

A MONTE CARLO STUDY OF TWO
NONPARAMETRIC STATISTICS WITH
COMPARISONS OF TYPE I ERROR
RATES AND POWER

By

CHIN-HUEY LEE

Bachelor of Business Administration
National Chung-Hsing University
Taipei, Taiwan
1994

Master of Business Administration
Pittsburg State University
Pittsburg, KS
1997

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2007

A MONTE CARLO STUDY OF TWO
NONPARAMETRIC STATISTICS WITH
COMPARISONS OF TYPE I ERROR
RATES AND POWER

Dissertation Approved:

Dr. Janice Miller

Dissertation Adviser

Dr. Laura Barnes

Dr. Katye Perry

Dr. Diane Montgomery

Dr. A. Gordon Emslie

Dean of the Graduate College

ACKNOWLEDGEMENTS

I wish to express my sincere thanks to my dissertation advisor, Dr. Janice Miller, for her time and patience. My dissertation would not have been accomplished without her support and guidance. To Dr. Laura Barnes, my committee chair, thank you for taking the initiative to guide me to the REMS doctoral program. To Dr. Katye Perry, thank you for your consistent encouragement. I appreciate your willingness to help me think through issues regarding my life and my study. I really learn a lot from you. You are my mentor and friend. To Dr. Diane Montgomery, thank you for the positive encouragement and serving as my committee member.

Next, I acknowledge all my friends and colleagues in the REMS program, the College of Education, and the Department of Statistics. Without their encouragement, I would not have survived.

In addition, I would also extend my appreciation to my parents-in-law and brother-in-law for their thoughtful support. I also acknowledge my father. He makes me feel that I am important to him and encourages me to accomplish my goals. I also dedicate this effort to my mother.

Last but certainly not least, I want to express my sincere and deepest appreciation and love to my husband, Dr. Cheng-Chih Liu, for his patience, encouragement, and love through the highs and the lows of this endeavor. My two lovely children, John and Amy, thank you for your patience and being such wonderful companies through my years of study.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
Problem Statement	2
Purpose of Study	3
Research Questions	4
Significance of the Study	4
Definition of Terms	5
Assumptions	6
Restrictions	6
Organization of the Study	7
II. REVIEW OF LITERATURE	
Overview	8
Historical Framework of the Tests	8
Introduction	8
The Mann-Whitney Test	9
The Kolmogorov-Smirnov Two-Sample Test	10
Current Use of the MW test and the KS-2 Test in Research	11
Theoretical Development of the Tests	15
Introduction	15
The Mann-Whitney Test	15
Assumptions and Data Arrangements	16
Applicable Hypotheses	18
Formulas of the Test Statistic, Sample Size, and Decision Rules	20
Examples to Demonstrate the Calculation of Test Statistics	29
The Mann-Whitney Test Used in This Study	36
Selecting Sample Sizes	39
Issue of Ties	40
The Kolmogorov-Smirnov Two-Sample Test	41
Assumptions and Data Arrangements	41
Applicable Hypotheses	43
Formulas of the Test Statistic, Sample Size, and Decision Rules	44
Examples to Demonstrate the Calculation of Test Statistics	46
The Kolmogorov-Smirnov Two-Sample Test Used in This Study	49
Selecting Sample Sizes	52
Issue of Ties	52

Heterogeneity of Variance, Skewness and Kurtosis	54
Introduction.....	54
Heterogeneity of variance	54
Skewness and Kurtosis	55
Method of Selecting Population Distribution	57
Issues related to the Mann-Whitney test (MW test)	60
Issues related to the Kolmogorov-Smirnov two-sample test (KS-2 test)	75
Comparisons between the MW test and the KS-2 test.....	79
Summary	82
III. RESEARCH METHODOLOGY	
Introduction.....	84
Simulation Overview	85
Populations.....	86
Sampling	89
Test Statistics	95
Simulation Steps	101
Summary	103
IV. RESULTS	104
Introduction.....	104
Findings.....	104
Research Question 1	104
Research Question 2	125
Research Question 3	134
Research Question 4	137
Summary	145
V. DISCUSSIONS.....	147
Introduction.....	147
General Conclusions	148
Theoretical Implications	150
Sample Size.....	151
Heterogeneity of Variance	154
Difference in Skewness.....	155
Difference in Kurtosis.....	157
Practical Implications.....	158
Method to Simulate Statistical Power.....	158
Advice to Researchers.....	159
Limitations of the Study.....	161
Recommendations for Future Research	163
REFERENCES	166

APPENDICES	172
APPENDIX I: COEFFICIENTS OF FLEISHMAN’S POWER FUNCTION ..	173
APPENDIX II: E-MAIL CONTACT WITH CONOVER, W. J.	175
APPENDIX III: A SAMPLE OF SAS SYNTAX FOR GENERATING POPULATION DISTRIBUTIONS	176
APPENDIX IV: A SAMPLE OF SAS SYNTAX FOR SAMPLING PROCEDURE.....	177
APPENDIX V: HISTOGRAMS OF POPULATION DISTRIBUTIONS (N = 20000)	178
APPENDIX VI: TABLES OF FINDINGS.....	184

LIST OF TABLES

TABLE	PAGE
1: Table of Simulation Combinations	3
2: The use of the Mann-Whitney, the Kolmogorov-Smirnov two-sample test and other nonparametric statistical techniques	12
3: Table Values for the Kolmogorov-Smirnov two-sample test when sample sizes from either simple group are greater than 25	46
4: Coefficients used in Fleishmen's power function (1978) with $\mu = 0$; $\sigma = 1$	88
5: Summary of Conditions for Monte Carlo Simulations	92
6: Type I Error Rates: Only Sample sizes Differ between Two Samples (SD Ratio = 1)	106
7: Power of Normal Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$)	183
8: Power of Platykurtic Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$)	185
9: Power of Normal Platykurtic Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$)	187
10: Power of Leptokurtic 1 Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$)	189
11: Power of Leptokurtic 2 Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$)	191
12: Power of Leptokurtic 3 Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$)	193
13: Power of Skewed and Leptokurtic 1 Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$)	195
14: Power of Skewed and Leptokurtic 2 Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$)	197

LIST OF TABLES (CONT.)

TABLE		PAGE
15:	Power of Uniform-Like Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$).....	199
16:	Power of Logistic-Like Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$).....	201
17:	Power of Double Exponential-Like Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$).....	203
18:	Power of Skewed-Leptokurtic Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$).....	205
19:	Power of Skewed Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$).....	207
20:	Power of Skewed and Platykurtic 1 Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$).....	209
21:	Power of Skewed and Platykurtic 2 Distributions when sample sizes differ and SD Ratio $\neq 1$ ($\alpha = .05$).....	211
22:	Power of Normal Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$).....	213
23:	Power of Platykurtic Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$).....	214
24:	Power of Normal Platykurtic Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$).....	215
25:	Power of Leptokurtic_1 Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$).....	216
26:	Power of Leptokurtic_2 Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$).....	217
27:	Power of Leptokurtic_3 Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$).....	218
28:	Power of Skewed Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$).....	219

LIST OF TABLES (CONT.)

TABLE	PAGE
29: Power of Skewed and Platykurtic_1 Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)	220
30: Power of Skewed and Platykurtic_2 Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)	221
31: Power of Skewed and Leptokurtic_1 Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)	222
32: Power of Skewed and Leptokurtic_2 Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)	223
33: Power of Skewed-Leptokurtic Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)	224
34: Power of Uniform-Like Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)	225
35: Power of Logistic-Like Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)	226
36: Power of Double Exponential-Like Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)	227
37 Power; Only Skewness Ratios Are Different ($\gamma_{11} \neq \gamma_{12}$ & $\alpha = .05$)	135
38: Power; Only Kurtosis Ratios Are Different ($\gamma_{21} \neq \gamma_{22}$ & $\alpha = .05$)	138
39: Summary of the Conditions to Use the MW or the KS-2 Test	160

LIST OF FIGURES

FIGURE	PAGE
1: Power of the Normal Distribution when Sample Sizes Differ and SD ratios = 4 & 3	111
2: Power of the Normal Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4	112
3: Power of the Platykuritic Distribution when Sample Sizes Differ and SD ratios = 4 & 3	112
4: Power of the Platykuritic Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4	112
5: Power of the Normal Platykuritic Distribution when Sample Sizes Differ and SD ratios = 4 & 3	113
6: Power of the Normal Platykuritic Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4	113
7: Power of the Leptokurtic Distribution when Sample Sizes Differ and SD ratios = 4 & 3	113
8: Power of the Leptokurtic 1 Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4	114
9: Power of the Leptokurtic 2 Distribution when Sample Sizes Differ and SD ratios = 4 & 3	114
10: Power of the Leptokurtic 2 Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4	114
11: Power of the Leptokurtic 3 Distribution when Sample Sizes Differ and SD ratios = 4 & 3	115
12: Power of the Leptokurtic 3 Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4	115

LIST OF FIGURES (CONT.)

FIGURE	PAGE
13: Power of the Skewed and Leptokurtic 1 Distribution when Sample Sizes Differ and SD ratios = 4 & 3	115
14: Power of the Skewed and Leptokurtic 1 Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4	116
15: Power of the Skewed and Leptokurtic 2 Distribution when Sample Sizes Differ and SD ratios = 4 & 3	116
16: Power of the Skewed and Leptokurtic 2 Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4	116
17: Power of the Uniform-Like Distribution when Sample Sizes Differ and SD ratios = 4 & 3	117
18: Power of the Uniform-Like 1 Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4	117
19: Power of the Logistic-Like Distribution when Sample Sizes Differ and SD ratios = 4 & 3	117
20: Power of the Logistic-Like 1 Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4	118
21: Power of the Double Exponential -Like Distribution when Sample Sizes Differ and SD ratios = 4 & 3	118
22: Power of the Double Exponential -Like Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4	118
23: Histogram of the Skewed-Leptokurtic distribution (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score) ratios = 4 & 3	120
24: Power of the Skewed-Leptokurtic Distribution when Sample Sizes Differ and SD ratios = 4, 3, 2	120
25: Power of the Skewed-Leptokurtic Distribution when Sample Sizes Differ and SD ratios = 1/2, 1/3 & 1/4	120
26: Histogram of the Skewed distribution (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)	121

LIST OF FIGURES (CONT.)

FIGURE	PAGE
27: Histogram of the Skewed and Platykurtic_1 (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)	122
28: Histogram of the Skewed and Platykurtic_2 (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)	122
29: Power of Skewed Distribution when Sample Size Differ and SD ratios = 4, 3 & 2	122
30: Power of Skewed Distribution when Sample Size Differ and SD ratios = ½, 1/3, & 1/4	123
31: Power of Skewed and Platykurtic 1 Distribution when Sample Size Differ and SD ratios = 4, 3 & 2	123
32: Power of Skewed and Platykurtic 1 Distribution when Sample Size Differ and SD ratios = ½, 1/3, & 1/4	123
33: Power of Skewed and Platykurtic 2 Distribution when Sample Size Differ and SD ratios = 4, 3 & 2	124
34: Power of Skewed and Platykurtic 2 Distribution when Sample Size Differ and SD ratios = ½, 1/3, & 1/4	124
35: Power of the Normal Population when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$	126
36: Power of the Platykurtic Population when Only SD Ratios Are Different with Sample Size = (50, 50)and $\alpha = .05$	127
37: Power of the Normal Platykurtic Population when Only SD Ratios Are Different with Sample Size = (50, 50)and $\alpha = .05$	127
38: Power of the Leptokurtic_1 Population when Only SD Ratios Are Different with Sample Size = (50, 50)and $\alpha = .05$	127
39: Power of the Leptokurtic_2 Population when Only SD Ratios Are Different with Sample Size = (50, 50)and $\alpha = .05$	128
40: Power of the Leptokurtic_3 Population when Only SD Ratios Are Different with Sample Size = (50, 50)and $\alpha = .05$	128

LIST OF FIGURES (CONT.)

FIGURE	PAGE
41: Power of the Uniform-Like Population when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$	128
42: Power of the Logistic-Like Population when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$	129
43: Power of the Double Exponential-Like Population when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$	129
44: Histogram of the Skewed and Leptokurtic_1 (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)	130
45: Histogram of the Skewed and Leptokurtic_2 (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)	131
46: Histogram of the Skewed- Leptokurtic (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)	131
47: Power of the Skewed Population with ONLY SD Ratios Are Different and Sample Size = (50, 50) and $\alpha = .05$	131
48: Power of the Skewed and Platykurtic_1 Population with ONLY SD Ratios Are Different and Sample Size = (50, 50) and $\alpha = .05$	132
49: Power of the Skewed and Platykurtic_2 Population with ONLY SD Ratios Are Different and Sample Size = (50, 50), (25, 25) and $\alpha = .05$	132
50: Power of the Skewed and Leptokurtic_1 Population with ONLY SD Ratios Are Different and Sample Size = (50, 50) and $\alpha = .05$	132
51: Power of the Skewed and Leptokurtic_2 Population with ONLY SD Ratios Are Different and Sample Size = (50, 50) and $\alpha = .05$	133
52: Power of the Skewed-Leptokurtic Population with ONLY SD Ratios Are Different and Sample Size = (50, 50), (25, 25) and $\alpha = .05$	133
53: Histogram; Normal Population Distribution (Skewness = 0.00, Kurtosis = 0.00) and $\alpha = .05$	178
54: Histogram; Platykurtic Population Distribution (Skewness = 0.00, Kurtosis = 0.00) and $\alpha = .05$	178

LIST OF FIGURES (CONT.)

FIGURE	PAGE
55: Histogram; Normal Platykurtic Population Distribution (Skewness = 0.00, Kurtosis = -1.00) and $\alpha = .05$	179
56: Histogram; Leptokurtic_1 Population Distribution (Skewness = 0.00, Kurtosis = 1.00) and $\alpha = .05$	179
57: Histogram; Leptokurtic_2 Population Distribution (Skewness = 0.00, Kurtosis = 2.00) and $\alpha = .05$	180
58: Histogram; Leptokurtic_3 Population Distribution (Skewness = 0.00, Kurtosis = 3.75) and $\alpha = .05$	180
59: Histogram; Uniform-Like Population Distribution (Skewness = 0.00, Kurtosis = -1.20) and $\alpha = .05$	181
60: Histogram; Logistic-Like Population Distribution (Skewness = 0.00, Kurtosis = 1.30) and $\alpha = .05$	181
61: Histogram; Double Exponential-Like Population Distribution (Skewness = 0.00, Kurtosis = 3.00) and $\alpha = .05$	182

CHAPTER ONE

INTRODUCTION

In the educational and social behavioral sciences, the two-sample statistical comparison is one of the most important procedures in hypothesis testing. Based on the difference in the nature of the population distributions, many different parametric and nonparametric statistical tests are available to use under different assumptions (Buning, 2001). Most of the data related to research questions in the educational, social, and behavioral sciences are primarily ordinal in nature and distribution-free (Cliff & Keats, 2003; Keselman & Cribbie, 1997). Micerri (1989) investigated more than 400 large-sample data sets by performing parametric tests. He concluded that only 28.4% of the distributions in the educational or educational psychological fields were relatively symmetric, and that 30.7% were extremely asymmetric. Most of population distributions in those studies did not meet the assumption of normality. Authors of textbooks in education, psychology, and other related fields also recommend the use of nonparametric statistical tests when assumptions are violated, particularly, normality and homogeneity of variance (Zimmerman, 1998). Nonparametric statistics are more powerful than parametric statistics when the data are not normally distributed. Among various nonparametric statistical tests for comparing two populations, the Kolmogorov-Smirnov two-sample test (KS-2) and the Mann-Whitney test (MW) are the most often cited in nonparametric statistics textbooks published since 1956 (Fahoome &

Sawilowsky, 2000). The two tests are often in direct competition when a researcher is choosing an analytic technique for data analyses.

Research data are measured mostly with scales in the educational and behavioral sciences. Response options for the items for instruments are usually rank-ordered. Thus, these scores are at least ordinal in nature (Cliff & Keats, 2003). Both the KS-2 test and the MW test utilize ranks to analyze ordinal data (Conover, 1999; Daniel, 1990; Higgins, 2004; Krishnaiah & Sen, 1984; Pratt & Gibbons, 1981; Sheskin, 2000; Siegel & Castellan, 1988). They are both used to detect whether two independent samples are from the same population (Siegel & Castellan, 1988), or whether two populations for two independent samples are identical (Conover, 1999).

Problem Statement

When educational and social-behavioral researchers perform the MW test and the KS-2 test for the same data sets, the results for these two methods may remain the same under one condition. However, results may differ due to different shapes of the two population distributions, unequal population variances, or unequal sizes between two samples (Lee, 2005). However, there are limited studies that compare the MW test and the KS-2 test to determine the scenario of applying either one of these two nonparametric statistical techniques. Studies focusing on the conditions of population distributions, unequal population variances, and unequal sizes between two samples to evaluate Type I error rates and statistical power are in demand.

Purpose of Study

Even though both the KS-2 test and the MW test detect group differences, they may produce significantly different results with the same data sets. This may be due to differences in sizes between two samples, heterogeneity of variance, or skewness and kurtosis of population distributions. Therefore, the main purpose of this study was to compare the MW test with the KS-2 test through a Monte Carlo simulation. This study investigated conditions when these two tests produce different results, and thus different interpretations of the same data. The following considerations were assessed (Table 1):

1. Both equal and unequal sizes in large and small samples,
2. Heterogeneity of variance between two samples,
3. Different skewness between two samples,
4. Different kurtosis between two samples.

Similarities in Type I error rate and power were explored under these considerations, and overlapping characteristics were reported. Guidelines were developed to aid researchers' data analysis in applied educational settings.

Table 1: Table of Simulation Combinations

Sample Size	Equal	Unequal
Condition	Unequal population variance Difference in Skewness only Difference in Kurtosis only	Equal population variance Unequal population variance
Simulation	Statistical power	Type I error rates Statistical power

Research Questions

This study compared the Type I rates and power estimates of the KS-2 test and the MW test when performing a test under an alternative hypothesis that there were differences between two sampled population distributions. Thus, there were several research questions to guide the considerations of the research:

Question 1: If only sample sizes differ between two samples,

- a. Is there any difference in Type I error rate for these two nonparametric techniques?
- b. Is there any difference in power for these two nonparametric techniques?

Question 2: If only the heterogeneity of variance between two populations exists, is there any difference in power for these two nonparametric techniques?

Question 3: If the nature of the underlying population distributions varies in skewness only, is there any difference in power for these two nonparametric techniques?

Question 4: If the nature of the underlying population distributions varies in kurtosis only, is there any difference in power for these two nonparametric techniques?

Significance of the Study

This study was developed in order to provide guidelines for educational and social-behavioral researchers as they perform nonparametric data analyses. These results can help researchers to determine the nonparametric statistical methods they should adopt when choosing between the KS-2 test and the MW test under the specific conditions of concern.

Definition of Terms

1. Monte Carlo simulation: A procedure using random samples from known populations of simulated data to track a statistic's behavior (Mooney, 1997)
2. Nonparametric test: Defined as inferential statistical test that evaluates categorical/nominal data and ordinal/rank-order data (Sheskin, 2000).
3. Kolmogorov-Smirnov two sample test: A nonparametric statistical test that is employed with ordinal (rank-order) data in a hypothesis testing situation involving two independent samples (Sheskin, 2000)
4. Mann-Whitney test: A nonparametric statistical test that is employed with ordinal (rank-order) data in a hypothesis testing situation involving independent samples (Sheskin, 2000).
5. Type I error: The likelihood of rejecting a true null hypothesis (Sheskin, 2000).
6. Ties (in Rank): Equal values that are resolved by assigning each of the items which are ties the mean of the ranks they jointly occupy (Freund & Williams, 1966).
7. Power: Also called statistical power; a measure of the sensitivity of a statistical test; it is used to detect effects of a specific size by providing variance and sample size of a study. It is $1 - \beta$ (β is the Type II error rate) (Vogt, 2005).
8. Sample Size: The number of observations in a sample (Freund & Williams, 1966).
9. Variance: A measure of dispersion or the spread of scores in a distribution of scores (Vogt, 1993).
10. Skewness: The degree to which scores are clustered on one side of the central tendency and trail out on the other (Vogt, 2005).
11. Kurtosis: The relative peakedness or flatness of a distribution of scores (Vogt, 2005).

Assumptions

Both the KS-2 test and the MW test were used to assess general differences for two independent samples. It was assumed that samples for performing these two nonparametric statistical tests met their general assumptions (presented in CHAPTER TWO along with their formulas and test statistics). Furthermore, the tests performed one determined condition at a time (both equal and unequal sizes for large and small samples, heterogeneous variance, skewness, and kurtosis). The condition of ties was ignored in this study.

Restrictions

The following were identified restrictions of this research:

1. This research assessed only two independent sample comparisons using the MW test and the KS-2 test.
2. The simulations of this study were limited to
 - (1) Specific formulas of generating types of population distributions,
 - (2) The nominal Type I error rates (α) for comparisons at 0.5,
 - (3) Specific sizes for selecting samples from population distributions,
 - (4) Selected coefficients of skewness and kurtosis for generating population distributions,
 - (5) Ratios of variances between two simulated sample distributions,
 - (6) Formulas of test statistics for the KS-2 test and the MW test,
 - (7) Tied scores within and between sample distributions were ignored.

Organization of the Study

In summary, Chapter One provided an introduction of this study including statement of purpose, research questions, significance of the study, definition of terms, assumptions, as well as restrictions. Chapter Two introduces the review of literature, historical framework, theoretical developments including assumptions, data arrangement, and formulas related to the KS-2 test and the MW test. There were two examples used to demonstrate different methods of calculating test statistics of the MW test and the KS-2 test. The chapter also presents introductions of heterogeneity of variance, skewness, and kurtosis. Lastly, the method of selecting population distributions is described. Chapter Three proposes the research method and develops the statistical framework used in this study. Chapter Four presents the findings and results of the Monte Carlo simulations. Finally, Chapter Five summarizes the findings and discusses conclusions and implications. Recommendations are also described for both statistical theory and practice.

CHAPTER TWO

REVIEW OF LITERATURE

Overview

This chapter presents a review of literature related to this study. It includes: (1) historical framework of the tests: the Mann-Whitney test (MW test) and the Kolmogorov-Smirnov two-sample test (KS-2 test); (2) theoretical development of the tests: the Mann-Whitney test (MW test) and the Kolmogorov-Smirnov two-sample test (KS-2 test) which includes data definition, assumptions, hypotheses, and test statistics, sample size selection, and the issue of ties for both the MW test and the KS-2 test; (3) homogeneity of variance, skewness and kurtosis; (4) methods of selecting population distributions; (5) issues related to the MW test; (6) issues related to the KS-2; and (7) comparisons between the MW test and the KS-2 test.

Historical Framework of the Tests

Introduction

The Mann-Whitney test and the Kolmogorov-Smirnov two-sample tests are two of the nonparametric statistical techniques described in most nonparametric statistical and distribution-free textbooks. In order to have a better understanding of these two nonparametric techniques, the historical framework of the MW test and the KS-2 test is discussed in the next two sections. Lastly, the use of nonparametric statistical techniques including the MW and the KS-2 tests from 1995 to 2006 is summarized and tabulated to help readers observe how the MW and the KS-2 tests apply in current research.

The Mann-Whitney Test

Wilcoxon used a popular rank sum as a group test statistic under the condition of equal sample sizes in 1945 (Daniel, 1990). In 1947, Mann and Whitney proposed a slightly different version of the test that would apply to both equal and unequal sample sizes and provided tables for small sample sizes (Conover, 1999). Researchers (e.g., Gibbons & Chakraborti, 2003) mentioned that Mann-Whitney test is equivalent to the Wilcoxon rank-sum test since both tests employ ordinal data (rank-order) from independent and continuous population distributions. Siegel and Castellan (1988) called the test the Wilcoxon-Mann-Whitney test because Wilcoxon, Mann, and Whitney independently presented a nonparametric test with similar principles. Daniel (1990) referred to it as the Mann-Whitney–Wilcoxon test due to the equivalent statistical procedures between the Mann-Whitney and the Wilcoxon tests. Thus, the Mann-Whitney test was an improvement to the Wilcoxon test.

Although the MW test is one of the nonparametric techniques used to detect differences with the general two-sample problem under a null hypothesis of identical populations, Gibbons and Chajraborti (2003) as well as Neave and Worthington (1988) concluded that the MW test is most effective when testing the alternative hypothesis that the two populations are the same except for a difference between two location parameters. Freund and Williams (1966) defined location parameters as the parameters ones that attempt to locate the center of a population or a sample.

The Mann-Whitney (MW) test was explored in this study since Mann and Whitney offered similar versions of the test for both equal and unequal sample sizes and provided tables for small sample sizes.

The Kolmogorov-Smirnov Two-Sample Test

In 1933, Kolmogorov developed a one-sample goodness-of-fit test for ordinal data (Conover, 1999). A goodness-of-fit test is a statistical test to detect whether a model fits a data set or matches a theoretical expectation (Vogt, 2005). Sprent and Smeeton (2001) suggested that researchers should completely specify the underlying continuous distribution when performing the Kolmogorov goodness-of-fit test. Conover pointed out that the Kolmogorov one-sample test works well for goodness of fit when the sample size is small. In 1939, Smirnov modified the Kolmogorov's test and developed a nonparametric test for a two-sample scenario (Marascuilo & McSweeney, 1977). Conover referred to the KS-2 test as the Smirnov test even though it was an application of the Kolmogorov one-sample test. Daniel (1990) pointed out that the KS-2 test was developed by Smirnov but carried the name of Kolmogorov because of its similarities to the Kolmogorov one-sample test. Daniel (1990) proposed the KS-2 test as a general or omnibus test since "it is sensitive to differences of all types that may exist between two distributions." (p. 330) Higgins (2004) also proposed the KS-2 test as an omnibus test that is used to detect differences among sample groups despite the nature the differences. Siegel and Castellan (1988) defined the Kolmogorov-Smirnov two-sample test as:

A test of whether two independent samples have been drawn from the same population (or from populations with the same distribution). The two-tailed test is sensitive to any kind of difference in the distributions from which the two samples were drawn- difference in location (central tendency), in dispersion, in skewness, etc. The one-tailed test is used to decide whether or not the data values in the population from which one of

the samples was drawn are stochastically larger than the values of the population from which the other sample was drawn (p.144).

The Kolmogorov-Smirnov two-sample (KS-2) test was explored in this study since it is a well known nonparametric statistical technique.

In summary, both the MW test and the KS-2 test work in a similar way with the alternative hypothesis that there are differences between two sampled population distributions. The MW test seems to work efficiently when testing two populations with different locations. The KS-2 test is sensitive to general differences not only in location but also in variations and shapes of the distributions.

Current Use of the MW test and the KS-2 Test in Research

To explore how researchers have applied the MW test and the KS-2 test in their research, EBSCO Host Research Databases were assessed for five areas. Areas of reference included educational, psychological, educational psychology, social and behavioral, and health and medical related fields. The researcher selected these fields to examine articles that applied nonparametric statistical techniques. As shown in Table 1, the number of articles that used nonparametric statistical technique for analysis is extensive. Two thousand eight hundred twenty-eight full-text articles from peer-reviewed journals analyzed data with nonparametric statistical techniques from January 1995 to August 2006. Overall, 121 articles were located where researchers analyzed their data with the MW test, while the KS-2 test was used in 47 articles. Other nonparametric techniques were used in 2660 articles with analytical techniques for one-sample, two-sample, and multiple-sample situations. Examples of other nonparametric statistical techniques used in these studies were Chi-Square, the Sign test, the Spearman rank-order, and the Kruskal-Wallis. Over this eleven-year period, the MW test was

applied more often than the KS-2 test in those studies utilizing nonparametric statistical techniques. It should be noted that the MW and the KS-2 combined were used by researchers in these fields more often than all other nonparametric statistics.

Table 2: The Use of the Mann-Whitney (MW) and the Kolmogorov-Smirnov Two-sample (KS-2) Tests and Other Nonparametric Statistical Techniques, 1995-2006

Area of Interest	MW	KS-2	Other	Total
Educational	16	6	37	59
Psychological	10	0	8	18
Education	2	0	15	17
Psychological				
Social Behavioral	10	5	6	21
Health & Medical	83	36	2594	2713
Total (without Health & Medical)	38	11	66	115
Total (including Health & Medical)	121	47	2660	2828

*Examples of journals:

- Educational - Journal of Research in Music Education
Journal of Higher Education Policy and Management
- Psychological - Psychological Reports
Psychological Sciences & Social Sciences
- Education Psychological - British of Journal Educational Psychology
Applied Measurement in Education
- Social Behavioral - Humanities and Social Science
Sociological Methods and Research
- Health & Medical - American Journal of Health Promotion
Brain Research

When considering area of interest, Table 2 revealed that the MW test was applied more frequently in health and medical or medically related fields than in any other area. In other words, 83 out of 2713 health and medical articles utilized Mann-Whitney test. Thirty-eight articles that employed the MW test were from educational, psychological, education psychological, and social behavioral areas. Similarly, the KS-2 test was utilized in 36 articles from the health and medical journals. Only six articles applied the KS-2 technique in the educational field. There was no article using the KS-2 test in the psychological or educational psychological fields between winter 1995 and summer 2006. Overall, 11 journal articles using the KS-2 test are from the educational, psychological, educational psychological, and the social behavioral areas. Once again, when nonparametric statistics were used for statistical analyses, researchers in health and medical related fields applied these techniques more often than researches in the educational, psychological, education psychological, and social behavioral fields. As noted in the table, the choice of a nonparametric statistical technique appeared to depend on the area of research.

Most of the research articles reviewed here utilized the MW test by comparing the medians of two samples to detect whether there was any difference between two populations. Some articles applied the MW test to test group differences without specifying whether the median was used for the comparisons. There were some articles applying the MW test for simulations to test predetermined conditions for their hypotheses. Several articles used the KS-2 test to examine whether there was any general difference between two populations. The KS-2 test was utilized to assess whether or not there were differences in the shape of two population distributions. The KS-2 test was further applied to check whether two populations fitted each other. Simulation studies using the KS-2 test for evaluating

hypotheses were employed in some of the articles. One of the simulation studies modified the KS-2 test for multiple populations (two or more), while the other study was fully explored in a later section in this chapter, as it was directly related to this study.

The main similarities of the reviewed articles that applied either the MW test or the KS-2 test included small sample sizes in their data. Most of the sample sizes used for these two techniques were less than 30. Especially in the health and medical related articles, there were some articles with sample sizes of 10 or less. It has been suggested (Siegel & Castellan, 1988) that sample size be somehow the main consideration for using these two nonparametric techniques. When comparing two-group differences in locations, researchers tend to use the MW test. The KS-2 test is typically employed when making general difference comparisons when researchers want to see whether there is any difference between two populations in general.

In summary, Table 2 represented the occurrence of the MW test and the KS-2 test in current research using nonparametric statistical techniques. When examining the peer-reviewed journal articles in the educational, psychological, education psychological, and social behavioral areas, 38 out of 115 reviewed articles applied the MW test. However, only 11 out of these 115 peer-reviewed articles applied the KS-2 test in these fields. This raised the question why researchers seemed to favor employing the MW test rather than the KS-2 test. Furthermore, how do researchers chose between these two tests? When researchers decided to use a nonparametric statistical analysis, did they look at the nature of their research questions? Or did researchers think of the nature of their data? To explore these issues, it was necessary to explore the theoretical backgrounds of these two nonparametric

statistical techniques. Also of concern were the assumptions about the use of the MW test and the KS-2 test, and data definitions for applying these two techniques.

Theoretical Development of the Tests

Introduction

When performing a Monte Carlo simulation, the null model from which the random samples are drawn should be correctly specified. Further, formulas of test statistics should also be specified for the simulation. By examining several nonparametric statistics text books (Bradley, 1968; Conover, 1999; Daniel, 1990; Gibbons & Chakraborti, 2003; Higgins, 2004; Krishnaiah & Sen, 1984; Marascuilo & McSweeney, 1977; Pratt & Gibbons, 1981; Sheskin, 2000; Siegel & Castellan, 1988), formulas of statistics for both the MW test and the KS-2 test were reviewed. These are presented in the following sections, and they provide the parameters needed to conduct the Monte Carlo simulation.

The Mann-Whitney Test

The Mann-Whitney test may be applied to detect whether the two independent samples are drawn from two different populations when researchers measure their variables on at least ordinal scales. When applying the MW test, researchers should understand: 1) the assumptions about this test and the procedures of setting up data sets, 2) the types of hypotheses applicable, 3) the formulas for calculating test statistics, the definitions of sample sizes, and the decision rules for performing the test. This section presents various approaches from different textbook authors in order to help the researcher gain more in-depth understanding about the MW test. A further consideration includes 4) two examples (one small-sample and one large-sample example) to be presented to calculate test statistics

introduced by various textbook authors. The researcher also recommends 5) the Mann-Whitney test used in this study. Finally, discussions on 6) selecting sample sizes discussed by various textbook authors and 7) issues on ties related to the MW test are shown in this section too.

1) Assumptions and Data Arrangements

There are several assumptions that researchers should be aware of in order to perform the MW test. Based upon suggestions from Bradley (1968), Daniel (1990) and Conover (1999), they are as follows: First, each sample score is randomly selected from the population it represents. Then, two sample score sets are mutually independent or independent of one another. Lastly, the measurement employed is at least an ordinal scale.

Sheskin (2000) and Daniel (1990) proposed another assumption, that the originally observed variable is a continuous variable. In addition, Sheskin (2000) suggested, “the underlying populations from which the samples are derived are identical in shape” (p. 289). Daniel (1990) also pointed out that “the distributions of the two populations differ only with respect to location, if they differ at all” (p. 90). This study will adapt the suggestions from these previous researchers as assumptions for performing a Monte Carlo simulation on the MW test.

After developing data to meet the assumptions for performing the MW test, researchers should start to arrange the data set in order to calculate the test statistic of the statistical analysis. Daniel (1990) provided a way to configure the data set:

Let X_1, X_2, \dots, X_{n_1} denote the random sample scores size n_1 with unknown median M_x from population 1; let Y_1, Y_2, \dots, Y_{n_2} denote the random sample scores size n_2

with unknown median M_y from population 2; assign the ranks 1 to $n_1 + n_2$ to the observations from the smallest to the largest; let $N = n_1 + n_2$.

Conover (1999) used a slightly different presentation of data arrangement by removing the medians in each group. He specified the following:

Let X_1, X_2, \dots, X_{n_1} denote the random sample scores size n from population 1; let Y_1, Y_2, \dots, Y_{n_2} denote the random sample scores size m from population 2; assign the ranks 1 to $n_1 + n_2$ to the observations from the smallest to the largest; let $N = n_1 + n_2$. Also, $R(X_i)$ and $R(Y_j)$ are the ranks assigned to X_i and Y_j , where i is equal to 1, 2, ..., n_1 and j is equal to 1, 2, ..., n_2 .

Sheskin (2000), Marascuilo and McSweeney (1977), and Pratt and Gibbons (1981) provided a method similar to Conover's (1999), and specified using the sum ranks:

Let X_1, X_2, \dots, X_{n_1} denote the random sample scores size n_1 from population 1; let Y_1, Y_2, \dots, Y_{n_2} denote the random sample scores size n_2 from population 2; where the number of X 's is larger than the number of Y 's; assign the ranks 1 to $n_1 + n_2$ to the observations from the smallest to the largest; let $N = n_1 + n_2$ where:

$\sum R_1$ is the sum of the ranks of the sample of the first sample group.

$\sum R_2$ is the sum of the ranks of the sample of the second sample group.

From the assumptions and data arrangements provided by various researchers, this researcher found that Daniel (1990) focused the MW test on detecting location differences between two populations since he provided medians for both sample groups. Other researchers were more likely to use the MW test to determine whether the two samples were drawn from the different populations.

This study will use the method of data arrangement by Sheskin (2000), Marascuilo and McSweeney (1977), and Pratt and Gibbons (1981) which specified using the sum ranks.

2) Applicable Hypotheses

Vogt (2005) defined hypothesis as “a tentative answer to a research question” (p.146). This means that there is a statement of the relationship between the variables that are studied in a research question. There are two parts of any hypothesis: the null and alternative hypotheses. The null hypothesis is a statement to describe that there is no relationship between the populations or variables that researchers intend to compare. The alternative hypothesis is a statement to point out that there is a relationship between the populations or variables. If researchers do not specify the type or direction of difference between the populations or variables, the alternative hypothesis is a non-directional or two-tailed hypothesis. Since this study will detect general differences between two samples, a non-directional hypothesis is applied for the comparison.

When researchers set up their research questions and decide to apply the MW test for statistical analysis, they should state their null and alternative hypothesis in order to perform the statistical test. Various textbook authors propose different formats of null and alternative hypotheses based upon the way of arranging the data. For example, Daniel (1990), Marascuilo and McSweeney (1977), Siegel and Castellan (1988), and Sheskin (2000) used population medians to represent the relationship between two tested populations. As a consequence, the null and alternative hypotheses are:

Non-directional hypotheses:

Null hypothesis H_0 : $M_x = M_y$; or there is no difference between the medians of two populations.

Alternative hypothesis H_a : $M_x \neq M_y$; or there is a difference between the medians of two populations.

Where M_x is the median of the population associated with the variable X and M_y is the median of the population associated with the variable Y .

Conover (1999) and Gibbons and Chakraborti (2003) used the population distributions to express the hypothesis statements. Bradley (1968) also proposed similar hypothesis statements:

Non-directional hypotheses:

Null hypothesis H_0 : $F(x) = G(x)$ for all x ; or there are no differences between two populations.

Alternative hypothesis H_a : $F(x) \neq G(x)$ for some x ; or there are some differences between two populations.

Where $F(x)$ is the population distribution corresponding to the variable of X and $G(x)$ is the population distribution corresponding to the variable of Y .

In the format of alternative hypothesis, Conover (1999) pointed out that the MW test is also sensitive to test the mean differences between two populations. Therefore, the other way to express the alternative hypothesis is:

For non-directional test:

Null hypothesis H_0 : $E(X) = E(Y)$ or the mean of population X is not equal the mean of population Y .

Alternative hypothesis H_a : $E(X) \neq E(Y)$ or the mean of population X is not equal the mean of population Y .

In summary, there are several ways to express the null and alternative hypotheses; they are developed to serve different research interests and research questions. When researchers want to detect the general differences between two populations, they can use the method proposed by Bradley (1968), Conover (1999), and Gibbons and Chakraborti (2003). If researchers are more likely to detect the differences in location between two populations, the methods supplied by Bradley (1968), Daniel (1990), Marascuilo and McSweeney (1977), Pratt and Gibbons (1981), and Sheskin (2000) or the other form by Conover (1999) can be adapted. Moreover, when researchers want to testify whether the two ranked distributions have the same probability, the format by Gibbons and Chakraborti (2003) may be suggested. Siegel and Castellan (1988) concluded that Mann-Whitney test can be utilized with all three research questions.

This study seeks to detect the general difference between two populations; therefore, the non-directional alternative hypothesis will be applied to the research. The null and alternative hypotheses are:

Ho: $F(x) = G(x)$ for all x ; or there are no differences between two populations.

Ha: $F(x) \neq G(x)$ for some x ; or there are some differences between two populations.

3) Formulas of the Test Statistic, Sample Size, and Decision Rules

Vogt (2005) defined test statistics as “statistics used to test a finding for statistical significance” (p.323). Freund and Williams (1966) described the term test statistic as “a statistic on which the decision whether to accept or reject a given hypothesis is based” (p. 110). However, due to the different data arrangements and forms of hypotheses, different textbooks provide slightly different forms of the test statistic for a statistical test. Therefore, after hypotheses are formed, the next step is to find appropriate test statistics to either support

or refuse the hypotheses. The following four sections introduce different formats of test statistics for small and large samples that have been developed by textbook authors.

Conover's Test Statistic T

Conover (1999) used T and T_1 as the test statistic for evaluating hypotheses. He used one formula (T) when “there are no ties or just a few ties” (p.272) and the other (T_1) when “there are many ties” (p. 273). A tie means a situation when there are samples that have exactly the same values as other values. Conover (1999) suggested assigning the average of the ranks (called mid-rank) to all the equal values. The formula for the test statistic T with no or a few ties is:

$$T = \sum_{i=1}^{n_1} R(X_i), \text{ where the } R(X_i) \text{ is the rank associated with the variable } X_s \text{ in}$$

population 1.

The formula for the test statistic T_1 with many ties is:

$$T_1 = \frac{T - n_1 \frac{N+1}{2}}{\sqrt{\frac{n_1 n_2}{N(N-1)} \sum_{i=1}^N R_i^2 - \frac{n_1 n_2 (N+1)^2}{4(N-1)}}, \text{ where } \sum_{i=1}^N R_i^2 \text{ is the sum of the squares of all}$$

N of the ranks or average ranks actually used in both samples. $N = n_1 + n_2$.

The above two formulas apply as test statistics when both samples are no more than 20 ($n_1 \leq 20$ and $n_2 \leq 20$). Conover (1999) proposed another method to find the approximate p value by calculating the standard normal Z score for the test statistic T. If one uses a non-directional alternative hypothesis, the p-value for the situation of no ties is:

$$\text{p-value} = 2 \times P\left(Z \leq \frac{T + \frac{1}{2} - n_1 \frac{N+1}{2}}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}}\right), \text{ where } Z \text{ is a standard normal variable. Where}$$

the p-value is twice as small as the value of $2 \times P(Z \leq T)$ or $2 \times P(Z \geq T)$. When there are some ties, T is substituted with T_1 .

When either one of the sample with sizes is more than 20 ($n_1 > 20$ or $n_2 > 20$) and there are no ties in either sample, Conover (1999) proposed another formula for the large sample approximation. The test statistic formula is:

$$\omega_p \cong \frac{n_1(N+1)}{2} + z_p \sqrt{\frac{n_1 n_2 (N+1)}{12}}, \text{ where } Z_p \text{ is the standard } Z \text{ value with the}$$

associated upper quantile p (ω_p).

Freund and Williams (1966) explained the term, quantile, as “a value at or below which lies a given fraction of a set of data” (p. 44).

Conover (1999) used the same formula for both small and large samples with ties in the samples. The only difference for the large-sample situation is that T_1 is compared with the standard normal Z , not the table values as used with the small samples. The decision rule is to reject the null hypothesis with a fixed nominal type I error rate (α) if T or T_1 is less than the tabled quantile value (ω_p) with both e small and large sample sizes.

The issue of ties appears confusing in the literature. Further, there is no clear definition for “a few ties” and “many ties”. Conover (2005) suggested that if there are ties in the samples, researchers should always become conservative and use the formulas for the situation of ties. In addition, it is not clear about the definition of lower and upper quantiles. Moreover, previous researchers do not suggest using the formula of approximate p-value when there are small samples and the underlying population distributions are not normal. In

this study, therefore, the formula of approximate p-value proposed by Conover (1999) does not apply in any of the calculations in the examples to be demonstrated.

Daniel's Test Statistics T

Daniel (1990) proposed a different formula of test statistic for evaluating hypotheses compared to Conover (1999). He used information associated with population one to calculate the test statistic T. When there are ties in either sample, the mean of the rank will be assigned to the tied values. He claimed that no matter the size of the median (location parameter) between population one and two, “depending on the null hypothesis, either a sufficiently small or a sufficiently large sum of ranks assigned to sample observations from population one causes us to reject the null hypothesis” (p.91). The test statistic T when both samples are no more than 20 ($n_1 \leq 20$ and $n_2 \leq 20$) is:

$$T = S - \frac{n_1(n_1 + 1)}{2}; \text{ where } S \text{ is the sum of the ranks assigned to the samples from}$$

population 1, and n_1 is the sample size of group 1.

This formula is used no matter whether ties exist or not.

The decision rules for the small sample sizes with the (α) level of significance (nominal Type I error rate) is:

When the alternative hypothesis is non-directional ($H_a: M_x \neq M_y$), reject H_0 if the calculated T test statistic is less than the table value $w_{(\alpha/2)}$ or greater than $w_{1-(\alpha/2)}$ which is given by $n_1 n_2 - w_{(\alpha/2)}$

When either sample size is greater than 20 ($n_1 > 20$ or $n_2 > 20$) and there are no ties, the formula for the normal approximation is:

$$Z = \frac{T - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

When there are ties across groups, the formula may be adjusted under the denominator of the formula above by the correction for ties which is $\frac{n_1 n_2 (\sum t^3 - \sum t)}{12(n_1 + n_2)(n_1 + n_2 - 1)}$; where t is the number of ties for a given rank. Once, corrected for ties, the large-sample approximation formula is:

$$Z = \frac{T - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} - \frac{n_1 n_2 (\sum t^3 - \sum t)}{12(n_1 + n_2)(n_1 + n_2 - 1)}}$$

The decision rule is:

If the calculated absolute Z is greater than the tabled Z value at the $(\alpha/2)$ level, then reject the null hypothesis $H_0: M_x = M_y$. If the calculated absolute Z is greater than the tabled Z value with the α level, then reject the null hypothesis.

Daniel's T test statistic is suggested when the purpose of the research is to compare the location parameters between two populations. He did not mention whether the test statistics could be applied for determining the general differences between two populations. Dealing with the situation of ties under the small sample sizes, he did not provide any adjustment. Instead, Daniel (1990) (as cited in Noether (1967) used Noether's suggestion that "the adjustment has a negligible effect unless a large proportion of observations are tied or there are ties of considerable extent" (p. 93). However, Daniel is very conservative about the large sample approximation when ties exist across groups. Daniel suggested neglecting ties if they

exist within the same group since “there is no effect on test statistic when ties happen within groups” (p. 94).

Test Statistic U

Bradley (1968), Marascuilo and McSweeney (1977), Pratt and Gibbons (1981), and Sheskin (2000) proposed similar formulas of test statistics for evaluating hypotheses that are different from the test statistics proposed by Daniel (1990) and Conover (1999). They use U and U' (or U_1 and U_2) to derive the test statistics. When both sample sizes are less than or equal to 20 ($n_1 \leq 20$ and $n_2 \leq 20$) for Bradley (1968), Pratt and Gibbons (1981), and Sheskin (2000), or when both sample sizes are less than or equal to 10 ($n_1 \leq 10$ and $n_2 \leq 10$) for Marascuilo and McSweeney (1977), the test statistic is:

$\sum R_1$ is the sum of the ranks of the sample expected to have the smaller sum.

$\sum R_2$ is the sum of the ranks of the sample expected to have the larger sum.

$$U_1 = n_1 \times n_2 + \frac{n_1(n_1 + 1)}{2} - \sum R_1$$

$$U_2 = n_1 \times n_2 + \frac{n_2(n_2 + 1)}{2} - \sum R_2 \quad \text{or} \quad = n_1 \times n_2 - U_1$$

Where $U_1 + U_2 = n_1 \times n_2$;

The smallest U statistic is tested for significance.

The decision rule is:

If the observed U is less than or equal to the tabled U_{critical} ($U \leq U_{\text{critical}}$) at the specific level of significance (α), the null hypothesis is rejected. When this occurs, there is a significant difference between these two populations.

When sample sizes are large ($n_1 > 20$ or $n_2 > 20$), Sheskin (2000) suggested using a similar formula similar to Daniel's (1990). Marascuilo and McSweeney (1977) proposed the same formula when either or both sample sizes in the groups are greater than 10. The normal approximation is:

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

The decision rules are the same as the one proposed by Daniel's (1990). They are:

If the calculated absolute Z is greater than the tabled Z value at the $(\alpha/2)$ level, then the null hypothesis $H_0: M_x = M_y$ is rejected. If the calculated absolute Z is greater than the tabled Z value with the α level, then reject the null hypothesis.

Bradley (1968) and Pratt and Gibbons (1981) did not provide any adjustment for the situation of the existence of ties for either cases of small or large sample sizes. Instead, they suggested using the average rank (called mid-rank) method. The average rank method assigns each sample with the same average rank value and then applying this average rank value to the test statistic formulas. Marascuilo and McSweeney (1977) and Sheskin (2000) did not provide any adjustment to the situation of the existence of ties for the small sample test. When the sample sizes are large (greater than 20), they proposed applying a method similar to the method proposed by Daniel (1990). It is as follows:

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} \frac{n_1 n_2 [\sum (t_i^3 - t_i)]}{12(n_1 + n_2)(n_1 + n_2 - 1)}}$$

where t is the number of ties for a given rank. The decision rules for the statistical analysis are the same as the ones used by Daniel (1990). They are:

If the calculated absolute Z is greater than the tabled Z value at the $(\alpha/2)$ level, then the null hypothesis $H_0: M_x = M_y$ is rejected. If the calculated absolute Z is greater than the tabled Z value with the α level, then reject the null hypothesis.

The formulas for small samples proposed by Marascuilo and McSweeney (1977), Pratt and Gibbons (1981), and Sheskin (2000) are used to test whether two populations are the same except for a shift in location (median, or mean). Bradley (1968) provided the same formulas for the hypothesis of testing location differences between two populations and the one of testing whether the two populations are identical. Similarly, there is no clear description of when to apply the formulas for the existence of ties.

Test Statistic W

Siegel and Castellan (1988) pointed out that the Mann-Whitney test may be used to check whether two independent samples were drawn from the same population. The formulas for test statistics are somewhat similar to the ones proposed by Conover (1999). Test statistics W is used when the sample sizes are less than or equal to 10 ($n_1 \leq 10$ and $n_2 \leq 10$), and is as follows:

$W_x = \sum R_1$; the sum of the ranks of multiple variables of X s from population 1s

$W_y = \sum R_2$; the sum of the ranks of multiple variables of Y s from population 2s

$W_x + W_y = \frac{N(N+1)}{2}$, where $N = n_1 + n_2$

The smaller value of W_x and W_y is used as the test statistic.

The decision rule is:

If the probability of the observed W found in the Table As less than the specific level of significance (α), the null hypothesis is rejected and there is a significant difference between these two populations.

When the sample size is more than 10 ($n_1 > 10$ or $n_2 > 10$) or when one of the sample sizes is 3 or 4 and the other is more than 12, the formula for the normal approximation is used, which is:

$$Z = \frac{W_x \pm 0.5 - \frac{n_1(N+1)}{2}}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}}, \text{ where } W_x = \sum R_1.$$

Siegel and Castellan (1988) suggested assigning each tied value with the average rank (called mid-rank) and applying the test statistic formulas for the samples less than or equal to 10. If either or both samples are greater than 10, they suggested the following normal approximation formula:

$$Z = \frac{W_x \pm 0.5 - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2}{N(N-1)} \left[\frac{N^3(N-1)}{12} - \frac{\sum_{j=1}^g (t_j^3 - t_j)}{12} \right]}}, \text{ where } t_j \text{ is the number of the tied ranks in}$$

the j th grouping.

The decision rule is:

If calculated absolute Z is greater than the tabled Z value with the $\alpha/2$ level, then reject the null hypothesis.

Siegel and Castellan (1988) suggested that the test statistics be applied to investigate whether two independent samples have been drawn from the same population or whether the two populations have the same medians. The test statistics are also used to test whether the

probability of population X greater than population Y ($P(X>Y)$) is the same as the probability of population X less than population Y ($P(X<Y)$) which is equal to 0.5. On the issue of ties, Siegel and Castellan did not specify the minimum number of ties in order to use the formula for the ties situation.

4) Examples to Demonstrate the Calculation of Test Statistics

There were two examples designed by the researcher in order to present the different ways of calculating test statistics from the methods proposed by various textbook authors. These examples were presented to aid understanding and allow for a comparison of the differences among various formulas.

Example One: A Small Sample for Each Group

Score values were as follows,

Sample 1: 17, 36, 18, 40, 52; $n_1 = 5$

Sample 2: 15, 37, 23, 32, 43, 50; $n_2 = 6$

The research question was designed to detect whether these two samples were drawn from identical populations at the α level of 0.05. Thus, the null and alternative hypotheses were:

Ho: There was no difference between the two populations.

Ha: There was a difference between the two populations.

Data arrangement was as follows:

Score X	17	18		36	40	50	
$R_{n1}(x)$	2	3		6	8	10	
Score Y	15		23	32	37	43	52
$R_{n2}(x)$	1		4	5	7	9	11

$R_{n1}(x)$ and $R_{n2}(x)$ were the ranks assigned to X_i and Y_j , where i was equal to 1, 2, ..., n_1 and j was equal to 1, 2, ..., n_2 .

Calculations of the test statistics from four methods that were previously introduced in the “Formulas of the Test Statistic and Decision Rules” section were demonstrated:

Method 1: Conover’s Test Statistic T

$T = \sum_{i=1}^{n_1} R(X_i)$, where the $R(X_i)$ was the rank associated with the variable X scores in population 1.

$T = 2 + 3 + 6 + 8 + 10 = 29$; the quantile value at the α level of 0.05 ($\omega_{0.05}$) was 21.

The calculated test statistic T was not less than the tabled quantile value at the nominal Type I error rate (α) of 0.05. Therefore, the null hypothesis cannot be rejected.

Method 2: Daniel’s Test Statistic T

$T = S - \frac{n_1(n_1 + 1)}{2}$; where S was the sum of the ranks assigned to the samples from population 1.

$S = 2 + 3 + 6 + 8 + 10 = 29$

$T = 30 - \frac{5 \times (5+1)}{2} = 15$; the tabled quantile value of the two-tailed test at the α level of

0.05 ($\alpha_{0.05/2}$) was 4.

T was not less than the table value; therefore, the null hypothesis cannot be rejected.

Method 3: Test Statistic U

$\sum R_1$ was the sum of the ranks of the sample expected to have the smaller sum.

$\sum R_2$ was the sum of the ranks of the sample expected to have the larger sum.

$$U_1 = n_1 \times n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1$$

$$U_2 = n_1 \times n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2 \quad \text{or } = n_1 \times n_2 - U_1; \text{ where } U_1 + U_2 = n_1 \times n_2.$$

The smallest U statistic was tested for significance.

$$\sum R_1 = 2 + 3 + 6 + 8 + 10 = 29; \quad U \text{ (or } U_1) = 5 \times 6 + \frac{5 \times (5+1)}{2} - 29 = 16$$

$$\sum R_2 = 1 + 4 + 5 + 7 + 9 + 11 = 37; \quad U' \text{ (or } U_2) = 5 \times 5 + \frac{6 \times (6+1)}{2} - 37 = 16$$

$$\text{or } U' \text{ (or } U_2) = 5 \times 6 - 16 = 14$$

U = 14; the tabled quantile value at the α level of 0.05 ($\alpha_{0.05/2}$) equals 4.

T was not less than the table value; therefore, the null hypothesis cannot be rejected.

Method 4: Test Statistic W

$W_x = \sum R_1$; it was the sum of the ranks of variables of Xs from population 1.

$W_y = \sum R_2$; it was the sum of the ranks of variables of Ys from population 2.

$$W_x + W_y = \frac{N(N+1)}{2}, \text{ where } N = n_1 + n_2$$

$$W_x = \sum R_1 = 2 + 3 + 6 + 8 + 10 = 29$$

$$W_y = \sum R_2 = 1 + 4 + 5 + 7 + 9 + 11 = 37$$

The test statistic W was 29; its p -value from the table provided by Siegel and Castellan (1988) was 0.5346. Since the nominal Type I error rate (α) was 0.05, the null hypothesis cannot be rejected.

In conclusion, by comparing the results of the tests from the different calculations of the test statistic, the same conclusion was reached with all tests in this example. Therefore, it may be that the same result will occur from the different calculations by various textbook authors. It appears that researchers can decide to use the test statistic that best suits their research needs.

Example Two: A Large Sample for Either Group

Score values were as follows,

Sample 1: 8, 17, 36, 18, 40, 52, 38, 59, 31, 68; $n_1 = 10$

Sample 2: 3, 15, 25, 48, 37, 65, 6, 57, 42, 35, 11, 23, 32, 43, 50, 51, 62, 74, 20,
69, 44, 9, 39, 47, 66, 55; $n_2 = 26$

The research question was developed to detect whether these two samples were drawn from the identical populations at the α level of 0.05. Thus, the null and alternative hypotheses were:

H_0 : There was no difference between two the populations.

H_a : There was a difference between two the populations.

Data arrangement was as follows:

X	8		17	18			31		36	38				
R _{n1} (x)	3		7	8			12		15	17				
Y	3	6	9	11	15		20	23	25	32	35	37	39	
R _{n2} (x)	1	2	4	5	6		9	10	11	13	14	16	18	
X	40						52	59				68		
R _{n1} (x)	19						27	29				34		
Y	42	43	44	47	48	50	51	55	57	62	65	66	69	74
R _{n2} (x)	20	21	22	23	24	25	26	28	30	31	32	33	35	36

R_{n1}(x) and R_{n2}(x) were the ranks assigned to X_i and Y_j, where i was equal to 1, 2, ..., n₁ and j was equal to 1, 2, ..., n₂. They were:

$$R(x_1) = 3 + 7 + 8 + 12 + 15 + 17 + 19 + 27 + 29 + 34 = 171$$

$$R(x_2) = 1 + 2 + 4 + 5 + 6 + 9 + 10 + 12 + 13 + 14 + 16 + 18 + 20 + 21 + 22 + 23 + 24 + 25 + 26 + 28 + 30 + 31 + 32 + 33 + 35 + 36 = 496$$

Method 1: Conover's Test Statistic T

$T = \sum_{i=1}^n R(X_i)$, where the R(X_i) was the rank associated with the variable X scores in population 1.

$$T = 171$$

$$\omega_p \cong \frac{n_1(N+1)}{2} + z_p \sqrt{\frac{n_1 n_2 (N+1)}{12}}, \text{ where } Z_p \text{ was the standardized Z value with the}$$

associated upper quantile p.

$$n_1 = 10, n_2 = 26, N = 10 + 26 = 36$$

$$\begin{aligned}\omega p &\cong \frac{n_1(N+1)}{2} + z_p \sqrt{\frac{n_1 n_2 (N+1)}{12}} = \frac{10(36+1)}{2} + z_p \sqrt{\frac{10 \times 26 \times (36+1)}{12}} \\ &= 185 + 1.96 \times 28.32 \\ &\approx 240\end{aligned}$$

T was less than ωp , so the null hypothesis was retained.

Method 2: Daniel's Test Statistic T

$$Z = \frac{T - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}, \text{ and } T = S - \frac{n_1(n_1 + 1)}{2}; \text{ where } S \text{ was the sum of the ranks}$$

assigned to the samples from population 1.

$$S = 171$$

$$T = S - \frac{n_1(n_1 + 1)}{2} = 171 - \frac{10 \times (10 + 1)}{2} = 116$$

$$Z = \frac{T - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} = \frac{116 - \frac{10 \times 26}{2}}{\sqrt{\frac{10 \times 26 \times (10 + 26 + 1)}{12}}} = \frac{-14}{28.32} \approx -0.49.$$

From the standard normal Z table, the probability (p-value) of $Z \geq -0.49$ was about 0.6879.

The p-value was greater than the level of significance ($\alpha = .05$), thus, the null hypothesis cannot be rejected.

Method 3: Test Statistic U

$$U \text{ (or } U_1) = n_1 \times n_2 + \frac{n_1(n_1 + 1)}{2} - \sum R_1 = 10 \times 26 + \frac{10 \times (10 + 1)}{2} - 171 = 144$$

$$U' \text{ (or } U_2) = n_1 \times n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2 = n_1 \times n_2 - U \text{ (or } U_1) = 10 \times 26 - 144$$

$$= 116$$

$$U = 116$$

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} = \frac{116 - \frac{10 \times 26}{2}}{\sqrt{\frac{10 \times 26 \times (10 + 26 + 1)}{12}}} = \frac{-14}{28.32} \approx -0.49.$$

From the standard normal Z table, the probability (p-value) of $Z \geq -0.53$ was about 0.6879.

The p-value was greater than the level of significance ($\alpha = .05$), so the null hypothesis cannot be rejected.

Method 4: Test Statistic W

$$Z = \frac{W_x \pm 0.5 - \frac{n_1(N+1)}{2}}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}}; \text{ where } W_x = \sum R_1.$$

$$Z = \frac{171 \pm 0.5 - \frac{10 \times (36+1)}{2}}{\sqrt{\frac{10 \times 26 \times (36+1)}{12}}} = \frac{-14 \pm 0.5}{28.32} \quad Z \approx -0.48 \text{ or } -0.51$$

From the standard normal Z table, the probability (p-value) of $Z \geq -0.48$ or $Z \geq -0.51$ was about 0.6844 and 0.6950, respectively. The p-value was greater than the level of significance ($\alpha = .05$); therefore, the null hypothesis cannot be rejected.

Again, from this large-sample example, in comparing the results of the tests from the different calculations of the test statistic, the same conclusion was reached with all tests in

this example. Therefore, it was shown that researchers may use any formula for the normal large-sample approximation introduced by various textbook authors to obtain the same result.

5) The Mann-Whitney Test Used in This Study

Based upon the conclusion provided by Siegel and Castellan (1988), The MW test can be used to test the general difference between two populations, the location of the two populations (means or median), and the equivalent probabilities of the two populations.

Presented below are the summarized and modified: *1) assumptions and data arrangements, 2) hypotheses, and 3) formulas of test statistics and decision rules for small and large sample sizes* for the MW test that were used for this study.

1) Assumptions and Data Arrangements

The assumptions for applying the MW test were as follows:

- (1) Each sample score has been randomly selected from the population it represents.
- (2) The originally observed sample score was a continuous variable.
- (3) Two sample scores were randomly selected and score sets were mutually independent.
- (4) The measurement scale employed was at least ordinal.

Data Arrangement shows the expression of arranging data after we get the data sets were obtained for use with the MW test technique.

Let X_1, X_2, \dots, X_{n_1} denote the random sample scores size n_1 with an expected smaller sum of ranks.

Let X_1, X_2, \dots, X_{n_2} denote the random sample scores size n_2 with an expected larger sum of ranks.

Assign the ranks 1 to $(n_1 + n_2)$ to the observations from the smallest to the largest.

Let $N = n_1 + n_2$.

2) *Applicable Hypotheses*

Because this research was designed to detect the alternative hypothesis that there were differences between two sampled population distributions, the non-directional hypothesis (two-tailed test) of the test was:

H_0 : $F(x) = G(x)$ for all x ; or there was no difference between the two populations.

H_a : $F(x) \neq G(x)$ for some x ; or there were some differences between the two populations.

Where $F(x)$ was the population distribution function of the sum of the ranks of the sample expected to have the smaller sum, and $G(x)$ the population distribution function of the sum of the ranks of the sample expected to have the larger sum.

3) *Formulas of Test Statistics and Decision Rules for Small and Large Sample Sizes*

Test statistics are used to calculate the value needed to perform the hypothesis test. Because of the ease of understanding and calculating the formula as well as consistent with the procedure in SAS/NPAR1WAY, the test statistic used in this research is adapted from the Test statistics W method proposed by Siegel and Castellan (1988).

Small Sample Size in each group ($n_1 \leq 20$; $n_2 \leq 20$)

$W_x = \sum R_i$; the sum of the ranks of multiple variables of X_s from population 1s

$W_y = \sum R_2$; the sum of the ranks of multiple variables of Ys from population 2s

$$W_x + W_y = \frac{N(N+1)}{2}, \text{ where } N = n_1 + n_2$$

The smaller value of W_x and W_y is used as the test statistic.

The decision rule is:

If the probability of the observed W found in the table is less than the specific level of significance (α), the null hypothesis is rejected and there is a significant difference between these two populations.

When the sample size is more than 20 ($n_1 > 20$ or $n_2 > 20$), the formula for the normal approximation is used, which is:

$$Z = \frac{W_x \pm 0.5 - \frac{n_1(N+1)}{2}}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}}, \text{ where } W_x = \sum R_1.$$

The decision rule is:

If calculated absolute Z is greater than the tabled Z value with the $\alpha/2$ level, then reject the null hypothesis.

Siegel and Castellan (1988) suggested that the test statistics be applied to investigate whether two independent samples have been drawn from the same population or whether the two populations have the same medians. The test statistics are also used to test whether the probability of population X greater than population Y ($P(X>Y)$) is the same as the probability of population X less than population Y ($P(X<Y)$) which is equal to 0.5. On the issue of ties, Siegel and Castellan did not specify the minimum number of ties in order to use the formula for the ties situation.

In summary, the researcher suggests the following steps to execute the MW test. First, give two sample score sets, X with the size of n_1 and Y with the size of n_2 , with $N = n_1 + n_2$. Second, combine the observations from these two groups into a single group, and then assign the rank from one to N to the observation from the smallest to the largest. Third, let R_1 represent the smaller sum of the ranks of the observations for the first group, and let R_2 serve as the larger sum of the ranks of the observations for the second group. Fourth, use formulas to calculate the test statistic or p-value. Fifth, use the tabled value or calculate the p-value and detect whether the test statistic reaches the level of significance. Lastly, draw conclusions based on the findings in step five.

6) Selecting Sample Sizes

This section introduced methods of selecting sample sizes by researchers who proposed test statistics for investigating the null hypothesis in section three formulas of the test statistic and decision rules. Various textbooks provided tables with different pairs of equal and unequal sample sizes and the associated critical values used to assess statistical significances. Neave and Worthington (1988) provided critical values for all sample size combinations up to 25 per group. Marascuilo and McSweeney (1977) provided critical values tables for equal and unequal sample size groups from (1, 1) to (10, 10). Siegel and Castellan (1988) included lower and upper-tail probability of W_x for sample size groups from (3, 3) to (10, 10). Conover (1999) and Daniel (1991), Bradley (1968), Pratt and Gibbons (1981), and Sheskin (2000) provided critical values tables for equal and unequal sample size groups from (2, 2) to (20, 20). Due to different formulas used in calculating test statistics, the critical values were slightly different for each formula. Therefore, when researchers decided on the formulas for

calculating test statistics, it was appropriate to adopt the associated critical value table for the sample sizes in order to perform the statistical analysis.

Samples in this research included small ($n_1 \leq 20$; $n_2 \leq 20$) and large ($n_1 > 20$; $n_2 > 20$) sizes, with equal and unequal conditions. The specific sizes of both samples were introduced in CHAPTER THREE.

7) Issue of Ties

Tied scores are always an issue for all nonparametric statistics. The issue of ties, as pointed out from researchers who proposed formulas of test statistics in Formulas of the Test Statistic and Decision Rules section, must be solved for this study. When there were some samples with the same values (ties that occur in the same sample), or when ties existed between two samples, researchers such as Conover (1999), Bradley (1968), and Pratt and Gibbons (1981) suggested assigning an average of the ranks (mid-rank) to those observations. However, Siegel and Castellan (1988), Neave and Worthington (1988), and Conover (1999) pointed out that variability in the sets of ranks are affected by tied ranks. They suggested using a formula for tie correction as a compromise to the problem. However, no researcher clearly defined when to use the test statistic formulas of tied conditions. Conover (1999) used the phrase of “if there are many ties” for the situation to use the test statistic of ties without quantifying the “many”. Others used “when there are ties” in applying the test statistic. In personal communication with Conover (2005), Conover suggested when ties existed, the formulas for “many ties” should be used especially if various number of ties were manipulated during the simulation process. Due to a lack of clarity among the definitions of ties for the various authors, this study did not address this issue of ties. In other words, tied scores were not considered in this study.

The Kolmogorov-Smirnov Two-Sample Test

The Kolmogorov-Smirnov two-sample test (the KS-2 test) is one of the nonparametric statistical techniques for comparing two sample cumulative distribution functions to detect whether there are any differences between two population distributions for two samples Conover (1999). Daniel (1990) and Higgins (2004) wrote that the KS-2 test was also referred to as a general or omnibus test for testing whether the populations of two independent samples were identical. Siegel and Castellan (1988) and Marascuilo and McSweeney (1977) also concluded that when any non-directional alternative hypothesis was tested, the KS-2 test was sensitive to any distributional difference.

When conducting the KS-2 test, researchers should understand: 1) the assumptions about this test and the procedures of setting up data sets, 2) the types of applicable hypotheses, 3) the formulas of calculating test statistics and the definitions of sample sizes, and the decision rules of performing the test. This section presented various approaches from different textbook authors in order to help the researcher understand more about the KS-2 test. A further consideration included 4) two examples (one small-sample and one large-sample example) to calculate test statistics of the KS-2 test introduced by various textbook authors. The researcher also recommends 5) the KS-2 test used in this study. Finally, discussions on 6) selecting sample sizes as considered by various textbook authors and 7) the issues of ties related to the KS-2 test.

1) Assumptions and Data Arrangements

There were some assumptions that researchers should be sensitive about in order to perform the KS-2 test. Based upon the suggestions from Conover (1999), they are as follows:

- (1) Each sample has been randomly selected from the population it represented.
- (2) The measurement scale employed is at least ordinal.
- (3) Two samples are mutually independent.
- (4) The originally observed variable is a continuous variable.

Bradley (1968) also revealed that the sizes of sampled populations are infinite and no tied observations occurred in the samples. Marascuilo and McSweeney (1977) also pointed out that in order to eliminate tied observations, the continuity on variables was necessary. Daniel (1990) and Sheskin (2000) only assumed that samples were independent and random, and the data were measured on at least ordinal scale.

After identifying the assumptions of applying the KS-2 test, researchers should know how to define the data set in order to perform this test. Daniel (1990), Conover (1999), and Sheskin (2000) defined the data in the following ways.

- (1) Let $S_1(x)$ be the empirical distribution function based upon the random sample scores of X_1, X_2, \dots, X_{n_1} .
- (2) Determine the cumulative probabilities for each value of X_1, X_2, \dots, X_{n_1} .
- (3) Let $S_2(x)$ be the empirical distribution function based upon the random sample scores of Y_1, Y_2, \dots, Y_{n_2} .
- (4) Determine the cumulative probabilities for each value of Y_1, Y_2, \dots, Y_{n_2} .

Daniel (1990) also defined $S_1(x)$ and $S_2(x)$ as:

$$S_1(x) = \frac{(\text{number of observed X's} \leq x)}{n_1}, \text{ and } S_2(x) = \frac{(\text{number of observed Y's} \leq y)}{n_2}.$$

Siegel and Castellan (1988) had similar data definition but specified $S_1(x)$ and $S_2(x)$ to be the cumulative distribution:

$$\text{Let } S_1(x) = \frac{K}{n_1} \text{ and } S_2(x) = \frac{K}{n_2}, \text{ where } K \text{ is the number of data less than or equal to } X$$

in the first sample set and Y in second sample set.

Higgins (2004) had the same definitions but changed $S_1(x)$ to $F_1(W)$ and $S_2(x)$ to $F_2(W)$.

These assumptions and data arrangement as they related to this study are discussed in a later section “The Kolmogorov-Smirnov two-sample Test Used in This Study”.

2) Applicable Hypotheses

When researchers determined their research questions and perform the hypothesis tests, the first step is to define the null and alternative hypotheses relating to the research questions. Marascuilo and McSweeney (1977), Daniel (1990), Conover (1999), and Sheskin (2000) proposed similar formats of non-directional alternative hypotheses. They were shown below:

Non-directional (two-sided) test:

Ho: $F(x) = G(x)$ for all x ; from $-\infty$ to $+\infty$

Ho: There are no differences between two populations.

Ha: $F(x) \neq G(x)$ for at least one value of x ;

Ha: There are some differences between two populations.

Marascuilo and McSweeney (1977) and Conover (1999) explained that the hypothesis test detects the general difference between two populations. Once the null hypothesis was

rejected, the difference would be between the location parameter (mean or median), the scale parameter (standard deviation), the skewness, or kurtosis.

This research investigated the general differences between two populations, and did not compare whether one was superior the other. Therefore, non-directional hypotheses were applied to the study, that was:

Ho: $F(x) = G(x)$ for all x ; from $-\infty$ to $+\infty$; or there is no difference between the two populations.

Ha: $F(x) \neq G(x)$ for at least one value of x ; or there are some differences between the two populations.

3) Test Statistics and Decision Rules for the Testing the Hypotheses

When executing the hypothesis test, the most important step was to calculate the test statistic and determine whether the null hypothesis was rejected or retained. Hence, this section presented the formulas for test statistics and the decision rules to test the hypotheses. Neave (1988) stated that “the Kolmogorov-Smirnov method uses the maximum vertical difference between two cumulative population distribution functions (cdf’s) as the test statistics” (p. 149). Higgins (2004) explained that “the Kolmogorov-Smirnov statistic was the maximum absolute value of the difference between the two sample cdf’s.” (p. 57) Neave and Higgins proposed the same method to find the test statistic for the Kolmogorov-Smirnov method. They both used the maximum absolute difference between two cumulative sample distribution functions as the test statistic. Bradley (1968), Conover (1999), Daniel (1990), and Siegel and Castellan (1988) all pointed out that this test can be used with both equal and unequal sample sizes. Textbooks reviewed by the researcher presented similar format of the test statistics; therefore, only one demonstration of the test statistic in small and large sample

sizes found from various textbooks was presented in this section. The test statistics for both small sample sizes (n_1 or n_2 was no more than 25) and large-sample sizes were provided as follows:

Small Sample Size ($n_1 \leq 25$ and $n_2 \leq 25$)

No matter whether the two samples are equal or unequal, when the sample sizes are less than or equal to 25 in both groups ($n_1 \leq 25$ and $n_2 \leq 25$), the test statistic is presented as:

D_{n_1, n_2} is the maximum absolute difference between the two empirical distribution functions or cumulative distribution functions, and

$$D_{n_1, n_2} = \max_x |S_{n_1}(x) - S_{n_2}(x)|$$

The decision rule for the hypothesis test is:

If the observed $D_{m, n}$ is greater than or equal to the tabled $D_{n_1, n_2 \text{ critical}}$ ($D_{n_1, n_2} \geq D_{n_1, n_2 \text{ critical}}$) at the specific level of significance (α), the null hypothesis is rejected.

Therefore, there is a significant difference between these two populations.

Large Sample Size ($n_1 > 25$ or $n_2 > 25$)

Textbooks reviewed by this researcher proposed similar format of the test statistics. The test statistic when either one sample or both samples are larger than twenty five ($n_1 > 25$ or $n_2 > 25$) is shown as:

$$D_{n_1, n_2} = \max_x |S_{n_1}(x) - S_{n_2}(x)|$$

Critical D_{n_1, n_2} is calculated with a formula based on various significance level (α).

When the significance level is α , the critical value is:

$$D_{n_1, n_2} = \text{table value (K)} \times \sqrt{\frac{n_1 + n_2}{n_1 n_2}}.$$

Table value (K) is displayed in Table 3, shown below.

Table 3: Table Values for the Kolmogorov-Smirnov two-sample test when sample sizes from either simple group are greater than 25

Two-tailed	Significance Level (α)				
	0.20	0.10	0.05	0.02	0.01
Table Value (K)	1.07	1.22	1.36	1.52	1.63

The decision rule for the hypothesis test is:

If the observed D_{n_1, n_2} is greater than or equal to the calculated and tabled $D_{n_1, n_2 \text{ critical}}$ ($D_{n_1, n_2} \geq D_{n_1, n_2 \text{ critical}}$) at the specific level of significance (α), the null hypothesis is rejected. Therefore, there is a significant difference between these two populations.

4) Examples to Demonstrate the Calculation of the Test Statistic

Presented next were the examples which were demonstrated in the MW test. This may help readers examine differences in performing the MW and the KS-2 tests.

Example One: A Small Sample for Each Group

Score values were as follows:

Sample 1: 17, 36, 18, 40, 52; $n_1 = 5$

Sample 2: 15, 37, 23, 32, 43, 50; $n_2 = 6$

The research question was designed to detect whether these two samples were drawn from identical populations at the α level of 0.05. Thus, the null and alternative hypotheses were:

Ho: There was no difference between two populations.

Ha: There was a difference between two populations.

Data arrangement was as follows:

X	17	18			36		40		50		
$S_{n1}(x)$	0	0.2	0.4	0.4	0.4	0.6	0.6	0.8	0.8	1.0	1.0
Y	15		23	32		37		43		52	
$S_{n2}(x)$.017	0.17	0.17	0.33	0.5	0.5	0.67	0.67	0.83	0.83	1.0
D	.017	0.03	0.13	0.07	0.1	0.1	0.07	0.13	0.03	0.17	0

$$S_1(x) = \frac{(\text{number of observed X's} \leq x)}{n_1}, \text{ and } S_2(x) = \frac{(\text{number of observed Y's} \leq y)}{n_2}.$$

$$D_{n1,n2} = \max_x |S_{n1}(x) - S_{n2}(x)| \\ = 0.13$$

Table value $D_{n1,n2; 0.05} = 0.667$

0.13 was not greater than 0.667, so the null hypothesis can not be rejected. Therefore, it was concluded that there was no difference between the two populations.

Example Two: A Large Sample for Either Group

Score values were as follows,

Sample 1: 8, 17, 36, 18, 40, 52, 38, 59, 31, 68; $n_1 = 10$

Sample 2: 3, 15, 25, 48, 37, 65, 6, 57, 42, 35, 11, 23, 32, 43, 50, 51, 62, 74, 20, 69, 44, 9, 39, 47, 66, 55; $n_2 = 26$

The research question was developed to detect whether these two samples were drawn from the identical populations at the α level of 0.05. Thus, the null and alternative hypotheses were:

Ho: There was no difference between two the populations.

Ha: There was a difference between two the populations.

Data arrangement was as follows:

X	8			17			18			31		
$S_{n1}(x)$	0	0	.1	.1	.1	.1	.2	.3	.3	.3	.3	.4
Y	3	6	9		11	15			20	23	25	
$S_{n2}(x)$.038	.076	.076	.114	.152	.190	.190	.190	.228	.226	.304	.304
D	.038	.076	.024	-.014	-.052	-.090	.01	.11	.062	.014	-.004	.096
X	36			38			40					
$S_{n1}(x)$.4	.4	.5	.5	.6	.6	.7	.7	.7	.7	.7	.7
Y	32	35	37		39			42	43	44	47	48
$S_{n2}(x)$.342	.38	.418	.456	.456	.494	.494	.532	.570	.608	.646	.684
D	.058	.02	.082	.024	.144	.106	.206	.168	.130	.092	.054	.016
X	52			59			68					
$S_{n1}(x)$.7	.7	.8	.8	.9	.9	.9	.9	.9	1	1	1
Y	50	51	55		57	62			65	66	69	
$S_{n2}(x)$.722	.760	.760	.798	.836	.836	.874	.912	.950	.950	.988	1
D	-.022	-.060	.04	.002	.064	.064	.026	-.012	-.050	.050	.012	0

$$S_1(x) = \frac{(\text{number of observed X's} \leq x)}{n_1}, \text{ and } S_2(x) = \frac{(\text{number of observed Y's} \leq y)}{n_2}.$$

$$D_{n1,n2} = \max_x |S_{n1}(x) - S_{n2}(x)| \\ = 0.206$$

the critical $D_{n_1, n_2} = 0.500$

0.206 was not greater than 0.500, so the null hypothesis can not be rejected.

If the formula for large samples was applied, and $\alpha = 0.05$,

$$\text{the critical } D_{n_1, n_2} = \text{table value (K)} \times \sqrt{\frac{n_1 + n_2}{n_1 n_2}} = 1.36 \times \sqrt{\frac{10 + 26}{10 \times 26}} \approx 0.506$$

0.206 was not greater than 0.506, so the null hypothesis was retained. Therefore, it was concluded that there was no difference between the two populations.

5) The Kolmogorov-Smirnov two-sample Test Used in This Study

After reviewing various textbooks, the following elements were recommended for applying the KS-2 test. These same elements were proposed by all; they were described as *1) assumptions and data arrangements, 2) hypotheses, and 3) formulas of test statistics and decision rules for small and large sample sizes.*

1) Assumptions and Data Arrangements

Assumptions similar to Conover's (1999) were suggested for this study. There were four assumptions as followed:

- (1) Each sample has been randomly selected from the population it represented.
- (2) The measurement scale employed was at least ordinal.
- (3) The originally observed variable was a continuous variable.
- (4) Two samples were mutually independent.

Data arrangement proposed by Siegel and Castellan (1988) were modified and used in this study:

Let $S_1(x)$ be the cumulative distribution probability function (cdf's) based upon the random sample scores of X_1, X_2, \dots, X_{n_1} .

Determine the $S_1(x)$ for each value of X_1, X_2, \dots, X_{n_1} , let $S_1(x) = \frac{K}{n_1}$.

Let $F(x)$ be the population that the sample of X 's were randomly drawn from.

Let $S_2(x)$ be the cumulative distribution probability function (cdf's) based upon the random sample scores of Y_1, Y_2, \dots, Y_{n_2} .

Determine $S_2(x)$ for each value of Y_1, Y_2, \dots, Y_{n_2} , let $S_2(x) = \frac{K}{n_2}$.

Let $G(x)$ be the population that the sample of Y 's were randomly drawn from.

D_{n_1, n_2} was symbolized as the test statistic for the KS-2 test. It was the maximum absolute difference between the two empirical distribution functions or cumulative distribution functions.

2) *Applicable Hypotheses*

Since this study compared whether there were any general differences between the two populations, a non-directional hypothesis test was presented.

The null and alternative hypotheses were:

Ho: there were no differences between two populations, or

Ho: $F(x) = G(x)$ for all x ; from $-\infty$ to $+\infty$.

Ha: there were some differences between two populations, or

Ha: $F(x) \neq G(x)$ for at least one value of x .

3) Formulas of Test Statistics and Decision Rules for Small and Large Sample Sizes.

Formulas of the test statistic (D_{n_1, n_2}) for both small and large sample conditions as well as decision rules for the testing the hypotheses were presented. In order to be consistent in the definition of sample sizes to compare with the MW test, the researcher will use a size of 20 as the boundary to define small and large sample sizes.

Small Sample Size ($n_1 \leq 20$ or $n_2 \leq 20$)

When both samples were no more than 20 ($n_1 \leq 20$ or $n_2 \leq 20$), the test statistic of the KS-2 test was:

$$D_{n_1, n_2} = \max_x |S_{n_1}(x) - S_{n_2}(x)|$$

The decision rule of the hypothesis test was:

If the observed $D_{m, n}$ was greater than or equal to the tabled D_{n_1, n_2} critical ($D_{n_1, n_2} \geq D_{n_1, n_2}$ critical) at the specific level of significance (α), the null hypothesis was rejected.

Therefore, there was a significant difference between these two populations.

Large Sample Size ($n_1 > 20$ or $n_2 > 20$)

When either or both samples were larger than 20: ($n_1 > 20$ or $n_2 > 20$), the test statistic of the KS-2 test was:

$$D_{n_1, n_2} = \max_x |S_{n_1}(x) - S_{n_2}(x)|$$

Critical D_{n_1, n_2} was calculated with a formula based on various significance level (α).

When the significance level was α ,

$$\text{the critical } D_{n_1, n_2} = \text{table value (K)} \times \sqrt{\frac{n_1 + n_2}{n_1 n_2}}.$$

The decision rule was:

When the observed D_{n_1, n_2} was greater than or equal to the tabled $D_{n_1, n_2 \text{ critical}}$ ($D_{n_1, n_2} \geq D_{n_1, n_2 \text{ critical}}$) at the specific level of significance (α), the null hypothesis was rejected. Therefore, there was a significant difference between these two populations.

In summary, when researchers decide to apply the KS-2 test as their statistical analysis, the following steps were proposed to employ the test: first, rank sample scores from each of the two sample distributions in their own cumulative frequency distribution. Second, for each listed variable, determine the difference between the two-sample cumulative distributions by subtracting the two cumulative relative frequencies. Third, find the largest difference in either direction. Fourth, use the Tabled value and detect whether the test statistics reach the significance level. Fifth, draw conclusions based on the finding from step Four.

6) Selecting sample sizes:

Different textbook authors provided tables with different pairs of equal and unequal sample sizes and the associated critical values. Gibbons and Chakraborti (2003) included critical values tables for selected sample size groups from (2, 2) to (8, 8) with equal and unequal sizes, and (9, 9) to (20, 20) with equal sample sizes. Marascuilo and McSweeney (1977), Siegel and Castellan (1988), and Sheskin (2000) included critical values tables for both equal and unequal sample size groups from (3, 3) to (25, 25). Conover (1999) and Daniel (1990) provided critical values tables for selected unequal sample size groups from (1, 9) to (16, 20).

In this research, both small and large sizes of samples included both equal and unequal conditions. The specific size elections for the two samples were described in CHAPTER THREE.

7) Issue of Ties:

When there were observations in both samples having the same score values (or tied scores), researchers proposed different ways to deal with this situation. For example, Bradley (1968) and Marascuilo and McSweeney (1977) assumed that originally observed variable was a continuous variable implying that no tied observations occurred in the samples. Siegel and Castellan (1988), Conover (1999), Sheskin (2000), and Higgins (2004) did not discuss the issue of ties, while Daniel (1990) claimed that there was no problem when tied scores were presented within the same sample group. It was complicated if the tied condition happens between two sample groups. To simplify this problem, Daniel (1990) and Schroer and Trenkler (1995) proposed that if there are any ties shown between two samples, the probability of the tied value is zero (0). Then, they suggest using a pair chart for the diagonal line to calculate the difference from the chart. However, it is complicated to draw the path and the diagonal line. Neave and Worthington (1988) pointed out that ties might cause severe differences only when they occurred in the area of the maximum difference. They proposed two methods to calculate the maximum difference. The first one was to assign the probability of zero to the tied sample values. They concluded that the difference only shows in the calculation at the end of ties. Similarly, they also suggested using pair chart to check whether the results were the same as of the previous step. The second method was to average the calculated D for the sample with the same score values. They claimed, “This method can be tricky to apply if there were a lot of ties” (p. 155). As a result, to eliminate the difficulty of defining the positions of “the end of ties” and “a lot of ties”, this study did not discuss tied scores for the sample data.

Heterogeneity of Variance, Skewness, and Kurtosis

Introduction

When researchers apply any parametric statistic to their study, they assume that the data are drawn from normal populations and the variances among the populations are equal to one another. When there are violations about the assumptions, the nonparametric statistical analytic techniques usually will be applied to replace the parametric ones (Conover, 1999). Heterogeneity of variance, skewness, and kurtosis are considered violations of the assumptions for parametric statistics. Discussions about these violations are presented below:

Heterogeneity of variance

Homogeneity of variance is one of the assumptions that must be satisfied when performing any parametric statistic (Conover, 1999; Pedhazur & Schmelkin, 1991; Sheskin, 2000; Siegel & Castellan, 1988). If this assumption is not met, nonparametric statistical tests are typically introduced for statistical analysis (Gibbons & Chakraborti, 1991). Vogt (2005) defined homogeneity of variance as a condition that populations from which samples have been drawn do not have similar or equal variance. Zinnerman's research (2004) revealed that nonparametric tests of location, such as the Wilcoxon-Mann-Whitney rank test, were affected by unequal variances of two samples. When the ratios between two population standard deviations were increased from 1 to 2, Type I error rate increased significantly when the population distributions are normal and non-normal distributions. Non-normal distributions such as lognormal, gamma, Gumbel, Weibull, have a power function shape. Moreover, when the population standard deviation ratios were increased from 2 to 3, Type I error rates of these populations became more liberal (Type I error rates are greater than the significance level α). Therefore, it may be necessary to detect the variances between two

populations if researchers decide to assess Type I error rates for any two-sample statistical test. Penfield (1994) also supported the argument and suggested examining the equality of the variance assumption and the level of skewness about the data sets when performing a two-sample location test. This is particularly true when the Type I error rate and power are evaluated.

There are various methods for indicating homogeneity of variance between two samples when performing Monte Carlo simulations. Penfield (1994) used the ratio between two

population variances $\frac{\sigma_1^2}{\sigma_2^2}$ as an index for homogeneity of variance. The symbol σ_1^2

represents the population variance of the first sample group and σ_2^2 is the population

variance of the second sample group. Gibbons and Chakraborti (1991) and Zimmerman

(2003; 2004) also used the ratio between two population standard deviations $\frac{\sigma_1}{\sigma_2}$ as an

indicator for homogeneity of variance. The two indicators $\frac{\sigma_1^2}{\sigma_2^2}$ and $\frac{\sigma_1}{\sigma_2}$ have the same effect

when applied to Monte Carlo simulations since the first ratio is the squared value of the

second one. Therefore, this researcher decided to use $\frac{\sigma_1}{\sigma_2}$ as an index for detecting the

violation of the assumption of homogeneity of variance.

Skewness and Kurtosis

Skewness and kurtosis are assessed to detect shapes of a distribution (Balakrishnan & Nevzorov, 2003; Joanest & Gill, 1998). Sheskin (2000) and Vogt (2005) defined skewness as a measure which reflects the degree to which a score distribution is asymmetrical or symmetrical. When data are symmetrical, researchers usually assume data are normally

distributed. According to the definition by Vogt (2005), kurtosis is an indicator of the degree to which a score distribution is peaked. Sheskin (2000) revealed that the reason for measuring kurtosis is to verify whether data are derived from a normally distributed population. In 1895, Pearson first developed a set of measures of skewness and kurtosis (as cited in Balakrishman and Nevzorov, 2003). They are given by:

$$\text{Skewness: } \gamma_1 = \frac{\beta_3}{\sqrt{\beta_2^3}},$$

where β_3 is the third central moment of the population distribution function.

β_2 is the second central moment of the population distribution function.

$$\text{Kurtosis: } \gamma_2 = \frac{\beta_4}{\beta_2^2}$$

where β_4 is the fourth central moment of the population distribution function.

β_2 is the second central moment of the population distribution function.

Based upon Sheskin's explanation (2000), "the word moment is employed to represent to the sum of the deviations from the mean in reference to sample size" (p.10). Balakrishman and Nevzorov (2003) provided the formula of the nth central moment (β_n) of a continuous variable X which is defined as $\beta_n = E(X-EX)^n$. Based on this formula, the first central moment of the population distribution function is derived as $\beta_1 : \beta_1 = E(X-EX)^1$. After this calculation, β_1 , population mean (μ), is obtained. Using a similar procedure, the second central moment of the population distribution function β_2 is obtained. This is the population variance (σ^2). Fleishman (1978) and Joanest and Gill (1998) proposed exact formulas based on Pearson's work to find skewness and kurtosis. However, Bai and Ng (2005), Sheskin

(2000) and Algina, Olejnik, and Ocanto (1989) replaced the β_3 , β_4 , and β_2 by μ_3 , μ_4 , and σ^2 , and changed the formulas of skewness and kurtosis to $\gamma_1 = \frac{\mu_3}{\sigma^3}$ and $\gamma_2 = \frac{\mu_4}{\sigma^4}$.

Balakrishman and Nevzorov (2003) pointed out that distributions with $\gamma_2 > 3$ are leptokurtic distributions; those with $\gamma_2 < 3$ are platykurtic distributions; others with $\gamma_2 = 3$ are mesokurtic distributions (including the normal distribution). Moreover, distributions with $\gamma_1 > 0$ are positively skewed distributions; those with $\gamma_1 < 0$ are negatively skewed distributions; Algina, Olejnik, and Ocanto (1989) suggested that distributions with $\gamma_1 = 0$ and $\gamma_2 = 3$ are normal distributions. Sheskin (2000) defined leptokurtic as a score distribution that tends to be clustered much more closely around the mean with a high degree of peakedness. A platykurtic distribution is one where the score distribution tends to be spread out more from the mean with a low degree of peakedness. A mesokurtic distribution has a moderate degree of peakedness and is represented by a normal distribution that is a bell-shaped curve.

Skewness and kurtosis are significant indicators for describing shape characteristics of a score distribution. When researchers decide to perform any statistical test, skewness and kurtosis are important considerations about whether the population distributions are normal or non-normal. This helps researchers determine whether to use parametric or nonparametric statistical analytic techniques. This study will detect how the MW test and the KS-2 perform in terms of Type I error rate and statistical power under various degrees of skewness and kurtosis. The strategy is fully explained in CHAPTER THREE.

Method of Selecting Population Distributions

In evaluating two-sample statistical tests, many researchers have developed methods to simulate samples from population distributions. A population distribution, according to

Sheskin's definition (2000), is a shape of arranging a group of variables that share something in common with one another. Based on the characteristics of population distributions, researchers have explored Type I error rates and statistical power when comparing various two-sample statistical tests. For example, Blair and Higgins (1985) compared the power of the paired-sample test with Wilcoxon's signed-ranks test among normal, lognormal, mixed-normal, exponential, mixed- exponential, uniform, double- exponential, truncated normal, Chi-square, and Cauchy population distributions. MacDonald (1999) investigated statistical power and Type I error rates between two samples for the Student t test and the Wilcoxon rank sum test (the Mann-Whitney test) across normal, mixed-normal, and exponential population distributions. Zimmerman (2001b) examined Type I and Type II error rates between two samples among the Student t test, the t tests on rank and the MW test when the sample sizes are the same. Zimmerman (2001b) detected these two-sample statistical tests for normal, mixed-normal, exponential, Laplace, and Cauchy population distributions. These researchers used the known population distributions to examine statistical power and Type I and Type II error rates in parametric and nonparametric two-sample statistical tests.

Fleishman (1978) developed a power function as a distribution generating method to help researchers produce widely different distributions and to simulate empirical distributions.

The formula is as follows:

$$Y = a + [(d \times X + c) \times X + b] \times X,$$

where Y is a distribution dependent on the constants.

X is a random variate normally distributed with the mean zero and unit standard deviation 1, or N (0, 1).

a is constant, a = -c, b, c, and d values which were generated by Fleishman and are found in APPENDIX I. The coefficients of a, b, c, and d in APPENDIX I can be found with the restrictions that the mean, variance, skewness, and kurtosis are 0, 1, γ_1 and γ_2 . This simulation formula was adopted by researchers such as Penfield (1994) to detect Type I error rates between two sample tests in parametric and nonparametric statistics research.

In APPENDIX I, the measures of skewness and kurtosis are calculated by the formulas provided by Fleishman (1978):

$$\text{Where the measure of skewness} = \gamma_1 = \frac{\beta_3}{\sigma^3};$$

$$\text{the measure of kurtosis} = \gamma_2 = \frac{\beta_4}{\sigma^4} - 3$$

This population distribution generating function has been applied in Monte Carlo studies for detecting Type I error rates and statistical power by various researchers. Olejnik and Algina (1987) and Algina, Olejnik, and Ocanto (1989) adopted Fleishman's power function (1978) to create observations on both normal and non-normal distributions and used these to estimate Type I error rates and power for the O'Brien test, the Brown-Forsythe test, the Fligner-Killeen test and two Tiku's tests. These tests are other nonparametric statistical two-sample tests of scale difference (such as difference in variances). In Algina, Olejnik, and Ocanto's 1989 study, twelve distributions were generated by different degrees of skewness and kurtosis. Penfield (1994) applied Fleishman's power function to investigate Type I error rates and power for the Student t test, the MW test, vander Waerden Normal Score (NS) test, and Welchi-Aspin-Satterthwaite (W) test. About nineteen population distributions were generated in that study. As shown here, researchers adopted Fleishman's power function to

investigate statistical power and Type I error rates under different shapes of population distributions when the focus of their research was to detect power and Type I error rates with various degrees of skewness and kurtosis, as well as for testing the differences between variances of samples.

Given that one purpose of this study is to detect Type I error rates and statistical power under various differences in variances, skewness, and kurtosis between two samples, Fleishman's power function will be utilized to generate different distributions along with various ratios of skewness and kurtosis for the Monte Carlo simulation. Moreover, this study will adopt the coefficient of skewness and kurtosis as applied in Penfield (1994) and Algina, Olejnik, and Ocanto (1989) to investigate Type I error rates and statistical power between the MW and the KS-2 tests.

Issues Related to the Mann-Whitney Test

When investigating the Mann-Whitney (MW) test, Type I error rates and statistical power are two of the most important criteria to determine whether the statistical test is conservative or liberal in the decision-making of hypothesis testing. Research related to these two issues was explored and is presented below.

Type I Error Rates

Type I error rate is the probability of rejecting a true null hypothesis. When researchers perform hypothesis tests, one of the main goals is to find out Type I error rates for making decisions in statistical inference. Many studies here investigated Type I error rates. For example, Gibbons and Chakraborti (1991) investigated Type I error rates of the Mann-Whitney test and the Student t test with normal distributions with the conditions of equal ($n_1 =$

$n_2 = 10$) and unequal ($n_1 = 4, n_2 = 16$) sizes in small samples. They also considered one equal ($\sigma_1 = \sigma_2$) and four sets of unequal population standard deviations: (1) $\sigma_1 = 2.5\sigma_2$; (2) $\sigma_1 = 5\sigma_2$; (3) $\sigma_2 = 2.5\sigma_1$; (4) $\sigma_2 = 5\sigma_1$. There were two findings with equal population standard deviations between two samples. First, it was found that when both sample sizes were 10 and the significance level (α) was 0.0432, Type I error rate was about 0.0458 which is a little greater than the significance level (α). However, when sample sizes were unequal ($n_1 = 4, n_2 = 16$) and the significance level (α) was 0.05, Type I error rate was about 0.0457 which is a little less than the significance level (α). It was concluded that the MW test is more conservative with the condition of unequal small sample sizes than equal ones. Based on the definition from Gibbons and Chakraborti (1991), when Type I error rate is less than the significance level (α), the test is a conservative test.

There are several findings from Gibbons and Chakraborti (1991) when the population standard deviations were not equal between two samples. When both sample sizes were the same ($n_1 = n_2 = 10$), it was found that when the significance level (α) was 0.0432, Type I error rates of the MW test were changed from 0.0559 with $\sigma_1 = 2.5\sigma_2$ to 0.0691 with $\sigma_1 = 5\sigma_2$. Similar results were found for the other two sets of unequal variances (0.0565 with $\sigma_2 = 2.5\sigma_1$, and 0.0749 with $\sigma_2 = 5\sigma_1$). It was shown that Type I error rates of all four sets of the unequal standard deviations were greater than the significance level (α). Similarly, when sample sizes were not equal ($n_1 = 4, n_2 = 16$) and the significance level (α) was 0.05, Type I error rates were about 0.1271, 0.1439 with the population standard deviations of $\sigma_1 = 2.5\sigma_2$ and $\sigma_1 = 5\sigma_2$, respectively. However, when the population standard deviations were $\sigma_2 = 2.5\sigma_1$ and $\sigma_2 = 5\sigma_1$, Type I error rates were about 0.0075, and 0.0109, respectively, when

comparing with the significance level (α) of 0.05. It was found that the MW test was more conservative (Type I error rates were less than the significance level α) within the condition of unequal small sample sizes. When the smaller sample had the smaller population standard deviation and the larger sample had the larger one, Type I error rates were less than the significance level α , and, the MW test became conservative.

There were several conclusions drawn from Gibbons and Chakraborti's 1991 study. First, Type I error rates of the MW test were very close to the significance level (α) when the sizes of two samples were small and equal regardless of population standard deviations. Second, Type I errors were much greater than the significance level (α) when small sample sizes were unequal, especially when the smaller sample was associated with the larger population standard deviations and the larger size with a smaller population standard deviations. However, the MW test became much more conservative (Type I errors were much less than the significance level α) when the smaller sample had a smaller standard deviation.

Penfield (1994) investigated Type I error rates of the Student's t test, the MW test, the van der Waerden Normal Scores (NS) test, and the Welch-Aspin-Sattertheaite (W) test from normal and non-normal distributions. Data in this study were generated by Fleishman's power function (1978) with various degrees of skewness (S) and kurtosis (K). Penfield examined three sets of equal sample sizes: (5, 5), (10, 10), and (20, 20) when both equal and unequal population variances were applied and the significance levels (α) were 0.056, 0.052, and 0.05, respectively. He also examined two sets of unequal sample sizes: (5, 15) and (10, 20) with α of 0.053 and 0.05, respectively and both conditions of equal and unequal

population variances. The ranges for 19 pairs of skewness and kurtosis (K, S) were from (0, -1) to (1.5, 3.5).

This research revealed several findings with regards to the MW test. First, when there were equal population variances between two samples, Type I error rates for all level of skewness and kurtosis (K, S) were close to the significance levels (α) for all equal pairs of samples in this study. Second, when there were unequal sample sizes (5, 15) and (10, 20) with equal population variances between two samples, Type I error rates were acceptable at α of 0.053 and 0.05 for all levels of skewness and kurtosis (K, S). Third, when two samples had equal sizes but different variances ($\sigma_1^2 = 4\sigma_2^2$), Type I error rates were greater than the significance levels (α) at levels of skewness. Moreover, as the level of skewness increased, Type I error rates increased significantly. Fourth, when two samples were unequal in both sizes and population variances, Type I error rates were greater than the significance levels (α) when the larger population variances were associated with the smaller sample sizes at all levels of skewness. However, when the larger population variances were associated with the larger sample sizes, Type I error rates were much less than the significance levels (α) at all level of skewness.

In conclusion, the MW test was very liberal (Type I error rates were greater than the significance level α) when two samples had equal sizes but different variances despite the sample sizes and levels of skewness. When both samples were of the same size, as the level of skewness increased, the actual Type I rates increased significantly. The MW test became extremely liberal when one of the two samples had the larger variance and the smaller size. On contrary, the test was very conservative (Type I error rates were less than the significance level α) when the larger sample held a larger variance.

The findings from Penfield's research confirmed the conclusions proposed from Gibbons and Chakraborti's investigation in 1991 that the MW test was conservative (the Type I error rate was less than the significance level α) in terms of Type I error rates when two samples had small unequal sample sizes, and the larger size of the two samples had the larger population variance. The test was liberal (Type I error rates were greater than the significance level α) when the smaller sample had a larger variance.

Kasuya (2001) investigated Type I error rates of the MW test when the variances of two populations were not equal. He used the ratios of two population standard deviations to simulate the results of the MW test under equal and unequal sample sizes ($n_1 = 25, n_2 = 15$; $n_1 = n_2 = 20$; and $n_1 = 30, n_2 = 10$). Simulations were separately performed with the populations from normal and uniform distributions. Results revealed that in the normal distribution, when the sample size of the two samples were unequal ($n_1 = 25, n_2 = 15$ and $n_1 = 30, n_2 = 10$), as the standard deviation ratio (SD ratio) between two populations was increased from 0.2 to 4, Type I error rates increased from 0.05 to 0.11 ($n_1 = 25, n_2 = 15$) and 0.02 to 0.14 ($n_1 = 30, n_2 = 10$). When the sample sizes were equal, Type I error rates were decreased from 0.08 to 0.05 when the SD ratio between two normally distributed populations changed from 0.2 to 1.6. However, when the SD ratio changed from 1.6 to 4.0, Type I rates increased from 0.05 to 0.08. Similar results were found when the populations were from uniform distributions.

Thus, conclusions drawn from Kasuya (2001) also confirmed that the MW test inflated Type I error rate when the variances differed between two samples with equal and unequal sample sizes. This study also supported Penfield (1994) and Gibbons and Chakraborti's investigation in 1991 that when the larger size of the two samples had the larger population

standard deviation, Type I errors became much less than the significance level (α) and the MW test was extremely conservative.

Zimmerman has studied Type I error rates and the power of nonparametric tests over two decades since 1980s, particularly the MW test. In 1985, Zimmerman proposed a simulation study of the MW test with the assumptions of (1) normal (binomial distribution) and non-normal population distributions (uniform distribution), (2) equal and small sample sizes ($n_1 = n_2 = 5$), (3) equal and unequal ($\sigma_1 = 4\sigma_2$) population variances, and (4) the significance level (α) was 0.056. The results of this study were that when the two population variances were the same, Type one error rates were less than the significance level (0.053 for the normal distribution and 0.055 for the non-normal distribution). However, when populations variances were unequal ($\sigma_1 = 4\sigma_2$), Type I error rates of the MW test were greater than the 0.056 significance level (0.070 for the normal distribution and 0.076 for the non-normal distribution).

In conclusion, under the condition of small and equal sizes between two samples, the MW test was conservative when the population variances were the same regardless of the population distributions. The MW test became liberal when the population variances were unequal but the sample sizes were small and equal with both normal and non-normal distributions. However, the study was conducted only comparing one set of sample sizes in two pairs of population variances under two population distributions.

Zimmerman (1987) expanded the study only in the normal distribution with the assumptions of: (1) three pairs of small sample sizes with one equal ($n_1 = n_2 = 10$) and two unequal ($n_1 = 16, n_2 = 4$ and $n_1 = 4, n_2 = 16$), (2) one pair of equal population variances ($\sigma_1 = \sigma_2$) and one with extremely unequal population variances ($\sigma_1 = 5\sigma_2$), and (3) the significance

level (α) was 0.05. It was found that when the assumption of homogeneity of variances was met, Type I error rates were 0.041, 0.049, and 0.048 for sample sizes of $n_1 = n_2 = 10$, $n_1 = 16$, $n_2 = 4$ and $n_1 = 4$, $n_2 = 16$, respectively. These Type I error rates were all less than the significance level of 0.05. Moreover, when the first population standard deviation was five times as large as the second population standard deviation ($\sigma_1 = 5\sigma_2$), only the sample size combination of $n_1 = 16$ and $n_2 = 4$ had a very small Type I error (0.006). Type I error rates of other two pairs of sample size combinations were all greater than the 0.05 significance level (0.075 for $n_1 = n_2 = 10$ and 0.134 for $n_1 = 4$, $n_2 = 16$).

Based on this research, it appeared that when two populations had the same small sample sizes, the MW test was liberal (Type I error rate exceeded the significance level) with extremely unequal variances. When the sample size was large with much larger variance than the other sample, the MW test became very conservative (the Type I error rate was less than the significance level). On the other hand, the MW test was liberal when the sample size was small with much larger variance than the other sample. Gibbons and Chakraborti (1991), Penfield (1994), and Kasuya (2001) all confirmed this finding in their studies in later years which was discussed in an earlier section.

In 1990, Zimmerman and Zumbo investigated Type I error rates of the MW test with normal, uniform, exponential, Cauchy, and mix-normal distributions for the two populations. They examined two sets of small and equal sample sizes ($n_1 = n_2 = 8$ and $n_1 = n_2 = 16$). Nine sets of differences between two population standard deviations ($\sigma_1 - \sigma_2 = 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, \text{ and } 4.0$) were also examined for Type I error rates. It was found that when the difference between the two population standard deviations was zero and both sample sizes were eight, Type I error rates for normal, uniform, exponential, Cauchy, and mix-normal

distributions were 0.051, 0.052, 0.052, 0.047, and 0.048, respectively. As the population standard deviation differences increased, Type I error rates for all five kinds of population distributions increased too. Moreover, when both sample sizes increased from 8 to 16 and the standard deviation difference was zero, all five distributions had Type I error rates less than 0.05. Similarly, when the standard deviation difference was increased, Type I error rates were raised as well.

This study by Zimmerman and Zumbo (1990a) concluded that when the two samples had equal sizes and were less than 20, the MW test was conservative since Type I error rates were less than the significance level of 0.05 with normal, uniform, exponential, Cauchy, and mix-normal population distributions. When homogeneous of variances was violated, the MW test became liberal in any of these five population distributions. It was suggested that the MW test was powerful for both normal and non-normal distributions when the two samples had small and equal sizes and population variances of these two samples were the same. However, when two samples had equal sample sizes but population variances of these two samples differed from each other, the MW test was not powerful with both normal and non-normal population distributions.

In 1998, Zimmerman started to examine Type I error rates of the MW test with the normal population distribution under the conditions of both equal and unequal sample sizes with both equal and unequal population standard deviations. The significance level was 0.05 for this study. The pairs of sample sizes (n_1, n_2) were (40, 20), (20, 40), (20, 20), (30, 30), and (40, 40). Ratios of two population standard deviations ($\frac{\sigma_1}{\sigma_2}$) were used to examine Type

I error rates; $\frac{\sigma_1}{\sigma_2}$ ratios were 1, 2, 3, and 4, respectively. It was found that Type I error rates

were 0.049, 0.049, and 0.050 which were very close to the significance level of 0.05 when the ratio $\frac{\sigma_1}{\sigma_2}$ was 1 (equal population variance) and the combinations of these two sample sizes were (20, 20), (30, 30), and (40, 40) respectively. When $\frac{\sigma_1}{\sigma_2}$ ratios increased, Type I error rates of these three combinations of equal sample sizes became greater than the 0.05 significance level. When the pair combination of the two samples was (20, 40), Type I error rates were greater than 0.05 significance level regardless of the ratios of two population standard deviations ($\frac{\sigma_1}{\sigma_2}$). However, when the pair combination of the two samples changed to (40, 20), the Type I error rate was greater than 0.05 when the $\frac{\sigma_1}{\sigma_2}$ ratio was equal to one. Type I error rates became less than 0.05 significance level as ratios of $\frac{\sigma_1}{\sigma_2}$ became greater than one (unequal variances).

In conclusion, with normal population distributions and large sample size scenarios, the MW test was liberal (Type I error rates exceed the significance level) with the assumption of homogeneity of variances when the sizes of the two samples were unequal to one the other. The MW test was conservative (Type I error rates are less than the significance level) with the assumption of unequal population variances when the sample with large size had large $\frac{\sigma_1}{\sigma_2}$ ratios.

Zimmerman (2000) proposed a Type I error rate investigation for both large and small equal sample sizes ($n_1 = n_2 = 4, 5, 6, 7, 8, 20, 40,$ and 80) with population standard deviation

ratios ($\frac{\sigma_1}{\sigma_2}$) from 1.0 to 4.0 in increments of 0.5. The study included three α significance levels for each pair of sample size combination. When the sample sizes were four ($n_1 = n_2 = 4$), the significance levels α were 0.028, 0.058, and 0.114. When the sample sizes were five ($n_1 = n_2 = 5$), the significance levels α were 0.016, 0.056, and 0.096. When the sample sizes were 6 ($n_1 = n_2 = 6$), the significance levels α were 0.016, 0.042, and 0.094. When the sample sizes were 7 ($n_1 = n_2 = 7$), the significance levels α were 0.012, 0.054, and 0.104. When the sample sizes were 8 ($n_1 = n_2 = 8$), the significance levels α were 0.010, 0.050, and 0.104. When the sample sizes were 20, 40, and 80 ($n_1 = n_2 = 20, 40, \text{ and } 80$), the significance levels α were 0.01, 0.05, and 0.10.

It was found that Type I error rates were less than or equal to the significance levels α for all pairs of sample size combinations when the $\frac{\sigma_1}{\sigma_2}$ ratio was equal to one. When the $\frac{\sigma_1}{\sigma_2}$ ratio increased, Type I error rates became greater than the significance levels α for both small and large sample sizes.

The results revealed that the MW test was mildly conservative when homogeneity of variances existed with the normal distribution and both samples were equal regardless of the sizes of these samples. The MW test became liberal when the condition of homogeneity of variances was violated for all sizes of equal samples. However, only the normal distribution for the two populations was examined in this study.

In order to assess Type I error rates of the MW test with both normal and non-normal population distributions, Zimmerman examined Type I error rates of the MW test for 11

different population distributions in 2003 and 25 different population distributions in 2004 for both small and large sample sizes with different ratios of population standard deviations. Further, in 2003, Zimmerman examined Type I error rates of the MW test with both three pairs of small and equal size combinations ($n_1 = n_2 = 6, 8, \text{ and } 10$) and six pairs of large and equal sample size combinations ($n_1 = n_2 = 20, 30, 60, 90, 120, \text{ and } 200$). The population standard deviation ratios ($\frac{\sigma_1}{\sigma_2}$) were 1.0, 1.1, 1.2 which were equal to or had small differences between two population variances. Three levels of significance were considered ($\alpha = 0.009, 0.041, \text{ and } 0.093$). It was found that, at all three levels of significance, Type I error rates were slightly inflated as the $\frac{\sigma_1}{\sigma_2}$ ratios increased from 1.0 to 1.2 in both small ($n_1 = n_2 = 6, 8, \text{ and } 10$) and large samples ($n_1 = n_2 = 20, 30, 60, 90, 120, \text{ and } 200$) regardless of the type of population distributions.

In conclusion, the MW test was slightly conservative (Type I error rates were less than significance) when homogeneity of variances existed ($\frac{\sigma_1}{\sigma_2} = 1$) for small and large sample sizes with the normal and non-normal population distributions. The MW test was conservative when homogeneity of variances was slightly violated with normal distributions and large sample sizes. However, the MW test was liberal (Type I error rates exceed the significance level) when homogeneity of variances was slightly violated with non-normal population distributions regardless of the sizes of these two equal samples.

In 2004, Zimmerman investigated Type I error rates of the MW test with both four pairs of large and equal size combinations ($n_1 = n_2 = 20, 25, 50, \text{ and } 80$) with population standard

deviation ratios ($\frac{\sigma_1}{\sigma_2}$) of 1.0, 1.25, 2.0 and 3.0. There were 25 normal and non-normal population distributions examined. Three levels of significance were considered (0.01, 0.05, and 0.10). It was found that when both sample sizes were 25, Type I error rates were close or equal to the significance levels as the $\frac{\sigma_1}{\sigma_2}$ ratio was equal to 1. Type I error rates exceeded the significance levels as the $\frac{\sigma_1}{\sigma_2}$ ratios changed to 2 and 3. Especially when the populations were exponential, gamma, and the Weibull distributions, Type I error rates increased dramatically. Type I error rates also inflated as the $\frac{\sigma_1}{\sigma_2}$ ratios changed from 1.25 to 2.0 and the pairs of equal sample sizes increased from 20 to 80. This was particularly the case with the Weibull population distribution.

In conclusion, the MW test was slightly conservative when homogeneity of variances existed ($\frac{\sigma_1}{\sigma_2}=1$) for large and equal sample sizes with the normal and non-normal population distributions. It became liberal when there was no existence of homogeneity of variances when sample sizes were equal and large with normal and non-normal population distributions. The MW test was extremely liberal, especially, when populations were non-normal distributions. This indicated that researchers should reconsider whether the MW test is appropriate under conditions such as sample sizes, population variances, and shapes of the population distributions.

Statistical Power Estimates

Statistical power is another important criteria for making decisions in statistical inference. Statistical power is the probability of correctly rejecting a false null hypothesis. Shavelson (1988) stated that statistical power is used to point out the probability of detecting a difference if the difference actually exists. Researchers might hope to have high statistical power when performing any statistical test.

In power comparisons of small and equal sample sizes ($n_1 = n_2 = 10$), Gibbons and Chakraborti (1991) found that the MW test had similar power with the Student's t test when the population variances were equal to each other ($\sigma_1 = \sigma_2$). The MW test was more powerful than the Student's t test when there were extremely unequal variances ($\sigma_1 = 5\sigma_2$, and $\sigma_2 = 5\sigma_1$) between two samples. The results revealed that when the sample sizes are small and equal, the MW test was more powerful than the Student's test as the assumption of homogeneity of variances was violated.

Penfield (1994) examined statistical power of the Student's t test, the MW test, the van der Waerden Normal Scores (NS) test, and the Welch-Aspin-Sattertheaite (W) test from normal and non-normal distributions. Data in this study were generated by Fleishman's power function (1978) with various degrees of skewness (S) and kurtosis (K). Penfield considered three sets of equal sample sizes: (5, 5), (10, 10), and (20, 20) and two sets of unequal sample sizes: (5, 15) and (10, 20) with both conditions of equal and unequal population variances. The ranges for 19 pairs of skewness and kurtosis (K, S) were from (0, -1) to (1.5, 3.5).

It was found that when sizes for both samples were five ($n_1 = n_2 = 5$) and the pairs of skewness and kurtosis (S, K) were (0.5, -.05), (1, 3), the power of the MW test and the van der Waerden Normal Scores (NS) test was the same and greater than the Student's t test.

When the pairs of skewness and kurtosis were (1.5, 2.5) and (1.5, 3), the MW test and NS test were the desired tests. When the two sample sizes were (10, 10), (20, 20), (5, 15) and (10, 20) and pairs of S, K were (0.5, -0.5), (0.5, 3), (1, .5) (1, 3) (1.5, 2.5) and (1.5, 3), the MW test was preferred to the other tests. When variances were unequal ($\sigma_1 = 2\sigma_2$), the MW test had more power only with the sample sizes of (10, 20) and the combinations of skewness and kurtosis were (1, 0.5), (1, 3), (1.5, 2.5) and (1.5, 3).

In conclusion, the MW test was powerful when the samples were small with equal and unequal sizes. The MW test was also powerful when the population distributions had various degrees of skewness and kurtosis. It was suggested by Penfield (1994) that the MW test had more power in the small equal and unequal sample sizes and non-normal population distributions.

Zimmerman (1985) investigated statistical power estimates between the MW test and the Student's t test in the normal distribution under the conditions of equal small sample sizes ($n_1 = n_2 = 5$) and both equal and unequal ($\sigma_1 = 4\sigma_2$) population variances. It was found that, in the condition of small and equal sample sizes, the Student's t test was more powerful than the MW test for both equal and unequal variances.

In 1987, Zimmerman examined the power of the MW test and the Student's t test in the normal distribution with the assumptions of (1) three pairs of small sample sizes with one equal ($n_1 = n_2 = 10$) and two unequal ($n_1 = 16, n_2 = 4$ and $n_1 = 4, n_2 = 16$), (2) one pair of equal population variances ($\sigma_1 = \sigma_2$) and one with extremely unequal population variances ($\sigma_1 = 5\sigma_2$), and (3) a significance level (α) of 0.05. It was found that the MW test was more powerful only under the condition of unequal and small sample sizes ($n_1 = 16, n_2 = 4$) when the extremely unequal population variances ($\sigma_1 = 5\sigma_2$) existed.

The results from the 1985 and 1987 studies by Zimmerman revealed that the MW test is more powerful with a normal distribution when the two samples had small and unequal sizes, and when the sample with the larger size had a larger population variance. However, it appeared that the comparisons of sample sizes and population variances were limited. One might question paired comparisons of sample sizes and population variances that were not in the range used in this investigation.

In 1990, Zimmerman and Zumbo investigated the power estimates of the MW test and the Student t test with normal, uniform, exponential, Cauchy, and mix-normal distributions for two populations. They examined two sets of small and equal sample sizes ($n_1 = n_2 = 8$ and $n_1 = n_2 = 16$). Nine sets with differences between two population standard deviations ($\sigma_1 - \sigma_2 = 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5,$ and 4.0) were also examined for statistical power estimates. It was found that the MW test had more power than the Student t test under exponential, Cauchy, and mixed-normal distributions.

In 2003, Zimmerman examined the power estimates of the MW test and the Student t test with both three pairs of small and equal sample size combinations ($n_1 = n_2 = 6, 8,$ and 10) and six pairs of large and equal sample size combinations ($n_1 = n_2 = 20, 30, 60, 90, 120,$ and 200) for 11 different population distributions. The population standard deviation ratios ($\frac{\sigma_1}{\sigma_2}$) were 1.0, 1.1, 1.2 which were equal to, or had small differences between two population variances.

It was found that when the population standard deviation ratio ($\frac{\sigma_1}{\sigma_2}$) was 1.1, the MW test was more powerful than the Student t test at the sample size combinations of $n_1 = n_2 = 10$ with populations of exponential, lognormal, and skewed binomial distributions. As the sample sizes increased, the power of the MW test also increased. As a result, it was

suggested that the MW test had more power than the Student's t test when selected samples had small or large equal sizes and limited non-normal distributions. When the samples were large, the MW test had less power in most non-normal distributions and normal distributions.

In the current study, the researcher decided to examine Type I error rates and power estimates of the MW test with populations of selected normal and non-normal distributions. Fleishman's power function will be used for generating those selected normal and non-normal distributions since the coefficients of skewness and kurtosis can be defined through this power function. Pair combinations of two samples will include conditions of equal and unequal, as well as small and large sizes. The specific sizes of pair combinations will be presented in CHAPTER THREE.

Issues Related to the Kolmogorov-Smirnov Two-Sample Test

When investigating the Kolmogorov-Smirnov two sample test (KS-2), Type I error rates and statistical power are major focus to assess whether the statistical test is conservative or liberal in the hypothesis testing. Research related to Type I error rates and statistical power were explored and presented below.

Type I Error Rates

Even though the Type I error rate is one of the important criterion of examining a statistical test, there was limited research to detect Type I error rates for the KS-2 test with a non-directional (two-tailed) hypothesis in peer-reviewed journals or in nonparametric statistical textbooks. In the KS-2 test study by Sackrowitz and Samuel- Cahn (1999), it used expected p values to replace Type I error rates and examined conditions directional (one-tailed) hypothesis. However, in the educational and social behavioral research fields, most

researchers tend to be conservative and use a non-directional (two-tailed) hypothesis to define research questions. Type I error rates are significantly important when performing hypothesis testing. Moreover, there is a lack of research in the KS-2 test with a non-directional (two-tailed) hypothesis test for detecting general differences between two samples. Due to this critical need, Type I error rates of the KS-2 test were explored in this study.

Statistical Power Estimates

When comparing the power efficiency of the KS-2 test to other statistical tests under a non-directional (two-tailed) hypothesis, the Student's *t* was the one that was often used to be evaluated with the KS-2 test. By comparing the power efficiency between the Student's *t* test and the KS-2 test, the KS-2 test had higher power efficiency when sample sizes were small (Siegel & Castellan, 1988). When performing comparisons of the power efficiency of the KS-2 test with other nonparametric statistical tests, the chi-square and the median tests were the ones that were often used in comparison. For example, in assessing the chi-square test and the KS-2 test, or the median test and the KS-2 test, the KS-2 test was more powerful than any of these two tests regardless of the sample sizes (Siegel & Castellan, 1988). Textbook authors made comments about power estimates in the KS-2 test. Sprent and Smeeton (2001) claimed that the KS-2 test may have less power estimates than other tests when detecting mean differences between two distributions. Siegel and Castellan (1988) pointed out that the KS-2 test is more powerful for small samples. Power estimates may be slightly reduced when samples are increased in size. However, neither Siegel and Castellan (1988), nor Sprent and Smeeton (2001) specified the number of sample sizes that were used to perform the comparisons. They also did not describe in detail the kind of population distributions and sample sizes used to obtain these results.

In the 1990's, researchers such as Wilcox, Baumgartner, WeiB, and Shindler examined the power of the KS-2 test along with some other parametric and nonparametric statistical techniques for non-directional hypotheses (two-tailed test). Wilcox (1997) examined the power of the KS-2 tests and Student's t test when the sample sizes were 25 with mean differences of 0.6, 0.8 and 1.0 for a normal distribution, 1.0 for a mix-normal distribution, and 0.6 for both exponential and lognormal distributions. It was found that, at the nominal Type I error rate (α) of 0.05, the KS-2 tests had smaller power (0.384, 0.608, and 0.814) than Student's t test (0.529, 0.778, and 0.925) when population distributions were normal regardless of the population mean differences. The KS-2 test had greater power (0.688, 0.866 and 0.666, respectively) than the Student's t test when the populations were mix-normal, exponential and lognormal.

In conclusion, as population distributions become non-normal, statistical power of the KS-2 test was increased when the sizes were 25 in each sample when mean differences occurred between two samples. However, there was no simulation under the consideration of no mean differences. No consideration of changing population variances was examined in the research that was reviewed.

Baumgartner, WeiB, and Shindler (1998) detected the statistical power of the KS-2 test along with the Student's t test, the Wilcoxon test, the Cramer-von Miss test and one new rank test they proposed at the nominal Type I error rate (α) of 0.05. Four simulations were performed in this research. The first simulation compared these parametric and nonparametric statistical tests when the sizes of both samples were 10 ($n_1 = n_2 = 10$) and mean differences but equal variances existed between two populations with normal distributions. It was found that the KS-2 test was the less powerful among these evaluated

statistical tests when population distributions were normal, with mean differences between two populations. The second simulation detected power functions of the KS-2 test, the Wilcoxon test, the Cramer-von Miss test, and the proposed new rank test with both sample sizes of 10 ($n_1 = n_2 = 10$) and the normal distribution. These two samples had no population mean differences but population variances were different. It was found that when the KS-2 test was compared with the Wilcoxon test and the Cramer-von Miss test, the KS-2 was the most powerful test among these three nonparametric statistical tests. A third simulation examined power functions of the KS-2 test, the Wilcoxon test, the Cramer-von Miss test and the new proposed rank test with both sample sizes of 10 ($n_1 = n_2 = 10$) and the exponential distribution. It was found that the KS-2 test had the least power estimates among those tests. One last comparison simulated power estimates of the KS-2 test, Cramer-von Miss test, and the proposed new rank test with large sample sizes (from $n_1 = n_2 = 50$ to 1200) and the underlying populations were normal, with the mean of 0 and standard deviation of $\frac{1}{12}$ and uniformly distributed in the interval of -0.5 and 0.5. Findings indicated that the KS-2 test was the least powerful among these tests, especially with a simulated sample size of more than 800. It appears that the power estimates of the KS-2 test are inferior to the other two tests when populations are large and uniformly distributed.

In Baumgartner, WeiB, and Shindler's (1998) study, the following conclusions were drawn. The KS-2 was not powerful under the conditions of equal sample sizes (both small and large) and normal distributions with no difference between underlying population variances. The KS-2 was powerful when sample sizes were small and equal with a normal distribution and variance differences between the underlying populations. Even though this study added homogeneity of variances into consideration when performing simulations, these

conclusions seem limited and not enough to generate results for other non-normal population distributions without simulating different skewness and kurtosis for the shapes of the underlying population distributions.

Based upon prior research, it seems like these studies considered the situations of equal sample sizes only. No conditions of unequal sample sizes were simulated to estimate statistical power of the KS-2 two-tailed test. Moreover, under the consideration of equal sizes, the numbers of paired size combinations of two samples might not be sufficient enough for researchers to generalize conclusions based upon fewer cases of equal sample sizes. Furthermore, the KS-2 test is one of the nonparametric statistical techniques for determining general differences between two populations when the population distributions are non-normal. These researches seemed to mainly focus on power estimates in normal distributions. Only a few non-normal population distributions, such as mix-normal, exponential, and lognormal, were investigated along with normal distributions. No study related reported Type I error rates for non-directional hypotheses. Therefore, this study will perform Monte Carlo simulations of Type I error rates and power estimates for the KS-2 test with equal and unequal sample sizes in both small and large samples. Non-normal population distributions with different degrees of skewness and kurtosis will be considered in these simulations. The specific considerations will be described in detail in CHAPTER THREE.

Comparisons between the MW test and the KS-2 test

As noted, various researches have explored Type I error rates and power estimates for parametric and nonparametric techniques, such as the Student's t test and the Mann-Whitney test. However, there appears to be limited related research to detect Type I error rates for the KS-2 test. Several researchers have performed statistical power comparisons varying only in

location with normal distributions for the KS-2 test. Siegel and Castellan (1988) even suggested that the KS-2 test was more powerful than the Wilcoxon- MW test with the scenario of very small sample sizes.

Dixon (1954) detected power estimates of the MW test and the KS-2 test under small sizes and normal population distribution conditions. This study showed that when sample sizes are equal and small ($n_1 = n_2 = 2, 3, 4,$ and 5), the power estimates are the same between the MW test and the KS-2 test with the α level of $\frac{1}{3}, \frac{1}{10}, \frac{1}{70}$ and $\frac{1}{126}$, respectively. Schroer and Trenkler (1995) simulated power functions for the KS-2 test, Student's t test, and the MW test in normal, Cauchy, lognormal, and logistic distributions under equal ($n_1 = n_2 = 8$ and $n_1 = n_2 = 15$) and unequal sample sizes ($n_1 = 12, n_2 = 4,$ and $n_1 = 18, n_2 = 12$). It was found that when underlying population distributions were asymmetric or had extreme values or outliers, the KS-2 test had better power than the other assessed statistical tests regardless of the equality of sample sizes.

The conclusion drawn from these two studies was that when the two independent samples had equal and small sample sizes with an underlying population of normal distribution, power estimates of the MW test and the KS-2 test were very similar or even the same. However, when the population distributions for both samples became non-normal, power estimates of the KS-2 test were better than the MW test.

Schroer and Trenkler (1995) also compared the power of the KS-2, the MW, the Cramer-von Mises test and another new test they proposed in three non-normal distributions (Pareto, lognormal, and Singh-Maddalas) with large sample sizes ($n_1 = n_2 = 25$). It was found that the KS-2 test had the smallest power in both the Pareto distribution and the Singh-Maddala

distribution. The KS-2 test had higher power than the MW test when the population was lognormal and Singh-Maddala distributions.

In conclusion, as noted here, the KS-2 was not superior or inferior to the MW test in statistical power with some non-normal population distributions. The shape of the population distributions might be the essential determination of statistical power estimation for these two nonparametric statistical tests.

Baumgartner, WeiB, and Shindler (1998) investigated the statistical power function of the KS-2 test, the MW test, along with other parameter and nonparametric tests when the underlying populations were normal distributions. When both sample sizes were equal to 10 and the population variances were 1 for both samples, power estimates of the MW test were superior to the KS-2 test regardless of the differences in population mean. However, when there was no difference between the mean of two populations but population variances did vary, power estimates of the KS-2 test were better than the MW test when the two samples were size 10 and the nominal Type I error rate (α) was 0.05.

Fahoome (1999) investigated the smallest equal-sample sizes for large-sample approximations of the MW test, the KS-2 test, and other nineteen nonparametric tests for single-sample, two-sample, and multiple-sample conditions with minimal Type I error inflation or loss of power. He also compared differences in the statistics between large-sample approximations and tabulated critical value if the comparisons were appropriate. This research simulated data for normal, smooth symmetric, extreme asymmetric, extreme bimodal, and multimodal lumpy distributions from Miccerri data sets (1989). It was found that the KS-2 test performed inconsistent by when either approximate or critical p-value were closer to the nominal Type I error rate (α). Critical 0.01 p-values were better for normal and

multimodal lumpy distributions. Approximated 0.01 and 0.05 p-values were better for smooth symmetric and extreme asymmetric distribution data sets. The KS-2 test did not perform well on the Miccerri data sets. There was no value of suggested smallest equal sample sizes for large sample approximations with nominal Type I error rates of 0.01 or 0.05. All four Miccerri distributions performed well with critical p-values for the MW test. When determining the smallest equal sample sizes for large sample approximations, with these four data sets, there were several suggestions based upon various distributions. For normal distribution, the suggested sample sizes were 25 for α of 0.01 and 14 for α of 0.05. For extreme asymmetric distribution, the suggested smallest equal-sample sizes were 44 for α of 0.01 and 17 for α of 0.05. When multimodal distribution occurred, the smallest equal-sample sizes were 29 for α of 0.01 and 10 for α of 0.05. There was no value of the smallest equal-sample sizes with smooth symmetric data sets with α of 0.01. When α was 0.05, the smallest equal-sample size was 17.

Summary

In this chapter, the historical development of the Mann-Whitney test and the Kolmogorov-Smirnov two-sample test was reviewed. The theoretical framework of these two tests including data definition, assumptions, hypotheses, and test statistics from various textbooks were also examined. Sample size selections and the issue of tied conditions were investigated through the literature. Examples developed by the researcher were implemented for performing the calculation of test statistics of the MW test and the KS-2 test, as suggested by various textbooks. In this study, heterogeneity of variances, skewness, and kurtosis of population distributions will be main considerations when performing the Monte Carlo simulations, therefore, these considerations were also reviewed and presented. Selecting

population distributions was another key concern for this study; thus, methods of selecting populations were examined from the literature. Finally, issues of Type I error rates and power estimates as related to the MW test and the KS-2 test were reviewed to guide the researcher in selecting sample size combinations of the two independent samples as well as ratios of population standard deviations (SD ratios) when executing Monte Carlo simulations. Overall, when comparing Type I error rates and power estimates between the Mann-Whitney test and the Kolmogorov-Smirnov two-sample test, especially, under the non-directional alternative hypothesis, there is little related peer-reviewed literature to discuss this issue.

In conclusion, this literature review provides a foundation for understanding elements to perform this simulation study. It helps this research clearly define the conditions, such as sample size combinations, SD ratios, ratios of skewness and kurtosis, to form population distributions by using Fleishmen's power function (1978). This will serve to appropriately execute the simulations and to aid in resolving the research questions of this study.

CHAPTER THREE
RESEARCH METHOD

Introduction

This chapter presents the populations and sampling methods used to determine the simulated subjects for this study. Sample sizes for both the Mann-Whitney (MW) and the Kolmogorov-Smirnov two-sample (KS-2) tests are discussed. Formulas for these two statistical tests are presented. The SAS computer program that was utilized to perform the Monte Carlo simulation techniques is discussed. Formulated test statistics for small samples and large sample approximations for the MW test and the KS-2 test were planned in this chapter. Methods of selecting population distributions, simulated data sets related to sample size combinations, ratios between two population standard deviations, and levels of nominal Type I error are introduced as these were needed to compare the actual Type I error rates and statistical power estimates of the two nonparametric statistical tests. There were 15 population distributions, 12 sets of sample size combinations, and 7 different ratios of standard deviation. Exactly 20,000 replications per condition were executed for a total of 1380 conditions (840 for the first research question, 360 for the second question, 36 for the third research question, and 144 for the fourth research question) examining Type I error rates and statistical power for the MW test and the KS-2 test when applicable. Moreover, the steps for performing this simulation study are described in this chapter.

Simulation Overview

Since this was a simulation, there were no human subjects used in this study. Population distributions of two independent sample sets were strictly generated by the computer, a Dell IBM compatible computer with the CPU processor of Pentium dual core 2.80 GHz, along with the program using SAS version 8.2 ("Statistical analysis system," 1999). The RANNOR procedure in SAS was used to generate random numbers from a normal distribution with a mean of zero and a variance of one which was required in the Fleishmen's power transformation method (1978) of generating population distributions (Fan, Felsovalyi, Sivo, & Keenan, 2003). After generating the sample sets, the PROC NPAR1WAY procedure was used to perform actual Type I error rates and power simulations. Simulated data were used to analyze Type I error rates and power for both the MW test and the KS-2 test under conditions determined by the researcher in Table 3. A SAS syntax program was written by the researcher in order to generate populations and sampling distributions, and for calculating each test statistic. A sample of the SAS syntax for this study was provided in APPENDIX II.

The calculated test statistics were evaluated utilizing both tabled critical values and asymptotic approximated critical values. The nominal Type I error rates, alpha (α), for each sample size was 0.05 as was used in Carolan and Tebbs (2005). The actual Type I error rates (exact p-values) were computed for both small sample tabled values and large-sample approximations.

Populations

Since this was a simulation study for comparing the MW and the KS-2 tests in two independent sample conditions, the first step to perform simulations was to determine population distributions associated with these two samples. Therefore, it was important for the researcher to develop populations for simulations. More importantly, a method to consistently generate populations in order to produce reliable population distributions for sampling data sets and performing Monte Carlo simulations was crucial.

Method of Generating the Populations

This section described the method of generating population distributions used for this study. This section also introduced the types of population distributions used for simulating the comparisons of the MW and the KS-2 tests. Fleishman's power function (1978) was utilized for generating population data sets for the simulations in this study. Fleishman (1978) developed a power function as a population distribution generating method for creating widely different distributions and simulating empirical distributions. The formula was as follows:

$$Y = a + ((dX+c)X+b)X$$

This was presented as formula (12) and introduced in CHAPTER TWO. Based on Fleishman's definitions, the X was a random variate, normally distributed with a mean of zero and unit standard deviation of 1, or $N(0, 1)$, and coefficient a equals negative c . The variable X was generated using the SAS/RANNOR program. The coefficients a , b , c , and d were defined based upon the associated conditions of the study, such as means, standard

deviations, and pairings of skewness, and kurtosis. A sample of the SAS syntax was provided in APPENDIX II.

An essential step was to define the population distributions in comparing the MW and the KS-2 tests. Since one of the research questions examined Type I error rates and statistical power when degrees of skewness and kurtosis for population distributions were varied, it was necessary to find populations based on Fleishman's power function. Among the population distributions that were used in this study, twelve population distributions were utilized by Algina, Olejnik, and Ocanto (1989) and three population distributions (uniform-like, logistic-like, and exponential-like) were used by Penfield (1994). Therefore, a total of fifteen population distributions were investigated to examine these two nonparametric statistical techniques. Based on Fleishman's work (1978), the following table listed the pairings of skewness and kurtosis and the coefficients b , c , and d with a mean of 0 and standard deviation of 1. This listed information was used in this study in order to generate population distributions for the two sample sets to perform Monte Carlo simulations. Neither Fleishman nor Penfield provided the coefficients b , c , and d for the uniform-like and logistic-like distributions. Therefore, these coefficients, reported in Table 4, were calculated using Fleishman's formula with Mathematica 5.0 software (Wolfram, 2003).

In Table 4, there were three leptokurtic distributions with same skewness ratios but different degrees of kurtosis. There were also two skewed and platykurtic distributions with different degrees of skewness and kurtosis. Moreover, two different skewed and leptokurtic distributions were determined by the same kurtosis ratios but different skewness ratios.

Table 4: Coefficients used in Fleishmen's power function (1978) with $\mu = 0$; $\sigma = 1$.

Distribution	Skewness	Kurtosis	a	b	c	d
	(γ_1)	(γ_2)				
Normal ¹	0.00	0.00	0.00	1.0000000	0.00	0.00
Platykurtic ¹	0.00	-0.50	0.00	1.0767327	0.00	-0.0262683
Normal Platykurtic ¹	0.00	-1.00	0.00	1.2210010	0.00	-0.0801584
Leptokurtic ¹	0.00	1.00	0.00	0.9029766	0.00	0.0313565
Leptokurtic ¹	0.00	2.00	0.00	0.8356646	0.00	0.0520574
Leptokurtic ¹	0.00	3.75	0.00	0.7480208	0.00	0.0778727
Skewed ¹	0.75	0.00	-0.1736300	1.1125146	0.1736300	-0.0503344
Skewed and platykurtic ¹	0.50	-0.50	-0.1201561	1.1478491	0.1201561	-0.0575035
Skewed and platykurtic ¹	0.25	-1.00	-0.0774624	1.2634128	0.0774624	-0.1000360
Skewed and leptokurtic ¹	0.75	3.75	-0.0856306	0.7699520	0.0856306	0.0693486
Skewed and leptokurtic ¹	1.25	3.75	-0.1606426	0.8188816	0.1606426	0.0491652
Skewed-leptokurtic ¹	1.75	3.75	-0.3994967	0.9296605	0.3994967	-0.0364670
Uniform-like ²	0.00	-1.20	0.00	1.2237300	0.00	-0.0636881
Logistic-like ²	0.00	1.30	0.00	0.8807330	0.00	0.0382866
Double exponential-like ²	0.00	3.00	0.00	0.7823562	0.00	0.06790456

Note: 1 indicated distributions adopted from Algina, Olejnik, and Ocanto (1989).

2 indicated distributions adopted from Penfield (1994).

Sampling

After defining population distributions for the two samples, the size of the two samples was defined for this simulation study. Moreover, the condition related to sampling such as ratios between the two population variances, which these two samples were generated from, also affected these two samples. Therefore, it was important to introduce pair combinations of the two sample sizes and ratio combinations of two standard deviations between these two populations ($\frac{\sigma_1}{\sigma_2}$) when implementing this Monte Carlo simulation study. Furthermore, sampling procedures for simulations were described in this section.

Sample Size Determination

Because most reviewed studies in the literature performed simulations in both equal and unequal sample size scenarios, some significant findings were uncovered when examining Type I error rates and power for the MW test or the KS-2 test in the condition of equal sample sizes. Even though the first research question was to detect Type I error rates and power only when sample sizes were not equal to each other, the equal sample size condition was also simulated in this study. Due to the nominal Type I error (significance level) of this study of 0.05, the selected sample size combinations were based on the literature. Both equal and unequal sample size conditions were examined since statistical tests may behave differently under these sample size conditions. Small equal sample size combinations included (8, 8) and (16, 16), as used in Zimmerman and Zumbo (1990). The smallest sample size combination of (8, 8) was used in both studies by Zimmerman and Zumbo (1990) and Schroer and Trenkler (1995) as the smallest sample sets with the significance level (α) of 0.05. In Zimmerman and Zumbo's study, it was found that the actual Type I error rate of the

MW test was close to 0.05 when population variances were not the same. Similarly, Schroer and Trenkler (1995) used (8, 8) as the smallest sample sets to simulate statistical power of the MW test and the KS-2 test with the significance level (α) of 0.05.

Large equal sample size combinations were (25, 25) and (50, 50) as suggested in Baumgrater, WeiB, and Shindler (1998). In Baumgrater, WeiB, and Shindler's 1998 study, there was no result to explain the performance of Type I error rates and statistical power when no mean differences existed between two samples with the sizes of (25, 25) and (50, 50). However, it was crucial for current study since one of the considerations was to detect Type I error rates and statistical power when two samples were equal and large in sizes with no concern of any mean differences. Unequal sample size combinations included (4, 16) and (16, 4) in Zimmerman (1985), (10, 20) in Penfield (1994), and (30, 10) in Kasuya (2001). The researcher also investigated the conditions of (20, 10) and (10, 30) in order to compare with Zimmerman's study. Two other size combinations of (50, 100) and (100, 50) also were used to detect Type I error rates and statistical power when the differences between two sample sizes that were at least 50.

Since these collections of both equal and unequal sample size combinations were used in the MW test, the simulation results presented here either validated or revoked the literature. Moreover, the combinations were selected to allow for comparisons of the results for the MW test and the KS-2 test to draw conclusions for the research questions of this study.

Ratios of Two Standard Deviation Conditions (SD ratios or $\frac{\sigma_1}{\sigma_2}$)

One of the research questions in this study involved examining Type I error rates and power estimates of the MW test and the KS-2 test with the condition of heterogeneity of

variances of the populations. Unequal standard deviations between two populations were used in the study. The considered SD ratios ($\frac{\sigma_1}{\sigma_2}$) were 1, 2, 3, and 4 from Zimmerman (1998). Further, $\frac{1}{2}$, $\frac{1}{3}$, and $\frac{1}{4}$ were examined based on the researcher's interest of the idea of variance ratios by Gibbons and Chakraborti (1991). The selected SD ratios were used in the simulation along with other conditions to compare results for the MW test and the KS-2 test, and to draw conclusions based on the research questions of this research in consideration with the literature.

Sampling Procedure

After the fifteen population distributions were simulated from Fleishmen's power function with the associated coefficients (a, b, c, and d), the desired samples were randomly generated based on the determined conditions. These conditions were sample sizes and ratios of standard deviations between the two populations distributions listed in Table AV. After specifying the pair combinations of sample sizes and ratios between the two population standard deviations, the SAS/RANNOR procedure were implemented to generate two sample data sets and then the comparison of the MW and the KS-2 tests were performed by the SAS/NPAR1WAY procedure. A sample of SAS syntax was in APPENDIX III. Overall, the design of the simulation followed the elements of first part of each research question:

Question 1: If only sample sizes differ between two samples,

- a. Is there any difference in Type I error rates for these two nonparametric techniques?

The main concern of this research question was sample sizes, so the simulation was performed under the same population distributions and equal SD ratio between two samples.

Table 5: Summary of Conditions for Monte Carlo Simulations

	Distribution	Skewness (γ_1)	Kurtosis (γ_2)	Sample Size	SD Ratio	Simulation
					$\frac{\sigma_1}{\sigma_2}$	
1	Normal	0.00	0.00	(8, 8)	4	Type I Rate
2	Platykurtic	0.00	-0.50	(16, 16)	3	Power Estimates
3	Normal Platykurtic	0.00	-1.00	(25, 25)	2	---
4	Leptokurtic ¹	0.00	1.00	(50, 50)	1	---
5	Leptokurtic ²	0.00	2.00	(4, 16)	$\frac{1}{2}$	---
6	Leptokurtic ³	0.00	3.75	(16, 4)	$\frac{1}{3}$	---
7	Skewed	0.75	0.00	(10, 20)	$\frac{1}{4}$	---
8	Skewed and platykurtic ¹	0.50	-0.50	(20, 10)	---	---
9	Skewed and platykurtic ²	0.25	-1.00	(10, 30)	---	---
10	Skewed and leptokurtic ¹	0.75	3.75	(30, 10)	---	---
11	Skewed and leptokurtic ²	1.25	3.75	(50, 100)	---	---
12	Skewed-leptokurtic	1.75	3.75	(100, 50)	---	---
13	Uniform-like	0.00	-1.20	---	---	---
14	Logistic-like	0.00	1.30	---	---	---
15	Double exponential-like	0.00	3.00	---	---	---

Therefore, it was involved 15 distributions \times 8 sample sizes \times 1 ratio of between two standard deviations \times 1 run (Type I error rate) for a total of 120 conditions.

b. Is there any difference in power for these two nonparametric techniques?

This research question was not only sample sizes but changed in SD ratios, so the simulation was executed under the same population distributions but different sample sizes and SD ratios. Then, the simulation conditions were involved 15 distributions \times 8 sample sizes \times 6 ratios of between two standard deviations \times 1 run (Power) for a total of 720 conditions.

Question 2: If only the heterogeneity of variance between two populations exists, is there any difference in power for these two nonparametric techniques?

The considerations of this research question were different population variances in the same population distributions and equal sample sizes, so the simulation involved 15 distributions \times 4 sample sizes \times 6 ratios of between two standard deviations \times 1 run (Power) for a total of 360 conditions.

Question 3: If the nature of the underlying population distributions varies in skewness only, is there any difference in power for these two nonparametric techniques?

The third research question involved different skewness but the same kurtosis under the conditions of equal sample sizes and SD ratios. Among these 15 population distributions, shown in Table 4, normal and skew distributions with the same degrees of kurtosis ($\gamma_2 = 0$) were the first two population distributions for the comparison. For example, one pair of population distributions \times 4 sample sizes \times 1 ratio of between two standard deviations \times 1 run (Power) for a total of four conditions for this paired populations. Platykurtic and skewed

and platykurtic¹ with the same degrees of kurtosis ($\gamma_2 = -.050$) were the second two population distributions for the comparison, so 1 pair of population distributions \times 4 sample sizes \times 1 ratio of between two standard deviations \times 1 run (Power) for a total of four conditions for this paired set of populations. Normal platykurtic and skewed and platykurtic² with the same degrees of kurtosis ($\gamma_2 = -1.00$) was the third two population distributions for the comparison, so 1 pair of population distributions \times 4 sample sizes \times 1 ratio of between two standard deviations \times 1 run (Power) for a total of four conditions for this paired set of populations. The footnote notation in this section indicated the associated population distribution shown in Table 4. Lastly, four distributions (Leptokurtic³, Skewed and leptokurtic¹, Skewed and leptokurtic² and Skewed-leptokurtic) with the same degrees of kurtosis ($\gamma_2 = 3.75$) but with different skewness were used to perform pair-wise comparisons. Therefore, 6 paired population distributions \times 4 sample sizes \times 1 ratio of between two standard deviations \times 1 run (Power) for a total of 24 conditions for this paired set of populations. In conclusion, a total of 36 ($4 + 4 + 4 + 24 = 36$) conditions were performed for examining the third research question.

Question 4: If the nature of the underlying population distributions varies in kurtosis only, is there any difference in power for these two nonparametric techniques?

The fourth research question considered different kurtosis but the same skewness, sample sizes, and SD ratios. Among these 15 population distributions in Table 4, exactly 9 populations had the same skewness ($\gamma_1 = 0$) but vary in kurtosis. These nine were pair-wise compared to fulfill the fourth research question. So, $36 (= \frac{9 \times 8}{2})$ paired population distributions \times 4 sample sizes \times 1 ratio of between two standard deviations \times 1 run (Power)

for a total of 144 conditions for this paired set of populations. Two other population distributions also had the same skewness ($\gamma_1 = 0.75$) but varying in kurtosis, so 1 pair of population distribution \times 4 sample sizes \times 1 ratio of between two standard deviations \times 1 run (Power) for a total of 4 conditions for this paired set of populations. Therefore, a total of 148 ($144 + 4 = 148$) conditions were performed for examining the fourth research question.

Exactly 20,000 replications per condition were employed to simulate the Type I error rate and power of both tests. The nominal Type I error rate (α) for this study was 0.05 and was used for the comparisons with actual Type I error rates. Thus, the performance of the MW and the KS-2 tests under each evaluated condition were examined. All simulated data were rounded to three digits.

Test Statistics

Formulas of the test statistic for both small and large samples for the MW test and the KS-2 test were listed in CHAPTER TWO. Each of the two tests was applied to the generated data samples. Two-tailed tests will investigate statistical differences between two simulated samples under each determined condition (832 conditions in total) at the nominal alpha level (α) of 0.05 by the SAS/NPAR1WAY program.

The Mann-Whitney Test Used in This Study

Based upon the literature reviewed in CHAPTER TWO, this researcher summarized and modified *1) assumptions and data arrangements, 2) hypotheses, and 3) formulas of test statistics and decision rules for small and large sample sizes* and presented the MW test that will be used for this study.

1) Assumptions and Data Arrangements

The assumptions for applying the MW test are as follows:

- (1) Each sample score has been randomly selected from the population it represents.
- (2) The originally observed sample score is a continuous variable.
- (3) Two sample scores are randomly selected and score sets are mutually independent.
- (4) The measurement scale employed is at least ordinal.

The data Arrangement shows the expression of arranging data after the data sets are obtained to use with the MW test technique.

Let X_1, X_2, \dots, X_{n_1} denote the random sample scores size n_1 with an expected smaller sum of ranks.

Let X_1, X_2, \dots, X_{n_2} denote the random sample scores size n_2 with an expected larger sum of ranks.

Assign the ranks 1 to $(n_1 + n_2)$ to the observations from the smallest to the largest.

Let $N = n_1 + n_2$.

2) Applicable Hypotheses

Because this research is designed to detect the alternative hypothesis that there are differences between two sampled population distributions, the non-directional hypothesis (two-tailed test) of the test is:

$H_0: F(x) = G(x)$ for all x , or there is no difference between the two populations.

$H_a: F(x) \neq G(x)$ for some x , or there are some differences between the two populations.

Where $F(x)$ is the population distribution function of the sum of the ranks of the sample expected to have the smaller sum, and $G(x)$ the population distribution function of the sum of the ranks of the sample expected to have the larger sum.

3) *Formulas of Test Statistics and Decision Rules for Small and Large Sample Sizes*

Test statistics are used to calculate the value needed to perform the hypothesis test. Because of the ease of understanding and calculating the formula and consistent with the procedure in SAS/NPAR1WAY , the test statistic used in this research is adapted from the Test statistics W method proposed by Siegel and Castellan (1988).

Small Sample Size in each group ($n_1 \leq 20$; $n_2 \leq 20$)

$W_x = \sum R_1$; the sum of the ranks of multiple variables of Xs from population 1s

$W_y = \sum R_2$; the sum of the ranks of multiple variables of Ys from population 2s

$$W_x + W_y = \frac{N(N+1)}{2}, \text{ where } N = n_1 + n_2$$

The smaller value of W_x and W_y is used as the test statistic.

The decision rule is:

If the probability of the observed W found in the table is less than the specific level of significance (α), the null hypothesis is rejected and there is a significant difference between these two populations.

When the sample size is more than 20 ($n_1 > 20$ or $n_2 > 20$), the formula for the normal approximation is used, which is:

$$Z = \frac{W_x \pm 0.5 - \frac{n_1(N+1)}{2}}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}}, \text{ where } W_x = \sum R_1.$$

The decision rule is:

If calculated absolute Z is greater than the tabled Z value with the $\alpha/2$ level, then reject the null hypothesis.

Siegel and Castellan (1988) suggested that the test statistics be applied to investigate whether two independent samples have been drawn from the same population or whether the two populations have the same medians. The test statistics are also used to test whether the probability of population X greater than population Y ($P(X>Y)$) is the same as the probability of population X less than population Y ($P(X<Y)$) which is equal to 0.5. On the issue of ties, Siegel and Castellan did not specify the minimum number of ties in order to use the formula for the ties situation.

The Kolmogorov-Smirnov Two-Sample Test Used in This Study

After reviewing the literature as presented in CHAPTER TWO, the following elements are recommended for applying the KS-2 test: 1) *assumptions and data arrangements*, 2) *hypotheses*, and 3) *formulas of test statistics and decision rules for small and large sample sizes*.

1) *Assumptions and Data Arrangements*

Assumptions similar to Conover (1999) are suggested for this study. There are four assumptions as followed:

- (1) Each sample has been randomly selected from the population it represented.

- (2) The measurement scale employed is at least ordinal.
- (3) The originally observed variable is a continuous variable.
- (4) Two samples are mutually independent.

The data arrangement proposed by Siegel and Castellan (1988) was modified and used in this study:

Let $S_1(x)$ be the cumulative distribution probability function (cdf's) based upon the random sample scores of X_1, X_2, \dots, X_{n_1} .

Determine the $S_1(x)$ for each value of X_1, X_2, \dots, X_{n_1} , let $S_1(x) = \frac{K}{n_1}$.

Let $F(x)$ be the population that the sample of X 's are randomly drawn from.

Let $S_2(x)$ be the cumulative distribution probability function (cdf's) based upon the random sample scores of Y_1, Y_2, \dots, Y_{n_2} .

Determine $S_2(x)$ for each value of Y_1, Y_2, \dots, Y_{n_2} , let $S_2(x) = \frac{K}{n_2}$.

Let $G(x)$ be the population that the sample of Y 's are randomly drawn from.

D_{n_1, n_2} is symbolized as the test statistic for the KS-2 test. It is the maximum absolute difference between the two empirical distribution functions or cumulative distribution functions.

2) *Applicable Hypotheses*

Because this research is designed to detect the alternative hypothesis that there are differences between two sampled population distributions, the non-directional hypothesis (two-tailed test) of the test is:

H_0 : there is no difference between two populations, or

$F(x) = G(x)$ for all x ; from $-\infty$ to $+\infty$.

H_a : there are some differences between two populations, or

$F(x) \neq G(x)$ for at least one value of x .

3) Formulas of Test Statistics and Decision Rules for Small and Large Sample Sizes.

Formulas of the test statistic (D_{n_1, n_2}) for both small and large sample conditions as well as decision rules for the testing the hypotheses are presented. To be consistent in the definition of sample sizes for comparison with the MW test, a size of 20 was selected as the boundary to define small and large sample sizes.

Small Sample Size ($n_1 \leq 20$ or $n_2 \leq 20$)

When both samples are no more than 20 ($n_1 \leq 20$ or $n_2 \leq 20$), the test statistic of the KS-2 test is:

$$D_{n_1, n_2} = \max_x |S_{n_1}(x) - S_{n_2}(x)|$$

The decision rule of the hypothesis test is:

If the observed $D_{m, n}$ is greater than or equal to the tabled $D_{n_1, n_2 \text{ critical}}$ ($D_{n_1, n_2} \geq D_{n_1, n_2 \text{ critical}}$) at the specific level of significance (α), the null hypothesis is rejected.

Therefore, there is a significant difference between these two populations.

Large Sample Size ($n_1 > 20$ or $n_2 > 20$)

When either or both samples are larger than 20: ($n_1 > 20$ or $n_2 > 20$), the test statistic of the KS-2 test is:

$$D_{n_1, n_2} = \max_x |S_{n_1}(x) - S_{n_2}(x)|$$

Critical D_{n_1, n_2} is calculated with a formula based on various significance level (α).

When the significance level is α ,

$$\text{the critical } D_{n_1, n_2} = \text{table value (K)} \times \sqrt{\frac{n_1 + n_2}{n_1 n_2}}.$$

The decision rule is:

When the observed D_{n_1, n_2} is greater than or equal to the tabled $D_{n_1, n_2 \text{ critical}}$ ($D_{n_1, n_2} \geq D_{n_1, n_2 \text{ critical}}$) at the specific level of significance (α), the null hypothesis is rejected.

Therefore, a significant difference probably exists between these two populations.

Simulation Steps

In order to assist in performing the Monte Carlo simulation study for a two-tailed test, the simulation steps were described here to avoid any confusion in executing the simulations.

These six steps included:

Step 1

Use Fleishmen's power function (1978) with $\mu = 0$; $\sigma = 1$ for generating these 15 population distributions. The coefficients provided in Table AII and were used and 15 population distributions were generated by executing the SAS/RANNOR program.

Step 2

Determine the null and alternative hypotheses for each comparison and the significance levels for each comparison ($\alpha = 0.05$). Then, determine the formulas of test statistic U for the MW test and test statistic D for the KS-2 test (described in the section of Test Statistics).

Step 3

Generate two independent random samples of size n_1 and n_2 , respectively, from the sixteen population distributions with the specified ratios of the two population standard deviations ($\frac{\sigma_1}{\sigma_2}$). The pair combinations of sample sizes and ratios of population standard deviations were listed in Table 4.

Step 4

Calculate the values of test statistics of the MW test (U) and the KS-2 test (D), based on the generated two independent samples in Step 3.

Step 5

Compare W with critical W and D with critical D and determine whether to reject or retain the null hypothesis (H_0) by the decision rules in the section of Test Statistics by utilizing SAS/NPAR1WAY procedure.

Step 6

About 20,000 replications per condition 1 were required when performing this simulation. (Computer was automatically to repeat the first five steps 20,000 times and count the total number of times H_0 is rejected for the MW and the KS-2 tests, and obtain the proportion of rejections for each test by using SAS/RANNOR procedure.) Gibbons and Chakraborti (1991) noted that “theses proportions provide estimates of the probability information of rejection by the respective tests for particular configuration of means, variances, and sample sizes” (p. 261).

Summary

In this Monte Carlo simulation study, the researcher examined Type I error rates and statistical power when applicable under each predetermined condition. There are 15 population distributions, 12 sets of sample size combinations, and 7 different ratios of standard deviation. Exactly 20,000 replications per condition were executed for a total of 1380 conditions (840 for the first research question, 360 for the second question, 36 for the third research question, and 144 for the fourth research question). Moreover, the steps for performing this simulation study were also described in this chapter.

The SAS/RANNOR procedure was used to generate sample data sets for population distributions. These distributions came from Fleishmen's power function (1978) by using the coefficients listed in Table 4. A summary of the types of population distributions, combinations of the sample sizes, and ratios of standard deviation (SD ratios) to be used here were listed in Table 5. Suggested formulas of test statistics for the MW test and the KS-2 test were also presented. The SAS/NPAR1WAY procedure was used to simulate Type I error rates and statistical power for the MW and the KS-2 tests in each condition when applicable. Moreover, simulation steps were used and followed when performing Monte Carlo simulations in order to eliminate any confusion when the researcher performs the simulations. The results of simulations were presented in CHAPTER IV.

CHAPTER FOUR

RESULTS

Introduction

In this chapter, estimated Type I error rates and statistical power for the Mann-Whitney test (MW) and Kolmogorov-Smirnov two-sample test (KS-2) under various conditions are presented and discussed at the significance level (α) of .05. In order to help the researcher better understand the shape of fifteen populations discussed in this study, figures of these fifteen population distributions that were simulated are presented in Appendix IV: Histograms of fifteen population distributions. Furthermore, tables and figures are provided based on the order of the research questions in this study. In research questions one and two, only crucial tables and scatter plots are used to display the results of these simulations. The complete results of tables for these two questions are presented in Appendix V: Tables of findings. Findings of research questions three and four are displayed as tables and are presented in this chapter.

Findings

Findings of this study are presented based upon the arrangement of the research questions. Significant findings are provided for research questions one and two. The results of research questions three and four are presented later this chapter.

Research Question 1: If only sample sizes differ between two samples,

- a. Is there any difference in Type I error rate for these two nonparametric techniques?

In this research question, the researcher simulated the conditions that two samples were from the same population distribution with the same SD ratios, but they differed in sample sizes. Exactly eight pairs of unequal sample sizes from the same population distribution (about 15 population distributions in total) and SD ratio of 1 were simulated. The MW and the KS-2 tests were performed to examine the simulated Type I error rates for both nonparametric statistical techniques.

Table 5 illustrates simulated Type I error rates for both the MW and the KS-2 tests after performing the simulations. The table shows that when sample sizes were unequal and small, such as (4, 16) and (16, 4), the simulated Type I error rates were less than 0.020 for all 15 population distributions when performing the KS-2 test. When samples were unequal and increased by size, the simulated Type I error rates were also raised. Overall, estimated Type I error rates for these fifteen population distributions were less than the significance level (α) of 0.05 in the KS-2 test. When the MW test was executed for the simulated samples, it was found that the range of estimated Type I error rates for these fifteen populations was between 0.046 and 0.055. Most of the estimated Type I error rates were less than the α level of 0.05. It seemed that there was no increase in estimated Type I error rates as the sample sizes increased. Both the KS-2 test and the MW test were found to yield consistent results across these 15 population distributions.

After investigating the results of the MW test and the KS-2 test, it appeared that the KS-2 test was a more conservative test than the MW test based upon the conditions discussed in the first part of the of the first research question.

Table 6: Type I Error Rates: Only Sample sizes Differ between Two Samples (SD Ratio = 1)

POPULATION	SAMPLE SIZE	TYPE I ERROR		POPULATION	SAMPLE SIZE	TYPE I ERROR	
		MW	KS-2			MW	KS-2
Normal	(4, 16)	.048*	.014*	Platykurtic	(4, 16)	.052	.016*
	(16, 4)	.051	.015*		(16, 4)	.049*	.014*
	(10, 20)	.050*	.030*		(10, 20)	.049*	.029*
	(20, 10)	.051	.029*		(20, 10)	.049*	.028*
	(10, 30)	.048*	.036*		(10, 30)	.049*	.039*
	(30, 10)	.051	.038*		(30, 10)	.048*	.035*
	(50, 100)	.052	.043*		(50, 100)	.049*	.039*
	(100, 50)	.050*	.043*		(100, 50)	.047*	.041*
Normal Platykurtic	(4, 16)	.049*	.014*	Leptokurtic_1	(4, 16)	.049*	.014*
	(16, 4)	.050*	.014*		(16, 4)	.050*	.015*
	(10, 20)	.050*	.028*		(10, 20)	.050*	.028*
	(20, 10)	.048*	.029*		(20, 10)	.048*	.028*
	(10, 30)	.047*	.035*		(10, 30)	.051	.039*
	(30, 10)	.051	.038*		(30, 10)	.051	.038*
	(50, 100)	.052	.043*		(50, 100)	.048*	.040*
	(100, 50)	.051	.041*		(100, 50)	.049*	.040*

* indicated the simulated Type I Error Rate was less than the significance level (α) of .05.

Table 6 CONT: Type I Error Rates: Only Sample sizes Differ between Two Samples (SD Ratio = 1)

POPULATION	SAMPLE SIZE	TYPE I ERROR		POPULATION	SAMPLE SIZE	TYPE I ERROR	
		MW	KS-2			MW	KS-2
Leptokurtic_2	(4, 16)	.048*	.013*	Leptokurtic_3	(4, 16)	.051	.014*
	(16, 4)	.051	.014*		(16, 4)	.051	.016*
	(10, 20)	.049*	.029*		(10, 20)	.049*	.029*
	(20, 10)	.049*	.029*		(20, 10)	.048*	.028*
	(10, 30)	.052	.038*		(10, 30)	.050*	.036*
	(30, 10)	.051	.038*		(30, 10)	.051	.038*
	(50, 100)	.047*	.040*		(50, 100)	.050*	.041*
	(100, 50)	.048*	.041*		(100, 50)	.051	.037*
Skewed	(4, 16)	.050*	.016*	Skewed and Platykurtic_1	(4, 16)	.049*	.015*
	(16, 4)	.050*	.015*		(16, 4)	.051	.015*
	(10, 20)	.047*	.029*		(10, 20)	.052	.030*
	(20, 10)	.048*	.030*		(20, 10)	.049*	.029*
	(10, 30)	.053	.037*		(10, 30)	.051	.038*
	(30, 10)	.051	.039*		(30, 10)	.048*	.037*
	(50, 100)	.051	.039*		(50, 100)	.047*	.038*
	(100, 50)	.051	.041*		(100, 50)	.049*	.040*

* indicated the simulated Type I Error Rate was less than the significance level (α) of .05.

Table 6 CONT.: Type I Error Rates: Only Sample sizes Differ between Two Samples (SD Ratio = 1)

POPULATION	SAMPLE SIZE	TYPE I ERROR		POPULATION	SAMPLE SIZE	TYPE I ERROR	
		MW	KS-2			MW	KS-2
Skewed and Platykurtic_2	(4, 16)	.047*	.014*	Skewed and Leptokurtic_1	(4, 16)	.050*	.016*
	(16, 4)	.050*	.014*		(16, 4)	.049*	.015*
	(10, 20)	.047*	.030*		(10, 20)	.052	.030*
	(20, 10)	.048*	.029*		(20, 10)	.050*	.030*
	(10, 30)	.048*	.036*		(10, 30)	.049*	.037*
	(30, 10)	.051	.037*		(30, 10)	.049*	.038*
	(50, 100)	.048*	.040*		(50, 100)	.049*	.040*
	(100, 50)	.050*	.041*		(100, 50)	.055	.045*
Skewed and Leptokurtic_2	(4, 16)	.050*	.014*	Skewed- Leptokurtic	(4, 16)	.051	.016*
	(16, 4)	.050*	.014*		(16, 4)	.053	.015*
	(10, 20)	.046*	.028*		(10, 20)	.048*	.028*
	(20, 10)	.050*	.030*		(20, 10)	.046*	.029*
	(10, 30)	.050*	.035*		(10, 30)	.051	.039*
	(30, 10)	.050*	.037*		(30, 10)	.051	.039*
	(50, 100)	.049*	.040*		(50, 100)	.047*	.038*
	(100, 50)	.051	.039*		(100, 50)	.048*	.039*

* indicated the simulated Type I Error Rate was less than the significance level (α) of .05.

Table 6 CONT.: Type I Error Rates: Only Sample sizes Differ between Two Samples (SD Ratio = 1)

POPULATION	SAMPLE SIZE	TYPE I ERROR		POPULATION	SAMPLE SIZE	TYPE I ERROR	
		MW	KS-2			MW	KS-2
Uniform-Like	(4, 16)	.051	.015*	Logistic-Like	(4, 16)	.051	.015*
	(16, 4)	.049*	.015*		(16, 4)	.050*	.015*
	(10, 20)	.050*	.030*		(10, 20)	.049*	.030*
	(20, 10)	.047*	.028*		(20, 10)	.047*	.027*
	(10, 30)	.050*	.038*		(10, 30)	.046*	.036*
	(30, 10)	.051	.034*		(30, 10)	.050*	.037*
	(50, 100)	.050*	.041*		(50, 100)	.051	.041*
	(100, 50)	.051	.040*		(100, 50)	.050*	.039*
Double Exponential-Like	(4, 16)	.049*	.015*				
	(16, 4)	.051	.016*				
	(10, 20)	.047*	.027*				
	(20, 10)	.049*	.029*				
	(10, 30)	.050*	.037*				
	(30, 10)	.048*	.038*				
	(50, 100)	.048*	.042*				
	(100, 50)	.051	.042*				

* indicated the simulated Type I Error Rate was less than the significance level (α) of .05.

Research Question 1: If only sample sizes differ between two samples,

b. Is there any difference in power for these two nonparametric techniques?

When investigating part (b) of question one, the researcher decided to change the SD ratios. This was because when the SD ratios of the two samples were not equal to 1, the shapes of the population distributions were not the same between two samples. In other words, it was assumed that the null hypothesis $H_0: F(X) = G(X)$ was violated. Therefore, the p-values yielded from the MW test and the KS-2 test with differences in SD ratios served as statistical power (the probability of rejecting the false null hypothesis). The simulations were performed based upon the same population distributions but the SD ratios and the sample sizes were changed. The complete set of statistical power for all 15 population distributions are given in Tables 7 to 21 in APPENDIX V: Tables of Findings.

From Tables 7 to 21 in APPENDIX V, it was found that when the two independent samples were unequal, regardless of the sizes in 14 population distributions (except for the Skewed-Leptokurtic distribution, Figures 23), the estimated statistical power values of performing the MW test were all small. The values of statistical power for the MW test were from 0.006 to 0.300. The values of the statistical power for the KS-2 test were various based on the SD ratios and the sizes of the two samples. The range of the power of the KS-2 test was from 0.005 to 1.0. Figures 1 to 22 are the scatter plots of estimated statistical power for the MW and KS-2 tests when the SD ratios were 4, 3, $\frac{1}{3}$, and $\frac{1}{4}$ for 11 populations (except the Skewed, Skewed and Platykurtic_1, Skewed and Platykurtic_2, and Skewed-Leptokurtic distributions). These figures represented statistical power values based upon Tables 7 to 17 in APPENDIX V.

The other important finding from the majority of the population distributions (except the Skewed, Skewed and Platykurtic_1, Skewed and Platykurtic_2, and Skewed-Leptokurtic distributions) was: when the population standard deviation between two samples was very different and the size of the unequal samples were much different, such as from (4, 16) to (100, 50), the KS-2 test had higher statistical power than the MW test. Moreover, as the sample sizes were increased, the estimated statistical power of the MW test was consistent and did not increase. However, statistical power of the KS-2 test was dramatically increased as the sample sizes increased. When the sample sizes became (50, 100) and (100, 50), the estimated statistical power approached 1.0. In some conditions, for example, sample size = (100, 50) and SD ratio = 4 or $\frac{1}{4}$ in the Platykurtic distribution, the estimated statistical power was equal to one.

Figure 1: Power of the **Normal Distribution** when Sample Sizes Differ and SD ratios = 4 & 3
 Power

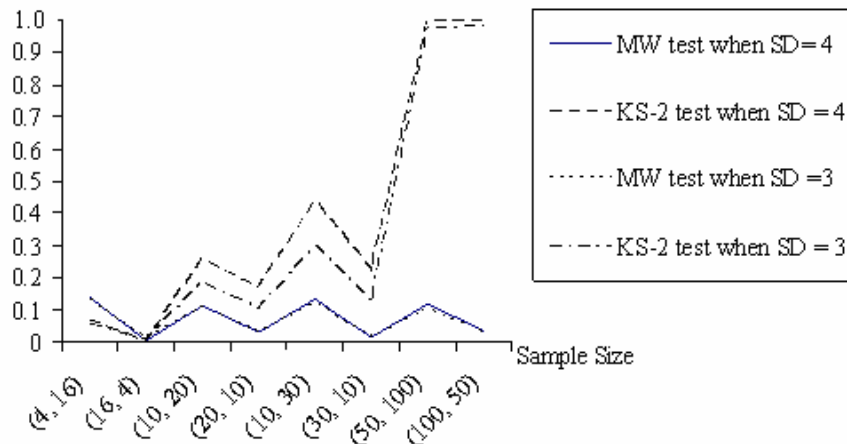


Figure 2: Power of the **Normal Distribution** when Sample Sizes Differ and SD ratios = 1/3 & 1/4

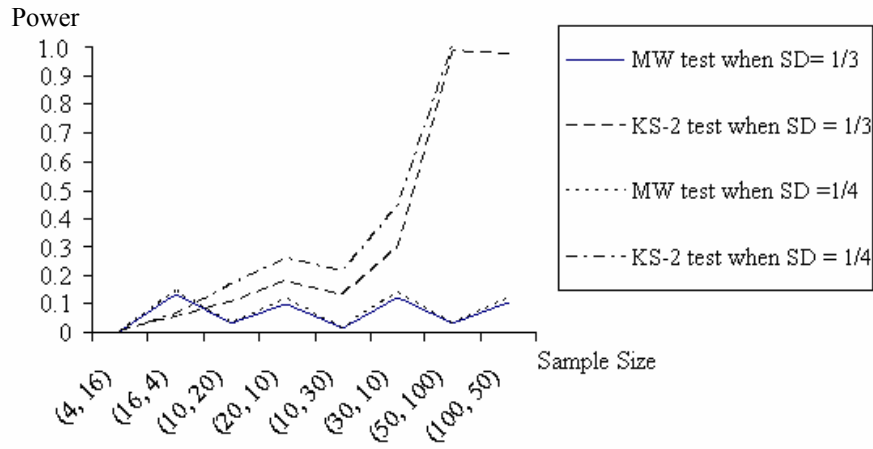


Figure 3: Power of the **Platykurite Distribution** when Sample Sizes Differ and SD ratios = 4 & 3

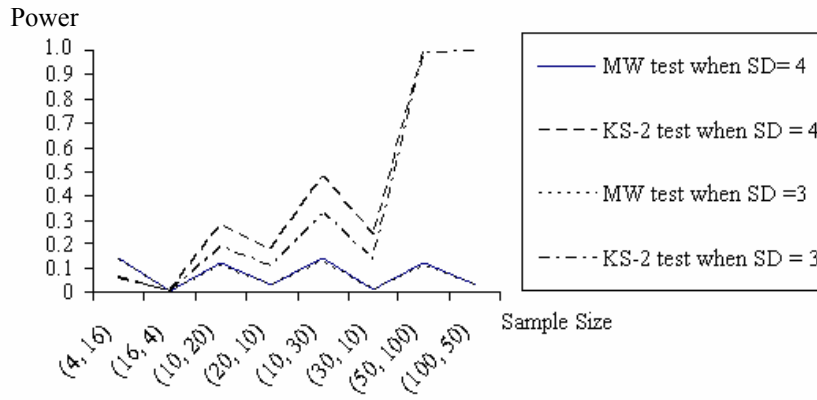


Figure 4: Power of the **Platykurite Distribution** when Sample Sizes Differ and SD ratios = 1/3 & 1/4

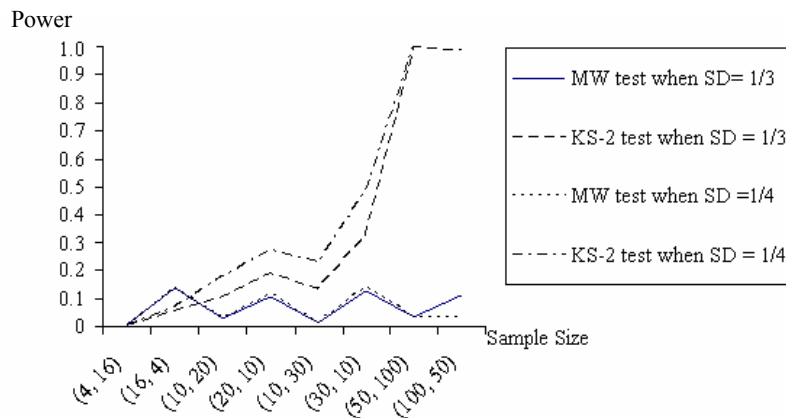


Figure 5: Power of the **Normal Platykurtic Distribution** when Sample Sizes Differ and SD ratios = 4 & 3

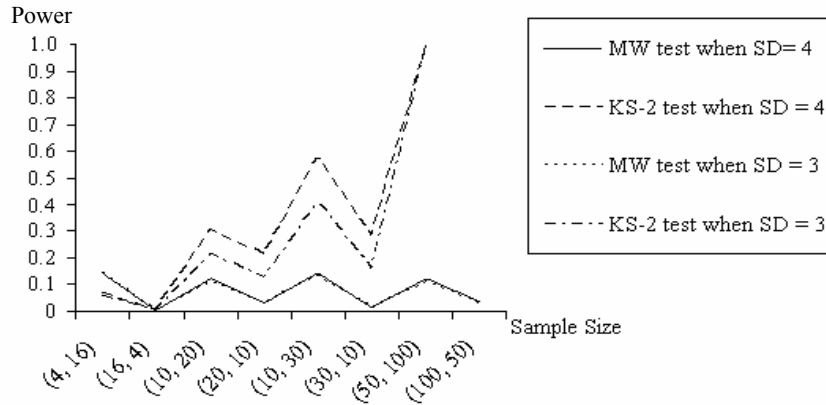


Figure 6: Power of the **Normal Platykurtic Distribution** when Sample Sizes Differ and SD ratios = 1/3 & 1/4

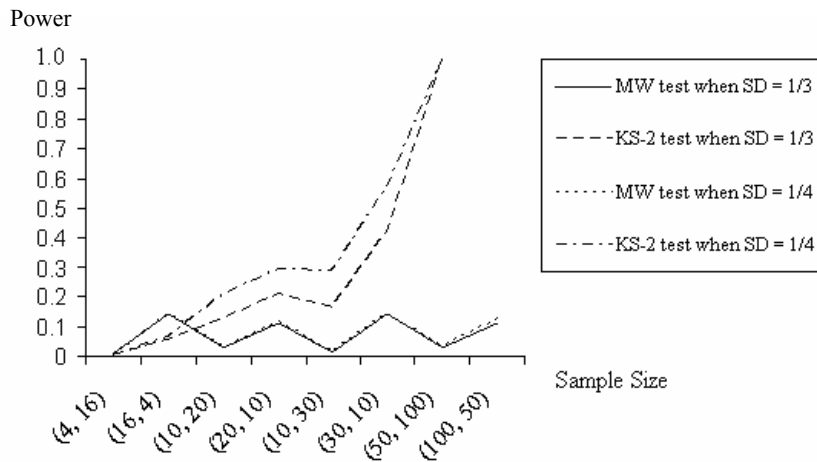


Figure 7: Power of the **Leptokurtic 1Distribution** when Sample Sizes Differ and SD ratios = 4 & 3

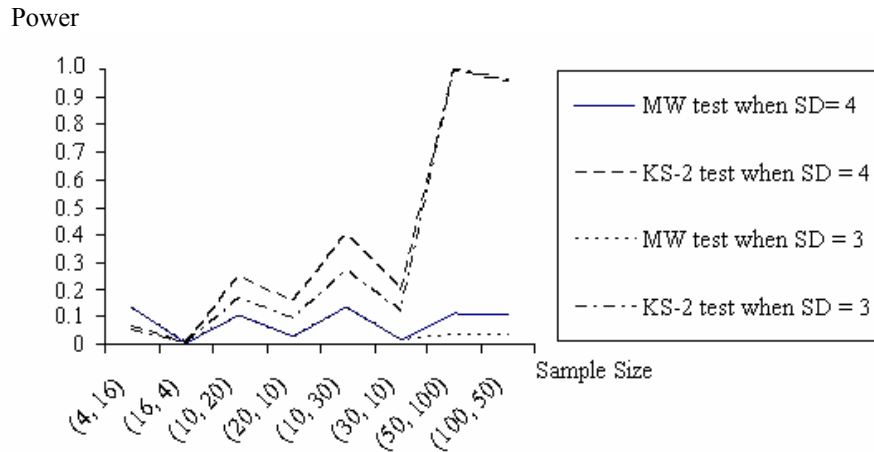


Figure 8: Power of the **Leptokurtic 1 Distribution** when Sample Sizes Differ and SD ratios = 1/3 & 1/4

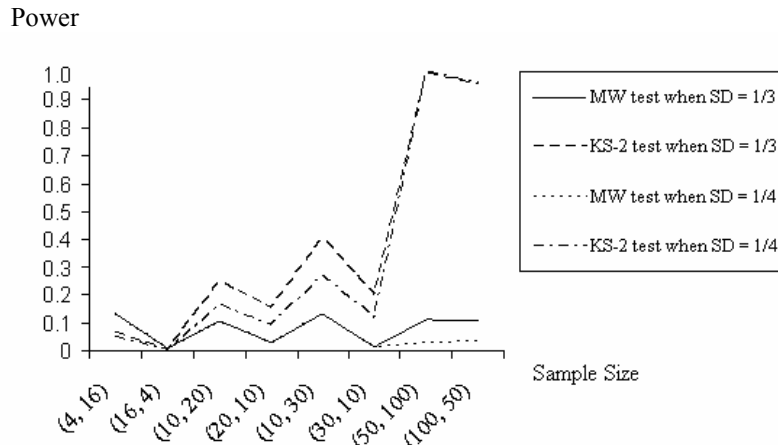


Figure 9: Power of the **Leptokurtic 2 Distribution** when Sample Sizes Differ and SD ratios = 4 & 3

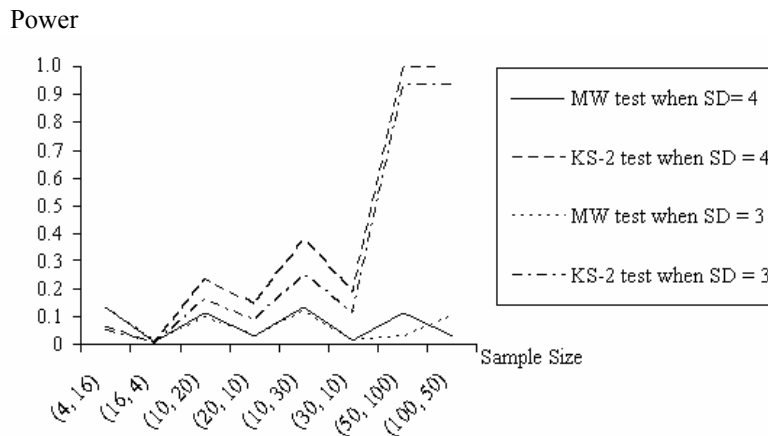


Figure 10: Power of the **Leptokurtic 2 Distribution** when Sample Sizes Differ and SD ratios = 1/3 & 1/4

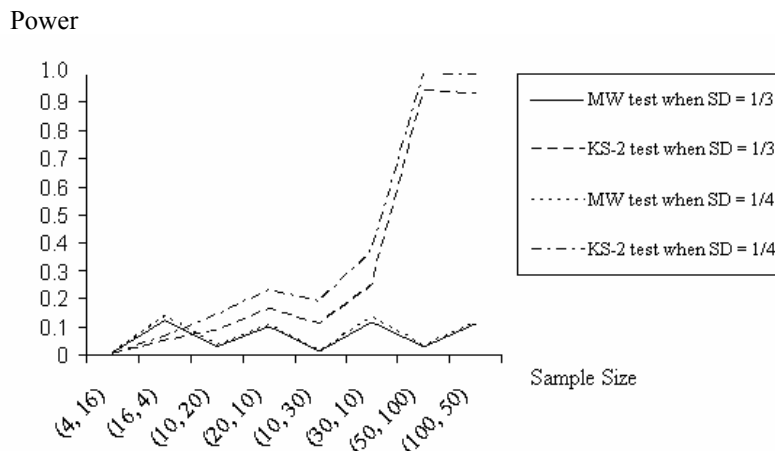


Figure 11: Power of the **Leptokurtic 3 Distribution** when Sample Sizes Differ and SD ratios = 4 & 3

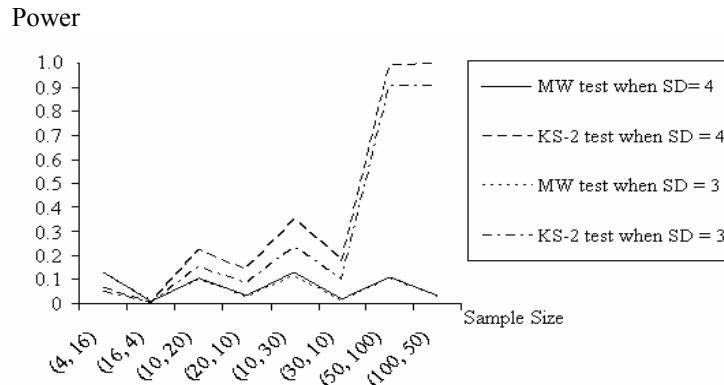


Figure 12: Power of the **Leptokurtic 3 Distribution** when Sample Sizes Differ and SD ratios = 1/3 & 1/4

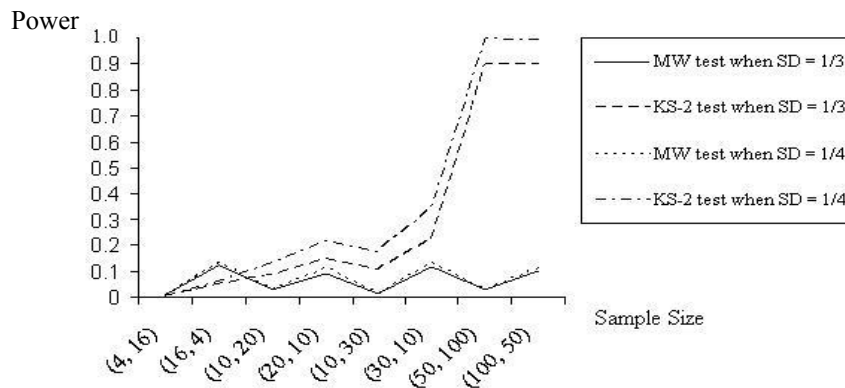


Figure 13: Power of the **Skewed and Leptokurtic 1 Distribution** when Sample Size Differ and SD ratios = 4 & 3

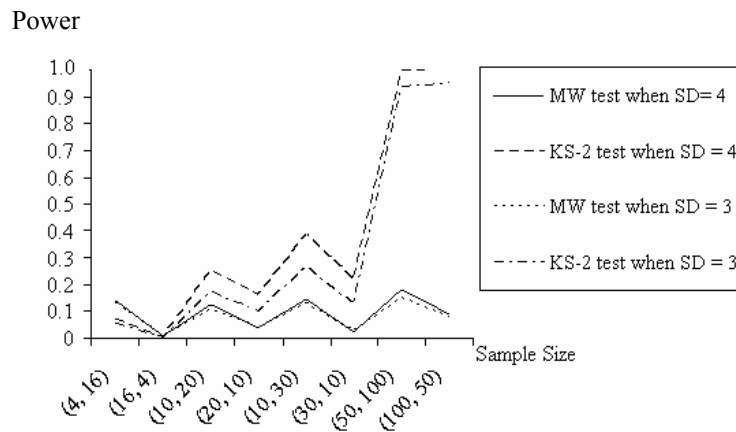


Figure 14: Power of the **Skewed and Leptokurtic 1** Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4

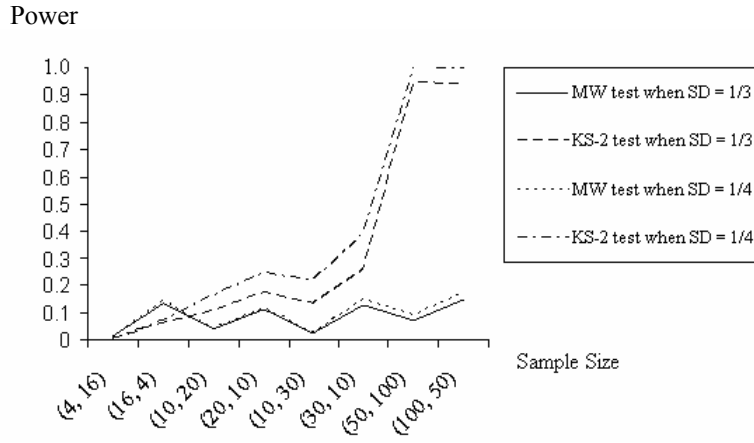


Figure 15: Power of the **Skewed and Leptokurtic 2** Distribution when Sample Sizes Differ and SD ratios = 4 & 3

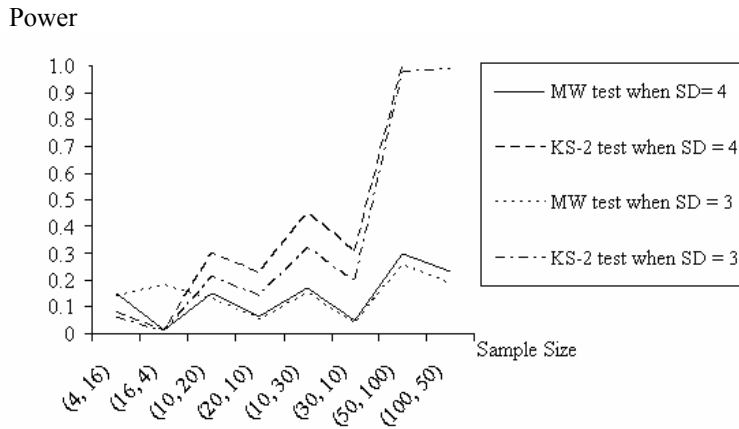


Figure 16: Power of the **Skewed and Leptokurtic 2** Distribution when Sample Sizes Differ and SD ratios = 1/3 & 1/4

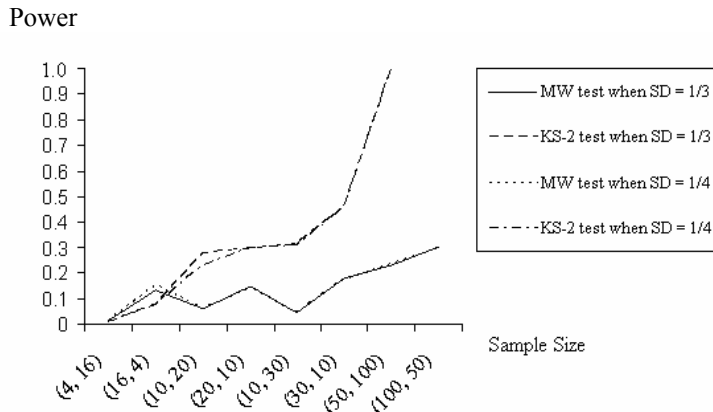


Figure 17: Power of the **Uniform-Like Distribution** when Sample Sizes Differ and SD ratios = 4 & 3

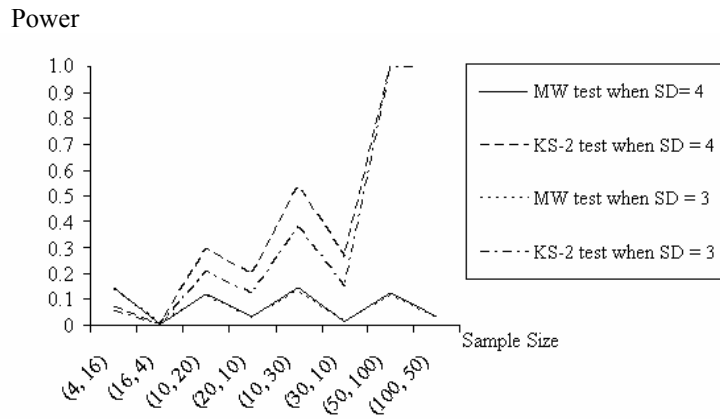


Figure 18: Power of the **Uniform-Like Distribution** when Sample Sizes Differ and SD ratios = 1/3 & 1/4

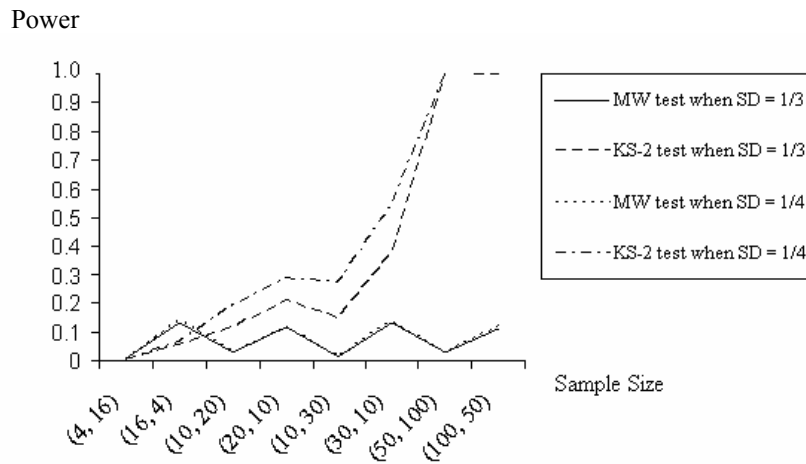


Figure 19: Power of the **Logistic-Like Distribution** when Sample Sizes Differ and SD ratios = 4 & 3

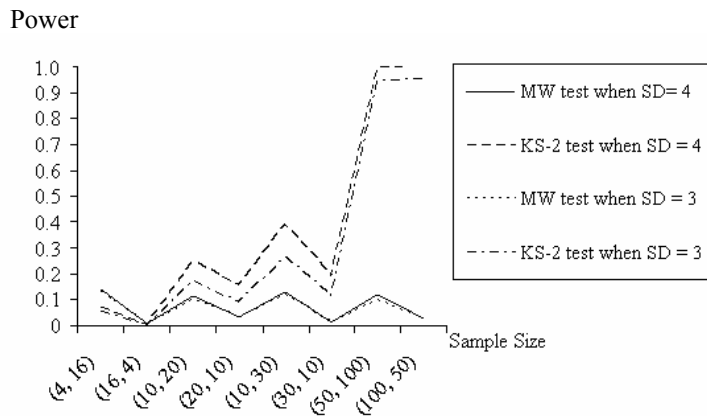


Figure 20: Power of the **Logistic-Like Distribution** when Sample Sizes Differ and SD ratios = 1/3 & 1/4

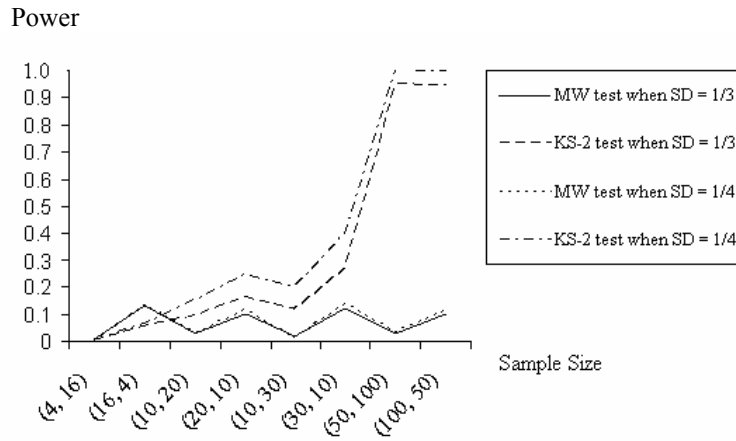


Figure 21: Power of the **Double Exponential -Like Distribution** when Sample Sizes Differ and SD ratios = 4 & 3

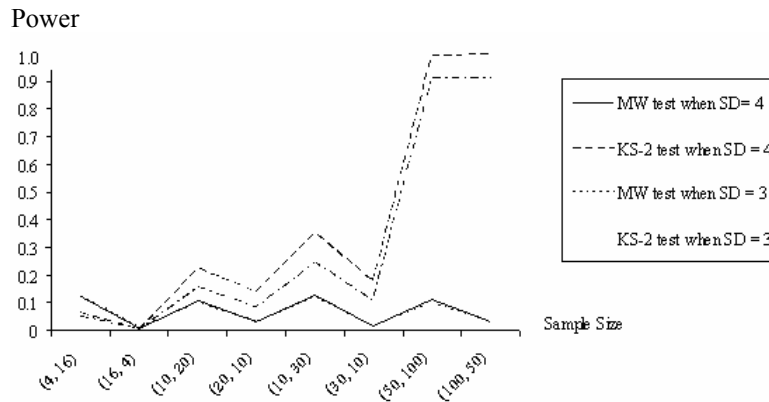
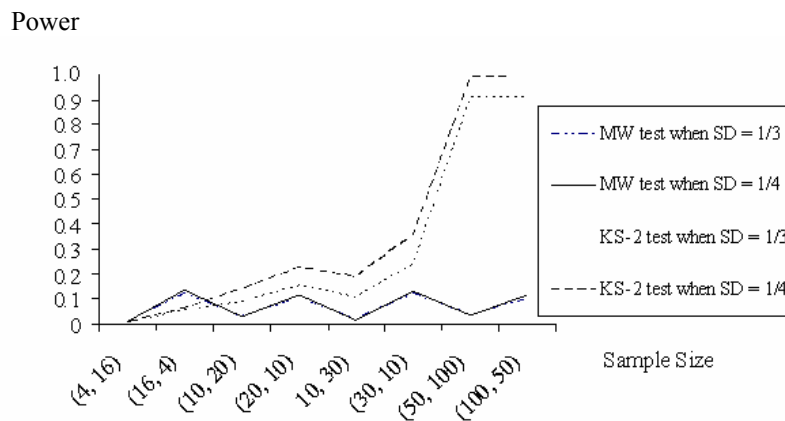


Figure 22: Power of the **Double Exponential -Like Distribution** when Sample Sizes Differ and SD ratios = 1/3 & 1/4



Figures 23 to 31 are the scatter plots of the simulated statistical power of the MW test and the KS-2 test when SD ratios were 4, 3, 2, and $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$ for the Skewed, Skewed and Platykurtic_1, Skewed and Platykurtic_2, and the Skewed-Leptokurtic distributions. These graphs show that the estimated statistical power of both the MW test and the KS-2 test were small when sample sizes were (4, 16) and (16, 4). The estimated power of the KS-2 test was smaller than or close to the MW test when sample sizes were (4, 16) and (16, 4).

When the two underlying populations are heavily skew to the left (the Skewed-Leptokurtic distribution, Figures 23), the estimated statistical power values of performing the MW test were all small (Figures 24 and 25 and Table 18 in APPENDIX V). The range of statistical power for the MW test was from 0.006 to 0.300. When the sample sizes increased to (50, 100) and (100, 50), the estimated statistical power at all six SD ratios were between 0.544 and 0.733 in a Skewed-Leptokurtic distribution. Moreover, when the sample with smaller size had larger population standard deviations, the estimated statistical power was greater than the condition of larger samples with small population standard deviations. When performing the KS-2 test on the same simulated sample sets, it was found that when sample sizes were unequal and small, the estimated statistical power was small too. When sample sizes were increased, the estimated statistical power increased as well. When the sample sizes were either (50, 100) or (100, 50), statistical power was substantially large and close to 1 when the SD ratios were 4, 3, $\frac{1}{3}$, and $\frac{1}{4}$. The range of estimated statistical power was between 0.911 and 1.0 for these conditions.

Figure23: Histogram of the **Skewed-Leptokurtic distribution** (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)

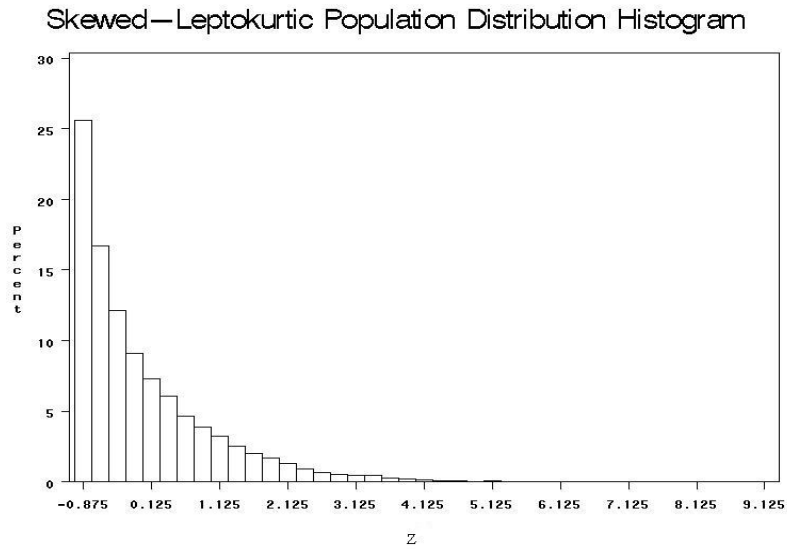


Figure 24: Power of the **Skewed-Leptokurtic Distribution** when Sample Size Differs and SD ratios = 4, 3 & 2

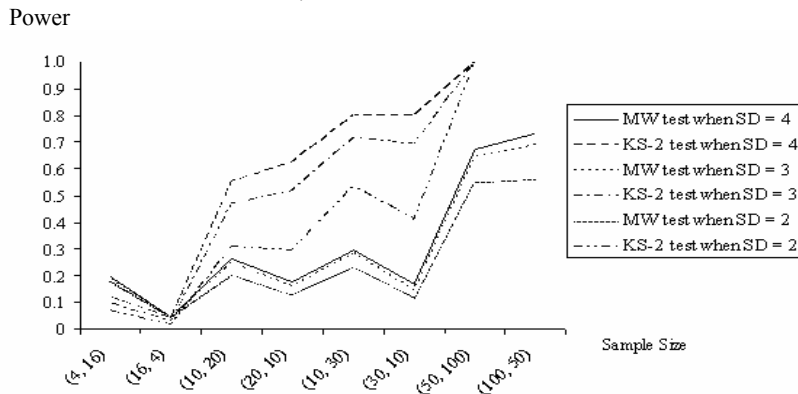
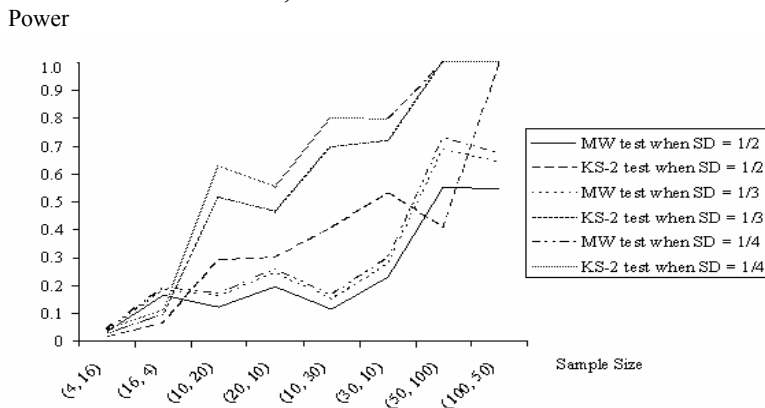


Figure 25: Power of the **Skewed-Leptokurtic Distribution** when Sample Size Differs and SD ratios = 1/2, 1/3 & 1/4



When the two underlying populations are positively skewed (Skewed, Skewed and Platykurtic_1, and Skewed and Platykurtic_2 distributions, as demonstrated in Figures 26 to 28), with increasing sample sizes with unequal samples such as (50, 100) and (100, 50), changes in statistical power of the KS-2 test was at least 0.90 as the population standard deviation (SD ratios) were either 2 or $\frac{1}{2}$. As SD ratios became very different between the two populations ($3, \frac{1}{3}, 4, \text{ and } \frac{1}{4}$), statistical power was almost equal to one. The graphs of statistical power for these four population distributions with SD ratios of ($4, 3, 2, \frac{1}{2}, \frac{1}{3}, \text{ and } \frac{1}{4}$) are presented in Figures 29 to 34. The simulated results of statistical power for these three population distributions at conditions of eight different sample size combinations with six SD ratios are displayed in APPENDIX V: Tables 19 to 21.

Figure26: Histogram of the **Skewed distribution** (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)

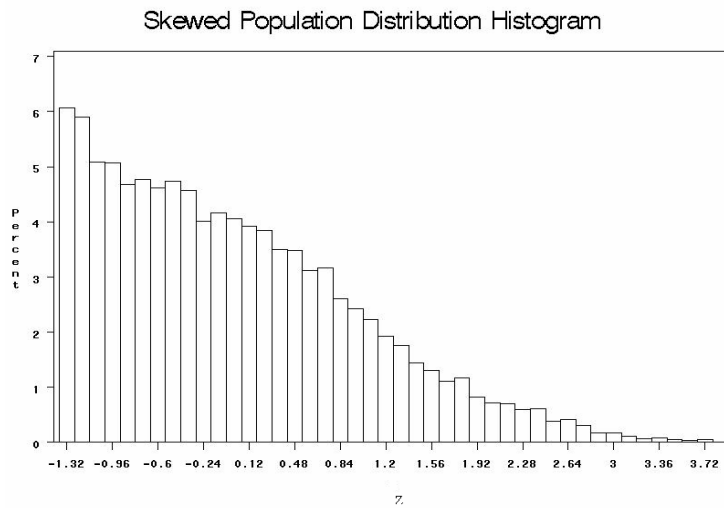


Figure27: Histogram of the **Skewed and Platykurtic_1 distribution** (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)

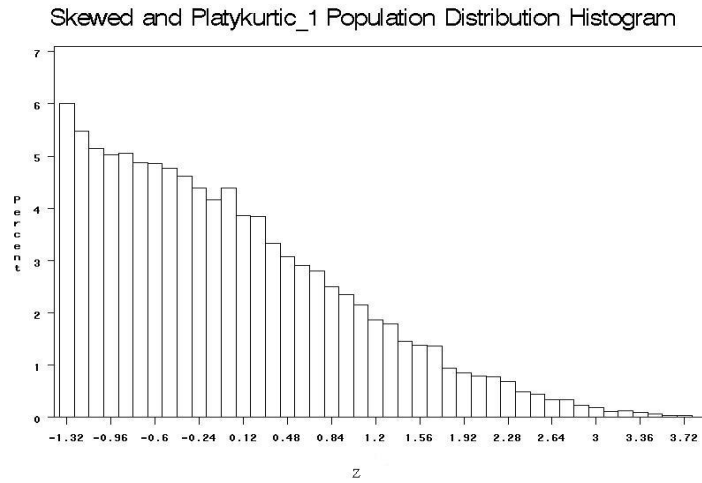


Figure28: Histogram of the **Skewed and Platykurtic_2 distribution** (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)

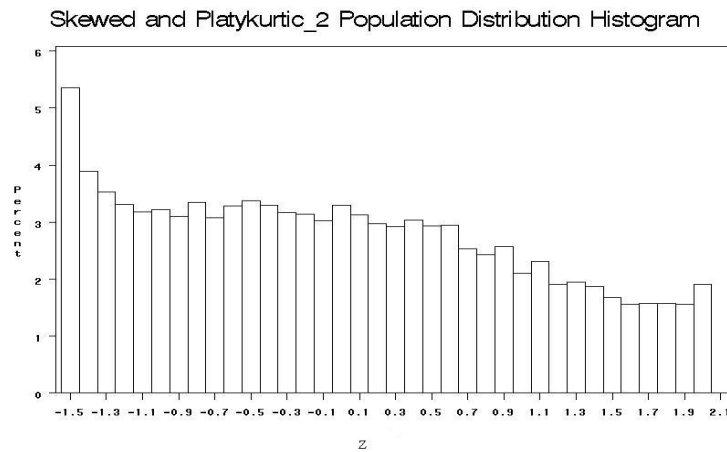


Figure 29: Power of **Skewed Distribution** when Sample Size Differs and SD ratios = 4, 3, & 2

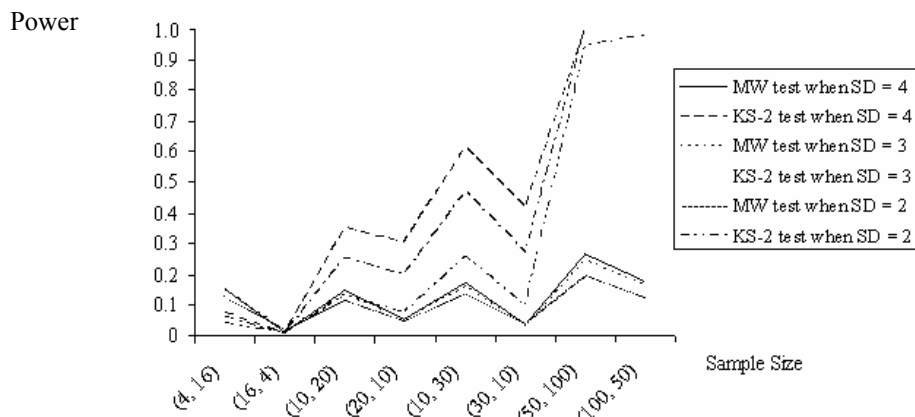


Figure 30: Power of **Skewed Distribution** when Sample Size Differs and SD ratios = 1/2, 1/3, & 1/4

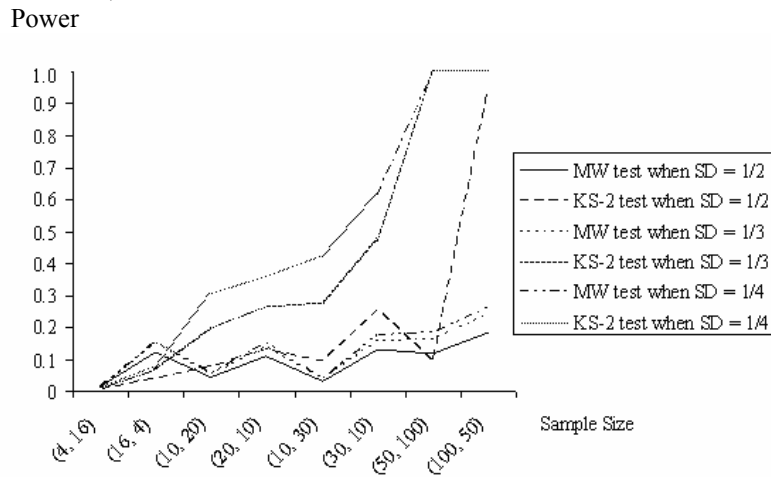


Figure 31: Power of **Skewed and Platykurtic 1 Distribution** when Sample Size Differs and SD ratios = 4, 3, & 2

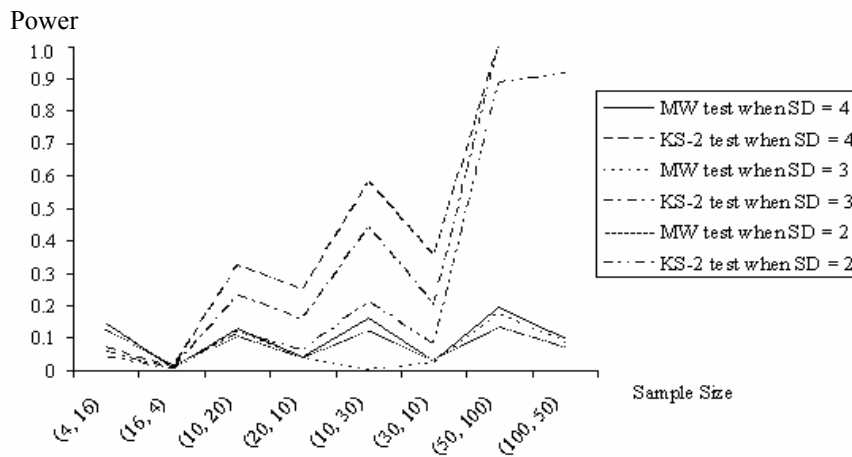


Figure 32: Power of **Skewed and Platykurtic 1 Distribution** when Sample Size Differs and SD ratios = 1/2, 1/3, & 1/4

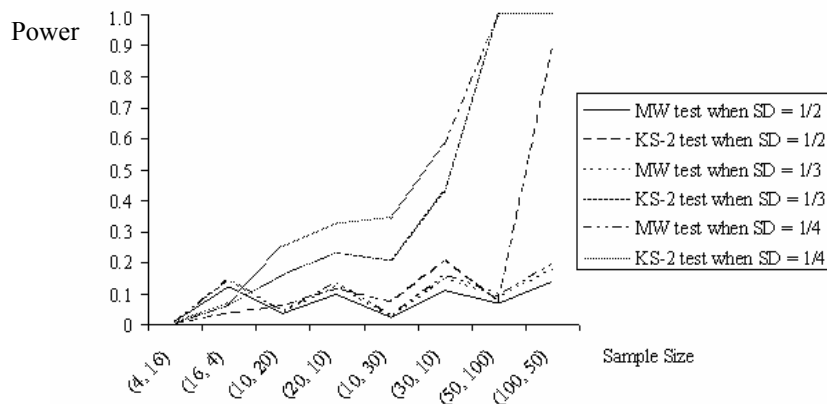


Figure 33: Power of **Skewed and Platykurtic 2** Distribution when Sample Size Differs and SD ratios = 4, 3, & 2

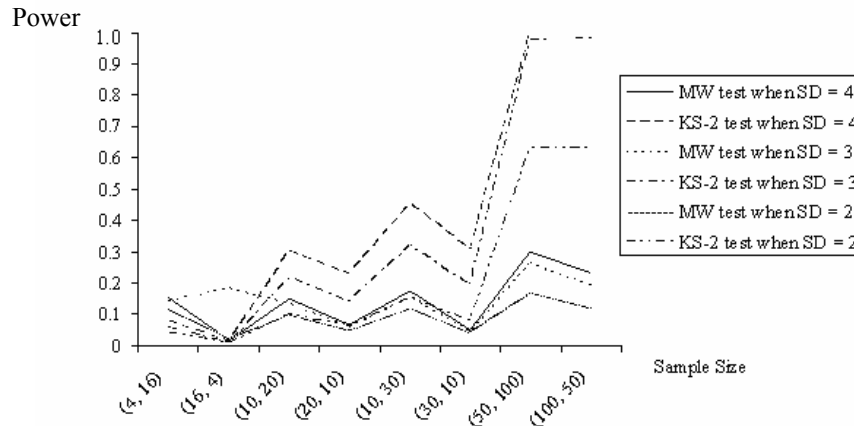
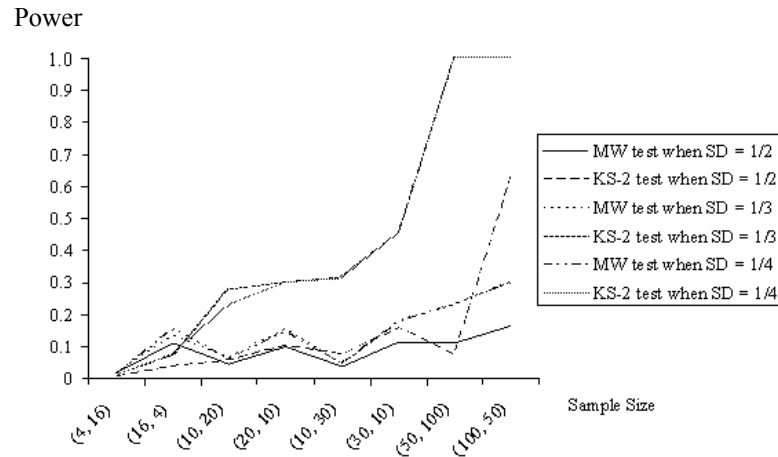


Figure 34: Power of the **Skewed and Platykurtic 2** Distribution when Sample Size Differs and SD ratios = 1/2, 1/3, & 1/4



Research Question 2: If only the heterogeneity of variance between two populations exists, is there any difference in power for these two nonparametric techniques?

Tables 22 to 36 in APPENDIX V display the simulated results of statistical power regarding the MW test and the KS-2 test with six different SD ratios and four pairs of equal sample sizes drawn from fifteen population distributions. When simulating statistical power for the MW test across four different pairs of equal sample sizes under the condition of heterogeneity of variance, it was found that there were no significant differences in statistical power among four pairs of equal samples sizes, (8, 8), (16, 16), (25, 25) and (50, 50) when the SD ratios were 4 and $\frac{1}{4}$, 3 and $\frac{1}{3}$, 2 and $\frac{1}{2}$, in nine out of fifteen populations (Tables 22 to 28, and 35 to 37 in APPENDIX V). The simulated statistical power of the MW test for this research question tended to be small. The range of statistical power for SD ratios of 4 and $\frac{1}{4}$, 3 and $\frac{1}{3}$, and 2 and $\frac{1}{2}$ were between .068 and .083, .059 and .080, and .052 and .061, respectively. This indicated that when there were large differences in the population standard deviations between two underlying population distributions (such as SD ratio = 4 or $\frac{1}{4}$), there was slightly more statistical power than when samples were drawn from the same population distribution but with a small difference in the population standard deviations (such as SD ratio = 2 and $\frac{1}{2}$).

In contrast, the trends of simulated statistical power of the KS-2 test with six different SD ratios and four pairs of equal sample sizes drawn from these nine population distributions were much different than the MW test with two samples under the same conditions. It appeared that as sample sizes were small and equal, and regardless of the differences in SD

ratios, statistical power of the MW test and the KS-2 test were small too. When sample sizes began to increase under the condition of the same SD ratio, statistical power of the KS-2 test increased spectacularly. However, the power of the MW test remained similar with the same condition. When the two samples increased in size and the population SD ratios between the two underlying population distributions were 2 and $\frac{1}{2}$, the range of statistical power of the MW test was small under these two conditions. When there was an increase in sample sizes and population SD ratios, statistical power of the KS-2 test increased extensively. Conversely, the power of the MW test was alike with the same condition. With a sample size of (50, 50) and population standard deviation ratios that were significantly different from each other (SD ratios = 4 and $\frac{1}{4}$), the range of statistical power of the KS-2 test was between .945 and 1.0.

Figures 35 to 43 show the tendencies in statistical power of the MW test and the KS-2 test under the conditions of sample size (50, 50) and SD ratios of $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$, 2, 3, and 4 in nine population distributions.

Figure 35: Power of the **Normal Population** when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$

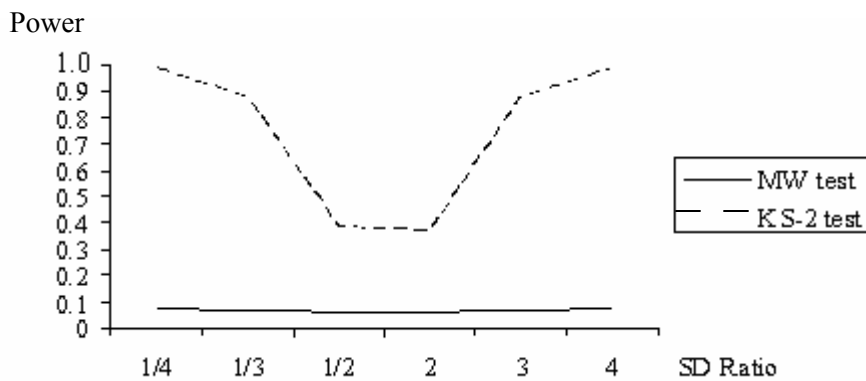


Figure 36: Power of the **Platykurtic Population** when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$

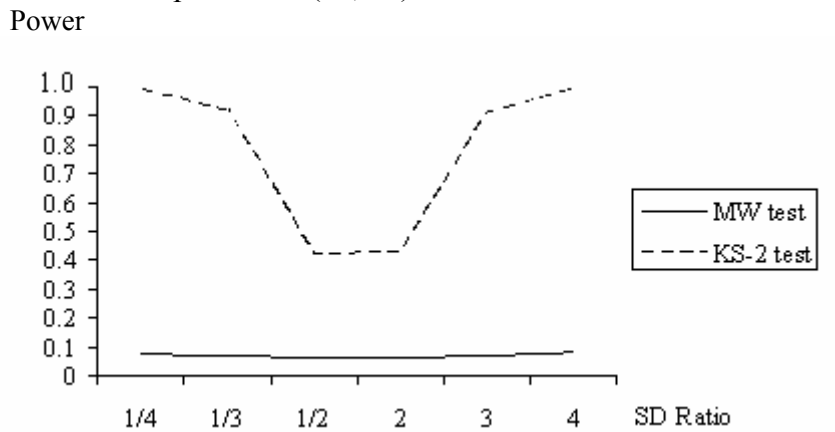


Figure 37: Power of the **Normal Platykurtic Population** when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$

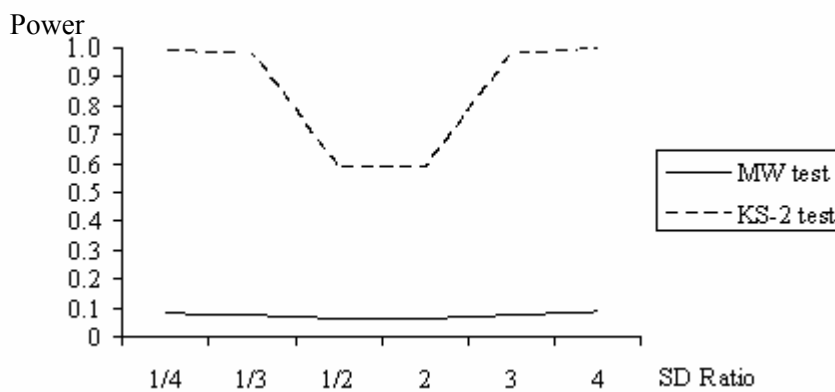


Figure 38: Power of the **Leptokurtic_1 Population** when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$

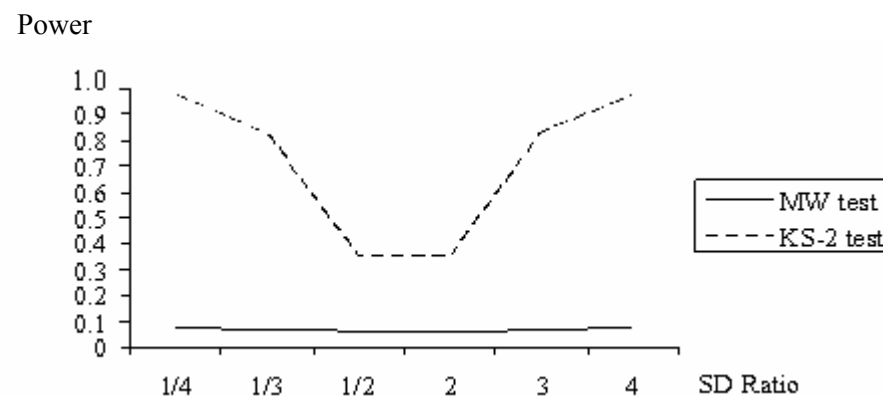


Figure 39: Power of the **Leptokurtic_2 Population** when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$

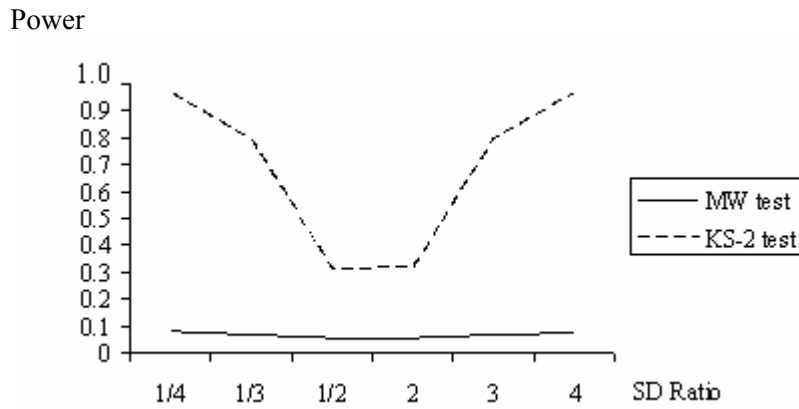


Figure 40: Power of the **Leptokurtic_3 Population** when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$

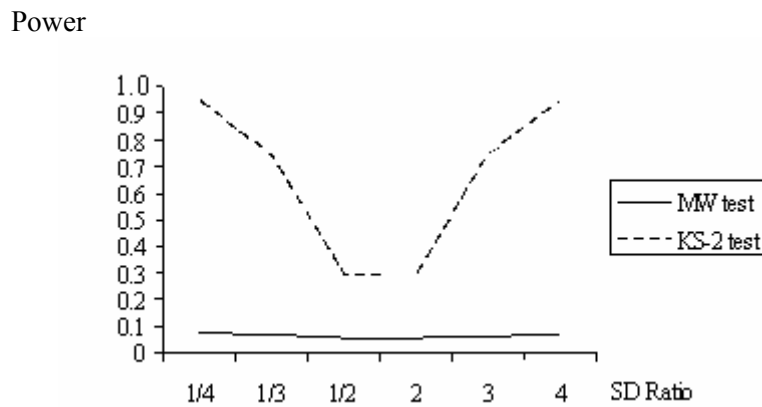


Figure 41: Power of the **Uniform-Like Population** when Only SD Ratios Are Different with sample Size = (50, 50) and $\alpha = .05$

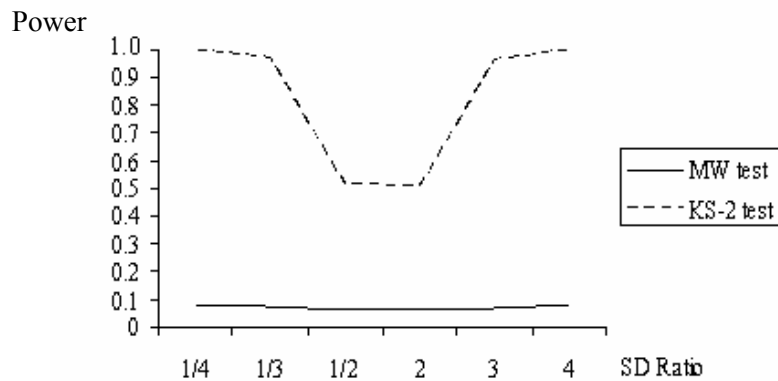


Figure 42: Power of the **Logistic-Like Population** when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$

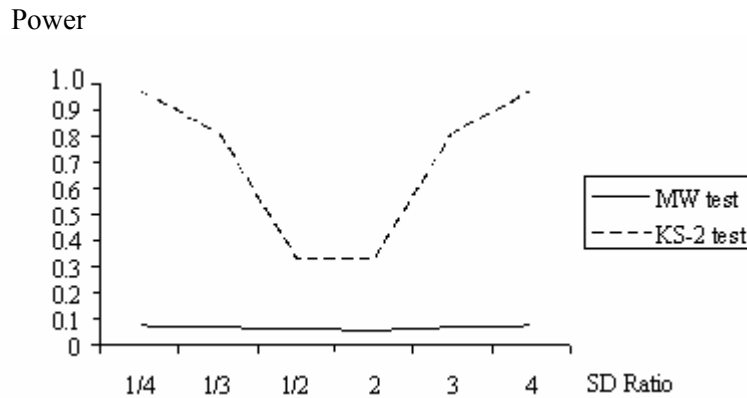
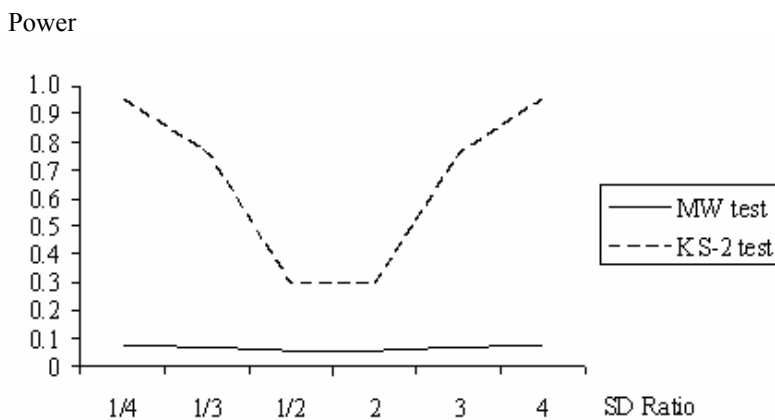


Figure 43: Power of the **Double Exponential-Like Population** when Only SD Ratios Are Different with Sample Size = (50, 50) and $\alpha = .05$



When two equal-sized samples were drawn from following six positively skewed population distributions (Skewed, Skewed and Platykurtic_1, Skewed and Platykurtic_2, Skewed and Leptokurtic_1, Skewed and Leptokurtic_2, and Skewed- Leptokurtic distributions; see Figures 26 to 28, and 44 to 46), statistical power for both the MW test and the KS-2 test increased as the sizes in both samples increased. However, statistical power for the MW test was slightly raised as the sample sizes increased in spite of the differences in the population standard deviations between the two samples. The statistical power of the MW test for conditions of all four pairs of equal sample sizes and six SD ratios were small and

less than 0.250 in these six population distributions except the Skewed- Leptokurtic distribution (Tables 29 to 34 in APPENDIX V). Conversely, statistical power for the KS-2 test significantly increased as the sample sizes changed from (8, 8) to (50, 50). As the difference between the two population standard deviations became more severe, such as SD ratio = 4 and $\frac{1}{4}$, statistical power turned out to be stronger. When the size of two samples was (50, 50), the range of statistical power with SD ratios of 4 and $\frac{1}{4}$ under these population distributions was between 0.967 and 1.0 (Figures 47 to 51).

When the two samples were drawn from a Skewed- Leptokurtic distribution and the sizes were (25, 25) and (50, 50), statistical power of the KS-2 test across six SD ratios was almost equal to 1.0 (Figure 52 and APPENDIX V: Table 34). The statistical power of the KS-2 test was increased considerably when sizes of two samples were increased across six different SD ratios.

Figure 44: Histogram of the **Skewed and Leptokurtic_1** distribution (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)

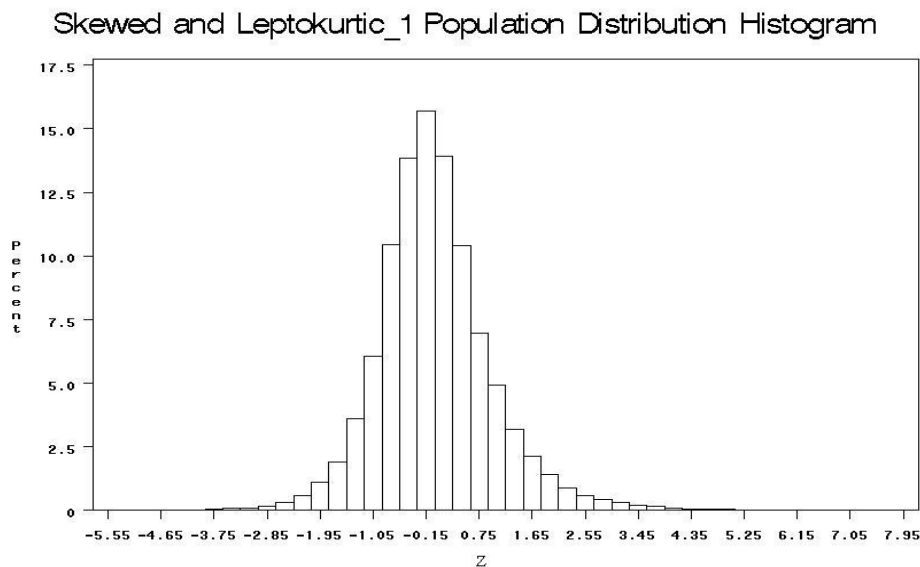


Figure 45: Histogram of the **Skewed and Leptokurtic_2 distribution** (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)

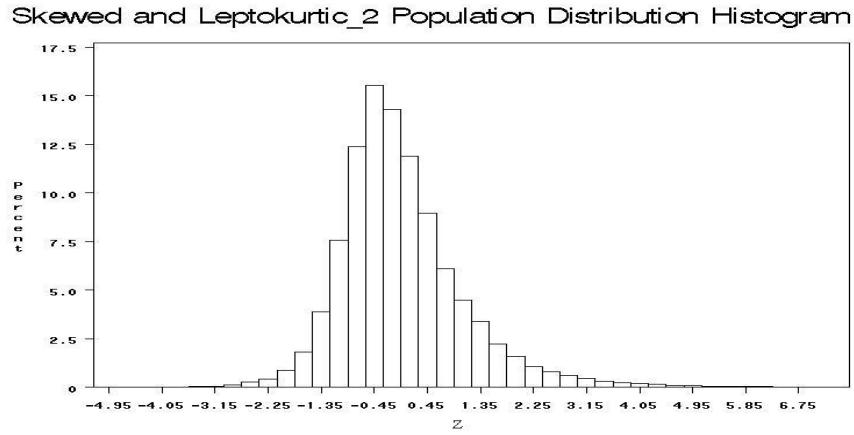


Figure 46: Histogram of the **Skewed- Leptokurtic distribution** (N =20, 000, Y-axis is the relative frequency, X-axis is the Z score)

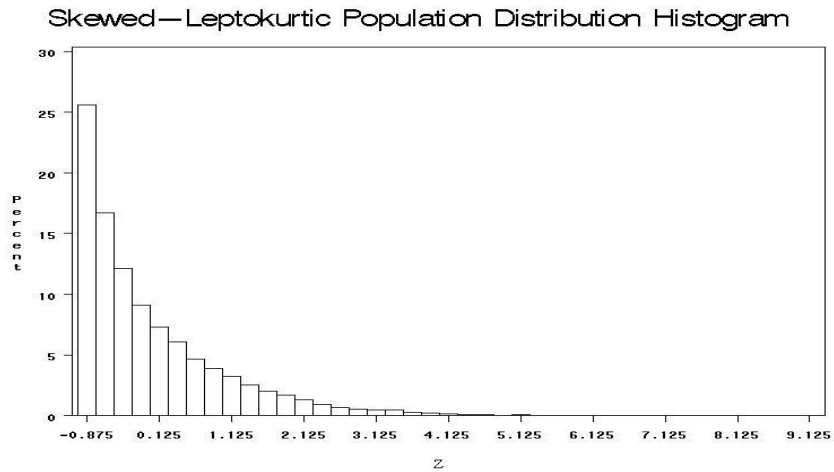


Figure 47: Power of the **Skewed Population** with ONLY SD Ratios Are Different and Sample Size = (50, 50) and $\alpha = .05$

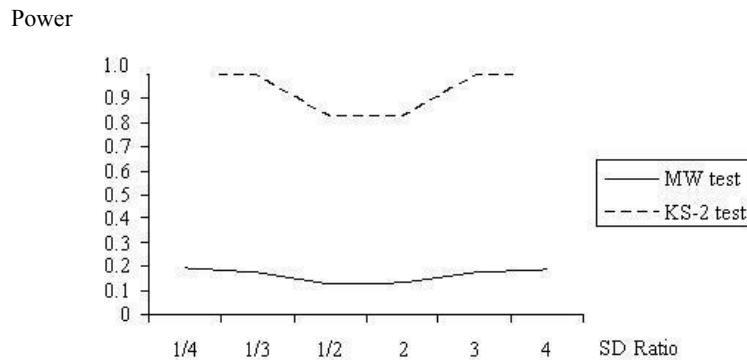


Figure 48: Power of the **Skewed and Platykurtic_1 Population** with ONLY SD Ratios Are Different and Sample Size = (50, 50) and $\alpha = .05$

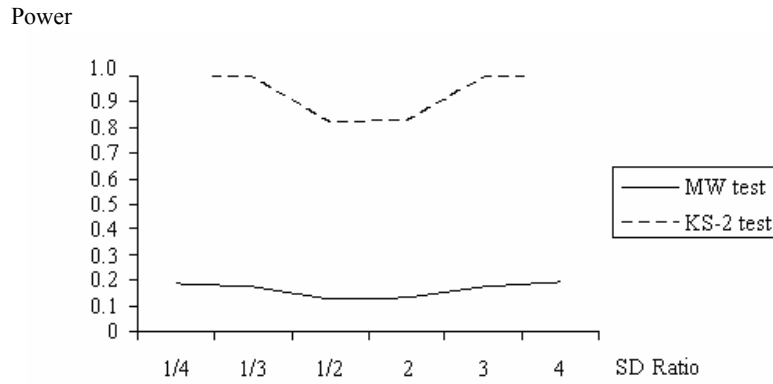


Figure 49: Power of the **Skewed and Platykurtic_2 Population** with ONLY SD Ratios Are Different and Sample Size = (50, 50), (25, 25) and $\alpha = .05$

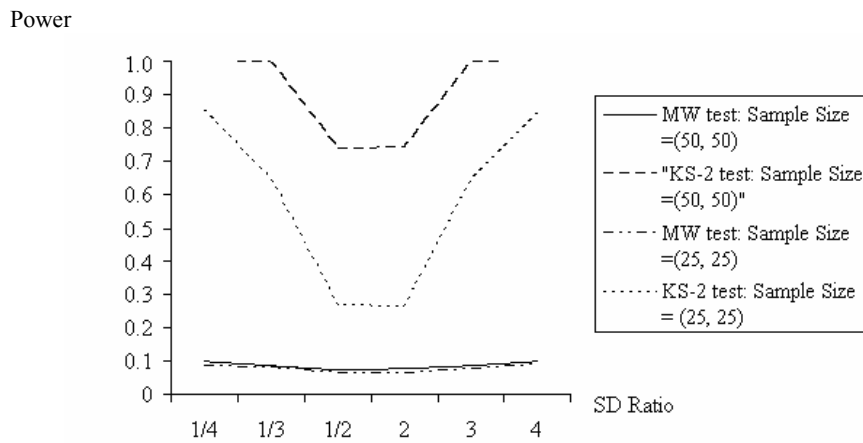


Figure 50: Power of the **Skewed and Leptokurtic_1 Population** with ONLY SD Ratios Are Different and Sample Size = (50, 50) and $\alpha = .05$

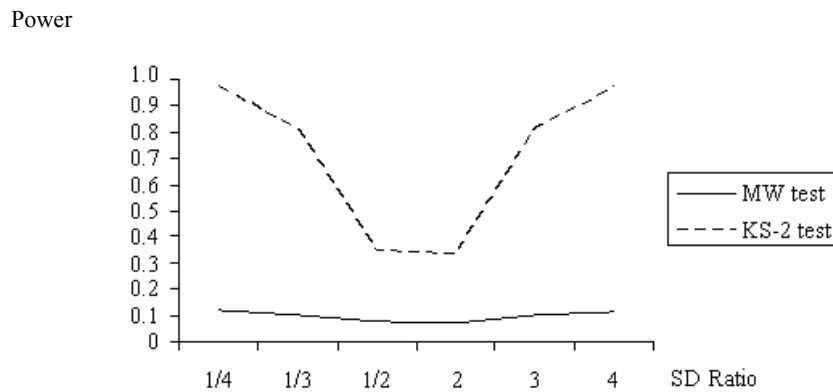


Figure 51: Power of the **Skewed and Leptokurtic_2 Population** with ONLY SD Ratios Are Different and Sample Size = (50, 50) and $\alpha = .05$

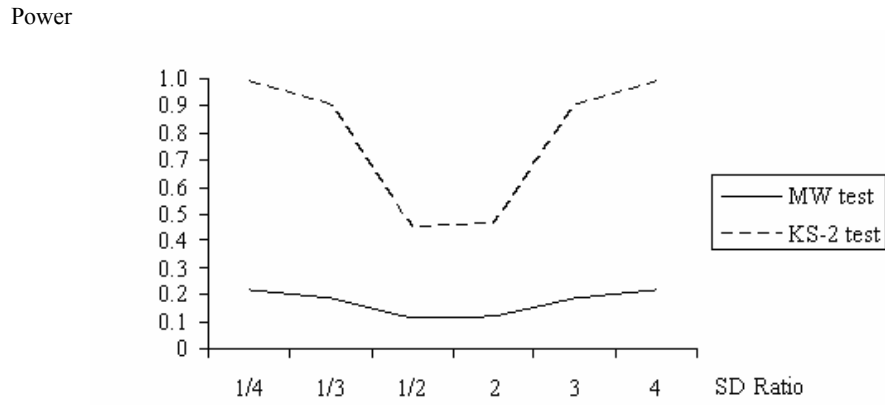
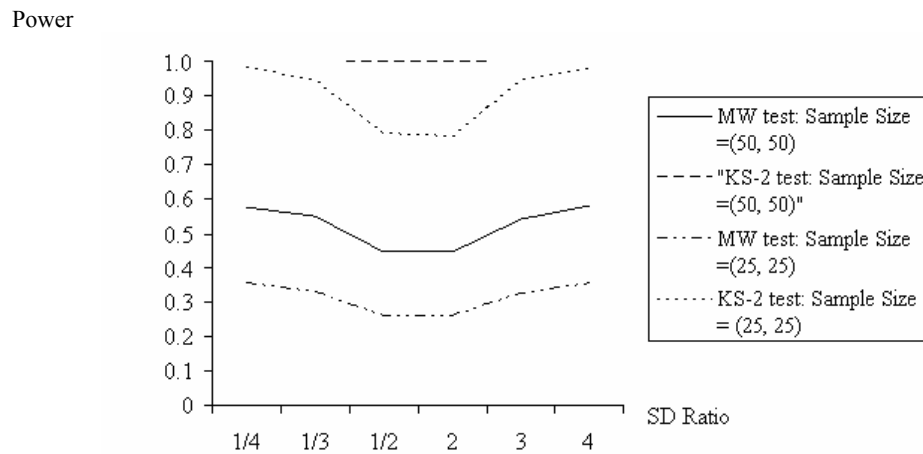


Figure 52: Power of the **Skewed-Leptokurtic Population** with ONLY SD Ratios Are Different and Sample Size = (50, 50), (25, 25) and $\alpha = .05$



Research Question 3: If the nature of the underlying population distributions varies in skewness only, is there any difference in power for these two nonparametric techniques?

This research question allowed for a comparison of statistical power with the same sample sizes and kurtosis coefficients but with different skewness coefficients. When two samples were drawn from the population distributions with different degrees of skewness but equal kurtosis ratios, statistical power for the MW test was small and almost the same as sample sizes were increased. The range of statistical power of the MW test for all 36 combinations was between 0.049 and 0.082.

When the KS-2 test was applied to these 36 combinations under the conditions of this research question, it was found that statistical power was smaller than the power for the MW test when the sample size was (8, 8) in 36 simulations. The statistical power of sample size (8, 8) in the KS-2 test ranged from 0.018 to 0.042 which was smaller than the range in the MW test (from 0.047 to 0.071). In these groups with four different combinations of equal sample sizes, most simulations showed that when the degree of the skewness changed in the large sample set, the statistical power of the KS-2 test was higher than the power of the MW test. However, when the two populations had the same degree of kurtosis -0.50 but differed in the degree of skewness, 0.00 and 0.50, statistical power of the KS-2 test was still smaller than the MW test for both small and large samples. Similar findings were found under the conditions that kurtosis was 3.75 and skewness between the two population distributions were (0.75, 1.25) and (0.00, 0.75) for both small and large sample sizes. However, when the kurtosis ratio was -1.00 and the skewness ratios for the two underlying population distributions were 0.00 and 0.25, the MW test had more power than the KS-2 test regardless

of size of the two samples. Similar results applied to conditions that kurtosis ratio of 3.75 and the skewness ratios were (0.75, 1.75) and (0.00, 0.75) for the two underlying population distributions. The complete statistical power values of the MW test and the KS-2 test for this research question are presented in Table 37.

Table 37: Power; Only Skewness Ratios Are Different ($\gamma_{11} \neq \gamma_{12}$ & $\alpha = .05$)

POPULATION	SAMPLE SIZE	POWER		POPULATION	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
Normal Vs. Skewed	Kurtosis = 0.00 $\gamma_{1N} = 0.00; \gamma_{1S} = 0.75$			Leptokurtic_3 Vs. Skewed and Leptokurtic	Kurtosis = 3.75 $\gamma_{1L3} = 0.00; \gamma_{1SL3} = 1.75$		
	(8, 8)	.051	.021		(8, 8)	.071	.042
	(16, 16)	.053	.048		(16, 16)	.086	.114
	(25, 25)	.057	.055		(25, 25)	.105	.162
	(50, 50)	.065	.082		(50, 50)	.163	.351
. Platykurtic Vs Skewed and Platykurtic_1	Kurtosis = -.50 $\gamma_{1P} = 0.00; \gamma_{1PS1} = 0.50$			Skewed and Leptokurtic_1 vs Skewed and Leptokurtic_2	Kurtosis = 3.75 $\gamma_{1SL1} = 0.75; \gamma_{1SL2} = 1.25$		
	(8, 8)	.054	.023		(8, 8)	.050	.020
	(16, 16)	.051	.041		(16, 16)	.050	.039
	(25, 25)	.053	.045		(25, 25)	.053	.040
	(50, 50)	.056	.058		(50, 50)	.057	.055

Table 37 CONT.: Power; Only Skewness Ratios Are Different ($\gamma_{11} \neq \gamma_{12}$ & $\alpha = .05$)

POPULATION	SAMPLE SIZE	POWER		POPULATION	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
Normal Platykurtic Vs. Skewed and Platykurtic_2	Kurtosis = -1.00 $\gamma_{1NP} = 0.00; \gamma_{1SP2} = 0.25$			Skewed and Leptokurtic_1 Vs. Skewed and Leptokurtic_3	Kurtosis = 3.75 $\gamma_{1SP1} = 0.75; \gamma_{1SP3} = 1.75$		
	(8, 8)	.047	.018		(8, 8)	.047	.018
	(16, 16)	.046	.035		(16, 16)	.046	.035
	(25, 25)	.053	.038		(25, 25)	.053	.038
	(50, 50)	.050	.046		(50, 50)	.050	.046
Leptokurtic_3 Vs. Skewed and Leptokurtic_1	Kurtosis = 3.75 $\gamma_{1L3} = 0.00; \gamma_{1SL1} = 0.75$			Skewed and Leptokurtic_2 Vs. Skewed and Leptokurtic_3	Kurtosis = 3.75 $\gamma_{1SL2} = 1.25; \gamma_{1SL3} = 1.75$		
	(8, 8)	.049	.018		(8, 8)	.049	.018
	(16, 16)	.050	.041		(16, 16)	.050	.041
	(25, 25)	.057	.044		(25, 25)	.057	.044
	(50, 50)	.057	.055		(50, 50)	.057	.055
Leptokurtic_3 Vs. Skewed and Leptokurtic_2	Kurtosis = 3.75 $\gamma_{1L3} = 0.00; \gamma_{1SL2} = 1.25$						
	(8, 8)	.057	.025				
	(16, 16)	.058	.054				
	(25, 25)	.068	.064				
	(50, 50)	.082	.102				

Research Question 4: If the nature of the underlying population distributions varies in kurtosis only, is there any difference in power for these two nonparametric techniques?

This research question allowed for a comparison of statistical power with the same sample sizes and skewness coefficients but different kurtosis coefficients. When two samples were drawn from the population distributions with different degrees of kurtosis but equal skewness ratios, statistical power for the MW test was small and almost the same regardless of increases in sample sizes such as (8, 8) to (50, 50). The range of statistical power of the MW test for all 144 combinations was between 0.043 and 0.065. It was shown that statistical power of the MW test in this research question was very consistent across all levels of sample size when kurtosis ratios changed but skewness ratios remained the same.

When the KS-2 test was applied to the same simulated samples, it was found that statistical power was smaller than power for the MW test with this small sample size (8, 8). The range of statistical power of sample size (8, 8) in the KS-2 test ranged from 0.018 to 0.022. The range of statistical power in the MW test was from 0.047 to 0.054.

When the two samples had the same skewness but the difference in kurtosis was smaller than 2.0 in most of the comparisons, statistical power for the MW test was higher than the power for the KS-2 test across four equal-sized pairs of samples. However, if the difference in kurtosis between the two populations become substantial, statistical power for the KS-2 test was larger than power for the MW test especially in the two samples with large sets of sizes.

Complete statistical power values for the MW test and the KS-2 test for this research question are presented in Table 38.

Table 38: Power; Only Kurtosis Ratios Are Different ($\gamma_{21} \neq \gamma_{22}$ & $\alpha = .05$)

POPULATION	SAMPLE SIZE	POWER		POPULATION	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
Normal Vs. Platykurtic	Skewness = 0.00 $\gamma_{2N} = 0.00; \gamma_{2P} = -0.50$			Normal Vs. Logistic-Like	Skewness = 0.00 $\gamma_{2N} = 0.00; \gamma_{2LL} = 1.30$		
	(8, 8)	.051	.018		(8, 8)	.048	.018
	(16, 16)	.045	.033		(16, 16)	.047	.037
	(25, 25)	.050	.035		(25, 25)	.051	.040
	(50, 50)	.050	.041		(50, 50)	.048	.046
. Normal Vs. Normal Platykurtic	Skewness = 0.00 $\gamma_{2N} = 0.00; \gamma_{2NP} = -1.00$			Normal Vs. Leptokurtic_2	Skewness = 0.00 $\gamma_{2N} = 0.00; \gamma_{2L2} = 2.00$		
	(8, 8)	.048	.019		(8, 8)	.050	.020
	(16, 16)	.050	.039		(16, 16)	.047	.038
	(25, 25)	.048	.040		(25, 25)	.053	.044
	(50, 50)	.052	.052		(50, 50)	.050	.055
Normal Vs. Leptokurtic_1	Skewness = 0.00 $\gamma_{2N} = 0.00; \gamma_{2L1} = 1.00$			Normal Vs. Double Exponential- Like	Skewness = 0.00 $\gamma_{2N} = 0.00; \gamma_{2DEL} = 3.00$		
	(8, 8)	.050	.019		(8, 8)	.052	.021
	(16, 16)	.048	.037		(16, 16)	.050	.041
	(25, 25)	.050	.037		(25, 25)	.053	.044
	(50, 50)	.048	.041		(50, 50)	.050	.055

Table 38 CONT.: Power; Only Kurtosis Ratios Are Different ($\gamma_{21} \neq \gamma_{22}$ & $\alpha = .05$)

POPULATION	SAMPLE SIZE	POWER		POPULATION	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
. Normal Vs. Leptokurtic_3	Skewness = 0.00 $\gamma_{2N} = 0.00; \gamma_{2L3} = 3.75$			Platykurtic Vs. Leptokurtic_1	Skewness = 0.00 $\gamma_{2P} = -0.50; \gamma_{2L1} = 1.00$		
	(8, 8)	.049	.019		(8, 8)	.051	.019
	(16, 16)	.047	.041		(16, 16)	.046	.037
	(25, 25)	.052	.047		(25, 25)	.050	.041
	(50, 50)	.052	.067		(50, 50)	.048	.045
. Platykurtic Vs. Normal Platykurtic	Skewness = 0.00 $\gamma_{2P} = -0.50; \gamma_{2NP} = -1.00$. Platykurtic Vs. Leptokurtic_2	Skewness = 0.00 $\gamma_{2P} = -0.50; \gamma_{2L2} = 2.00$		
	(8, 8)	.050	.018		(8, 8)	.049	.019
	(16, 16)	.045	.036		(16, 16)	.048	.041
	(25, 25)	.050	.038		(25, 25)	.048	.043
	(50, 50)	.052	.046		(50, 50)	.048	.055
Normal Vs. Uniform-Like	Skewness = 0.00 $\gamma_{2N} = 0.00; \gamma_{2UL} = -1.20$			Platykurtic Vs. Leptokurtic_3	Skewness = 0.00 $\gamma_{2P} = -0.50; \gamma_{2L3} = 3.75$		
	(8, 8)	.050	.018		(8, 8)	.051	.022
	(16, 16)	.050	.040		(16, 16)	.048	.045
	(25, 25)	.051	.043		(25, 25)	.051	.052
	(50, 50)	.053	.053		(50, 50)	.050	.079

Table 38 CONT.: Power; Only Kurtosis Ratios Are Different ($\gamma_{21} \neq \gamma_{22}$ & $\alpha = .05$)

POPULATION	SAMPLE	POWER		POPULATION	SAMPLE	POWER	
	SIZE	MW	KS-2		SIZE	MW	KS-2
Platykurtic Vs. Uniform-Like	Skewness = 0.00 $\gamma_{2P} = -0.50; \gamma_{2UL} = -1.20$			Normal Platykurtic Vs. Uniform-Like	Skewness = 0.00 $\gamma_{2NP} = -1.00; \gamma_{2UL} = -1.20$		
	(8, 8)	.049	.018		(8, 8)	.050	.019
	(16, 16)	.047	.036		(16, 16)	.045	.033
	(25, 25)	.049	.038		(25, 25)	.049	.036
	(50, 50)	.049	.044		(50, 50)	.049	.040
Platykurtic Vs. Logistic-Like	Skewness = 0.00 $\gamma_{2P} = -0.50; \gamma_{2LL} = 1.30$			Normal Platykurtic Vs. Logistic-Like	Skewness = 0.00 $\gamma_{2NP} = -1.00; \gamma_{2LL} = 1.30$		
	(8, 8)	.053	.021		(8, 8)	.048	.018
	(16, 16)	.048	.038		(16, 16)	.048	.042
	(25, 25)	.047	.038		(25, 25)	.048	.050
	(50, 50)	.049	.047		(50, 50)	.049	.070
Platykurtic I Vs. Double Exponential- Like	Skewness = 0.00 $\gamma_{2NP} = -0.50; \gamma_{2DEL} = 3.00$			Normal Platykurtic Vs. Double Exponential- Like	Skewness = 0.00 $\gamma_{2NP} = -1.00; \gamma_{2L1} = 3.00$		
	(8, 8)	.052	.021		(8, 8)	.052	.021
	(16, 16)	.050	.044		(16, 16)	.050	.041
	(25, 25)	.049	.048		(25, 25)	.053	.044
	(50, 50)	.051	.068		(50, 50)	.050	.055

Table 38 CONT.: Power; Only Kurtosis Ratios Are Different ($\gamma_{21} \neq \gamma_{22}$ & $\alpha = .05$)

POPULATION	SAMPLE SIZE	POWER		POPULATION	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
Normal Platykurtic Vs. Leptokurtic_1	Skewness = 0.00 $\gamma_{2NP} = -1.00; \gamma_{2L1} = 1.00$			Leptokurtic_1 Vs. Leptokurtic_2	Skewness = 0.00 $\gamma_{2L1} = 1.00; \gamma_{2L2} = 2.00$		
	(8, 8)	.051	.021		(8, 8)	.047	.018
	(16, 16)	.045	.043		(16, 16)	.043	.034
	(25, 25)	.053	.049		(25, 25)	.050	.037
	(50, 50)	.052	.070		(50, 50)	.049	.039
Normal Platykurtic Vs. Leptokurtic_2	Skewness = 0.00 $\gamma_{2NP} = -1.00; \gamma_{2L2} = 2.00$			Leptokurtic_1 Vs. Leptokurtic_3	Skewness = 0.00 $\gamma_{2L1} = 1.00; \gamma_{2L3} = 3.75$		
	(8, 8)	.050	.020		(8, 8)	.051	.019
	(16, 16)	.052	.050		(16, 16)	.048	.038
	(25, 25)	.053	.055		(25, 25)	.049	.039
	(50, 50)	.052	.084		(50, 50)	.050	.049
Normal Platykurtic Vs. Leptokurtic_3	Skewness = 0.00 $\gamma_{2NP} = -1.00; \gamma_{2L3} = 3.75$			Leptokurtic_1 Vs. Uniform-Like	Skewness = 0.00 $\gamma_{2L1} = 1.00; \gamma_{2UL} = -1.20$		
	(8, 8)	.051	.023		(8, 8)	.051	.019
	(16, 16)	.046	.051		(16, 16)	.048	.045
	(25, 25)	.050	.066		(25, 25)	.050	.047
	(50, 50)	.052	.117		(50, 50)	.051	.070

Table 38 CONT.: Power; Only Kurtosis Ratios Are Different ($\gamma_{21} \neq \gamma_{22}$ & $\alpha = .05$)

POPULATION	SAMPLE	POWER		POPULATION	SAMPLE	POWER	
	SIZE	MW	KS-2		SIZE	MW	KS-2
Leptokurtic_1 Vs. Logistic-Like	Skewness = 0.00 $\gamma_{2L1} = 1.00; \gamma_{2LL} = 1.30$			Leptokurtic_3 Vs. Uniform-Like	Skewness = 0.00 $\gamma_{2L3} = 3.75; \gamma_{2UL} = -1.20$		
	(8, 8)	.049	.019		(8, 8)	.051	.021
	(16, 16)	.048	.036		(16, 16)	.050	.055
	(25, 25)	.052	.038		(25, 25)	.054	.070
	(50, 50)	.048	.040		(50, 50)	.051	.114
Leptokurtic_1 Vs. Double Exponential- Like	Skewness = 0.00 $\gamma_{2L1} = 1.00; \gamma_{2DEL} = 3.00$			Leptokurtic_3 Vs. Logistic-Like	Skewness = 0.00 $\gamma_{2L3} = 3.75; \gamma_{2LL} = 1.30$		
	(8, 8)	.050	.018		(8, 8)	.048	.020
	(16, 16)	.047	.036		(16, 16)	.048	.036
	(25, 25)	.049	.037		(25, 25)	.047	.036
	(50, 50)	.054	.046		(50, 50)	.047	.048
Leptokurtic_2 Vs. Leptokurtic_3	Skewness = 0.00 $\gamma_{2L2} = 2.00; \gamma_{2L3} = 3.75$			Leptokurtic_3 Vs. Double Exponential- Like	Skewness = 0.00 $\gamma_{2L3} = 3.75; \gamma_{2DEL} = 3.00$		
	(8, 8)	.047	.020		(8, 8)	.050	.019
	(16, 16)	.046	.034		(16, 16)	.046	.034
	(25, 25)	.051	.038		(25, 25)	.048	.038
	(50, 50)	.050	.042		(50, 50)	.049	.041

Table 38 CONT.: Power; Only Kurtosis Ratios Are Different ($\gamma_{21} \neq \gamma_{22}$ & $\alpha = .05$)

POPULATION	SAMPLE SIZE	POWER		POPULATION	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
Leptokurtic_2 Vs. Uniform-Like	Skewness = 0.00 $\gamma_{2L2} = 2.00; \gamma_{2L3} = -1.20$			Uniform-Like Vs. Logistic-Like	Skewness = 0.00 $\gamma_{2UL} = -1.20; \gamma_{2LL} = 1.30$		
	(8, 8)	.051	.022		(8, 8)	.051	.021
	(16, 16)	.050	.048		(16, 16)	.046	.044
	(25, 25)	.051	.057		(25, 25)	.051	.050
	(50, 50)	.051	.088		(50, 50)	.048	.074
Leptokurtic_2 Vs. Logistic-Like	Skewness = 0.00 $\gamma_{2L2} = 2.00; \gamma_{2L3} = 1.30$			Uniform-Like Vs. Double Exponential- Like	Skewness = 0.00 $\gamma_{2UL} = -1.20; \gamma_{2DEL} = 3.00$		
	(8, 8)	.051	.019		(8, 8)	.055	.024
	(16, 16)	.047	.034		(16, 16)	.049	.050
	(25, 25)	.051	.037		(25, 25)	.051	.063
	(50, 50)	.051	.040		(50, 50)	.052	.102
Leptokurtic_2 Vs. Double Exponential- Like	Skewness = 0.00 $\gamma_{2L2} = 2.00; \gamma_{2L3} = 3.75$			Logistic-Like Vs. Double Exponential- Like	Skewness = 0.00 $\gamma_{2LL} = 1.30; \gamma_{2DEL} = 3.00$		
	(8, 8)	.051	.019		(8, 8)	.049	.019
	(16, 16)	.046	.036		(16, 16)	.048	.036
	(25, 25)	.050	.034		(25, 25)	.049	.037
	(50, 50)	.051	.042		(50, 50)	.050	.043

Table 38 CONT.: Power; Only Kurtosis Ratios Are Different ($\gamma_{21} \neq \gamma_{22}$ & $\alpha = .05$)

POPULATION	SAMPLE SIZE	POWER	
		MW	KS-2
Skewed Vs. Skewed leptokurtic_1	Skewness = 0.75		
	$\gamma_{2S} = 0.00; \gamma_{2SL1} = 3.75$		
	(8, 8)	.054	.024
	(16, 16)	.052	.050
	(25, 25)	.058	.063
	(50, 50)	.065	.106

Summary

This chapter presented the results and findings of the simulations for the study. There were four research questions addressed. Among these four research questions, results of both Type I error rates and statistical power were discussed in the first research question. Results of statistical power were expressed for research questions two through four. The significance level (α) of 0.05 was applied when performing the MW test and the KS-2 test to the simulated data sets.

In the findings of Type I error rates between these two nonparametric statistical techniques under the conditions of the first research question, most simulated Type I error rates for both the MW test and the KS-2 test were less than 0.05. The KS-2 test had typically lower Type I error rates than the MW test. In other words, Type I error rates for the KS-2 test tended to be less than the rates for the MW test.

The study of statistical power for the second part of the first research question indicated that both the KS-2 test and the MW test had small statistical power when the sample sizes were small and unequal in spite of differences in the SD ratios of the populations. When the sample sizes were large and unequal such as (50, 100) and (100, 50) and the differences in population standard deviations were considerably large (such as SD ratio = 4 or $\frac{1}{4}$), statistical power of the KS-2 test for all 15 populations was close to 1.0. Moreover, when the shapes of the underlying populations were positively skewed with sample sizes (50, 100) and (100, 50), statistical power of the KS-2 test was much more sensitive than the MW test when population standard deviation ratios were not equal to 1.0 (such as SD ratio = 4, 3, 2, $\frac{1}{2}$, $\frac{1}{3}$, or $\frac{1}{4}$). The findings of the second research question yielded similar results in statistical power

when performing the KS-2 test for four pairs of independent samples with equal sizes (8, 8), (16, 16), (25, 25), and (50, 50). The statistical power of the MW test under the conditions of the second research question was found to be consistently small in fifteen population distributions across different levels of SD ratios with the defined four pairs of equal sample sizes.

Even though the considerations for research questions three and four were not the same, the results of statistical power for these two questions were very similar. Both statistical power values for the MW test and the KS-2 tests were small. The results of statistical power of the MW test for both research questions showed that statistical power was small and almost the same, being consistent across all four pairs of two equal-sized independent samples despite the changes in either the kurtosis or skewness ratios. The KS-2 test produced slightly different results when compared with the MW test. Statistical power was relatively small when the sample sizes were small and equal, such as (8, 8). As the sample sizes increased, statistical power also increased. When the two underlying population distributions had the same kurtosis but greatly differed in skewness, statistical power for the KS-2 test was higher than the power of the MW test while the two equal-sized samples increased in size. Similar results were found when the two samples had fixed skewness but different kurtosis in their population distributions. When the two equal-sized samples had the same skewness but different kurtosis in their underlying population distributions, statistical power for the KS-2 test was higher than the power of the MW test as the two equal sized samples increased in size.

CHAPTER FIVE

DISCUSSIONS

Introduction

The Mann-Whitney (MW) test and the Kolmogrov-Smirnov two sample test (KS-2) are nonparametric statistical tests used to detect whether there is a general difference between two samples when the two underlying population distributions are distribution-free. The focus of this study was to examine and compare Type I error rates and statistical power between the Mann-Whitney (MW) and the Kolmogrov-Smirnov two sample (KS-2) tests when the two samples had different population variances or various degrees of kurtosis and skewness. This study also compared Type I error rates and power, if applicable, when the two samples were of different sizes.

This chapter provides the general conclusions of the study. In addition, theoretical implications, practical implications, limitations, and recommendations for future research are presented. Conclusions are proposed based upon the findings of Type I error rates and statistical power between the two tests. Next, theoretical implications are provided through a comparison of the literature in accordance with the research questions. Practical implications provide suggestions for practice, such as the method of simulating statistical power in this study, and criteria for selecting between the MW and the KS-2 tests. Limitations of this study are provided. Finally, recommendations for future research are presented in this chapter.

General Conclusions

This study examined Type I error rates and statistical power differences between the Mann-Whitney (MW) and the Kolmogrov-Smirnov two-sample (KS-2) tests. Simulations were conducted to examine power comparisons between the KS-2 and the MW tests by performing 20, 000 replications per condition. Variations in sample, the underlying population distributions varying in variance, and skewness and kurtosis were utilized. Simulations were performed to investigate four research questions. The first research question was directly applicable to assess Type I error rates. The results highlight differences between the KS-2 test and the MW test, important to those performing these nonparametric statistical hypothesis tests for general differences between populations.

When examining Type I error rates for the MW test and the KS-2 test with different sample sizes but same SD ratio and population distributions between two samples, the study showed that Type I error rates for the KS-2 test were much less than the rates for the MW test. Type I errors for the MW test were close to the nominal value ($\alpha=0.05$). However, Type I error rates for the KS-2 test were much less than the nominal value. This implied that when researchers perform significance tests in detecting general differences between two samples with the same underlying population distributions, the KS-2 test is more likely to result in rejecting the null hypothesis. This is not true, however, when mean differences are detected. Researchers will have a greater chance of finding a difference between two samples with the same underlying population distributions when applying the KS-2 test rather than the MW test.

When sample sizes are unequal and small with the same underlying population distributions, the MW test has more statistical power than the KS-2 test regardless of the

population SD ratios. The KS-2 test has more statistical power than the MW test under the condition that population variances and the sample sizes are greatly different from one another, regardless of the underlying population distributions. It was also discovered that when sample sizes were small and unequal, both the MW test and the KS-2 test had small statistical power. The MW test was slightly more powerful than the KS-2 test in spite of the SD ratios. It is suggested that when the size of the two samples is small and unequal, the MW test is more powerful than the KS-2 test. As the sample size increases and remains unequal, the KS-2 test is more powerful than the MW test.

When the two underlying population distributions differed in skewness, statistical power for both the MW and the KS-2 test was small. When sample size was small, regardless of the differences in skewness between the two population distributions, the KS-2 had smaller statistical power than the MW test. When the skewness between the two populations became different and sizes for both samples were large, the KS-2 test had more power than the MW test.

When only the degree of kurtosis was different between the two population distributions, the MW and the KS-2 test had small statistical power. When sample size was small and the degree of kurtosis between the two population distributions was ignored, the KS-2 test had smaller statistical power than the MW test. The KS-2 test had slightly more statistical power as the degree of the kurtosis become very different between the two populations in comparisons of two samples with large sizes.

In conclusion, the KS-2 test is smaller than the MW test in comparison of the type I error rates in unequal sample sets. Moreover, when population variances vary between two samples, the KS-2 test has more statistical power than the MW test. Furthermore, the power

of the KS-2 test exceeded the power of the MW test in large sample settings when either one of the following conditions existed:

1. The difference in the Skewness in populations between the two samples was more than 0.5 with the same kurtosis and variance.
2. The difference in the Kurtosis in populations between the two samples was more than 2.0 with the same skewness and variance.

Theoretical Implications

This study investigated Type I error rates and statistical power differences between the Mann-Whitney (MW) and the Kolmogrov-Smirnov two sample (KS-2) tests under various conditions. The simulated findings can be meshed with the literature, as guided by the research questions listed below:

Question 1: If only **sample sizes** differ between two samples,

- a. Is there any difference in Type I error rate for these two nonparametric techniques?
- b. Is there any difference in power for these two nonparametric techniques?

Question 2: If only the **heterogeneity of variance** between two populations exists, is there any difference in power for these two nonparametric techniques?

Question 3: If the nature of the underlying population distributions varies in **skewness** only, is there any difference in power for these two nonparametric techniques?

Question 4: If the nature of the underlying population distributions varies in **kurtosis** only, is there any difference in power for these two nonparametric techniques?

Sample Size

The first research question detected Type I error rates and statistical power in the MW and the KS-2 tests when two samples varied in sample size. Eight different pairs of equal sample sizes were simulated. When only sample sizes were different between the two samples, with same SD ratios and same degrees of skewness and kurtosis in the two underlying population distributions, it was found that Type I error rates for both the MW and the KS-2 tests were all small and mostly less than the nominal significance level of (α) 0.05. When detecting Type I error rates for the MW test, results in the present simulation study were similar to findings reported in the literature. For example, when the two samples had sizes of (4, 16), the Type I error rate for the MW test (the Type I error rate = 0.048) in the normal distribution was less than the nominal rate (nominal $\alpha = 0.05$). This finding was similar to the values reported by Zimmerman (1987) and Gibbons and Chakraborti's (1991). Zimmerman (1987) reported a Type I error rate of 0.048 with an α of 0.05. Gibbons and Chakraborti's study (1991) found a error rate of 0.048 with the same α of 0.05. Moreover, when detecting the Type I error rate for the MW test with sample size (16, 4), the Type I error rate of 0.051 was greater than the nominal significance level of 0.05. However, Zimmerman (1987) had a Type I error rate of 0.049 for the same conditions, which was less than the significance level of 0.05. When sample size increased to (30, 10), Type I error rates of the MW test became 0.051 which was inflated and greater than the nominal significance level of 0.05. Conversely, Kasuya (2001) reported a Type I error rate of 0.0495 which was slightly less than the α of 0.05.

Based on a review of literature, there is a lack of research on the KS-2 test with a non-directional hypothesis test when investigating the general difference between two samples. In

order to fill this gap, the current study simulated Type I error rates for the KS-2 test in eight pairs of unequal sample sizes and fifteen population distributions. The present study discovered that the Type I error rates for all eight pairs of two samples in fifteen population distributions were less than the nominal level α of 0.05. Type I error rates for the KS-2 test were less than and close to the nominal significance level when the two samples with the same underlying population distributions differed by size. Also of interest was that if the sample sizes were small and unequal (both sizes were no more than 20), Type I error rates were extremely small and at most 0.03. As the two unequal samples increased their sizes, such as the (50, 100) and (100, 50), Type I error rates were at most 0.045.

In general, it appears that the Type I error rates for the KS-2 test are much more lower than the rates for the MW test. This occurs when the two sample sizes are small and unequal, and homogeneity of variance exists in normal and fourteen non-normal population distributions. For two samples with large unequal sizes, Type I error rates for both tests approached the nominal significance level of 0.05.

This study also investigated statistical power for both the KS-2 test and the MW test with the same eight pairs of unequal sample size combinations in fifteen population distributions. These distributions differed in SD ratios. It was discovered that both statistical power for the MW test and the KS-2 test was very small. The MW test was more powerful than the KS-2 test under the condition of small and unequal sample sizes with a normal population distribution. As the sizes increased, the statistical power of the KS-2 test became superior to the MW test. When the sample sizes were large and relatively unequal, and the SD ratios were extremely small or extremely large, the power of the KS-2 test was close to one. Similar results were found in the 14 non-normal distributions discussed in the CHAPTER IV.

The present Monte Carlo study discovered that when one of the two samples is tremendously different from the other in sample size, the KS-2 test is more powerful than the MW test under the condition that population variances are greatly different, regardless of the underlying population distributions. This finding did not support the figures of power functions provided in Schroer and Trenkler's (1995) study. Those figures showed that the MW test had better power than the KS-2 test, when the condition of small and equal sample sizes existed. When comparing statistical power between the MW test and the KS-2 test, most research literature discussed this issue under the condition of different population variances but equal sizes between the two samples. Some research literature discussed this issue for a directional hypothesis test. There was limited literature comparing statistical power between the MW test and the KS-2 test for non-directional hypothesis tests, when the two samples were different in size and population variance with the same underlying population distribution.

In general, the present simulation study showed that the KS-2 test had smaller Type I error rates than the MW test when two samples differ in size with homogeneity of population variance. The KS-2 test had less power than the MW test when sample sizes were small and unequal. The value of statistical power for the KS-2 test was greater than the value for the MW test as sample size become large and unequal to one another. This was true with a violation of homogeneity of population variance.

Heterogeneity of Variance

The second research question examined statistical power for the MW and the KS-2 tests when the two samples were only different in population variances. This study simulated statistical power in the conditions of equal sample size and the same underlying population

distributions (15 population distributions) between two samples that differed in SD ratios.

The four pairs of equal sample sizes were (8, 8), (16, 16), (25, 25) and (50, 50). The

statistical power for the MW and the KS-2 tests were produced when the SD ratios were $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$, 2, 3, and 4.

This present study found that the MW test had very little but consistent statistical power across the four simulated pairs with equal sample sizes. This was true when population variances of the two samples were not the same in normal and non-normal population distributions. Even though population variances were greatly different between two samples, statistical power for the MW test changed only slightly. By reviewing literature, it was found that the results of statistical power for the MW test were similar with the study by Gibbons and Chakraborti (1991) in generating statistical power for the MW test in the normal distribution. For example, the statistical power for the MW test was 0.070 with sample size of (8, 8) and a SD ratio of 4 in the present study. The power was changed to 0.056 when the SD ratio changed to 2. Similarly, Gibbons and Chakraborti (1991) had a statistical power for the MW test of 0.0691 with sample size of (10, 10) and a SD ratio of 5. The power became 0.0559 with the same size with a SD ratio of 2.5.

The current study found that the KS-2 test was much more powerful than the MW test in fifteen population distributions. Findings were simulated for both small and large samples when population variances were different between two samples. This current study agreed with Siegel and Castellan (1988) and Baumgartner, WeiB, and Shindler's study (1998) that the KS-2 test was more powerful for small samples when population variances were not equal. For example, in Baumgartner, WeiB, and Shindler's study (1998), a figure of

simulated power functions demonstrated that the KS-2 test had more power than the MW test with a sample size of (10, 10) and an increase in population variance in normally distributed samples. The current study found that the KS-2 test and the MW test had a statistical power of 0.082 and 0.070, respectively, with a SD ratio of 4. As sample sizes increased, the power of the KS-2 test increased too. Moreover, the present study provided evidence that the KS-2 test was much more powerful than the MW test with large sample sizes when the population variances were extremely different between the two samples. When the SD ratios were extremely large or extremely small with a large sample size, statistical power for the KS-2 test was substantially large in both normal and non-normal population distributions.

In conclusion, when the condition of heterogeneity of variance between two populations existed in the two equal-sized small samples, the KS-2 test and the MW test had similar statistical power. However, the KS-2 test had much greater statistical power than the MW test when sample sizes were equal and large.

Difference in Skewness

The third research question investigated statistical power of the MW and the KS-2 tests between two equal-sized samples with different degrees of skewness in their underlying population distributions. It was assumed that the two samples were from populations with the same kurtosis and SD ratios (SD ratio = 1) but different skewness. The four pairs of equal sample sizes used for simulations were (8, 8), (16, 16), (25, 25) and (50, 50).

Simulation results suggested that both the MW and KS-2 tests had small statistical power regardless of the differences in skewness between the two underlying population distributions. When the two sample sizes were small, the MW test had more power than the KS-2 test despite differences in skewness between the two underlying population

distributions. As sample sizes increased and skewness between the two populations were separated from one the other, the KS-2 test had more power than the MW test in most of the simulations.

Literature such as Penfield (1994) pointed out that the MW test had more power than other nonparametric two-sample tests (the van der Wredden Normal Score (NS) test and the Welch-Aspin-Sattertheaite (W) test) under various degrees of the kurtosis and skewness. However, Penfield (1994) did not compare statistical power between the MW test and the KS-2 test when there is only different skewness between the two populations. The present study provided evidence that the MW test had more power than the KS-2 test in specified ratios of skewness and kurtosis when the two samples were small and had the same sizes and SD ratio. When the size of the two samples started to increase, the KS-2 test became superior to the MW test regarding statistical power in most of the comparisons when the two underlying populations had two different skewness ratios with the same kurtosis.

Overall, the KS-2 test and the MW test had small statistical power when only skewness ratios varied for both small and large equal sized samples. The MW test was more powerful than the KS-2 test when sample size was small regardless of the difference in skewness. As the difference in skewness between the two populations became more than 0.5 in large sample settings, such as (50, 50), the KS-2 test became superior of the MW test in statistical power.

Difference in Kurtosis

The last research question considered the statistical power of the MW and the KS-2 tests between two samples equal in size but different in degrees of kurtosis in their underlying population distributions. It was assumed that the two samples were from populations with

the same kurtosis and SD ratios (SD ratio = 1) but with different skewness. Simulated sample sizes were (8, 8), (16, 16), (25, 25) and (50, 50). Findings suggested that the values of statistical power were small for the MW test and the KS-2 test under the simulation conditions.

By comparing statistical power between the MW test and the KS-2 test, this study concluded that when sample sizes were small, statistical power of the MW test was small and superior to the KS-2 test in spite of the difference in the degrees of kurtosis. If the differences in the kurtosis between two population distributions were more than 2.0 and sample sizes increased, the KS-2 test had more power than the MW test in most of the simulations. When the skewness ratio was zero and the kurtosis ratios of two population distributions were apart from one another, power of the KS-2 test was inferior to the MW test as the two samples increased in size.

It is difficult to locate literature which focuses on the comparison between the MW test and the KS-2 test and power estimates when only differences in the degree of kurtosis exist with normal and non-normal population distributions. This simulation study is unique in presenting evidence that the MW test had slightly more statistical power than the KS-2 test in two-sample comparisons, when the two underlying population distributions had the same skewness but differed mildly in kurtosis. The KS-2 test had more statistical power when the two underlying population distributions had the same skewness but the difference in kurtosis was more than 2.0 with large and equal sample sizes like (50, 50).

Generally, the KS-2 test and the MW test had small statistical power when only kurtosis varied for both small and large equal sized samples. The MW test had more statistical power than the KS-2 test when sample size was equal and small regardless of the difference in

kurtosis. When the difference of kurtosis between the two populations was more than 2.0 in large sample settings, such as (50, 50), the KS-2 test became more powerful than the MW test.

Practical Implications

This study presents two main practical implications. First, an explanation is provided concerning why the effect size was not appropriate for performing statistical power simulations. Next, this study has provided guidelines for researchers who choose between the MW test and the KS-2 test for hypothesis testing.

Method to Simulate Statistical Power

When estimating the power of a statistical test, most researchers, such as Cohen (1988) and Murphy and Myors (1998) suggest that statistical power relies on the significance level (α) and effect size. Effect size (d) is a function of the difference between two population means divided by the population variance. Equal variance is required in finding an effect size. The formula for effect size provided by Cohen (1988) is:

$$\text{Effect size (d)} = \frac{|\mu_B - \mu_A|}{\sigma};$$

where μ_B and μ_A are population means for the two samples;

σ is the population standard deviation for either sample (equal variance is assumed)

In this study, a hypothesis test was used to evaluate whether there was a general difference between the two samples. Heterogeneity of variance, difference in skewness, and difference in kurtosis were investigated through the simulations. The main focus of this simulation study was not in the difference in means between two samples. Moreover, when

either of the populations of the two samples changed in variance, skewness, and/or kurtosis, the two population distributions also change. Therefore, effect size was not applicable for determining statistical power in this simulation study.

Advice to Researchers

The present simulation study used a predetermined significance level of 0.05 to assess statistical power. The method of finding statistical power for either the MW test or the KS-2 test involved determining the proportion of the number of hypothesis tests under the determined condition with statistical significance (p-value less than the significance level) out of the total number of replications. The larger the proportion, the greater the statistical power for the MW or the KS-2 test.

The MW test and the KS-2 test are both nonparametric statistical techniques used to perform a hypothesis test on determining a general difference between two populations. The current simulation study presented suggestions for researchers in determining which one of these two nonparametric statistical techniques should be applied (Also in Table 38):

- (1) When the two samples are different in sample sizes only, the KS-2 test is the recommended statistical test. The KS-2 test is much more lower on Type I error rates than the MW test. Moreover, the KS-2 test has more statistical power than the MW test under this condition. In the other words, when researchers use the KS-2 technique for hypothesis testing, the KS-2 test is more sensitive to rejecting the null hypothesis; moreover, the finding from the hypothesis test is more likely to generalize from sample data back to populations.
- (2) When the two samples are different in population variance only, the KS-2 test has more statistical power than the MW test when the sample sizes are large for the two

- samples. When the size of the two samples is small, and if the population SD ratios are extremely large or small, the MW has greater statistical power than the KS-2 test.
- (3) When the two samples differ in the degree of skewness between two the underlying population distributions, the MW test has more statistical power than the KS-2 test in small samples. This is true regardless of the differences in the degree of skewness. The KS-2 test has more statistical power when the difference in the degree of the skewness is more than 0.50 with large samples.
- (4) When the two samples differ in the degree of kurtosis between two underlying population distributions, the MW test has more statistical power than the KS-2 test in small and large samples when the difference in the of the degree of kurtosis is at most 2.0. The KS-2 test has more statistical power when the difference in the degree of the kurtosis is more than 2.0 in large samples.

Table 39: Summary of the Conditions to Use the MW or the KS-2 Test

Condition 1: Unequal Sample Size		
Sample Size	Population SD Ratio	Test
$n_1 < n_2$	1	KS-2
(4, 16)	1/4, 1/3, 3, 4	MW
(10, 20), (10, 30) (50, 100)	1/4, 1/3, 3, 4	KS-2

Table 39 Cont.: Summary of the Conditions to Use the MW or the KS-2 Test

Condition 2: Equal Sample Size		
1. Population SD Ratio	Sample Size	Test
1/4, 1/3, 3, 4	$n_1 = n_2 = 8$	KS-2
	$n_1 = n_2 > 8$	KS-2
1/2, 2	$n_1 = n_2 = 8$	MW
	$n_1 = n_2 > 8$	KS-2
2. Differences of two Skewness Ratios	Sample Size	Test
≤ 0.5	$n_1 = n_2 \leq 25$	MW
	$n_1 = n_2 > 25$	MW
> 0.5	$n_1 = n_2 \leq 25$	MW
	$n_1 = n_2 > 25$	KS-2
3. Differences of two Kurtosis Ratios	Sample Size	Test
≤ 2	$n_1 = n_2 \leq 25$	MW
	$n_1 = n_2 > 25$	MW
> 2	$n_1 = n_2 \leq 25$	MW
	$n_1 = n_2 > 25$	KS-2

Limitations of the Study

There are several limitations for the current Monte Carlo simulation study. First, the issues of tied data are excluded in this study. Researchers, such as Siegel and Castellan (1988), Neave and Worthington (1988), and Conover (1999), revealed that variability in the sets of ranks are affected by tied ranks. They suggested using a tie correction formula as a

compromise to the problem when performing the MW test. However, researchers have not provided clarity of the definition of ties and when to use the test statistic formulas of tied conditions. Similar but more complicated discussions have taken place when the KS-2 test was performed under the tied condition. Some researcher, such as Bradley (1968) and Marascuilo and McSweeney (1977) claimed that an originally observed variable was a continuous variable implying that no tied observations occurred in their samples. Daniel (1990) claimed that there was no problem when tied scores were presented within the same sample group while complications arose when the tied condition happened between two sample groups. Other researchers, such as Siegel and Castellan (1988), Conover (1999), Sheskin (2000), and Higgins (2004), did not discuss the issue of ties. Due to a lack of clarity among the definition of ties for the various notable authors, this study did not address the issue of ties. In other words, tied scores were not considered in this study.

Next, pairs of equal and unequal sample sizes were selected for inclusion based on literature. The purpose of such sample selection was to compare the simulation results in Type I errors and statistical power between the MW test and the KS-2 test in conjunction with previous literature. However, there are many different pairs of sample sizes other than the ones in this current research. Sample sizes are often selected by researchers because of their individual research settings.

Lastly, values of skewness and kurtosis were limited, too. The selected degrees of skewness and kurtosis were based upon the 15 population distributions utilized in this study. Other unknown, named non-normal distributions existed, along with degrees of skewness and kurtosis which vary due to the shape of the data distribution.

Recommendations for Future Research

This study was designed to explore Type I error rates and statistical power between the KS-2 test and the MW test under specific and separate conditions: (1) unequal sample size, (2) heterogeneity of variance, (3) difference in skewness, and (4) difference in kurtosis between the two underlying population distributions. When two underlying populations differ in their distributions, examining statistical power becomes essential in statistical tests. Murphy and Myers (1998) clearly describe how statistical power affects researchers in the decision making process:

Studies with too little statistical power can frequently lead to erroneous conclusions. In particular, they will very often lead to the incorrect conclusion that findings reported in a particular study are not likely to be true in a broader population. (p. 1)

Murphy and Myers (1998) pointed out the importance of statistical power in the social and behavioral sciences when researchers perform statistical tests for their study. When statistical power is too small, the results of the hypothesis tests may not be generalizable to the population.

There is substantial research on statistical power between the MW test and parametric statistical techniques, such as the Student's t test. However, when researchers try to determine whether to use either the MW test or the KS-2 test for evaluating a general difference between two samples, there is inadequate research on Type I error rates and statistical power between these two tests to support a decision. This research performed simulations under predetermined conditions for only one of the effects under fifteen population distributions between the KS-2 and the MW tests. It is hoped that future researchers are aided in strengthening their decision to perform either of these two

nonparametric statistical tests in their studies. However, the reported results were simulated based on a limited number of conditions. The simulations were also executed one condition at a time. Future research can expand simulations in the areas suggested below:

- (1) Interaction effects: If two or more of the effects (such as different population variances, different degrees of the skewness and kurtosis) explored in this study occur simultaneously, what is the statistical power for the MW test or the KS-2 test? The present study simulated statistical power for the MW test and the KS-2 test when only one of the following conditions occurs: heterogeneity of variance, difference in skewness, or difference in kurtosis. It is possible that two underlying non-normal populations differ in variance and skewness, variance and kurtosis, skewness and kurtosis, or even in variance, skewness, and kurtosis when the two samples differ in size. It is recommended that the conditions explored here to be combined, and interaction effects might be analyzed.
- (2) Sample sizes: This study examined four pairs of equal sample sizes and eight pairs of unequal sample sizes. However, there are still many pairs of equal and unequal sample sizes that should be simulated. Such an assessment might assist researchers in finding a nonparametric statistical test between the MW and the KS-2 tests with a higher statistical power. A higher statistical power may ensure the chance of generalizing the findings of the hypothesis test to setting with larger populations.
- (3) Skewness and Kurtosis: The present study simulated statistical power with 15 populations and equal samples in size. When comparing statistical power, only some values of different degrees of kurtosis and skewness in these populations were used for the simulations. There are non-normal distributions with various degrees of

kurtosis and skewness, other than the ones explored in this current study. Future researchers can perform Monte Carlo simulations and compare statistical power for the MW test and the KS-2 test under various degrees of kurtosis and skewness and combinations of the unequal sample size condition to help researchers select the most powerful two-sample nonparametric test, either the MW test or the KS-2 test.

In conclusion, the Mann-Whitney and the Kolmogrov-Smirnov two samples nonparametric statistical tests are known for the hypothesis tests of general difference between two samples. They are utilized when samples are violated the assumption of normality in the populations and the measurement of samples is at least ordinal. This current study compared the statistical power and Type I errors (if applicable) between these two nonparametric techniques. The study revealed that the KS-2 test was more powerful than the MW test when the two samples have unequal size. The KS-2 test had smaller Type I error rates than the MW test under this condition too. The MW test had slightly more statistical power the KS-2 test under the condition of small and equal-sized samples. However, when the two equal samples were large and at least 25 with the underlying non-normal populations, the KS-2 test had more power than the MW test.

Furthermore, there are still areas the need future research to fill the gap such as comparison the statistical power for the KS-2 and the MW tests when two unequal-sizes samples with different population variance, skewness, or kurtosis. The optimal goal of this study is to provide guidelines for researchers in strengthening their decision when selecting either of these two nonparametric statistical tests in their studies.

References

- Algina, J., Olejnik, S., & Ocanto, R. (1989). Type I error rates and power estimates for selected two-sample tests of scale. *Journal of Educational Statistics, 14*(4), 373-384.
- Bai, J., & Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics, 23*(1), 49-60.
- Balakrishnan, N., & Nevzorov, V. B. (2003). *A primer on statistical distributions*. Hoboken, New Jersey: A John Wiley & Sons, Inc.
- Baumgartner, W., WeiB, P., & Shindler, H. (1998). A nonparametric test for the general two-sample problem. *Biometrics, 54*, 1129-1135.
- Blair, R. C., & Higgins, J. J. (1985). Comparison of the power of the paired samples t-test to that of Wilcoxon's sign-ranks test under various population shapes *Psychological Bulletin, 97*(1), 119-128.
- Bradley, J. V. (1968). *Distribution-free statistical tests* Englewood Cliffs, N.J.: Prentice-Hall
- Buning, H. (2001). Kolmogorov-Simrnov and Cramer-Von Mises type two-sample tests with various weight functions. *Communication Statistics, 30*(4), 847-865.
- Carolan, C. A., & Tebbs, J. A. (2005). Nonparametric test for and against likelihood ratios ordering in the two-sample problem. *Biometrika, 92*(1), 159-171.
- Cliff, N., & Keats, J. A. (2003). *Ordinal measurement in the behavioral sciences* (2 ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Hillsdale, N. J.: Lawrence Erlbaum Associates.

- Conover, W. J. (1999). *Practical nonparametric statistics* (3 ed.). New York: John Wiley & Sons, Inc.
- Conover, W. J. (2005). Practical nonparametric statistics. In C. H. Lee (Ed.).
- Daniel, W. W. (1990). *Applied nonparametric statistics* (2 ed.). Boston: PWS-Kent Publishing Company.
- Dixon, W. J. (1954). Power under normality of several nonparametric tests. *Annals of Mathematical Statistics*, 20, 393-403.
- Fahoome, G. (1999). *A Monte Carlo study of twenty-one nonparametric statistics with normal and nonnormal data*. Unpublished doctoral dissertation, Wayne State University, Detroit, MI.
- Fahoome, G., & Sawilowsky, S. S. (2000, April 24-28). *Review of twenty nonparametric statistics and their large sample approximations*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Fan, X., Felsovalyi, A., Sivo, S. A., & Keenan, S. C. (2003). *SAS for Monte Carlo studies: A guide for quantitative researchers*. Cary, NC: SAS Publishing.
- Fleishman, A. I. (1978). A method for simulating non-normal distribution. *Psychometrika*, 43, 521-532.
- Freund, J. E., & Williams, F. J. (1966). *Dictionary/outline of basic statistics*. New York: Dover Publications, Inc. .
- Gibbons, J. D., & Chakraborti, S. (1991). Comparisons of Mann-Whitney, student's t, and alternate t test for means of normal distributions. *Journal of Experimental Education*, 59(3), 258-267.

- Gibbons, J. D., & Chakraborti, S. (2003). *Nonparametric statistical inference* (4th ed.). New York, NY: Marcel Dekker, Inc.
- Higgins, J. J. (2004). *Introduction to modern nonparametric statistics*. Pacific Grove: Thomson Learning, Inc.
- Joanest, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *The Statistician* 47(1), 183-189.
- Kasuya, E. (2001). Mann-Whitney test when variances are unequal. *Animal Behaviour*, 61(6), 1247-1249.
- Keselman, H. J., & Cribbie, R. (1997). Specialized tests for detecting treatment effects in the two-sample problems. *Journal of Experimental Education*, 65(4), 355-366.
- Krishnaiah, P. R., & Sen, P. K. (1984). *Nonparametric methods* (Vol. 4). New York: Elsevier Science Publishers B. V.
- Lee, C. H. (2005, April 22). *Factors affecting student learning outcomes among engineering students in statistics courses* Paper presented at the Twenty-Third Annual Oklahoma Psychological Society Spring Research Conference, Edmond, OK.
- MacDonald, P. (1999). Power, Type I and Type III error rates of parametric and nonparametric statistical tests. *Journal of Experimental Education*, 67(4), 367-379.
- Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for social science*. Monterey, CA: Brooks/Cole Publishing Company.
- Micerri, T. (1989). The Unicorn, normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156-166.
- Mooney, C. Z. (1997). *Monte Carlo simulation*. Thousand Oaks, CA: SAGE Publications, Inc.

- Murphy, K. R. & Myors, B. (1998). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Neave, H. R., & Worthington, P. L. (1988). *Distribution-free tests*. London: Unwin Hyman Ltd.
- Noether, G. (1967). *Elements of nonparametric statistics*. New York: Wiley.
- Olejnik, S. F., & Algina, J. (1987). Type I error rates and power estimates of selected parametric and nonparametric tests of scale. *Journal of Educational Statistics*, 12(1), 45-61.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Penfield, D. A. (1994). Choosing a two-sample location test. *Journal of Experimental Education* 62(4), 343-350.
- Pratt, J. W., & Gibbons, J. D. (1981). *Concepts of nonparametric theory*. New York: Springer-Verlag.
- Sackrowitz, H., & Samuel-Cahn, E. (1999). P-values as random variables -expected p values. *The American Statistician*, 53(4), 326-331.
- Schroer, G., & Trenkler, D. (1995). Exact and randomization distributions of Kolmogorov-Smirnov tests two or three samples. *Computational Statistics & Data Analysis*, 20, 185-202.
- Shavelson, R. J. (1988). *Statistical reasoning for the behavioral sciences* (2nd ed.). Boston: Allyn and Bacon, Inc.

- Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures*.
Boca Raton: Chapman & Hall/CRC.
- Siegel, S., & Castellan, J. N. J. (1988). *Nonparametric statistics for the behavioral sciences*
(2 ed.). Boston, Massachusetts: McGraw-Hill.
- Sprent, P., & Smeeton, N. C. (2001). *Applied Nonparametric Statistical Methods* (3rd ed.).
New York Chapman & Hall /CRC.
- Statistical analysis system. (1999). Cary, NC: SAS Institute Inc.
- Vogt, W. P. (1993). *Dictionary of statistics and methodology: A nontechnical guide for the
social sciences*. Newbury Park, CA: Sage Publications. .
- Vogt, W. P. (2005). *Dictionary of statistics and methodology: A nontechnical guide for the
social sciences* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Wilcox, R. R. (1997). Some practical reasons for reconsidering the Kolmogorov-Smirnov
test. *British Journal of Mathematical & Statistical Psychology*, *50*(1), 9-20.
- Wolfram, S. (2003). Mathematica 5.0. Champaign, IL Wolfram Research, Inc.
- Zimmerman, D. W. (1985). Power functions of the t test and Mann-Whitney U test under
violation of parametric assumptions. *Perceptual and Motor Skills*, *61*, 467-470.
- Zimmerman, D. W. (1987). Comparative power of Student t test and Mann-Whitney U test
for unequal sample sizes and variances. *The Journal of Experimental Education*, *55*,
171-174.
- 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. (2000). Statistical significance levels of nonparametric test biased by heterogeneous variance of treatment groups. *The Journal of General Psychology* 127(4), 354-364.

Zimmerman, D. W. (2001b). Mimicking properties of nonparametric rank tests using scores that are not ranks. *The Journal of General Psychology*, 120(4), 509-516.

Zimmerman, D. W. (2003). A warning about the large-sample Wilcoxon-Whitney test. *Understanding Statistics* 2(4), 267-280.

Zimmerman, D. W. (2004). Inflation of type I error rates by unequal variance associated with parametric, nonparametric, and rank-transformation tests. *Psicologica*, 25, 103-133.

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

APPENDICES

APPENDIX I: Coefficients of Fleishman's power function (1978)

Skew	Kurto- sis	B	C	D
1.75	3.75	0.92966052480111	0.39949667453766	-0.03646699281275
1.50	3.75	0.86588620352314	0.22102762101262	0.02722069915809
1.50	3.50	0.88690855456083	0.23272187792846	0.01875401444244
1.50	3.25	0.91023877496903	0.24780864411835	0.00869952997029
1.50	3.00	0.93620992090360	0.26831868322542	-0.00368190099903
1.50	2.75	0.96443747224458	0.29807621191230	-0.01963521430303
1.50	2.50	0.99209856718687	0.34526935903177	-0.04181526211241
1.25	3.75	0.81888156132542	0.16064255561731	0.04916517172492
1.25	3.50	0.83472669039047	0.16546665419634	0.04385221308384
1.25	3.25	0.85174062710067	0.17101073821620	0.03803066692496
1.25	3.00	0.87016387886005	0.17749222807992	0.03157509494526
1.25	2.75	0.89031839050274	0.18523508277808	0.02430713561023
1.25	2.50	0.91264314105424	0.19474622768576	0.01596248199126
1.25	2.25	0.93774043576005	0.20686671601473	0.00613024990315
1.25	2.00	0.96640616806420	0.22308878847471	-0.00586255218100
1.25	1.75	0.99949784644724	0.24624842887675	-0.02117724378041
1.25	1.50	1.03732397122554	0.28227102596774	-0.04209052633812
1.00	3.75	0.78942074416451	0.11942383662867	0.06153961924505
1.00	3.50	0.80290583376385	0.12210992461489	0.05733551785617
1.00	3.25	0.81713276543078	0.12508112045759	0.05284904949443
1.00	3.00	0.83221632289426	0.12839670935047	0.04803205907079
1.00	2.75	0.84830145553715	0.13213547065430	0.04282253816015
1.00	2.50	0.86557488958491	0.13640488393449	0.03713875125893
1.00	2.25	0.88428277118735	0.14135625914686	0.03086993391983
1.00	2.00	0.90475830311225	0.14721081863342	0.02386092280190
1.00	1.75	0.92746633976156	0.15430725098288	0.01588548300086
1.00	1.50	0.95307689770618	0.16319427626410	0.00659736974453
1.00	1.25	0.98258511915167	0.17482469452982	-0.00456507744552
1.00	1.00	1.01748518639311	0.19099508385633	-0.01857699796908
1.00	0.75	1.05993380621160	0.21543408088777	-0.03728846051332
1.00	0.50	1.11465523356736	0.25852489125964	-0.06601339414569
0.75	3.75	0.76995202064185	0.08563059561704	0.06934855449019
0.75	3.50	0.78217273051806	0.08725259129727	0.06568498605230
0.75	3.25	0.79495262685357	0.08901550782281	0.06182200554749
0.75	3.00	0.80836339881256	0.09094289213403	0.05773230755921
0.75	2.75	0.82249224466377	0.09306441724945	0.05338234897766
0.75	2.50	0.83744678260912	0.09541813650752	0.04873027147556
0.75	2.25	0.85336207930094	0.09805385102577	0.04372288092319
0.75	2.00	0.87041098768531	0.10103830525054	0.03829112262516
0.75	1.75	0.88881983583405	0.10446351079607	0.03234306422362
0.75	1.50	0.90889310938952	0.10846068760906	0.02575256208705
0.75	1.25	0.93105392309623	0.11322488108796	0.01834005494883
0.75	1.00	0.95591357125244	0.11906128313604	0.00983810049833
0.75	0.75	0.98439732894675	0.12647935041464	-0.00017482979206
0.75	0.50	1.01798354640471	0.13640251351290	-0.01241224193515
0.75	0.25	1.05917362852414	0.15068875788687	-0.02819626089809
0.75	0.0	1.11251460048528	0.17363001955694	-0.05033444870926
0.75	-0.25	1.20392340617686	0.22758947506748	-0.09549567396576
0.50	3.75	0.75739984777977	0.05552444121576	0.07425915142054
0.50	3.50	0.76890587541111	0.05647215540722	0.07088677148643

APPENDIX I CONT.: Coefficients of Fleishman's power function

Skew	Kurto- sis	B	C	D
0.50	3.25	0.78088173005011	0.05749287097856	0.06735271683459
0.50	3.00	0.79338100476375	0.05859728796468	0.06363759352080
0.50	2.75	0.80646754404870	0.05979852701132	0.05971815189213
0.50	2.50	0.82021829990300	0.06111289719250	0.05556617644075
0.50	2.25	0.83472726718530	0.06256098771565	0.05114694309845
0.50	2.00	0.85011102914029	0.06416925946524	0.04641702467833
0.50	1.75	0.86651677519629	0.06597243296920	0.04132108400060
0.50	1.50	0.88413424213468	0.06801719309367	0.03578703942948
0.50	1.25	0.90321412393338	0.07036816659914	0.02971850575197
0.50	1.00	0.92409763318404	0.07311802793159	0.02298245181387
0.50	0.75	0.94726632241948	0.07640557409735	0.01538797196646
0.50	0.50	0.97343106918044	0.08045036185716	0.00664738328997
0.50	0.25	1.00370252335312	0.08562503291528	-0.00370088000554
0.50	0.0	1.03994603972583	0.09262357406250	-0.01646085654705
0.50	-0.25	1.08559667905205	0.10290996902235	-0.03319706659066
0.50	-0.50	1.14784905722603	0.12015606910630	-0.05750353451604
0.25	3.75	0.75031534111078	0.02734119591845	0.07699282409939
0.25	3.50	0.76144830727079	0.02778212551548	0.07376857545917
0.25	3.25	0.77300829583485	0.02825487458003	0.07040005844916
0.25	3.00	0.78504099113665	0.02876378789438	0.06687116600052
0.25	2.75	0.79760024256974	0.02931412174287	0.06316282101938
0.25	2.50	0.81075018336126	0.02991231084290	0.05925218604949
0.25	2.25	0.82456809276114	0.03056633874422	0.05511158940232
0.25	2.00	0.83914834011794	0.03128626308577	0.05070703595619
0.25	1.75	0.85460794420601	0.03208497913365	0.04599609338072
0.25	1.50	0.87109461567493	0.03297936179585	0.04092481046466
0.25	1.25	0.88879874777889	0.03399203130579	0.03542308246001
0.25	1.00	0.90797193683084	0.03515419180007	0.02939742137986
0.25	0.75	0.92895681403887	0.03651041219964	0.02271917644022
0.25	0.50	0.95223758733324	0.03812714596039	0.01520430356261
0.25	0.25	0.97853113001303	0.04010900967596	0.00657629354597
0.25	0.0	1.00896426283423	0.04263274479965	-0.00360752773660
0.25	-0.25	1.04545395821482	0.04602657996297	-0.01611868374910
0.25	-0.50	1.09162984652106	0.05098546424880	-0.03246963121043
0.25	-0.75	1.15546858231190	0.05928145029513	-0.05617881116691
0.25	-1.00	1.26341280092760	0.07746243900117	-0.10003604502301
0.0	3.75	0.74802080799221	0.0	0.07787271610187
0.0	3.50	0.75903729021108	0.0	0.07469419122736
0.0	3.25	0.77046795694613	0.0	0.07137653241549
0.0	3.00	0.78235622045349	0.0	0.06790455640586
0.0	2.75	0.79475308530197	0.0	0.06426034643397
0.0	2.50	0.80771907418732	0.0	0.06042254280525
0.0	2.25	0.82132681354781	0.0	0.05636538554628
0.0	2.00	0.83566457198565	0.0	0.05205739701455
0.0	1.75	0.85084120886649	0.0	0.04745952834774
0.0	1.50	0.86699326941512	0.0	0.04252248423852
0.0	1.25	0.88429545439108	0.0	0.03718274611280
0.0	1.00	0.90297659829926	0.0	0.03135645239684
0.0	0.75	0.92334504635701	0.0	0.02492958648521
0.0	0.50	0.94583093702434	0.0	0.01774144664586
0.0	0.25	0.97106090002478	0.0	0.00955505507423
0.0	0.0	1.00000000000000	0.0	0.0
0.0	-0.25	1.03424763182041	0.0	-0.01154929007313
0.0	-0.50	1.07673274256343	0.0	-0.02626832123859
0.0	-0.75	1.13362194989244	0.0	-0.04673170311060
0.0	-1.00	1.22100956933052	0.0	-0.08015837236135
-0.25	3.75	0.75031534111078	-0.02734119591845	0.07699282409939

APPENDIX II: E-mail Contact with Conover, W. J.

Subject: A question about one formula in your book

October 4, 2005

Chin-Huey,

There are no guidelines as to when we cross over from “just a few ties” to “many ties.” I suggest that if you are in doubt, you should use the formulas for “many ties” especially if this is a simulation where the number of ties will vary.

Dr. Conover

APPENDIX III: A Sample of SAS Syntax for Generating Population Distributions

```
DM "OUTPUT; CLEAR; LOG; CLEAR;";
OPTIONS LS=80 PS=75 NODATE;

*****;
** AUTHOR: Chin-Huey Lee **;
** DATE: 10-26-06 **;
** UPDATE: **;
** PURPOSE: GENERATE NORMAL POPULATION DISTRIBUTIONS AND HISTOGRAM **;
** REPLICATED 20000 SAMPLES of SIZE (8, 8) **;
*****;

*1. GENERATE NORMAL POPULATION & HISTOGRAM;
DATA NORMAL;
  DO I=1 TO 20000;
    Y = RANNOR (0);
    OUTPUT;
  END;
RUN;
title 'Normal Population Distribution Histogram';
PROC UNIVARIATE data=Normal;
  var Y;
  histogram ;
RUN;
QUIT;
```

APPENDIX IV: A Sample of SAS Syntax for Sampling Procedure

```
DM "OUTPUT; CLEAR; LOG; CLEAR;";
OPTIONS LS=80 PS=75 NODATE;

*****;
** AUTHOR: Chin-Huey Lee **;
** DATE: 10-26-06 **;
** UPDATE: **;
** PURPOSE: GENERATE NORMAL POPULATIONS with SD ratio = 1 **;
** REPLICATED 20000 SAMPLES of SIZE (4, 16) **;
** NOTES: **;
*****;

DATA NORMALS1;
DO I=1 TO 20000;
    DO J = 1 TO 4;
        Y = RANNOR (0);
        OUTPUT;
    END;
    DO K = 1 TO 16;
        Y = RANNOR (0);
        OUTPUT;
    END;
END;
RUN;

DATA SETNS1;
SET NORMALS1;
IF J < 5 THEN J =1;
IF J = 5 THEN J =2;
KEEP I J Y;
RUN;

*PERFORM MAN-WHITNEY AND KS-2 TESTS;
PROC NPAR1WAY WILCOXON EDF CORRECT = NO NOPRINT;
BY I;
CLASS J;
VAR Y;
OUTPUT OUT = OUTPUTNS1 WILCOXON EDF;
RUN;
PROC EXPORT DATA=OUTPUTNS1

OUTFILE="F:\DISSERTATION\OUTPUT1\EQUAL_SD\SAMPLE1\OUTPUTNS1.XLS"
        DBMS= EXCEL REPLACE;

RUN;

*SAVE P-VALUE FOR THE MANN-WHITNEY AND THE KS-2 TESTS;
DATA P_VALUE_NS1;
SET OUTPUTNS1;
KEEP P2_WIL P_KSA;
RUN;

*FIND TYPE I ERROR RATE;
DATA MWTYPEI_NS1;
    SET P_VALUE_NS1;
    RETAIN COUNT 0;
    IF P2_WIL < = .05 THEN COUNT = COUNT +1;
    RENAME COUNT = MW_COUNT;
RUN;
```

```
DATA KSTYPEI_NS1;
SET P_VALUE_NS1;
RETAIN COUNT 0;
  IF P_KSA < = .05 THEN COUNT = COUNT +1;
  RENAME COUNT =KS_COUNT;
RUN;
DATA TYPEI_NS1;
MERGE MWTYPEI_NS1 KSTYPEI_NS1;
RUN;
PROC EXPORT DATA=TYPEI_NS1

OUTFILE="F:\DISSERTATION\OUTPUT1\EQUAL_SD\SAMPLE1\TYPEI_N1S1.XLS"
        DBMS= EXCEL REPLACE;
RUN;
QUIT;
```


APPENDIX V: Histograms of Population Distributions (N = 20000; Y-axis: relative frequency, X-axis: Z score)

Figure 53: Histogram; Normal Population Distribution (Skewness = 0.00, Kurtosis = 0.00)

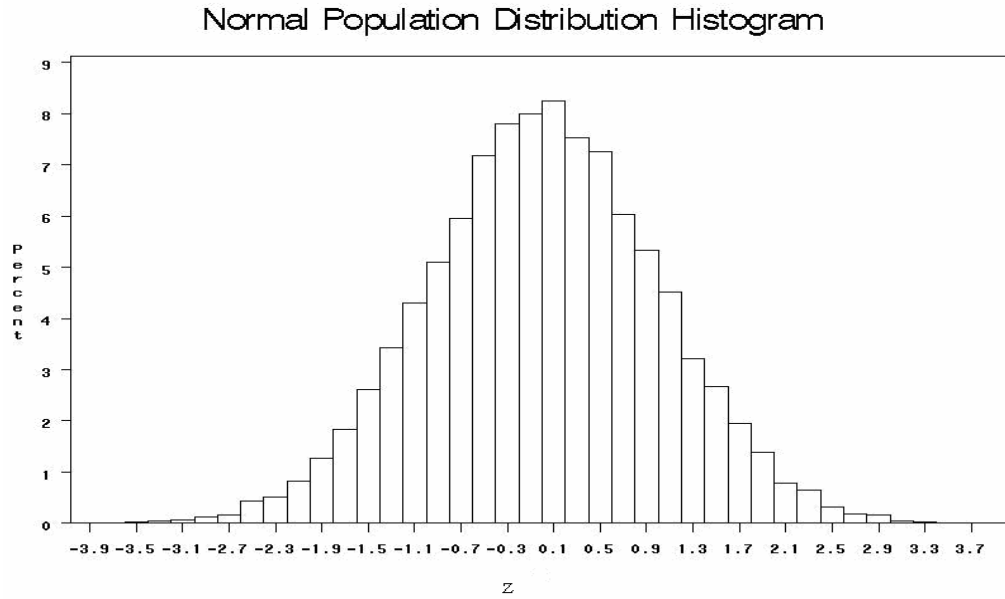


Figure 54: Histogram; Platykurtic Population Distribution (Skewness = 0.00, Kurtosis = -0.50)

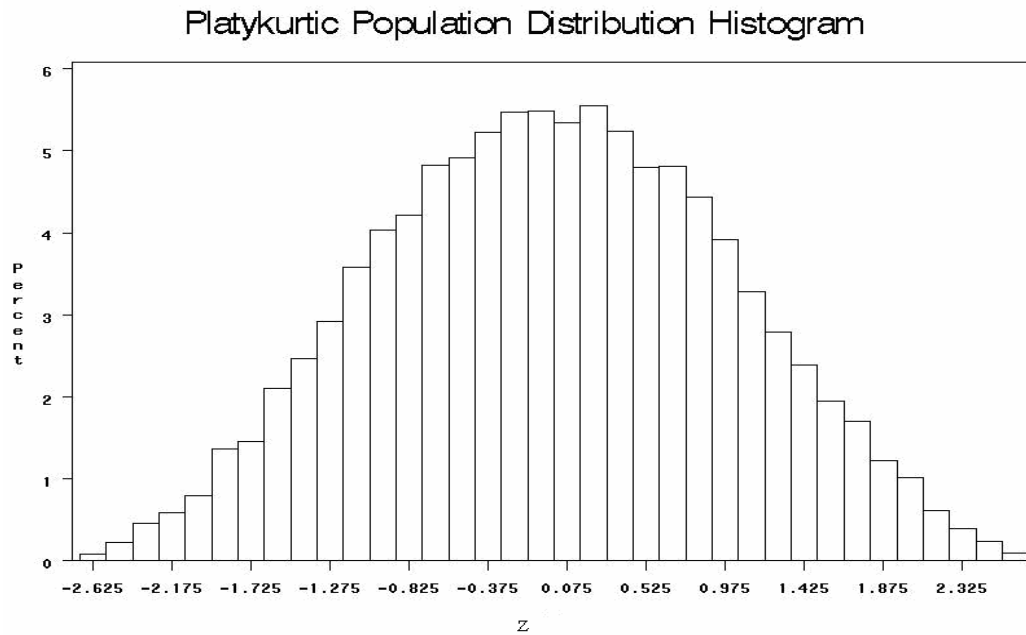


Figure 55: Histogram; Normal Platykurtic Population Distribution (Skewness = 0.00, Kurtosis = -1.00)

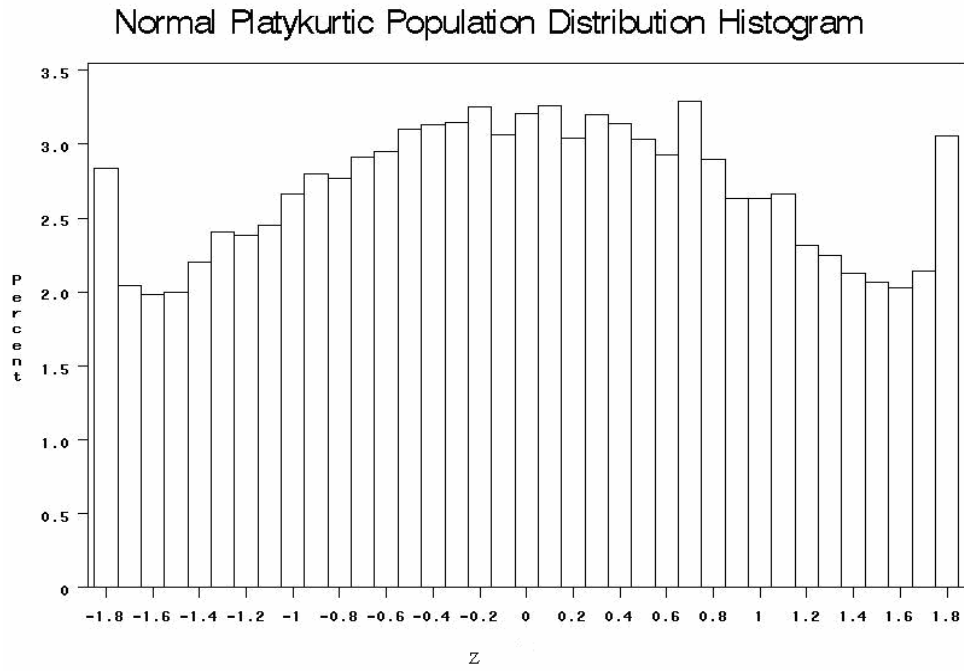


Figure 56: Histogram; Leptokurtic_1 Population Distribution (Skewness = 0.00, Kurtosis = 1.00)

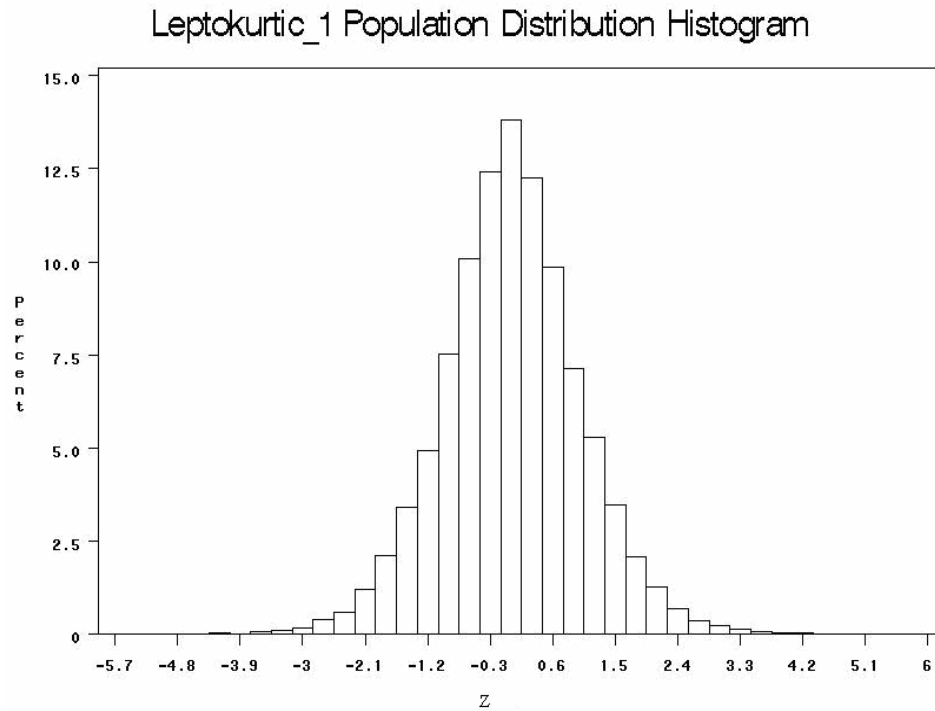


Figure 57: Histogram; Leptokurtic_2 Population Distribution (Skewness = 0.00, Kurtosis = 2.00)

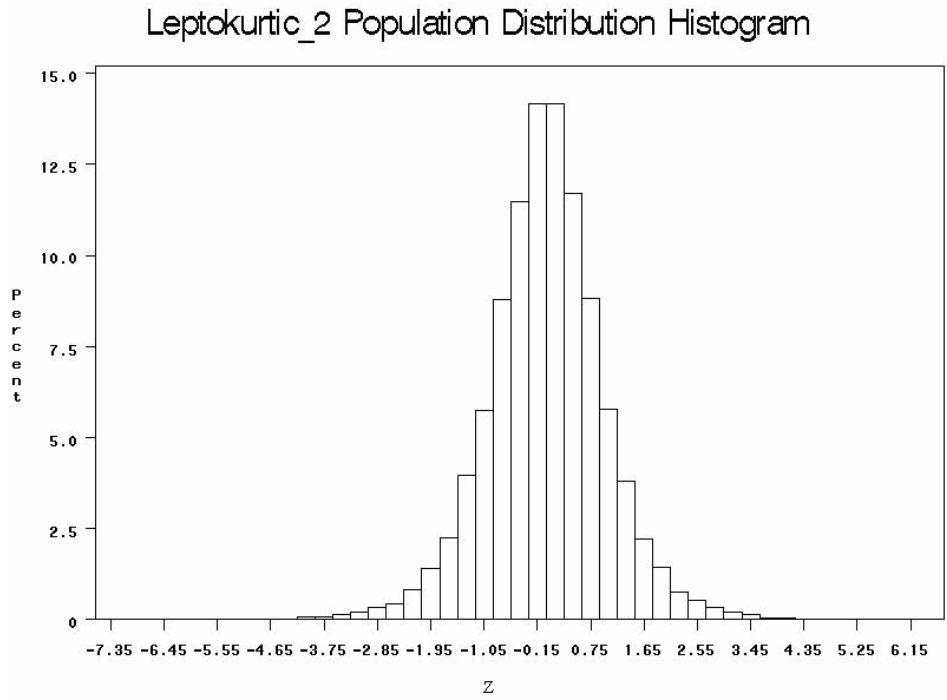


Figure 58: Histogram; Leptokurtic_3 Population Distribution (Skewness = 0.00, Kurtosis = 3.75)

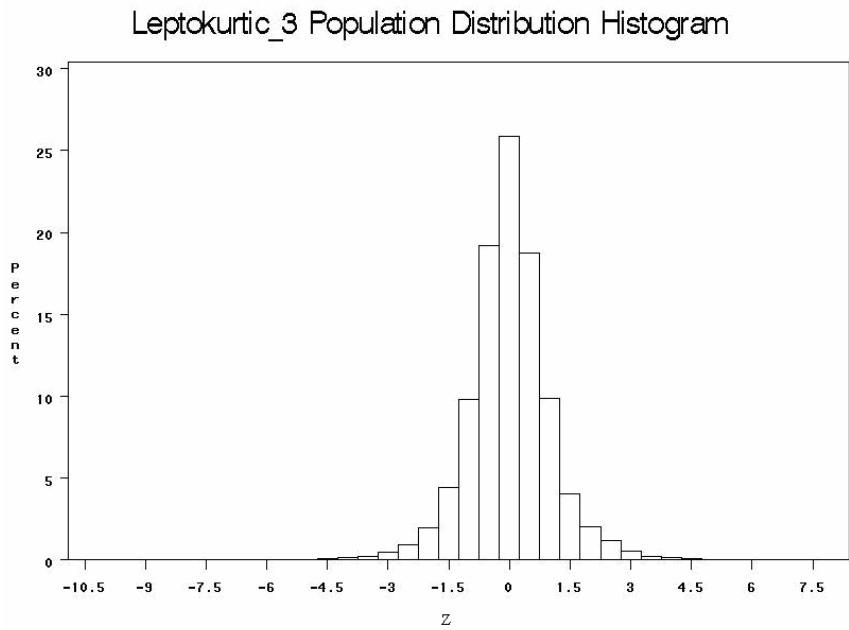


Figure 59: Histogram; Uniform-Like Population Distribution (Skewness = 0.00, Kurtosis = -1.20)

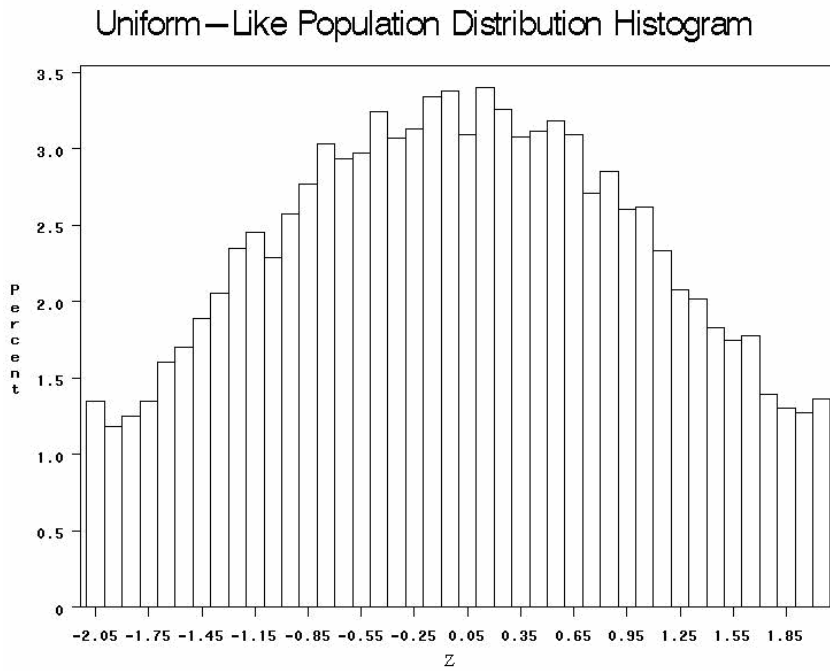


Figure 60: Histogram; Logistic-Like Population Distribution (Skewness = 0.00, Kurtosis = 1.30)

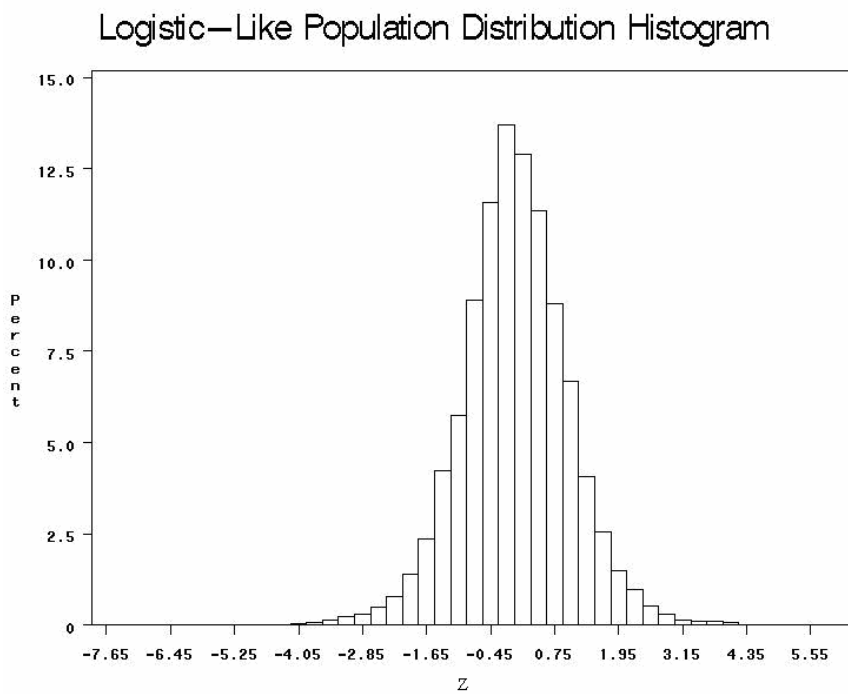
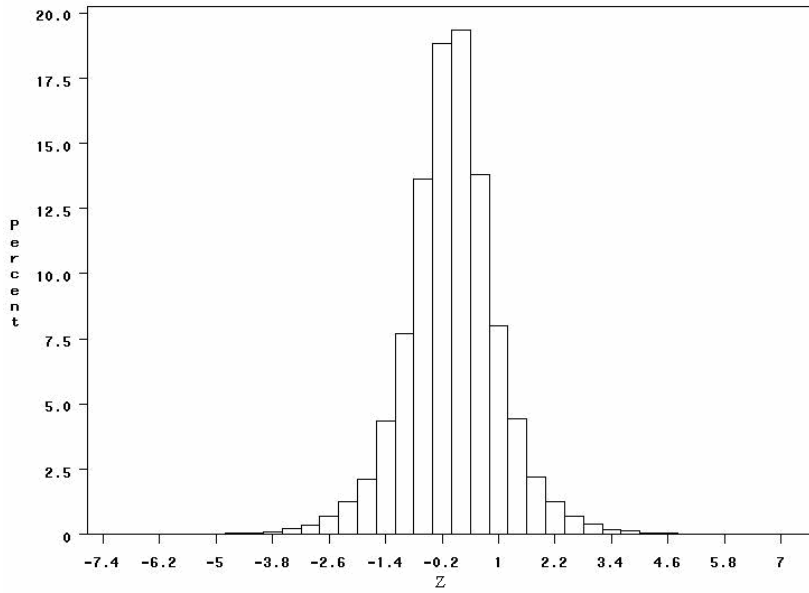


Figure 61: Histogram; Double Exponential-Like Population Distribution (Skewness = 0.00, Kurtosis = 3.00)

Double Exponential-Like Population Distribution Histogram



APPENDIX VI: Tables of Findings

Table 7: Power of Normal Distributions When Sample Sizes Differ and SD Ratio $\neq 1$
 $(\alpha = .05)$

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.141	.068	(16, 4)	4	.007	.005
	3	.132	.055		3	.010	.005
	2	.111	.039		2	.015	.006
	$\frac{1}{2}$.015	.006		$\frac{1}{2}$.113	.037
	$\frac{1}{3}$.008	.005		$\frac{1}{3}$.133	.055
	$\frac{1}{4}$.006	.005		$\frac{1}{4}$.145	.070
(10, 20)	4	.114	.260	(20, 10)	4	.033	.170
	3	.107	.185		3	.031	.103
	2	.088	.089		2	.032	.047
	$\frac{1}{2}$.033	.049		$\frac{1}{2}$.084	.088
	$\frac{1}{3}$.032	.105		$\frac{1}{3}$.102	.178
	$\frac{1}{4}$.033	.171		$\frac{1}{4}$.117	.261

Table 7 CONT.: Power of Normal Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.136	.436	(30, 10)	4	.016	.222
	3	.124	.297		3	.017	.126
	2	.100	.137		2	.021	.057
	$\frac{1}{2}$.022	.056		$\frac{1}{2}$.100	.139
	$\frac{1}{3}$.017	.130		$\frac{1}{3}$.127	.297
	$\frac{1}{4}$.015	.214		$\frac{1}{4}$.139	.439
(50, 100)	4	.119	.999	(100, 50)	4	.034	.999
	3	.105	.973		3	.030	.981
	2	.084	.570		2	.031	.525
	$\frac{1}{2}$.033	.519		$\frac{1}{2}$.088	.572
	$\frac{1}{3}$.032	.982		$\frac{1}{3}$.108	.971
	$\frac{1}{4}$.034	.999		$\frac{1}{4}$.124	.999

Table 8: Power of Platykurtic Distribution When Sample Sizes Differ and SD Ratio $\neq 1$
 $(\alpha = .05)$

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.141	.066	(16, 4)	4	.008	.006
	3	.138	.059		3	.009	.005
	2	.112	.037		2	.015	.006
	$\frac{1}{2}$.014	.007		$\frac{1}{2}$.116	.038
	$\frac{1}{3}$.009	.006		$\frac{1}{3}$.140	.056
	$\frac{1}{4}$.007	.005		$\frac{1}{4}$.141	.070
(10, 20)	4	.121	.281	(20, 10)	4	.032	.177
	3	.111	.189		3	.032	.110
	2	.085	.091		2	.032	.048
	$\frac{1}{2}$.031	.051		$\frac{1}{2}$.089	.095
	$\frac{1}{3}$.031	.107		$\frac{1}{3}$.107	.191
	$\frac{1}{4}$.033	.182		$\frac{1}{4}$.118	.272

Table 8 CONT.: Power of Platykurtic Distribution When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.143	.480	(30, 10)	4	.016	.234
	3	.128	.323		3	.016	.137
	2	.105	.151		2	.019	.062
	$\frac{1}{2}$.021	.058		$\frac{1}{2}$.101	.143
	$\frac{1}{3}$.018	.136		$\frac{1}{3}$.131	.328
	$\frac{1}{4}$.016	.234		$\frac{1}{4}$.142	.483
(50, 100)	4	.126	.999	(100, 50)	4	.035	1
	3	.109	.985		3	.031	.993
	2	.087	.623		2	.032	.581
	$\frac{1}{2}$.033	.586		$\frac{1}{2}$.089	.627
	$\frac{1}{3}$.034	.993		$\frac{1}{3}$.112	.986
	$\frac{1}{4}$.035	.999		$\frac{1}{4}$.035	1

Table 9: Power of Normal Platykurtic Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.144	.070	(16, 4)	4	.007	.006
	3	.143	.059		3	.008	.005
	2	.121	.041		2	.012	.006
	$\frac{1}{2}$.012	.006		$\frac{1}{2}$.124	.040
	$\frac{1}{3}$.008	.006		$\frac{1}{3}$.145	.059
	$\frac{1}{4}$.006	.006		$\frac{1}{4}$.146	.069
(10, 20)	4	.124	.304	(20, 10)	4	.033	.216
	3	.115	.216		3	.030	.127
	2	.092	.106		2	.031	.053
	$\frac{1}{2}$.030	.050		$\frac{1}{2}$.087	.100
	$\frac{1}{3}$.031	.127		$\frac{1}{3}$.115	.213
	$\frac{1}{4}$.032	.213		$\frac{1}{4}$.119	.296

Table 9 CONT.: Power of Normal Platykurtic Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.144	.574	(30, 10)	4	.017	.283
	3	.136	.410		3	.016	.159
	2	.111	.191		2	.021	.063
	$\frac{1}{2}$.019	.066		$\frac{1}{2}$.111	.192
	$\frac{1}{3}$.016	.164		$\frac{1}{3}$.145	.421
	$\frac{1}{4}$.018	.287		$\frac{1}{4}$.145	.572
(50, 100)	4	.125	1	(100, 50)	4	.035	1.000
	3	.115	.999		3	.033	.999
	2	.092	.819		2	.035	.805
	$\frac{1}{2}$.032	.805		$\frac{1}{2}$.097	.824
	$\frac{1}{3}$.033	.999		$\frac{1}{3}$.115	.999
	$\frac{1}{4}$.035	1		$\frac{1}{4}$.130	1.000

Table 10: Power of Leptokurtic 1Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.136	.066	(16, 4)	4	.008	.006
	3	.129	.052		3	.010	.006
	2	.107	.037		2	.017	.006
	$\frac{1}{2}$.017	.006		$\frac{1}{2}$.108	.039
	$\frac{1}{3}$.009	.006		$\frac{1}{3}$.130	.054
	$\frac{1}{4}$.008	.006		$\frac{1}{4}$.142	.069
(10, 20)	4	.110	.247	(20, 10)	4	.032	.156
	3	.102	.167		3	.031	.095
	2	.082	.084		2	.033	.044
	$\frac{1}{2}$.031	.045		$\frac{1}{2}$.084	.085
	$\frac{1}{3}$.031	.045		$\frac{1}{3}$.105	.170
	$\frac{1}{4}$.030	.153		$\frac{1}{4}$.116	.250

Table 10 CONT.: Power of Leptokurtic 1Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.135	.402	(30, 10)	4	.016	.200
	3	.130	.267		3	.017	.120
	2	.097	.129		2	.024	.055
	$\frac{1}{2}$.022	.053		$\frac{1}{2}$.096	.128
	$\frac{1}{3}$.018	.118		$\frac{1}{3}$.122	.261
	$\frac{1}{4}$.015	.200		$\frac{1}{4}$.135	.399
(50, 100)	4	.115	.997	(100, 50)	4	.033	.999
	3	.106	.953		3	.035	.960
	2	.081	.508		2	.034	.467
	$\frac{1}{2}$.033	.468		$\frac{1}{2}$.085	.520
	$\frac{1}{3}$.034	.959		$\frac{1}{3}$.103	.949
	$\frac{1}{4}$.033	.999		$\frac{1}{4}$.119	.998

Table 11: Power of Leptokurtic 2 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.134	.063	(16, 4)	4	.008	.005
	3	.128	.054		3	.009	.006
	2	.103	.036		2	.019	.006
	$\frac{1}{2}$.018	.007		$\frac{1}{2}$.102	.036
	$\frac{1}{3}$.010	.006		$\frac{1}{3}$.125	.051
	$\frac{1}{4}$.008	.005		$\frac{1}{4}$.138	.067
(10, 20)	4	.114	.233	(20, 10)	4	.030	.148
	3	.100	.162		3	.030	.090
	2	.077	.081		2	.034	.046
	$\frac{1}{2}$.031	.043		$\frac{1}{2}$.082	.085
	$\frac{1}{3}$.031	.090		$\frac{1}{3}$.102	.165
	$\frac{1}{4}$.034	.146		$\frac{1}{4}$.110	.233

Table 11 CONT.: Power of Leptokurtic 2 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.133	.373	(30, 10)	4	.016	.190
	3	.120	.252		3	.018	.112
	2	.096	.126		2	.024	.056
	$\frac{1}{2}$.024	.055		$\frac{1}{2}$.093	.124
	$\frac{1}{3}$.018	.112		$\frac{1}{3}$.120	.249
	$\frac{1}{4}$.016	.192		$\frac{1}{4}$.134	.371
(50, 100)	4	.117	.996	(100, 50)	4	.033	.998
	3	.104	.934		3	.033	.934
	2	.083	.476		2	.034	.422
	$\frac{1}{2}$.033	.425		$\frac{1}{2}$.084	.479
	$\frac{1}{3}$.031	.939		$\frac{1}{3}$.111	.930
	$\frac{1}{4}$.034	.998		$\frac{1}{4}$.116	.995

Table 12: Power of Leptokurtic 3 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.133	.065	(16, 4)	4	.008	.005
	3	.126	.052		3	.010	.005
	2	.102	.035		2	.016	.006
	$\frac{1}{2}$.020	.007		$\frac{1}{2}$.103	.037
	$\frac{1}{3}$.011	.006		$\frac{1}{3}$.125	.053
	$\frac{1}{4}$.008	.005		$\frac{1}{4}$.135	.063
(10, 20)	4	.109	.223	(20, 10)	4	.033	.141
	3	.100	.151		3	.029	.085
	2	.078	.079		2	.031	.043
	$\frac{1}{2}$.034	.044		$\frac{1}{2}$.077	.079
	$\frac{1}{3}$.029	.087		$\frac{1}{3}$.095	.149
	$\frac{1}{4}$.032	.135		$\frac{1}{4}$.111	.219

Table 12 CONT.: Power of Leptokurtic 3 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.133	.349	(30, 10)	4	.019	.176
	3	.118	.231		3	.017	.102
	2	.090	.119		2	.024	.052
	$\frac{1}{2}$.025	.054		$\frac{1}{2}$.093	.118
	$\frac{1}{3}$.018	.108		$\frac{1}{3}$.117	.229
	$\frac{1}{4}$.015	.174		$\frac{1}{4}$.133	.344
(50, 100)	4	.111	.991	(100, 50)	4	.033	.995
	3	.106	.906		3	.032	.902
	2	.082	.435		2	.034	.380
	$\frac{1}{2}$.035	.387		$\frac{1}{2}$.080	.442
	$\frac{1}{3}$.032	.897		$\frac{1}{3}$.101	.898
	$\frac{1}{4}$.031	.994		$\frac{1}{4}$.114	.992

Table 13: Power of Skewed and Leptokurtic 1 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.142	.070	(16, 4)	4	.010	.007
	3	.132	.056		3	.012	.006
	2	.113	.039		2	.018	.007
	$\frac{1}{2}$.018	.006		$\frac{1}{2}$.107	.038
	$\frac{1}{3}$.013	.007		$\frac{1}{3}$.136	.060
	$\frac{1}{4}$.010	.007		$\frac{1}{4}$.143	.070
(10, 20)	4	.126	.254	(20, 10)	4	.042	.164
	3	.109	.172		3	.039	.103
	2	.082	.085		2	.041	.051
	$\frac{1}{2}$.036	.048		$\frac{1}{2}$.083	.089
	$\frac{1}{3}$.040	.107		$\frac{1}{3}$.112	.175
	$\frac{1}{4}$.042	.167		$\frac{1}{4}$.119	.247

Table 13 CONT.: Power of Skewed and Leptokurtic 1 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.149	.386	(30, 10)	4	.024	.219
	3	.131	.262		3	.025	.129
	2	.101	.129		2	.027	.061
	$\frac{1}{2}$.027	.061		$\frac{1}{2}$.101	.130
	$\frac{1}{3}$.026	.132		$\frac{1}{3}$.130	.260
	$\frac{1}{4}$.028	.218		$\frac{1}{4}$.151	.392
(50, 100)	4	.181	.996	(100, 50)	4	.092	.998
	3	.152	.936		3	.080	.951
	2	.109	.505		2	.057	.470
	$\frac{1}{2}$.058	.470		$\frac{1}{2}$.109	.506
	$\frac{1}{3}$.073	.944		$\frac{1}{3}$.151	.939
	$\frac{1}{4}$.092	.998		$\frac{1}{4}$.176	.996

Table 14: Power of Skewed and Leptokurtic 2 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.152	.079	(16, 4)	4	.016	.011
	3	.141	.060		3	.184	.008
	2	.113	.041		2	.020	.007
	$\frac{1}{2}$.021	.008		$\frac{1}{2}$.111	.040
	$\frac{1}{3}$.018	.012		$\frac{1}{3}$.135	.075
	$\frac{1}{4}$.016	.011		$\frac{1}{4}$.155	.078
(10, 20)	4	.150	.302	(20, 10)	4	.065	.228
	3	.134	.214		3	.058	.143
	2	.097	.099		2	.047	.062
	$\frac{1}{2}$.047	.059		$\frac{1}{2}$.101	.105
	$\frac{1}{3}$.064	.277		$\frac{1}{3}$.152	.301
	$\frac{1}{4}$.063	.232		$\frac{1}{4}$.145	.300

Table 14 CONT.: Power of Skewed and Leptokurtic 2 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.175	.453	(30, 10)	4	.049	.306
	3	.154	.318		3	.043	.192
	2	.117	.154		2	.036	.078
	$\frac{1}{2}$.039	.077		$\frac{1}{2}$.116	.157
	$\frac{1}{3}$.048	.317		$\frac{1}{3}$.179	.458
	$\frac{1}{4}$.047	.310		$\frac{1}{4}$.177	.459
(50, 100)	4	.300	.999	(100, 50)	4	.234	1
	3	.260	.977		3	.190	.985
	2	.166	.632		2	.115	.631
	$\frac{1}{2}$.110	.625		$\frac{1}{2}$.165	.631
	$\frac{1}{3}$.232	1		$\frac{1}{3}$.302	.999
	$\frac{1}{4}$.236	1		$\frac{1}{4}$.300	.999

Table 15: Power of Uniform-Like Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.146	.070	(16, 4)	4	.007	.006
	3	.137	.055		3	.008	.005
	2	.120	.038		2	.013	.006
	$\frac{1}{2}$.012	.006		$\frac{1}{2}$.120	.041
	$\frac{1}{3}$.008	.005		$\frac{1}{3}$.136	.057
	$\frac{1}{4}$.007	.006		$\frac{1}{4}$.144	.069
(10, 20)	4	.124	.296	(20, 10)	4	.033	.199
	3	.113	.206		3	.032	.120
	2	.090	.099		2	.032	.052
	$\frac{1}{2}$.031	.052		$\frac{1}{2}$.088	.097
	$\frac{1}{3}$.030	.116		$\frac{1}{3}$.117	.209
	$\frac{1}{4}$.031	.198		$\frac{1}{4}$.120	.289

Table 15 CONT.: Power of Uniform-Like Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.146	.540	(30, 10)	4	.015	.265
	3	.134	.383		3	.016	.149
	2	.109	.175		2	.019	.059
	$\frac{1}{2}$.020	.059		$\frac{1}{2}$.106	.170
	$\frac{1}{3}$.016	.149		$\frac{1}{3}$.132	.376
	$\frac{1}{4}$.017	.272		$\frac{1}{4}$.141	.544
(50, 100)	4	.126	1	(100, 50)	4	.035	1
	3	.117	.996		3	.032	.999
	2	.090	.731		2	.032	.703
	$\frac{1}{2}$.031	.701		$\frac{1}{2}$.093	.734
	$\frac{1}{3}$.033	.999		$\frac{1}{3}$.111	.997
	$\frac{1}{4}$.033	1		$\frac{1}{4}$.126	1

Table 16: Power of Logistic-Like Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.141	.071	(16, 4)	4	.008	.006
	3	.130	.054		3	.009	.005
	2	.104	.037		2	.017	.007
	$\frac{1}{2}$.017	.007		$\frac{1}{2}$.104	.036
	$\frac{1}{3}$.010	.006		$\frac{1}{3}$.132	.057
	$\frac{1}{4}$.006	.005		$\frac{1}{4}$.133	.065
(10, 20)	4	.117	.247	(20, 10)	4	.033	.155
	3	.103	.171		3	.034	.093
	2	.083	.084		2	.033	.047
	$\frac{1}{2}$.031	.044		$\frac{1}{2}$.084	.086
	$\frac{1}{3}$.031	.096		$\frac{1}{3}$.102	.164
	$\frac{1}{4}$.031	.155		$\frac{1}{4}$.116	.248

Table 16 CONT.: Power of Logistic-Like Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.134	.390	(30, 10)	4	.015	.193
	3	.124	.266		3	.017	.118
	2	.100	.131		2	.023	.055
	$\frac{1}{2}$.024	.056		$\frac{1}{2}$.097	.125
	$\frac{1}{3}$.019	.119		$\frac{1}{3}$.124	.266
	$\frac{1}{4}$.016	.201		$\frac{1}{4}$.137	.397
(50, 100)	4	.122	.997	(100, 50)	4	.032	.999
	3	.104	.944		3	.031	.951
	2	.083	.502		2	.034	.451
	$\frac{1}{2}$.032	.455		$\frac{1}{2}$.085	.507
	$\frac{1}{3}$.032	.951		$\frac{1}{3}$.104	.945
	$\frac{1}{4}$.035	.999		$\frac{1}{4}$.117	.997

Table 17: Power of Double Exponential-Like Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.124	.066	(16, 4)	4	.010	.006
	3	.125	.052		3	.011	.006
	2	.102	.036		2	.017	.006
	$\frac{1}{2}$.019	.007		$\frac{1}{2}$.100	.035
	$\frac{1}{3}$.012	.006		$\frac{1}{3}$.122	.053
	$\frac{1}{4}$.009	.005		$\frac{1}{4}$.136	.063
(10, 20)	4	.109	.226	(20, 10)	4	.032	.139
	3	.099	.155		3	.033	.084
	2	.079	.080		2	.033	.046
	$\frac{1}{2}$.034	.045		$\frac{1}{2}$.078	.080
	$\frac{1}{3}$.030	.087		$\frac{1}{3}$.100	.156
	$\frac{1}{4}$.030	.139		$\frac{1}{4}$.116	.228

Table 17 CONT: Power of Double Exponential -Like Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.130	.351	(30, 10)	4	.018	.177
	3	.120	.244		3	.017	.108
	2	.094	.119		2	.025	.054
	$\frac{1}{2}$.023	.053		$\frac{1}{2}$.097	.122
	$\frac{1}{3}$.019	.107		$\frac{1}{3}$.119	.242
	$\frac{1}{4}$.016	.186		$\frac{1}{4}$.132	.357
(50, 100)	4	.113	.992	(100, 50)	4	.033	.996
	3	.102	.912		3	.033	.913
	2	.083	.452		2	.033	.404
	$\frac{1}{2}$.034	.403		$\frac{1}{2}$.082	.449
	$\frac{1}{3}$.034	.913		$\frac{1}{3}$.101	.911
	$\frac{1}{4}$.034	.996		$\frac{1}{4}$.116	.992

Table 18: Power of Skewed-Leptokurtic Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.197	.119	(16, 4)	4	.043	.041
	3	.187	.095		3	.040	.026
	2	.174	.069		2	.034	.014
	$\frac{1}{2}$.036	.015		$\frac{1}{2}$.165	.065
	$\frac{1}{3}$.040	.028		$\frac{1}{3}$.186	.095
	$\frac{1}{4}$.044	.041		$\frac{1}{4}$.192	.113
(10, 20)	4	.262	.552	(20, 10)	4	.176	.627
	3	.248	.472		3	.161	.517
	2	.199	.305		2	.125	.293
	$\frac{1}{2}$.123	.293		$\frac{1}{2}$.195	.300
	$\frac{1}{3}$.160	.514		$\frac{1}{3}$.248	.466
	$\frac{1}{4}$.175	.627		$\frac{1}{4}$.259	.553

Table 18 CONT.: Power of Skewed-Leptokurtic Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.295	.800	(30, 10)	4	.168	.801
	3	.281	.719		3	.143	.692
	2	.231	.534		2	.113	.410
	$\frac{1}{2}$.115	.409		$\frac{1}{2}$.229	.532
	$\frac{1}{3}$.149	.695		$\frac{1}{3}$.278	.719
	$\frac{1}{4}$.164	.799		$\frac{1}{4}$.300	.798
(50, 100)	4	.675	1	(100, 50)	4	.733	1
	3	.647	1		3	.689	1
	2	.544	1		2	.556	1
	$\frac{1}{2}$.555	1		$\frac{1}{2}$.548	1
	$\frac{1}{3}$.688	1		$\frac{1}{3}$.641	1
	$\frac{1}{4}$.730	1		$\frac{1}{4}$.675	1

Table 19: Power of Skewed Distributions When Sample Sizes Differ and SD Ratio $\neq 1$
 $(\alpha = .05)$

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.152	.076	(16, 4)	4	.013	.010
	3	.151	.064		3	.012	.009
	2	.126	.042		2	.016	.007
	$\frac{1}{2}$.015	.007		$\frac{1}{2}$.125	.041
	$\frac{1}{3}$.014	.007		$\frac{1}{3}$.151	.065
	$\frac{1}{4}$.014	.011		$\frac{1}{4}$.155	.079
(10, 20)	4	.149	.351	(20, 10)	4	.056	.304
	3	.134	.256		3	.055	.201
	2	.113	.133		2	.045	.078
	$\frac{1}{2}$.045	.078		$\frac{1}{2}$.111	.131
	$\frac{1}{3}$.052	.196		$\frac{1}{3}$.139	.262
	$\frac{1}{4}$.058	.305		$\frac{1}{4}$.153	.358

Table 19 CONT.: Power of Skewed Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.172	.615	(30, 10)	4	.040	.421
	3	.162	.471		3	.034	.270
	2	.134	.258		2	.033	.097
	$\frac{1}{2}$.032	.096		$\frac{1}{2}$.132	.256
	$\frac{1}{3}$.035	.276		$\frac{1}{3}$.161	.475
	$\frac{1}{4}$.038	.424		$\frac{1}{4}$.178	.618
(50, 100)	4	.266	1	(100, 50)	4	.179	1
	3	.245	1		3	.166	1
	2	.195	.945		2	.122	.979
	$\frac{1}{2}$.120	.980		$\frac{1}{2}$.186	.946
	$\frac{1}{3}$.164	1		$\frac{1}{3}$.244	1
	$\frac{1}{4}$.185	1		$\frac{1}{4}$.265	1

Table 20: Power of Skewed and Platykurtic 1 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.145	.072	(16, 4)	4	.009	.008
	3	.142	.059		3	.010	.006
	2	.123	.041		2	.014	.006
	$\frac{1}{2}$.014	.006		$\frac{1}{2}$.122	.039
	$\frac{1}{3}$.010	.006		$\frac{1}{3}$.145	.061
	$\frac{1}{4}$.010	.008		$\frac{1}{4}$.147	.071
(10, 20)	4	.129	.324	(20, 10)	4	.043	.247
	3	.120	.231		3	.039	.157
	2	.102	.117		2	.037	.062
	$\frac{1}{2}$.037	.061		$\frac{1}{2}$.098	.115
	$\frac{1}{3}$.040	.161		$\frac{1}{3}$.123	.232
	$\frac{1}{4}$.044	.251		$\frac{1}{4}$.137	.327

Table 20 CONT.: Power of Skewed and Platykurtic 1 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.163	.580	(30, 10)	4	.027	.351
	3	.153	.438		3	.024	.204
	2	.120	.211		2	.027	.079
	$\frac{1}{2}$.024	.076		$\frac{1}{2}$.113	.205
	$\frac{1}{3}$.025	.206		$\frac{1}{3}$.148	.430
	$\frac{1}{4}$.028	.346		$\frac{1}{4}$.159	.582
(50, 100)	4	.195	1	(100, 50)	4	.098	1
	3	.175	.999		3	.089	1
	2	.134	.886		2	.072	.917
	$\frac{1}{2}$.070	.918		$\frac{1}{2}$.138	.887
	$\frac{1}{3}$.089	1		$\frac{1}{3}$.177	.999
	$\frac{1}{4}$.098	1		$\frac{1}{4}$.192	1

Table 21: Power of Skewed and Platykurtic 2 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(4, 16)	4	.146	.074	(16, 4)	4	.007	.007
	3	.136	.054		3	.008	.005
	2	.124	.039		2	.012	.006
	$\frac{1}{2}$.013	.006		$\frac{1}{2}$.123	.039
	$\frac{1}{3}$.010	.006		$\frac{1}{3}$.146	.058
	$\frac{1}{4}$.009	.007		$\frac{1}{4}$.151	.075
(10, 20)	4	.134	.328	(20, 10)	4	.039	.243
	3	.118	.233		3	.037	.155
	2	.096	.114		2	.033	.063
	$\frac{1}{2}$.035	.065		$\frac{1}{2}$.095	.114
	$\frac{1}{3}$.034	.152		$\frac{1}{3}$.117	.234
	$\frac{1}{4}$.037	.244		$\frac{1}{4}$.124	.321

Table 21 CONT.: Power of Skewed and Platykurtic 2 Distributions When Sample Sizes Differ and SD Ratio $\neq 1$ ($\alpha = .05$)

Sample Size	SD Ratio	Power		Sample Size	SD Ratio	Power	
		MW	KS-2			MW	KS-2
(10, 30)	4	.155	.622	(30, 10)	4	.020	.340
	3	.145	.461		3	.021	.200
	2	.118	.226		2	.020	.071
	$\frac{1}{2}$.022	.070		$\frac{1}{2}$.118	.226
	$\frac{1}{3}$.018	.196		$\frac{1}{3}$.139	.458
	$\frac{1}{4}$.019	.337		$\frac{1}{4}$.151	.615
(50, 100)	4	.154	1	(100, 50)	4	.057	1
	3	.139	1		3	.051	1
	2	.111	.917		2	.048	.942
	$\frac{1}{2}$.044	.941		$\frac{1}{2}$.110	.913
	$\frac{1}{3}$.051	1		$\frac{1}{3}$.136	1
	$\frac{1}{4}$.056	1		$\frac{1}{4}$.150	1

Table 22: Power of Normal Populations with ONLY SD Ratios Are Different
(SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWE7R		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.070	.082	$\frac{1}{2}$	(8, 8)	.061	.035
	(16, 16)	.074	.358		(16, 16)	.053	.098
	(25, 25)	.077	.643		(25, 25)	.058	.152
	(50, 50)	.075	.989		(50, 50)	.060	.383
3	(8, 8)	.066	.056	$\frac{1}{3}$	(8, 8)	.069	.060
	(16, 16)	.064	.224		(16, 16)	.063	.223
	(25, 25)	.070	.409		(25, 25)	.070	.414
	(50, 50)	.067	.880		(50, 50)	.069	.877
2	(8, 8)	.056	.033	$\frac{1}{4}$	(8, 8)	.071	.083
	(16, 16)	.052	.093		(16, 16)	.072	.344
	(25, 25)	.058	.153		(25, 25)	.072	.639
	(50, 50)	.058	.375		(50, 50)	.074	.987

Table 23: Power of Platykurtic Populations with ONLY SD Ratios Are Different
(SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.069	.085	$\frac{1}{2}$	(8, 8)	.057	.033
	(16, 16)	.072	.376		(16, 16)	.055	.098
	(25, 25)	.078	.685		(25, 25)	.063	.171
	(50, 50)	.080	.994		(50, 50)	.059	.418
3	(8, 8)	.069	.061	$\frac{1}{3}$	(8, 8)	.066	.061
	(16, 16)	.066	.236		(16, 16)	.066	.236
	(25, 25)	.067	.444		(25, 25)	.071	.451
	(50, 50)	.067	.915		(50, 50)	.068	.918
2	(8, 8)	.060	.035	$\frac{1}{4}$	(8, 8)	.069	.085
	(16, 16)	.056	.100		(16, 16)	.071	.375
	(25, 25)	.060	.167		(25, 25)	.075	.695
	(50, 50)	.059	.425		(50, 50)	.076	.994

Table 24: Power of Normal Platykurtic Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.076	.097	$\frac{1}{2}$	(8, 8)	.059	.034
	(16, 16)	.072	.444		(16, 16)	.057	.117
	(25, 25)	.075	.794		(25, 25)	.057	.203
	(50, 50)	.083	1.000		(50, 50)	.063	.586
3	(8, 8)	.072	.070	$\frac{1}{3}$	(8, 8)	.066	.068
	(16, 16)	.080	.292		(16, 16)	.066	.288
	(25, 25)	.075	.569		(25, 25)	.070	.569
	(50, 50)	.073	.981		(50, 50)	.071	.982
2	(8, 8)	.061	.038	$\frac{1}{4}$	(8, 8)	.076	.100
	(16, 16)	.057	.121		(16, 16)	.073	.444
	(25, 25)	.060	.207		(25, 25)	.079	.800
	(50, 50)	.064	.583		(50, 50)	.081	.992

Table 25: Power of Leptokurtic_1 Populations with ONLY SD Ratios Are Different
(SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.070	.078	$\frac{1}{2}$	(8, 8)	.057	.032
	(16, 16)	.070	.318		(16, 16)	.057	.093
	(25, 25)	.074	.588		(25, 25)	.059	.137
	(50, 50)	.072	.974		(50, 50)	.060	.345
3	(8, 8)	.063	.053	$\frac{1}{3}$	(8, 8)	.067	.055
	(16, 16)	.062	.203		(16, 16)	.064	.203
	(25, 25)	.061	.371		(25, 25)	.065	.359
	(50, 50)	.067	.828		(50, 50)	.065	.824
2	(8, 8)	.058	.034	$\frac{1}{4}$	(8, 8)	.073	.079
	(16, 16)	.052	.089		(16, 16)	.070	.316
	(25, 25)	.057	.142		(25, 25)	.074	.588
	(50, 50)	.057	.345		(50, 50)	.071	.974

Table 26: Power of Leptokurtic_2 Populations with ONLY SD Ratios Are Different
(SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.071	.075	$\frac{1}{2}$	(8, 8)	.055	.029
	(16, 16)	.069	.297		(16, 16)	.053	.087
	(25, 25)	.073	.551		(25, 25)	.056	.133
	(50, 50)	.073	.963		(50, 50)	.057	.314
3	(8, 8)	.064	.052	$\frac{1}{3}$	(8, 8)	.062	.051
	(16, 16)	.060	.185		(16, 16)	.059	.186
	(25, 25)	.068	.345		(25, 25)	.069	.347
	(50, 50)	.068	.791		(50, 50)	.066	.789
2	(8, 8)	.058	.032	$\frac{1}{4}$	(8, 8)	.070	.076
	(16, 16)	.055	.087		(16, 16)	.069	.300
	(25, 25)	.056	.133		(25, 25)	.072	.548
	(50, 50)	.057	.315		(50, 50)	.079	.962

Table 27: Power of Leptokurtic_3 Populations with ONLY SD Ratios Are Different
(SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.071	.072	$\frac{1}{2}$	(8, 8)	.057	.031
	(16, 16)	.068	.280		(16, 16)	.052	.083
	(25, 25)	.068	.504		(25, 25)	.059	.126
	(50, 50)	.070	.944		(50, 50)	.056	.290
3	(8, 8)	.067	.051	$\frac{1}{3}$	(8, 8)	.065	.051
	(16, 16)	.062	.179		(16, 16)	.062	.182
	(25, 25)	.069	.318		(25, 25)	.063	.306
	(50, 50)	.064	.742		(50, 50)	.066	.737
2	(8, 8)	.054	.031	$\frac{1}{4}$	(8, 8)	.065	.066
	(16, 16)	.055	.084		(16, 16)	.070	.279
	(25, 25)	.056	.122		(25, 25)	.071	.506
	(50, 50)	.055	.288		(50, 50)	.073	.943

Table 28: Power of Skewed Populations with ONLY SD Ratios Are Different
 (SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.091	.122	$\frac{1}{2}$	(8, 8)	.070	.046
	(16, 16)	.109	.535		(16, 16)	.081	.173
	(25, 25)	.129	.858		(25, 25)	.094	.338
	(50, 50)	.191	1.000		(50, 50)	.130	.823
3	(8, 8)	.081	.089	$\frac{1}{3}$	(8, 8)	.081	.088
	(16, 16)	.085	.382		(16, 16)	.096	.385
	(25, 25)	.122	.697		(25, 25)	.121	.697
	(50, 50)	.177	.994		(50, 50)	.174	.994
2	(8, 8)	.069	.047	$\frac{1}{4}$	(8, 8)	.089	.124
	(16, 16)	.079	.176		(16, 16)	.106	.530
	(25, 25)	.097	.337		(25, 25)	.133	.861
	(50, 50)	.135	.823		(50, 50)	.197	1.000

Table 29: Power of Skewed and Platykurtic_1 Populations with ONLY SD Ratios
Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.090	.123	$\frac{1}{2}$	(8, 8)	.071	.046
	(16, 16)	.108	.532		(16, 16)	.078	.171
	(25, 25)	.140	.895		(25, 25)	.095	.331
	(50, 50)	.195	1.000		(50, 50)	.131	.820
3	(8, 8)	.083	.087	$\frac{1}{3}$	(8, 8)	.084	.087
	(16, 16)	.100	.382		(16, 16)	.095	.381
	(25, 25)	.123	.689		(25, 25)	.123	.691
	(50, 50)	.175	.994		(50, 50)	.176	.995
2	(8, 8)	.070	.045	$\frac{1}{4}$	(8, 8)	.090	.127
	(16, 16)	.080	.175		(16, 16)	.106	.526
	(25, 25)	.096	.333		(25, 25)	.135	.864
	(50, 50)	.133	.823		(50, 50)	.188	1.000

Table 30: Power of Skewed and Platykurtic_2 Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.078	.107	$\frac{1}{2}$	(8, 8)	.065	.041
	(16, 16)	.081	.490		(16, 16)	.058	.135
	(25, 25)	.089	.846		(25, 25)	.066	.267
	(50, 50)	.100	1.000		(50, 50)	.074	.737
3	(8, 8)	.071	.071	$\frac{1}{3}$	(8, 8)	.072	.079
	(16, 16)	.071	.336		(16, 16)	.074	.336
	(25, 25)	.078	.644		(25, 25)	.084	.646
	(50, 50)	.088	.994		(50, 50)	.088	.994
2	(8, 8)	.064	.041	$\frac{1}{4}$	(8, 8)	.077	.079
	(16, 16)	.063	.140		(16, 16)	.079	.498
	(25, 25)	.066	.265		(25, 25)	.087	.851
	(50, 50)	.077	.742		(50, 50)	.098	1.000

Table 31: Power of Skewed and Leptokurtic_1 Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.071	.077	$\frac{1}{2}$	(8, 8)	.059	.031
	(16, 16)	.085	.323		(16, 16)	.059	.097
	(25, 25)	.095	.577		(25, 25)	.065	.145
	(50, 50)	.117	.970		(50, 50)	.077	.349
3	(8, 8)	.070	.059	$\frac{1}{3}$	(8, 8)	.069	.058
	(16, 16)	.072	.205		(16, 16)	.073	.209
	(25, 25)	.081	.367		(25, 25)	.084	.372
	(50, 50)	.102	.809		(50, 50)	.103	.811
2	(8, 8)	.057	.033	$\frac{1}{4}$	(8, 8)	.075	.077
	(16, 16)	.057	.091		(16, 16)	.084	.322
	(25, 25)	.067	.147		(25, 25)	.098	.578
	(50, 50)	.073	.337		(50, 50)	.122	.967

Table 32: Power of Skewed and Leptokurtic_2 Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.090	.102	$\frac{1}{2}$	(8, 8)	.068	.039
	(16, 16)	.115	.404		(16, 16)	.073	.119
	(25, 25)	.144	.680		(25, 25)	.088	.190
	(50, 50)	.221	.988		(50, 50)	.118	.454
3	(8, 8)	.084	.073	$\frac{1}{3}$	(8, 8)	.082	.072
	(16, 16)	.097	.267		(16, 16)	.096	.268
	(25, 25)	.125	.477		(25, 25)	.123	.474
	(50, 50)	.186	.904		(50, 50)	.191	.902
2	(8, 8)	.065	.037	$\frac{1}{4}$	(8, 8)	.089	.100
	(16, 16)	.076	.118		(16, 16)	.114	.406
	(25, 25)	.087	.192		(25, 25)	.144	.682
	(50, 50)	.122	.461		(50, 50)	.222	.988

Table 33: Power of Skewed-Leptokurtic Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.153	.243	$\frac{1}{2}$	(8, 8)	.117	.112
	(16, 16)	.252	.795		(16, 16)	.185	.485
	(25, 25)	.357	.976		(25, 25)	.261	.787
	(50, 50)	.580	1.000		(50, 50)	.446	.997
3	(8, 8)	.146	.193	$\frac{1}{3}$	(8, 8)	.143	.196
	(16, 16)	.228	.693		(16, 16)	.232	.693
	(25, 25)	.326	.945		(25, 25)	.331	.942
	(50, 50)	.540	1.000		(50, 50)	.549	1.000
2	(8, 8)	.115	.114	$\frac{1}{4}$	(8, 8)	.158	.246
	(16, 16)	.175	.472		(16, 16)	.251	.787
	(25, 25)	.258	.781		(25, 25)	.357	.978
	(50, 50)	.445	.997		(50, 50)	.575	1.000

Table 34: Power of Uniform-Like Populations with ONLY SD Ratios Are Different
(SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.074	.094	$\frac{1}{2}$	(8, 8)	.061	.036
	(16, 16)	.075	.423		(16, 16)	.054	.108
	(25, 25)	.075	.764		(25, 25)	.060	.188
	(50, 50)	.077	.998		(50, 50)	.064	.510
3	(8, 8)	.070	.066	$\frac{1}{3}$	(8, 8)	.069	.067
	(16, 16)	.068	.275		(16, 16)	.068	.271
	(25, 25)	.070	.525		(25, 25)	.071	.514
	(50, 50)	.070	.963		(50, 50)	.073	.967
2	(8, 8)	.059	.035	$\frac{1}{4}$	(8, 8)	.072	.093
	(16, 16)	.056	.110		(16, 16)	.073	.427
	(25, 25)	.061	.183		(25, 25)	.078	.769
	(50, 50)	.060	.507		(50, 50)	.080	.999

Table 35: Power of Logistic-Like Populations with ONLY SD Ratios Are Different
(SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.071	.079	$\frac{1}{2}$	(8, 8)	.056	.031
	(16, 16)	.071	.312		(16, 16)	.055	.090
	(25, 25)	.071	.574		(25, 25)	.057	.138
	(50, 50)	.075	.970		(50, 50)	.058	.328
3	(8, 8)	.064	.053	$\frac{1}{3}$	(8, 8)	.067	.054
	(16, 16)	.061	.200		(16, 16)	.065	.200
	(25, 25)	.062	.357		(25, 25)	.067	.352
	(50, 50)	.067	.809		(50, 50)	.067	.812
2	(8, 8)	.058	.033	$\frac{1}{4}$	(8, 8)	.067	.071
	(16, 16)	.054	.090		(16, 16)	.072	.313
	(25, 25)	.056	.139		(25, 25)	.073	.574
	(50, 50)	.055	.331		(50, 50)	.074	.972

Table 36: Power of Double Exponential-Like Populations with ONLY SD Ratios Are Different (SD Ratio $\neq 1$ & $\alpha = .05$)

SD RATIO	SAMPLE SIZE	POWER		SD RATIO	SAMPLE SIZE	POWER	
		MW	KS-2			MW	KS-2
4	(8, 8)	.070	.074	$\frac{1}{2}$	(8, 8)	.058	.033
	(16, 16)	.069	.289		(16, 16)	.053	.084
	(25, 25)	.073	.530		(25, 25)	.057	.127
	(50, 50)	.075	.949		(50, 50)	.053	.292
3	(8, 8)	.062	.049	$\frac{1}{3}$	(8, 8)	.064	.052
	(16, 16)	.062	.181		(16, 16)	.062	.184
	(25, 25)	.067	.320		(25, 25)	.066	.323
	(50, 50)	.067	.755		(50, 50)	.067	.760
2	(8, 8)	.056	.030	$\frac{1}{4}$	(8, 8)	.069	.072
	(16, 16)	.054	.085		(16, 16)	.069	.286
	(25, 25)	.054	.129		(25, 25)	.071	.524
	(50, 50)	.055	.293		(50, 50)	.076	.950

VITA

Chin-Huey Lee

Candidate for the Degree of

Doctor of Philosophy

Thesis: A MONTE CARLO STUDY OF TWO NONPARAMETRIC STATISTICS
WITH COMPARISONS OF TYPE I ERROR RATES AND POWER

Major Field: Educational Psychology

Biographical:

Education:

Bachelor of Business Administration from National Chung-Hsing
University Taipei, Taiwan in 1994

Master of Business Administration from Pittsburg State University,
Pittsburg, KS in 1997

Educational Specialist from Pittsburg State University, Pittsburg, KS, in
1999

Completed the requirements for the Doctor of Philosophy degree in
Educational Psychology: Research, Evaluation, Measurement, and
Statistics at Oklahoma State University in May 2007

Experience:

Mathematics teacher, Wha-Wen after School Program, Taipei, Taiwan,
1991-1995; Assistant Accountant, Lin Accounting Firm, Taipei, Taiwan,
1995-1996; Graduate Teaching Assistant, Department of Statistics,
College of Art and Science, Oklahoma State University 2002-2004;
Graduate Teaching Associate, School of Educational Studies, College of
Education, Oklahoma State University 2004-present.

Professional Memberships:

American Educational and Research Association (AERA)

American Evaluation Association (AEA)

American Statistical Association (ASA)

Psychometric Society

Name: Chin-Huey Lee

Date of Degree: May, 2007

Institution: Oklahoma State University

Location: Stillwater, Oklahoma

Title of Study: A MONTE CARLO STUDY OF TWO NONPARAMETRIC
STATISTICS WITH COMPARISONS OF TYPE I ERROR RATES
AND POWER

Pages in Study: 228

Candidate for the Degree of Doctor of Philosophy

Major Field: Educational Psychology

Scope and Method of Study: The Mann-Whitney (MW) test and the Kolmogorov-Smirnov two sample test (KS-2) are nonparametric statistical tests used to detect whether there is a general difference between two samples when the two underlying population distributions are distribution-free. The focus of this study was to examine and compare Type I error rates and statistical power between the MW and the KS-2 tests when the two samples had different population variances or various degrees of kurtosis and skewness. This study also compared Type I error rates and power, if applicable, when the two samples were of different sizes. Simulations in SAS program were conducted to simulate various conditions to examine Type I error rates and statistical powers for these two nonparametric statistical tests. There were 15 population distributions, 12 sets of sample size combinations, and 7 different ratios of standard deviation. Exactly 20,000 replications per condition were executed for a total of 1380 conditions.

Findings and Conclusions: The study revealed that the KS-2 test is smaller than the MW test in comparisons of the type I error rates in unequal sample sets. The MW test had slightly more statistical power than the KS-2 test under the condition of small and equal-sized samples. Moreover, when population variances vary between two samples, the KS-2 test has more statistical power than the MW test. Furthermore, the power of the KS-2 test exceeded the power of the MW test in large sample settings when either one of the following conditions existed: (1) The difference in the Skewness ratios in populations between the two samples was more than 0.5 with the same kurtosis and variance. (2) The difference in the Kurtosis ratios in populations between the two samples was more than 2.0 with the same skewness and variance. Theoretical and practical implications, limitations of the study, are discussed, as well as recommendations for future research.

ADVISER'S APPROVAL: Dr. Janice Miller
