol or consent form require review and approval by the Human Investigation Committee (HIC) **BEFORE** implementation.

You are also required to submit a written description of any adverse reactions or unexpected events on the appropriate form (Adverse Reaction and Unexpected Event Form) within the specified time frame (see HIC policy).

- **Based on the Expedited Review List, revised November, 1998**
- C: Shlomo Sawilowski, Ph.D., 351 EDU

Revised 1/99

WAYNE STATE UNIVERSITY

HUMAN INVESTIGATION COMMITTEE
4201 ST ANTOINE BOULEVARD - UHC-6G,
DETROIT MICHIGAN 48201
PHONE: (313) 577-1628
FAX: (313) 993-7122
HIC WEBSITE: WWW.HIC.WAYNE.EDU

NOTICE OF EXPEDITED CONTINUATION APPROVAL

TO: Bruce Robert Fay
Education, Evaluation & Research
30580 Springland St.
Farmington Hills, MI 48334

FROM: Peter A. Lichtenberg, Ph.D. Peter A. Lichtenberg
Chairman, Behavioral Institutional Review Board (B03)

DATE: February 18, 2002

RE: Re-review of Protocol #: 03-33-01(B03)-ER "A Monte Carlo Computer Simulation Study of the Power Properties of 14 Nonparametric/Distribution-free Statistics Under Six Different Methods of Resolving Tied Ranks When Applied to Normal and Non-normal Data Distributions" No Funding Requested

The above-referenced protocol and Continuation Form, submitted on January 23, 2002, were **APPROVED** following Expedited Review by the Chairman of the Wayne State University Institutional Review Board (B03) for the period of **February 18, 2002 through February 17, 2003.**

EXPIRATION DATE: February 17, 2003

This approval does not replace any departmental or other approvals that may be required.

Federal regulations require that all research be reviewed at least annually. **It is the Principal Investigator's responsibility to obtain review and continued approval before the expiration date.** You may not continue any research activity beyond the expiration date without HIC approval.

- If you wish to have your protocol approved for another year, please submit a completed Continuation Form at least six weeks before the expiration date. It may take up to six weeks from the time of submission to the time of approval to process your continuation request.
  Failure to receive approval for continuation before the expiration date will result in the automatic suspension of the approval of this protocol on the expiration date. Information collected following suspension is unapproved research and can never be reported or published as research data.

- If you do not wish continued approval, please submit a completed Closure Form when the study is terminated.

All changes or amendments to your protocol or consent form require review and approval by the Human Investigation Committee (HIC) **BEFORE** implementation.

You are also required to submit a written description of any adverse reactions or unexpected events on the appropriate form (Adverse Reaction and Unexpected Event Form) within the specified time frame.

WAYNE STATE UNIVERSITY

HUMAN INVESTIGATION COMMITTEE
4201 St Antoine Boulevard - UHC-6G,
Detroit Michigan 48201
Phone: (313) 577-1628
FAX: (313) 993-7122
HIC website: www.hic.wayne.edu

NOTICE OF EXPEDITED CONTINUATION APPROVAL

TO: Bruce Robert Fay
30580 Springland
Farmington Hills, MI 48334

FROM: Ellen Barton, Ph.D. E Barton
Vice-Chair, Behavioral Institutional Review Board (B3)

DATE: January 14, 2003

RE: **NEW HIC#:** 033301B3F
Study Title: A Monte Carlo Computer Simulation Study of the Power Properties of 14 Nonparametric/Distribution-free Statistics Under Six Different Methods of Resolving Tied Ranks When Applied to Normal and Non-normal Data Distributions.
Sponsor: No Funding Requested

The above-referenced protocol and Continuation Form, submitted on January 5, 2003 were **APPROVED** following Expedited Review by the Vice Chair of the Wayne State University Institutional Review Board (B03) for the period of **January 14, 2003 through January 13, 2004.**

MARK YOUR CALENDAR!
Deadline for Re-review: Monday, December 1, 2003
To be reviewed by the Chair or his/her designee and reported to the December convened B3 meeting

This approval does not replace any departmental or other approvals that may be required.

Federal regulations require that all research be reviewed at least annually. **It is the Principal Investigator's responsibility to obtain review and continued approval before the expiration date.** You may not continue any research activity beyond the expiration date without HIC approval.

- If you wish to have your protocol approved for another year, please submit a completed Continuation Form at least six weeks before the expiration date. It may take up to six weeks from the time of submission to the time of approval to process your continuation request.
Failure to receive approval for continuation before the expiration date will result in the automatic suspension of the approval of this protocol on the expiration date. Information collected following suspension is unapproved research and can never be reported or published as research data.

- If you do not wish continued approval, please submit a completed Closure Form when the study is terminated.

All changes or amendments to your protocol or consent form require review and approval by the Human Investigation *Committee (HIC)* **BEFORE** *implementation.*

You are also required to submit a written description of any adverse reactions or unexpected events on the appropriate form (Adverse Reaction and Unexpected Event Form) within the specified time frame.

APPENDIX B

MICCERI (1986) DISTRIBUTIONS

Tables B-1 through B-4 provide scores, frequencies, cumulative frequencies and cumulative (probability) distribution functions for each of the four Micceri (1986) distributions used in this study. Figures B-1 through B-8 provide, in pairs, the corresponding graphs of the frequency distribution and cumulative (probability) distribution functions.

Table B – 1
Micceri Smooth Symmetric Data Set

Score	freq	cf	cdf
0	11	11	0.002047
1	21	32	0.005953
2	27	59	0.010977
3	58	117	0.021767
4	100	217	0.040372
5	115	332	0.061767
6	165	497	0.092465
7	206	703	0.130791
8	252	955	0.177674
9	278	1233	0.229395
10	348	1581	0.294140
11	384	1965	0.365581
12	419	2384	0.443535
13	443	2827	0.525953
14	439	3266	0.607628
15	391	3657	0.680372
16	372	4029	0.749581
17	289	4318	0.803349
18	264	4582	0.852465
19	216	4798	0.892651
20	189	4987	0.927814
21	143	5130	0.954419
22	93	5223	0.971721
23	66	5289	0.984000
24	39	5328	0.991256
25	30	5358	0.996837
26	17	5375	1.000000

Figure B – 1
Micceri Smooth Symmetric Data Set
Frequency Distribution

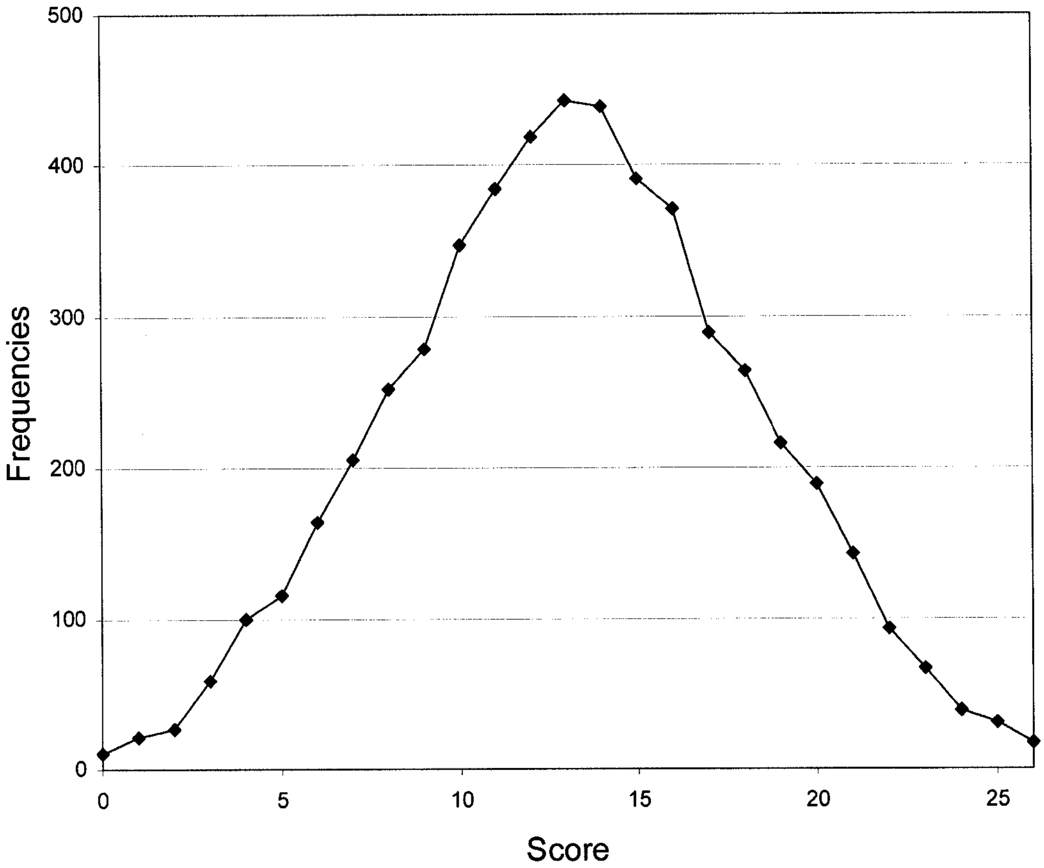


Figure B – 2
Micceri Smooth Symmetric Data Set
Cummulative Distribution Function

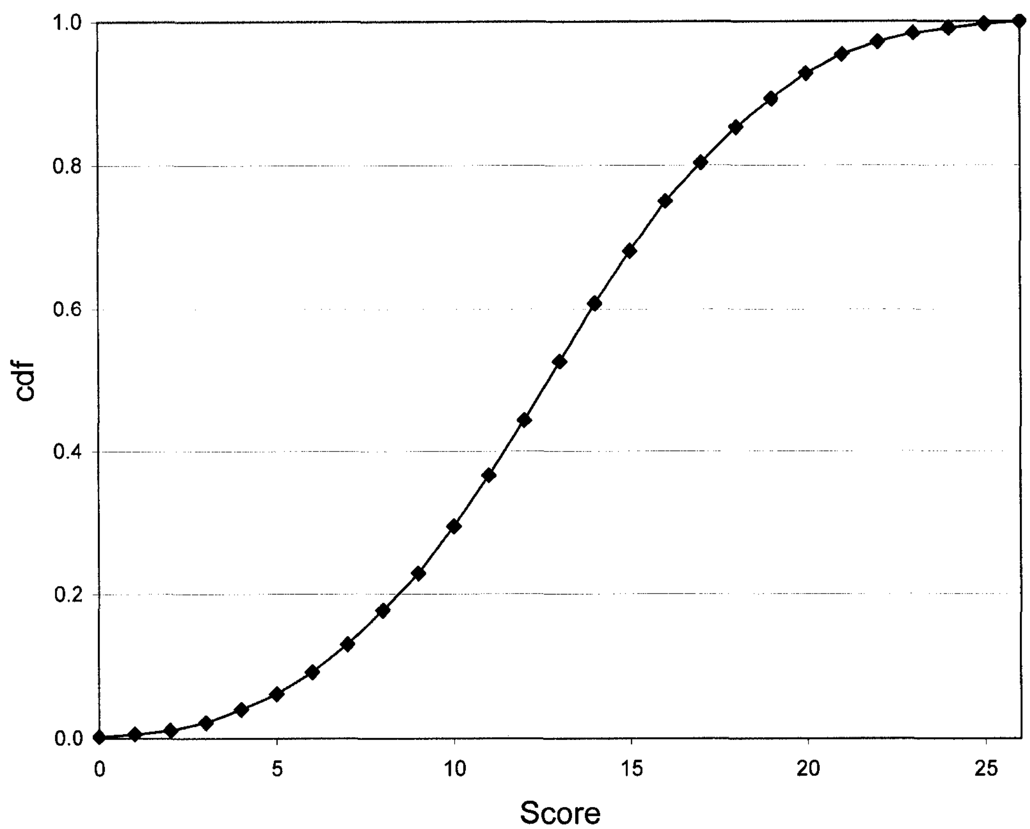


Table B – 2
Micceri Extreme Asymmetric Data Set (Achievement)

Score	freq	cf	cdf
0	0	0	0.000000
1	0	0	0.000000
2	0	0	0.000000
3	0	0	0.000000
4	6	6	0.002168
5	10	16	0.005780
6	9	25	0.009032
7	12	37	0.013367
8	18	55	0.019870
9	25	80	0.028902
10	19	99	0.035766
11	26	125	0.045159
12	37	162	0.058526
13	34	196	0.070809
14	39	235	0.084899
15	46	281	0.101517
16	43	324	0.117052
17	61	385	0.139090
18	48	433	0.156431
19	64	497	0.179552
20	80	577	0.208454
21	87	664	0.239884
22	95	759	0.274205
23	105	864	0.312139
24	144	1008	0.364162
25	153	1161	0.419436
26	198	1359	0.490968
27	261	1620	0.585260
28	297	1917	0.692558
29	430	2347	0.847905
30	421	2768	1.000000

Figure B – 3
Micceri Extreme Asymmetric Data Set (Achievement)
Frequency Distribution

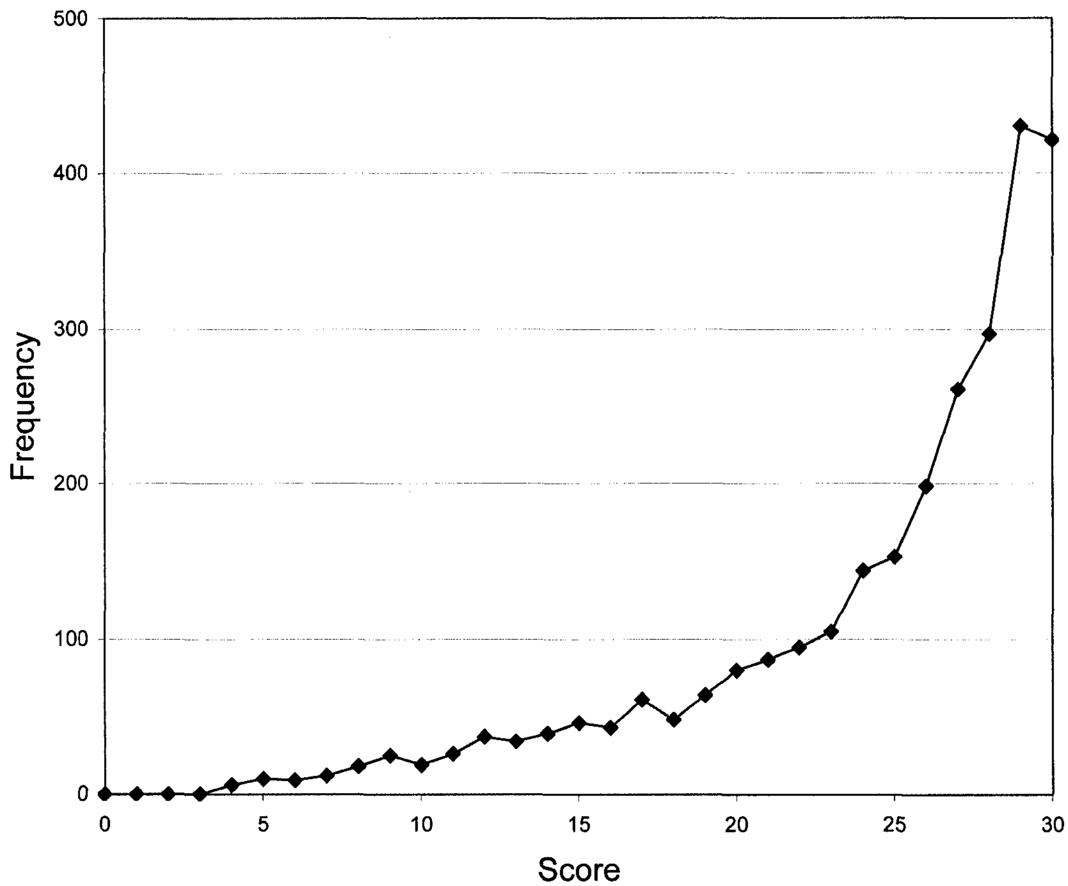


Figure B – 4

Micceri Extreme Asymmetric Data Set (Achievement)
Cummulative Distribution Function

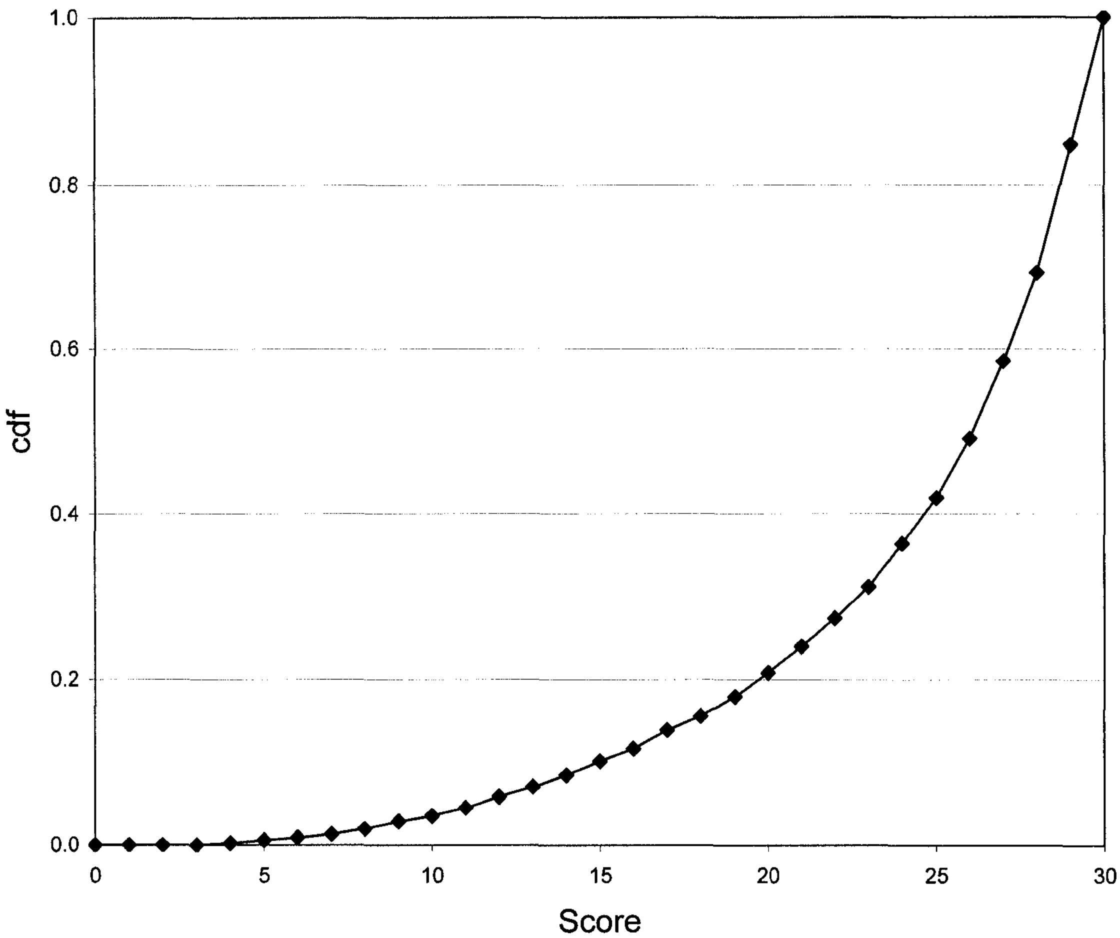


Table B – 3
Micceri Extreme Bimodal Data Set (Psychometric)

Score	freq	cf	cdf
0	3	3	0.004511
1	233	236	0.354887
2	58	294	0.442105
3	37	331	0.497744
4	154	485	0.729323
5	180	665	1.000000
6	0	665	1.000000
7	0	665	1.000000

Figure B – 5
Micceri Extreme Bimodal Data Set (Psychometric)
Frequency Distribution

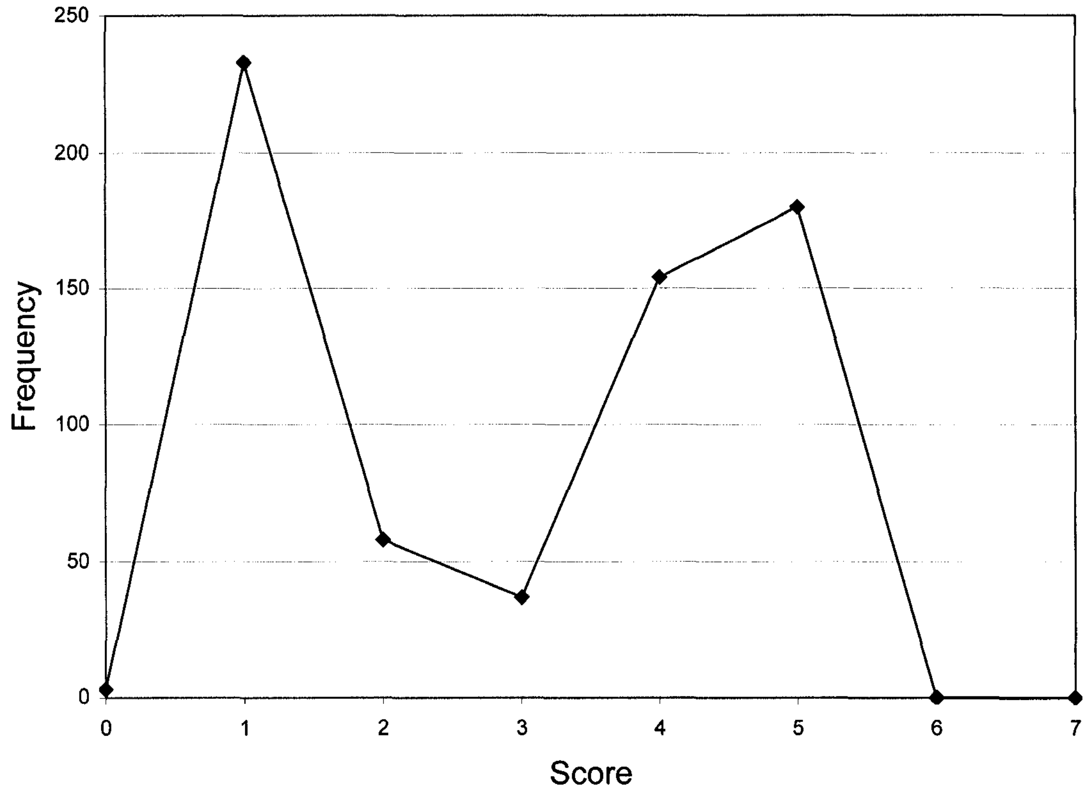


Figure B – 6
Micceri Extreme Bimodal Data Set (Psychometric)
Cummulative Distribution Function

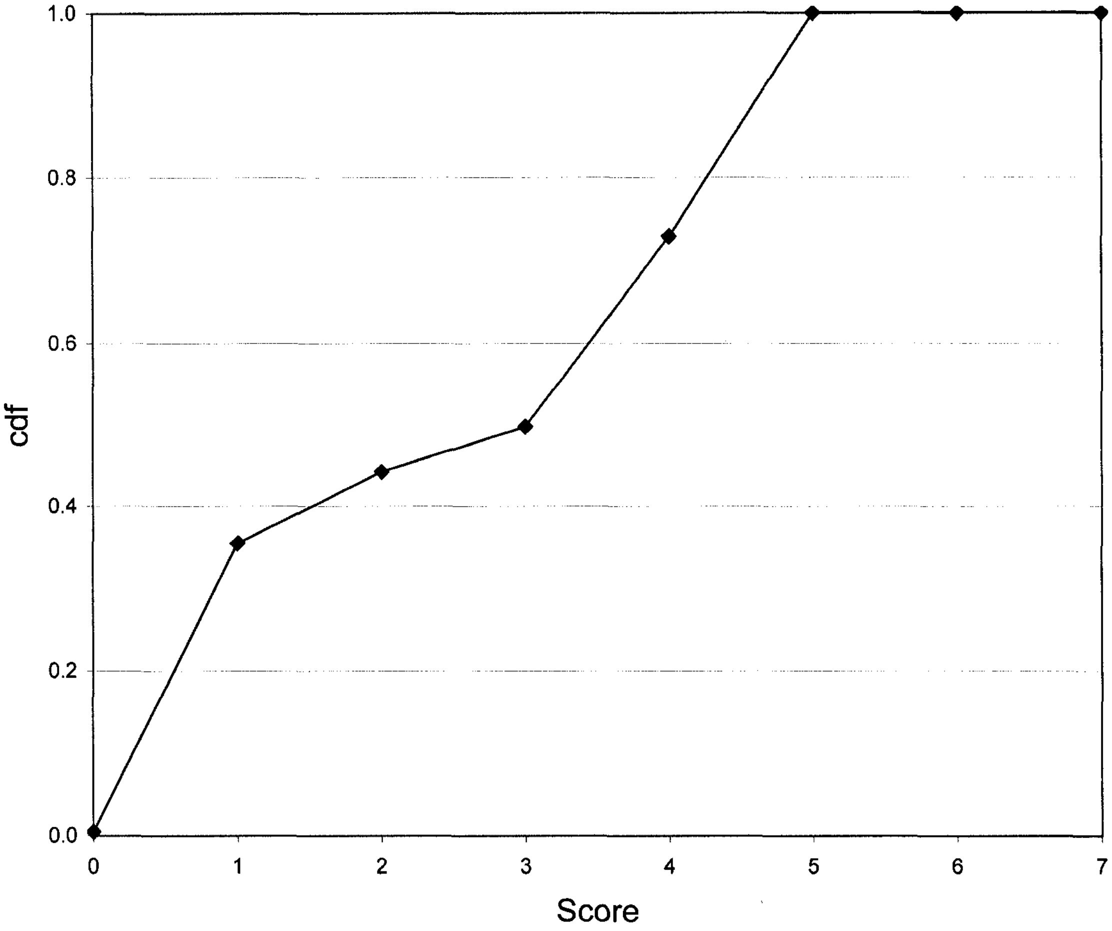


Table B – 4
Micceri Multimodal Lumpy Data Set (Achievement)

Score	freq	cf	cdf
0	5	5	0.010707
1	8	13	0.027837
2	8	21	0.044968
3	3	24	0.051392
4	8	32	0.068522
5	6	38	0.081370
6	3	41	0.087794
7	9	50	0.107066
8	12	62	0.132762
9	18	80	0.171306
10	11	91	0.194861
11	23	114	0.244111
12	22	136	0.291221
13	24	160	0.342612
14	20	180	0.385439
15	15	195	0.417559
16	18	213	0.456103
17	12	225	0.481799
18	9	234	0.501071
19	10	244	0.522484
20	10	254	0.543897
21	7	261	0.558887
22	8	269	0.576017
23	10	279	0.597430
24	3	282	0.603854
25	5	287	0.614561
26	10	297	0.635974
27	9	306	0.655246
28	3	309	0.661670
29	10	319	0.683084
30	6	325	0.695931
31	11	336	0.719486
32	15	351	0.751606
33	13	364	0.779443
34	15	379	0.811563
35	10	389	0.832976
36	12	401	0.858672
37	17	418	0.895075
38	10	428	0.916488
39	6	434	0.929336
40	11	445	0.952891
41	9	454	0.972163
42	6	460	0.985011
43	7	467	1.000000

Figure B – 7
Micceri Multimodal Lumpy Data Set (Achievement)
Frequency Distribution

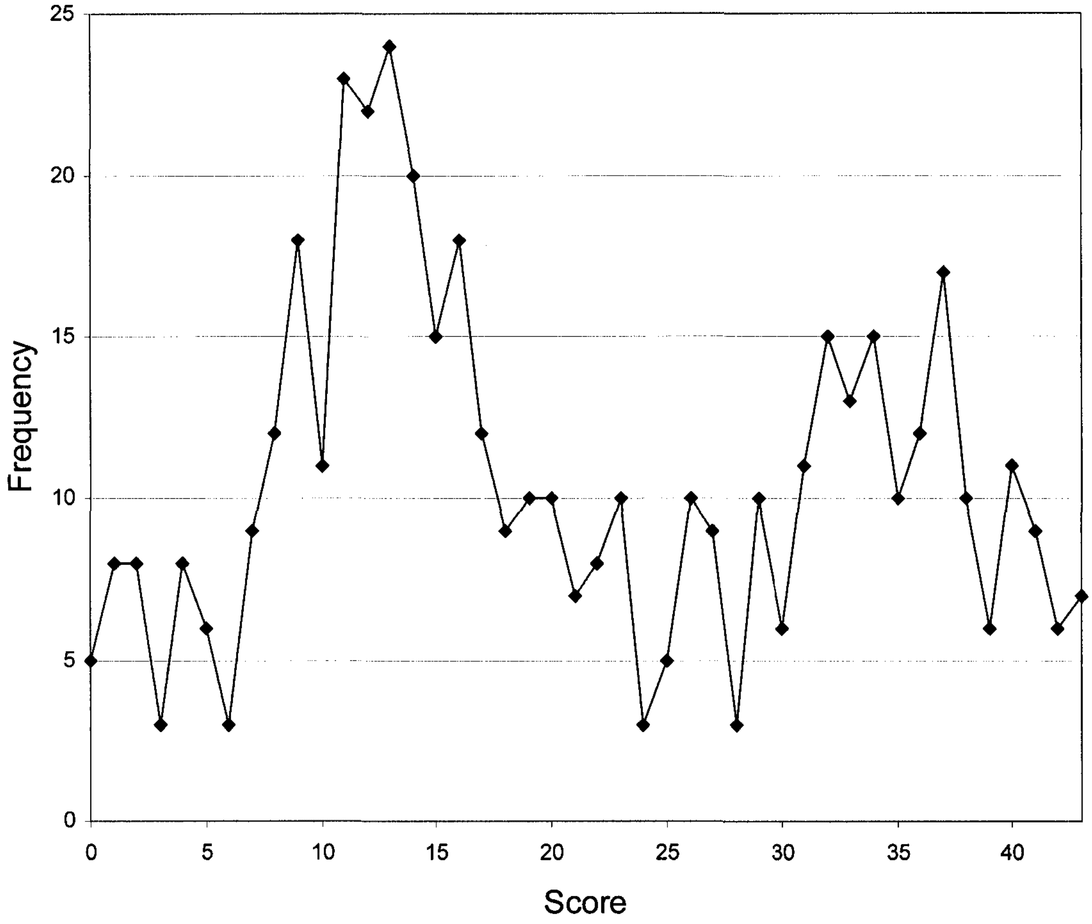
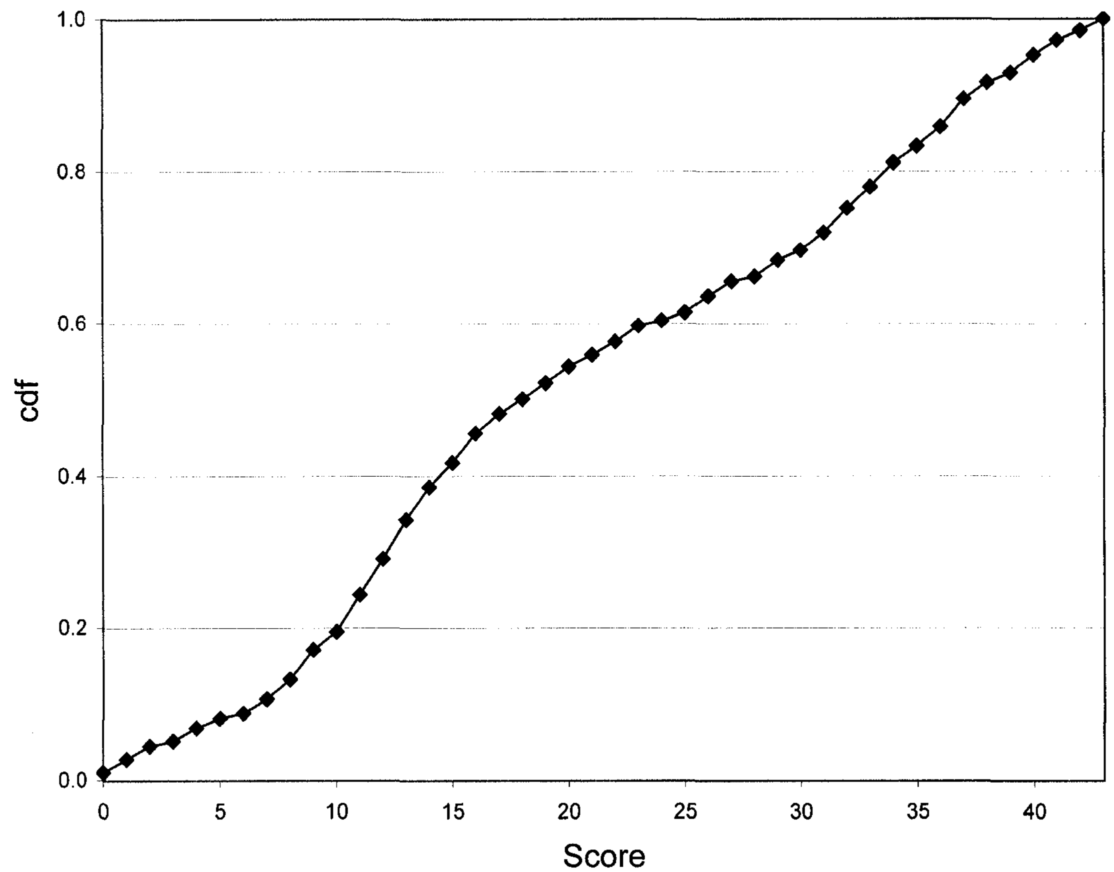


Figure B – 8
Miccei Multimodal Lumpy Data Set (Achievement)
Cummulative Distribution Function



APPENDIX C

RANKING METHODS OF POWER ANALYSIS

Results of the power and Type III error studies were analyzed by ranking methods. The first method involved ranking the obtained powers across methods for each unique combination of test, nominal alpha level, directionality, nominal effect size multiplier and initial sample size. For a particular test, nominal alpha level and directionality, these ranks were then analyzed within each distribution on a single Excel® spreadsheet by finding the mean ranks across: 1) nominal effect size multipliers; 2) initial sample sizes, and; 3) both nominal effect size multipliers and initial sample sizes. These analyses were performed on the power results rounded to four, three and two decimal places. Table C – 1 is an example of such a spreadsheet for the Wilcoxon-Mann-Whitney Test at α .01, 1-sided, four decimal places.

The other type of analysis consisted of counting the number of first place finishes for each of the three ways of calculating mean ranks described above. Tables C – 2, C – 3 and C – 4 provide examples based on the mean ranks in Table C – 1.

Table C – 1

Analysis of Power Results, Wilcoxon-Mann-Whitney Test
alpha .01, 1-sided, 4 decimals

NESM		0.2				0.5				0.8				1.2				row rank sums				row mean ranks			
Distribution		EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS	EA	EB	ML	SS
ISS	Mthd																								
6	3a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.5		3.5	3.5	2.5	3.5	3.5	3.5	10.0	7.0	14.0	14.0	2.5	3.5	3.5	3.5
	4			2.0	1.0		1.0	1.0	1.5			1.0	1.0		2.0	1.0	1.0		3.0	5.0	4.5		1.5	1.3	1.1
	5	1.0		1.0	2.0	1.0	2.0	2.0	1.5	1.0		2.0	2.0	1.0	1.0	2.0	2.0	4.0	3.0	7.0	7.5	1.0	1.5	1.8	1.9
	6a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.5		3.5	3.5	2.5	3.5	3.5	3.5	10.0	7.0	14.0	14.0	2.5	3.5	3.5	3.5
	cs	6.0		10.0	10.0	6.0	10.0	10.0	10.0	6.0		10.0	10.0	6.0	10.0	10.0	10.0	24.0	20.0	40.0	40.0				
12	3a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.5		3.5	3.5	2.5	3.5	3.5	3.5	10.0	7.0	14.0	14.0	2.5	3.5	3.5	3.5
	4			2.0	2.0		1.0	1.5	2.0			2.0	1.5		2.0	1.0	1.0		3.0	6.5	6.5		1.5	1.6	1.6
	5	1.0		1.0	1.0	1.0	2.0	1.5	1.0	1.0		1.0	1.5	1.0	1.0	2.0	2.0	4.0	3.0	5.5	5.5	1.0	1.5	1.4	1.4
	6a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.5		3.5	3.5	2.5	3.5	3.5	3.5	10.0	7.0	14.0	14.0	2.5	3.5	3.5	3.5
	cs	6.0		10.0	10.0	6.0	10.0	10.0	10.0	6.0		10.0	10.0	6.0	10.0	10.0	10.0	24.0	20.0	40.0	40.0				
18	3a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.5		3.5	3.5	2.5	2.5	3.5	3.5	10.0	6.0	14.0	14.0	2.5	3.0	3.5	3.5
	4			2.0	2.0		1.0	1.5	2.0			2.0	1.0		4.0	1.0	1.0		5.0	6.5	6.0		2.5	1.6	1.5
	5	1.0		1.0	1.0	1.0	2.0	1.5	1.0	1.0		1.0	2.0	1.0	1.0	2.0	2.0	4.0	3.0	5.5	6.0	1.0	1.5	1.4	1.5
	6a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.5		3.5	3.5	2.5	2.5	3.5	3.5	10.0	6.0	14.0	14.0	2.5	3.0	3.5	3.5
	cs	6.0		10.0	10.0	6.0	10.0	10.0	10.0	6.0		10.0	10.0	6.0	10.0	10.0	10.0	24.0	20.0	40.0	40.0				
24	3a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.5		3.5	3.5	2.5	3.5	3.5	2.5	10.0	7.0	14.0	13.0	2.5	3.5	3.5	3.3
	4			2.0	2.0		1.0	1.0	2.0			1.0	1.5		2.0	1.0	1.0		3.0	5.0	6.5		1.5	1.3	1.6
	5	1.0		1.0	1.0	1.0	2.0	2.0	1.0	1.0		2.0	1.5	1.0	1.0	2.0	4.0	4.0	3.0	7.0	7.5	1.0	1.5	1.8	1.9
	6a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.5		3.5	3.5	2.5	3.5	3.5	2.5	10.0	7.0	14.0	13.0	2.5	3.5	3.5	3.3
	cs	6.0		10.0	10.0	6.0	10.0	10.0	10.0	6.0		10.0	10.0	6.0	10.0	10.0	10.0	24.0	20.0	40.0	40.0				
30	3a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.0		3.5	3.5	1.5	1.5	3.5	2.0	8.5	5.0	14.0	12.5	2.1	2.5	3.5	3.1
	4			2.0	2.0		1.0	1.0	2.0			1.0	1.0		4.0	1.0	2.0		5.0	5.0	7.0		2.5	1.3	1.8
	5	1.0		1.0	1.0	1.0	2.0	2.0	1.0	2.0		2.0	2.0	3.0	3.0	2.0	4.0	7.0	5.0	7.0	8.0	1.8	2.5	1.8	2.0
	6a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.0		3.5	3.5	1.5	1.5	3.5	2.0	8.5	5.0	14.0	12.5	2.1	2.5	3.5	3.1
	cs	6.0		10.0	10.0	6.0	10.0	10.0	10.0	6.0		10.0	10.0	6.0	10.0	10.0	10.0	24.0	20.0	40.0	40.0				
col rank sums	3a	12.5		17.5	17.5	12.5	17.5	17.5	17.5	12.0		17.5	17.5	11.5	14.5	17.5	15.0	48.5	32.0	70.0	67.5				
	4			10.0	9.0		5.0	6.0	9.5			7.0	6.0		14.0	5.0	6.0		19.0	28.0	30.5				
	5	5.0		5.0	6.0	5.0	10.0	9.0	5.5	6.0		8.0	9.0	7.0	7.0	10.0	14.0	23.0	17.0	32.0	34.5	<— table rank sums			
	6a	12.5		17.5	17.5	12.5	17.5	17.5	17.5	12.0		17.5	17.5	11.5	14.5	17.5	15.0	48.5	32.0	70.0	67.5				
	cs	30.0		50.0	50.0	30.0	50.0	50.0	50.0	30.0		50.0	50.0	30.0	50.0	50.0	50.0	120.0	100.0	200.0	200.0				
		0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0				
col mean ranks	3a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.4		3.5	3.5	2.3	2.9	3.5	3.0					2.4	3.2	3.5	3.4
	4			2.0	1.8		1.0	1.2	1.9			1.4	1.2		2.8	1.0	1.2						1.9	1.4	1.5
	5	1.0		1.0	1.2	1.0	2.0	1.8	1.1	1.2		1.6	1.8	1.4	1.4	2.0	2.8	table mean ranks --->				1.2	1.7	1.6	1.7
	6a	2.5		3.5	3.5	2.5	3.5	3.5	3.5	2.4		3.5	3.5	2.3	2.9	3.5	3.0					2.4	3.2	3.5	3.4

Table C – 2

Analysis of Mean Ranks of Power Results, Wilcoxon-Mann-Whitney Test, 2 Groups, Number of First Place Finishes by Method and Distribution Across Nominal Alpha (.01, .05) and Directionality (1-s, 2-s), 4 decimals, by Initial Sample Size Across Nominal Effect Size Multiplier

		METHOD				Row
ISS	DIST	3a	4	5	6a	Checksum
6	EA	0		4.0	0	4.0
	EB	0	2.5	1.5	0	4.0
	ML	0	2.0	2.0	0	4.0
	SS	0	2.0	2.0	0	4.0
12	EA	0		4.0	0	4.0
	EB	0	2.0	2.0	0	4.0
	ML	0	0	4.0	0	4.0
	SS	0	0	4.0	0	4.0
18	EA	0		4.0	0	4.0
	EB	0	0.5	3.5	0	4.0
	ML	0	2.5	1.5	0	4.0
	SS	0	1.5	2.5	0	4.0
24	EA	0		4.0	0	4.0
	EB	0	2.0	2.0	0	4.0
	ML	0	4.0	0	0	4.0
	SS	0	2.5	1.5	0	4.0
30	EA	0		4.0	0	4.0
	EB	0	1.5	2.5	0	4.0
	ML	0	4.0	0	0	4.0
	SS	0	2.5	1.5	0	4.0
First Place Finishes	EA	0		20.0	0	20.0
	EB	0	8.5	11.5	0	20.0
	ML	0	12.5	7.5	0	20.0
	SS	0	8.5	11.5	0	20.0
Total 1st place for mthd across dist		0.0	29.5	50.5	0.0	80.0
# of mean ranks		20	15	20	20	Total 1st
2 alphas x 2 sides		4	4	4	4	
Max 1st place for method		80	60	80	80	
Proportion of Max 1st for method		0.000	0.492	0.631	0.000	
Proportion of Total 1st		0.000	0.369	0.631	0.000	1.000

Table C – 3

Analysis of Mean Ranks of Power Results, Wilcoxon-Mann-Whitney Test, 2 Groups, Number of First Place Finishes by Method and Distribution Across Nominal Alpha (.01, .05) and Directionality (1-s, 2-s), 4 decimals, by Nominal Effect Size Across Initial Sample Size

		METHOD				Row
NESM	DIST	3a	4	5	6a	Checksum
0.2	EA	0		4.0	0	4.0
	EB					
	ML	0	0	4.0	0	4.0
	SS	0	0	4.0	0	4.0
0.5	EA	0		4.0	0	4.0
	EB	0	4.0	0	0	4.0
	ML	0	4.0	0	0	4.0
	SS	0	0	4.0	0	4.0
0.8	EA	0		4.0	0	4.0
	EB					
	ML	0	3.5	0.5	0	4.0
	SS	0	3.0	1.0	0	4.0
1.2	EA	0		4.0	0	4.0
	EB	0	0	4.0	0	4.0
	ML	0	4.0	0	0	4.0
	SS	0	4.0	0	0	4.0
First Place Finishes	EA	0		16.0	0	16.0
	EB	0	4.0	4.0	0	8.0
	ML	0	11.5	4.5	0	16.0
	SS	0	7.0	9.0	0	16.0
Total 1st place for mthd across dist		0.0	22.5	33.5	0.0	56.0
# of mean ranks		14	10	14	14	Total 1st
2 alphas x 2 sides		4	4	4	4	
Max 1st place for method		56	40	56	56	
Proportion of Max 1st for method		0.000	0.563	0.598	0.000	
Proportion of Total 1st		0.000	0.402	0.598	0.000	1.000

Table C – 4

Analysis of Mean Ranks of Power Results, Wilcoxon-Mann-Whitney Test, 2 Groups, Number of First Place Finishes by Method and Distribution Across Nominal Alpha (.01, .05) and Directionality (1-s, 2-s), 4 decimals, Across Nominal Effect Size Multiplier and Initial Sample Size

	METHOD				Row
DIST	3a	4	5	6a	Checksum
EA	0		4.000	0	4.000
EB	0.333	0.500	2.833	0.333	4.000
ML	0	3.500	0.500	0	4.000
SS	0	2.000	2.000	0	4.000
Total 1st place for mthd across dist	0.333	6.000	9.333	0.333	16.000
# of mean ranks	4	3	4	4	Total 1st
2 alphas x 2 sides	4	4	4	4	
Max 1st place for method	16	12	16	16	
Proportion of Max 1st for method	0.021	0.500	0.583	0.021	
Proportion of Total 1st	0.021	0.375	0.583	0.021	1.000

APPENDIX D

GLOSSARY

Alpha(α): The *a priori* probability of a Type I error, also referred to as nominal alpha or significance level that a researcher has selected as acceptable in applying an inferential test. See also Type I Error, Type II Error, Type III error and Power.

Assumptions: The necessary and sufficient mathematical conditions on which a test is based and which, when met, ensure the validity of the result. When a test is used, it is 'assumed' that these conditions are met, which may or may not be the case. The validity of the result is not guaranteed when the assumptions are not met, although it may, in fact, still be valid.

Alternative hypothesis or H_1: The hypothesis to which a statistical test is most sensitive when the null hypothesis (H_0) is false. Alternative hypotheses may be non-directional or directional. In the Fisherian tradition, an alternative hypothesis is not needed or used. One merely supports or fails to support the null hypothesis, the alternative hypothesis being taken as 'not H_0'.

Asymmetrical distribution: A distribution (probability or frequency) that lacks symmetry about some central value.

Beta(β)-error: See Type II Error.

Bimodal: A distribution (probability or frequency) with two modes.

Binomial distribution: Everitt (1998) described this as "The distribution of the number of 'successes', X, in a series of *n*-independent *Bernoulli trials* where the probability of success at each trial is p and the probability of failure is $q = 1 - p$" (p. 34-35) where a Bernoulli trial is "A set of n independent binary variables in which the *j*th observation is either a 'success' or a 'failure', with the probability of success, *p*, being the same for all trials" (p. 30).

Ceiling effect: Refers to a situation in which many observations occur near the upper limit of a variable with finite, restricted range. Distributions of such scores are often skewed left and thus highly asymmetric.

Chi-squared distribution: Described by Everitt (1998) as:

> The probability distribution, *f*(*x*), of a random variable defined as the sum of squares of a number (*v*) of independent standard normal variables and given by $f(x) = \frac{1}{2^{v/2}\Gamma(v/2)} x^{(v-2)/2} e^{-x/2}$, $x > 0$, i.e., a *standard gamma distribution* with $\alpha = v/2$. The *shape parameter*, *v*, is usually known as the degrees of freedom of the distribution. (p. 59)

Contaminated normal distribution: A distribution formed from the "combination of two normal distributions having the same mean but different variances" (Everitt, 1998, p. 75). They are commonly used in certain types of Monte Carlo studies.

Continuous variable: A variable whose true values are represented by the domain of real numbers. When referring to sampling processes this is sometimes stated the probability of drawing any specific value is zero. In practice, it is taken to mean a variable whose underlying values can always be uniquely distinguished given sufficient precision of measurement.

Critical region: Also known as the rejection region. Everitt (1998) wrote:

> The values of a test statistic that lead to rejection of a null hypothesis. The size of the critical region is the probability of obtaining an outcome

belonging to this region when the null hypothesis is true, i.e., the probability of a type I error. (p. 86)

Critical value: The value or values of a test statistic that separate the critical region of the sampling distribution from the non-critical region. When referring to tables of critical values for a particular test it is the value of the test statistic for a particular alpha level and directionality against which the obtained value of the statistic is compared in order to determine whether or not to support or fail to support the null hypothesis.

Cumulative distribution function (cdf): See cumulative frequency distribution.

Cumulative frequency: The total count of observations up to and including a particular value or class interval (Vogt, 1999).

Cumulative frequency distribution (cfd): The complete display of cumulative frequencies for a sample of observations. "The empirical equivalent of the cumulative probability distribution" (Everitt, 1998, p. 88).

Cumulative probability distribution (cpd): See probability distribution.

Degrees-of-freedom (*df*): "An elusive concept that occurs throughout statistics. Essentially the term means the number of independent units of information in a sample relevant to the estimation of a parameter or calculation of a statistic" (Everitt, 1998, p. 95). "The number of pieces of information that can vary independently of one another" (Vogt, 1999, p. 75).

Dependent variable: Also known as the criterion, predicted, or response variable. Vogt (1999) described this as "(a) The presumed effect in a study, so called because it 'depends' on another variable. (b) The variable whose values are predicted by the independent variable, whether or not caused by it" (p. 78).

Directional alternative hypothesis: Also known as a directional or one-sided test. See one-sided test.

Discrete variables: Variables that can take on only certain values, in contradistinction to continuous variables. All measured data is inherently discrete due to limitations of precision in measurement.

Dispersion: A general reference to the degree to which observations differ from a measure of central tendency. Also known as spread. More specifically, see variance.

Density function: See probability density function (pdf).

Effect: Generally refers to a change in a dependent variable or variables that is associated with a change in one or more independent variables, often as a result of the application of 'treatment' to one or more groups of subjects.

Effect size: The amount of effect, generally stated in terms of the standard deviation of the population, i.e., $c\sigma$, where c is a constant multiplier and σ is the population standard deviation. Effect sizes for the social sciences are often quite small, with $0 \leq c \leq 2$ typically. A rule of thumb is $c = 0.2$ (small), $c = 0.5$ (medium), $c = 0.8$ (large), and $c = 1.2$ (large) (Cohen, 1988; Sawilowsky & Blair, 1992). The power of a test is functionally related to effect size, nominal alpha, and sample size.

Errors of the third kind: "Giving the right answer to the wrong question! (Not to be confused with Type III error.)" (Everitt, 1998, p. 116)

Exact test: See randomization test.

Expected value ($E(x)$): The mean of a random variable x.

Extreme asymmetric (distribution): Generally a highly skewed but essentially monotonic distribution. Specifically, the name given to the Micceri (1986) empirical frequency distribution considered representative of an achievement measure.

Extreme bimodal (distribution): In general any distribution exhibiting exactly two modes that dominate all other data values to an extreme degree. Specifically, the name given to the Micceri (1986) empirical frequency distribution considered representative of a psychometric measure obtained from the use of a Likert scale or other discrete valued instrument with severely restricted range.

Finite population: A population with a countable membership. In practice this means a population in which, if samples are drawn without replacement, the population will eventually be exhausted.

Floor effect: Refers to a situation in which many observations occur near the lower limit of a variable with finite, restricted range. Distributions of such scores are often skewed right and thus highly asymmetric.

Frequency distribution: "The division of a sample of observations into a number of classes, together with the number of observations in each class...Essentially the empirical equivalent of the probability distribution" (Everitt, 1998, p. 132). A histogram is normally used as the graphical representation of a frequency distribution.

Gaussian distribution: See normal distribution.

Goodness-of-fit: The concept of how close two (or more) things are to being identical.

Goodness-of-fit statistics: "Measures of the agreement between a set of sample observations and the corresponding values predicted from some model of interest" (Everitt, 1998, p. 144).

H_0: See null hypothesis.

H_1: See alternative hypothesis.

Heterogeneous: A general term used to indicate that a collection of objects is very dissimilar on one or more characteristics, in contradistinction to homogeneous.

Homogeneous: A general term used to indicate that a collection of objects is very similar on one or more characteristics, in contradistinction to heterogeneous.

Hypothesis testing: A general term for the use of a statistical procedure to infer whether sampled data might reasonably have come from a population with certain characteristics.

IMSL STAT/LIBRARY: "A Fortran subprogram library of statistical and mathematical procedures from Visual Numerics Inc., Houston, TX.

Independence: "Essentially, two events are said to be independent if knowing the outcome of one tells us nothing about the other" (Everitt, 1998, p. 161). Mathematically, two events are independent if $P(A \text{ and } B) = P(A) \times P(B)$.

Independent variable: Also known as a predictor, explanatory, manipulated or controlled variable, as distinguished from a dependent variable.

In true experiments it is the variable(s) that the experimenter can directly manipulate in an effort to determine the effect, if any, on the dependent variable(s).

Inference: "The process of drawing conclusions about an unknown population on the basis of measurements or observations made on a sample of individuals presumed to come from that population.

Infinite population: An uncountable collection of objects. In practice this represents the idea of a population corresponding to the set of real numbers such that, if the population is sampled without replacement, the population will never be exhausted.

Large sample method: "Any statistical method based on an approximation to a normal distribution or other probability distribution that becomes more accurate as sample size increases" (Everitt, 1998, p. 183).

Level of significance: See alpha and significance level.

Linear Congruential Generator (LCG): A method for generating (pseudo) random numbers originally proposed by Lehmer (1948) (as cited in Gentle, 1998, p. 6). The basic form is a congruence relationship, given by the recursive formula $x_i \equiv (ax_{i-1} + c) \bmod m$, with $0 \leq x_i < m$ (Gentle, 1998, p. 6) where x_i, a, c and m are all non-negative integers. With $c \neq 0$ this is also known as a mixed congruential generator while $c = 0$ is known as a multiplicative congruential generator. The resulting sequence of numbers is converted to rational numbers in the interval (0,1) by $u_i = x_i/m$ (Gentle, 1998, p. 6). Extensions include the multiple recursive generator (MRG) given by Gentle (1998) as

$x_i = (a_1 x_{i-1} + a_2 x_{i-2} + \cdots + a_k x_{i-k}) \bmod m$ (p. 22). This can also be extended to generate pseudorandom vectors in the form of a matrix congruential generator (MCG) given by $x_i \equiv (Ax_{i-1} + C) \bmod m$ (Gentle, 1998, p. 26) where the x_i are vectors of length *d* and *A* and *C* are *d* x *d* matrices. Finally, Gentle (1998) described extension to a multiple recursive matrix generator given by $x_i \equiv (A_1 x_{i-1} + A_2 x_{i-2} + \cdots + A_k x_{i-k}) \bmod m$ (p. 26) with terms previously defined.

Lower-tail: The lower (left hand) portion of a probability distribution used to determine if a statistic falls in the rejection region for a result in that direction.

Matrix Congruential Generator (MCG): See Linear Congruential Generator.

Mean: One of the measures of central tendency, often referred to (ambiguously) as the 'average.' Everitt (1998) defined the mean as "A measure of location or central value for a continuous variable" (p. 209). See also expected value. The basic calculation for a sample of observations x_1, x_2, ..., x_n is given by $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Everitt (1998) added that the mean is "Most useful when the data have a symmetric distribution and do not contain outliers" (p. 209).

Median: Everitt (1998) defined the median as:

> The value in a set of ranked observations that divides the data into two parts of equal size. When there is an odd number of observations the median is the middle value. When there is an even number of observations the measure is calculated as the average of the two central values. Provides a measure of location of a sample that is suitable for asymmetric distributions and is also relatively insensitive to the presence of outliers. (p. 210)

Measures of location: Also known as measures of central tendency. Parameters and statistics that attempt to locate the center of a population or sample. See mean, median, and mode.

Measure of scale: Also known as measures of dispersion or spread. Parameters and statistics that attempt to quantify the degree to which data is spread out or clustered together, often with respect to a measure of location. Variance, standard deviation, and interquartile range are common examples.

Mid-range: The mean of the smallest and largest values in a sample. For symmetric distributions the mean and median are the same and the mid-range may estimate them quite efficiently.

Mid-rank: The rank assigned to all members of a subset of a sample when those members have equal data values resulting in tied ranks. It is the mean (median, mid-range) of the set of ranks the tied-for values would have been assigned if they had been distinct. For example, in ranking the data set {1,3, 5, 5, 5} from 1 to 5 the five data points would be assigned the ranks {1, 2, 4, 4, 4} as 4 is the mid-rank for the three tied data values which would normally occupy ranks 3, 4, and 5. This can become problematic with a sample such as {1, 3, 5, 5, 5, 5} where the mid-rank for positions 3 through 6 is now between 4 and 5. As such it would be assigned the value 4.5, which is not an integer, thus giving the ranks {1, 2, 4.5, 4.5, 4.5, 4.5} for the six data values.

Mis-interpretation of P-values: Everitt (1998) offered the following caution with regard to interpretation of p-values: "A P-value is commonly interpreted in a variety of ways that are incorrect. Most common are that it is the probability of the null hypothesis, and that it is the probability of the data having arisen by chance" (p. 213). See the entry for P-value for the correct interpretation.

Mode: One or more data values from a set of data values that occurs more frequently than any others, either globally or locally. Related to the idea of a global or relative maximum of a function.

Monte Carlo methods: Gentle (1998) described Monte Carlo simulations, in their greatest generality, as "the use of experiments with random numbers to evaluate a mathematical expression" (p. 131). More specifically related to statistics is the use of Monte Carlo methods for both exploratory and confirmatory work on new and existing statistics. Gentle (1998) indicated that Monte Carlo studies "often provide a significant amount of the available knowledge of the properties of statistical techniques, especially under various alternative models" (p. 178). Studies of the Type I error and power properties of statistical tests in response to violation of assumptions are of this later type.

Multimodal distribution: A distribution (probability or frequency) with two or more modes. Multimodality can arise from heterogeneous populations comprised of distinct homogeneous sub-populations. It can also arise in situations where there is digit preference.

Multimodal lumpy (distribution): The name implies a discrete distribution with more than one mode and with some values occurring relatively more often than other nearby values, resulting in a distribution that is 'lumpy' rather than smooth. Specifically, the name given to the Micceri (1986) empirical frequency distribution representative of certain types of achievement measures.

Multiple Recursive Generator (MRG): See Linear Congruential Generator.

Nominal alpha (α): See alpha (α).

Nominal significance level: Synonymous with alpha, nominal alpha, level of significance, and significance level. See alpha.

Noncentral distribution: Everitt (1998) defined this as:

> A series of probability distributions each of which is an adaptation of one of the standard sampling distributions such as the chi-squared distribution, the *F*-distribution or Student's *t*-distribution for the distribution of some test statistic under the alternative hypothesis. Such distributions allow the power of the corresponding hypothesis to be calculated. (p. 230)

Nonnormal distribution: Any distribution that is not a Normal distribution.

Normal approximation: Everitt (1998) described this as "A normal distribution with mean np and variance $np(1 - p)$ that acts as an approximation to a binomial distribution as n, the number of trials, increases. The term p represents the probability of a 'success' on any trial" (p. 233). With reference to some nonparametric/distribution-free statistics, it refers to one of two situations, either a) that the sampling distribution of the statistic approaches a normal distribution with increasing sample size, or b) that a normal approximation formula is available that converts the test statistic to a form that is approximately normally distributed. In either case, the inferential test can then be performed by referring the resulting value to the table of critical values based on the normal distribution.

Normal distribution: "A probability distribution, $f(x)$, of a random variable, X, that is assumed by many statistical methods" (Everitt, 1998, p. 233). For mean μ and variance σ^2, the normal distribution is defined by the equation

$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\forall x,\ -\infty < x < \infty$. The distribution is fully determined when μ and σ^2 are specified. When graphed it produces a 'bell-shaped' curve. Since all possible events are represented by this distribution, the area under the curve, down to the *x*-axis, is equal to 1.

Normal random variable: A random variable that is normally distributed.

Normal variate: See normal random variable.

Normal theory: Statistical theories based on the assumption that populations consist of normally distributed random variables and/or that statistics of interest are normally distributed.

Normality: A term used to indicate that some variable of interest is a normal random variable.

Null distribution: The probability distribution of a test statistic when the null hypothesis is true.

Null hypothesis: The hypothesis of 'no difference', 'no association', or 'no effect'; the hypothesis that all samples were drawn from the same or identical populations. Note that in the Fisherian tradition, an alternative hypothesis was not required. The result of a test either supported or failed to support the null hypothesis. Failure to support was taken as an indication that there was a difference between the samples that was unlikely to have occurred by chance, i.e., as an artifact of sampling, with the resulting inference that the samples were representative of populations that differed in some way.

One-sided test: Synonymous with a directional alternative hypothesis. A significance test for which the alternative hypothesis is directional, for example, that one population mean is greater than another. The choice between a one-sided and two-sided test must be made before data is collected and examined and, in general, requires *a priori* support for the research question the null and alternative hypotheses are designed to answer. A one-sided test is more powerful than a two-sided test, all else equal, if the effect occurs in the direction anticipated. However, a one-sided test has no power to detect an effect in the other direction, regardless of magnitude.

One-tailed test: See one-sided test.

Ordered alternative hypothesis: An alternative hypothesis that specifies a specific *a priori* order for the populations based on whatever property of the sample data the test is sensitive to.

Omnibus alternative hypothesis: The alternative hypothesis that the populations from which the samples were drawn are not all the same, without regard to direction, order, or specific source of the difference, based on whatever property of the sample data the test is sensitive to.

Omnibus test: A test utilizing an omnibus alternative hypothesis.

Order statistics: The identification of a value in a ranked set of observations with its position, i.e., the *r*th largest value in a sample is called the *r*th-order statistic.

Ordinal variable: A variable that is measured on an ordinal scale. Ordinal scales require only the ability to determine greater than (or less than) so that observations can be ranked with respect to some characteristic. Differences at different points of the scale are not necessarily equivalent, however, so that nothing is known about the metric distance between the observations, if indeed the concept of metric distance even applies to the observations. Ordinal variables can arise as a result of ranking measurements made on a higher scale (interval or ratio), or they may originate as ranks, as when a judge places objects of study in rank order by assigning ranks to each object.

Outlier: Everitt (1998) gave the following description of an outlier:

> An observation that appears to deviate markedly from the other members of a sample in which it occurs...More formally the term refers to an observation which appears to be inconsistent with the rest of the data, relative to an assumed model. Such extreme observations may be reflecting some abnormality in the measured characteristic of the subject, or they may result from an error in the measurement or recording. (p. 240)

Mean and median based criteria can be used to classify observations as outliers. Three standard deviations from the mean, or 1.5 times the interquartile range from the median, are typical but arbitrary conventions. The mean is sensitive to outliers, however, whereas the median is not. In general, rank-based (order) statistics are much less sensitive to the presence of outliers than tests based on the mean. When outliers occur as a result of errors in measurement or recording they misrepresent the true situation and should rightly be corrected or eliminated. When they arise naturally and correctly they become more problematic. On the one hand they are not 'typical' of the bulk of the data and may distort the test statistic and resulting inference. On the other hand, on what basis are they to be dismissed, other than the fact that they are statistically inconvenient?

Parameter: A characteristic of a population or a model. The mean and variance of a normal distribution, for instance, completely specify its shape. Parameters are rarely known for real populations. One of the major purposes of inferential parametric statistics is to make estimates of, or draw conclusions about, population parameters based on sample data. The ability to do this depends, in turn, on the assumption that the underlying populations have specific distributional forms for which the parameters are appropriate. See parametric methods.

Parametric hypothesis: "A hypothesis concerning the parameters of a distribution. For example, the hypothesis that the mean of a population equals the mean of a second population, when the populations are assumed to have a normal distribution" (Everitt, 1998, p. 244).

Parametric methods: "Procedures for testing hypotheses about parameters in a population described by a specified distributional form, often, a normal distribution. Student's t-test is an example of such a method" (Everitt, 1998, p. 244). Such methods stand in contrast to nonparametric methods, which do not estimate or draw inferences about population parameters..

Parametric statistic: A statistic whose purpose is to estimate a population parameter. The sample mean, $\bar{x}$ estimates the population mean μ, and the sample variance s^2, estimates the population variance, σ^2, when properly adjusted for sample size.

PDF: An abbreviation for probability density function. See probability distribution. (Note to be confused with 'portable document format' from Adobe Software, Inc.)

Permutation test: Synonymous with randomization test.

Population: In statistics this term is used for any collection of units, finite or infinite, which are the objects of study based on some observable, measurable characteristic(s). Populations often consist of people, but can just as well be animals, events, institutions, physical objects or processes. In most research, the population(s) are theoretical in that the researcher cannot, as a practical matter, observe or measure all members of the population. If they could, inferential statistics would be unnecessary and descriptive statistics would suffice. Since it is necessary to use samples in place of populations, one of the central issues in inferential statistics is sampling, i.e., employing methodologies to obtain samples that are representative of the target population (to the extent possible). Another central issue concerns the nature and extent of assumptions about the form of the population attached to a particular statistical test.

Power: Everitt (1998) touched on several aspects of power:

> The probability of rejecting the null hypothesis when it is false. Power gives a method of discriminating between competing tests of the same hypothesis, the test with the higher power being preferred. It is also the basis of procedures for estimating the sample size needed to detect an effect of a particular magnitude. (p. 259)

Power can be thought of as the ability of a test to detect an effect if it is present. It is equal to $1-\beta$, where β is the Type II error of the test. It depends in a complicated way on a number of factors, including the chosen significance level, sample size, and actual effect size (Cohen, 1988). For a given significance level and sample size, a particular test will have more power to detect larger effect sizes and less power to detect smaller ones. For a fixed alpha and effect size, larger samples will give a test more power. For a fixed sample size and effect size, a higher alpha (higher risk of committing a Type I error) results in a lower Type II error rate and thus more power.

Precision: "A term applied to the likely spread of the estimates of a parameter in a statistical model. Measured by the standard error of the estimator; this can be decreased, and hence precision increased, by using a larger sample size" (Everitt, 1998, p. 260).

Probability: Vogt (1999) defined probability as "(a) The likelihood that a particular event or relationship will occur; the proportion of tries that are successes. More formally, out of all possible outcomes, the proportionate expectation of a given outcome" (p. 222). Probabilities take on real number values on the interval [0,1]. A probability of zero means an event cannot occur. A probability of 1 means an event is certain, i.e., it must happen every time.

Probabilities come in several flavors. Subjective probabilities are estimates, guesses or hunches based on a 'feeling' about a situation but unsupported by precise computation. Empirical probabilities are based actual counts of observed events divided by the total number of possible events. Theoretical probabilities are based on mathematical calculations of expected or predicted likelihood. Conditional probabilities apply to situations in which the probability of some event, B, depends on (is conditioned by) a preceding event,

A. Finally, joint probability refers to the probability of a simultaneous occurrence of two or more events. See Vogt (1999, p. 222-223) for more complete descriptions.

Probability density: See probability distribution.

Probability density function (pdf): See probability distribution.

Probability distribution: Everitt (1998) gave the following description:

> For a discrete random variable, a mathematical formula that gives the probability of each value of the variable. ... For a continuous random variable, a curve described by a mathematical formula which specifies, by way of areas under the curve, the probability that the variable falls within a particular interval. ... In both cases the term probability density may also be used. (A distinction is sometimes made between 'density' and 'distribution', when the latter is reserved for the probability that the random variable falls below some value. In this dictionary, however, the latter will be termed the cumulative probability distribution and probability distribution and probability density used synonymously.) (p. 262)

Pseudorandom numbers: Gentle (1998) stated that pseudorandom numbers are those that arise from a deterministic process but appear to be randomly drawn from some known distribution. Because the process is deterministic the sequence of numbers will eventually repeat. Computer programs that produce pseudorandom number sequences are known as random number generators, the 'pseudo' being understood. The quality of these generators depends on the algorithms used and the word length of the processor. A very long sequence of numbers with good randomness characteristics that also provides smooth, complete conformance to the intended distribution of values are the hallmarks of a quality random number generator.

P-value: "The probability of the observed data (or data showing a more extreme departure from the null hypothesis) when the null hypothesis is true"

(Everitt, 1998, p. 268). See also misinterpretation of P-value, significance test and significance level.

Random: "Events that are unpredictable because their occurrence is unrelated to their characteristics" (Vogt, 1999, p. 233). "Governed by chance; not completely determined by other factors. Non-deterministic" (Everitt, 1998, p. 274). Most inferential statistical procedures require random selection of subjects (random sampling) and random assignment of subjects to groups. See below.

Random allocation (assignment): Vogt (1999) noted that "random assignment increases internal validity" (p. 233). Everitt (1998) characterized random allocation as:

> A method for forming treatment and control groups particularly in the context of a clinical trial. Subjects receive the active treatment or placebo on the basis of the outcome of a chance event, for example, tossing a coin. The method provides an impartial procedure for allocation of treatments to individuals, free from personal biases, and ensures a firm footing for the application of significance tests and most of the rest of the statistical methodology likely to be used. Additionally the method distributes the effects of concomitant variables, both observed and unobserved, in a statistically acceptable fashion. (p. 274

Random number: Random numbers or random variates are simulated realizations of random variables (Gentle, 1998, p. 3). Truly random processes are difficult to generate artificially such as with a computer. See pseudorandom numbers.

Random sample, random sampling: Also referred to as a simple random sample or equiprobability sample. The sampling process is also known as random selection. Vogt (1999) asserted that "random sampling increases external validity...and reduces the likelihood of bias" (p. 233 and 235). Random sampling is the best way to ensure samples that are representative of the population from which they are drawn. Everitt described a random sample as "Either a set of *n* independent and identically distributed random variables, or a sample of *n* individuals selected from a population in such a way that each sample of the same size is equally likely" (p. 276). In other words, a truly random sample requires that all members of the population are available for selection and are equally likely to be selected on any given draw. This requirement is rarely met in practice. Variations on simple random sampling include stratified random sampling, cluster sampling, and quota sampling (Vogt, 1999, p. 235).

Random variable: Also known as a stochastic variable. "A variable, the values of which occur according to some specified probability distribution" (Everitt, 1998, p. 276). "A variable that varies in ways the researcher does not control; a variable whose values are randomly determined" (Vogt, 1999, p. 235).

Random variation: "The variation in a data set unexplained by identifiable sources" (Everitt, 1998, p. 276).

Randomization tests: Procedures for determining statistical significance directly from sample data by comparing the value of the statistic for the actually obtained samples with values of the statistic that result from all possible permutations of that data. If the proportion of resulting statistics that is less than the obtained statistic is smaller than some *a priori* significance level then the result is deemed significant at that level.

Ranking: The process of arranging a data set into either ascending or descending order.

Rank order statistics: Statistics based only on the ranks of the sample observations, for example the Wilcoxon signed-rank and rank-sum tests. See order statistics and ordinal variables.

Ranks: The numbers, usually integers starting from 1, that correspond to the relative positions of the members of a sample that has been arranged in increasing or decreasing order with respect to some characteristic that admits of greater than (or less than) comparison. See order statistics, ordinal variables and rank order statistics.

Rectangular distribution: See uniform distribution.

Relative efficiency: "The ratio of the variances of two possible estimates of a parameter or the ratio of the sample sizes required by two statistical procedures to achieve the same power" (Everitt, 1998, p. 283).

Relative frequency: See probability.

Resistant statistics: Measures and statistical tests that are relatively un-influenced by unusual or extreme scores such as outliers (Vogt, 1999).

Robust statistics: Measures and statistical tests that maintain Type I error within a close approximation of nominal alpha and maintain their power to detect the non-null conditions represented by the alternative hypothesis when one or more of the assumptions of the test has been violated. Robustness is not an absolute concept. The nature of robustness is complex, depending on the particular combination of assumptions that have been violated and the degree of violation.

Sample: A subset of a population chosen by some process for the purpose of investigating (inferring) particular properties of the population from an analysis of the sample data.

Sample size: The number of objects in a sample. A distinction is sometimes made between the total sample size and the per group sample sizes if there is more than one sample group. Per group sample sizes may or may not be equal depending on the design of the investigation and the requirements of the analytical procedures and statistical tests. Sample size has a direct effect on the power of statistical test to detect an effect of a specified size.

Sample space: A set of data points that represents possible outcomes of an experiment. A list of all possible samples of given size that can be drawn from a particular population (Vogt, 1999).

Sampling: The act of selecting a subset of a population for study. The major concern in sampling is that the objects of study or units of analysis selected are as representative of the larger population as possible. See random sampling.

Sampling distribution (of a statistic): "The probability distribution of a statistic calculated from a random sample of a particular size" (Everitt, 1998, p. 293). It is a theoretical frequency distribution of all the possible values of a statistic, and their associated probabilities of occurrence, for specific sample size (Vogt, 1999). The whole of inferential statistics depends on sampling distributions of statistics. It is the sampling distribution of a statistic that allows the determination as to whether the value of the statistic for the obtained sample might reasonably have occurred by chance or not (Vogt, 1999).

Sampling error: A theoretical concept, it is the difference between the value of a sample result and the population characteristic being estimated, which in general is not known. Good sampling procedures help ensure that this error is small and bounded. In Monte Carlo studies the population characteristics are known and so it is possible to determine the sampling error.

Sampling variation (or variability): The variation in the value of a statistic between samples of the same size drawn from the same population. Sampling variation is used in determining the bounds of sampling error.

Sampling with and without replacement: When sampling with replacement an object is returned to the population (after examining it) before the next object is drawn, thus insuring that the probability for selection for every object remains the same from draw to draw and that the sampling is independent. When sampling without replacement an object is not returned to the population after selection. For infinite populations this does not affect the independence of the sampling as the probabilities of selection remain unchanged from draw to draw. In a finite population, however, sampling without replacement means the probabilities for selection at each draw are conditional upon the selections that have preceded it.

Scale parameter: "A general term used for a parameter of a probability distribution that determines the scale of measurement. For example, the variance parameter in a normal distribution" (Everitt, 1998, p. 295).

Shape parameter: "A general term for a parameter of a probability function that determines the 'shape' (in a sense distinct from location and scale) of the distribution within a family of shapes associated with a specified type of variable" (Everitt, 1998, p. 303). Examples include skew & kurtosis.

Significance level: The arbitrary *a priori* probability at which the null hypothesis fails to be accepted or supported. The .05 level is conventional in much of social and behavioral science because it provides a balance of an acceptably small Type I error rate, sufficient power to detect a meaningful effect, should one exist, and reasonable sample sizes. The 0.01 level is also commonly used, especially in situations where the consequences of Type I errors are high, such as in some medical applications. Indeed, if it is much more important to guard against a 'false positive' (Type II error) than to correctly detect an effect, significance levels of .001 may be employed. In such cases, however, larger sample sizes are needed to maintain the power of the test. The legitimacy of all inferential statistical tests require that the significance level be established prior to the collection, examination, or testing of the data.

Significance test: A statistical procedure applied to some characteristic of a sample that results in the value of a test statistic with its associated P-value. The P-value is then compared to nominal alpha (significance level) to determine whether or not the null hypothesis is supported.

Simple random sampling: As described in Everitt (1998):

> A form of sampling design in which *n* distinct units are selected from the *N* units in the population in such a way that every possible combination of *n* units is equally likely to be the sample selected. ... Designs other than this one may also give each unit equal probability of being included, but

only here does each possible sample of n units have the same probability. (p. 306).

Simulation: In its most general sense a simulation is a recreation of something. More specifically, computer simulations are concerned with recreating a real process through mathematical and logical modeling. When that modeling also includes random processes that require the generation of pseudorandom numbers, it becomes an experiment (Gentle, 1998). Such simulations are usually referred to as Monte Carlo methods or studies.

Skew (or skewness): The lack of symmetry in a probability distribution. A distribution with most of its values to the left and a long, thin tail to the right has positive skew. When most of the data is to the right and there is a long, thin tail to the left, the distribution as negative skew. Skew is calculated using the second (μ_2) and third (μ_3) moments about the mean (μ_1) by the formula

$s = \frac{\mu_3}{\mu_2^{3/2}}$. For symmetrical distributions, s = 0.

Smooth symmetric: A general description of a distribution in which adjacent frequencies are relatively close (smooth) and equally distributed to either side of a central point (symmetric). Specifically, the name given to the Micceri (1986) empirical frequency distribution most representative of these characteristics.

Spread: See dispersion.

Standard deviation: A measure of the spread in a set of observations equal to the square root of the variance. See variance.

Standard error: "The standard deviation of the sampling distribution of a statistic" (Everitt, 1998, p. 318).

Standard error of the (?): "For example, the standard error of the mean of n observations is $\sigma/\sqrt{n}$ where σ^2 is the variance of the original observations" (Everitt, 1998, p. 318). Standard error of the estimate is the more general term.

Statistic: A numerical characteristic of a sample. For example, the sample mean, sample variance, sample median, or sample interquartile range. See parameter for comparison.

Stochastic ordering: See univariate directional ordering.

Student's *t*-distribution: As given by Everitt (1998, p. 323), it is the distribution of the variable $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ where $\bar{x}$ is the arithmetic mean of n observations from a normal distribution with mean μ and sample standard deviation s. For $v = n-1$, the distribution is given by the formula

$$f(t) = \frac{\Gamma\left\{\frac{1}{2}(v+1)\right\}}{(v\pi)^{1/2}\,\Gamma\left\{\frac{1}{2}v\right\}}\left(1+\frac{t^2}{v}\right)^{-\frac{1}{2}(v+1)}$$

. The shape of $f(x)$ is symmetric and varies with v (n), approaching normality as n approaches infinity. For small values of v $f(x)$ has heavier tails and does not rise as high at its maximum point.

Student's *t*-Tests: A family of tests about means of normally distributed populations. They are uniformly most powerful tests when their assumptions are met, which is rarely (if ever) the case. The single sample t-test tests the mean of a sampled population against a hypothesized value. The two-independent-samples t-test compares the means of two independently selected samples. The matched pairs t-test is used with two-related-samples situations, such a pre- and post-test scores.

The test statistic for the two-independent-samples version, given below, has a Student's t-distribution with $n_1 + n_2 - 2$ degrees of freedom when the null hypothesis is true. For samples of size n_1, n_2 with sample means $\bar{x}_1$, $\bar{x}_2$, sample variances s_1^2, s_2^2 and common variance s^2, given by

$s^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$, the test statistic is $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$.

Sturdy statistics: See robust statistics.

Symmetrical distribution: A distribution (probability or frequency) whose values are symmetrical about some central value.

Target population: The intended (theoretical) population about which it is desired to make an inference. To the extent that the population actually sampled differs from the target population inferences may not be valid for the target population.

Test statistic: A statistic used to assess the plausibility of a particular hypothesis in relation to some target population. See sampling distribution.

Tied observations: Two or more observations in a sample that have the same value on some characteristic of interest.

Tied ranks: When ranking the elements of set, some values of which are the same, those elements end up 'tied' for the assignment of the subset of ranks that occur at that point in the sequence.

Two-sided test: Synonymous with a non-directional alternative hypothesis. A significance test for which the alternative hypothesis does not postulate a specific direction for the effect, e.g., that one population mean is greater than another. The choice between a one-sided and two-sided test must be made before data is collected and examined. Lacking *a priori* support for a directional expectation, the test should be two-sided. Two-sided tests are less powerful than one-sided tests, but can detect an effect in either direction. See omnibus test and omnibus alternative hypothesis.

Type I error: Incorrectly rejecting (failing to accept) a true null hypothesis; concluding that populations were different, or that the treatments were effective, when they were not. (Reject when true is not Type II.)

Type II error: Incorrectly accepting (supporting) the null hypothesis when it is not true; concluding that the populations were not different, or that the treatments were not effective, when in fact they were. Failing to reject a false null hypothesis.

Type III error: This term is not seen much in the literature. It has been used to indicate a significant result in the wrong tail. "It has been suggested by a number of authors that this term be used for identifying the poorer of two treatments as the better" (Everitt, 1998, p. 338). Not to be confused with errors of the third kind.

Uniform distribution: Everitt (1998) described this as:

> The probability distribution, $f(x)$, of a random variable having constant probability over an interval. Specifically given by $f(x) = 1/(\beta - \alpha)$ where $\alpha < x < \beta$. The mean of the distribution is $(\alpha + \beta)/2$ and the variance is $(\beta - \alpha)2/12$. The most commonly encountered such distribution is one in which the parameters α and β take values 0 and 1, respectively. (p. 340)

Note that for these values of α and β the distribution is uniform on the interval (0,1).

Uniformly most powerful test: "A test of a given hypothesis that is (at least) as powerful as another for all values of the parameter under consideration, and more powerful for at least one value of the parameter" (Everitt, 1998, p. 340).

Unimodal: A distribution (probability or frequency) having only a single mode. The normal distribution and Student's *t*-distribution are examples.

Univariate data: Data involving a single measurement on a common characteristic of each subject, object or unit of analysis.

Univariate directional orderings: "A term applied to the question of whether one distribution is in some sense to the right of another" (Everitt, 1998, p. 341). For random variables X and Y, stochastic ordering is often taken to mean that Y is stochastically greater than X if their cumulative probability distributions, F and G, satisfy the relationship $F(t) \geq G(t), \forall t$.

Upper-tail: The upper (right hand) portion of a probability distribution used to determine if a statistic falls in the rejection region for a result in that direction.

U-shaped distribution: A distribution (probability or frequency) shaped more or less like the letter U without implying symmetry. Such a distribution has its greatest frequencies at the two extremes of the range of the variable and relatively few frequencies near its center.

Variable: Some common characteristic that differs from object to object or from time to time for the same object. Also, a symbolic placeholder that can assume different numerical values, such as in an algebraic expression or a computer program.

Variance: A measure of the spread or dispersion in a set of scores (Vogt, 1999). The standard deviation is the square root of the variance. "In a population, the second moment about the mean" (Everitt, 1998, p. 343).

An unbiased estimator of the population value is derived from a sample as the sum of the squared deviations from the sample mean divided by the number of observations minus one. For a sample with n observations, $x_1, x_2, \ldots, x_n$ and sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, the variance is given by $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$.

Zero-difference scores: The score that results when forming difference scores by subtraction of two equal data values. Zero difference scores can result in tied ranks and are problematic for some tests.

BIBLIOGRAPHY

Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Algina, J., Oshima, T. C., & Lin, W. Y. (1994). Type I error rates for Welch's test and James's second-order test under nonnormality and inequality of variance when there are two groups. *Journal of Educational and Behavioral Statistics*, *19*(4), 275-291.

Ansari, A. R., & Bradley, R. A. (1960). Rank-sum test for dispersions. *Annals of Mathematical Statistics*, *31*, 1174-1189.

Arbuthnot, J. (1710). An argument for divine providence, taken from the constant regularity observed in the births of both sexes. *Philosophical Transactions*, *27*, 186-190.

Arnold, J. C., & Briley, T. S. (1973). A distribution free test for extreme reactions. *Educational and Psychological Measurement*, *33*(2), 301-309.

Bergmann, R., Ludbrook, J., & Spooren, W. P. J. M. (2000). Different outcomes of the Wilcoxon-Mann-Whitney test from different statistics packages. *The American Statistician*, *54*(1), 72-77.

Berry, K. J., & Mielke, P. W., Jr. (1997). Exact and approximate probability values for the Terpstra-Jonckheere test against ordered alternatives. *Perceptual and Motor Skills*, *85*, 107-111.

Berry, K. J., & Mielke, P. W., Jr. (2000). Exact and Monte Carlo resampling procedures for the Wilcoxon-Mann-Whitney and Kruskal-Wallis tests. *Perceptual and Motor Skills Special Issue*, *91*(3 Part 1), 749-754.

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. (undated). MOTHER random number generator from BFRA utility module version 2.1. Obtained from G. F. Fahoome of Wayne State University.

Blair, R. C., & Higgins, J. J. (1980). A comparison of the power of Wilcoxon's rank-sum statistic to that of Student's t statistic under various nonnormal distributions. *Journal of Educational Statistics*, *5*(4), 309-335.

Blair, R. C., & Higgins, J. J. (1985). Comparison of the power of the paired samples t test to that of Wilcoxon's signed-ranks test under various population shapes. *Psychological Bulletin*, *97*(1), 119-128.

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

Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, *40*, 318-334.

Box, G. E. P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, *25*, 290-302.

Box, G. E. P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics*, *25*, 484-498.

Box, G. E. P., & Andersen, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption. *Journal of the Royal Statistical Society*, *XVII*(1), 1-34.

Boyer, C. B. (1985). *A history of mathematics*. Princeton, N.J.: Princeton University Press.

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

Bradley, J. V. (1975). The optimal–pessimal paradox. *Human Factors*, *17*(4), 321-327.

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

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

Bridge, P. D., & Sawilowsky, S. S. (1997). Revisiting the t test on ranks as an alternative to the Wilcoxon rank-sum test. *Perceptual and Motor Skills*, *85*, 399-402.

Chacko, V. J. (1963). Testing homogeneity against ordered alternatives. *Annals of Mathematical Statistics*, *34*, 945-956.

Cliff, N. (1989). Ordinal consistency and ordinal true scores. *Psychometrika*, *54*(1), 75-91.

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, *114*(3), 494-509.

Cliff, N. (1996a). Answering ordinal questions with ordinal data using ordinal statistics. *Multivariate Behavioral Research*, *31*(3), 331-350.

Cliff, N. (1996b). *Ordinal methods for behavioral data analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

COMPAQ. (1999 Sep). *Compaq FORTRAN language reference manual* (4th ed., Vol. 1). Houston, TX: Compaq Computer Corporation.

COMPAQ. (2000). *Visual Fortran* (Version 6.5) [Fortran]. Houston, TX: Compaq Computer Corporation.

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: John Wiley & Sons, Inc.

Day, C. A. (1972). *Fortran techniques with special reference to non-numerical applications*. London: Cambridge University Press.

de Moivre, A. (1733). *Approximatio ad Summam Terminorum Binomii* $(a + b)^n$ *in Seriem expansi.*

de Moivre, A. (1738). *The Doctrine of Chances* (2nd ed.).

Deng, L.-Y., & Lin, D. K. J. (2000). Random number generation for the new century. *The American Statistician*, *54*(2), 145-150.

Deshpande, J. V., Gore, A. P., & Shanubhogue, A. (1995). *Statistical analysis of nonnormal data*. New York: John Wiley & Sons, Inc.

Elmore, P. B., & Woehlke, P. L. (1998, Apr 13-17). *Twenty years of research methods employed in "American Educational Research Journal," "Educational Researcher," and "Review of Educational Research."*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Emmons, N. J., Stallings, W. M., & Layne, B. H. (1990, Apr 16-20). *Statistical methods used in "American Educational Research Journal," "Journal of Educational Psychology," and "Sociology of Education" from 1972 through 1987*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Etzel, M., & Dickinson, K. (1999). *Digital Visual Fortran programmer's guide*. Boston: Butterworth-Heinemann.

Everitt, B. S. (1998). *The Cambridge dictionary of statistics*. Cambridge, UK: Cambridge University Press.

Fahoome, G. F. (1999). *A Monte Carlo study of twenty-one nonparametric statistics with normal and nonnormal data*. Unpublished Dissertation, Ph.D., Wayne State University, Detroit, MI.

Fahoome, G. F. (2002). Review of twenty nonparametric statistics and their large sample approximations. *Journal of Modern Applied Statistical Methods*, *1*(2), 248-268.

Fay, B. R. (2002). JMASM4: Critical values for four nonparametric and/or distribution-free tests of location for two independent samples. *Journal of Modern Applied Statistical Methods*, *1*(2), 489-517.

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*(4), 789-799.

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

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*(4), 521-532.

Fligner, M. A., & Policello, G. E., II. (1981). Robust rank procedures for the Behrens-Fisher problem. *Journal of American Statistical Association*, *76*, 162-174.

Freidlin, B., & Gastwirth, J. L. (2000). Should the median test be retired from general use? *The American Statistician*, *54*(3), 161-164.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of American Statistical Association*, *32*, 675-701.

Geary, R. C. (1947). Testing for normality. *Biometrika*, *34*, 209-242.

Gentle, J. E. (1998). *Random number generation and Monte Carlo methods*. New York: Springer-Verlag.

Gibbons, J. D. (1985). *Nonparametric methods for quantitative analysis* (2nd ed. Vol. 2). Columbus, Ohio: American Sciences Press, Inc.

Gibbons, J. D., & Chakraborti, S. (1991). Comparisons of the Mann-Whitney, Student's t, and alternate t tests for means of normal distributions. *Journal of Experimental Education*, *59*(3), 258-267.

Gibbons, J. D., & Chakraborti, S. (1992). *Nonparametric statistical inference* (3rd ed. Vol. 131). New York: Marcel Dekker, Inc.

Hájek, J., & Šidák, Z. (1967). *Theory of rank tests*. New York: Academic Press.

Hájek, J., Šidák, Z., & Sen, P. K. (1999). *Theory of rank tests* (2nd ed.). San Diego, CA: Academic Press.

Harwell, M. R. (1988). Choosing between parametric and nonparametric tests. *Journal of Counseling and Development, 67*(1), 35-38.

Harwell, M. R. (1990). A general approach to hypothesis testing for nonparametric tests. *Journal of Experimental Education*, *58*(2), 143-156.

Harwell, M. R. (1992). Summarizing Monte Carlo results in methodological research. *Journal of Educational Statistics*, *17*(4), 297-313.

Headrick, T. C., & Sawilowsky, S. S. (1999a). Errata: "Simulating correlated multivariate nonnormal distributions: extending the Fleishman power method." *Psychometrika*, *64*(2), 251.

Headrick, T. C., & Sawilowsky, S. S. (1999b). Simulating correlated multivariate nonnormal distributions: extending the Fleishman power method. *Psychometrika*, *64*(1), 25-35.

Headrick, T. C., & Sawilowsky, S. S. (2000). Weighted simplex procedures for determining boundary points and constants for the univariate and multivariate power methods. *Journal of Educational and Behavioral Statistics*, *25*(4), 417-436.

Hettmansperger, T. P. (1998). "The goals and strategies of robust methods": Comment. *British Journal of Mathematical and Statistical Psychology*, *51*(1), 41-42.

Hettmansperger, T. P., McKean, J. W., & Sheather, S. J. (2000). Robust Nonparametric Methods. *Journal of the American Statistical Association*, *95*(452), 1308-1312.

Hogben, L. (1957). *Statistical theory: the relationship of probability, credibility and error*. London: George Allen & Unwin, Ltd.

Hollander, M. (1963). A nonparametric test for the two-sample problem. *Psychometrika*, *28*(4), 395-403.

Hollander, M., & Wolfe, D. (1999). *Nonparametric statistical methods* (2nd ed.). New York: John Wiley & Sons, Inc.

Hotelling, H., & Pabst, M. R. (1936). Rank correlation and tests of significance involving no assumptions of normality. *Annals of Mathematical Statistics*, *7*, 29-43.

Iman, R. L., Quade, D., & Alexander, D. A. (1975). Exact probability levels for the Kruskal-Wallis test. *Selected Tables in Mathematical Statistics, Volume III* (1975, reprinted 1988), 329-384. Harter, H. L. & Owen, D. B., co-editors, Providence, RI: American Mathematical Society (edited by the Institute of Mathematical Statistics).

IMSL. (1997). Stat/Library: FORTRAN subroutines for statistical applications. Houston, TX: Visual Numerics, Inc.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, *38*, 324-329.

James, G. S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratio of the population variances are unknown. *Biometrika*, *67*, 85-92.

Jonckheere, A. R. (1954). A distribution free *k*-sample test against ordered alternatives. *Biometrika*, *41*, 133-145.

Kelley, D. L., Sawilowsky, S. S., & Blair, R. C. (1994, Oct 12-15). *Comparison of ANOVA, McSweeney, Bradley, Harwell-Serlin, and Blair-Sawilowsky tests in the balanced 2x2x2 layout*. Paper presented at the Annual Meeting of the Midwestern Educational Research Association, Chicago, IL.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*, 81-93.

Keselman, H. J., Cribbie, R., & Zumbo, B. D. (1997). Specialized tests for detecting treatment effects in the two-sample problem. *Journal of Experimental Education*, *65*(4), 355-366.

Killian, C. R., & Hoover, H. D. (1974, Apr). *An investigation of selected two-sample hypothesis testing procedures when sampling from empirically based test score models*. Paper presented at the 59th Annual Meeting of the American Educational Research Association, Chicago, IL.

Knuth, D. E. (1997). *The art of computer programming volume 1: Fundamental algorithms* (3rd ed.). Reading, MA: Addison Wesley Longman.

Knuth, D. E. (1998a). *The art of computer programming volume 2: Seminumerical algorithms* (3rd ed.). Reading, MA: Addison Wesley Longman.

Knuth, D. E. (1998b). *The art of computer programming volume 3: Sorting and searching* (2nd ed.). Reading, MA: Addison Wesley Longman.

Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, *4*, 83-91.

Kreitzberg, C. B., & Shneiderman, B. (1972). *The elements of FORTRAN style: Techniques for effective programming*. New York: Harcourt Brace Jovanovich, Inc.

Kruskal, W. H. (1952). A nonparametric test for the several sample problem. *Annals of Mathematical Statistics*, *23*, 525-540.

Kruskal, W. H. (1957). Historical notes on the Wilcoxon unpaired two-sample test. *Journal of American Statistical Association*, *52*, 356-360.

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

Law, V. J. (1983). *ANSI FORTRAN 77: An introduction to structured software design*. Dubuque, IA: Wm. C. Brown Company Publishers.

Legendre, A.-M. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes* (H. W. Henry A. Ruger, Trans.). Paris.

Lehmann, E. L. (1998). *Nonparametrics: Statistical methods based on ranks* (1st (Revised) ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

Lindquist, E. F. (1953). *Design and analysis of experiments in psychology and education*. Boston, MA: Houghton Mifflin.

MacDonald, P. (1999). Power, type I, and type III error rates of parametric and nonparametric statistical tests. *Journal of Experimental Education*, *67*(4), 369-379.

Mack, G. A., & Wolfe, D. A. (1981). k-sample rank tests for umbrella alternatives. *Journal of American Statistical Association*, *76*, 175-181.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*, 50-60.

May, R. B., & Konkin, P. (1970). A nonparametric test of an ordered hypothesis for k independent samples. *Educational and Psychological Measurement*, *30*(2), 251-257.

Mehta, C. R., & Patel, N. R. (1999). StatXact (Version 4.0). Cambridge, MA: Cytel Software Corporation.

Micceri, T. (1986, Nov). *A futile search for that statistical chimera of normality*. Paper presented at the Annual meeting of the Florida Educational Research Association 1986, Tampa, FL.

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

Mielke, P. W., Jr., & Berry, K. J. (2000). The Terpstra-Jonchkheere test for ordered alternatives: Randomized probability values. *Perceptual and Motor Skills*, *91*, 447-450.

Moses, L. E. (1952). A two-sample test. *Psychometrika*, *17*, 239-247.

Mosteller, F. (1948). A *k*-sample slippage test for an extreme population. *Annals of Mathematical Statistics*, *19*, 58-65.

Mosteller, F., & Tukey, J. W. (1950). Significance levels for a *k*-sample slippage test. *Annals of Mathematical Statistics*, *21*, 216-230.

Neave, H. R. (1981). *Elementary statistics tables*. London: Unwin Hyman Ltd.

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

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Odeh, R. E. (1971). On Jonckheere's k-sample test against ordered alternatives. *Technometrics*, *13*, 912-918.

Organick, E. I. (1966). *A FORTRAN IV primer*. Reading, MA: Addison-Wesley Publishing Company, Inc.

Pearson, K. (1895). Contributions to the mathematical theory of evolution: II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society*, *Ser. A, 186*, 343-414.

Penfield, D. A. (1994). Choosing a two-sample location test. *Journal of Experimental Education*, *62*(4), 343-360.

Porter, T. M. (1986). *The rise of statistical thinking, 1820–1900*. Princeton: Princeton University Press.

Puri, M. L. (1965). Some distribution-free k-sample rank tests for homogeneity against ordered alternatives. *Communications (on) Pure Applied Mathematics*, *18*, 51-63.

Puri, M. L., & Sen, P. K. (1993). *Nonparametric methods in multivariate analysis*. Malabar, FL: Krieger Publishing Company. (Reprint edition with substantial corrections. Original edition 1971.)

Rayner, J. C. W., & Best, D. J. (2001). *A contingency table approach to nonparametric testing*. Boca Raton, FL: Chapman & Hall / CRC.

Rosenbaum, S. (1953). Tables for a nonparametric test of dispersion. *Annals of Mathematical Statistics*, *24*, 663-668.

Rosenbaum, S. (1954). Tables for a nonparametric test of location. *Annals of Mathematical Statistics*, *25*, 146-150.

Rosenbaum, S. (1965). On some two-sample non-parametric tests. *Journal of American Statistical Association*, *60*, 1118-1126.

Rust, S. W., & Fligner, M. A. (1984). A modification of the Kruskal-Wallis statistic for the generalized Behrens-Fisher problem. *Communication (on) Statistical Theory Methods*, *13*, 2013-2028.

Savage, R. I. (1953). Bibliography of nonparametric statistics and related topics. *Journal of American Statistical Association*, *48*, 844-906.

Savage, R. I. (1962). *Bibliography of nonparametric statistics*. Cambridge, MA: Harvard University Press.

Sawilowsky, S. S. (1990). Nonparametric tests 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 (Psychologie Canadienne)*, *34*(4), 432-439.

Sawilowsky, S. S. (2002). Fermat, Schubert, Einstein, and Behrens-Fisher: The probable difference between two means when $\sigma_1^2 \neq \sigma_2^2$: *Journal of Modern Applied Statistical Methods, 1*(2), 461-472.

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

Sawilowsky, S. S., Blair, R. C., & Micceri, T. (1990). A PC FORTRAN subroutine library of psychology and education data sets. *Psychometrika*, *55*(4), 729.

Sawilowsky, S. S., & Brown, M. T. (1991). On using the t test on ranks as an alternative to the Wilcoxon test. *Perceptual and Motor Skills*, *72*(Part 1), 860-862.

Sawilowsky, S. S., & Hillman, S. B. (1991, Aug 16-20). *Sample size tables, "t" test, and a prevalent psychometric distribution*. Paper presented at the 99th Annual Meeting of the American Psychological Association, San Francisco, CA.

Sawilowsky, S. S., & Hillman, S. B. (1992). Power of the independent samples t test under a prevalent psychometric measure distribution. *Journal of Consulting and Clinical Psychology*, *60*(2), 240-243.

Sawilowsky, S. S., & Markman, B. S. (1989). *GUIL"t"*. Detroit, MI: Wayne State University.

Scheffé, H. (1943). Statistical inference in the nonparametric case. *Annals of Mathematical Statistics*, *14*, 305-332.

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

Serlin, R. C. (2000). Testing for robustness in Monte Carlo studies. *Psychological Methods*, *5*(2), 230-240.

Sheskin, D. J. (1997). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, Fl: CRC Press.

Siegal, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Siegal, S., & Tukey, J. W. (1960). A nonparametric sum of ranks procedure for relative spread in unpaired samples. *Journal of American Statistical Association*, *55*, 429-445.

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

Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin of Mathematics University of Moscow 2*, *2*, 3-14.

Smirnov, N. V. (1948). Tables for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, *19*, 280-281.

Smith, D. E. (Ed.). (1959). *A Source Book in Mathematics*. Mineola, N.Y.: Dover Publications, Inc.

Sparks, J. N. (1967). *Effects of Inapplicability of the continuity condition upon the probability distributions of selected statistics and their implications for research in education. Final report.* (RIE SYN71840): Pennsylvania State Univ., University Park.

Sprent, P., & Smeeton, N. C. (2001). *Applied nonparametric statistical methods* (3rd ed.). Boca Raton, Fl: Chapman & Hall / CRC.

Srisukho, D., & Marascuilo, L. A. (1974, Apr). *Monte Carlo study of the power of H-test compared to F-test when population distributions are different in form.* Paper presented at the Annual Meeting of the American Education Research Association, Chicago, IL.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677-680.

Stevens, S. S. (1968). Measurement, statistics, and the schemapiric view. *Science*, *161*(3844), 849-856.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: The Belknap Press of Harvard University Press.

Student [Gosset, W. S.] (1908). The probable error of a mean. *Biometrika*, *6*, 1-25.

Tadikamalla, P. R. (1980). On simulating non-normal distributions. *Psychometrika*, *45*(2), 273-279.

Terpstra, T. J. (1952). *A non-parametric k sample test and its connection with the H-test* (S-92 (VP2)). Amsterdam: Mathematical Center.

Tukey, J. W. (1959). A quick, compact, two-sample test to Duckworth's specifications. *Technometrics, 1*, 31-48.

van den Brink, W. P., & van den Brink, S. G. (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*(2), 183-189.

Vargha, A., & Delaney, H. D. (1998). The Kruskal-Wallis test and stochastic homogeneity. *Journal of Educational and Behavioral Statistics*, *23*(2), 170-192.

Vogt, P. W. (1999). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Wald, A., & Wolfowitz, J. (1940). On a test of whether two samples are from the same population. *Annals of Mathematical Statistics*, *11*, 147-162.

Wald, A., & Wolfowitz, J. (1943). An exact test for randomness in the non-parametric case based on serial correlation. *Annals of Mathematical Statistics*, *14*, 378-388.

Wald, A., & Wolfowitz, J. (1944). Statistical tests based on permutations of the observations. *Annals of Mathematical Statistics*, *15*, 358-372.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, *29*, 350-362.

Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing.* New York: Wiley.

Wilcox, R. R. (1991). Testing whether independent treatment groups have equal medians. *Psychometrika*, *56*(3), 381-395.

Wilcox, R. R. (1994). A one-way random effects model for trimmed means. *Psychometrika*, *59*(3), 289-306.

Wilcox, R. R. (1995). A paradigm for low power and misleading measures of effect size? *Review of Educational Research*, *65*(1), 51-77.

Wilcox, R. R. (1996). *Statistics for the social sciences*. San Diego, CA: Academic Press.

Wilcox, R. R. (1997). Some practical reasons for reconsidering the Kolmogorov-Smirnov test. *British Journal of Mathematical and Statistical Psychology*, *50*, 9-20.

Wilcox, R. R. (1998a). The goals and strategies of robust methods. *British Journal of Mathematical and Statistical Psychology*, *51*, 1-39.

Wilcox, R. R. (1998b). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, *53*(3), 300-314.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, *1*, 80-83.

Wilks, S. S. (1942). Statistical prediction with special reference to the problem of tolerance limits. *Annals of Mathematical Statistics*, *13*, 400-409.

Zimmerman, D. W. (1987). Comparative power of Student t test and Mann-Whitney U test for unequal sample size and variances. *Journal of Experimental Education*, *55*(3), 171-174.

Zimmerman, D. W. (1992). Failure of the Mann-Whitney test: A note of the simulation study of Gibbons and Chakraborti (1991). *Journal of Experimental Education*, *60*(4), 359-364.

Zimmerman, D. W. (1993). Increasing the power of rank tests by reducing the number of ranks. *Journal of Experimental Education*, *61*(3), 271-277.

Zimmerman, D. W. (1995). Increasing the power of nonparametric tests by detecting and downweighting outliers. *Journal of Experimental Education*, *64*(1), 71-78.

Zimmerman, D. W. (1998). Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. *Journal of Experimental Education*, *67*(1), 55-68.

Zimmerman, D. W. (1999). Type I error probabilities of the Wilcoxon-Mann-Whitney test and Student t test altered by heterogeneous variances and equal sample sizes. *Perceptual and Motor Skills*, *88*, 556-558.

Zimmerman, D. W., & Zumbo, B. D. (1989). A note on rank transformations and comparative power of the Student t-test and Wilcoxon-Mann-Whitney test. *Perceptual and Motor Skills*, *68*(3, Part 2), 1139-1146.

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*(1), 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. *Journal of General Psychology*, *117*(4), 425-436.

Zimmerman, D. W., & Zumbo, B. D. (1992a). Correction for nonindependence of sample observations in ANOVA "F" tests. *Journal of Experimental Education*, *60*(4), 367-381.

Zimmerman, D. W., & Zumbo, B. D. (1992b). Parametric alternatives to the Student t test under violation of normality and homogeneity of variance. *Perceptual and Motor Skills*, *74*(3, Part 1), 835-844.

Zumbo, B. D., & Zimmerman, D. W. (1991). *Levels of measurement and the relation between parametric and nonparametric statistical tests: A review of recent findings* (Working paper 91-1). Ottawa, CA: Edumetrics Research Group, University of Ottawa.

Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology (Psychologie Canadienne)*, *34*(4), 390-400.

Zwick, R. (1985). Nonparametric one-way multivariate analysis of variance: A computational approach based on the Pillai-Bartlett trace. *Psychological Bulletin*, *97*(1), 148-152.

ABSTRACT

A MONTE CARLO COMPUTER STUDY OF THE POWER PROPERTIES OF SIX DISTRIBUTION-FREE AND/OR NONPARAMETRIC STATISTICAL TESTS UNDER VARIOUS METHODS OF RESOLVING TIED RANKS WHEN APPLIED TO NORMAL AND NONNORMAL DATA DISTRIBUTIONS

by

BRUCE ROBERT FAY

December 2003

Advisor: Shlomo Sawilowsky, Ph.D.

Major: Education Evaluation & Research

Degree: Doctor of Philosophy

A study of the power properties of six nonparametric / distribution-free statistics was conducted via computer simulations using Monte Carlo techniques. These statistics utilized rank based procedures and shared an underlying assumption of population continuity such that samples were assumed to have no equal data values (zero difference scores, tied ranks). This assumption is almost never met in practice, however, especially with data in the social and behavioral sciences. The properties of such tests, in particular their power to detect real effects when they

exist, is altered under such conditions compared to their theoretical performance when underlying assumptions are met. The nature and extent of these departures are often mathematically intractable, depending on complex interactions of particular combinations and degrees of violation. These tests have not been as extensively studied under violation of their assumptions as they need to be. Both theoretical and empirical data distributions were used for sampling, including the normal distribution for reference. Four of the distributions described by Micerri (1986, 1989) as being typical of the data found in social and behavioral science were used as the source of discrete, nonnormal data. Pure location shift effects were introduced, with integral amounts of shift for the Micceri (1986, 1989) distributions in order to get equal data values between groups. A number of methods for dealing with this situation have been suggested in the literature, but there are few empirical studies. This study looked at various methods for dealing with equal data values (tied ranks). The best overall results, across all tests and combinations of simulation parameters, were obtained by randomly resolving ties. The method of dropping ties and reducing the sample size performed very poorly and should be avoided. The pattern and occurrence of ties was also investigated, along with sampling adequacy. Critical values and associated probabilities were also generated for four of the tests.

AUTOBIOGRAPHICAL STATEMENT

BRUCE ROBERT FAY

Bruce R. Fay is an assessment consultant with the Wayne County Regional Educational Service Agency in southeast Michigan, where he works with K-12 public school systems in the areas of assessment of student achievement, program evaluation, accreditation and accountability. Prior to joining Wayne RESA, Bruce taught high school mathematics and physics for nine years. He received his initial secondary certification in 1991 and his Master of Arts in Teaching in 1993, both from Wayne State University. He has also taught Philosophy of Education for Wayne State University as an adjunct faculty member. In 2001, Bruce received his certification in Reality Therapy, Choice Theory and Lead-management, from the William Glasser Institute.

Bruce owned a commercial / industrial photographic business for five years before deciding to become an educator. Prior to that he held management, marketing and engineering positions in several companies dealing with computer-based industrial electronics and process control. He has a Master of Science in Electrical Engineering and a Bachelor of Science in Electrical Engineering, both from the University of Missouri – Columbia. Before taking up the study of engineering, Bruce spent two and half years at the Oberlin College Conservatory of Music studying music composition, theory and the French Horn. Bruce was first introduced to computers and programming while at Oberlin. For his M.S.E.E. project he developed a control systems simulation program to model complex physical systems by simultaneously solving sets of differential equations.

Bruce is married with two children. His wife, Linda, is a C.P.A. and a graduate of the school of business at the University of Missouri – Columbia. His son, Brendan, holds a Bachelor of Arts degree in Mathematics and History of Art from the University of Michigan Ann Arbor and is currently pursuing a Ph.D. in History of Art at Harvard University in Cambridge, Massachusetts. His daughter, Meghan, just received a Bachelor of Arts degree in History and Women's Studies from the University of Michigan Ann Arbor and is presently working in Ann Arbor. Bruce and Linda have a Golden Retriever named Einstein and two cats, Copernicus and Tycho Brahe. When he isn't busy working or going to school, Bruce enjoys working on his house and traveling. He is a licensed private pilot, but no longer flies actively.