12-1992

A Critical Examination of the Use of Preliminary Tests in Two-Sample Tests of Location

Kimberly Tucker Perry
Western Michigan University

A CRITICAL EXAMINATION OF THE USE OF PRELIMINARY TESTS IN TWO-SAMPLE TESTS OF LOCATION

by

Kimberly Tucker Perry

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Mathematics and Statistics

Western Michigan University
Kalamazoo, Michigan
December 1992

A CRITICAL EXAMINATION OF THE USE OF PRELIMINARY TESTS IN TWO-SAMPLE TESTS OF LOCATION

Kimberly Tucker Perry, Ph.D.

Western Michigan University, 1992

The purpose of this dissertation was to explore the appropriateness of testing the equality of two means using either a *t* test, the Welch test, or the Mann-Whitney-Wilcoxon test for two independent samples based on the results of using two classes of preliminary tests. One class of preliminary tests determines whether the population variances differ, and the other class ascertains if the underlying distributions are symmetric or skewed. The F-ratio test and the Levene test (using the median) were compared as preliminary tests for variance homogeneity; and the D'Agostino S_U and the Triples tests were also compared as preliminary tests of symmetry/asymmetry. The simulation results were used in the formulation of an expert system procedure in which both classes of preliminary tests were incorporated in the test of equality of two means. Depending on the results of the two preliminary tests, the *t* test, Welch test, or Mann-Whitney-Wilcoxon test of means was selected. The performance of the expert system was evaluated relative to Type I error only. No power results were reported.

The performance of an expert system was also evaluated with respect to robustness and the frequency at which the expert system selected the most appropriate means test.

The results reported here indicate that a preliminary test procedure for equality of variance should be used rather than using the Mann-Whitney-Wilcoxon test or t test for cases where nothing is known a priori about variance equality. A second conclusion is that a preliminary test procedure should be run at a significance level ≥ 0.25.

Finally, suggestions for improvements in the overall expert system are given.

Order Number 9306539

A critical examination of the use of preliminary tests in two-sample tests of location

Perry, Kimberly Tucker, Ph.D.

Western Michigan University, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

ACKNOWLEDGEMENTS

I wish to thank my doctoral chair, Dr. Michael Stoline for his counsel and guidance during the entire period of my dissertation research and preparation. Dr. Stoline has unceasingly given his time, interest, and encouragement on innumerable occasions, and I am most grateful to him.

I would also like to thank Ann Edwards for her skillful and steadfast word processing in preparing this manuscript.

Finally, I must extend my unending gratitude to my husband, Ric, for his support, understanding, encouragement, and continual sacrifices made during my graduate study. My family, my husband's family, and my close friends have also been an additional source of encouragement on numerous occasions.

Kimberly Tucker Perry

TABLE OF CONTENTS

CHAPTER

Table of Contents--Continued

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Robustness of the t test, the Welch Test, and the Mann-Whitney-Wilcoxon Test

In practice, the two-sample t test is widely used to test the equality of two means. However, it is well-known that the assumptions of independence (which will not be discussed in this dissertation), variance homogeneity and normality must be met for the two-sample t test to perform well. It has been shown in the literature (Zimmerman and Williams, 1989) that the Type I error rate for testing the equality of two means for normally distributed populations is an increasing function of the ratio $R = \sigma_1/\sigma_2$ when sample sizes are unequal and indirectly paired (when the smaller sample size is associated with the larger variance); and it is a decreasing function of the ratio $R = \sigma_1/\sigma_2$ when direct pairing is employed (when the smaller sample size is associated with the smaller variance). Little dependence of the true Type I error rate on the ratio of σ_1/σ_2 is seen under equitable sample sizes. Only when both the variances and the sample sizes are unbalanced is the observed significance level affected to any measurable degree.

An alternative method for testing the equality of two means for the case of variance heterogeneity is the Welch test, which is known to sustain excellent control over the probability of the Type I error for those normal cases with extreme variance heterogeneity and imbalanced sample sizes (Zimmerman & Williams, 1989). However this method, like the *t* test, is liberal (rejects too often) in the indirect pairing cases for skewed distributions (Gans, 1981; Murphy, 1976).

The Mann-Whitney-Wilcoxon test is a distribution free test that is not based on normality and has been used for testing equal location parameters. The Mann-Whitney-Wilcoxon test is known to be robust (preserves the stated Type I error rate) for the case of asymmetric distributions with variance homogeneity (Snedecor & Cochran, 1967). However for the case of symmetric distributions, Murphy (1976) found the Mann-Whitney-Wilcoxon test to be more stable with respect to Type I error than the *t* test but still unacceptable when the variances are unequal, especially for the cases of indirect pairing. Murphy (1976) states that non-normality is not a serious problem when the distributions are symmetric, but none of the three above tests (the *t* test, the Welch test and the Mann-Whitney-Wilcoxon test) are robust under extreme skewness.

Based on these results for testing the equality of means, we conclude the following:

1. The *t* test is robust when the distribution is symmetric and the variances are equivalent.

2. The Welch test is robust when the distribution is symmetric and the variances are unequal.

3. The Mann-Whitney-Wilcoxon test is robust when the distribution is asymmetric and the variances are equivalent.

4. None of the above three methods are robust when the distribution is asymmetric and the variances are unequal.

Therefore it would be useful to design a system, an expert system, which would use the results from a preliminary test of variance homogeneity and a preliminary test of symmetry/skewness to determine which of the three tests, the *t* test, the Welch test, or the Mann-Whitney-Wilcoxon test, should be used to test the hypothesis H_o: $\mu_1 = \mu_2$.

A great deal of uncertainty surrounds the use of a preliminary test for testing variance homogeneity. This is discussed in Section 1.2. The idea of incorporation of a preliminary test for testing symmetry/skewness has not been investigated for the case of testing for means equality. Selecting methodology for testing the hypothesis of symmetry/skewness is discussed in Section 1.3. Section 1.4 contains a statement of the purposes of this dissertation work based on the results from the literature review cited in Sections 1.2 and 1.3.

1.2 Tests of Variances Used as Preliminary Tests

1.2.1 Introduction

The goal of the preliminary test for variance heterogeneity is to select an appropriate method for testing equality of means. In other words, this preliminary test would advise the researcher when to avoid using mean tests that are sensitive to variance heterogeneity.

The selection of preliminary tests of H_o: $\sigma_1^2 = \sigma_2^2$ is now discussed. The F-ratio test is one of the oldest methods for testing the hypothesis H_o: $\sigma_1^2 = \sigma_2^2$. However, the F-ratio test lacks robustness when the underlying distributions are non-normal (Box, 1953). For this reason some statisticians do not recommend its use as a preliminary test of H_o: $\sigma_1^2 = \sigma_2^2$.

Many other methods for testing variance homogeneity have been developed, namely several by Levene, which are more robust with respect to normality. Section 1.2.2 contains a discussion of four simulation studies which address this question. The four studies are: (1) Brown and Forsythe (1974), (2) Conover, M.E. Johnson, and M.M. Johnson (1981), (3) Loh (1987), and (4) O'Brien (1979). The L_{50}, the Levene test using the median, was found to be robust for the non-normal cases and recommended by Conover et al. (1981). Therefore, the Levene test using the median might be a more robust preliminary test procedure than the F-ratio test.

Some researchers have conducted studies using the F-ratio test as a preliminary test for variance homogeneity. Four such simulation studies are discussed in Section 1.2.3. The four studies are: (1) Gans (1981), (2) Moser, Stevens, and Watts (1989), (3) C.A. Markowski and E.P. Markowski (1991), and (4) Lauer and Han (1974). The preliminary test procedure using the F-ratio test is not recommended in any of the above four simulation studies except in the Lauer and Han (1974) study. Gans (1981) and Moser et al. (1989) used only traditional significance levels ranging from 0.05 to 0.25 for testing variance homogeneity. The significance level used in the Markowski and Markowski study (1991) was not disclosed. In the Lauer and Han (1974) study, the significance level was chosen based on the configuration of n_1 and n_2. Only one distribution, the normal distribution, was examined. Their results showed that the preliminary F-ratio test was more robust than the *t* test or the Welch test used alone. It is noted that Gurland and McCullough (1962) also conducted a simulation study similar to Lauer and Han (1974) but used several different statistics for testing means. The significance level for the preliminary test was chosen in a fashion similar to one used in the Lauer and Han (1974) study. Again the preliminary F-ratio test procedure was recommended over the means test used alone with respect to robustness.

It is of interest to examine the performance of the preliminary F-ratio test procedure using higher significance levels than were used in the Gans (1981) and Moser et al. (1989) studies for testing of variance homogeneity preceding a test of means.

Two simulation studies have been reported using the Levene test as a preliminary test procedure: (1) Olejnik (1987) and (2) Keselman, Games, and Clinch (1979). Olejnik (1987) conducted a study where the Levene test using the median was compared to the O'Brien procedure as a preliminary test procedure preceding the means test (Section 1.2.4). It is noted that Olejnik (1987) used only significance levels of 5% and 10% for testing variance homogeneity in the preliminary test procedure. His results showed the Levene test and the O'Brien procedure used as preliminary tests of variance homogeneity were only slightly more robust than using the means test alone. The Keselman et al. (1979) simulation study compared six preliminary test procedures for variance homogeneity preceding the test for equality of four means. One of the six tests for variance homogeneity was the Levene test using the median. This study will not be discussed in Section 1.2.4 as it was designed to examine four samples instead of two samples which is beyond the focus of this dissertation work. However, it is noted that the authors concluded that none of the preliminary test procedures at the 5% significance level were uniformly robust. Keselman et al. (1979) recommend using a robust test for equality of means such as the Brown and Forsythe procedure, which is based on the Welch test, without employing a preliminary test of variance homogeneity.

As was conjectured in regards to the preliminary F-ratio test procedure, there is interest in the performance of the L_{50} test as preliminary test procedure where a higher significance level is used for testing variance homogeneity.

1.2.2 Methods for Testing Variance Homogeneity

Four simulation studies for testing variance homogeneity using the Levene test are discussed in this section. These are: (1) Brown and Forsythe (1974), (2) Conover et al. (1981), (3) Loh (1987), and (4) O'Brien (1979). In each of the four studies the L_{50} (Levene test using the median) was recommended.

Let $x_{11}, \ldots, x_{1n_1}$ be a random sample with sample size of n_1 from a distribution denoted $f_1(x; \mu_1, \sigma_1)$; and $x_{21}, \ldots, x_{2n_2}$ be a random sample with sample size of n_2 from a distribution denoted $f_2(x; \mu_2, \sigma_2)$. The two samples are assumed to be independent. Let the sample mean and sample variance for $x_{i1}, \ldots, x_{in_i}$ be denoted as $\overline{x_i}$ and s_i^2 for i = 1, 2, respectively.

Brown and Forsythe (1974) Study

Brown and Forsythe (1974) examined six tests for equality of variances, H_o: $\sigma_1^2 = \sigma_2^2$. These tests are described below:

1. F-ratio test, with the upper α critical values F_{α, df_1, df_2} taken from the Snedecor F-table with degrees of freedom $df_1 = n_1 - 1$ and $df_2 = n_2 - 1$

for the significance level α.

2. L_0, the Levene test in which a one-way analysis of variance procedure is performed on the z_{ij} values where $z_{ij} = | x_{ij} - \bar{x}_i |$ which is compared to the upper α critical value $F_{\alpha, 1, n_1 + n_2 - 2}$. Alternate formulations considered by Levene are replacement of z_{ij} by $\sqrt{z_{ij}}$ or by log (z_{ij}). Since, in the authors' empirical sampling, both are less powerful than using z_{ij}, only the L_0's results are reported.

3. L_{50}, the Levene test in which the calculation is based on z_{ij} where $z_{ij} = | x_{ij} - \tilde{x}_i |$ and $\tilde{x}_i$ is the median of $\{ x_{ij}, j = 1, \ldots, n_i \}$. L_{50} is compared to the upper α value $F_{\alpha, 1, n_1 + n_2 - 2}$.

4. L_{10}, the Levene test in Test 2 with the mean replaced by the ten percent trimmed mean with the upper α value $F_{\alpha, 1, n_1 + n_2 - 2}$.

5. Layard (1973) χ^2 test statistic, which is a function of the kurtosis.

6. Miller's jackknife procedure (1968), generalized by Layard (1973).

The alternative formulations of Levene's test statistic were found to be robust under nonnormality. Brown and Forsythe (1974) also concluded that the equality of variances for long-tailed distributions can best be tested by the L_{10} statistic, and by L_{50} for the asymmetric distributions.

<u>Conover, M.E. Johnson, and M.M. Johnson (1981) Study</u>

Conover, M.E. Johnson, and M.M. Johnson (1981) conducted a simulation study comparing fifty-six tests for equality of variance. This study included the most popular and most useful parametric and nonparametric tests available for testing the equality of k variances ($k \geq 2$) assuming the means are unknown. Comparisons were made under the null hypothesis (for robustness) and under the alternative hypothesis (for power). Simulations were conducted with respect to different distributions, sample sizes, means, and variances.

Their results show that many of the extensively used tests, such as Bartlett test, Cochran test, and F-ratio test, have uncontrolled risk of Type I errors when the populations possess asymmetric and heavy-tailed characteristics. Conover et al. (1981) also found the more popular nonparametric tests showed erratic error rates when the population means are unknown. Thus, they felt that it was important to find tests for variances that show more stable error rates and more acceptable power levels.

Three tests were found to be preferable in terms of robustness and power. One of the three tests recommended is the L_{50} test. All three selected tests performed well when applied to the large set of oil and gas lease bidding data discussed in the paper.

Loh (1987) Study

In 1987, Loh examined various modifications of the Levene test for testing homogeneity of variance. The author evaluated six different tests.

These are:

1. L_{50}, the Levene test using the median.

2. Satterthwaite's method, where the test statistic is the same as the L_{50} test, but its value is referred to the F-distribution with 1 and ν degrees of freedom (instead of 1 and $n_1 + n_2 - 2$), where

$$\nu = \frac{\sum_i u_i}{\sum u_i^2 \nu_i^{-1}}, \quad u_i = \sum_j (z_{ij} - \bar{z}_i)^2, \text{ and } \nu_i = n_i - 1. \tag{1.1}$$

3. Data-based transformation, where a power transformation of the z_{ij}, denoted as z_{ij}^p, is performed before applying the L_{50} test, where p is some real number.

4. Satterthwaite's method combined with the data-based transformation.

5. An *Exact* test based on z_{ij}, which is the L_{50} test made exact by computer simulation where each x_{ij} is independent and normal (0, 1).

6. An *Exact* test based on z_{ij}^p, which is the Exact test (Test 5) applied to z_{ij}^p instead of z_{ij}.

The six tests were compared in a simulation experiment using five distributions and fours sets of group sizes. Distributions ranged from symmetric to asymmetric, and light to heavy tails.

Loh's (1987) results show that the L_{50} test and the *Exact* test based on z_{ij} were the best overall performers, with the *Exact* test preferred over the L_{50} test for symmetric distributions and vice versa for asymmetric distributions.

Based on the above cited literature, the L_{50} test is a viable robust choice as a preliminary test of variances.

O'Brien (1979) Study

O'Brien (1979) conducted a simulation study which compared the O'Brien procedure to the L_{50} test. As in the case of the L_{50} test, the O'Brien procedure transforms the original data x_{ij} to create a dependent variable z_{ij}, which is used in an one-way analysis of variance procedure. The transformation z_{ij} is defined as follows:

(1.2) $$z_{ij} = \frac{(w+n_i-2)\, n_i\, (x_{ij}-\bar{x}_i)^2 - w s_i^2 (n-1)}{(n_i-1)(n_i-2)}$$

where w is between 0 and 1 depending on the shape of the population distribution (O'Brien recommends $w = 0.5$).

O'Brien's (1979) results show that for the case of light-tailed distributions, such as the uniform distribution, this procedure is more powerful than the L_{50} test, but less powerful with the heavy-tailed distributions.

1.2.3 F-ratio Test Used as a Preliminary Test

Discussion is limited in this section to four simulation studies which use a preliminary F-ratio test to test variance homogeneity, and based on the F-ratio test outcome, use the *t* test, the Welch test, and/or the Mann-Whitney-Wilcoxon test to test the equality of two means. The four studies discussed are: (1) Gans (1981), (2) Moser et al. (1989), (3) Markowski and Markowski (1991), and (4) Lauer and Han (1974).

Gans (1981) Study

Gans (1981) used the following preliminary test procedure. The equality of variances is tested first using the F-ratio test and if the variances are found to be equivalent, the *t* test is used to test the equality of means; otherwise the Welch test is used. He explored the performance of this preliminary test procedure using the normal distribution, the uniform distribution, and the (single) exponential distribution. The significance levels of 0.05, 0.10, and 0.20 were used for the preliminary F-ratio test.

For the cases of the normal distribution and the uniform distribution, the results show the *t* test rejects too rarely (conservative). When using the preliminary F-ratio test procedure, the conservative bias was counteracted yielding a more robust test. For the cases where the *t* test is liberal (namely the indirect pairing cases), the preliminary F-ratio test procedure improved upon the *t* test, but only slightly as the preliminary F-ratio test procedure was still liberal. The author noted this is especially true for the moderate indirect pairing cases ($R = \sigma_1/\sigma_2 = 0.67$). Using the 5% preliminary F-ratio test procedure, the null rejection rate was as high as 12% for the cases of moderate indirect pairing. Gans (1981) observed that the particular significance level (0.05, 0.10, and 0.20) used for the preliminary F-ratio test procedure has relatively little effect in reducing the bias in the resulting means tests. Gans (1981) noted that as the significance level of the preliminary test procedure is increased, the resulting bias is decreased. Gans (1981) also noted that the Welch test used alone is generally robust.

For the skewed exponential distribution, Gans' (1981) results showed that the *t* test and the Welch test both lacked robustness. For the skewed exponential cases, the *t* test was biased even for the equal sample size cases. The Welch test also yielded biased results, but was less biased than the *t* test. The skewed exponential results using the preliminary F-ratio test procedure at 5%, 10%, or 20% were found to be less biased than the *t* test, and generally comparable to the Welch test results.

Based on these results, Gans (1981) recommends that the Welch test be used directly without a preliminary F-ratio test for variance homogeneity. Gans (1981) reported that the Welch test is superior to both the *t* test and the preliminary F-ratio test procedure when the sample sizes are unequal, and as good as these procedures when the sample sizes are equal.

<u>Moser, Stevens, and Watts (1989) Study</u>

Moser, Stevens, and Watts (1989) report a simulation study similar to that of Gans (1981). Like Gans (1981), the F-ratio test is selected as the preliminary test with significance levels of 5% and 25%. However, only the normal distribution case was examined in this study. If the variances were found to be unequal, Satterthwaite's approximate *F* test for testing H_o: $\mu_1 = \mu_2$ was performed; otherwise, the *t* test was used. Satterthwaite's approximate *F* test is identical to the Welch test; however, the degrees of freedom are calculated in a different manner than the Welch test.

The Moser et al. (1989) results are analogous to those reported in the Gans (1981) study. When $n_1 = n_2$, the *t* test, Satterthwaite's approximate *F* test, and the preliminary F-ratio test procedures (at 5% and 25%) show almost identical simulated null rejection rates for all values of $R = \sigma_1/\sigma_2$. When the sample sizes differ, Satterthwaite's approximate *F* test is more robust than both the *t* test and the preliminary F-ratio test procedures, regardless of the *R* value.

Based on these results, Moser et al. (1989) recommend that Satterthwaite's approximate *F* test be used directly without preliminary testing since it is more robust than both the *t* test and the preliminary F-ratio procedures.

Markowski and Markowski (1991) Study

C.A. Markowski and E.P. Markowski (1991) evaluated the preliminary F-ratio test procedure using a variety of configurations of sample size. They assumed that the samples are drawn from normal distributions and from other symmetric distributions, as well as skewed distributions. The significance level of the F-ratio test for testing variance homogeneity was not disclosed in the paper. The *t* test is used if the hypothesis of equality of variance is not rejected, and the Welch test is used if the preliminary test suggests that the equality of variance assumption is violated. The performance of the preliminary F-ratio test procedure was compared only to the *t* test in this study (i.e., the unconditional Welch test was not discussed).

For the normal distribution cases with equal sample sizes, the authors state the *t* test is robust and that the preliminary F-ratio test procedure is unnecessary.

When the sample sizes differ, relatively small deviations from equal variances (e.g., $R = \sigma_1/\sigma_2 = 0.7$ or 1.4) yielded biased results when using the *t* test. The preliminary F-ratio test procedure also lacked robustness for these cases.

Markowski and Markowski (1991) also investigated the performance of the preliminary F-ratio test procedure for the two symmetric contaminated normal distribution cases. For the unequal sample size cases, the preliminary F-ratio test is consistently liberal because it uses the *t* test too often. For most situations where sample sizes are balanced, the preliminary F-ratio test procedure yielded robust results counteracting the bias seen when using the *t* test alone.

For skewed distributions, (exponential distribution and the chi-square distribution), the preliminary F-ratio test procedure showed large discrepancies between the nominal and estimated actual Type I error levels, regardless of the sample size configurations.

Like Gans (1981) and Moser et al. (1989), Markowski and Markowski (1991) concluded that the preliminary F-ratio test procedure does not warn when the two-sample *t* test is inappropriate for testing the equality of two means and hence, do not recommend use of a preliminary test of variance.

Lauer and Han (1974) Study

Lauer and Han (1974) also conducted a simulation study to investigate methods for test equality of means after using a preliminary F-ratio test for equality of variances in two normal populations. The preliminary F-ratio test procedure is the same as described in the other three studies; however, the degrees of freedom for the Welch test were obtained by using an approximation given in Cochran (1964) which interpolates between t-table values. Important to this study is that the authors allowed the significance level of the F-ratio test to depend on n_1 and n_2 for the normal distributions examined. The significance level was chosen to satisfy an optimality criterion which is derived in the 1974 paper, assuming normality. They concluded a gain with respect to robustness was realized in every instance using the preliminary F-ratio test procedure and hence, recommend use of a preliminary test of variance.

1.2.4 The Levene Test Used as a Preliminary Test

Olejnik (1987) conducted a simulation study comparing two preliminary test procedures to the unconditional ANOVA (analysis of variance) test for testing the equality of two means. Since only two sample populations were used, the unconditional ANOVA test is just the squared t test statistic t^2.

Five distribution types were examined: (1) normal, (2) platykurtic, (3) skewed, (4) leptokurtic, and (5) skewed/leptokurtic. For the platykurtic case, the distribution has a flatter top than the normal distribution (e.g., rectangular) with a negative coefficient of kurtosis. In contrast, the leptokurtic distribution is usually taller and slimmer (peaked) than the normal distribution with a positive coefficient of kurtosis.

To assess the preliminary test procedures, Olejnik used the following strategy. First, the equality of means was tested yielding the Type I error rate using the *t* test (i.e., no preliminary test for variance homogeneity) with n = 1000 simulation runs. Secondly, the Type I error rate for the preliminary test procedure was obtained in the following manner. The L_{50} test and the O'Brien test procedure were computed to test for variance homogeneity at the 5% and 10% significance levels for all 1000 simulation runs. The succeeding *t* test was computed only in n^* cases ($n^* \leq n$) where n^* is the number of tests where variance equality was not rejected; thereby, yielding a Type I error rate for the preliminary test procedure.

For the equal variance cases, Olejnik's (1987) results show both preliminary test procedures are robust, except for the skewed/leptokurtic distributions. For skewed/leptokurtic distributions, both preliminary test procedures are conservative.

For the unequal variance cases, Olejnik's (1987) results using the *t* test concur with what has been previously cited. For the direct pairing cases, the *t* test is conservative; whereas, it is liberal for the indirect pairing cases. The results of <u>both</u> preliminary test procedures using significance levels of 5% and 10% for the analysis of variance equality are similar. When the *t* test for means is conservative, the preliminary test procedures are conservative, and when the *t* test for means is liberal, the preliminary test procedures are also liberal. It was also demonstrated as the difference in sample size increases, the degree of liberalism and conservatism also increases. Their results show that when the tests are liberal, the preliminary test procedures are more liberal than the *t* test. Therefore, Olejnik (1987) concludes there is very little achieved when using these preliminary test procedures.

1.3 Tests for Skewness/Symmetry

Murphy (1976), Gans (1981), and Olejnik (1987) all concluded skewness adversely affects the robustness of the *t* test and the Welch test. Therefore, a second class of preliminary testing, a test of symmetry/skewness, will be examined in this dissertation. Two test methods were selected, the D'Agostino S_U test for skewness and the Triples test for symmetry.

1.3.1 D'Agostino's Skewness Test

D'Agostino's test is a test of normality versus non-normality, which is sensitive to skewed nonnormal alternatives. A sketch of this procedure is now described.

A simple and very powerful test of normality against skewed alternatives (Shapiro, Wilk, & Chen, 1968) is the test based on the standardized third sample moment,

(1.3)
$$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left[\frac{1}{n}\sum_{i=1}(x_i-\bar{x})^2\right]^{3/2}}$$

The test of normality is rejected if g_1 is large in absolute value when compared to a table of critical values (Pearson & Hartley, 1966). However, Pearson and Hartley's (1966) table is extremely scant for critical values starting at $n = 25$, since it is tabulated for increments of $n = 25$ or larger. These tables are appropriate for two-sided tests with significance levels of 0.02 and 0.10, and one-sided tests with levels of 0.01 and 0.05.

D'Agostino (1970), using a transformation due to Johnson (1949), transforms the distribution of g_1 to that of the standard normal distribution with mean zero and variance of one. Details of this transformation are outlined in Chapter II. Results from D'Agostino's Monte Carlo simulations for $n < 25$ and checks with an existing table of Pearson and Hartley (1966) for $n \geq 25$ show that the accuracy of the transformation is very good.

1.3.2 Triples Test

The second approach is a Triples test described in a paper by Randles, Fligner, Policello, and Wolfe (1980). Let $x_1, \ldots, x_n$ denote a random sample from a continuous population where i, j, k are distinct integers such that $1 \leq i, j, k \leq n$. The Triples test is an asymptotically distribution-free procedure which examines each triple (x_i, x_j, x_k). If the middle observation is closer to the smallest observation than it is to the largest observation, then a "right triple" is formed (looks skewed to the right). If the middle observation is closer to the largest observation than it is to the smallest observation, then a "left triple" is formed (looks skewed to the left). The Triples test statistic is a function of the numbers of right triples and left triples. The procedure used to obtain the test statistic is outlined in Chapter II.

Randles et al. (1980) compared three procedures for testing whether a univariate population is symmetric about some unspecified value compared to an immense class of asymmetric distribution alternatives. The Triples test was compared to Gupta's skewness test (Gupta, 1967) and Gupta's nonparametric procedure (Gupta, 1967). Their results show that the Triples test is superior to either competitor for testing the hypothesis of symmetry, even for sample sizes as small as 20, while possessing good power for detecting asymmetric alternative distributions (Randles et al., 1980).

The performance of the D'Agostino skewness test and the Triples test as preliminary tests for symmetry/skewness is examined in Chapter VI. As previously stated, the goal of incorporating a preliminary test of symmetry/skewness is to alert the researcher when the means tests sensitive to asymmetry, like the *t* test and the Welch test, should be avoided.

1.4 Summary of Research to be Conducted

The purpose of this dissertation is to further explore the appropriateness of testing the equality of two means using either a two-sample *t* test, the Welch test, or the Mann-Whitney-Wilcoxon test based on the results of using two classes of preliminary tests. One class of preliminary tests determines whether the population variances differ, and the other class ascertains if the underlying distributions are symmetric or skewed. The "commonly used" F-ratio test and the "recommended" L_{50} test (hereafter denoted Levene test) will be compared and assessed as preliminary tests for variance homogeneity. The D'Agostino S_U test and the Triples test will be compared and assessed as preliminary tests of symmetry/asymmetry. These assessments will be based on seven population distributions: normal, uniform, double exponential, logistic, lognormal, gamma and contaminated normal. These were arbitrarily chosen based on their symmetric and skewed population characteristics. Unequal variance and unequal sample size cases under direct pairing and indirect pairing will be examined. The nominal significance levels of 0.05, 0.25, 0.50, and 0.75 for both classes of preliminary tests will be evaluated.

The three objectives of this dissertation work are:

1. Comparison of the F-ratio test to the Levene test as a preliminary

test for testing variance homogeneity preceding a means test. Based on this comparison, a preliminary test for variance homogeneity at a significance level α^* will be chosen and implemented into an expert system.

2. Comparison of the D'Agostino S_U test and the Triples test as tests of symmetry/asymmetry. Based on this comparison, a preliminary test for symmetry/asymmetry at a significance level α^{**} will be chosen and implemented into an expert system.

3. Evaluation of the performance of an expert system which selects a method for testing H_o: $\mu_1 = \mu_2$ based on two classes of preliminary tests: (1) the F-ratio or the Levene test at a significance level of α^* for the test of variance homogeneity; and (2) the D'Agostino S_U test or the Triples test at a significance level of α^{**} for the test of symmetry/asymmetry.

The purpose of the expert system is to yield a robust test for testing the hypothesis H_o: $\mu_1 = \mu_2$ when the data analyst does not know whether the variances are homogeneous and whether symmetry is present. Based on the results of the two preliminary tests, the *t* test, the Welch test or the Mann-Whitney-Wilcoxon test is selected to test the equality of means.

The expert system is constructed in the following way:

Case I: If $\sigma_1 = \sigma_2$ and symmetry is concluded, then the *t* test is used.

Case II: If $\sigma_1 \neq \sigma_2$ and symmetry is concluded, then the Welch test is used.

Case III: If $\sigma_1 = \sigma_2$ and symmetry is rejected, then the Mann-

Whitney-Wilcoxon test is used.

Case IV: If $\sigma_1 \neq \sigma_2$ and symmetry is rejected, then the Welch test is used.

It is noted that robust methods exist for testing H_o: $\mu_1 = \mu_2$ for Cases I-III, but no robust method exists for Case IV.

The expert system procedure is outlined in Figure 1.

		Test of Symmetry	
		Nonsignificant	Significant
Test of Variance Homogeneity	Nonsignificant	*t* test	Mann-Whitney-Wilcoxon test
	Significant	Welch test	Welch test

Figure 1. Expert System.

Results from Chapters III, IV and V are used to assess which of the two preliminary tests of variance homogeneity, along with a specific significance level are recommended for use in the expert system procedure. In Chapter VI the performance of the two methods for testing skewness/symmetry is judged and, based on these findings, the most robust testing method will be implemented into an expert system. The overall performance of this expert system is discussed in Chapter VII. A summary and recommendations for future study follow in Chapter VIII.

CHAPTER II

METHODOLOGY

2.1 Introduction

This chapter contains the details describing the two-sample methodology used to test the equality of means, variance homogeneity and skewness/symmetry under selected distributions.

Let $x_{11}, \ldots, x_{1n_1}$ be a random sample with sample size of n_1 from a distribution denoted $f_1(x; \mu_1, \sigma_1)$; and $x_{21}, \ldots, x_{2n_2}$ be a random sample with sample size of n_2 from a distribution denoted $f_2(x; \mu_2, \sigma_2)$. The two samples are assumed to be independent. Let the sample mean and sample variance for $x_{i1}, \ldots, x_{in_i}$ be denoted as $\overline{x_i}$ and s_i^2 for $i = 1, 2$, respectively.

2.2 Testing the Equality of Means

The t test, the Welch test, and the Mann-Whitney-Wilcoxon test procedures of H_o: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 \neq \mu_2$, are now described.

The t test is the given as

(2.1a)
$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{s^2 (1/n_1 + 1/n_2)}} ,$$

(2.1b) $$\text{where } s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1+n_2-2)}$$

is the pooled estimate of σ^2, assuming $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

The Welch test statistic is

(2.2a) $$t_w = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{(s_1^2/n_1)+(s_2^2/n_2)}} ,$$

which uses Satterthwaite's (1946) approximation for the degrees of freedom:

(2.2b) $$df = \frac{(s_1^2/n_1+s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1)+(s_2^2/n_2)^2/(n_2-1)} .$$

The Mann-Whitney-Wilcoxon statistic is

(2.3) $$z = \frac{S-n_1(n_1+1)/2-n_1n_2/2}{\sqrt{n_1n_2(n_1+n_2+1)/12}} ,$$

where S is the sum of the ranks assigned to the sample observations from group 1, and z is an approximate normal deviate.

The α-level tests of H_o: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 \neq \mu_2$ are $|t| > t_{\alpha/2,\, n_1+n_2-2}$, $|t_w| > t_{\alpha/2,\, df}$, and $|z| > z_{\alpha/2}$ for the t test, the Welch test, and the Mann-Whitney-Wilcoxon test, respectively, where z_α is the upper α-point of the standard unit normal distribution and $t_{\alpha,r}$ is the upper α-point of a t

distribution with r degrees of freedom.

2.3 Testing the Equality of Variances

The F-ratio test and the Levene test of H_o: $\sigma_1^2 = \sigma_2^2$ vs. H_1: $\sigma_1^2 \neq \sigma_2^2$ are now described, assuming the sampling conditions described in Section 2.1 hold.

The F-ratio α-level test statistic is

(2.4) $$F = (\text{larger } s_1^2, s_2^2) \,/\, (\text{smaller } s_1^2, s_2^2) > F_{\alpha,\, df_L,\, df_S}$$

where $F_{\alpha,\, df_L,\, df_S}$ is the upper α-point of an F random variable where df_L is the degrees of freedom (i.e sample size minus one) from the sample which has the larger sample variance, and df_S is the degrees of freedom from the sample which has the smaller sample variance.

The Levene α-level test is

(2.5) $$L = \frac{\sum n_i (z_{i.} - z_{..})^2}{\sum\sum (z_{ij} - z_{i.})^2 / (n_1 + n_2 - 2)} > F_{\alpha,\, 1,\, n_1 + n_2 - 2},$$

which is the one-way analysis of variance F-test computed on the z_{ij} values, where $z_{ij} = |x_{ij}\text{-median of group } i|$.

2.4 Testing of Symmetry Versus Skewness

The D'Agostino test of skewness and the Triples test of symmetry are described first for a general random sample $x_1, \ldots, x_n$ from some distribution $f(x; \mu, \sigma)$ before generalization to the two samples described in Section 2.1. It is convenient to let $\bar{x}$ denote the sample mean of $x_1, \ldots, x_n$ and to let the sample estimates of $\sqrt{\beta_1}$, the third standardized moment, and β_2, the fourth standardized moment, be denoted as

(2.6) $$\sqrt{b_1} = m_3 / m_2^{3/2} \ ,$$

(2.7) $$\text{and } b_2 = m_4 / m_2^2,$$

(2.8) $$\text{where } m_k = \sum (x_i - \bar{x})^k / n \quad \text{for } k = 2, 3, 4.$$

The D'Agostino S_U skewness statistic is obtained in the following manner. First, compute $\sqrt{b_1}$ from the sample data. Secondly compute $Z(\sqrt{b_1})$, where

(2.9a) $$Z(\sqrt{b_1}) = \delta \ln(Y/a + [(Y/a)^2 + 1]^{1/2}),$$

(2.9b) $$Y = \sqrt{b_1}\left[\frac{(n+1)(n+3)}{6(n-2)}\right],$$

(2.9c) $$W^2 = -1 + [2(\beta_2(\sqrt{(b_1)}) - 1)]^{1/2} \ ,$$

(2.9d) $$\beta_2(\sqrt{b_1}) = \frac{3(n^2 + 27n - 70)(n+1)}{(n-2)(n+5)(n+7)(n+9)} \ ,$$

(2.9e) $$\delta = 1/\sqrt{\ln W} \quad \text{and} \quad a = 2/(W^2 - 1)^{1/2} \ .$$

The α-level D'Agostino test of skewness is:

$$Z(\sqrt{b_1}) > z_\alpha. \tag{2.10}$$

$Z(\sqrt{b_1})$ is approximately $n(0, 1)$ under the null hypothesis of population normality for cases where $n > 8$ (D'Agostino, Belanger, & D'Agostino, 1990).

For the Triples test of symmetry, we again let $x_1, \ldots, x_n$ denote the random sample. The test rejects H_o of symmetry if $|T_1| > t_{n,(\alpha/2)}$ where

$$T_1 = n^{1/2}\hat{\eta}/\hat{\sigma}_n , \tag{2.11a}$$

$$\hat{\eta} = \frac{\{(\text{number of right triples}) - (\text{number of left triples})\}}{\left[3\binom{n}{3}\right]} \tag{2.11b}$$

and $\hat{\sigma}_n$ is the standard deviation of $\hat{\eta}$. The statistic $\hat{\eta}$ is calculated as

$$\hat{\eta} = \binom{n}{3}^{-1} \sum_{i<j<k} f^*(X_i, X_j, X_k) \tag{2.12}$$

where $f^*(X_i, X_j, X_k) = \{\text{sign}(X_i + X_j - 2X_k) + \text{sign}(X_i + X_k - 2X_j) + \text{sign}(X_j + X_k - 2X_i)\}/3$ and $\text{sign}(u) = -1, 0,$ or 1 as $u <, =,$ or > 0.

To compute var $(\hat{\eta})$, let

(2.13a) $$\frac{\hat{\sigma}_n^2}{n} = \binom{n}{3}^{-1} \sum_{c=1}^{3} \binom{3}{c}\binom{n-3}{3-c} \xi_c$$

(2.13b) where ξ_c = var $[f_c^*(X_1, \ldots, X_c)]$.

Then $\xi_1 = var[f_1^*(X_1)]$, with $f_1^*(x) = E[f^*(x, X_2, X_3)]$, yields

(2.14a) $$\hat{\xi}_1 = \frac{1}{n} \sum_{i=1}^{n} (\hat{f}_1^* (X_i) - \hat{\eta})^2 \text{ , where}$$

(2.14b) $$\hat{f}_1^*(X_i) = \frac{1}{\binom{n-1}{2}} \sum \sum_{\substack{j<k \\ j \neq i \neq k}} f^*(X_i, X_j, X_k) \text{ .}$$

Similarly,

(2.15a) $$\hat{\xi}_2 = \frac{1}{\binom{n}{2}} \sum \sum_{j<k} (\hat{f}_2^*(X_j, X_k) - \hat{\eta})^2 \text{ , where}$$

(2.15b) $$\hat{f}_2^* (X_j, X_k) = \frac{1}{n-2} \sum_{\substack{i=1 \\ i \neq j \neq k \\ i \neq k}} f^*(X_i, X_j, X_k) \text{ ,}$$

(2.16) and $\hat{\xi}_3 = \frac{1}{9} - \hat{\eta}^2$.

2.5 Selected Configurations of Distributions, Sample Sizes and Variance Ratios Used in the Simulation

Type I error rates for testing the homogeneity of means were simulated under a variety of conditions. Seven distributions were arbitrarily chosen based on their different population characteristics. These distributions are classified into three groups: (1) symmetric, (2) asymmetric, and (3) outlier model. Chapter III examines the use of testing for variance homogeneity preceding the test of equality of means, H_o: $\mu_1 = \mu_2$ for the four symmetric distributions: (1) normal, (2) uniform, (3) double exponential, and (4) logistic; Chapter IV examines this test for the two asymmetric distributions: (1) lognormal and (2) gamma; and Chapter V for the outlier model, the contaminated normal.

To evaluate the performance of the preliminary test of variance homogeneity, the following standard deviation ratios $R = \sigma_1 / \sigma_2$ are used: 0.25, 0.50, 0.67, 1.0, 1.5, 2.0, and 4.0. Clearly the sample variances are equal when $R = 1$. Sample size configurations (n_1:n_2) used in the simulations are: 5:5, 5:10, 5:20, 10:10, 10:20, 10:40, and 20:20. This allows for both direct and indirect pairings to be examined.

Direct pairing occurs when R = 0.25, 0.50 and 0.67 holds with the imbalanced samples 5:10, 5:20, 10:20, and 10:40. Severe direct pairing is defined to occur in those above situations when R = 0.25 and 0.50; whereas, moderate direct pairing occurs with the R = 0.67 cases. An example of a moderate direct pairing case is as follows. Let group 1 be distributed with mean $\mu_1 = 0$ and standard deviation $\sigma_1 = 0.67$ with a sample size of $n_1 = 5$; and group 2 be distributed with mean $\mu_2 = 0$ and standard deviation $\sigma_2 = 1.0$ with sample size of $n_2 = 10$. Then the ratio $R = \sigma_1/\sigma_2 = 0.67$ and the group with the smaller σ ($\sigma_1 = 0.67$) is estimated from the group with the smaller sample size ($n_1 = 5$).

Indirect pairing occurs when $R = 1.5$, 2.0 and 4.0 holds with the imbalanced sample sizes 5:10, 5:20, 10:20, and 10:40. Severe indirect pairing is defined to occur for the above cases when $R = 2.0$ and 4.0; and moderate indirect pairing for the $R = 1.5$ cases. An example of moderate indirect pairing is as follows. Let group 1 be distributed with mean $\mu_1 = 0$ and standard deviation $\sigma_1 = 1.50$ with a sample size of $n_1 = 5$; and group 2 be distributed with mean $\mu_2 = 0$ and standard deviation $\sigma_2 = 1.0$ with sample size of $n_2 = 10$. Then the ratio $R = \sigma_1/\sigma_2 = 1.50$ and the group with the smaller σ ($\sigma_2 = 1.00$) is estimated from the group with the larger sample size ($n_2 = 10$).

As stated in Chapter I, it is of interest to examine the performances of the F-ratio test and the Levene test as preliminary tests of variance homogeneity using nontraditional significance levels. Therefore, we have arbitrarily chosen significance levels of 0.50 and 0.75 in addition to the traditional 0.05 and 0.25 levels. In all cases, the means are set equal, so that performance is evaluated relative to Type I error only. All tests are two-tailed. No power results are reported.

2.6 Generation of Random Realizations

This section contains an outline of how the random realizations are generated for each specified distribution. As before let $x_{11}, \ldots, x_{1n_1}$ be a random sample of size n_1 from the distribution $f_1(x; \mu_1, \sigma_1)$; and $x_{21}, \ldots, x_{2n_2}$ be a random sample of size n_2 from the distribution $f_2(x; \mu_2, \sigma_2)$, where it is assumed that the two samples are independent.

The random realizations from the standardized distribution f_2 $(x; \mu_2, \sigma_2)$ are generated for each of the selected distributions. For the first sample, f_1 $(x; \mu_1, \sigma_1)$, the random realizations are generated in the same fashion, but shape parameters and scale parameters are adjusted to yield the desired standard deviation ratio $R = \sigma_1/\sigma_2$.

The IMSL random number generator RNSET, which initializes the seed, is used in all of the simulations.

2.6.1 Normal Distribution

In the case of the normal distribution, both sample means are set to zero, $\mu_1 = \mu_2 = 0$. For the second of the two samples, the distribution $f_2(x; \mu_2, \sigma_2)$ is initialized as a normal (0, 1). The FORTRAN function RNNOF was used to generate the normal (0, 1) random numbers. RNNOF is the function form of RNNOR which generates standard normal (Gaussian) numbers using an inverse CDF technique. For the first sample, random normal (0,1) numbers are generated and then multiplied by a scale factor R to yield random realizations with distribution $f_1(x; 0, R)$.

2.6.2 Uniform Distribution

Let x be uniform (a, b) with mean $\mu = (a + b)/2$ and standard deviation $\sigma = (b - a)/\sqrt{12}$. As in the case of the normal distribution, we set both means equal to zero, $\mu_1 = \mu_2 = 0$ for the uniform distribution cases. For the second of two samples, the distribution $f_2(x;\ \mu_2,\ \sigma_2)$ is arbitrarily designated as a uniform (-1/2, 1/2) distribution yielding a mean $\mu_2 = 0$ and standard deviation $\sigma_2 = 1/\sqrt{12}$. To achieve the desired R ratio, random realizations from a uniform $(-R/2, R/2)$ distribution are generated for the $f_1(x;\ \mu_1,\ \sigma_1)$ distribution yielding mean $\mu_1 = 0$ and a standard deviation $\sigma_1 = R/\sqrt{12}$.

Uniform random realizations for a uniform $(-R/2, R/2)$ distribution are constructed in the following fashion. First, random numbers u_i from a uniform (0,1) distribution are generated using the FORTRAN function RNUN. Then, the uniform $(-R/2, R/2)$ random realizations are:

(2.17) $$x_i = R(u_i - 1/2).$$

2.6.3 Double Exponential Distribution

Let x have the double exponential probability density function $f(x)$ where

(2.18) $$f(x) = \frac{\exp[-|x|]}{2}, \quad -\infty < x < \infty .$$

The mean and variance are

(2.19) $$\mu = E(x) = 0 \text{ and}$$

(2.20) $$\sigma^2 = \text{Var}(x) = 2.$$

To simulate x for distribution $f_2(x;\ 0,\ \sqrt{2}\ )$, we use the following transformation:

(2.21) $$x = (y_1 - y_2)/2$$

where y_1 and y_2 are two independent chi-square random variables, each with two degrees of freedom.

For the first sample, let

(2.22) $$x = R(y_1 - y_2)/2$$

to yield realizations with distribution $f_1(x;\ \mu_1,\ \sigma_1) = f_1(x;\ 0,\ \sqrt{2}R)$.

The FORTRAN function RNCHI generates random values from a chi-squared distribution with df degrees of freedom. In our case $df = 2$ and the chi-squared random number y is generated as

(2.23) $$y = -2 \ln (u)$$

where u is an independent random number from a uniform (0,1) distribution (see Section 2.6.2).

2.6.4 Logistic Distribution

Let $f(x)$ represent the probability density function for a logistic distribution

(2.24) $$f(x) = \frac{e^x}{(1+e^x)^2} \quad where \; -\infty \le x \le \infty \; . \; \cdot$$

The mean and variance are

(2.25) $$\mu = E(x) = 0 \text{ and}$$

(2.26) $$\sigma^2 = Var(x) = 3/\pi^2$$

For the second of two samples, the standard deviation is preset to equal 1, so that $f_2(x;\, 0,\, \sigma_2) = f_2(x;\, 0,\, 1)$. The random numbers x_i for this logistic distribution are generated using the transformation

(2.27) $$x_i = \frac{\sqrt{3}}{\pi} \log\left(\frac{u_i}{1-u_i}\right)$$

where u_i is uniform (0,1).

The transformation that yields realizations for the first sample $f_1(x;\, 0,\, \sigma_1) = f_1(x;\, 0,\, R)$ is:

(2.28) $$x_i = \frac{\sqrt{3}\,R}{\pi} \log\left(\frac{u_i}{1-u_i}\right) \,,$$

where u_i is uniform (0,1).

2.6.5 Lognormal Distribution

The probability density function for the lognormal distribution with parameters a and b is:

(2.29) $$f(x)=\frac{1}{b\,x\,\sqrt{2\pi}}\exp\left(-\frac{1}{2b^2}(\ln x-a)^2\right) \text{ for } x>0\,.$$

The mean μ, variance σ^2, and coefficient of skewness are

(2.30) $$\mu = E(x)=\exp\left(a+\frac{b^2}{2}\right)$$

(2.31) $$\sigma^2 = Var(x)=w(w-1)\exp(2a)\text{ , and}$$

(2.32) $$\text{coefficient of skewness} = (w+2)(w-1)^{1/2}$$

where $w = \exp(b^2)$. Let x be $n(a, b)$, then $y = e^x$ has the lognormal probability density function in (2.29).

Three lognormal distributions are selected due to their degree of skewness. In each case, $a_2 = 0$ in the second of the two samples from distribution $f_2(x; \mu_2, \sigma_2)$ denoted as lognormal (a_2, b_2). Assuming the means of the two samples to be equal $\mu_1 = \mu_2$ results in the conditions of:

(2.33) $$a_1=\frac{1}{2}(b_2^2-b_1^2) \text{ and}$$

$$(2.34) \qquad R^2 = \frac{\exp(2a_1)\exp(b_1^2)\,[\exp(b_1^2)-1]}{\exp(b_2^2)\,[\exp(b_2^2)-1]} \, .$$

The three b_2 parameter values chosen for the $f_2(x; \mu_2, \sigma_2)$ lognormally distributed sample are: (1) $b_2 = 0.4$, (2) $b_2 = 1.0$, and (3) $b_2 = 1.75$. The coefficients of skewness for these cases are 1.3, 6.2, and 105.6, respectively. The case of $b_2 = 0.4$ is denoted as slight skewness, $b_2 = 1.0$ as moderate skewness, and $b_2 = 1.75$ as heavy skewness. Lognormal distributions with slight skewness ($b_2 = 0.4$) and heavy skewness ($b_2 = 1.75$) are displayed in Figure 2. The slightly skewed lognormal distribution ($b_2 = 0.4$) in Figure 2 looks similar to a normal bell-shaped curve but is slightly skewed to the right. The heavy skewed lognormal ($b_2 = 1.75$) peaks rapidly leaving a short left tail and than decreases much slower, resulting in a long right tail.

Substituting (2.33) into (2.34) yields

$$(2.35) \qquad b_1^2 = \ln(R^2(0.1735) + 1).$$

$$(2.36) \qquad b_1^2 = \ln(R^2(1.7183) + 1).$$

$$(2.37) \qquad b_1^2 = \ln(R^2(20.3809) + 1).$$

for the $f_1(x; \mu_1, \sigma_1)$ distribution samples for the case of slight, moderate, and heavy skewness, respectively.

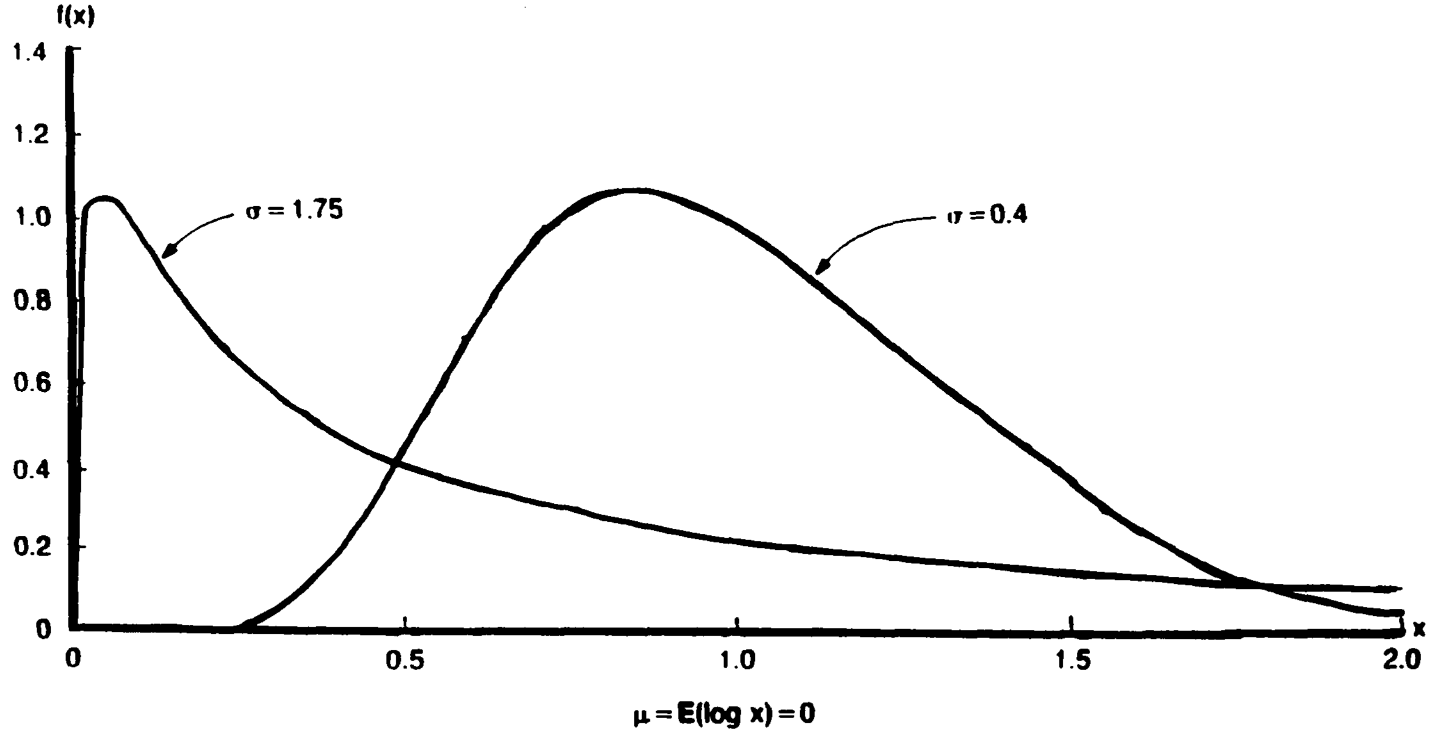

Figure 2. **Lognormal Distributions (Rothschild & Logothetis, 1986)**

Source: **Rothschild, V., & Logothetis, N. (1986).** <u>**Probability Distributions**</u>**, p. 32**

Table 1

Parameter Values to Generate Lognormal Samples

	Slight Skewness		Moderate Skewness		Heavy Skewness	
$R = \sigma_1 / \sigma_2$	a_1	b_1	a_1	b_1	a_1	b_1
0.25	0.0746	0.1039	0.4490	0.3194	1.1205	0.9064
0.50	0.0588	0.2061	0.3213	0.5978	0.6275	1.3444
0.67	0.0425	0.2739	0.2141	0.7561	0.3726	1.5223
1.00	0.0000	0.4000	0.0000	1.0000	0.0000	1.7500
1.50	-0.0848	0.5741	-0.2912	1.2579	-0.3923	1.9614
2.00	-0.1835	0.7260	-0.5317	1.4365	-0.6753	2.1007
4.00	-0.5843	1.1527	-1.1748	1.8302	-1.3639	2.4063

Table 1 displays the parameters values a_1 and b_1 used in the first sample $f_1(x; \mu_1, \sigma_1)$ to form the specified standard deviation ratio.

The FORTRAN function RNLNL is used to create the random realizations for a lognormal distribution $f(x; \mu, \sigma)$ using the transformation $y = e^x$ where x is $n(a, b)$ (IMSL, 1989a).

2.6.6 Gamma Distribution

The probability density function for the gamma distribution with shape parameter α and scale parameter β is

(2.38) $$f(x) = \frac{1}{\Gamma(\alpha)\,\beta^{\alpha}} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \quad \textit{where } x>0,\ \alpha>0,\ \beta>0$$

with mean $\alpha\beta$, variance $\alpha\beta^2$ and coefficient of skewness $2/\sqrt{\alpha}$.

Two gamma distributions are examined in this section, one with shape parameter equal to 3 and unit scale parameter (denoted as G(3,1)), and the other with shape parameter equal to 2 and unit scale parameter (denoted as G(2,1)). Figure 3 displays the G(3,1) and G(2,1) distributions. As seen in Figure 3, the G(3,1) distribution is only slightly skewed (coefficient of skewness = 1.5) whereas the skewness is more pronounced in the G(2,1) distribution (coefficient of skewness = 1.41).

The gamma random realizations are generated using RNGAM (IMSL Routine) which yields random numbers with shape parameter α and unit scale parameter ($\beta = 1$). To formulate the random numbers from a $G(\alpha,\beta)$ distribution, the numbers generated from RNGAM are then multiplied by the scale parameter β.

Applying the condition of equal means results in

(2.39) $$\alpha_1\beta_1 = \alpha_2\beta_2$$

For the case of the G(3,1) distribution where $\alpha_2 = 3$ and $\beta_2 = 1$, the results of (2.39) reduce to

(2.40) $$\alpha_1 = 3/\beta_1,$$

and it follows that:

(2.41) $$\beta_1 = R^2$$

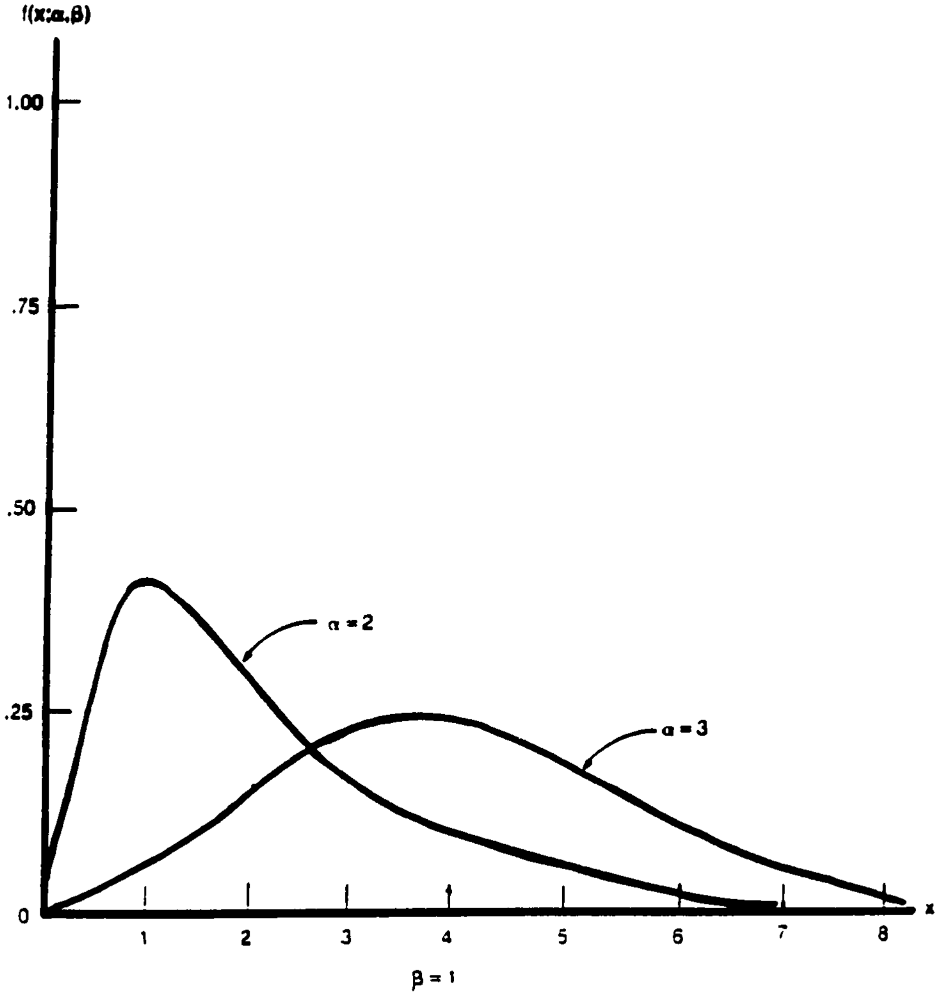

Figure 3. Gamma Distributions (Rothschild & Logothetis, 1986)

Source: Rothschild, V., & Logothetis, N. (1986). <u>Probability Distributions,</u> p. 42

(2.42) $\text{and } \alpha_1 = 3/R^2.$

For the case of the G(2,1) distribution where $\alpha_2 = 2$ and $\beta_2 = 1$, the results of (2.39) reduce to

(2.43) $\alpha_1 = 2/\beta_1,$

and it follows that:

(2.44) $\beta_1 = R^2$

(2.45) $\text{and } \alpha_1 = 2/R^2.$

Table 2 lists the α_1 and β_1 used in the first sample to achieve each R value.

Table 2

Parameter Values to Generate Gamma Samples

	G (3,1)		G (2,1)	
R	α_1	β_1	α_1	β_1
0.25	48.0000	0.0625	32.0000	0.0625
0.50	12.0000	0.2500	8.0000	0.2500
0.67	6.6830	0.4489	4.4553	0.4489
1.00	3.0000	1.0000	2.0000	1.0000
1.50	1.3333	2.2500	0.8889	2.2500
2.00	0.7500	4.0000	0.5000	4.0000
4.00	0.1875	16.0000	0.1250	16.0000

2.6.7 Contaminated Normal: Outlier Model

The contaminated normal has observations which usually follow a normal distribution but where occasionally something goes wrong with the experiment so that an occasional observation is a gross error. Under this model, $f(x)$ takes the form

(2.46) $$f(x) = (1 - p)n(\mu_a, \sigma_a) + pn(\mu_b, \sigma_b)$$

where $0 \le p < 1$, $n(\mu, \sigma)$ is the normal distribution function with mean μ and standard deviation σ. The former normal distribution, $n(\mu_a, \sigma_a)$ denotes the "parent" and the latter normal distribution, $n(\mu_b, \sigma_b)$ denotes the "child" or the "outlier". The mean and variance of the contaminated normal is

(2.47) $$\mu = (1-p)\mu_a + p\mu_b$$

(2.48) and $$\sigma^2 = (1-p)(\mu_a^2 + \sigma_a^2) + p(\mu_b^2 + \sigma_b^2) + [(1-p)\mu_a + p\mu_b]^2 .$$

The contaminated normal numbers are generated in the following way. RNUN is used to generate a uniform random number u, as described in Section 2.6.2, and RNNOF generates a normal (0,1) random variable, as described in Section 2.6.1. If the uniform random number u is less than (1 - p), then the corresponding normal (0,1) random number is transformed into a $n(\mu_a, \sigma_a)$ value, otherwise it is transformed to a $n(\mu_b, \sigma_b)$ value.

Two types of contaminated normal (0,1) distributions are investigated. The first has 95% of the observations in the parent distribution and 5% in the child with differing means; resulting in a skewed distribution. In the second contaminated normal (0,1), 90% of the observations are in the parent and 10% in the child with the mean in both the parent and child set to zero; resulting in a symmetric distribution. Since the contaminated normal mean is set to zero, the variance becomes

$$\sigma^2 = (1 - p)(\mu_a^2 + \sigma_a^2) + p(\mu_b^2 + \sigma_b^2). \tag{2.49}$$

The random numbers for the contaminated normal (0,1) with $p = .05$ are generated as follows. In the second sample the parent is chosen to be a $n(-0.15, 0.31019)$ and the child as a $n(2.85, 3.1019)$ thus resulting in a contaminated normal (0,1). If the uniform random number is less than 0.95 then the corresponding normal (0,1) random number is transformed to the parent, $n(-0.15, 0.31019)$, if not, a $n(2.85, 3.1019)$ is created.

For the first of the two samples, the parent remains as $n(-0.15, 0.31019)$ and the mean of the child is fixed at 2.85. Thus, the ratio R is obtained by changing the standard deviation of the child by the following equality,

$$\sigma_c^2 = (R^2 - .5189)/0.05. \tag{2.50}$$

For the symmetric contaminated normal (0, 1) with $p = 0.10$, the means of the parent and child are both set to zero. In the case of the second sample, $f_2(x; 0, 1)$, the parent is initialized as n $(0, \sqrt{.5})$ and the child as n $(0, \sqrt{5.5})$ thereby resulting in a symmetric contaminated normal (0, 1). In this case, when the uniform random number is less than 0.90, the corresponding normal (0, 1) random number is transformed to a parent $n(0, \sqrt{.5})$ value; otherwise to a n $(0, \sqrt{5.5})$ value.

The parent remains as n $(0, \sqrt{.5})$ and the mean of the child is again fixed at 0 for the sample from the $f_1(x; 0, R)$ distribution. The ratio R results by changing the standard deviation ratio of the child by the following equation,

$$\sigma_c^2 = (R^2 - 0.45)/0.1. \qquad (2.51)$$

For the direct pairing cases, the unbalanced sample size configurations are reversed so $n_1 > n_2$ resulting in $n_1{:}n_2$ = 10:5, 20:5, 20:10, and 40:10. By choosing $n_1 > n_2$, the σ_c used for the moderate indirect pairing cases can then be also used for the moderate direct pairing cases. The same is true for the severe direct pairing cases.

Table 3 gives the σ_c values used to yield the R values in the symmetric and asymmetric cases.

Table 3

Parameter Values to Generate Contaminated Normal Samples

$R = \sigma_1 / \sigma_2$	σ_c for $p = 0.10$ (Symmetric)	σ_c for $p = 0.05$ (Asymmetric)
0.25	12.4700	17.5961
0.50	5.9582	8.3440
0.67	4.2426	5.8840
1.00	2.3452	3.1019
1.50	4.2426	5.8840
2.00	5.9582	8.3440
4.00	12.4700	17.5961

CHAPTER III

SYMMETRIC DISTRIBUTIONS

3.1 Introduction

This chapter contains an examination of the performance of the preliminary variance equality test procedures for the four symmetric distributions: normal, uniform, double exponential, and logistic. Presented are the simulation results for testing the hypothesis of H_o: $\mu_1 = \mu_2$ after utilizing the two types of preliminary test of variance homogeneity, the F-ratio test and the Levene test, for each of the four symmetric distributions.

Preliminary testing is handled in the following manner. When the preliminary test for variance homogeneity is found to be not significant at the specified significance level, the *t* test is employed; otherwise the Welch test is used. The test of equality of means is conducted at a 5% significance level. In the following tables, the symbols F(α) and L(α) represent the F-ratio test and the Levene test respectively, tested at the α level of significance in the parentheses.

Tables 4-7 present the simulated null rejection rates, where the proportion of rejections is expressed as a percent for each of the four symmetric distributions (normal, uniform, double exponential, and logistic).

Each entry in the table is the result of ten thousand simulation runs.

For each distribution, the results are given for the seven selected sample size combinations $n_1{:}n_2$ = 5:5, 5:10, 5:20, 10:10, 10:20, 10:40, and 20:20. For each of the seven sample size combinations, the simulated null rejection rate is reported for the specified ratio $R = \sigma_1/\sigma_2$. The R values were selected to yield three cases of direct pairing (R = 0.25, 0.50, and 0.67), three cases of indirect pairing (R = 1.50, 2.00, and 4.00), in addition to the case of equal variance (R = 1). Five testing procedures are evaluated: (1) the t test without the preliminary test for variance homogeneity, denoted as t; (2) the Welch test without the preliminary test for variance homogeneity, denoted as W; (3) the Mann-Whitney-Wilcoxon test without the preliminary test for variance homogeneity, denoted as MW; (4) the preliminary test procedure (described above) using the F-ratio test for testing variance homogeneity at significance levels of 0.05, 0.25, 0.50, and 0.75, denoted as F(α); and (5) the preliminary test procedure using the Levene test for testing variance homogeneity at significance levels of 0.05, 0.25, 0.50, and 0.75 denoted as L(α). A testing procedure is defined to be robust, liberal or conservative as follows:

Robust: If the simulated null rejection rate is > 4% and ≤ 6%.

Conservative: If the simulated null rejection rate is ≤ 4%.

Liberal: If the simulated null rejection rate is > 6%.

3.2 Normal Distribution

The simulated null rejection rates for the normal distribution are presented in Table 4. Results in Table 4 show that the *t* test is generally robust in the case of equal sample sizes regardless of the degree of variance heterogeneity. The *t* test yields liberal results when both samples are size 5 and $R = 0.25$ or 4.00. As cited in the published literature (Chapter I), the *t* test is biased when $R \neq 1$ and the sample sizes are unequal. The *t* test true significance levels tend to be too large (liberal) when the smaller sample size is associated with the larger variance (i.e., indirect pairing) and too small (conservative) when the smaller sample is associated with the smaller variance (i.e., direct pairing). The conservative bias increases as the *R* value decreases, whereas the liberal bias increases as the *R* value increases. The Welch test is robust regardless of the sample size configuration and the variance disparities. The observed simulated null rejection rates resulting from the Welch test ranged from 3.95% to 5.34%. For the case of unequal sample sizes and unequal variances, the Mann-Whitney-Wilcoxon test is a more robust procedure than the *t* test but not as robust as the Welch test. For equal sample sizes, regardless of variance conditions, the *t* test and the Welch test are more robust than the Mann-Whitney-Wilcoxon test. The results in Table 4 support the statements cited in the published literature discussed in Chapter I.

Table 4

Normal Distribution: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	7.35	5.86	5.24	4.97	5.24	5.87	7.31
		F(.05)	5.68	5.53	5.03	4.83	5.06	5.62	5.81
		F(.25)	4.94	4.88	4.60	4.45	4.79	4.99	5.07
		F(.50)	4.85	4.60	4.36	4.24	4.50	4.75	4.95
		F(.75)	4.80	4.53	4.28	4.08	4.48	4.65	4.93
		L(.05)	7.17	5.82	5.18	4.94	5.21	5.85	7.13
		L(.25)	5.38	5.23	4.79	4.74	5.00	5.28	5.52
		L(.50)	4.89	4.74	4.48	4.34	4.61	4.85	5.06
		L(.75)	4.81	4.53	4.32	4.08	4.48	4.67	4.94
		W	4.77	4.48	4.21	3.95	4.35	4.62	4.93
		MW	6.87	6.10	5.69	5.48	5.73	6.40	7.25
5	10	*t*	1.52	2.10	2.82	5.06	8.37	11.41	17.81
		F(.05)	3.94	2.93	3.28	5.15	7.63	9.05	6.58
		F(.25)	4.70	4.18	4.27	5.39	6.29	6.61	4.98
		F(.50)	4.79	4.66	4.65	5.33	5.78	5.64	4.74
		F(.75)	4.81	4.70	4.78	5.17	5.45	5.22	4.68
		L(.05)	2.64	2.66	3.15	5.20	8.20	10.67	11.86
		L(.25)	4.65	4.04	4.27	5.63	7.39	8.13	6.34
		L(.50)	4.78	4.67	4.72	5.65	6.54	6.52	5.08
		L(.75)	4.81	4.72	4.81	5.41	5.80	5.66	4.80
		W	4.81	4.65	4.74	4.96	5.21	5.08	4.65
		MW	2.55	2.54	2.85	4.10	5.63	6.84	8.28

Table 4--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	20	*t*	0.13	0.62	1.36	5.01	12.26	18.85	31.53
		F(.05)	4.30	2.72	2.55	5.31	10.37	11.58	6.69
		F(.25)	4.85	4.50	4.41	6.06	7.85	7.45	5.31
		F(.50)	4.88	4.91	4.88	6.16	6.91	6.00	5.09
		F(.75)	4.88	4.91	4.97	5.69	6.06	5.50	4.99
		L(.05)	2.97	2.08	2.36	5.32	11.50	15.33	10.86
		L(.25)	4.86	4.63	4.52	6.64	9.97	10.60	6.52
		L(.50)	4.88	4.88	5.00	6.72	8.36	8.07	5.57
		L(.75)	4.88	4.90	4.97	6.13	7.13	6.40	5.21
		W	4.88	4.89	4.87	5.23	5.34	5.15	4.92
		MW	0.93	1.54	2.44	4.91	8.72	11.32	17.69
10	10	*t*	5.96	5.36	5.18	4.99	5.10	5.46	6.14
		F(.05)	4.85	4.97	5.07	4.95	4.94	5.00	4.95
		F(.25)	4.83	4.82	4.91	4.90	4.82	4.90	4.91
		F(.50)	4.81	4.79	4.91	4.82	4.78	4.87	4.90
		F(.75)	4.81	4.79	4.89	4.81	4.76	4.86	4.90
		L(.05)	5.02	5.13	5.12	4.97	5.03	5.20	5.21
		L(.25)	4.83	4.85	4.94	4.91	4.88	4.96	4.93
		L(.50)	4.81	4.81	4.92	4.84	4.79	4.88	4.90
		L(.75)	4.81	4.79	4.89	4.80	4.76	4.86	4.90
		W	4.81	4.79	4.89	4.79	4.74	4.86	4.90
		MW	7.76	5.92	5.53	5.14	5.33	5.99	7.95

Table 4--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	1.09	1.78	2.70	4.98	8.57	11.59	16.85
		F(.05)	4.74	4.00	3.65	5.03	7.12	6.95	5.22
		F(.25)	4.76	4.80	4.49	5.20	5.91	5.66	5.18
		F(.50)	4.76	4.87	4.77	5.21	5.48	5.28	5.17
		F(.75)	4.76	4.86	4.90	5.15	5.24	5.24	5.17
		L(.05)	4.56	3.47	3.24	5.03	7.70	8.25	5.60
		L(.25)	4.76	4.68	4.42	5.32	6.36	6.28	5.19
		L(.50)	4.76	4.85	4.77	5.34	5.73	5.55	5.18
		L(.75)	4.76	4.86	4.84	5.20	5.38	5.34	5.17
		W	4.76	4.86	4.88	5.10	5.14	5.23	5.17
		MW	3.11	3.08	3.54	4.94	7.02	8.83	12.00
10	40	*t*	0.07	0.49	1.47	4.85	12.02	18.34	29.24
		F(.05)	4.96	3.96	3.28	5.10	8.35	7.22	4.82
		F(.25)	4.96	4.78	4.56	5.47	6.09	5.41	4.81
		F(.50)	4.96	4.85	4.97	5.43	5.38	4.98	4.81
		F(.75)	4.96	4.86	5.02	5.07	5.10	4.87	4.81
		L(.05)	4.94	3.61	3.07	5.11	9.35	9.19	4.97
		L(.25)	4.96	4.76	4.64	5.69	7.12	6.26	4.81
		L(.50)	4.96	4.86	4.98	5.62	5.96	5.36	4.81
		L(.75)	4.96	4.86	5.02	5.28	5.41	5.02	4.81
		W	4.96	4.86	5.00	4.88	4.91	4.82	4.81
		MW	0.90	1.54	2.59	4.77	8.33	10.65	15.32

Table 4--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
20	20	*t*	5.55	5.11	4.95	5.10	5.23	5.56	5.54
		F(.05)	4.99	4.92	4.89	5.08	5.20	5.21	5.01
		F(.25)	4.99	4.89	4.84	5.06	5.18	5.21	5.01
		F(.50)	4.99	4.88	4.84	5.04	5.18	5.21	5.01
		F(.75)	4.99	4.88	4.84	5.04	5.18	5.21	5.01
		L(.05)	4.99	4.95	4.90	5.07	5.20	5.24	5.01
		L(.25)	4.99	4.89	4.85	5.06	5.18	5.21	5.01
		L(.50)	4.99	4.88	4.84	5.04	5.18	5.21	5.01
		L(.75)	4.99	4.88	4.84	5.03	5.18	5.21	5.01
		W	4.99	4.88	4.84	5.03	5.18	5.21	5.01
		MW	7.83	5.77	5.29	4.72	5.24	5.88	7.48

The results using the preliminary F-ratio and Levene test procedures for normal samples are as follows. In the situation of R = 1 (including balanced and unbalanced sample sizes), all the preliminary F-ratio and Levene test procedures at any significance level are robust except for the n_1:n_2 = 5:20 cases. For this case, only the F(.05) L(.75), and L(.05) test procedures are robust. The preliminary Levene test procedure tends to be more liberal (maximum of 6.72%) than that of the preliminary F-ratio test procedure (maximum of 6.16%) for the α = 0.25 and 0.50 cases.

For the R = 0.67 cases where the sample sizes are equal, the preliminary F-ratio and Levene test procedures at any significance level are robust. For the moderate direct pairing cases where the *t* test and the MW test are too conservative, the F(.05) and the L(.05) test procedure are too conservative as the *t* test is used too often. The preliminary F-ratio and Levene test procedures are robust for moderate direct pairing cases for choices of $\alpha \geq 0.25$.

For the R = 0.25 and 0.50 cases where the sample sizes are equal, the preliminary F-ratio and Levene test procedures at any significance level are robust. For the severe direct pairing cases, the same patterns as described above for the moderate direct pairing cases are seen. The simulated null rejection rates tend to be conservative when the preliminary F-ratio and Levene test procedures are used at a significance level of 5%.

The *t* test and the MW test are both generally liberal for the moderate indirect pairing cases. This is also true when the preliminary F-ratio and Levene tests are used at a 5% or 25% significance level, as the preliminary test procedures use the *t* test too often. For the n_1:n_2 = 5:20 cases, neither the preliminary F-ratio nor the Levene test procedure is robust. It is noted that the preliminary Levene test procedure is slower to counteract the liberalism of the *t* test than the preliminary F-ratio test procedure at any significance level. For the cases of R =1.50 with equal sample sizes, all test procedures are robust.

The *t* test, MW test, F(.05) and L(.05) test procedures are liberal for the severe indirect pairing cases. The F(.25) and the L(.25) test procedures are also liberal for many cases. Again it is noted that the preliminary Levene test procedure is slower to counteract the liberalism of the *t* test than the preliminary F-ratio test procedure at any significance level. The F(.50) and F(.75) test procedures are robust in all unbalanced samples sizes cases. For the R = 2.00 and 4.00 cases with equal sample sizes, all preliminary F-ratio and Levene test procedures are robust with one exception. The 5% preliminary Levene test procedure is liberal in the case of $R = 4$ where $n_1 = n_2 = 5$.

Overall, the F(.05) and the L(.05) test procedures are not robust for the unequal sample size cases as the *t* test is used too often. However, preliminary test procedures conducted at a higher significance level do counteract the bias of the *t* test. The preliminary F-ratio and Levene test procedures for cases where $\alpha > 0.05$ demonstrate robustness in some cases, but none of the procedures are robust for all cases. As previously noted, the preliminary Levene test is slower to counteract the bias of the *t* test than the preliminary F-ratio test procedures at any significance level.

3.3 Uniform Distribution

Table 5 presents the simulated null rejection rates for the uniform distribution which are presented in the same format as those for the normal case in Table 4. The results are very similar to those stated with the normal distribution where the *t* test is robust for cases of unequal variances as long as the sample sizes are equal, but seriously biased when the sample sizes are unequal. The bias observed under conditions of indirect and direct pairing are similar for the uniform cases as those reported for the normal distribution. The MW test is more robust than the *t* test for cases with unequal sample sizes and unequal variances; otherwise the *t* test is more robust than the MW test. The Welch test is generally more robust than the *t* test and the MW test regardless of both the sample size and variance disparities.

The results using the preliminary F-ratio and Levene test procedures are as follows. When the variances are equal, the preliminary F-ratio and Levene test procedures are robust except for the sample size cases of n_1:n_2 = 5:10 and 5:20. For these two cases, the preliminary F-ratio and Levene test procedures tend to be liberal. The preliminary Levene test procedures (maximum of 7.60%) are more liberal than that of the preliminary F-ratio test procedures (maximum of 7.20%), as was the case for normal samples.

For R = 0.67 cases, all preliminary test procedures are robust when the sample sizes are equivalent. For the situations of moderate direct pairing, both the F(.05) and the L(.05) test procedures yield conservative results as the conservative *t* test is used too often. Except for the case of n_1 = 5 and n_2 = 20, the preliminary F-ratio and Levene test procedures at significance levels of 0.50 and 0.75 are robust. These procedures are slightly liberal with a maximum simulated rejection rate of 6.18% for the

Table 5

Uniform Distribution: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	8.51	6.78	5.77	5.31	5.94	6.97	8.89
		F(.05)	7.37	6.52	5.59	5.11	5.75	6.75	7.86
		F(.25)	6.53	6.01	5.27	4.72	5.43	6.28	6.74
		F(.50)	6.39	5.74	4.99	4.49	5.21	6.03	6.66
		F(.75)	6.38	5.67	4.87	4.47	5.14	5.91	6.66
		L(.05)	8.34	6.74	5.75	5.27	5.90	6.95	8.84
		L(.25)	7.11	6.35	5.47	4.96	5.63	6.61	7.48
		L(.50)	6.48	5.97	5.13	4.63	5.33	6.16	6.81
		L(.75)	6.39	5.69	4.89	4.45	5.15	5.95	6.66
		W	6.38	5.61	4.78	4.38	5.05	5.84	6.65
		MW	6.99	6.45	6.07	5.48	6.04	7.07	7.46
5	10	*t*	1.78	2.31	3.03	5.25	9.00	12.44	18.40
		F(.05)	4.24	3.26	3.64	5.45	8.59	11.10	7.30
		F(.25)	5.19	4.41	4.68	5.88	7.78	8.74	6.31
		F(.50)	5.26	4.83	5.05	6.10	7.16	7.34	6.19
		F(.75)	5.26	4.90	5.12	6.07	6.78	6.85	6.17
		L(.05)	2.82	2.83	3.52	5.55	9.00	12.22	14.23
		L(.25)	4.91	4.27	4.71	6.19	8.71	10.66	7.70
		L(.50)	5.24	4.82	5.12	6.44	8.16	8.89	6.44
		L(.75)	5.26	4.90	5.17	6.29	7.28	7.61	6.22
		W	5.26	4.87	5.05	5.88	6.41	6.65	6.15
		MW	2.67	2.38	2.65	4.10	6.40	8.05	8.57

Table 5--Continued

			Ratio = σ_2/σ_1						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	20	*t*	0.19	0.62	1.38	4.93	12.76	19.84	31.09
		F(.05)	4.40	3.08	3.17	5.31	12.11	15.15	7.41
		F(.25)	5.02	5.00	5.55	6.68	10.44	9.56	6.64
		F(.50)	5.03	5.28	6.15	7.20	9.04	7.81	6.57
		F(.75)	5.03	5.31	6.18	7.04	7.85	7.25	6.56
		L(.05)	3.30	3.35	3.92	5.93	12.86	18.88	12.16
		L(.25)	5.02	5.12	5.98	7.53	12.66	14.71	7.52
		L(.50)	5.03	5.29	6.17	7.60	11.30	10.74	6.85
		L(.75)	5.03	5.30	6.11	7.18	9.43	8.39	6.64
		W	5.03	5.31	6.08	6.69	7.05	6.80	6.55
		MW	0.88	1.18	1.68	4.91	10.43	13.63	21.14
10	10	*t*	6.28	5.70	5.34	5.10	5.33	5.61	6.35
		F(.05)	5.25	5.38	5.28	5.03	5.22	5.33	5.17
		F(.25)	5.24	5.25	5.10	4.98	5.10	5.12	5.17
		F(.50)	5.24	5.25	5.06	4.94	4.99	5.10	5.17
		F(.75)	5.24	5.24	5.02	4.89	4.98	5.09	5.17
		L(.05)	5.47	5.54	5.32	5.03	5.33	5.47	5.48
		L(.25)	5.25	5.29	5.11	4.97	5.12	5.21	5.18
		L(.50)	5.24	5.25	5.05	4.95	5.03	5.11	5.17
		L(.75)	5.24	5.25	5.04	4.90	4.98	5.09	5.17
		W	5.24	5.24	5.01	4.88	4.98	5.09	5.17
		MW	8.52	6.30	5.73	5.14	5.81	6.72	8.56

Table 5--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	1.19	1.18	2.62	5.11	8.89	12.17	16.70
		F(.05)	4.95	3.91	3.54	5.19	8.08	7.58	5.35
		F(.25)	4.95	4.80	4.57	5.38	6.51	5.96	5.35
		F(.50)	4.95	4.86	4.91	5.50	5.85	5.62	5.35
		F(.75)	4.95	4.87	4.97	5.46	5.63	5.61	5.35
		L(.05)	4.83	3.63	3.61	5.35	8.45	9.44	5.58
		L(.25)	4.95	4.71	4.60	5.60	7.32	6.72	5.35
		L(.50)	4.95	4.85	4.92	5.65	6.50	5.95	5.35
		L(.75)	4.95	4.87	4.96	5.58	5.96	5.68	5.35
		W	4.95	4.87	4.96	5.36	5.55	5.59	5.35
		MW	3.24	2.79	3.05	4.94	8.19	10.25	12.91
10	40	*t*	0.09	0.45	1.32	4.72	12.43	18.68	29.34
		F(.05)	5.13	4.20	3.22	4.88	10.12	7.05	5.19
		F(.25)	5.13	4.95	4.99	5.40	6.83	5.35	5.19
		F(.50)	5.13	5.00	5.33	5.64	5.74	5.14	5.19
		F(.75)	5.13	5.00	5.34	5.58	5.33	5.11	5.18
		L(.05)	5.13	4.32	4.12	5.31	11.19	9.74	5.21
		L(.25)	5.13	4.97	5.18	5.96	8.50	6.42	5.19
		L(.50)	5.13	5.00	5.33	5.93	7.04	5.49	5.19
		L(.75)	5.13	5.00	5.34	5.60	6.01	5.23	5.19
		W	5.13	5.00	5.32	5.31	5.22	5.11	5.18
		MW	0.83	1.15	1.82	4.77	9.96	12.53	16.29

Table 5--Continued

			Ratio = σ_2/σ_1						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
20	20	*t*	5.71	5.39	5.15	5.13	5.29	5.34	5.64
		F(.05)	5.18	5.12	5.11	5.12	5.19	5.10	5.09
		F(.25)	5.18	5.11	5.05	5.11	5.14	5.09	5.09
		F(.50)	5.18	5.11	5.05	5.10	5.13	5.09	5.09
		F(.75)	5.18	5.11	5.05	5.10	5.13	5.09	5.09
		L(.05)	5.18	5.16	5.10	5.12	5.23	5.12	5.09
		L(.25)	5.18	5.11	5.05	5.10	5.15	5.10	5.09
		L(.50)	5.18	5.11	5.05	5.10	5.13	5.09	5.09
		L(.75)	5.18	5.11	5.04	5.10	5.12	5.09	5.09
		W	5.18	5.11	5.04	5.10	5.12	5.09	5.09
		MW	8.45	6.45	5.64	4.72	5.60	6.32	8.05

case of $n_1 = 5$ and $n_2 = 20$. It is noted that the Welch test is also slightly liberal for this case. The F(.25) and the L(.25) test procedures are robust for all cases.

The preliminary test procedures using the F-ratio test and the Levene test are generally robust for the $R = 0.25$ and 0.50 cases with equal sample sizes. For the $n_1 = 5$ and $n_2 = 5$ case, the preliminary F-ratio and Levene test procedures tend to be liberal. This is not unexpected because both the *t* test and the Welch test tend to be liberal for this case. The L(.05) test procedure is more conservative than the F(.05) test procedure. The other preliminary test procedures for choices of $\alpha \geq 0.25$ are generally robust using either the Levene test or the F-ratio test. It is noted that the preliminary Levene test procedures are slower to counteract the conservatism of the t test than the preliminary F-ratio test procedures.

For the $R = 1.50$ cases with equal sample sizes, all preliminary test procedures are robust. For the n_1:n_2 = 5:10 and 5:20 cases, all preliminary F-ratio and Levene test procedures are liberal. This is expected since both the t test and the Welch test are liberal. As the significance level of the preliminary test of variance is increased, the liberal bias is decreased; however, the bias is always less when using the Welch test alone. For the n_1:n_2 = 10:20 and 10:40 cases, the F(.50) and the F(.75) test procedures are both robust.

For the $R = 2.00$ and 4.00 cases where the sample sizes are equal to 10 or 20, all preliminary test procedures using the F-ratio test and the Levene test are robust. All the preliminary test procedures tend to be liberal when $R = 4$ and $R = 2$ when $n_1 = n_2 = 5$. For the imbalanced cases of n_1:n_2 = 10:20 and 10:40, the preliminary F-ratio test procedures for $\alpha >$ .05 and the preliminary Levene test procedure for $\alpha > 0.25$ are robust. For the other two imbalanced cases: n_1:n_2 = 5:10 and 5:20, both the t test and the Welch test are liberal; although the t test is more liberal than the Welch test. Therefore, both of the preliminary test procedures are liberal, regardless of the significance level used in testing the equality of variances. The preliminary Levene test procedure is slower to counteract the large liberal bias of the t test than is the preliminary F-ratio test procedure.

In summary, the F(.05) and the L(.05) test procedures are not recommended as they are too heavily influenced by the biased t test for the case of unequal sample sizes. The preliminary Levene test procedures are slower to counteract the bias of the t test than the preliminary F-ratio test procedures. The preliminary F-ratio test procedure for $\alpha > .05$, the L(.50) and the L(.75) test procedures are generally robust for those cases where the Welch test is robust. However, for those cases where the t test and the Welch test are liberal, the preliminary F-ratio and Levene test procedures are also liberal.

3.4 Double Exponential Distribution

Simulation results for the double exponential distribution are displayed in Table 6. This table is similar in format to those presented in Table 4 for normal cases. For the equal sample size cases, the t test and the MW test perform well and tend to be slightly more robust than the Welch test, which is sometimes conservative (when $n_1 = n_2 = 5$). It is generally observed that the MW test, like the t test, is conservative for the direct pairing cases and liberal for the indirect pairing cases. However, it is noted for both the direct and indirect pairing cases, the MW test is less biased than the t test. The Welch test is observed to be slightly conservative for both direct and indirect pairing cases.

The results using the preliminary F-ratio and Levene test procedures are now discussed. The F(.05) and L(.05) test procedures are robust as they use the robust t test for the cases of equal variances. The preliminary test procedures where $\alpha > 0.05$ are generally robust except for the n_1:n_2 = 5:5 and 5:10 cases, where they tend to be slightly conservative (minimum simulated null rejection rate of 3.62%).

The results for the R = 0.25, 0.50 and 0.67 cases are similar. For the case of $n_1 = n_2 = 5$, the F(.05) and the L(.05) test procedures are generally robust; whereas, the preliminary test procedures at $\alpha > 0.05$ are liberal. For the case where both sample sizes are 10 or 20 (10:10 and 20:20), all preliminary test procedures are robust. All preliminary test procedures are conservative for the $n_1 = 5$ and $n_2 = 10$ cases, as both the t test and the Welch test are conservative for these cases. For the remaining direct pairing cases, the preliminary F-ratio and Levene test procedures are generally robust when $\alpha > 0.05$. The F(.05) and the L(.05) test procedures are generally conservative.

For the R = 1.50 cases where $n_1 = n_2 = 5$, the 5% and 25% preliminary F-ratio and Levene test procedures are robust, whereas the 50% and 75% preliminary F-ratio and Levene test procedures are slightly conservative (ranging from 3.71%-3.78%). For the other cases of balanced sample sizes, 10:10 and 20:20 with R = 1.50, the preliminary F-ratio and Levene test procedures are robust. For the indirect pairing cases, the preliminary F-ratio test procedures are robust except for the F(.05) test

Table 6

Double Exponential Distribution: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	t	5.40	4.90	4.65	4.56	4.74	5.09	5.69
		F(.05)	3.84	4.34	4.30	4.35	4.50	4.59	4.28
		F(.25)	3.32	3.72	3.81	3.96	4.04	4.01	3.73
		F(.50)	3.18	3.53	3.63	3.76	3.78	3.78	3.62
		F(.75)	3.15	3.46	3.46	3.62	3.73	3.72	3.61
		L(.05)	5.12	4.80	4.63	4.52	4.72	5.02	5.37
		L(.25)	3.71	4.01	4.14	4.21	4.36	4.36	4.05
		L(.50)	3.23	3.64	3.71	3.83	3.93	3.86	3.69
		L(.75)	3.18	3.45	3.52	3.67	3.71	3.73	3.60
		W	3.12	3.35	3.39	3.51	3.62	3.62	3.59
		MW	6.81	6.45	6.06	5.81	6.07	6.40	6.90
5	10	t	1.07	1.77	2.54	4.58	7.59	10.24	16.70
		F(.05)	3.11	2.57	3.01	4.26	5.69	6.35	5.41
		F(.25)	3.76	3.54	3.63	4.21	4.38	4.55	3.79
		F(.50)	3.82	3.86	3.79	4.00	3.93	4.05	3.38
		F(.75)	3.81	3.88	3.83	3.80	3.61	3.75	3.28
		L(.05)	1.90	2.13	2.77	4.52	6.95	8.78	10.15
		L(.25)	3.57	3.23	3.45	4.50	5.34	5.79	4.79
		L(.50)	3.81	3.84	3.83	4.41	4.38	4.57	3.60
		L(.75)	3.83	3.88	3.81	3.94	3.90	3.89	3.30
		W	3.79	3.85	3.69	3.52	3.48	3.53	3.23
		MW	2.32	2.59	2.87	3.86	5.26	6.26	7.96

Table 6--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	20	*t*	0.09	0.75	1.81	5.25	11.51	17.55	30.46
		F(.05)	3.92	2.65	2.94	5.15	7.65	8.48	5.62
		F(.25)	4.53	4.27	4.27	5.16	5.95	5.57	4.19
		F(.50)	4.57	4.56	4.68	4.76	4.99	4.66	3.80
		F(.75)	4.58	4.54	4.60	4.57	4.32	4.11	3.60
		L(.05)	1.54	1.27	2.10	5.06	9.58	12.29	10.27
		L(.25)	4.40	3.82	3.91	5.39	7.52	7.61	5.14
		L(.50)	4.58	4.54	4.70	5.39	5.95	5.59	4.12
		L(.75)	4.58	4.59	4.67	4.88	4.92	4.56	3.73
		W	4.58	4.48	4.34	4.14	3.93	3.79	3.45
		MW	1.09	2.09	3.13	5.20	7.63	9.71	14.68
10	10	*t*	5.14	4.54	4.42	4.41	4.57	4.84	5.18
		F(.05)	4.04	4.19	4.23	4.28	4.34	4.41	3.98
		F(.25)	4.02	4.04	4.13	4.19	4.19	4.31	3.94
		F(.50)	4.02	4.03	4.10	4.15	4.14	4.30	3.94
		F(.75)	4.02	4.03	4.07	4.14	4.13	4.29	3.94
		L(.05)	4.29	4.35	4.33	4.39	4.51	4.62	4.35
		L(.25)	4.04	4.10	4.17	4.23	4.29	4.34	3.94
		L(.50)	4.02	4.03	4.11	4.17	4.17	4.30	3.94
		L(.75)	4.02	4.02	4.09	4.15	4.14	4.29	3.94
		W	4.02	4.01	4.06	4.10	4.11	4.27	3.94
		MW	6.92	5.49	5.03	4.93	4.97	5.46	6.89

Table 6--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	0.83	1.66	2.56	4.93	7.87	10.56	16.01
		F(.05)	4.33	3.73	3.81	4.72	5.52	5.65	4.51
		F(.25)	4.39	4.40	4.50	4.78	4.85	4.59	4.35
		F(.50)	4.39	4.54	4.59	4.59	4.52	4.45	4.29
		F(.75)	4.39	4.55	4.63	4.56	4.35	4.28	4.27
		L(.05)	3.61	2.58	3.09	4.86	6.57	7.05	5.27
		L(.25)	4.35	4.12	4.01	4.79	5.05	5.04	4.39
		L(.50)	4.39	4.48	4.57	4.82	4.57	4.51	4.29
		L(.75)	4.39	4.54	4.64	4.56	4.42	4.33	4.26
		W	4.39	4.55	4.61	4.51	4.31	4.23	4.26
		MW	3.08	3.25	3.69	4.76	6.18	7.55	10.51
10	40	*t*	0.07	0.57	1.80	5.36	11.90	17.51	29.31
		F(.05)	4.74	3.91	3.67	5.06	6.81	6.68	4.62
		F(.25)	4.77	4.54	4.62	5.13	5.38	5.11	4.44
		F(.50)	4.77	4.65	4.84	5.01	4.92	4.72	4.42
		F(.75)	4.77	4.68	4.78	4.91	4.68	4.54	4.42
		L(.05)	4.04	1.99	2.41	5.09	8.44	8.70	5.00
		L(.25)	4.76	4.28	4.02	5.13	6.13	5.78	4.51
		L(.50)	4.77	4.60	4.72	5.25	5.05	5.00	4.42
		L(.75)	4.77	4.69	4.79	4.99	4.74	4.62	4.41
		W	4.77	4.68	4.68	4.74	4.55	4.51	4.41
		MW	1.09	2.13	2.92	5.15	7.73	9.73	14.24

Table 6--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
20	20	*t*	5.15	4.90	4.93	4.79	4.94	4.94	5.13
		F(.05)	4.51	4.61	4.80	4.71	4.84	4.71	4.52
		F(.25)	4.51	4.59	4.80	4.67	4.81	4.71	4.52
		F(.50)	4.51	4.59	4.80	4.67	4.81	4.71	4.52
		F(.75)	4.51	4.59	4.80	4.67	4.79	4.71	4.52
		L(.05)	4.51	4.69	4.86	4.77	4.87	4.81	4.52
		L(.25)	4.51	4.59	4.80	4.71	4.83	4.72	4.52
		L(.50)	4.51	4.59	4.80	4.67	4.82	4.71	4.52
		L(.75)	4.51	4.59	4.80	4.67	4.82	4.71	4.52
		W	4.51	4.59	4.80	4.67	4.79	4.71	4.52
		MW	6.66	5.28	5.00	4.82	4.88	5.34	6.66

procedure. The F(.05) test procedure uses the liberal *t* test too often. This liberal bias is also seen when using the L(.05) and the L(.25) test procedures.

Results for the R = 2.00 cases are similar to those stated above for the R = 1.50 cases. The results for the cases of R = 4 are as follows. When the sample sizes are equal, all preliminary test procedures using the F-ratio test and the Levene test are generally robust (minimum of 3.6%). The preliminary F-ratio and Levene test procedures are robust for the 10:20 and 10:40 cases. For the 5:10 cases, only the F(.05) test procedure and the L(.25) test procedures are robust. For the 5:20 cases with R = 4.00, the

F(.05), F(.25), L(.25), and L(.50) test procedures yield robust results.

In summary, the *t* test is conservative for cases with direct pairing. The F(.05) and the L(.05) test procedures are generally conservative, as they use the *t* test too often. The preliminary test procedures which use significance levels $\alpha \geq 0.25$ are more robust.

For the indirect pairing cases, the liberal bias of the *t* test is observed to increase as R increases. This observation is in agreement with the results published in the literature (Chapter I). The Welch test is somewhat conservative for the indirect pairing cases. As a result, the 25% and 50% preliminary test procedures generally yield robust results, as they are not too heavily influenced by the liberal *t* test or the conservative Welch test.

For the balanced sample size cases where $n_1 = n_2 > 5$, all procedures are generally robust.

3.5 Logistic Distribution

Using the same format as in the normal case, the simulated null rejection rates for the logistic distribution are displayed in Table 7. For each sample size and variance configuration, the Welch test is generally robust. The *t* test and MW test are generally conservative for the direct pairing cases and liberal for the indirect pairing cases. The MW test rejects too frequently for R values of 0.25, 0.50, 2.00, and 4.00 when the sample sizes are balanced. The *t* test is somewhat liberal in the severe indirect pairing cases with R = 2.00 and 4.00. For the R = 1 cases, the t test, Welch test, and MW test are generally robust.

Results using the preliminary F-ratio and Levene test procedures are now discussed. For the equal variance cases, regardless of sample size, all preliminary test procedures at any significance level are robust, with one exception. Only the 5% and 25% preliminary test procedures using the F-ratio test and the Levene test are robust for the $n_1 = n_2 = 5$ case.

For the R = 0.67 cases with $n_1 = n_2$, all preliminary F-ratio and Levene test procedures are robust. For the moderate direct pairing cases, all preliminary F-ratio procedures are again robust with the exception of the F(.05) test procedures, which uses the conservative t test too often. The preliminary Levene test procedures where $\alpha > 0.05$ are also generally robust.

For the R = 0.25 and 0.50 cases with $n_1 = n_2$, all preliminary F-ratio and Levene test procedures are robust with the exception of the L(.05) test procedure, which is liberal for the case $n_1 = n_2 = 5$. For the severe direct pairing cases, all preliminary F-ratio and Levene test procedures are again generally robust with the exception of the F(.05) and the L(.05) test procedures, which use the conservative t test too often.

For the moderate indirect pairing cases, the F(.05) and the L(.05) test procedures yield liberal results due to the influence of the liberal t test. This bias is also seen for these cases when using the L(.25) test procedure.

Table 7

Logistic Distribution: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	6.65	5.38	4.88	4.70	5.03	5.80	7.22
		F(.05)	5.12	4.82	4.70	4.53	4.78	5.39	5.65
		F(.25)	4.48	4.37	4.26	4.07	4.31	4.81	5.08
		F(.50)	4.42	4.09	4.13	3.80	4.16	4.55	4.98
		F(.75)	4.40	4.00	4.04	3.63	4.05	4.44	4.95
		L(.05)	6.35	5.32	4.86	4.66	5.02	5.76	6.98
		L(.25)	4.89	4.66	4.60	4.36	4.61	5.16	5.60
		L(.50)	4.46	4.20	4.23	3.92	4.24	4.70	5.06
		L(.75)	4.42	4.05	4.02	3.66	4.08	4.46	4.96
		W	4.40	3.96	3.91	3.53	3.90	4.38	4.92
		MW	7.12	6.05	5.63	5.63	5.95	6.48	7.52
5	10	*t*	1.43	2.05	2.96	4.87	8.26	11.16	17.35
		F(.05)	3.74	2.94	3.40	4.68	6.93	7.86	6.18
		F(.25)	4.56	3.87	4.15	4.71	5.65	5.57	4.61
		F(.50)	4.67	4.26	4.39	4.64	4.89	4.93	4.32
		F(.75)	4.69	4.30	4.29	4.49	4.65	4.61	4.19
		L(.05)	2.46	2.43	3.21	4.89	7.84	10.04	11.44
		L(.25)	4.38	3.82	4.05	5.02	6.63	7.17	5.76
		L(.50)	4.69	4.28	4.47	4.95	5.69	5.70	4.73
		L(.75)	4.69	4.32	4.37	4.73	5.02	4.99	4.34
		W	4.69	4.22	4.21	4.33	4.43	4.42	4.15
		MW	2.48	2.67	3.04	3.96	5.53	6.66	8.38

Table 7--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	20	*t*	0.15	0.71	1.60	4.69	11.52	17.50	30.59
		F(.05)	4.25	2.74	2.91	4.85	8.96	9.94	6.31
		F(.25)	4.74	4.36	4.63	5.64	6.69	6.91	4.95
		F(.50)	4.77	4.63	4.95	5.55	5.78	5.81	4.62
		F(.75)	4.77	4.66	4.88	5.30	5.17	5.10	4.49
		L(.05)	2.41	1.63	2.16	4.88	10.45	13.41	11.08
		L(.25)	4.69	4.23	4.44	5.93	8.61	9.41	5.96
		L(.50)	4.76	4.67	5.07	6.13	7.27	7.24	4.98
		L(.75)	4.76	4.67	4.86	5.52	6.00	5.93	4.68
		W	4.76	4.62	4.69	4.87	4.75	4.67	4.45
		MW	0.94	1.88	2.93	5.17	8.12	10.44	16.32
10	10	*t*	6.01	5.29	5.20	5.07	5.17	5.45	6.14
		F(.05)	4.95	5.05	5.05	4.96	5.00	5.04	4.97
		F(.25)	4.92	4.87	4.93	4.84	4.89	4.84	4.92
		F(.50)	4.91	4.83	4.86	4.82	4.86	4.79	4.92
		F(.75)	4.91	4.82	4.86	4.79	4.84	4.79	4.92
		L(.05)	5.19	5.15	5.14	5.02	5.06	5.27	5.17
		L(.25)	4.92	4.94	4.96	4.91	4.97	4.92	4.95
		L(.50)	4.91	4.83	4.88	4.86	4.89	4.80	4.92
		L(.75)	4.90	4.81	4.86	4.83	4.85	4.79	4.92
		W	4.90	4.81	4.86	4.77	4.82	4.79	4.92
		MW	7.50	6.15	5.76	5.54	5.69	6.24	8.38

Table 7--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	1.14	1.93	2.85	5.33	8.37	11.02	15.81
		F(.05)	5.05	4.20	4.00	5.38	6.54	6.44	4.99
		F(.25)	5.12	4.91	4.88	5.30	5.57	5.36	4.91
		F(.50)	5.13	5.05	5.07	5.24	5.24	5.12	4.88
		F(.75)	5.13	5.05	5.12	5.17	5.14	5.01	4.87
		L(.05)	4.72	3.27	3.48	5.38	7.26	7.68	5.42
		L(.25)	5.09	4.87	4.56	5.44	5.94	5.82	4.95
		L(.50)	5.13	5.05	5.10	5.39	5.51	5.27	4.89
		L(.75)	5.13	5.03	5.14	5.28	5.21	5.07	4.87
		W	5.13	5.04	5.04	5.16	5.08	5.01	4.87
		MW	3.57	3.41	3.78	4.94	6.82	8.09	11.28
10	40	*t*	0.06	0.53	1.56	4.76	11.14	17.52	29.38
		F(.05)	4.85	3.88	3.67	4.90	7.39	6.90	4.57
		F(.25)	4.85	4.54	4.73	5.16	5.81	5.09	4.55
		F(.50)	4.85	4.68	5.00	5.20	5.19	4.75	4.53
		F(.75)	4.85	4.70	5.00	5.09	4.93	4.61	4.53
		L(.05)	4.70	2.82	2.82	4.86	8.64	8.72	4.92
		L(.25)	4.85	4.49	4.58	5.31	6.46	5.83	4.57
		L(.50)	4.85	4.66	5.02	5.51	5.50	5.00	4.54
		L(.75)	4.85	4.69	5.01	5.25	5.05	4.71	4.52
		W	4.85	4.69	4.93	4.90	4.72	4.56	4.52
		MW	0.69	1.60	2.55	4.80	7.61	9.96	14.80

Table 7--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
20	20	*t*	5.70	5.58	5.49	5.31	5.16	5.15	5.59
		F(.05)	5.16	5.37	5.45	5.31	5.03	4.93	4.87
		F(.25)	5.16	5.35	5.43	5.29	5.02	4.91	4.87
		F(.50)	5.16	5.35	5.43	5.26	5.01	4.90	4.87
		F(.75)	5.16	5.35	5.43	5.25	5.01	4.90	4.87
		L(.05)	5.16	5.37	5.97	5.30	5.05	4.98	4.87
		L(.25)	5.16	5.36	5.43	5.30	5.02	4.93	4.87
		L(.50)	5.16	5.35	5.43	5.28	5.01	4.90	4.87
		L(.75)	5.16	5.35	5.43	5.26	5.01	4.90	4.87
		W	5.16	5.35	5.43	5.25	5.01	4.90	4.87
		MW	7.64	6.13	5.37	5.04	5.51	5.81	7.37

However, the F(.50) test procedure yields robust results. For the R = 1.50 cases with balanced sample sizes, all preliminary test procedures are robust regardless of the significance level used.

For the R = 2 cases with indirect pairing, the conclusions are the same as stated above for the moderate indirect pairing cases. For the R = 4 cases, with the exception of the F(.05) and the L(.05) test procedures, all procedures are robust regardless of the sample size configuration.

In summary, conducting a preliminary test procedure using either the F(.05) or the L(.05) test procedure does not counteract the bias resulting from the t test, and for some cases the influence of the bias of the t test is seen in the preliminary test procedure at a significance level of 0.25. However, the preliminary test procedures are robust for cases where $\alpha >$ 0.25.

3.6 Overall Performance for the Symmetric Distributions

In order to get a clearer picture of the overall performance of the procedures for varying degrees of variance heterogeneity, the results of the simulation for the four symmetric distributions are combined in Tables 8-12. Five groupings are defined depending on $R = \sigma_1/\sigma_2$. These are: (1) R = 0.25 and 0.50, (2) R = 0.67, (3) R = 1 (equal variance), (4) R = 1.5, and (5) R = 2.0 and 4.0. The proportion of rejections is expressed as a percent for the t test, the Welch test, the Mann-Whitney-Wilcoxon test, the preliminary F-ratio test procedures and the preliminary Levene test procedures. These proportions are tabulated for each R grouping over all (28) combinations of sample size pairs (7) and distributions (4) for the five categories listed below:

1. $x \leq 2.5$
2. $2.5 < x \leq 4.0$
3. $4.0 < x \leq 6.0$
4. $6.0 < x \leq 10.0$
5. $x > 10.0$

The "x" represents the percentage of rejections for testing H_o: $\mu_1 = \mu_2$

based on 10,000 simulations for each sample size configuration. Each entry in the following tables denotes the frequency at which $a < x \leq b$ occurs.

The outcome of a test procedure is defined as follows:

Extremely Conservative: If the simulated null rejection rate is ≤ 2.5.

Conservative: If the simulated null rejection rate is > 2.5 and ≤ 4.0.

Robust: If the simulated null rejection rate is > 4.0 and ≤ 6.0.

Liberal: If the simulated null rejection rate is > 6.0 and ≤ 10.0.

Extremely Liberal: If the simulated null rejection is > 10.0.

R = 1 (Equal Variance Cases)

A summary of the simulated rejection rates for the four symmetric distributions with the condition of equal variances is presented in Table 8. Table 8 shows, as anticipated, the *t* test is robust (all simulated rejection rates of > 4% and ≤ 6%). However, the preliminary F-ratio and Levene test procedures at all significance levels, the MW test, and the Welch test are also generally robust tests. None of the tests examined show simulated rejection rates ≤ 2.5% or > 10%. The F(.05) and the L(.05) test procedures are more robust than the preliminary test procedures where $\alpha > 0.05$ and are more robust than the Welch test and the MW test.

R = 0.67 (Includes the Moderate Direct Pairing Cases)

Table 9 summarizes the simulated rejection rates for the four

Table 8

Summary of Symmetric Distributions: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level Equal Variance: $R=\sigma_1/\sigma_2 = 1.00$

test	$x \leq 2.5$	$2.5< x \leq 4$	$4< x \leq 6$	$6< x \leq 10$	$x >10$	Total
t	0	0	28	0	0	28
F(.05)	0	0	28	0	0	28
F(.25)	0	1	25	2	0	28
F(.50)	0	3	22	3	0	28
F(.75)	0	3	23	2	0	28
L(.05)	0	0	28	0	0	28
L(.25)	0	0	25	3	0	28
L(.50)	0	2	22	4	0	28
L(.75)	0	3	22	3	0	28
W	0	4	23	1	0	28
MW	0	2	26	0	0	28

symmetric distributions for the $R = 0.67$ cases. Table 9 shows the *t* test, the MW test and the L(.05) test procedure can be extremely conservative. The F(.05) test procedure tends to be conservative. All other preliminary test procedures and the Welch test are robust for the $R = 0.67$ cases.

$R = 0.25$ and 0.50 (Includes the Severe Direct Pairing Cases)

The simulated null rejection rates for the four symmetric distributions for the $R = 0.25$ and 0.50 cases are summarized in Table 10.

The *t* test shows the most evidence of extreme conservatism for these

Table 9

Summary of Symmetric Distributions: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level $R=\sigma_1/\sigma_2$ = 0.67

test	$x \leq 2.5$	$2.5< x \leq 4$	$4< x \leq 6$	$6< x \leq 10$	$x >10$	Total
t	8	8	12	0	0	28
F(.05)	0	16	12	0	0	28
F(.25)	0	2	26	0	0	28
F(.50)	0	2	25	1	0	28
F(.75)	0	2	25	1	0	28
L(.05)	4	11	13	0	0	28
L(.25)	0	2	26	0	0	28
L(.50)	0	2	25	1	0	28
L(.75)	0	2	25	1	0	28
W	0	3	24	1	0	28
MW	3	13	10	2	0	28

cases. The MW test shows both liberal and extreme conservative trends. The F(.05) and the L(.05) test procedures are conservative as they use the *t* test too often. The Welch test is generally robust but can be slightly conservative for certain cases. All other preliminary test procedures are roughly comparable to the Welch test with respect to robustness.

R = 1.50 (Includes the Moderate Indirect Pairing Cases)

Table 11 summarizes the simulated null rejection rates for the four symmetric distributions for the *R* = 1.50 cases. The *t* test is severely

Table 10

Summary of Symmetric Distributions: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level $R=\sigma_1/\sigma_2$ = 0.25 and 0.50

test	$x \leq 2.5$	$2.5< x \leq 4$	$4< x \leq 6$	$6< x \leq 10$	$x > 10$	Total
t	32	0	18	6	0	56
F(.05)	0	19	35	2	0	56
F(.25)	0	5	49	2	0	56
F(.50)	0	4	51	1	0	56
F(.75)	0	5	50	1	0	56
L(.05)	10	14	28	4	0	56
L(.25)	0	5	49	2	0	56
L(.50)	0	4	51	1	0	56
L(.75)	0	4	51	1	0	56
W	0	5	50	1	0	56
MW	19	13	4	20	0	56

liberal. The MW test, F(.05), L(.05), F(.25), and L(.25) test procedures are somewhat liberal, but less liberal than the *t* test. The F(.50), F(.75), L(.50), L(.75) and the Welch procedures are each reasonably robust.

R = 2.00 and 4.00 (Includes the Severe Indirect Pairing Cases)

The simulated null rejection rates for the four symmetric distributions for the R = 2.00 and 4.00 cases are summarized in Table 12. The *t* test, L(.05) test procedure, and the MW test are severely liberal. The F(.05) and the L(.25) test procedures are liberal. It is noted that as the

Table 11

Summary of Symmetric Distributions: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level $R=\sigma_1/\sigma_2 = 1.50$

test	$x \le 2.5$	$2.5 < x \le 4$	$4 < x \le 6$	$6 < x \le 10$	$x > 10$	Total
t	0	0	12	8	8	28
F(.05)	0	0	14	11	3	28
F(.25)	0	0	20	7	1	28
F(.50)	0	2	23	3	0	28
F(.75)	0	2	23	3	0	28
L(.05)	0	0	12	12	4	28
L(.25)	0	0	15	12	1	28
L(.50)	0	1	20	6	1	28
L(.75)	0	2	22	4	0	28
W	0	4	22	2	0	28
MW	0	0	13	14	1	28

significance level of the preliminary test is increased, the liberal bias is decreased; and that the preliminary Levene test procedure is slower to counteract the liberal *t* test than the preliminary F-ratio test procedure. The F(.50), F(.75), L(.50), L(.75), and the Welch procedures are each robust, with little evidence of extreme liberalism or extreme conservatism.

3.7 Summary

In summary in the case of variance homogeneity, the *t* test is the most robust test. However, as seen in Table 8, all of the test procedures

Table 12

Summary of Symmetric Distributions: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level $R = \sigma_1/\sigma_2$ = 2.0 and 4.0

test	$x \leq 2.5$	$2.5 < x \leq 4$	$4 < x \leq 6$	$6 < x \leq 10$	$x > 10$	Total
t	0	0	17	7	32	56
F(.05)	0	1	32	20	3	56
F(.25)	0	3	44	9	0	56
F(.50)	0	5	45	6	0	56
F(.75)	0	6	45	5	0	56
L(.05)	0	0	28	13	15	56
L(.25)	0	1	38	14	3	56
L(.50)	0	4	43	8	1	56
L(.75)	0	6	44	6	0	56
W	0	7	44	5	0	56
MW	0	0	5	33	18	56

evaluated are generally robust for the equal variance cases. For the R = 0.25, 0.50, and 0.67 cases, where the *t* test is conservative, the F(.05) and the L(.05) test procedures are also conservative. The MW test yields conservative results in the case of unequal sample sizes and liberal results in the case of equal sample sizes. The preliminary test procedures where $\alpha > 0.05$ and the Welch test tend to be generally robust.

For the R = 1.50, 2.00 and 4.00 cases, where the *t* test and the MW test are extremely liberal, the F(.05) and the L(.05) test procedures are also liberal. It is noted that as the significance level of the preliminary test procedures is increased, the liberal bias is decreased. However, the preliminary Levene test procedures are slower to counteract the liberalism of the *t* test than the preliminary F-ratio test procedures. The preliminary test procedures where $\alpha > 0.25$ and the Welch test are each generally robust.

Based on the above simulation results, the 50% and 75% preliminary test procedures and the Welch test are generally robust procedures with respect to variance heterogeneity for testing H_o: $\mu_1 = \mu_2$ for the general families of symmetric distributions examined.

CHAPTER IV

ASYMMETRIC DISTRIBUTIONS

4.1 Introduction

This chapter contains an examination of the performance of the preliminary variance equality test procedures for the two asymmetric distributions, the lognormal and the gamma,. with varying degrees of skewness. Presented are the simulated null rejection rates for testing the hypothesis H_o: $\mu_1 = \mu_2$ after utilizing the two types of preliminary tests of variance homogeneity, the F-ratio test and the Levene test, for each of the asymmetric distributions.

Preliminary testing is handled in the following manner. When the preliminary test is found not to be significant at the specified level, the Mann-Whitney-Wilcoxon test is employed; otherwise the Welch test is used. The test of equality of means is conducted at a significance level of 5%. In the following tables, the symbols F(α) and L(α) represent the F-ratio test and the Levene test respectively, tested at the α level of significance in the parentheses.

Tables 13-17 present the simulated null rejection rates, where the proportion of rejections is expressed as a percent for each of the five classes of asymmetric distributions (three lognormals and two gamma). Each entry in the table is again the result of ten thousand simulation runs. The format of Tables 13-17 is the same as described in Chapter III for Tables 4-7. For each distribution, the results are given for each of the seven selected sample size configurations. For each of the seven sample size combinations, the simulated null rejection rate is reported for the three cases of direct pairing (R = 0.25, 0.50, and 0.67), three cases of indirect pairing (R = 1.50, 2.00, and 4.00), in addition to the case of variance homogeneity (R = 1). The test of variance homogeneity in the preliminary test procedures is conducted at significant levels of 0.05, 0.25, 0.50, and 0.75, which are the same levels previously used for the symmetric cases.

Based on the observed simulated null rejection rate, a testing procedure is robust, liberal, or conservative as defined in Chapter III.

4.2 Lognormal Distributions

To evaluate the performance of the preliminary test procedures for the lognormal case, three lognormal distributions were chosen: one with slight skewness (b_2 = 0.4), one with moderate skewness (b_2 = 1.0), and one with extreme skewness (b_2 = 1.75). Each of the three lognormal distributions are discussed separately in the following sections.

4.2.1 Lognormal: Slight Skewness

Table 13 presents the simulated results for the lognormal (0, 0.40) distribution. For the equal variance cases, where $n_1 = n_2 = 10$ or $n_1 = n_2 = 20$, the t test, the Welch test, the MW test, as well as the preliminary F-ratio and Levene test procedures are robust. However for the $n_1 = n_2 = 5$ cases, the t test, the MW test, the F(.05), the L(.05), and L(.25) test procedures are robust; whereas the other preliminary F-ratio and Levene test procedures tend to be conservative as the conservative Welch test is often used. For the cases where n_2 is twice as large as n_1 (5:10 and 10:20), all test procedures are robust. For the cases where n_2 is four times as large as n_1 (5:20 and 10:40), the only robust preliminary test procedures are the F(.05) and the L(.05) test procedures.

For the $R = 0.67$ cases where the sample sizes are equivalent, the Welch test, the t test, the preliminary F-ratio test procedures where $\alpha > 0.05$, and the preliminary Levene test procedures where $\alpha > 0.50$ are robust. The MW test and the other preliminary test procedures are liberal. For the moderate direct pairing cases, the Welch test and the preliminary test procedures where $\alpha > 0.05$ are generally robust. The t test and the MW test are generally both conservative.

For the $R = 0.50$ cases where sample sizes are balanced, the Welch test, the F(.50), the F(.75) and the L(.75) test procedures are generally robust. The MW test and the other preliminary test procedures are liberal. For the severe direct pairing cases, the Welch test is robust and the MW test is generally conservative. For some situations, the F(.05) and the L(.05) test procedures are conservative, but generally the preliminary test procedures are robust.

For the R = 0.25 cases where the sample sizes are balanced, all procedures are liberal. For the severe direct pairing cases, the Welch test and the preliminary test procedures, excluding the L(.05) test procedure, are generally robust. The MW test and the L(.05) test procedure are somewhat conservative.

For the R = 1.50 cases where the sample sizes are equal, the MW test is liberal. The t test, the Welch test, the F(.50) and the F(.75) test procedures are generally robust. The preliminary Levene test procedures at each significance level yield liberal results in several instances. For the moderate indirect pairing cases, all procedures are generally liberal.

For the R = 2.0 and 4.0 cases, regardless of sample size configuration, all procedures are generally liberal.

In summary, for the R = 0.67, 0.50, and 0.25 cases, the Welch test, the F(.50), the F(.75), and the L(.75) are generally robust. For the R = 1.50 cases where the sample sizes are balanced, the t test, the Welch test, the F(.50) and the F(.75) test procedures are generally robust. For the moderate indirect pairing and for R = 2.0 and 4.0 cases, all procedures are

Table 13

Lognormal (0, 0.40) Distribution: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	8.88	6.30	5.23	4.45	5.22	6.74	16.36
		F(.05)	7.66	6.61	5.56	4.64	5.39	7.16	16.55
		F(.25)	7.13	6.08	4.87	3.96	4.78	6.57	15.63
		F(.50)	6.91	5.75	4.53	3.57	4.40	6.03	15.17
		F(.75)	6.89	5.49	4.31	3.40	4.14	5.71	14.95
		L(.05)	8.39	7.01	5.97	5.16	6.07	7.83	17.16
		L(.25)	7.56	6.58	5.50	4.56	5.37	7.16	16.53
		L(.50)	7.03	6.06	4.76	3.73	4.57	6.37	15.63
		L(.75)	6.94	5.60	4.28	3.42	4.24	5.83	14.93
		W	6.88	5.19	4.08	3.21	3.91	5.44	14.40
		MW	8.51	7.06	5.99	5.21	6.13	7.86	17.23
5	10	*t*	2.76	3.13	3.54	4.68	7.75	10.72	22.70
		F(.05)	4.97	3.69	3.80	4.38	6.08	8.53	19.74
		F(.25)	5.81	4.28	4.24	4.98	6.39	8.90	19.78
		F(.50)	6.07	4.58	4.42	5.01	6.72	9.19	19.86
		F(.75)	6.06	4.72	4.39	4.97	6.73	9.24	19.77
		L(.05)	4.83	3.65	3.75	4.49	6.40	8.79	19.89
		L(.25)	6.31	4.65	4.52	5.28	6.83	9.06	19.99
		L(.50)	6.20	4.82	4.71	5.33	6.96	9.42	20.14
		L(.75)	6.07	4.86	4.58	5.10	6.84	9.43	20.24
		W	6.07	4.73	4.42	4.95	6.67	9.17	19.80
		MW	3.96	3.21	3.38	4.19	6.31	8.84	19.90

Table 13--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	20	*t*	0.43	1.13	2.03	5.08	10.34	16.26	32.37
		F(.05)	4.54	4.26	4.65	5.94	9.03	13.09	27.36
		F(.25)	5.23	4.90	5.68	7.21	9.69	13.31	25.35
		F(.50)	5.23	4.94	5.44	7.10	9.77	13.10	23.79
		F(.75)	5.24	4.80	5.30	6.79	9.39	12.49	22.39
		L(.05)	4.07	3.38	3.84	5.53	9.22	13.90	31.72
		L(.25)	5.55	5.28	5.89	7.26	10.11	14.06	29.25
		L(.50	5.26	5.10	5.85	7.43	10.11	13.96	27.34
		L(.75)	5.25	4.84	5.35	6.92	9.58	12.89	24.59
		W	5.24	4.58	4.95	6.40	8.82	11.62	21.16
		MW	2.26	2.48	3.09	5.19	9.26	14.27	32.19
10	10	*t*	7.24	6.24	5.67	4.68	5.45	7.42	16.01
		F(.05)	6.33	6.86	6.30	5.21	6.38	9.58	21.00
		F(.25)	6.20	6.14	6.01	4.91	5.93	8.50	17.97
		F(.50)	6.20	5.92	5.70	4.61	5.52	7.91	16.65
		F(.75)	6.21	5.78	5.40	4.46	5.26	7.49	15.95
		L(.05)	9.93	7.92	6.76	5.50	7.02	11.26	31.39
		L(.25)	6.69	7.38	6.58	5.35	6.86	10.85	29.41
		L(.50)	6.26	6.49	6.17	4.95	6.26	9.65	21.91
		L(.75)	6.22	6.02	5.72	4.56	5.70	8.21	17.80
		W	6.21	5.78	5.29	4.23	4.94	6.96	15.47
		MW	10.98	7.86	6.63	5.35	6.93	11.18	31.36

Table 13--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	2.01	2.74	3.30	4.93	8.11	11.81	22.83
		F(.05)	5.83	5.77	5.72	5.49	8.05	12.10	22.70
		F(.25)	5.88	5.35	5.48	5.82	7.92	10.90	19.35
		F(.50)	5.88	5.15	5.25	5.53	7.69	10.18	18.16
		F(.75)	5.88	5.05	4.98	5.38	7.30	9.55	17.59
		L(.05)	8.13	6.61	5.69	5.38	8.72	14.79	39.79
		L(.25)	5.97	6.33	6.01	5.92	8.80	13.74	29.41
		L(.50)	5.88	5.49	5.64	5.88	8.51	12.08	22.34
		L(.75)	5.88	5.20	5.15	5.55	7.71	10.68	19.03
		W	5.88	5.01	4.77	5.16	6.95	9.04	17.38
		MW	7.58	5.87	5.04	5.06	8.81	14.96	40.23
10	40	*t*	0.40	0.96	1.79	4.64	11.21	17.05	32.87
		F(.05)	5.19	5.41	5.76	5.86	9.28	13.71	23.65
		F(.25)	5.19	5.15	5.80	6.65	9.09	11.98	20.20
		F(.50)	5.19	5.06	5.44	6.33	8.74	11.07	18.99
		F(.75)	5.19	4.95	5.19	6.07	8.20	10.43	18.67
		L(.05)	6.24	5.75	5.04	5.34	10.26	17.74	40.79
		L(.25)	5.20	5.58	6.17	6.60	10.23	15.72	27.71
		L(.50)	5.19	5.10	5.68	6.52	9.79	13.63	22.30
		L(.75)	5.19	4.99	5.26	6.18	8.87	11.70	19.75
		W	5.19	4.89	4.94	5.66	7.63	9.80	18.46
		MW	4.49	3.76	3.63	4.85	10.57	18.50	46.87

Table 13--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
20	20	t	6.88	5.81	5.30	4.98	5.30	6.64	13.39
		F(.05)	6.30	6.59	6.39	5.53	7.31	10.92	16.60
		F(.25)	6.30	5.77	5.74	5.55	6.52	8.65	14.07
		F(.50)	6.30	5.58	5.45	5.26	5.90	7.64	13.50
		F(.75)	6.30	5.51	5.21	5.11	5.49	6.96	13.24
		L(.05)	7.04	9.52	7.41	5.44	8.44	16.50	50.47
		L(.25)	6.31	6.85	6.81	5.77	8.11	13.60	27.63
		L(.50)	6.30	5.98	6.00	5.62	7.06	10.16	15.96
		L(.75)	6.30	5.62	5.47	5.28	6.15	8.06	13.77
		W	6.30	5.52	5.11	4.79	5.13	6.50	13.19
		MW	14.98	9.81	7.06	4.97	7.85	16.23	52.14

generally liberal.

4.2.2 Lognormal: Moderate Skewness

The null rejection rates for the lognormal (0, 1) distribution are shown in Table 14. The results for the cases of variance homogeneity, direct pairing, and indirect pairing are now discussed. For the variance homogeneity and balanced sample size cases, the MW test is robust, whereas the t test and the Welch test are conservative. None of the preliminary F-ratio or Levene test procedures yield robust results for every case, but the L(.05) test procedure is the most robust with a maximum Type

I error rate of 6.11%. For the case where the sample sizes differ, the MW test is robust, whereas the Welch test results are inconsistent. The *t* test is robust for some cases but conservative in others. The L(.05) test procedure yields robust results for all cases. This is not unexpected because it primarily uses the robust MW test. The other preliminary F-ratio and Levene test procedures are comparable to the Welch test.

For the $R = 0.67$ cases with equal sample sizes, the *t* test and the Welch test are robust except for the $n_1 = n_2 = 5$ case. The MW test is always liberal for these cases. The preliminary F-ratio and Levene test procedures are primarily liberal as the MW test is used too often for the $n_1 = n_2 = 20$ case; and primarily conservative as the Welch test is used too often for the $n_1 = n_2 = 5$ case. For the moderate direct pairing cases, the *t* test is generally robust. The Welch test is robust when n_2 is four times as large as n_1 (5:20 and 10:40), but generally conservative when n_2 is twice as large as n_1. The MW test is robust for the n_1:n_2 = 5:10 and 5:20 cases but liberal for the n_1:n_2 = 10:20 and 10:40 cases. The results using the preliminary F-ratio and Levene test procedures are also inconsistent as none of the preliminary test procedures are uniformly robust.

For the $R = 0.25$ and 0.50 cases with balanced sample sizes, all procedures are generally liberal. For the $R = 2$ cases with unbalanced sample sizes, the MW test is liberal, whereas the Welch test is generally conservative. The results using the *t* test are inconsistent (sometimes liberal and sometimes conservative). The preliminary F-ratio and Levene test procedures are generally liberal except for the F(.75) test procedure. The F(.75) test procedure is robust for all cases. For the $R = 4$ cases with unbalanced sample sizes, the results using the t test are again inconsistent. All other procedures are generally liberal.

For the $R = 1.50$ cases where the sample sizes are balanced, the MW test is always liberal. The Welch test is robust only for the $n_1 = n_2 = 20$ case; otherwise it is conservative. The t test is robust except for the $n_1 = n_2 = 5$ case where the results are conservative. The results from the preliminary test procedures using the F-ratio test and the Levene test are also inconsistent yielding no uniformly robust test. For the moderate indirect pairing cases, the MW test, the Welch test, and the preliminary F-ratio and Levene test procedures are liberal. The results from the t test are inconsistent (robust for two cases, liberal for one case and conservative for one case).

For the $R = 2$ and 4 cases regardless of sample size configuration, all procedures are generally liberal.

In summary, for the moderately skewed lognormal cases, the MW test and the L(.05) test procedure are robust tests for the variance homogeneity cases. For the $R \neq 1$ cases, all procedures can be biased (generally liberal).

Table 14

Lognormal (0, 1) Distribution: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_r/σ_s						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	15.39	7.15	4.44	2.98	3.57	5.06	11.01
		F(.05)	14.70	7.74	4.78	2.80	3.60	5.19	10.34
		F(.25)	14.46	6.37	3.61	2.04	2.49	3.73	8.09
		F(.50)	14.24	5.89	3.22	1.82	2.16	3.29	7.44
		F(.75)	13.99	5.67	3.03	1.72	1.96	2.99	7.15
		L(.05)	15.12	9.67	6.87	5.04	6.66	9.42	20.19
		L(.25)	14.80	8.03	5.12	3.08	4.14	5.87	12.36
		L(.50)	14.44	6.18	3.38	1.85	2.21	3.27	7.38
		L(.75)	14.03	5.61	2.96	1.71	1.94	2.88	6.88
		W	13.85	5.39	2.81	1.65	1.84	2.85	6.80
		MW	15.14	9.75	7.02	5.21	6.98	9.83	21.20
5	10	*t*	8.83	6.36	5.08	3.77	3.87	4.76	9.50
		F(.05)	10.93	5.54	3.75	3.68	6.34	9.74	21.33
		F(.25)	10.31	4.74	3.46	3.62	6.84	10.51	21.55
		F(.50)	10.25	4.31	3.08	3.55	6.77	10.43	21.30
		F(.75)	10.20	3.99	2.90	3.48	6.69	10.31	20.94
		L(.05)	12.52	6.66	4.74	4.44	7.07	10.35	24.69
		L(.25)	13.27	6.04	4.02	4.27	7.64	11.84	25.29
		L(.50)	12.47	5.45	3.54	3.68	6.82	10.38	21.09
		L(.75)	10.85	4.45	3.11	3.51	6.69	10.24	20.57
		W	9.95	3.77	2.77	3.41	6.62	10.21	20.50
		MW	12.28	6.61	4.70	4.19	6.63	9.97	24.20

Table 14--Continued

n_1	n_2	Test	Ratio = σ_1/σ_2 .25	.50	.67	1.00	1.50	2.00	4.00
5	20	*t*	2.90	3.85	4.27	4.88	5.47	6.35	10.47
		F(.05)	9.39	7.52	6.71	8.83	14.06	20.16	36.48
		F(.25)	7.66	6.06	6.12	9.10	15.42	21.46	36.62
		F(.50)	7.02	5.01	5.45	8.79	15.23	21.00	35.44
		F(.75)	6.41	4.16	4.92	8.61	14.97	20.57	34.18
		L(.05)	12.83	7.42	5.05	5.15	9.65	15.94	35.14
		L(.25)	13.25	8.28	6.93	8.44	13.98	20.17	37.17
		L(.50)	9.79	7.21	6.50	9.17	15.47	21.32	36.41
		L(.75)	6.94	5.26	5.43	8.73	14.92	20.55	34.52
		W	5.73	3.36	4.38	8.38	14.74	20.17	33.27
		MW	12.56	7.35	5.20	5.19	9.57	15.94	35.14
10	10	*t*	13.41	7.44	5.36	3.52	4.64	6.94	14.82
		F(.05)	16.29	11.42	6.84	3.81	5.95	10.10	25.12
		F(.25)	14.19	9.83	6.05	3.37	5.15	8.43	20.61
		F(.50)	13.54	8.72	5.50	3.03	4.64	7.35	17.56
		F(.75)	13.19	7.86	4.98	2.84	4.21	6.45	15.28
		L(.05)	26.20	13.29	8.54	5.62	9.08	15.56	38.95
		L(.25)	22.76	12.97	7.60	4.23	6.97	12.12	32.20
		L(.50)	15.98	11.22	6.34	3.35	5.35	9.34	24.43
		L(.75)	13.89	9.03	5.30	2.98	4.51	7.55	18.32
		W	12.95	7.01	4.49	2.65	3.75	5.82	13.18
		MW	26.10	13.15	8.22	5.35	9.01	15.43	39.08

Table 14--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	6.38	5.19	4.58	3.71	4.83	6.78	15.36
		F(.05)	12.24	11.10	6.85	5.60	10.82	17.45	38.11
		F(.25)	10.32	8.66	5.51	5.12	10.03	15.69	33.43
		F(.50)	10.02	6.93	4.63	4.83	9.34	14.38	29.89
		F(.75)	9.82	5.64	3.78	4.60	8.95	13.13	26.82
		L(.05)	29.57	14.41	8.22	5.28	10.62	19.74	49.24
		L(.25)	22.12	13.91	7.77	6.07	12.14	20.69	47.73
		L(.50)	12.26	11.13	6.46	5.18	10.36	16.74	37.54
		L(.75)	10.06	7.87	4.84	4.70	9.29	14.26	30.42
		W	9.64	4.57	3.24	4.34	8.48	12.38	24.40
		MW	29.34	14.07	8.07	5.06	10.29	19.52	49.16
10	40	*t*	2.04	2.97	3.54	4.30	6.30	8.48	16.98
		F(.05)	8.70	11.30	8.83	9.57	16.73	25.40	46.90
		F(.25)	7.16	8.00	7.35	9.17	16.02	23.47	42.52
		F(.50)	6.92	6.30	6.26	8.79	15.34	21.77	38.20
		F(.75)	6.80	5.03	5.21	8.49	14.66	20.50	34.96
		L(.05)	32.24	14.58	7.65	4.68	12.34	24.19	56.85
		L(.25)	20.56	15.02	9.65	8.73	16.29	26.34	53.78
		L(.50)	8.33	11.65	8.52	9.18	16.29	24.67	46.94
		L(.75)	6.90	7.51	6.55	8.64	14.98	21.41	38.24
		W	6.74	3.94	4.54	8.14	14.00	19.10	31.97
		MW	32.03	14.27	7.76	4.85	12.48	24.35	57.27

Table 14--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
20	20	*t*	12.14	8.05	5.61	3.85	4.87	7.42	16.57
		F(.05)	12.82	15.79	9.74	4.82	8.70	16.01	36.77
		F(.25)	12.02	12.92	8.26	4.43	7.00	12.89	28.54
		F(.50)	11.86	10.90	7.21	4.10	6.04	10.48	23.76
		F(.75)	11.85	9.30	6.22	3.80	5.26	8.67	19.82
		L(.05)	39.48	22.72	12.20	6.11	12.58	25.23	65.52
		L(.25)	16.74	21.49	11.71	5.25	10.74	22.57	59.67
		L(.50)	12.26	16.90	9.74	4.49	8.19	16.68	42.75
		L(.75)	11.84	12.09	7.58	3.95	6.28	11.71	28.32
		W	11.83	7.88	5.37	3.47	4.46	7.11	16.04
		MW	44.83	22.35	11.63	4.97	11.97	24.83	65.47

4.2.3 Lognormal: Extreme Skewness

Table 15 presents the simulated null rejection rates for the lognormal (0, 1.75) distribution. The equal variance cases are discussed first. Results show the MW test and the L(.05) test procedure are generally robust for the equal variance cases regardless of sample size configuration. The other procedures tend to be conservative for the equal variance cases with $n_1 = n_2$. For those equal variance cases where n_2 is four times as large as n_1 (5:20 and 10:40), the *t* test is robust, whereas, the Welch test is liberal. The *t* test and the Welch test are conservative in all of the remaining situations. All preliminary F-ratio and Levene test procedures, except the L(.05) test procedures, are liberal for those equal variance cases where the Welch test is liberal and conservative for those cases where the Welch test is conservative.

For the $R = 0.67$ cases, the MW test is generally liberal in all sample sizes configurations. The t test is conservative for those cases where the sample sizes are equal. For the n_1:n_2 = 5:10 and 10:20 cases, the t test is robust whereas, for the n_1:n_2 = 5:20 and 10:40 cases, the t test is liberal. The Welch test is generally conservative in all cases. The preliminary F-ratio and Levene test procedures are robust for the n_1:n_2 = 5:20 and 20:20 cases. For the remaining sample size pair cases, neither the preliminary F-ratio nor Levene test procedures yield a uniformly robust test.

For the $R = 0.25$ and 0.50 cases, the MW test is liberal. The t test is also liberal but not as severe as the MW test. The Welch test is somewhat robust but generally liberal for the $R = 4$ cases and conservative for the $R = 2$ cases. There is no uniformly robust preliminary test procedure because either the Welch test or the MW test is used too often.

For the $R = 1.50$ cases, the MW test is liberal except for the $n_1 = 5$ and $n_2 = 10$ case, where it is robust. For cases where n_2 is four times as large as n_1 (5:20 and 10:40), the t test is robust, whereas the Welch test is liberal. The t test and the Welch test are generally conservative in all of the remaining situations. Some of the preliminary F-ratio and Levene test procedures yield robust results but they are sporadic. The preliminary test

Table 15

Lognormal (0, 1.75) Distribution: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	12.19	3.62	2.29	1.78	1.91	2.34	3.92
		F(.05)	11.64	3.25	1.98	1.26	1.47	1.87	3.12
		F(.25)	9.44	2.29	1.38	0.88	0.92	1.22	2.32
		F(.50)	8.76	2.06	1.19	0.78	0.82	1.03	2.07
		F(.75)	8.34	1.92	1.10	0.73	0.73	0.90	1.97
		L(.05)	20.72	9.12	6.25	4.87	6.29	8.05	14.16
		L(.25)	13.51	4.58	3.13	2.18	2.96	3.58	6.23
		L(.50)	8.84	2.01	1.12	0.76	0.77	0.99	1.99
		L(.75)	8.18	1.86	1.08	0.71	0.70	0.88	1.93
		W	8.03	1.79	1.04	0.72	0.70	0.88	1.92
		MW	21.43	9.71	6.70	5.21	6.64	8.49	15.27
5	10	*t*	14.82	6.89	4.61	2.97	2.07	1.94	1.98
		F(.05)	12.32	3.39	2.30	2.11	3.19	4.55	8.32
		F(.25)	9.11	2.43	1.78	1.96	3.17	4.43	8.21
		F(.50)	7.09	1.94	1.57	1.90	3.09	4.36	8.10
		F(.75)	5.97	1.62	1.47	1.83	3.06	4.31	8.02
		L(.05)	19.56	6.52	4.30	3.93	5.58	8.27	17.16
		L(.25)	13.62	3.61	2.62	3.16	5.17	7.36	14.74
		L(.50)	10.92	2.80	1.90	2.01	3.12	4.38	8.06
		L(.75)	7.57	2.05	1.55	1.86	3.04	4.31	7.99
		W	4.93	1.36	1.32	1.79	3.02	4.29	7.98
		MW	21.55	7.85	5.18	4.19	5.36	7.87	16.52

Table 15--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	20	*t*	10.90	7.70	6.60	5.53	4.56	3.98	3.03
		F(.05)	16.65	6.65	6.03	8.00	11.62	15.02	23.78
		F(.25)	11.40	4.97	5.24	7.74	11.57	14.90	23.41
		F(.50)	8.00	3.98	4.67	7.48	11.40	14.71	23.04
		F(.75)	5.20	3.23	4.27	7.35	11.30	14.55	22.82
		L(.05)	25.07	7.30	4.66	4.43	7.45	11.35	24.92
		L(.25)	21.75	6.88	5.56	7.08	11.31	15.64	28.86
		L(.50	18.10	6.34	5.89	8.11	12.08	15.74	25.59
		L(.75)	10.40	4.45	4.89	7.60	11.39	14.64	23.01
		W	2.68	2.49	3.95	7.20	11.19	14.49	22.60
		MW	27.41	9.31	6.14	5.19	7.65	11.40	24.96
10	10	*t*	15.40	4.83	3.03	1.91	2.28	3.13	6.10
		F(.05)	26.94	7.01	3.71	2.30	3.01	4.27	9.70
		F(.25)	22.20	5.76	2.95	1.79	2.50	3.39	7.63
		F(.50)	18.96	5.00	2.51	1.61	2.10	2.94	6.33
		F(.75)	16.58	4.32	2.34	1.47	1.81	2.51	5.27
		L(.05)	39.52	13.33	8.11	5.43	7.70	11.77	27.05
		L(.25)	34.58	8.79	4.56	2.98	4.21	6.28	15.10
		L(.50)	26.38	6.14	3.13	1.90	2.65	3.70	8.83
		L(.75)	19.46	4.71	2.50	1.59	2.03	2.74	6.27
		W	14.11	3.71	2.06	1.34	1.64	2.11	4.45
		MW	39.47	13.47	8.17	5.35	7.63	11.94	27.89

Table 15--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	13.97	6.65	4.58	2.85	2.25	2.28	3.18
		F(.05)	27.29	7.78	4.16	3.92	6.52	9.72	19.71
		F(.25)	19.57	5.40	3.08	3.54	6.11	8.80	17.27
		F(.50)	15.28	3.91	2.53	3.39	5.75	8.22	15.83
		F(.75)	11.70	2.86	2.17	3.21	5.51	7.84	14.72
		L(.05)	47.71	13.72	7.14	4.73	8.47	14.42	35.75
		L(.25)	42.41	9.88	5.26	5.10	9.01	14.15	31.33
		L(.50)	33.61	7.42	3.97	3.86	6.37	9.42	19.19
		L(.75)	19.87	4.71	2.83	3.31	5.68	8.19	15.97
		W	8.39	2.09	1.73	3.03	5.33	7.45	13.85
		MW	48.65	15.38	8.24	5.06	8.37	14.26	35.63
10	40	*t*	8.72	6.81	6.07	5.14	4.42	3.82	3.63
		F(.05)	26.52	10.10	7.71	9.15	14.25	19.28	33.07
		F(.25)	17.01	7.27	6.43	8.70	13.50	18.07	30.54
		F(.50)	11.68	5.79	5.76	8.41	13.26	17.62	29.07
		F(.75)	7.75	4.41	5.20	8.20	13.09	17.19	28.11
		L(.05)	57.31	14.62	6.46	4.01	9.27	17.05	42.83
		L(.25)	54.23	13.70	8.04	7.91	13.84	21.13	44.36
		L(.50)	43.60	11.58	7.83	9.13	14.38	19.88	35.38
		L(.75)	22.37	7.51	6.24	8.50	13.34	17.82	29.72
		W	4.54	3.15	4.67	8.06	12.88	16.93	27.26
		MW	58.54	16.98	8.18	4.85	9.56	17.22	42.93

Table 15--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
20	20	*t*	17.42	6.07	3.60	2.32	2.76	3.80	8.19
		F(.05)	37.00	12.24	5.61	2.98	4.06	7.35	17.92
		F(.25)	29.30	9.28	4.59	2.60	3.39	5.73	13.52
		F(.50)	24.46	7.80	4.03	2.37	2.92	4.59	10.77
		F(.75)	20.83	6.74	3.53	2.15	2.57	3.97	9.05
		L(.05)	67.22	23.76	11.30	5.95	10.10	18.33	49.04
		L(.25)	62.89	17.91	7.69	3.73	6.23	11.90	32.59
		L(.50)	45.76	12.51	5.42	2.83	4.10	7.43	20.34
		L(.75)	30.44	8.92	4.14	2.31	3.15	5.21	13.00
		W	17.11	5.58	3.11	1.95	2.24	3.29	7.30
		MW	67.18	23.26	10.55	4.97	9.47	17.92	48.88

procedures specifically use the Welch test or the MW test exclusively.

For the R = 2 and 4 cases, the MW test is severely liberal. The *t* test is always conservative for the R = 2 cases. For the R = 4 cases, the *t* test is conservative when the sample sizes are unequal and generally liberal when the sample sizes are identical. The Welch test is generally liberal when the sample sizes are unequal and generally conservative when the sample sizes are equivalent for the R = 2 and 4 cases. As seen for the R =.50 cases, the preliminary F-ratio and Levene test procedures yield robust results for some cases, but they are sporadic. The preliminary test procedures specifically use the Welch test or the MW test exclusively.

In summary, the MW test and the L(0.05) test procedure are robust for the variance homogeneity cases regardless of the sample size configuration. As the variance heterogeneity increases, the MW test becomes more liberal. The t test and the Welch test are liberal for some cases and conservative in other cases, with no apparent pattern. Various preliminary test procedures yield robust results for one case or another, but in no apparent pattern. The liberal or conservative bias does tend to increase for all procedures as the R values depart from one.

4.3 Gamma Distributions

Two gamma distributions are examined in this section, a gamma with shape parameter equal to 3 and unit scale parameter (slight skewness) denoted as G(3,1) and the other with shape parameter of 2 and unit scale parameter (moderate skewness) denoted as G(2,1).

4.3.1 Gamma (3,1) Distribution

Table 16 presents the simulated null rejection rates for the gamma (3,1) distribution. For the variance homogeneity cases with equal sample sizes, Table 16 shows all procedures are generally robust. For the $n_1 \neq n_2$ cases, the t test, the MW test, the F(.05), and the L(.05) test procedure are robust. The Welch test and all other preliminary test procedures are robust for the cases where n_2 is twice as large as n_1 (5:10 and 10:20), but liberal

Table 16

Gamma (3, 1) Distribution: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	9.36	6.95	5.48	5.03	5.56	8.32	25.74
		F(.05)	7.81	6.85	5.67	5.18	5.80	8.40	25.53
		F(.25)	7.30	6.19	5.12	4.54	5.19	7.53	24.38
		F(.50)	7.14	6.00	4.81	4.37	4.70	7.11	23.87
		F(.75)	7.05	5.84	4.58	4.06	4.48	6.74	23.42
		L(.05)	8.47	7.33	6.03	5.69	6.41	9.40	27.52
		L(.25)	7.61	6.79	5.57	5.10	5.83	8.36	25.44
		L(.50)	7.23	6.12	4.96	4.47	4.81	7.20	23.52
		L(.75)	7.14	5.90	4.64	4.12	4.52	6.75	22.99
		W	6.92	5.63	4.41	3.88	4.26	6.49	22.97
		MW	8.59	7.33	6.08	5.73	6.48	9.50	27.77
5	10	*t*	2.84	3.19	3.43	4.71	7.79	12.33	31.62
		F(.05)	4.98	4.08	3.53	4.29	6.90	10.94	31.58
		F(.25)	6.06	4.66	4.10	4.89	7.45	11.55	31.77
		F(.50)	6.22	4.87	4.20	5.11	7.71	11.88	31.37
		F(.75)	6.26	4.83	4.19	5.13	7.68	11.90	30.85
		L(.05)	4.81	3.98	3.51	4.25	7.22	11.22	31.62
		L(.25)	6.41	4.94	4.33	5.12	7.83	11.76	31.73
		L(.50)	6.32	5.05	4.37	5.40	7.84	12.08	31.44
		L(.75)	6.28	4.87	4.29	5.32	7.67	11.88	30.65
		W	6.25	4.73	4.11	5.14	7.58	11.67	30.08
		MW	3.82	3.58	3.07	3.99	7.09	11.12	31.68

Table 16--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	20	*t*	0.42	1.06	1.85	4.81	11.15	18.75	40.66
		F(.05)	4.66	4.10	4.47	5.67	9.50	16.06	38.89
		F(.25)	5.39	5.06	5.49	7.16	10.13	16.03	36.04
		F(.50)	5.48	5.06	5.54	7.31	10.37	15.70	34.26
		F(.75)	5.48	4.80	5.40	6.91	9.82	14.83	32.71
		L(.05)	4.15	3.50	3.69	5.59	9.85	17.04	49.03
		L(.25)	5.67	5.40	5.76	7.45	10.75	17.23	44.14
		L(.50)	5.50	5.15	5.75	7.51	10.70	16.71	40.09
		L(.75)	5.49	4.85	5.41	7.04	9.86	15.26	35.85
		W	5.49	4.66	5.06	6.58	9.13	13.44	31.04
		MW	2.15	2.43	2.83	5.02	10.03	17.51	49.26
10	10	*t*	7.85	6.46	5.28	4.87	5.22	7.87	20.41
		F(.05)	6.81	6.72	5.94	5.36	6.63	11.91	29.25
		F(.25)	6.75	6.08	5.66	5.19	6.10	10.01	24.45
		F(.50)	6.76	5.94	5.33	5.00	5.67	8.95	22.12
		F(.75)	6.77	5.90	5.19	4.71	5.41	8.24	20.99
		L(.05)	10.61	7.98	6.41	5.57	7.28	13.93	50.52
		L(.25)	7.23	7.29	6.38	5.48	7.15	13.59	49.52
		L(.50)	6.79	6.42	5.81	5.18	6.59	12.06	37.81
		L(.75)	6.77	6.12	5.35	4.86	5.78	9.57	27.57
		W	6.77	5.94	4.97	4.48	4.86	7.48	20.13
		MW	11.40	7.87	6.27	5.39	7.16	13.77	50.44

Table 16--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	t	1.75	2.39	3.02	4.67	8.01	12.20	27.59
		F(.05)	5.52	5.30	5.26	5.45	8.39	14.43	29.12
		F(.25)	5.53	5.00	5.19	5.79	7.87	12.45	24.80
		F(.50)	5.53	4.84	5.00	5.58	7.48	11.27	23.48
		F(.75)	5.53	4.87	4.77	5.34	6.99	10.16	22.80
		L(.05)	7.72	6.37	5.23	5.32	9.28	18.63	60.62
		L(.25)	5.61	5.82	5.68	5.88	9.24	17.83	49.52
		L(.50)	5.53	5.17	5.31	5.78	8.72	15.12	36.30
		L(.75)	5.53	4.99	4.91	5.47	7.53	12.47	28.91
		W	5.53	4.88	4.60	5.04	6.46	9.42	22.36
		MW	6.85	5.29	4.45	4.93	9.35	18.70	60.68
10	40	t	0.26	0.87	1.67	4.86	11.38	18.53	36.66
		F(.05)	5.38	5.78	5.86	5.99	10.69	16.29	28.20
		F(.25)	5.37	5.49	5.83	6.61	10.23	14.01	24.86
		F(.50)	5.37	5.33	5.48	6.60	9.77	12.46	23.92
		F(.75)	5.37	5.25	5.25	6.38	8.94	11.40	23.59
		L(.05)	6.11	6.02	5.16	5.42	11.61	22.04	60.38
		L(.25)	5.37	5.87	6.19	6.76	11.82	19.59	42.68
		L(.50)	5.37	5.42	5.73	6.86	11.34	16.96	34.10
		L(.75)	5.37	5.25	5.28	6.43	9.95	14.08	28.45
		W	5.37	5.21	4.94	6.11	8.05	10.27	23.47
		MW	4.27	3.98	3.75	4.85	11.87	22.62	66.10

Table 16--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
20	20	*t*	6.68	5.55	5.25	5.42	5.35	7.11	15.20
		F(.05)	6.20	6.20	6.38	5.94	8.41	13.26	17.45
		F(.25)	6.20	5.54	5.69	6.07	7.30	9.93	15.49
		F(.50)	6.20	5.35	5.22	5.93	6.47	8.50	15.11
		F(.75)	6.20	5.36	5.10	5.62	5.73	7.62	15.00
		L(.05)	6.71	9.06	7.20	5.81	9.79	21.74	78.59
		L(.25)	6.21	6.60	6.74	6.35	9.57	18.23	52.08
		L(.50)	6.20	5.71	5.92	6.12	8.20	13.01	28.11
		L(.75)	6.20	5.43	5.31	5.82	6.66	9.42	19.32
		W	6.20	5.34	5.04	5.26	5.18	6.95	14.93
		MW	14.06	9.33	6.73	5.36	9.31	21.23	78.91

for the cases where n_1:n_2 = 5:20 and 10:40.

For the R = 0.67 cases with equal sample sizes, the t test and the Welch test are robust, whereas the MW test is liberal. All the preliminary F-ratio test procedures are robust, except the F(.05) test procedure for the $n_1 = n_2 = 20$ case where it is liberal. The L(.50) and the L(.75) test procedures are robust for the $n_1 = n_2$ cases. For the $n_1 \neq n_2$ cases, all preliminary F-ratio and Levene test procedures are generally robust, except for the L(.05) test procedure which is sometimes conservative and sometimes liberal.

For the R = 0.25 and 0.50 cases with equal sample sizes, the t test and the MW test are generally liberal. The Welch test and the F(.50) and the L(.75) test procedures are robust for the R = 0.50 cases but liberal for the R = 0.25 cases. None of the preliminary Levene test procedures are consistently robust for the $n_1 = n_2$ cases. For the severe direct pairing cases, the t test is always conservative. The MW test is sometimes liberal, sometimes conservative, and sometimes robust, whereas the Welch test is generally robust for the severe direct pairing cases. The preliminary test procedures using the F-ratio test and the Levene test are generally robust, except for the L(.05) test procedure. For the R = 0.25 cases where $n_1{:}n_2$ = 5:10, the preliminary test procedures with $\alpha > 0.05$ tend to be liberal because the Welch test is often used.

For the R = 1.50 cases with equal sample sizes, the t test, the Welch test, and the F(.75) test procedure are robust, whereas the MW test and the other preliminary test procedures tend to be liberal. For the $n_1 \neq n_2$ cases, all procedures are liberal.

For the R = 2 and 4 cases, all procedures are liberal.

In summary, for the R = 1 cases, the t test, the MW test, the F(.05), and the L(.05) procedures are robust. For the R = 0.67 cases, the t test, the Welch test, and the preliminary test procedures where $\alpha > 0.05$ are generally robust. For the R = 0.50 cases with equal sample sizes, the Welch test, the F(.50) and the L(.75) test procedure are robust. None of the procedures are robust for the R = 0.25 cases with equal sample sizes. For the severe direct pairing cases, the Welch test and the preliminary test procedures where $\alpha > 0.05$ are generally robust. For the $R = 1.50$ cases with equal sample sizes, the t test, the Welch test, and the F(.75) test procedure are robust. For the $R = 1.50$ with $n_1 \neq n_2$ cases, all procedures are liberal. For the $R = 2$ and 4 cases, all procedures are liberal.

4.3.2 Gamma (2,1) Distribution

The simulated null rejection rates for the gamma (2,1) distribution are presented in Table 17. For the $R = 1$ cases with balanced sample sizes, Table 17 shows the t test and the MW test procedures are robust. The Welch test is conservative for the $n_1 = n_2 = 5$ cases. This conservatism is also seen in the preliminary F-ratio test procedure with $\alpha > 0.05$ and in the preliminary Levene test procedure with $\alpha > 0.25$. The F(.05) and L(.05) test procedures are robust. For the $n_1 \neq n_2$ cases, the t test is robust. The MW test is robust for all cases except $n_1{:}n_2 = 5{:}10$ where it is slightly conservative. The Welch test is liberal for the case where n_2 is four times larger than n_1; otherwise, the Welch test is robust. The L(.05) test procedure is the only preliminary F-ratio or Levene test procedure which is always robust (maximum of 6.08%).

For the $R = 0.67$ cases with equal sample sizes, the t test and the Welch test are robust, whereas the MW test is liberal. The F(.75) test procedure is the only preliminary F-ratio or Levene test procedure which

Table 17

Gamma (2, 1) Distribution: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	9.90	6.90	5.49	4.56	5.57	8.27	27.98
		F(.05)	8.47	6.96	5.75	4.69	5.96	8.81	27.02
		F(.25)	8.08	6.52	4.99	3.95	4.86	7.45	24.92
		F(.50)	7.95	6.11	4.60	3.59	4.29	6.75	23.85
		F(.75)	7.87	5.95	4.36	3.40	4.01	6.42	23.27
		L(.05)	9.01	7.44	6.26	5.27	7.15	10.79	32.82
		L(.25)	8.35	6.91	5.54	4.54	5.76	8.78	26.79
		L(.50)	8.04	6.41	4.89	3.74	4.50	6.85	23.53
		L(.75)	7.93	5.99	4.40	3.43	4.08	6.24	22.78
		W	7.81	5.71	4.10	3.23	3.81	6.05	22.71
		MW	9.10	7.50	6.32	5.34	7.24	10.89	33.82
5	10	*t*	3.18	3.40	3.69	4.80	7.87	11.54	33.69
		F(.05)	5.68	3.88	3.69	4.23	7.87	12.67	38.55
		F(.25)	6.63	4.39	4.02	4.89	8.68	13.50	38.64
		F(.50)	6.96	4.56	4.04	4.97	9.04	13.75	38.32
		F(.75)	7.03	4.65	3.93	4.99	9.00	13.74	37.36
		L(.05)	5.69	3.77	3.67	4.33	7.80	12.86	38.66
		L(.25)	7.34	4.65	4.35	5.15	8.93	13.73	38.73
		L(.50)	7.07	4.85	4.32	5.28	9.09	13.71	37.66
		L(.75)	7.04	4.77	4.06	5.18	8.97	13.50	36.58
		W	7.00	4.52	3.83	5.03	8.82	13.35	36.24
		MW	4.69	3.49	3.33	3.84	7.62	12.68	38.62

Table 17--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	20	*t*	0.63	1.37	2.30	4.84	10.55	16.94	40.66
		F(.05)	4.80	4.90	5.25	6.13	11.28	18.80	46.40
		F(.25)	5.49	5.43	6.23	7.85	12.83	19.36	43.98
		F(.50)	5.51	5.35	6.01	7.98	12.75	18.99	41.96
		F(.75)	5.49	5.10	5.71	7.62	12.15	17.76	39.79
		L(.05)	4.51	4.20	4.18	5.58	11.18	19.57	57.38
		L(.25)	5.92	5.87	6.46	8.02	12.80	20.22	51.97
		L(.50)	5.54	5.60	6.39	8.28	12.83	19.50	47.11
		L(.75)	5.49	5.19	5.87	7.87	12.18	17.92	41.98
		W	5.49	4.78	5.32	7.42	11.55	16.60	37.24
		MW	2.89	3.11	3.22	4.96	11.27	19.88	57.53
10	10	*t*	8.18	6.84	5.18	5.02	5.50	7.91	24.48
		F(.05)	7.31	7.97	6.28	5.54	7.66	14.03	37.99
		F(.25)	7.17	7.17	5.76	5.25	6.95	11.92	31.78
		F(.50)	7.14	6.73	5.43	4.94	6.18	10.24	28.44
		F(.75)	7.14	6.56	5.08	4.65	5.64	8.93	26.06
		L(.05)	12.87	9.29	6.85	5.76	8.40	16.95	62.56
		L(.25)	8.12	8.75	6.68	5.57	8.02	16.47	61.37
		L(.50)	7.19	7.77	6.16	5.16	7.26	14.26	48.36
		L(.75)	7.17	6.90	5.42	4.81	6.14	10.94	35.13
		W	7.17	6.32	4.77	4.52	5.06	7.48	24.13
		MW	13.23	9.13	6.68	5.46	8.13	16.79	62.57

Table 17--Continued

			Ratio = σ_y/σ_x						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	2.32	2.97	3.35	4.67	8.50	12.74	28.93
		F(.05)	6.10	6.71	6.19	6.13	10.71	18.41	36.32
		F(.25)	6.15	6.06	5.71	6.55	10.46	15.75	31.08
		F(.50)	6.15	5.79	5.44	6.29	9.68	14.23	28.76
		F(.75)	6.15	5.72	5.18	6.02	8.97	12.90	27.31
		L(.05)	9.99	7.90	6.21	5.69	11.34	23.29	70.10
		L(.25)	6.34	7.40	6.58	6.72	11.81	22.84	61.18
		L(.50)	6.16	6.38	5.93	6.52	11.30	19.74	46.49
		L(.75)	6.15	5.84	5.36	6.05	9.72	15.46	35.92
		W	6.15	5.61	4.83	5.65	8.19	11.09	26.13
		MW	9.35	6.76	5.28	5.15	11.17	23.30	70.08
10	40	*t*	0.41	1.02	1.75	5.03	11.45	18.91	39.43
		F(.05)	5.87	5.89	6.32	7.19	12.52	20.91	38.15
		F(.25)	5.85	5.14	6.01	7.70	12.25	18.32	33.39
		F(.50)	5.85	4.91	5.67	7.31	11.36	16.22	31.19
		F(.75)	5.85	4.82	5.27	6.83	10.28	14.40	30.06
		L(.05)	8.03	6.51	5.70	6.08	13.59	27.91	73.51
		L(.25)	5.87	5.96	6.52	7.63	13.95	25.79	56.88
		L(.50)	5.85	5.14	5.93	7.61	13.35	22.56	46.69
		L(.75)	5.85	4.84	5.37	7.05	11.42	17.86	37.76
		W	5.85	4.69	4.96	6.37	9.24	12.60	29.23
		MW	6.41	4.63	4.08	5.39	13.75	28.23	77.64

Table 17--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
20	20	*t*	7.37	6.36	5.57	4.79	5.71	7.67	18.86
		F(.05)	6.89	7.73	7.82	5.72	10.11	17.05	24.27
		F(.25)	6.89	6.53	6.82	5.73	8.64	12.74	20.49
		F(.50)	6.89	6.18	6.38	5.38	7.49	10.55	19.53
		F(.75)	6.89	6.11	5.92	5.02	6.45	8.86	18.99
		L(.05)	8.03	11.66	8.76	5.48	12.17	29.34	88.90
		L(.25)	6.90	8.22	8.21	5.88	11.92	26.35	71.66
		L(.50)	6.89	6.74	7.32	5.53	10.16	19.15	45.63
		L(.75)	6.89	6.27	6.39	5.13	7.70	12.87	30.60
		W	6.89	6.09	5.46	4.68	5.55	7.41	18.77
		MW	18.16	11.81	8.15	4.72	11.47	28.93	88.94

is always robust. Several of the preliminary test procedures are robust for the $n_1 = n_2 = 5$ and $n_1 = n_2 = 10$ cases, but become liberal for the $n_1 = n_2 = 20$ case because the MW test is more often used. For the moderate direct pairing cases, the *t* test is always conservative, whereas, the Welch test is primarily robust. The MW test is sometimes conservative and sometimes liberal. The F(.50), F(.75), and L(.75) test procedures are generally robust in all cases.

For the $R = 0.25$ and 0.50 cases with equal sample sizes, all procedures are generally liberal. For the severe direct pairing cases, the *t* test is conservative. The MW test tends to be conservative for the n_1:n_2 =

5:10 and 5:20 cases and liberal for the n_1:n_2 = 10:20 and 10:40 cases. The Welch test is generally robust. The F(.50), F(.75), and L(.75) test procedures are also generally robust. The preliminary F-ratio and Levene test procedure at α = 0.05 or 0.25 tend to be biased because the MW test is more often used.

For the R = 1.50 cases with equal sample sizes, the t test and the Welch test (minimum of 3.81%) are generally robust, whereas the MW test is severely liberal. None of the preliminary F-ratio and Levene test procedures are robust for the $n_1 = n_2 = 20$ case. For $n_1 = n_2 = 5$ and $n_1 = n_2 = 10$ cases, only the F(.75) test procedure is robust. For the moderate indirect pairing cases, all procedures are liberal.

For the R = 2 and 4 cases regardless of sample size combination, all procedures are liberal.

In summary, for the R = 1 cases, the t test, the MW and the L(.05) test procedure are generally robust regardless of sample size configurations. For the R = 0.67 cases with balanced sample sizes, the t test, the Welch test, and the F(.75) test procedure are robust. For the moderate direct pairing cases, the Welch test, F(.50), F(.75), and L(.75) are robust. All procedures are liberal for the R = 0.25 and 0.50 cases with equal sample sizes. For the severe direct pairing cases, the Welch test, the F(.50), F(.75), and L(.75) test procedures are generally robust. For the R = 1.50 cases with equal sample sizes, the Welch test, the t test, and the F(.75) are generally robust. For the moderate indirect pairing cases and the $R = 2$ and 4 cases, all procedures are liberal.

4.4 Overall Performance for the Asymmetric Distributions

As in Chapter III for the symmetric distributions, the results of the simulations for the lognormal and gamma distributions are combined in Tables 18-22 in order to evaluate the procedures for varying degrees of variance heterogeneity. The same five groupings (previously defined for the symmetric cases) are again formulated depending on $R = \sigma_1/\sigma_2$. These are: (1) $R = 0.25$ and 0.50, (2) $R = 0.67$, (3) $R = 1$ (equal variances), (4) $R = 1.50$, and (5) $R = 2.0$ and 4.0. The proportion of rejections is expressed as a percent for the t test, the Welch test, the MW test, the preliminary F-ratio test procedures, and the preliminary Levene test procedures which are tabulated for each R grouping over all (35) combinations of sample size pairs (7) and distributions (5). In Tables 18-22, the interval $x \leq 2.5$ is separated into two categories: $x \leq 1$ and $1 < x \leq 2.5$ to yield more information concerning trends for the severe conservative cases. Also, the interval $x > 10$ is separated into two categories: $10 < x \leq 20$ and $x > 20$ to examine in more detail the patterns for the severe liberal cases. The seven categories are listed below:

1. $x \leq 1$
2. $1 < x \leq 2.5$

3. $2.5 < x \leq 4$

4. $4 < x \leq 6$

5. $6 < x \leq 10$

6. $10 < x \leq 20$

7. $x > 20$

Recall that value "x" represents the percentage of rejections for testing H_o: $\mu_1 = \mu_2$ based on 10,000 simulations for each sample size configuration. Each entry of the following tables denotes the frequency at which $a < x \leq b$ occurs. The outcome of a test procedure is defined as follows:

Severely Conservative: If the simulated null rejection rate is ≤ 1.

Extremely Conservative: If the simulated null rejection rate is > 1 and ≤ 2.5.

Conservative: If the simulated null rejection rate is > 2.5 and ≤ 4.0.

Robust: If the simulated null rejection rate is > 4.0 and ≤ 6.0.

Liberal: If the simulated null rejection rate is > 6.0 and ≤ 10.0.

Extremely Liberal: If the simulated null rejection rate is > 10.0 and ≤ 20.0.

Severely Liberal: If the simulated null rejection rate is > 20.0.

$R = 1$ (Equal Variance Case)

A summary of the simulated null rejection rates for the asymmetric distributions for the $R = 1$ cases is presented in Table 18. For the $R = 1$ cases, the MW test and the L(.05) test procedure are robust. The t test is

Table 18

Summary of Asymmetric Distributions: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
Equal Variances: $R=\sigma_1/\sigma_2 = 1.0$

test	$x \leq 1$	$1< x \leq 2.5$	$2.5< x \leq 4$	$4< x \leq 6$	$6< x \leq 10$	$10< x \leq 20$	$x >20$	Total
t	0	3	7	25	0	0	0	35
F(.05)	0	3	5	20	7	0	0	35
F(.25)	1	3	6	13	12	0	0	35
F(.50)	1	4	5	14	11	0	0	35
F(.75)	1	4	6	13	11	0	0	35
L(.05)	0	0	1	32	2	0	0	35
L(.25)	0	1	4	17	13	0	0	35
L(.50)	1	3	6	13	12	0	0	35
L(.75)	1	4	6	13	11	0	0	35
W	1	4	7	14	9	0	0	35
MW	0	0	2	33	0	0	0	35

*Preliminary test procedures used Mann-Whitney-Wilcoxon test for equal variances and the Welch test for unequal variances. Table is based on the five asymmetric distributions lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3, 1), & G(2, 1)] and seven sample size pairs.

robust for approximately 70% (25 of 35) of the cases but it can be conservative and even extremely conservative for some of the $R = 1$ cases. None of the procedures are extremely liberal.

R = 0.67 (Includes Moderate Direct Pairing Cases)

Table 19 summarizes the simulated null rejection rates for the asymmetric distributions for the $R = 0.67$ cases. The Welch test, the F(.50), F(.75), and L(.75) are robust in approximately 70% (25 of 35) of the $R = 0.67$ cases. These procedures can be conservative and even extremely conservative for some $R = 0.67$ cases. The t test is conservative or extremely conservative for almost 50% (16 of 35) of the $R = 0.67$ cases. The MW test are liberal or extremely liberal in more than 50% (19 of 35) of these cases. Overall, the preliminary test procedures using the Levene test are much slower to counteract the liberalism of the MW test than are the preliminary F-ratio test procedures.

$R = 0.25$ and 0.50 (Includes Severe Direct Pairing Cases)

The simulated null rejection rates for the asymmetric distributions for the $R = 0.25$ and 0.50 cases are summarized in Table 20. For the $R = 0.25$ and 0.50 cases, the Welch test, the F(.50), F(.75), and L(.75) test procedures are the most robust. The t test can be severely conservative as well as extremely liberal. The MW test is primarily liberal or extremely

Table 19

Summary of Asymmetric Distributions: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
$R=\sigma_1/\sigma_2 = 0.67$

test	$x \le 1$	$1< x \le 2.5$	$2.5< x \le 4$	$4< x \le 6$	$6< x \le 10$	$10< x \le 20$	$x >20$	Total
t	0	7	9	17	2	0	0	35
F(.05)	0	2	5	14	14	0	0	35
F(.25)	0	2	4	20	9	0	0	35
F(.50)	0	2	4	25	4	0	0	35
F(.75)	0	4	5	25	1	0	0	35
L(.05)	0	0	5	11	17	2	0	35
L(.25)	0	0	2	14	18	1	0	35
L(.50)	0	2	4	19	10	0	0	35
L(.75)	0	3	3	25	4	0	0	35
W	0	4	6	25	0	0	0	35
MW	0	0	8	8	17	2	0	35

*Preliminary test procedures used Mann-Whitney-Wilcoxon test for equal variances and the Welch test for unequal variances. Table is based on the five asymmetric distributions lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3, 1), & G(2, 1)] and seven sample size pairs.

liberal for the R = 0.25 and 0.50 cases. Overall, the preliminary test procedures using the Levene test are much slower to counteract the liberalism of the MW test than are the preliminary F-ratio test procedures.

<u>R = 1.50 (Includes Moderate Indirect Pairing Cases)</u>

Table 21 summarizes the simulated null rejection rates for the asymmetric distributions for the R = 1.50 cases. All procedures tend to be liberal or extremely liberal for the R = 1.50 cases. Excluding the MW test, the L(.05), and the L(.25) test procedures, the procedures can be occasionally extremely conservative.

<u>R = 2 and 4 (Includes Severe Indirect Pairing Cases)</u>

The simulated null rejection rates for the asymmetric distributions for the R = 2 and 4 cases are summarized in Table 22. All procedures are extremely or severely liberal for at least 50% of the R = 2 and 4 cases.

4.5 Summary

In summary, for the case of variance homogeneity, the MW test and the L(.05) test procedure are robust. Although all procedures evaluated are fairly robust for the equal variance cases. The Welch test, the F(.50), F(.75), and L(.75) are robust in approximately 70% of the R=0.67 cases. For the R=0.25 and 0.50 cases, the Welch test, the F(.50), F(.75), and L(.75) test procedures are the most robust. The best procedure is the F(.75) procedure which is only robust for 50% of the R=0.25 and 0.50 cases. For the R=1.50, 2.0, and 4.0 cases, all procedures range from liberal to severely liberal. As the R values increase, the procedures tend to become more liberal.

Table 20

Summary of Asymmetric Distributions: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level $R=\sigma_1/\sigma_2$ = 0.25 and 0.50

test	$x \le 1$	$1 < x \le 2.5$	$2.5 < x \le 4$	$4 < x \le 6$	$6 < x \le 10$	$10 < x \le 20$	$x > 20$	Total
t	8	9	12	4	28	9	0	70
F(.05)	0	0	4	21	27	14	4	70
F(.25)	0	2	0	26	31	9	2	70
F(.50)	0	2	2	31	25	9	1	70
F(.75)	0	2	3	35	23	6	1	70
L(.05)	0	0	5	8	32	13	12	70
L(.25)	0	0	1	19	30	11	9	70
L(.50)	0	1	1	22	29	13	4	70
L(.75)	0	2	0	31	26	9	2	70
W	0	4	6	33	22	5	0	70
MW	0	4	9	6	22	16	13	70

*Preliminary test procedures used Mann-Whitney-Wilcoxon test for equal variances and the Welch test for unequal variances. Table is based on the five asymmetric distributions lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3, 1), & G(2, 1)] and seven sample size pairs.

Table 21

Summary of Asymmetric Distributions: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
$R=\sigma_1/\sigma_2$ = 1.50

test	$x \le 1$	$1< x \le 2.5$	$2.5< x \le 4$	$4< x \le 6$	$6< x \le 10$	$10< x \le 20$	$x >20$	Total
t	0	4	3	15	7	6	0	35
F(.05)	0	1	3	5	16	10	0	35
F(.25)	1	2	2	5	15	10	0	35
F(.50)	1	2	2	8	15	7	0	35
F(.75)	1	2	2	11	13	6	0	35
L(.05)	0	0	0	1	24	10	0	35
L(.25)	0	0	1	6	14	14	0	35
L(.50)	1	1	2	5	14	12	0	35
L(.75)	1	2	2	7	17	6	0	35
W	1	3	4	9	13	5	0	35
MW	0	0	0	1	24	10	0	35

*Preliminary test procedures used Mann-Whitney-Wilcoxon test for equal variances and the Welch test for unequal variances. Table is based on the five asymmetric distributions lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3, 1), & G(2, 1)] and seven sample size pairs.

Table 22

Summary of Asymmetric Distributions: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
$R = \sigma_1/\sigma_2 = 2.0$ and 4.0

test	$x \leq 1.0$	$1 < x \leq 2.5$	$2.5 < x \leq 4$	$4 < x \leq 6$	$6 < x \leq 10$	$10 < x \leq 20$	$x > 20$	Total
t	0	4	8	2	17	23	16	70
F(.05)	0	1	1	3	10	27	28	70
F(.25)	0	2	2	2	12	27	25	70
F(.50)	0	2	2	2	13	29	22	70
F(.75)	1	1	3	3	16	25	21	70
L(.05)	0	0	0	0	6	26	38	70
L(.25)	0	0	1	1	7	23	38	70
L(.50)	1	1	2	1	11	23	31	70
L(.75)	1	1	2	3	12	27	24	70
W	1	2	2	4	17	24	20	70
MW	0	0	0	0	7	25	38	70

*Preliminary test procedures used Mann-Whitney-Wilcoxon test for equal variances and the Welch test for unequal variances. Table is based on the five asymmetric distributions lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3, 1), & G(2, 1)] and seven sample size pairs.

CHAPTER V

CONTAMINATED NORMAL DISTRIBUTIONS

5.1 Introduction

This chapter examines the performance of the preliminary variance equality test procedures for the two contaminated normal distributions, one being symmetric with 10% of the random realizations in the child distribution; and the other being asymmetric with 5% of the random realizations in the child distribution. Presented are the simulated null rejection rates for testing the hypothesis H_o: $\mu_1 = \mu_2$ after utilizing the two types of preliminary tests of variance homogeneity for each of the contaminated normal distributions.

Preliminary testing for the symmetric contaminated normal distribution is conducted in the same manner as described in Chapter III for symmetric cases. When the preliminary test for variance homogeneity is found not to be significant the t test is employed; otherwise, the Welch test is used.

Preliminary testing for the asymmetric contaminated normal distribution is conducted in the same manner as described in Chapter IV for asymmetric cases. When the test for variance homogeneity is found to be not significant at the specified level, the Mann-Whitney-Wilcoxon test is employed; otherwise the Welch test is used.

For both the asymmetric and symmetric contaminated normal cases, the test of equality of means is conducted at a significance level of 5%.

Tables 23 and 24 present the simulated null rejection rates where the proportion of rejections is expressed as a percent for the symmetric and asymmetric contaminated normal distributions, respectively. Each entry is again the result of ten thousand simulation runs. The format of Tables 23 and 24 is the same format as described in Chapter III for Tables 4-7. For each distribution, the results are given for the seven selected sample size configurations. For each of the seven sample size combinations, the simulated null rejection rate is reported for the three cases of direct pairing, the three cases of indirect pairing, in addition to the case of variance homogeneity. The test of variance homogeneity in the preliminary test procedures is again conducted at significance levels of 0.05, 0.25, 0.50, and 0.75.

The results are presented, but not individually discussed for the symmetric and asymmetric cases. An overall assessment is made in the next section.

5.2 Overall Performance for the Contaminated Normal Distributions

To evaluate the overall performance of the test procedures for

Table 23

Symmetric Contaminated Normal with $p = 0.10$: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	0.25	0.50	0.67	1.00	1.50	2.00	4.00
5	5	t	3.07	3.48	3.76	4.15	3.76	3.48	3.07
		F(.05)	2.90	3.32	3.60	3.97	3.60	3.32	2.90
		F(.25)	2.49	2.81	3.09	3.50	3.09	2.81	2.49
		F(.50)	2.38	2.67	2.94	3.28	2.94	2.67	2.38
		F(.75)	2.35	2.61	2.87	3.20	2.87	2.61	2.35
		L(.05)	3.01	3.44	3.74	4.13	3.74	3.44	3.01
		L(.25)	2.75	3.15	3.41	3.79	3.41	3.15	2.75
		L(.50)	2.46	2.77	3.06	3.38	3.06	2.77	2.46
		L(.75)	2.36	2.62	2.91	3.24	2.91	2.62	2.36
		W	2.29	2.54	2.78	3.13	2.78	2.54	2.29
		MW	5.63	5.68	5.73	5.71	5.73	5.68	5.63
5	10	t	2.41	3.09	3.46	4.31	4.62	4.49	4.24
		F(.05)	2.27	3.00	3.41	4.00	3.48	3.09	2.41
		F(.25)	2.31	3.11	3.60	3.92	3.41	3.07	2.64
		F(.50)	2.34	3.27	3.78	4.02	3.46	3.16	2.80
		F(.75)	2.36	3.27	3.78	3.93	3.38	3.08	2.72
		L(.05)	2.36	3.04	3.41	4.30	4.44	4.22	3.47
		L(.25)	2.37	3.11	3.61	4.38	3.93	3.57	2.90
		L(.50)	2.37	3.24	3.78	4.26	3.79	3.41	2.87
		L(.75)	2.43	3.38	3.89	4.03	3.50	3.19	2.79
		W	2.32	3.18	3.73	3.82	3.25	3.00	2.66
		MW	4.03	4.09	4.12	4.02	4.13	4.16	4.20

			Ratio = σ_1/σ_2						
n_1	n_2	Test	0.25	0.50	0.67	1.00	1.50	2.00	4.00
5	20	*t*	1.68	2.53	3.24	4.94	7.87	9.84	14.95
		F(.05)	2.36	3.13	3.42	4.28	3.84	3.25	2.62
		F(.25)	2.76	3.80	4.43	4.99	4.09	3.94	3.36
		F(.50)	2.73	3.95	4.59	4.96	4.21	4.02	3.48
		F(.75)	2.63	3.82	4.46	4.71	4.05	3.79	3.34
		L(.05)	1.48	2.29	2.98	4.83	6.36	7.36	9.73
		L(.25)	1.51	2.42	3.30	5.15	4.93	4.55	3.80
		L(.50	2.47	3.63	4.34	5.38	4.63	4.28	3.67
		L(.75)	2.62	3.88	4.58	5.00	4.21	3.95	3.52
		W	2.55	3.70	4.23	4.41	3.80	3.53	3.18
		MW	4.46	4.64	4.77	4.92	5.19	5.25	5.41
10	10	*t*	2.41	3.22	3.64	4.34	3.64	3.22	2.41
		F(.05)	2.26	3.01	3.46	4.24	3.46	3.01	2.26
		F(.25)	2.15	2.90	3.36	4.05	3.36	2.90	2.15
		F(.50)	2.15	2.89	3.36	4.02	3.36	2.89	2.15
		F(.75)	2.15	2.89	3.35	4.01	3.35	2.89	2.15
		L(.05)	2.40	3.21	3.63	4.30	3.63	3.21	2.40
		L(.25)	2.29	3.08	3.53	4.17	3.53	3.08	2.29
		L(.50)	2.19	2.96	3.37	4.07	3.37	2.96	2.19
		L(.75)	2.15	2.89	3.35	4.03	3.35	2.89	2.15
		W	2.13	2.88	3.34	3.99	3.34	2.88	2.13
		MW	5.24	5.21	5.22	5.20	5.22	5.21	5.24

			Ratio = σ_1/σ_2						
n_1	n_2	Test	0.25	0.50	0.67	1.00	1.50	2.00	4.00
10	20	t	1.81	2.83	3.64	4.49	5.35	5.86	5.95
		F(.05)	2.35	3.58	4.09	4.22	3.49	3.09	2.12
		F(.25)	2.36	3.76	4.29	4.16	3.44	2.91	2.14
		F(.50)	2.36	3.82	4.30	4.18	3.37	2.88	2.17
		F(.75)	2.32	3.82	4.28	4.13	3.30	2.78	2.12
		L(.05)	1.73	2.75	3.51	4.40	4.99	5.22	4.45
		L(.25)	1.73	2.77	3.60	4.42	4.21	3.86	2.72
		L(.50)	2.19	3.38	3.98	4.24	3.75	3.24	2.37
		L(.75)	2.30	3.69	4.25	4.22	3.42	2.93	2.19
		W	2.27	3.78	4.25	4.04	3.20	2.74	2.06
		MW	5.06	5.14	5.23	4.83	4.96	4.99	5.08
10	40	t	0.49	1.54	2.59	5.25	9.54	13.00	20.71
		F(.05)	3.45	3.99	4.05	4.55	3.98	3.29	2.21
		F(.25)	3.56	4.35	4.56	4.71	3.79	3.22	2.37
		F(.50)	3.58	4.37	4.60	4.70	3.80	3.31	2.49
		F(.75)	3.57	4.33	4.53	4.56	3.74	3.30	2.50
		L(.05)	0.45	1.45	2.42	4.91	7.12	8.49	10.5
		L(.25)	0.78	2.04	3.11	4.87	4.87	4.59	3.15
		L(.50)	2.79	3.56	4.09	4.92	4.30	3.85	2.76
		L(.75)	3.37	4.08	4.37	4.73	4.03	3.42	2.55
		W	3.56	4.31	4.40	4.47	3.71	3.20	2.44
		MW	4.31	4.44	4.54	5.48	5.72	5.79	5.89

			Ratio = σ_1/σ_2						
n_1	n_2	Test	0.25	0.50	0.67	1.00	1.50	2.00	4.00
20	20	*t*	2.09	3.32	3.32	4.49	3.32	3.32	2.09
		F(.05)	1.85	3.13	3.13	4.47	3.13	3.13	1.85
		F(.25)	1.85	3.12	3.12	4.46	3.12	3.12	1.85
		F(.50)	1.85	3.12	3.12	4.45	3.12	3.12	1.85
		F(.75)	1.85	3.12	3.12	4.45	3.12	3.12	1.85
		L(.05)	2.08	3.31	3.31	4.49	3.31	3.31	2.08
		L(.25)	1.88	3.20	3.20	4.46	3.20	3.20	1.88
		L(.50)	1.85	3.13	3.13	4.46	3.13	3.13	1.85
		L(.75)	1.85	3.12	3.12	4.46	3.12	3.12	1.85
		W	1.85	3.12	3.12	4.45	3.12	3.12	1.85
		MW	4.74	4.71	4.71	4.59	4.71	4.71	4.74

varying degrees of variance heterogeneity, the results of the simulation for the symmetric contaminated normal and the asymmetric contaminated normal are combined across the seven selected sample size pairs in Tables 25-29 and Tables 30-34, respectively. The format used in Tables 25-29 and Tables 30-34 is the same format as used in Chapter IV for Tables 18-22 where the five groups are determined by the $R = \sigma_1/\sigma_2$ values. These are: (1) $R = 0.25$ and 0.50, (2) $R = 0.67$, (3) $R = 1$ (equal variances), (4) $R = 1.50$, (5) $R = 2.0$ and 4.0. The proportion of rejections is expressed as a percent for the *t* test, the Welch test, the MW test, the preliminary F-ratio test procedures, and the preliminary Levene test procedures, which are tabulated for each R grouping over <u>all</u> (7) sample size pairs. The interval

Table 24

Asymmetric Contaminated Normal with p = 0.05: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	0.25	0.50	0.67	1.00	1.50	2.00	4.00
5	5	t	3.44	3.48	3.53	3.53	3.53	3.48	3.44
		F(.05)	3.71	3.85	3.91	3.88	3.91	3.85	3.71
		F(.25)	3.31	3.36	3.43	3.45	3.43	3.36	3.31
		F(.50)	3.02	3.05	3.10	3.12	3.10	3.05	3.02
		F(.75)	2.85	2.90	2.94	2.94	2.94	2.90	2.85
		L(.05)	5.60	5.57	5.51	5.52	5.51	5.57	5.60
		L(.25)	5.18	5.14	5.03	5.03	5.02	5.14	5.18
		L(.50)	3.25	3.34	3.42	3.40	3.42	3.34	3.25
		L(.75)	2.88	2.93	2.99	2.99	2.99	2.93	2.88
		W	2.64	2.68	2.72	2.74	2.72	2.68	2.64
		MW	5.62	5.59	5.53	5.54	5.53	5.59	5.62
5	10	t	3.35	3.41	3.53	3.69	3.55	3.38	3.13
		F(.05)	2.59	2.65	2.76	2.52	2.60	2.51	2.39
		F(.25)	2.64	2.83	2.91	2.65	2.71	2.59	2.50
		F(.50)	2.77	2.95	3.05	2.63	2.70	2.59	2.51
		F(.75)	2.80	2.98	3.09	2.70	2.72	2.66	2.58
		L(.05)	4.13	4.11	4.15	3.94	4.10	4.12	4.13
		L(.25)	3.94	3.96	3.99	3.67	3.73	3.63	3.43
		L(.50)	3.37	3.50	3.52	3.11	3.13	3.07	2.94
		L(.75)	3.04	3.24	3.33	2.78	2.81	2.68	2.63
		W	2.85	3.03	3.16	2.67	2.64	2.60	2.56
		MW	4.13	4.11	4.15	3.96	4.08	4.13	4.15

			Ratio = σ_y/σ_x						
n_1	n_2	Test	0.25	0.50	0.67	1.00	1.50	2.00	4.00
5	20	*t*	5.40	5.88	6.15	6.11	7.36	8.76	11.88
		F(.05)	2.45	2.90	3.02	4.22	4.04	3.94	3.87
		F(.25)	2.47	2.99	3.11	4.38	4.20	4.08	4.08
		F(.50)	2.34	2.84	3.01	4.26	4.11	4.00	4.00
		F(.75)	2.17	2.64	2.86	3.99	3.88	3.79	3.77
		L(.05)	4.98	4.93	4.89	4.93	4.95	4.92	4.75
		L(.25)	4.82	4.82	4.76	4.93	4.81	4.65	4.52
		L(.50)	4.25	4.30	4.20	5.14	4.88	4.82	4.76
		L(.75)	2.47	2.82	3.05	4.21	4.04	3.96	3.99
		W	1.96	2.43	2.67	3.73	3.58	3.51	3.52
		MW	5.12	5.12	5.07	5.12	5.18	5.23	5.27
10	10	*t*	2.51	2.68	2.73	2.87	2.73	2.68	2.51
		F(.05)	2.45	2.77	2.80	2.87	2.80	2.77	2.45
		F(.25)	2.28	2.51	2.59	2.64	2.59	2.51	2.28
		F(.50)	2.28	2.47	2.52	2.60	2.52	2.47	2.28
		F(.75)	2.33	2.48	2.52	2.63	2.52	2.48	2.33
		L(.05)	5.30	5.26	5.20	5.15	5.20	5.26	5.30
		L(.25)	4.70	4.65	4.61	4.54	4.61	4.65	4.70
		L(.50)	2.64	3.00	2.99	2.97	2.99	3.00	2.64
		L(.75)	2.44	2.66	2.71	2.79	2.71	2.66	2.44
		W	2.27	2.42	2.47	2.59	2.47	2.42	2.27
		MW	5.32	5.28	5.22	5.17	5.22	5.28	5.32

			Ratio = σ_1/σ_2						
n_1	n_2	Test	0.25	0.50	0.67	1.00	1.50	2.00	4.00
10	20	*t*	2.91	3.10	3.33	3.05	2.80	2.71	2.77
		F(.05)	1.70	2.27	2.51	2.93	2.76	2.67	2.41
		F(.25)	1.65	2.10	2.36	2.82	2.68	2.54	2.35
		F(.50)	1.65	2.06	2.34	2.65	2.62	2.48	2.30
		F(.75)	1.56	1.96	2.24	2.61	2.53	2.47	2.32
		L(.05)	5.21	5.14	5.12	4.87	4.81	4.85	4.84
		L(.25)	4.45	4.40	4.37	4.38	4.29	4.23	3.82
		L(.50)	2.58	2.97	3.11	3.30	3.14	3.02	2.72
		L(.75)	1.75	2.26	2.45	2.85	2.81	2.69	2.48
		W	1.51	1.91	2.10	2.56	2.52	2.45	2.31
		MW	5.33	5.25	5.21	4.98	4.95	4.97	4.97
10	40	*t*	3.12	4.21	4.75	5.48	7.16	9.52	15.42
		F(.05)	1.82	2.94	3.46	5.48	5.39	5.26	4.80
		F(.25)	1.61	2.66	3.25	5.23	5.29	5.20	4.74
		F(.50)	1.45	2.48	2.93	5.14	5.26	5.08	4.70
		F(.75)	1.38	2.36	2.82	5.03	5.13	4.98	4.61
		L(.05)	4.89	4.75	4.73	4.72	4.73	4.90	4.81
		L(.25)	4.73	4.48	4.41	4.71	4.53	4.51	4.03
		L(.50)	3.70	4.05	4.27	5.58	5.55	5.32	4.98
		L(.75)	2.02	2.84	3.27	5.14	5.27	5.07	4.73
		W	1.31	2.23	2.79	4.86	5.00	4.84	4.54
		MW	5.03	4.93	4.95	5.13	5.20	5.39	5.52

			Ratio = σ_1/σ_2						
n_1	n_2	Test	0.25	0.50	0.67	1.00	1.50	2.00	4.00
20	20	*t*	1.91	2.30	2.39	2.73	2.39	2.30	1.91
		F(.05)	1.78	2.33	2.51	2.83	2.51	2.33	1.78
		F(.25)	1.77	2.16	2.42	2.57	2.42	2.16	1.77
		F(.50)	1.72	2.13	2.29	2.46	2.29	2.13	1.72
		F(.75)	1.75	2.14	2.21	2.44	2.21	2.14	1.75
		L(.05)	5.26	5.21	5.14	5.20	5.14	5.21	5.26
		L(.25)	4.17	4.16	4.23	4.54	4.23	4.16	4.17
		L(.50)	2.34	2.80	3.03	3.24	3.03	2.80	2.34
		L(.75)	1.83	2.33	2.47	2.67	2.47	2.33	1.83
		W	1.73	2.06	2.17	2.36	2.17	2.06	1.73
		MW	5.22	5.15	5.06	5.14	5.06	5.15	5.22

$x \le 2.5$ is separated into two categories: $x \le 1$ and $1 < x \le 2.5$ to yield more information concerning trends for the severe conservative cases. The interval $x > 10$ is separated into two categories: $10 < x \le 20$ and $x > 20$ in order to examine more closely the patterns for the severe liberal cases. The seven categories are listed below:

1. $x \le 1$
2. $1 < x \le 2.5$
3. $2.5 < x \le 4$
4. $4 < x \le 6$
5. $6 < x \le 10$
6. $10 < x \le 20$

7. $x > 20$

The value "x" represents the percentage of rejections for testing H_0: $\mu_1 = \mu_2$ based on 10,000 simulations for each sample size configuration. Each entry of the following tables denotes the frequency at which $a < x \leq b$ occurs. A test procedure is defined to be severely conservative, extremely conservative, conservative, robust, liberal, extremely liberal, or severely liberal as defined in Chapter IV.

5.2.1 Symmetric Contaminated Normal (0,1) with $p = 0.10$

$R = 1$ (Equal Variance Cases)

A summary of the simulated null rejection rates for the symmetric contaminated normal distribution for the equal variance cases is presented in Table 25. The t test, the MW test, and the L(.05) test procedure are robust. The other preliminary test procedures are generally robust. The Welch test is slightly more conservative than the other procedures.

$R = 0.25$, 0.50, and 0.67 (Includes All Direct Pairing Cases)

Results of the simulated null rejection rates for the symmetric contaminated normal distribution for the $R = 0.25$, 0.50, and 0.67 cases are summarized in Tables 27 and 26, respectively. The MW test is robust, whereas the other procedures are conservative or extremely conservative.

Table 25

Summary of Symmetric Contaminated Normal Distribution With p=0.10: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
Equal Variances: $R=\sigma_1/\sigma_2 = 1.00$

test	$x \leq 1.0$	$1< x \leq 2.5$	$2.5< x \leq 4$	$4< x \leq 6$	$6< x \leq 10$	$10< x \leq 20$	$x >20$	Total
t	0	0	0	7	0	0	0	7
F(.05)	0	0	2	5	0	0	0	7
F(.25)	0	0	2	5	0	0	0	7
F(.50)	0	0	2	5	0	0	0	7
F(.75)	0	0	2	5	0	0	0	7
L(.05)	0	0	0	7	0	0	0	7
L(.25)	0	0	2	5	0	0	0	7
L(.50)	0	0	1	6	0	0	0	7
L(.75)	0	0	1	6	0	0	0	7
W	0	0	3	4	0	0	0	7
MW	0	0	0	7	0	0	0	7

*Preliminary test procedures used t test for equal variances and the Welch test for unequal variances.

Table is based on the seven sample size pairs.

Table 26

Summary of Symmetric Contaminated Normal Distribution With p=0.10: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
$R=\sigma_1/\sigma_2 = 0.67$

test	$x \le 1.0$	$1 < x \le 2.5$	$2.5 < x \le 4$	$4 < x \le 6$	$6 < x \le 10$	$10 < x \le 20$	$x > 20$	Total
t	0	0	6	0	1	0	0	7
F(.05)	0	0	6	1	0	0	0	7
F(.25)	0	0	5	2	0	0	0	7
F(.50)	0	0	5	2	0	0	0	7
F(.75)	0	0	5	2	0	0	0	7
L(.05)	0	0	6	0	1	0	0	7
L(.25)	0	0	6	1	0	0	0	7
L(.50)	0	0	5	2	0	0	0	7
L(.75)	0	0	4	3	0	0	0	7
W	0	0	5	2	0	0	0	7
MW	0	0	0	7	0	0	0	7

*Preliminary test procedures used *t* test for equal variances and the Welch test for unequal variances.

Table is based on the seven sample size pairs.

Table 27

Summary of Symmetric Contaminated Normal Distribution With p=0.10: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
$R=\sigma_1/\sigma_2$ = 0.25 and 0.50

test	$x \leq 1.0$	$1< x \leq 2.5$	$2.5< x \leq 4$	$4< x \leq 6$	$6< x \leq 10$	$10< x \leq 20$	$x >20$	Total
t	1	7	6	0	0	0	0	14
F(.05)	0	5	9	0	0	0	0	14
F(.25)	0	5	8	1	0	0	0	14
F(.50)	0	5	8	1	0	0	0	14
F(.75)	0	5	8	1	0	0	0	14
L(.05)	1	8	5	0	0	0	0	14
L(.25)	1	8	5	0	0	0	0	14
L(.50)	0	7	7	0	0	0	0	14
L(.75)	0	5	8	1	0	0	0	14
W	0	5	8	1	0	0	0	14
MW	0	0	0	14	0	0	0	14

*Preliminary test procedures used *t* test for equal variances and the Welch test for unequal variances.

Table is based on the seven sample size pairs.

The conservatism increases as the degree of variance heterogeneity increases.

R = 1.50, 2.0, and 4.0 (Includes All Indirect Pairing Cases)

Results of the simulated null rejection rates for the symmetric contaminated normal distribution for the *R* = 1.50, 2.0, and 4.0 cases are summarized in Tables 28 and 29, respectively. The MW test is robust, whereas the other procedures are conservative or extremely conservative. The conservatism increases as the degree of variance heterogeneity increases.

5.2.2 Asymmetric Contaminated Normal (0,1) with *p* = 0.05

R = 1 (Equal Variance Cases)

A summary of the simulated null rejection rates for the asymmetric contaminated normal distribution for the equal variance cases is presented in Table 30. Table 30 shows the MW test, the L(.05), and the L(.25) test procedures are robust. The *t* test, the Welch test, the preliminary F-ratio test procedures, and the remaining preliminary Levene test procedures are somewhat conservative.

R = 0.25, 0.50, and 0.67 (Includes All Direct Pairing Cases)

Results of the simulated null rejection rates for the asymmetric contaminated normal distribution for the *R* = 0.25, 0.50, and 0.67 cases are

Table 28

Summary of Symmetric Contaminated Normal Distribution With p=0.10: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
$R=\sigma_1/\sigma_2$ = 1.50

test	$x \leq 1.0$	$1 < x \leq 2.5$	$2.5 < x \leq 4$	$4 < x \leq 6$	$6 < x \leq 10$	$10 < x \leq 20$	$x > 20$	Total
t	0	0	3	2	2	0	0	7
F(.05)	0	0	7	0	0	0	0	7
F(.25)	0	0	6	1	0	0	0	7
F(.50)	0	0	6	1	0	0	0	7
F(.75)	0	0	6	1	0	0	0	7
L(.05)	0	0	3	2	2	0	0	7
L(.25)	0	0	4	3	0	0	0	7
L(.50)	0	0	5	2	0	0	0	7
L(.75)	0	0	5	2	0	0	0	7
W	0	0	7	0	0	0	0	7
MW	0	0	0	7	0	0	0	7

*Preliminary test procedures used t test for equal variances and the Welch test for unequal variances.

Table is based on the seven sample size pairs.

Table 29

Summary of Symmetric Contaminated Normal Distribution With p=0.10: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level $R=\sigma_1/\sigma_2$ = 2.00 and 4.00

test	$x \leq 1.0$	$1 < x \leq 2.5$	$2.5 < x \leq 4$	$4 < x \leq 6$	$6 < x \leq 10$	$10 < x \leq 20$	$x > 20$	Total
t	0	2	4	4	1	2	1	14
F(.05)	0	5	9	0	0	0	0	14
F(.25)	0	5	9	0	0	0	0	14
F(.50)	0	5	8	1	0	0	0	14
F(.75)	0	5	9	0	0	0	0	14
L(.05)	0	2	5	3	3	1	0	14
L(.25)	0	2	10	2	0	0	0	14
L(.50)	0	4	9	1	0	0	0	14
L(.75)	0	4	10	0	0	0	0	14
W	0	5	9	0	0	0	0	14
MW	0	0	0	14	0	0	0	14

*Preliminary test procedures used t test for equal variances and the Welch test for unequal variances.

Table is based on the seven sample size pairs.

Table 30

Summary of Asymmetric Contaminated Normal Distribution With p=0.05: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
Equal Variances: $R=\sigma_1/\sigma_2 = 1.00$

test	$x \leq 1.0$	$1< x \leq 2.5$	$2.5< x \leq 4$	$4< x \leq 6$	$6< x \leq 10$	$10< x \leq 20$	$x > 20$	Total
t	0	0	5	1	1	0	0	7
F(.05)	0	0	5	2	0	0	0	7
F(.25)	0	0	5	2	0	0	0	7
F(.50)	0	1	4	2	0	0	0	7
F(.75)	0	1	5	1	0	0	0	7
L(.05)	0	0	1	6	0	0	0	7
L(.25)	0	0	1	6	0	0	0	7
L(.50)	0	0	5	2	0	0	0	7
L(.75)	0	0	5	2	0	0	0	7
W	0	1	5	1	0	0	0	7
MW	0	0	1	6	0	0	0	7

*Preliminary test procedures used Mann-Whitney Wilcoxon test for equal variances and the Welch test for unequal variances.

Table is based on the seven sample size pairs.

Table 31

Summary of Asymmetric Contaminated Normal Distribution With p=0.05: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
$R=\sigma_1/\sigma_2$ = 0.67

test	$x \le 1$	$1< x \le 2.5$	$2.5< x \le 4$	$4< x \le 6$	$6< x \le 10$	$10< x \le 20$	$x > 20$	Total
t	0	1	4	1	1	0	0	7
F(.05)	0	0	7	0	0	0	0	7
F(.25)	0	2	5	0	0	0	0	7
F(.50)	0	2	5	0	0	0	0	7
F(.75)	0	2	5	0	0	0	0	7
L(.05)	0	0	0	7	0	0	0	7
L(.25)	0	0	1	6	0	0	0	7
L(.50)	0	0	5	2	0	0	0	7
L(.75)	0	2	5	0	0	0	0	7
W	0	3	4	0	0	0	0	7
MW	0	0	0	7	0	0	0	7

*Preliminary test procedures used Mann-Whitney Wilcoxon test for equal variances and the Welch test for unequal variances.

Table is based on the seven sample size pairs.

Table 32

Summary of Asymmetric Contaminated Normal Distribution With p=0.05: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level $R=\sigma_1/\sigma_2$ = 0.25 and 0.50

test	$x \le 1.0$	$1< x \le 2.5$	$2.5< x \le 4$	$4< x \le 6$	$6< x \le 10$	$10< x \le 20$	$x >20$	Total
t	0	2	9	3	0	0	0	14
F(.05)	0	7	7	0	0	0	0	14
F(.25)	0	7	7	0	0	0	0	14
F(.50)	0	9	5	0	0	0	0	14
F(.75)	0	9	5	0	0	0	0	14
L(.05)	0	0	0	14	0	0	0	14
L(.25)	0	0	2	12	0	0	0	14
L(.50)	0	1	10	3	0	0	0	14
L(.75)	0	7	7	0	0	0	0	14
W	0	10	4	0	0	0	0	14
MW	0	0	0	14	0	0	0	14

*Preliminary test procedures used Mann-Whitney Wilcoxon test for equal variances and the Welch test for unequal variances.

Table is based on the seven sample size pairs.

Table 33

Summary of Asymmetric Contaminated Normal Distribution With p=0.05: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
$R=\sigma_1/\sigma_2 = 1.50$

test	$x \leq 1.0$	$1 < x \leq 2.5$	$2.5 < x \leq 4$	$4 < x \leq 6$	$6 < x \leq 10$	$10 < x \leq 20$	$x > 20$	Total
t	0	1	4	0	2	0	0	7
F(.05)	0	0	5	2	0	0	0	7
F(.25)	0	1	4	2	0	0	0	7
F(.50)	0	1	4	2	0	0	0	7
F(.75)	0	1	5	1	0	0	0	7
L(.05)	0	0	0	7	0	0	0	7
L(.25)	0	0	1	6	0	0	0	7
L(.50)	0	0	5	2	0	0	0	7
L(.75)	0	1	4	2	0	0	0	7
W	0	2	4	1	0	0	0	7
MW	0	0	0	7	0	0	0	7

*Preliminary test procedures used Mann-Whitney Wilcoxon test for equal variances and the Welch test for unequal variances.

Table is based on the seven sample size pairs.

Table 34

Summary of Asymmetric Contaminated Normal Distribution With p=0.05: Frequency of Simulated Null Rejection Rate (%)* With Nominal 5% Level
$R=\sigma_1/\sigma_2$ = 2.00 and 4.00

test	$x \leq 1.0$	$1 < x \leq 2.5$	$2.5 < x \leq 4$	$4 < x \leq 6$	$6 < x \leq 10$	$10 < x \leq 20$	$x > 20$	Total
t	0	2	8	0	2	0	0	14
F(.05)	0	5	7	2	0	0	0	14
F(.25)	0	5	5	4	0	0	0	14
F(.50)	0	6	6	2	0	0	0	14
F(.75)	0	6	6	2	0	0	0	14
L(.05)	0	0	0	14	0	0	0	14
L(.25)	0	0	3	11	0	0	0	14
L(.50)	0	1	9	4	0	0	0	14
L(.75)	0	4	8	2	0	0	0	14
W	0	6	6	2	0	0	0	14
MW	0	0	0	14	0	0	0	14

*Preliminary test procedures used Mann-Whitney Wilcoxon test for equal variances and the Welch test for unequal variances.

Table is based on the seven sample size pairs.

summarized in Tables 32 and 31, respectively. The MW test, the L(.05) and the L(.25) test procedures are robust for the R = 0.25, 0.50, and 0.67 cases. The t test, the Welch test, and the other preliminary test procedures are conservative and can be extremely conservative. As the degree of variance heterogeneity increases, the conservatism of these procedures increases.

R = 1.50, 2.0, and 4.0 (Includes All Indirect Pairing Cases)

Results of the simulated null rejection rates for the asymmetric contaminated normal distribution for the R = 1.50, 2.0, and 4.0 cases are summarized in Tables 33 and 34, respectively. The MW test, the L(.05), and the L(.25) test procedures are robust for these cases. The t test, the Welch test, and the other preliminary test procedures are conservative and can be extremely conservative. As the degree of variance heterogeneity increases, the conservatism of these test procedures increases.

5.3 Summary

For the symmetric contaminated normal distribution, the MW test, the t test, and the L(.05) test procedure are robust for the R = 1 cases. For the $R \neq 1$ cases, only the MW test is robust. The other procedures are conservative and can be extremely conservative. The degree of conservatism tends to increase as the degree of variance heterogeneity increases.

For the asymmetric contaminated normal distribution, the MW test, the L(.05), and the L(.25) test procedures are robust for the $R = 1$ and $R \neq 1$ cases. The other procedures tend to be conservative or extremely conservative for these cases. As the degree of variance heterogeneity increases, the conservatism of these procedures tends to increase.

Clearly, the MW test is robust for the contaminated normal cases for both the symmetric and asymmetric cases. No other procedure is as robust as the MW test for these outlier model cases. It would be of interest to evaluate the performance of other outlier models than those considered here. However, for dissertation purposes, this concludes the discussion of preliminary tests for variance homogeneity in detecting mean differences in outlier models.

CHAPTER VI

SELECTING COMPONENTS OF EXPERT SYSTEM

6.1 Introduction

One objective of this dissertation work is to evaluate the performance of an expert system which selects a method for testing H_o: $\mu_1 = \mu_2$ based on two classes of preliminary tests: (1) a test of variance homogeneity, and (2) a test of symmetry/asymmetry. A preliminary test for testing of symmetry/asymmetry is examined because skewness adversely affects the robustness of the *t* test and the Welch test even for the equal variance cases (Murphy, 1976; Gans, 1981; and Olejnik, 1987). The D'Agostino S_U test for skewness and the Triples test for symmetry were selected for examination, and their performance is compared in Section 6.2.

Based on this evaluation, a preliminary test for symmetry/asymmetry at a significance level α^{**} will be chosen and implemented into an expert system. Using the simulation results in Chapters III and IV, a preliminary test for variance homogeneity at a significance level α^{*} will be chosen and implemented into an expert system (Section 6.3). The results of an expert system which consolidates a symmetry test and a variance homogeneity test are discussed in Chapter VII. The outlier model cases are beyond the focus of this dissertation work and will not be included when evaluating the performance of the expert system.

6.2 Selecting Methodology for Testing of Symmetry

The robustness and the power of the D'Agostino S_U test for skewness at significance levels of 0.05, 0.25, 0.50, and 0.75, denoted D(α), and the Triples test for symmetry at significance levels of 0.05, 0.25, 0.50, and 0.75, denoted as T(α), are examined in this section. These two methods are evaluated for the one sample cases and the two sample cases, which are described in the sections below. To assess the Type I error, the simulated null rejection rates are examined for the four symmetric distributions (normal, uniform, double exponential, and logistic). The asymmetric distributions (lognormal and gamma) are used to investigate the power of the two procedures.

6.2.1 One Sample Case

For the one sample case, n random realizations are generated from a distribution $f(x; \mu, \sigma)$ where n = 5, 10, 20, or 40. The hypothesis of symmetry is tested using the D'Agostino S_U test and the Triples test. For the n = 5 case, only the Triples test can be conducted as the sample size must be greater than seven for the D'Agostino S_U test to be defined.

Robustness

Tables 35-38 display the simulated null rejection rates for testing symmetry with respect to the four symmetric distributions. These results are presented but are not individually discussed. We will focus on an overall assessment with respect to the D(.05), D(.25), T(.05), and T(.25) tests.

To evaluate the performance of the two methods for testing symmetry at significance levels $\alpha = 0.05$ and 0.25, the results of the simulation for the four symmetric distributions are combined in the following manner. The proportion of rejections is expressed as a percent for the D(.05) test and the T(.05) test which is tabulated over all sample sizes for the seven categories listed below:

1. $x \leq 1.0$
2. $1.0 < x \leq 2.5$
3. $2.5 < x \leq 4.0$
4. $4.0 < x \leq 6.0$
5. $6.0 < x \leq 10.0$
6. $10.0 < x \leq 20.0$
7. $x > 20.0$

The value "x" represents the percentage of rejections for testing H_o: symmetry based on 10,000 simulations for each sample size. Each entry in the following tables denotes the frequency at which $a < x \le b$ occurs. The percentage at which $a < x \le b$ occurs is also given because the D'Agostino S_U test results are based on 12 observations rather than 16 due to the constraint that n must be greater than 8.

The outcome of the D(.05) test and the T(.05) test is defined to be robust if the simulated null rejection rate is > 4.0 and ≤ 6.0.

For the D(.25) test and the T(.25) test, the proportion of rejections (%) is tabulated over <u>all</u> sample sizes for the seven categories listed below:

1. $x \le 17.5$
2. $17.5 < x \le 32.5$
3. $32.5 < x \le 37.5$
4. $37.5 < x \le 42.5$
5. $42.5 < x \le 62.5$
6. $62.5 < x \le 80.0$
7. $x > 80.0$

The outcome of the D(.25) test and the T(.25) test is defined to be robust if the simulated null rejection rate is > 17.5 and ≤ 32.5.

Table 39 presents frequency of simulated null rejection rate (%) for testing symmetry with significance levels of 0.05 and 0.25 for the four symmetric distributions. Results show that the T(.25) test is robust for 93.8% of the simulated cases. The D(.05), T(.05), and D(.25) tests lack robustness as robustness is seen in only about 33% of the symmetric cases

Table 35

Normal Distribution: Simulated Null Rejection Rate (%) of Symmetry for the One Sample Case

test	$n = 5$	$n = 10$	$n = 20$	$n = 40$
D(.05)	--	5.30	4.66	4.55
D(.25)	--	24.09	24.57	24.30
D(.50)	--	48.54	49.34	49.60
D(.75)	--	73.73	74.48	75.15
T(.05)	2.95	1.64	3.63	4.52
T(.25)	24.45	16.34	20.63	22.13
T(.50)	50.06	40.72	44.96	46.56
T(.75)	85.99	69.46	71.15	72.49

Table 36

Uniform Distribution: Simulated Null Rejection Rate (%) of Symmetry for the One Sample Case

test	$n = 5$	$n = 10$	$n = 20$	$n = 40$
D(.05)	--	1.88	0.67	0.20
D(.25)	--	15.46	10.85	7.98
D(.50)	--	39.44	34.61	30.55
D(.75)	--	69.08	65.36	62.74
T(.05)	3.07	1.86	5.43	6.26
T(.25)	21.86	19.18	23.81	24.97
T(.50)	52.21	44.93	47.54	49.17
T(.75)	88.71	71.89	73.00	74.38

Table 37

Double Exponential Distribution: Simulated Null Rejection Rate (%) of Symmetry for the One Sample Case

test	$n = 5$	$n = 10$	$n = 20$	$n = 40$
D(.05)	--	17.70	25.84	33.17
D(.25)	--	43.68	52.70	57.95
D(.50)	--	65.09	71.54	75.07
D(.75)	--	82.65	86.68	88.26
T(.05)	3.98	3.38	5.38	6.14
T(.25)	29.87	21.17	22.95	24.15
T(.50)	54.75	46.03	47.02	47.57
T(.75)	85.46	72.01	72.97	73.46

Table 38

Logistic Distribution: Simulated Null Rejection Rate (%) of Symmetry for the One Sample Case

test	$n = 5$	$n = 10$	$n = 20$	$n = 40$
D(.05)	--	10.05	13.46	16.99
D(.25)	--	31.61	37.49	41.75
D(.50)	--	54.91	60.11	63.42
D(.75)	--	77.17	80.39	82.65
T(.05)	3.19	2.12	4.53	4.88
T(.25)	26.46	17.83	21.69	23.35
T(.50)	51.57	42.53	45.86	47.82
T(.75)	85.59	70.16	72.04	73.51

Table 39

Summary of Symmetric Distributions: Frequency of Simulated Null Rejection Rate (%) for Symmetry Test With Nominal 5% and 25% Levels One Sample Case

Nominal 5% Level								
test	≤1.0	>1.0, ≤2.5	>2.5, ≤4.0	Robust >4.0, ≤6.0	>6.0, ≤7.5	>7.5, ≤10	>10, ≤20	>20
D(.05)	2 (16.7%)	1 (8.3%)	0 (0.0%)	3 (25.0%)	0 (0.0%)	0 (0.0%)	4 (33.3%)	2 (16.7%)
T(.05)	0 (0.0%)	3 (18.8%)	6 (37.5%)	5 (31.3%)	2 (12.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Nominal 25% Level								
test	≤17.5	Robust >17.5, ≤32.5	>32.5, ≤37.5	>37.5, ≤42.5	>42.5,≤52.5	>52.5, ≤62.5	>62.5,≤80.0	>80.0
D(.25)	3 (25.0%)	4 (33.3%)	1 (8.3%)	1 (8.3%)	1 (8.3%)	2 (16.7%)	0 (0.0%)	0 (0.0%)
T(.25)	1 (6.3%)	15 (93.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

Table is based on the four symmetric distributions (normal, uniform, double exponential, and logistic) and seven sample sizes (six sample sizes for D'Agostino S_U approximation).

(25.0%, 31.3%, and 33.3% respectively).

Results of Power Analysis

The results of a power comparison of the D'Agostino S_U test and the Triples test is now reported. For each of the asymmetric distributions, the null rejection rate (%) (equivalent to power for the asymmetric cases) is presented in Tables 40-44. The results presented in Tables 40-44 are not individually discussed. However, an overall assessment of these results is now discussed. The results of the simulation for the five asymmetric distributions were combined over all sample sizes for the eight categories below:

1. $x \leq 25.0$
2. $25.0 < x \leq 50.0$
3. $50.0 < x \leq 75.0$
4. $75.0 < x \leq 90.0$
5. $90.0 < x \leq 95.0$
6. $95.0 < x \leq 98.0$
7. $98.0 < x \leq 99.5$
8. $x > 99.5$

The value "x" represents the power to detect asymmetry based on 10,000 simulations for each sample size configuration. Each entry in the following tables denotes the frequency at which $a < x \leq b$ occurs. As in

Table 40

Lognormal (0, 0.40) Distribution: Simulated Null Rejection Rate (%) of Symmetry for the One Sample Case

test	*n* = 5	*n* = 10	*n* = 20	*n* = 40
D(.05)	--	18.43	37.17	67.56
D(.25)	--	45.30	67.07	89.67
D(.50)	--	66.91	82.62	96.09
D(.75)	--	84.06	91.82	98.65
T(.05)	4.02	5.65	25.24	58.44
T(.25)	28.56	32.99	59.00	84.80
T(.50)	54.80	59.78	78.18	93.63
T(.75)	87.35	80.63	90.02	97.64

Table 41

Lognormal (0, 1) Distribution: Simulated Null Rejection Rate (%) of Symmetry for the One Sample Case

test	*n* = 5	*n* = 10	*n* = 20	*n* = 40
D(.05)	--	53.80	86.63	99.49
D(.25)	--	78.50	97.09	99.97
D(.50)	--	89.64	99.23	100.00
D(.75)	--	95.55	99.81	100.00
T(.05)	9.51	31.47	82.01	99.35
T(.25)	44.65	70.85	95.96	99.95
T(.50)	68.77	86.93	98.90	99.99
T(.75)	91.57	94.53	99.73	100.00

Table 42

Lognormal (0, 1.75) Distribution: Simulated Null Rejection Rate (%) of Symmetry for the One Sample Case

test	$n = 5$	$n = 10$	$n = 20$	$n = 40$
D(.05)	--	76.91	98.09	100.00
D(.25)	--	92.94	99.83	100.00
D(.50)	--	97.66	99.98	100.00
D(.75)	--	99.19	100.00	100.00
T(.05)	20.17	62.68	97.57	100.00
T(.25)	60.26	90.04	99.77	100.00
T(.50)	80.71	96.64	99.98	100.00
T(.75)	95.09	98.79	100.00	100.00

Table 43

Gamma (3, 1) Distribution: Simulated Null Rejection Rate (%) of Symmetry for the One Sample Case

test	$n = 5$	$n = 10$	$n = 20$	$n = 40$
D(.05)	--	16.78	34.93	66.08
D(.25)	--	44.28	67.54	90.23
D(.50)	--	65.85	83.53	96.45
D(.75)	--	84.12	92.94	98.77
T(.05)	4.11	6.31	27.42	62.05
T(.25)	28.41	34.40	61.76	86.52
T(.50)	54.49	60.54	79.67	94.38
T(.75)	87.40	81.60	91.38	97.87

Table 44

Gamma (2, 1) Distribution: Simulated Null Rejection Rate (%) of Symmetry for the One Sample Case

test	*n* = 5	*n* = 10	*n* = 20	*n* = 40
D(.05)	--	21.76	47.16	80.52
D(.25)	--	52.34	78.15	96.03
D(.50)	--	72.78	90.29	98.96
D(.75)	--	87.82	96.17	99.68
T(.05)	5.11	9.30	41.77	79.13
T(.25)	31.82	42.52	74.40	95.02
T(.50)	57.94	67.54	88.43	98.41
T(.75)	88.52	85.58	95.32	99.50

Tables 39, the percentages are given as the D'Agostino S_U approximation test results are based on 15 observations rather than 20 due to the constraint that *n* must be greater than 8.

Table 45 shows that the T(.05) and D(.05) tests lack power. The power is ≤ 0.75 for 60% of the cases when using the D(.05) test and is ≤ 0.75 for 75% of the cases when using the T(.05) test. The D(.25) test tends to reject symmetry more often than the T(.25) test. The power is > .95 for 33% of the cases when using the D(.25) test compared to 25% of the cases when using the T(.25) test. However, the power is ≤ 0.50 for 35% of the cases when using the T(.25) test compared to approximately 13% when using the D(.25) test.

Table 45

Summary of Asymmetric Distributions: Frequency of Simulated Power Rate (%) for Symmetry Test With Nominal 5% and 25% Levels One Sample Case

test	≤25.0	>25.0, ≤50.0	>50.0, ≤75.0	>75.0, ≤90.0	>90.0, ≤95.0	>95.0, ≤98.0	>98.0, ≤99.5	>99.5
D(.05)	3 (20.0%)	3 (20.0%)	3 (20.0%)	3 (20.0%)	0 (0.0%)	0 (0.0%)	2 (13.3%)	1 (6.7%)
T(.05)	8 (40.0%)	4 (20.0%)	3 (15.0%)	2 (10.0%)	0 (0.0%)	1 (5.0%)	1 (5.0%)	1 (5.0%)
D(.25)	0 (0.0%)	2 (13.3%)	3 (20.0%)	3 (20.0%)	2 (13.3%)	2 (13.3%)	0 (0.0%)	3 (20.0%)
T(.25)	0 (0.0%)	7 (35.0%)	5 (25.0%)	2 (10.0%)	1 (5.0%)	2 (10.0%)	0 (0.0%)	3 (15.0%)

Table is based on the five asymmetric distributions [lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3,1), & G(2,1)] and four sample sizes (three sample sizes for D'Agostino S_U approximation).

Based on the results of Tables 39-45, the T(.25) test is recommended for the one sample case.

6.2.2 Two Sample Case

In the expert system, a preliminary test of symmetry will be conducted where both samples are individually tested. If the null hypothesis is rejected in either case, asymmetry is declared (denoted as the two sample case). In this section, the robustness and power of the D'Agostino S_U test and the Triples test for the two sample case is evaluated. Since the two samples are assumed to be independent and symmetry is rejected if at least one sample is significant at level α, the Prob(Type I error) for testing at significance level α is:

$$\text{Prob(Type I error)} = 1 - (1 - \alpha)^2. \quad (6.1)$$

Therefore, the Prob(Type I error) is 9.75% and 43.75% for the cases of α = 0.05 and 0.25, respectively.

For the two sample case, two samples are generated for the R = 1 case in the same manner as described for testing the hypothesis of variance homogeneity. The same sample size configurations used in Chapters III and IV are used for the two sample case.

Robustness

Table 46 presents a summary of the simulated null rejection rates for the two sample case using a format similar to the one used in Table 39. For the $\alpha = 0.05$ cases where the Prob(Type I error) is 9.75%, Table 46 shows the T(.05) test is more robust than the D(.05) test. The Type I error rate using the T(.05) test is between 7.5% and 10.0% for almost one third of the simulated cases compared to approximately 8% when using the D(.05) test. Table 46 shows at least 60% of the cases are between 6.0% and 10.0% when using the T(.05) test compared to approximately 8% when using the D(.05) test.

For the $\alpha = 0.25$ cases where the Prob(Type I error) is 43.75%, Table 46 shows that the T(.25) test is more robust than the D(.25) test. The Type I error rate using the T(.25) test is between 37.5% and 52.5% for over 70% of the simulated cases compared to 25% when using the D(.25) test.

Based on the results of Table 46, the Triples test is recommended over the D'Agostino S_U test for both $\alpha = 0.05$ and 0.25 cases. The T(.25) test tends to be more robust than the T(.05) test. It is noted that the percentage of occurrences are displayed due to the dissimilar observations counts utilized in the two methods. The D'Agostino S_U test can not be conducted if $n < 8$, and therefore, only 24 observations could be used rather than the 28 observations for the Triples test.

Table 46

Summary of Symmetric Distributions: Frequency of Simulated Null Rejection Rate (%) for Symmetry Test With Nominal 5% and 25% Levels Two Sample Case

Nominal 5%								
test	≤1.0	>1.0, ≤2.5	>2.5, ≤4.0	>4.0, ≤6.0	>6.0, ≤7.5	Robust >7.5, ≤10	>10, ≤20	>20
D(.05)	1 (4.2%)	4 (16.7%)	1 (4.2%)	2 (8.3%)	0 (0.0%)	2 (8.3%)	6 (25.0%)	8 (33.3%)
T(.05)	0 (0.0%)	0 (0.0%)	2 (7.1%)	7 (25.0%)	8 (28.6%)	9 (32.1%)	2 (7.1%)	0 (0.0%)
Nominal 25%								
test	≤17.5	>17.5, ≤32.5	>32.5, ≤37.5	Robust >37.5, ≤42.5	Robust >42.5,≤52.5	>52.5, ≤62.5	>62.5,≤80.0	>80.0
D(.25)	2 (8.3%)	7 (29.2%)	1 (4.2%)	0 (0.0%)	6 (25.0%)	4 (16.7%)	4 (16.7%)	0 (0.0%)
T(.25)	0 (0.0%)	1 (3.6%)	7 (25.0%)	15 (53.6%)	5 (17.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

Table is based on the four symmetric distributions (normal, uniform, double exponential, and logistic) and seven sample size pairs (six sample size pairs for D'Agostino S_U test).

Results of Power Analysis

Table 47 presents the results of the simulation for the five asymmetric distributions combined over all sample sizes using a format similar to the one used in Table 45. For the $\alpha = 0.05$ case, Table 47 shows that both the D(.05) test and the T(.05) tests lack power. The power is $\leq$ .50 for more than 35% of the simulated cases when using the D(.05) test and 57.2% when using the T(.05) test. The D(.05) test also concludes asymmetry more often as the power is > .995 for approximately 13% of the simulated cases compared to approximately 6% when using the T(.05) test.

For the $\alpha = 0.25$ cases, Table 47 shows the D(.25) test tends to be more powerful than the T(.25) test. The power is > 0.90 for 50% of the simulated cases when using the D(.25) test compared to 40% when using the T(.25) test.

Based on the simulation results for the one sample and two sample cases, the T(.25) test is recommended over the D(.05), D(.25), and T(.05) tests; and therefore will be implemented into the expert system for testing symmetry.

6.3 Selecting Methodology For Testing of Variance Homogeneity

Results from Chapter III for the $R = 0.25$, 0.50, and 0.67 symmetric cases showed that the F(.05) and the L(.05) test procedures were

Table 47

Summary of Asymmetric Distributions: Frequency of Simulated Power Rate (%) for Symmetry Test With Nominal 5% and 25% Levels Two Sample Case

test	≤25.0	>25.0, ≤50.0	>50.0, ≤75.0	>75.0, ≤90.0	>90.0, ≤95.0	>95.0, ≤98.0	>98.0, ≤99.5	>99.5
D(.05)	3 (10.0%)	8 (26.7%)	7 (23.3%)	4 (13.3%)	1 (3.3%)	2 (6.7%)	1 (3.3%)	4 (13.3%)
T(.05)	10 (28.6%)	10 (28.6%)	5 (14.3%)	4 (11.4%)	0 (0.0%)	1 (2.9%)	3 (8.6%)	2 (5.7%)
D(.25)	0 (0.0%)	2 (6.7%)	5 (16.7%)	8 (26.7%)	3 (10.0%)	3 (10.0%)	2 (6.7%)	7 (23.3%)
T(.25)	0 (0.0%)	2 (5.7%)	12 (34.3%)	7 (20.0%)	3 (8.6%)	3 (8.6%)	2 (5.7%)	6 (17.1%)

Table is based on the five asymmetric distributions [lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3,1), & G(2,1) and seven sample size pairs (six sample size pairs for D'Agostino S_U approximation).

conservative. For the R = 1.50, 2.0, and 4.0 symmetric cases, the F(.05), F(.25), L(.05) and the L(.25) test procedures were liberal. The preliminary test procedures where $\alpha > 0.25$ were generally robust for all cases.

For the asymmetric cases examined in Chapter IV, the F(.05) and the L(.05) test procedures are not recommended. Results for the R = 0.25, 0.50,and 0.67 asymmetric cases showed the F(.50), the F(.75), and the L(.75) test procedures were the best preliminary test procedures with respect to robustness and comparable to the Welch test. For the R = 1.50, 2.0, and 4.0 asymmetric cases, all procedures were generally liberal or severely liberal and as the R value increased, the procedures became more severely liberal.

It was arbitrarily decided to select one Levene and one F-ratio preliminary test procedure to be included in an expert system. Based on these conclusions, the F(.05) and the L(.05) test procedures are eliminated as they do not control the Type I error rate for both the symmetric and asymmetric cases examined. The F(.75) and the L(.75) test procedures were also eliminated as these test procedures are in essence the Welch test. Therefore, of the F-ratio test procedures, the F(.50) test procedure was chosen to be implemented into an expert system because it generally performs better than the F(.25) test procedure. The L(.25) test procedure was chosen to provide a contrast to the F(.50) test procedure where a modest significance level is used that is > .05.

CHAPTER VII

RESULTS FROM EXPERT SYSTEM

7.1 Introduction

In this chapter, the performance of the expert system is evaluated. The expert system consists of two components. First a preliminary test of symmetry is conducted where both samples are individually tested. The T(.25) test is used to test the null hypothesis of symmetry. If the null hypothesis is rejected in either case, asymmetry is declared. The second component consists of a test for variance homogeneity between the two sample groups. The F-ratio test at a significance level 0.50 and the Levene test at a significance level 0.25 are used to test the null hypothesis of variance homogeneity. The expert system is constructed in the following way:

Case I: If $\sigma_1 = \sigma_2$ and symmetry is concluded, then the *t* test is used.

Case II: If $\sigma_1 \neq \sigma_2$ and symmetry is rejected, then the Welch test is used.

Case III: If $\sigma_1 = \sigma_2$ and symmetry is rejected, then the MW test is used.

Case IV: If $\sigma_1 \neq \sigma_2$ and symmetry is rejected, then the Welch test is used.

It is noted that robust methods exist for testing H_o: $\mu_1 = \mu_2$ for Cases I-III, but no robust method exists for Case IV.

The Fprel test procedure denotes the results of the expert system using the T(.25) test and the 50% F-ratio test for testing variance homogeneity while the results using the T(.25) test and the 25% Levene test are denoted as the Lprel test procedure. Figures 4 and 5 show the flowchart of the expert system using the Fprel test procedure and the Lprel test procedure, respectively.

7.2 Results From Symmetric Distributions

Tables 48-51 present the simulated null rejection rates, where the proportion of rejections is expressed as a percent for each of the four symmetric distributions in a format similar to the one used in Tables 4-7 in Chapter III. The simulated null rejection rates for each of the four symmetric distributions are included for completeness, but discussion will be limited to the overall performance of the test procedures as described below.

To evaluate the overall performance of the procedures for varying degree of variance heterogeneity, the results of the simulation for the four symmetric distributions are combined in Tables 52-55 using the same format as used in Tables 18-22 in Chapter IV. The outcome of a test

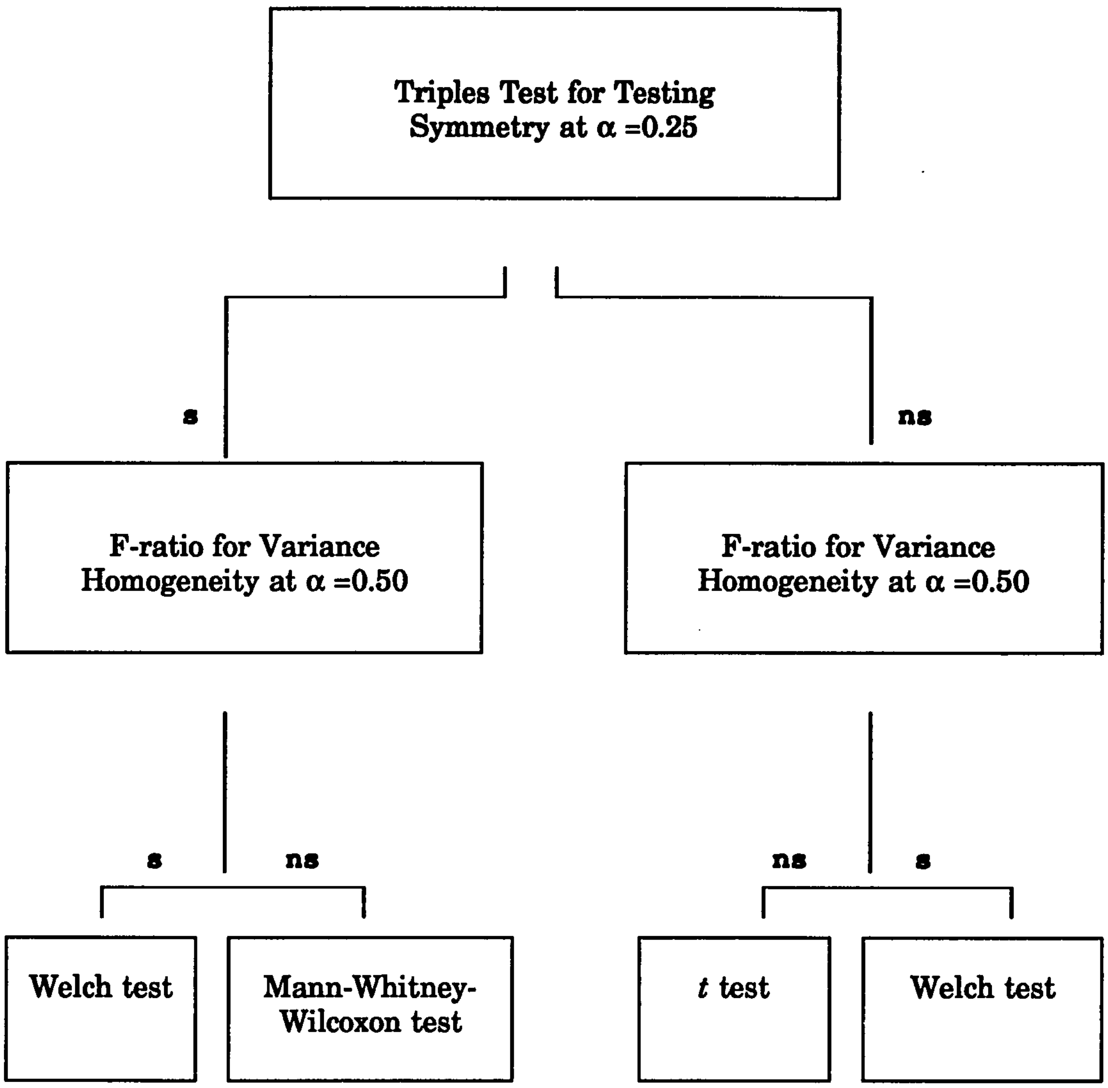

Figure 4. Expert System Using the Fprel Test Procedure.

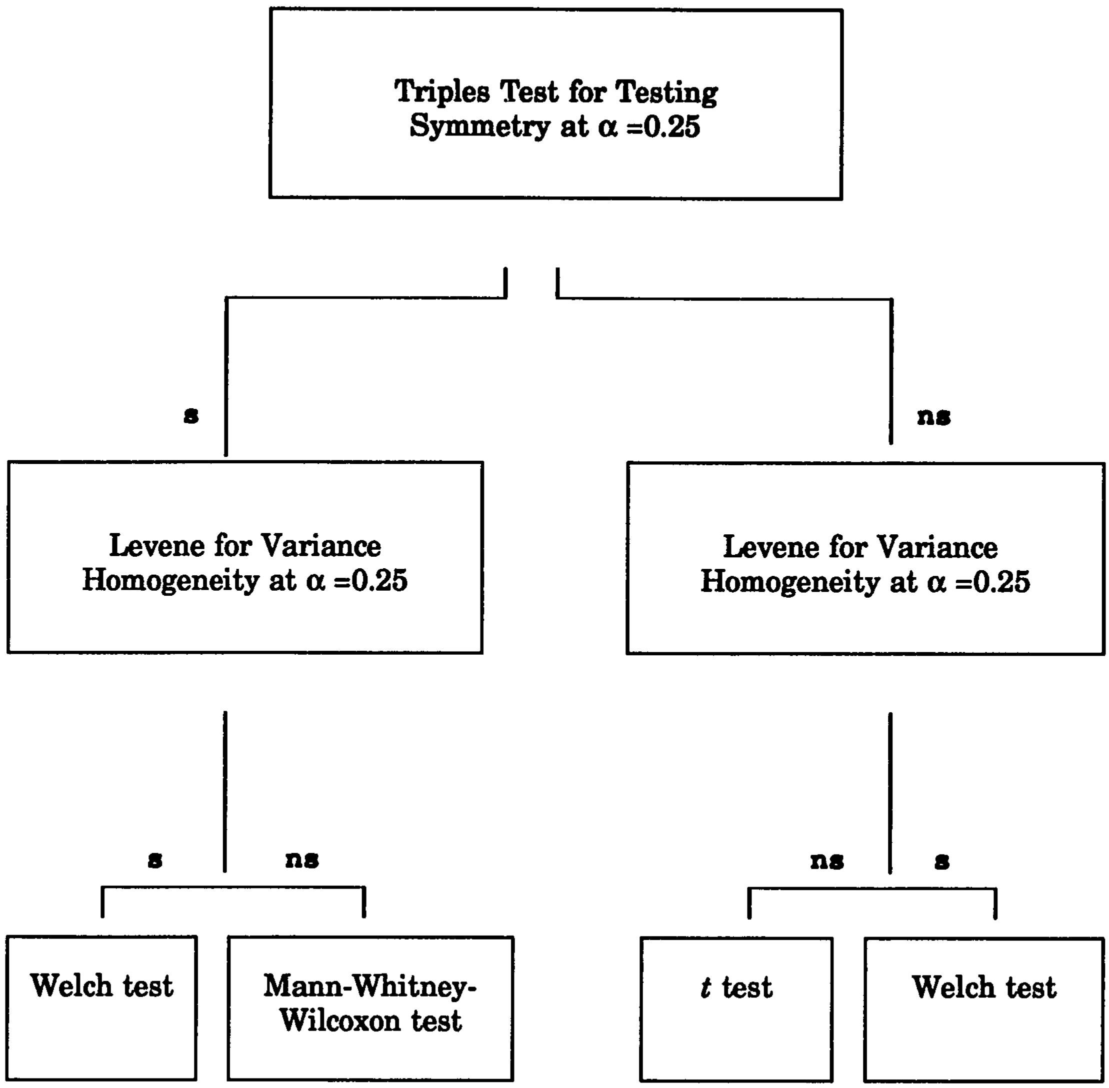

Figure 5. Expert System Using the Lprel Test Procedure.

Table 48

Normal Distribution Using the Expert System: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	7.35	5.86	5.24	4.97	5.24	5.87	7.31
		W	4.77	4.48	4.21	3.95	4.35	4.62	4.93
		MW	6.87	6.10	5.69	5.48	5.73	6.40	7.25
		Fprel	4.87	4.92	4.59	4.43	4.71	4.81	4.97
		Lprel	5.41	5.60	5.07	5.06	5.29	5.56	5.74
5	10	*t*	1.52	2.10	2.82	5.06	8.37	11.41	17.81
		W	4.81	4.65	4.74	4.96	5.21	5.08	4.65
		MW	2.55	2.54	2.85	4.10	5.63	6.84	8.28
		Fprel	4.78	4.63	4.60	5.15	5.63	5.44	4.70
		Lprel	4.69	4.15	4.23	5.41	6.91	7.45	5.81
5	20	*t*	0.13	0.62	1.36	5.01	12.26	18.85	31.53
		W	4.88	4.89	4.87	5.23	5.34	5.15	4.92
		MW	0.93	1.54	2.44	4.91	8.72	11.32	17.69
		Fprel	4.88	4.93	4.96	6.10	6.86	6.00	5.09
		Lprel	4.87	4.74	4.71	6.49	9.41	9.69	6.31
10	10	*t*	5.96	5.36	5.18	4.99	5.10	5.46	6.14
		W	4.81	4.79	4.89	4.79	4.74	4.86	4.90
		MW	7.76	5.92	5.53	5.14	5.33	5.99	7.95
		Fprel	4.81	4.83	4.88	4.90	4.89	4.84	4.90
		Lprel	4.86	5.07	5.17	5.09	5.20	5.13	4.95

Table 48--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	1.09	1.78	2.70	4.98	8.57	11.59	16.85
		W	4.76	4.86	4.88	5.10	5.14	5.23	5.17
		MW	3.11	3.08	3.54	4.94	7.02	8.83	12.00
		Fprel	4.76	4.88	4.79	5.34	5.47	5.30	5.17
		Lprel	4.77	4.77	4.56	5.48	6.32	6.19	5.18
10	40	*t*	0.07	0.53	1.46	4.93	12.12	18.28	29.75
		W	5.28	5.04	5.17	5.14	5.10	4.97	4.86
		MW	0.82	1.41	2.58	4.89	8.35	10.76	15.55
		Fprel	5.28	5.04	5.21	5.51	5.55	5.13	4.86
		Lprel	5.28	4.96	4.96	5.65	7.06	6.31	4.86
20	20	*t*	5.55	5.11	4.95	5.10	5.23	5.56	5.54
		W	4.99	4.88	4.84	5.03	5.18	5.21	5.01
		MW	7.83	5.77	5.29	4.72	5.24	5.88	7.48
		Fprel	4.99	4.88	4.87	4.95	5.18	5.21	5.01
		Lprel	4.99	4.86	4.98	5.03	5.26	5.26	5.01

Table 49

Uniform Distribution Using the Expert System: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	8.51	6.78	5.77	5.31	5.94	6.97	8.89
		W	6.38	5.61	4.78	4.38	5.05	5.84	6.65
		MW	6.99	6.45	6.07	5.48	6.04	7.07	7.46
		Fprel	6.39	6.00	5.43	4.93	5.54	6.42	6.67
		Lprel	6.64	6.40	5.78	5.16	5.79	6.81	6.95
5	10	*t*	1.78	2.31	3.03	5.25	9.00	12.44	18.40
		W	5.26	4.87	5.05	5.88	6.41	6.65	6.15
		MW	2.67	2.38	2.65	4.10	6.40	8.05	8.57
		Fprel	5.25	4.78	5.06	5.92	6.97	7.18	6.18
		Lprel	5.01	4.32	4.66	5.77	7.71	9.37	7.03
5	20	*t*	0.19	0.62	1.38	4.93	12.76	19.84	31.09
		W	5.03	5.31	6.08	6.69	7.05	6.80	6.55
		MW	0.88	1.18	1.68	4.91	10.43	13.63	21.14
		Fprel	5.03	5.29	6.18	7.25	9.02	7.69	6.57
		Lprel	5.02	5.20	6.02	7.42	12.08	13.86	7.38
10	10	*t*	6.28	5.70	5.34	5.10	5.33	5.61	6.35
		W	5.24	5.24	5.01	4.88	4.98	5.09	5.17
		MW	8.52	6.30	5.73	5.14	5.81	6.72	8.56
		Fprel	5.24	5.24	5.20	5.26	5.17	5.12	5.17
		Lprel	5.30	5.92	5.74	5.30	5.72	5.96	5.22

Table 49--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	1.19	1.80	2.62	5.11	8.89	12.17	16.70
		W	4.95	4.87	4.96	5.36	5.55	5.59	5.35
		MW	3.24	2.79	3.05	4.94	8.19	10.25	12.91
		Fprel	4.95	4.87	4.99	5.58	5.85	5.60	5.35
		Lprel	4.95	4.87	4.89	5.72	7.62	6.82	5.35
10	40	*t*	0.11	0.50	1.29	4.87	12.54	18.69	29.52
		W	5.09	5.05	5.31	5.34	5.35	5.15	5.21
		MW	0.74	1.01	1.68	4.89	10.03	12.84	16.18
		Fprel	5.09	5.05	5.35	5.82	5.78	5.18	5.21
		Lprel	5.09	5.03	5.25	6.16	8.62	6.34	5.21
20	20	*t*	5.71	5.39	5.15	5.13	5.29	5.34	5.64
		W	5.18	5.11	5.04	5.10	5.12	5.09	5.09
		MW	8.45	6.45	5.64	4.72	5.60	6.32	8.05
		Fprel	5.18	5.11	5.03	5.06	5.12	5.09	5.09
		Lprel	5.18	5.15	5.37	5.11	5.50	5.20	5.09

Table 50

Double Exponential Distribution Using the Expert System: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	5.40	4.90	4.65	4.56	4.74	5.09	5.69
		W	3.12	3.35	3.39	3.51	3.62	3.62	3.59
		MW	6.81	6.45	6.06	5.81	6.07	6.40	6.90
		Fprel	3.20	3.70	3.90	3.94	4.02	3.96	3.64
		Lprel	4.40	4.80	4.92	4.94	5.14	5.21	4.66
5	10	*t*	1.07	1.77	2.54	4.58	7.59	10.24	16.70
		W	3.79	3.85	3.69	3.52	3.48	3.53	3.23
		MW	2.32	2.59	2.87	3.86	5.26	6.26	7.96
		Fprel	3.81	3.80	3.77	3.95	3.90	3.98	3.41
		Lprel	3.78	3.41	3.56	4.40	5.10	5.55	4.57
5	20	*t*	0.09	0.75	1.81	5.25	11.51	17.55	30.46
		W	4.58	4.48	4.34	4.14	3.93	3.79	3.45
		MW	1.09	2.09	3.13	5.20	7.63	9.71	14.68
		Fprel	4.60	4.59	4.75	4.85	5.07	4.70	3.77
		Lprel	4.51	4.14	4.22	5.42	7.03	7.02	4.96
10	10	*t*	5.14	4.54	4.42	4.41	4.57	4.84	5.18
		W	4.02	4.01	4.06	4.10	4.11	4.27	3.94
		MW	6.92	5.49	5.03	4.93	4.97	5.46	6.89
		Fprel	4.01	4.06	4.13	4.06	4.13	4.37	3.96
		Lprel	4.08	4.23	4.35	4.32	4.48	4.54	4.07

Table 50--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	0.83	1.66	2.56	4.93	7.87	10.56	16.01
		W	4.39	4.55	4.61	4.51	4.31	4.23	4.26
		MW	3.08	3.25	3.69	4.76	6.18	7.55	10.51
		Fprel	4.39	4.51	4.43	4.40	4.58	4.42	4.28
		Lprel	4.39	4.22	4.01	4.57	4.94	4.99	4.38
10	40	*t*	0.08	0.56	1.60	5.10	12.18	18.00	29.89
		W	4.99	4.82	4.83	4.96	4.82	4.82	4.58
		MW	1.23	2.37	3.19	5.32	7.93	9.98	14.02
		Fprel	4.99	4.83	4.96	5.06	5.18	4.94	4.58
		Lprel	4.99	4.71	4.38	4.87	5.90	5.80	4.68
20	20	*t*	5.15	4.90	4.93	4.79	4.90	4.94	5.13
		W	4.51	4.59	4.80	4.67	4.79	4.71	4.52
		MW	6.66	5.28	5.00	4.82	4.88	5.34	6.66
		Fprel	4.51	4.51	4.72	4.42	4.64	4.63	4.52
		Lprel	4.51	4.54	4.55	4.34	4.50	4.58	4.52

Table 51

Logistic Distribution Using the Expert System: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	6.65	5.38	4.88	4.70	5.03	5.80	7.22
		W	4.40	3.96	3.91	3.53	3.90	4.38	4.92
		MW	7.12	6.05	5.63	5.63	5.95	6.48	7.52
		Fprel	4.48	4.21	4.27	4.06	4.32	4.66	4.99
		Lprel	5.34	5.26	5.06	4.90	5.20	5.54	5.70
5	10	*t*	1.43	2.05	2.96	4.87	8.26	11.16	17.35
		W	4.69	4.22	4.21	4.33	4.43	4.42	4.15
		MW	2.48	2.67	3.04	3.96	5.53	6.66	8.38
		Fprel	4.67	4.23	4.33	4.61	4.86	4.74	4.27
		Lprel	4.50	3.91	4.09	4.88	6.22	6.36	5.31
5	20	*t*	0.15	0.71	1.60	4.69	11.52	17.50	30.59
		W	4.76	4.62	4.69	4.87	4.75	4.67	4.45
		MW	0.94	1.88	2.93	5.17	8.12	10.44	16.32
		Fprel	4.77	4.67	4.97	5.69	5.83	5.76	4.59
		Lprel	4.71	4.40	4.66	6.06	8.13	8.72	5.81
10	10	*t*	6.01	5.29	5.20	5.07	5.71	5.45	6.14
		W	4.90	4.81	4.86	4.77	4.82	4.79	4.92
		MW	7.50	6.15	5.76	5.54	5.69	6.24	8.38
		Fprel	4.90	4.82	4.84	4.93	4.82	4.82	4.93
		Lprel	4.98	5.14	5.10	5.21	5.13	5.24	5.03

Table 51--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	1.14	1.93	2.85	5.33	8.37	11.02	15.81
		W	5.13	5.04	5.04	5.16	5.08	5.01	4.87
		MW	3.57	3.41	3.78	4.94	6.82	8.09	11.28
		Fprel	5.13	5.03	5.00	5.15	5.28	5.14	4.87
		Lprel	5.11	4.86	4.59	5.38	6.04	5.86	4.97
10	40	*t*	0.09	0.55	1.58	4.89	11.74	18.05	29.78
		W	5.16	5.02	5.16	4.93	4.85	4.71	4.56
		MW	0.82	1.54	2.69	4.89	8.07	10.45	15.05
		Fprel	5.16	5.00	5.25	5.19	5.25	4.94	4.56
		Lprel	5.16	4.84	4.89	5.24	6.52	6.05	4.57
20	20	*t*	5.70	5.58	5.49	5.31	5.16	5.15	5.59
		W	5.16	5.35	5.43	5.25	5.01	4.90	4.87
		MW	7.64	6.13	5.37	5.04	5.51	5.81	7.37
		Fprel	5.16	5.35	5.40	5.11	4.94	4.89	4.87
		Lprel	5.16	5.39	5.45	5.09	4.97	4.94	4.87

procedure is defined to be severely conservative, extremely conservative, conservative, robust, liberal, extremely liberal, or severely liberal as outlined in Chapter IV.

$R = 1$ (Equal Variance Cases)

A summary of the simulated null rejection rates for the four symmetric distributions for the equal variance cases is presented in Table

52. Table 52 shows, as anticipated, that the t test is robust. However, the other procedures are also generally robust. None of the procedures examined show simulated rejection rates ≤ 2.5% or > 10%.

R = 0.67 (Includes the Moderate Direct Pairing Cases)

Table 53 summarizes the simulated rejection rates for the four symmetric distributions for the R = 0.67 cases. Table 53 shows the t test and the MW test can be extremely conservative. The Welch test, the Fprel test procedure and the Lprel test procedure are robust for the R = 0.67 cases.

R = 0.25 and 0.50 (Includes the Severe Direct Pairing Cases)

The simulated null rejection rates for the four symmetric distributions for the R = 0.25 and 0.50 cases are summarized in Table 54. The t test tends to be severely conservative for these cases. Results using the MW test range from liberal to severely conservative. The Welch test, the Fprel test procedure, and the Lprel test procedure are comparable. They are generally robust but can be slightly conservative for some cases.

R = 1.50 (Includes the Moderate Indirect Pairing Cases)

Table 55 summarizes the simulated null rejection rates for the four symmetric distributions for the R = 1.50 cases. The t test is extremely

Table 52

Summary of Symmetric Distributions Using Expert System: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level
Equal Variances: $R=\sigma_1/\sigma_2 = 1.00$

test	$x \leq 1$	$1< x \leq 2.5$	$2.5< x \leq 4$	$4< x \leq 6$	$6< x \leq 10$	$10< x \leq 20$	$x >20$	Total
t	0	0	0	28	0	0	0	28
W	0	0	4	23	1	0	0	28
MW	0	0	2	26	0	0	0	28
Fprel	0	0	2	24	2	0	0	28
Lprel	0	0	0	24	4	0	0	28

Table is based on the four symmetric distributions (normal, uniform, double exponential, and logistic) and seven sample size pairs.

Table 53

Summary of Symmetric Distributions Using Expert System: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level
$R=\sigma_1/\sigma_2 = 0.67$

test	$x \le 1.0$	$1 < x \le 2.5$	$2.5 < x \le 4$	$4 < x \le 6$	$6 < x \le 10$	$10 < x \le 20$	$x > 20$	Total
t	0	8	8	12	0	0	0	28
W	0	0	3	24	1	0	0	28
MW	0	3	13	10	2	0	0	28
Fprel	0	0	2	25	1	0	0	28
Lprel	0	0	1	26	1	0	0	28

Table is based on the four symmetric distributions (normal, uniform, double exponential, and logistic) and seven sample size pairs.

Table 54

Summary of Symmetric Distributions Using Expert System: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level
$R=\sigma_1/\sigma_2$ = 0.25 and 0.50

test	$x \le 1$	$1 < x \le 2.5$	$2.5 < x \le 4$	$4 < x \le 6$	$6 < x \le 10$	$10 < x \le 20$	$x > 20$	Total
t	17	15	0	18	6	0	0	56
W	0	0	5	50	1	0	0	56
MW	6	13	13	4	20	0	0	56
Fprel	0	0	4	51	1	0	0	56
Lprel	0	0	3	51	2	0	0	56

Table is based on the four symmetric distributions (normal, uniform, double exponential, and logistic) and seven sample size pairs.

Table 55

Summary of Symmetric Distributions Using Expert System: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level
$R=\sigma_1/\sigma_2$ = 1.50

test	$x \leq 1$	$1 < x \leq 2.5$	$2.5 < x \leq 4$	$4 < x \leq 6$	$6 < x \leq 10$	$10 < x \leq 20$	$x > 20$	Total
t	0	0	0	12	8	8	0	28
W	0	0	4	22	2	0	0	28
MW	0	0	0	13	13	2	0	28
Fprel	0	0	1	24	3	0	0	28
Lprel	0	0	0	15	12	1	0	28

Table is based on the four symmetric distributions (normal, uniform, double exponential, and logistic) and seven sample size pairs.

Table 56

Summary of Symmetric Distributions Using Expert System: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level
$R=\sigma_1/\sigma_2$ = 2.00 and 4.00

test	$x \leq 1$	$1< x \leq 2.5$	$2.5< x \leq 4$	$4< x \leq 6$	$6< x \leq 10$	$10< x \leq 20$	$x >20$	Total
t	0	0	0	17	7	24	8	56
W	0	0	7	44	5	0	0	56
MW	0	0	0	5	32	18	1	56
Fprel	0	0	6	44	6	0	0	56
Lprel	0	0	0	39	16	1	0	56

Table is based on the four symmetric distributions (normal, uniform, double exponential, and logistic) and seven sample size pairs.

liberal. The MW test and the Lprel test procedure tend to be liberal. The Welch test and the Fprel test procedure are each reasonably robust.

R = 2.0 and 4.0 (Includes the Severe Indirect Pairing Cases)

The simulated null rejection rates for the four symmetric distributions for the R = 2.0 and 4.0 cases are summarized in Table 56. The t test can be severely liberal. The MW test is extremely liberal but not as severe as the t test. The Lprel test procedure is liberal. The Fprel test procedure and the Welch test are each robust with no evidence of extreme liberalism or extreme conservatism.

Based on the above simulation results, the Fprel test procedure and the Welch test are recommended as robust tests for testing the H_o: $\mu_1 = \mu_2$ for the symmetric cases examined.

7.3 Results From Asymmetric Distributions

Tables 57-61 present the simulated null rejection rates, where the proportion of rejections is expressed as a percent for each of the five asymmetric distributions in a format similar to the one used in Tables 4-7 in Chapter III. The simulated null rejection rates for each of the five asymmetric distributions are included for completeness, but discussion will be limited to the overall performance of the test procedures as described below.

Table 57

Lognormal (0, 0.40) Using the Expert System: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	8.88	6.30	5.23	4.45	5.22	6.74	16.36
		W	6.88	5.19	4.08	3.21	3.91	5.44	14.40
		MW	8.51	7.06	5.99	5.21	6.13	7.86	17.23
		Fprel	6.95	5.61	4.40	3.55	4.32	5.90	15.10
		Lprel	7.52	6.22	5.18	4.32	5.12	6.86	16.38
5	10	*t*	2.76	3.13	3.54	4.68	7.75	10.72	22.70
		W	6.07	4.73	4.42	4.95	6.67	9.17	19.80
		MW	3.96	3.21	3.38	4.19	6.31	8.84	19.90
		Fprel	6.07	4.75	4.54	5.11	6.88	9.45	20.43
		Lprel	5.97	4.70	4.69	5.40	7.11	9.39	20.96
5	20	*t*	0.43	1.13	2.03	5.08	10.34	16.26	32.37
		W	5.24	4.58	4.95	6.40	8.82	11.62	21.16
		MW	2.26	2.48	3.09	5.19	9.26	14.27	32.19
		Fprel	5.23	4.94	5.41	7.08	9.71	13.23	23.88
		Lprel	5.44	5.15	5.83	7.36	10.05	14.16	29.56
10	10	*t*	7.24	6.24	5.67	4.68	5.45	7.42	16.01
		W	6.21	5.78	5.29	4.23	4.94	6.96	15.47
		MW	10.98	7.86	6.63	5.35	6.93	11.18	31.36
		Fprel	6.21	5.99	5.63	4.54	5.40	7.89	16.68
		Lprel	6.46	6.98	6.25	5.04	6.53	10.32	28.22

Table 57--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	2.01	2.74	3.30	4.93	8.11	11.81	22.83
		W	5.88	5.01	4.77	5.16	6.95	9.04	17.38
		MW	7.58	5.87	5.04	5.06	8.81	14.96	40.23
		Fprel	5.88	5.18	5.20	5.51	7.61	10.17	18.19
		Lprel	5.92	6.05	5.86	5.92	8.45	13.33	28.90
10	40	*t*	0.27	0.89	1.82	4.71	11.34	17.31	32.96
		W	5.51	4.86	5.05	5.88	7.80	9.98	18.47
		MW	4.29	3.69	3.66	4.77	10.91	19.19	46.90
		Fprel	5.51	5.00	5.55	6.60	9.11	11.40	19.10
		Lprel	5.52	5.53	6.16	6.77	10.52	16.26	27.55
20	20	*t*	6.88	5.81	5.30	4.98	5.30	6.64	13.39
		W	6.30	5.52	5.11	4.79	5.13	6.50	13.19
		MW	14.98	9.81	7.06	4.97	7.85	16.23	52.14
		Fprel	6.30	5.58	5.43	5.24	5.91	7.63	13.50
		Lprel	6.31	5.59	6.60	5.70	8.06	13.48	27.57

Table 58

Lognormal (0, 1.0) Using the Expert System: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	15.39	7.15	4.44	2.98	3.57	5.06	11.01
		W	13.85	5.39	2.81	1.65	1.84	2.85	6.80
		MW	15.14	9.75	7.02	5.21	6.98	9.83	21.20
		Fprel	14.25	5.81	3.12	1.79	2.07	3.20	7.28
		Lprel	14.82	7.59	4.64	2.87	3.74	5.32	11.55
5	10	*t*	8.83	6.36	5.08	3.77	3.87	4.76	9.50
		W	9.95	3.77	2.77	3.41	6.62	10.21	20.50
		MW	12.28	6.61	4.70	4.19	6.63	9.97	24.20
		Fprel	10.34	4.42	3.13	3.58	6.78	10.43	21.26
		Lprel	12.57	6.09	4.08	4.27	7.45	11.64	24.73
5	20	*t*	2.90	3.85	4.27	4.88	5.47	6.35	10.47
		W	5.73	3.36	4.38	8.38	14.74	20.17	33.27
		MW	12.56	7.35	5.20	5.19	9.57	15.94	35.14
		Fprel	6.99	4.94	5.44	8.78	15.22	20.97	35.40
		Lprel	13.08	8.25	6.89	8.40	13.90	20.07	37.07
10	10	*t*	13.41	7.44	5.36	3.52	4.64	6.94	14.82
		W	12.95	7.01	4.49	2.65	3.75	5.82	13.18
		MW	26.10	13.15	8.22	5.35	9.01	15.43	39.08
		Fprel	13.52	8.63	5.37	3.01	4.63	7.31	17.52
		Lprel	21.33	12.41	7.30	4.11	6.85	11.91	31.91

Table 58--Continued

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	6.38	5.19	4.58	3.71	4.83	6.78	15.36
		W	9.64	4.57	3.24	4.34	8.48	12.38	24.40
		MW	29.34	14.07	8.07	5.06	10.29	19.52	49.16
		Fprel	10.05	6.93	4.63	4.82	9.34	14.37	29.89
		Lprel	21.92	13.90	7.75	6.06	12.12	20.66	47.71
10	40	*t*	2.03	3.01	3.54	4.43	6.47	8.57	17.32
		W	6.88	3.80	4.57	8.40	14.14	19.06	31.67
		MW	31.81	13.98	7.80	4.77	13.38	24.61	57.11
		Fprel	7.12	6.24	6.23	9.05	15.58	21.89	38.17
		Lprel	20.70	14.73	9.68	8.89	16.91	26.39	53.77
20	20	*t*	12.17	8.05	5.61	3.85	4.87	7.42	16.57
		W	11.83	7.88	5.37	3.47	4.46	7.11	16.04
		MW	44.83	22.35	11.63	4.97	11.97	24.83	65.47
		Fprel	11.87	10.88	7.20	4.10	6.04	10.48	23.76
		Lprel	16.63	21.45	11.70	5.25	10.74	22.57	59.67

Table 59

Lognormal (0, 1.75) Using the Expert System: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_2/σ_1						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	5	*t*	12.19	3.62	2.29	1.78	1.91	2.34	3.92
		W	8.03	1.79	1.04	0.72	0.70	0.88	1.92
		MW	21.43	9.71	6.70	5.21	6.64	8.49	15.27
		Fprel	8.72	2.01	1.18	0.79	0.82	1.03	2.05
		Lprel	12.84	4.21	2.86	2.04	2.67	3.31	5.87
5	10	*t*	14.82	6.89	4.61	2.97	2.07	1.94	1.98
		W	4.93	1.36	1.32	1.79	3.02	4.29	7.98
		MW	21.55	7.85	5.18	4.19	5.36	7.87	16.52
		Fprel	7.02	1.94	1.57	1.90	3.09	4.36	8.09
		Lprel	13.43	3.60	2.60	3.12	5.11	7.27	14.52
5	20	*t*	10.90	7.70	6.60	5.53	4.56	3.98	3.03
		W	2.68	2.49	3.95	7.20	11.19	14.49	22.60
		MW	27.41	9.31	6.14	5.19	7.65	11.40	24.96
		Fprel	8.00	3.98	4.67	7.48	11.40	14.71	23.04
		Lprel	21.75	6.87	5.55	7.07	11.31	15.63	28.85
10	10	*t*	15.40	4.83	3.03	1.91	2.28	3.13	6.10
		W	14.11	3.73	2.06	1.34	1.64	2.11	4.45
		MW	39.47	13.47	8.17	5.35	7.63	11.94	27.89
		Fprel	18.93	4.99	2.51	1.61	2.10	2.93	6.33
		Lprel	34.26	8.74	4.54	2.98	4.21	6.25	15.07

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	13.97	6.65	4.58	2.85	2.25	2.28	3.18
		W	8.39	2.09	1.73	3.03	5.33	7.45	13.85
		MW	48.65	15.38	8.24	5.06	8.37	14.26	35.63
		Fprel	15.28	3.91	2.53	3.39	5.75	8.22	15.83
		Lprel	42.40	9.88	5.26	5.10	9.01	14.15	31.33
10	40	*t*	9.09	7.15	6.39	5.29	4.54	4.04	3.59
		W	4.62	3.00	4.56	8.19	13.03	17.11	27.10
		MW	58.39	16.83	8.25	4.77	10.14	17.83	43.18
		Fprel	11.61	5.51	5.51	8.49	13.35	17.88	28.87
		Lprel	53.84	13.46	7.87	7.86	14.40	21.81	44.27
20	20	*t*	17.42	6.07	3.60	2.32	2.76	3.80	8.19
		W	17.11	5.58	3.11	1.95	2.24	3.29	7.30
		MW	67.18	23.26	10.55	4.97	9.47	17.92	48.88
		Fprel	24.46	7.80	4.03	2.37	2.92	4.59	10.77
		Lprel	62.89	17.91	7.69	3.73	6.23	11.90	32.59

Table 60

Gamma (3,1) Distribution Using the Expert System: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	*t*	t	9.36	6.95	5.48	5.03	5.56	8.32	25.74
		W	6.92	5.63	4.41	3.88	4.26	6.49	22.97
		MW	8.59	7.33	6.08	5.73	6.48	9.50	27.77
		Fprel	7.08	5.98	4.80	4.19	4.62	6.98	23.88
		Lprel	7.71	6.68	5.44	4.86	5.57	8.11	25.31
5	10	*t*	2.84	3.19	3.43	4.71	7.79	12.33	31.62
		W	6.25	4.73	4.11	5.14	7.58	11.67	30.08
		MW	3.82	3.58	3.07	3.99	7.09	11.12	31.68
		Fprel	6.25	4.93	4.32	5.33	7.88	11.90	31.52
		Lprel	6.13	4.88	4.55	5.41	8.04	11.78	31.99
5	20	*t*	0.42	1.06	1.85	4.81	11.15	18.75	40.66
		W	5.49	4.66	5.06	6.58	9.13	13.44	31.04
		MW	2.15	2.43	2.83	5.02	10.03	17.51	49.26
		Fprel	5.47	4.99	5.54	7.29	10.29	15.69	34.27
		Lprel	5.56	5.23	5.82	7.52	10.64	17.11	44.08
10	10	*t*	7.85	6.46	5.28	4.87	5.22	7.87	20.41
		W	6.77	5.94	4.97	4.48	4.86	7.48	20.13
		MW	11.40	7.87	6.27	5.39	7.16	13.77	50.44
		Fprel	6.77	6.03	5.31	4.86	5.57	8.91	22.12
		Lprel	7.06	7.03	6.08	5.23	6.79	13.20	49.42

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	1.75	2.39	3.02	4.67	8.01	12.20	27.59
		W	5.53	4.88	4.60	5.04	6.46	9.42	22.36
		MW	6.85	5.29	4.45	4.93	9.35	18.70	60.68
		Fprel	5.53	4.92	4.99	5.52	7.48	11.22	23.49
		Lprel	5.58	5.71	5.61	5.72	9.14	17.52	49.47
10	40	*t*	0.22	0.83	1.56	4.72	11.59	18.96	37.02
		W	5.57	5.13	5.06	6.11	8.28	10.49	23.35
		MW	4.22	3.81	3.66	4.70	12.23	22.83	66.39
		Fprel	5.57	5.27	5.52	6.67	9.91	12.46	23.83
		Lprel	5.57	5.66	6.16	6.79	12.01	19.73	43.08
20	20	*t*	6.68	5.55	5.25	5.42	5.34	7.11	15.20
		W	6.20	5.34	5.04	5.26	5.18	6.95	14.93
		MW	14.06	9.33	6.73	5.36	9.31	21.23	78.91
		Fprel	6.20	5.34	5.24	5.84	6.43	8.51	15.11
		Lprel	6.21	6.46	6.62	6.28	9.46	18.22	52.08

Table 61

Gamma (2,1) Distribution Using the Expert System: Simulated Null Rejection Rate (%) With Nominal 5% Level

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
5	*t*	t	9.90	6.90	5.49	4.56	5.57	8.27	27.98
		W	7.81	5.71	4.10	3.23	3.81	6.05	22.71
		MW	9.10	7.50	6.32	5.34	7.24	10.89	33.82
		Fprel	7.90	6.03	4.56	3.54	4.20	6.61	23.81
		Lprel	8.52	6.80	5.31	4.25	5.40	8.36	26.67
5	10	*t*	3.18	3.40	3.69	4.80	7.87	11.54	33.69
		W	7.00	4.52	3.83	5.03	8.82	13.35	36.24
		MW	4.69	3.49	3.33	3.84	7.62	12.68	38.62
		Fprel	7.01	4.70	4.13	5.06	9.06	13.81	38.35
		Lprel	6.96	4.73	4.42	5.35	8.88	13.63	38.72
5	20	*t*	0.63	1.37	2.30	4.84	10.55	16.94	40.66
		W	5.49	4.87	5.32	7.42	11.55	16.60	37.24
		MW	2.89	3.11	3.22	4.96	11.27	19.88	57.53
		Fprel	5.51	5.30	6.02	7.98	12.70	18.91	41.88
		Lprel	5.79	5.76	6.53	8.09	12.62	20.03	51.76
10	10	*t*	8.18	6.84	5.18	5.02	5.50	7.91	24.48
		W	7.17	6.32	4.77	4.52	5.06	7.48	24.13
		MW	13.23	9.13	6.68	5.46	8.13	16.79	62.57
		Fprel	7.16	6.63	5.30	4.89	6.05	10.14	28.44
		Lprel	7.76	8.12	6.29	5.45	7.73	16.09	61.32

			Ratio = σ_1/σ_2						
n_1	n_2	Test	.25	.50	.67	1.00	1.50	2.00	4.00
10	20	*t*	2.32	2.97	3.35	4.67	8.50	12.74	28.93
		W	6.15	5.61	4.83	5.65	8.19	11.09	26.13
		MW	9.35	6.76	5.28	5.15	11.17	23.30	70.08
		Fprel	6.15	5.82	5.39	6.28	9.68	14.21	28.76
		Lprel	6.29	7.37	6.49	6.71	11.71	22.63	61.16
10	40	*t*	0.32	1.03	1.90	5.26	11.32	18.62	38.98
		W	5.51	4.80	5.11	6.34	8.89	12.23	28.84
		MW	6.38	4.88	4.38	5.41	13.20	27.86	78.00
		Fprel	5.51	5.08	5.87	7.20	10.86	15.82	30.83
		Lprel	5.53	6.25	6.88	7.38	13.26	25.40	57.35
20	20	*t*	7.37	6.36	5.57	4.79	5.71	7.67	18.86
		W	6.89	6.09	5.46	4.68	5.55	7.41	18.77
		MW	18.16	11.81	8.15	4.72	11.47	28.93	88.94
		Fprel	6.89	6.21	6.36	5.33	7.47	10.55	19.53
		Lprel	6.89	8.11	8.11	5.87	11.91	26.35	71.66

To evaluate the overall performance of the procedures for varying degrees of variance heterogeneity, the results of the simulation for the five asymmetric distributions are combined in Tables 64-68 using the same format as used in Tables 18-22 in Chapter IV. The outcome of a test procedure is defined to be severely conservative, extremely conservative, conservative, robust, liberal, extremely liberal, or severely liberal as outlined in Chapter IV.

R = 1 (Equal Variance Cases)

A summary of the simulated null rejection rates for the five asymmetric distributions for the equal variance cases is presented in Table 62. The MW test is robust for the R = 1 cases. The t test is robust for approximately 70% (25 of 35) of the R = 1 cases but it can be extremely conservative. The Welch test, the Fprel test procedure, and the Lprel test procedure tend to be liberal or somewhat conservative. The Welch test and Fprel test procedure can be extremely conservative in some cases. None of the procedures are extremely liberal.

R = 0.67 (Includes Moderate Direct Pairing Cases)

Table 63 summarizes the simulated null rejection rates for the asymmetric distributions for the R = 0.67 cases. The Welch test and the Fprel test procedure are robust in approximately 70% (25 of 35) of the R = 0.67 cases. These procedures can be conservative and even extremely conservative for some R = 0.67 cases. The t test is conservative or extremely conservative for almost 50% (16 of 35) of the R = 0.67 cases. The MW test and the Lprel test procedure are liberal or extremely liberal in at least 50% of these cases. The Lprel test procedure is slower to counteract the liberalism of the MW test than is the Fprel test procedure.

Table 62

Summary of Asymmetric Distributions Using Expert System: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level
Equal Variances: $R=\sigma_1/\sigma_2 = 1.00$

test	$x \le 1$	$1< x \le 2.5$	$2.5< x \le 4$	$4< x \le 6$	$6< x \le 10$	$10< x \le 20$	$x >20$	Total
t	0	3	7	25	0	0	0	35
W	1	4	7	14	9	0	0	35
MW	0	0	2	33	0	0	0	35
Fprel	1	4	5	14	11	0	0	35
Lprel	0	1	4	17	13	0	0	35

Table is based on the five asymmetric distributions [lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3,1), & G(2,1)] and seven sample size pairs.

Table 63

Summary of Asymmetric Distributions Using Expert System: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level
$R = \sigma_1/\sigma_2 = 0.67$

test	$x \leq 1$	$1 < x \leq 2.5$	$2.5 < x \leq 4$	$4 < x \leq 6$	$6 < x \leq 10$	$10 < x \leq 20$	$x > 20$	Total
t	0	7	9	17	2	0	0	35
W	0	4	6	25	0	0	0	35
MW	0	0	8	8	17	2	0	35
Fprel	0	2	4	25	4	0	0	35
Lprel	0	0	2	15	17	1	0	35

Table is based on the five asymmetric distributions [lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3,1), & G(2,1)] and seven sample size pairs.

R = 0.25 and 0.50 (Includes Severe Direct Pairing Cases)

The simulated null rejection rates for the asymmetric distributions for the R = 0.25 and 0.50 cases are summarized in Table 64. For the R = 0.25 and 0.50 cases, none of the test procedures are robust. The Welch test and the Fprel test procedures tend to be liberal or extremely liberal, but these tests can be extremely conservative as well. The t test can be extremely liberal or severely conservative. The MW test tends to be severely liberal for the R = 0.25 and 0.50 cases.

R = 1.50 (Includes Moderate Indirect Pairing Cases)

Table 65 summarizes the simulated null rejection rates for the asymmetric distributions for the R = 1.50 cases. All procedures tend to be liberal or extremely liberal for the R = 1.50 cases. However, the t test is the most robust for these cases. The t test, the Welch test, and the Fprel test procedure can be occasionally extremely conservative.

R = 2 and 4 (Includes Severe Indirect Pairing Cases)

The simulated null rejection rates for the asymmetric distributions for the R = 2 and 4 cases are summarized in Table 66. All procedures are extremely or severely liberal for >55% of the R = 2 and 4 cases. These procedures can yield extremely conservative results as well.

Table 64

Summary of Asymmetric Distributions Using Expert System: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level
R=σ_1/σ_2 = 0.25 and 0.50

test	$x \leq 1$	$1 < x \leq 2.5$	$2.5 < x \leq 4$	$4 < x \leq 6$	$6 < x \leq 10$	$10 < x \leq 20$	$x > 20$	Total
t	8	9	12	4	28	9	0	70
W	0	4	6	33	22	5	0	70
MW	0	4	9	6	22	16	13	70
Fprel	0	2	2	30	26	9	1	70
Lprel	0	0	1	19	30	11	9	70

Table is based on the five asymmetric distributions [lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3,1), & G(2,1)] and seven sample size pairs.

Table 65

Summary of Asymmetric Distributions Using Expert System: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level
$R = \sigma_1/\sigma_2 = 1.50$

test	$x \le 1$	$1 < x \le 2.5$	$2.5 < x \le 4$	$4 < x \le 6$	$6 < x \le 10$	$10 < x \le 20$	$x > 20$	Total
t	0	4	3	15	7	6	0	35
W	1	3	4	9	13	5	0	35
MW	0	0	0	1	23	11	0	35
Fprel	1	2	2	8	15	7	0	35
Lprel	0	0	2	5	14	14	0	35

Table is based on the five asymmetric distributions [lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3,1), & G(2,1)] and seven sample size pairs.

Table 66

Summary of Asymmetric Distributions Using Expert System: Frequency of Simulated Null Rejection Rate (%) With Nominal 5% Level
$R=\sigma_1/\sigma_2$ = 2.00 and 4.00

test	$x \leq 1$	$1< x \leq 2.5$	$2.5< x \leq 4$	$4< x \leq 6$	$6< x \leq 10$	$10< x \leq 20$	$x >20$	Total
t	0	4	7	3	17	23	16	70
W	1	2	2	4	17	24	20	70
MW	0	0	0	0	7	25	38	70
Fprel	0	2	2	3	12	28	23	70
Lprel	0	0	1	2	6	22	39	70

Table is based on the five asymmetric distributions [lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3,1), & G(2,1)] and seven sample size pairs.

In summary for the $R = 1$ cases, all procedures are generally robust. For the $R = 0.67$ cases, the Welch test and the Fprel test procedure are reasonably robust. For the $R = 0.25$, 0.50, 1.50, 2.0 and 4.0 cases, all procedures tend to be liberal. The degree of liberal bias increases as the degree of variance heterogeneity increases.

7.4 Frequency (%) at Which Each Means Test is Used

In addition to the simulated null rejection rates, the expert system can report the frequency (%) at which each of the test procedures is used for a given sample size and R value. Results for the imbalanced case $n_1 = 10$ and $n_2 = 20$, and the balanced case $n_1 = n_2 = 20$ will be summarized for the normal distribution cases, the four symmetric distribution cases combined (including the normal distribution cases), and the five asymmetric distribution cases combined.

Tables 67-68, 69-70, and 71-72 summarize the frequency (%) at which each of the test procedures is used for the normal distribution cases, the four symmetric distribution cases combined, and the five asymmetric distribution cases combined, respectively. The format for Tables 67-72 is as follows. For each R value, the frequency at which the t test, the Welch-S test, the MW test, and the Welch-AS test was selected by the expert system is reported. In these tables, the t test, Welch-S, MW, and Welch-AS denote the following:

t test: The *t* test was used because the expert system concluded $\sigma_1 = \sigma_2$ and symmetry was accepted.

Welch-S: The Welch test was used because the expert system concluded $\sigma_1 \neq \sigma_2$ and symmetry was accepted.

MW: The MW test was used because the expert system concluded $\sigma_1 = \sigma_2$ and symmetry was rejected.

Welch-AS: The Welch test was used because the expert system concluded $\sigma_1 \neq \sigma_2$ and symmetry was rejected.

7.4.1 Normal Case

Tables 67-68 contain the frequency (%) at which each of the test procedures is used in the normal distribution cases for the balanced and imbalanced cases, respectively.

$R = 1$ (Includes Imbalanced and Balanced Cases)

For the $R = 1$ case with equal sample sizes, the *t* test is known to be robust. Results in Table 67 show the Lprel test procedure correctly selected the *t* test for approximately 50% of the simulations compared to approximately 33% when using the Fprel test procedure. The Welch-S test incorrectly was used twice as often by the Fprel test procedure than by the Lprel test procedure (31% and 15%, respectively).

For the $R = 1$ case with imbalanced sample sizes, the Fprel test

Table 67

Frequency (%) at Which Each Means Test is Used in the Normal Distribution $n_1 = 20$ and $n_2 = 20$ Case

		Fprel		
R	*t* test	Welch-S	MW	Welch-AS
0.25	0.00	63.05	0.00	36.95
0.50	0.77	62.28	0.47	36.48
0.67	8.16	54.89	5.23	31.72
1.00	31.69	31.36	18.43	18.52
1.50	8.62	54.43	4.76	32.19
2.00	0.80	62.25	0.44	36.51
4.00	0.00	63.05	0.00	36.95
		Lprel		
R	*t* test	Welch-S	MW	Welch-AS
0.25	0.00	63.05	0.01	36.94
0.50	3.91	59.14	2.65	34.30
0.67	21.14	41.91	13.20	23.75
1.00	48.26	14.79	29.21	7.74
1.50	21.02	42.03	13.53	23.42
2.00	3.86	59.19	2.45	34.50
4.00	0.00	63.05	0.00	36.95

Table 68

Frequency (%) at Which Each Means Test is Used
in the Normal Distribution
$n_1 = 10$ and $n_2 = 20$ Case

R	*t* test	Fprel Welch-S	MW	Welch-AS
0.25	0.00	65.24	0.00	34.76
0.50	2.70	62.54	1.50	33.26
0.67	15.04	50.20	8.28	26.48
1.00	32.86	32.38	17.47	17.29
1.50	13.78	51.46	6.96	27.80
2.00	3.56	61.68	1.64	33.12
4.00	0.02	65.22	0.00	34.76
R	*t* test	Lprel Welch-S	MW	Welch-AS
0.25	0.04	65.20	0.01	34.75
0.50	9.32	55.92	5.64	29.12
0.67	28.79	36.45	16.07	18.69
1.00	49.32	15.92	27.52	7.24
1.50	30.48	34.76	17.81	16.95
2.00	12.10	53.14	7.03	27.73
4.00	0.16	65.08	0.12	34.64

procedure correctly used the *t* test and incorrectly used the Welch-S test each for approximately 33% of the simulations (Table 68). The Lprel test procedure correctly selected the *t* test for approximately 50% of the simulations and incorrectly used the MW test for approximately 30% of the simulations.

R = 0.67 and 1.50 (Includes Imbalanced and Balanced Cases)

For the R = 0.67 and 1.50 cases with equal sample sizes, Table 67 shows the Fprel test procedure correctly selected the Welch-S test for approximately 55% of the simulations and the Welch-AS for approximately 32% of the simulations. The Lprel test procedure correctly selected the Welch-S test for approximately 42% of the simulations while the *t* test and the Welch-AS test were each selected for more than 20% of the simulations. Both test procedures incorrectly concluded asymmetry for approximately 37% of the normal cases with R = 0.67 and 1.50.

For the R = 0.67 and 1.50 cases with unequal sample sizes, Table 68 shows the Fprel test procedure correctly used the Welch-S test for approximately 50% of the simulations and incorrectly selected the Welch-AS test for approximately 25% of the simulations. The Lprel test procedure correctly selected the Welch-S test for slightly more than 33% of the simulations and incorrectly selected the *t* test for approximately 33% of the simulations. Similar to the $n_1 = n_2$ case, both the Fprel and the Lprel test procedures incorrectly concluded asymmetry for approximately 35% of the normal cases with R = 0.67 and 1.50.

R = 0.25, 0.50, 2.0 and 4.0 (Includes Imbalanced and Balanced Cases)

For the R = 0.25, 0.50, 2.0, and 4.0 cases, the Welch test is known to be robust. Tables 67-68 shows the Fprel and Lprel test procedures correctly used the Welch-S test for approximately 60% of the simulations regardless of sample size configuration. However, the Fprel and Lprel test procedures incorrectly concluded asymmetry for about 35-37% of the normal cases with R = 0.25, 0.50, 2.0, and 4.0.

7.4.2 Symmetric Cases

Tables 69-70 contain the frequency (%) at which each of the test procedures is used in the four symmetric distributions combined (including the normal distribution cases) for the balanced and imbalanced cases, respectively.

R = 1 (Includes the Imbalanced and Balanced Cases)

For the R = 1.00 case with equal sample sizes, the t test is known to be robust for the symmetric distributions. Results in Table 69 show that the Fprel test procedure correctly selected the t test for approximately 30% of the simulations compared to almost 50% when using the Lprel test procedure. The Welch-S test was incorrectly selected for approximately 30% of the simulations when using the Fprel test procedure compared to only 14% when using the Lprel test procedure.

For the R = 1.00 case with unequal sample sizes, Table 70 shows that the Lprel test procedure selected the t test for nearly 50% of the simulations compared to approximately 31% when using the Fprel test procedure. The Fprel test procedure incorrectly selected the Welch-S test for approximately 32% of the simulations, whereas the Welch-S test was incorrectly selected for 15% of the simulations when using the Lprel test procedure.

R = 0.67 and 1.50 (Includes the Imbalanced and Balanced Cases)

For the R = 0.67 and 1.50 cases with equal sample sizes, Table 69 shows the Fprel test procedure correctly selected the Welch-S test for more than 50% of the simulations compared to 40% when using the Lprel test procedure. The Lprel test procedure incorrectly selected the t test more frequently than the Fprel test procedure (21% versus 8%). Both test procedures incorrectly concluded asymmetry in about 40% of the R = 0.67 and 1.50 cases.

For the R = 0.67 and 1.50 cases with unequal sample sizes, Table 70 shows the Fprel test procedure correctly selected the Welch-S test for about 50% of the simulations compared to about 33% when using the Lprel test

Table 69

Frequency (%) at Which Each Means Test is Used in the Symmetric Distributions $n_1 = 20$ and $n_2 = 20$ Case

		Fprel		
R	*t* test	Welch-S	MW	Welch-AS
0.25	0.00	60.10	0.00	39.90
0.50	1.37	58.73	0.99	38.91
0.67	8.10	52.00	5.60	34.30
1.00	29.48	30.62	19.84	20.06
1.50	8.33	51.77	5.11	34.79
2.00	1.43	58.67	0.93	38.97
4.00	0.01	60.09	0.00	39.90
		Lprel		
R	*t* test	Welch-S	MW	Welch-AS
0.25	0.03	60.07	0.02	39.88
0.50	5.05	55.05	3.77	36.13
0.67	20.78	39.32	15.01	24.89
1.00	46.96	14.39	31.38	8.52
1.50	20.67	39.42	14.70	25.21
2.00	5.07	55.03	3.63	36.27
4.00	0.03	60.07	0.02	39.88

Table 70

Frequency (%) at Which Each Means Test is Used in the Symmetric Distributions $n_1 = 10$ and $n_2 = 20$ Case

		Fprel		
R	*t* test	Welch-S	MW	Welch-AS
0.25	0.02	62.40	0.02	37.56
0.50	3.02	59.40	2.05	35.53
0.67	13.35	49.08	8.42	29.15
1.00	30.70	31.72	18.43	19.15
1.50	13.00	49.43	7.49	30.08
2.00	4.23	58.20	2.37	35.20
4.00	0.09	62.34	0.06	37.51
		Lprel		
R	*t* test	Welch-S	MW	Welch-AS
0.25	0.20	62.23	0.16	37.42
0.50	10.26	52.17	7.11	30.46
0.67	27.63	34.79	18.04	19.54
1.00	47.27	15.15	29.63	7.94
1.50	29.65	32.78	19.50	18.07
2.00	12.63	49.79	8.44	29.14
4.00	0.43	61.99	0.32	37.26

procedure. The Lprel test procedure incorrectly selected the *t* test for about 29% of the simulations compared to approximately 13% when using the Fprel test procedure.

<u>*R* = 0.25, 0.50, 2, and 4 (Includes Imbalanced and Balanced Cases)</u>

For the $R = 0.25$, 0.50, 2.0, and 4.0 cases, the Welch test is known to be robust. Tables 69-70 shows the Fprel and the Lprel test procedures correctly used the Welch-S test for approximately 50-60% of the simulations regardless of the sample size configuration. The Welch-AS test was incorrectly used for approximately 30-40% of the simulations when using either the Fprel or the Lprel test procedure.

In summary, for the normal case and the combined symmetric cases, the Lprel test procedure correctly selected the *t* test more often than the Fprel test procedure for the $R = 1$ cases regardless of the sample size configuration. For the $R = 0.67$ and 1.50 cases, regardless of sample size configuration, the Fprel test procedure correctly selected the Welch-S test more frequently than did the Lprel test procedure. For the $R = 0.25$, 0.50, 2.0, and 4.0 cases, regardless of sample size configuration, the Fprel and the Lprel test procedures correctly used the Welch-S test for approximately 60% of the simulations. It is noted for the $R \neq 1$ cases, both the Fprel and the Lprel test procedures incorrectly concluded asymmetry for approximately 38% of the simulations.

7.4.3 Asymmetric Cases

Tables 71-72 contain the frequency (%) at which each of the test procedures is used in the five asymmetric distributions combined for the balanced and imbalanced cases, respectively.

R = 1 (Includes the Imbalanced and Balanced Cases)

For the $R = 1$ case with equal sample sizes, the MW test is known to be robust for the asymmetric distributions. Results in Table 71 show the Lprel test procedure correctly selected the MW test for approximately 69% of the simulations compared to approximately 25% when using the Fprel test procedure. The Fprel test procedure incorrectly selected the Welch-AS test (67%) in these homogeneous variance cases.

For the $R = 1$ case with unequal sample sizes, Table 72 shows the that Lprel test procedure correctly selected the MW test for approximately 65% of the simulations compared to approximately 24% when using the Fprel test procedure. As also seen for the balanced sample size case, the Fprel test procedure incorrectly selected the Welch-AS test in more than 60% of the cases.

R = 0.67 and 1.50 (Includes All Cases)

For the equal sample size cases, Table 71 shows the Lprel test

Table 71

Frequency (%) at Which Each Means Test is Used in the Asymmetric Distributions $n_1 = 20$ and $n_2 = 20$ Case

		Fprel		
R	*t* test	Welch-S	MW	Welch-AS
0.25	0.05	16.37	2.88	80.70
0.50	0.78	12.91	6.96	79.35
0.67	2.65	9.09	14.73	73.53
1.00	3.03	4.65	24.95	67.36
1.50	0.90	2.39	16.67	80.04
2.00	0.27	1.14	10.78	87.81
4.00	0.03	0.12	7.04	92.81
		Lprel		
R	*t* test	Welch-S	MW	Welch-AS
0.25	0.10	16.32	16.53	67.05
0.50	2.17	11.52	32.50	53.81
0.67	5.32	6.42	50.47	37.79
1.00	5.29	2.39	68.81	23.51
1.50	1.70	1.60	60.61	36.09
2.00	0.56	0.86	52.83	45.75
4.00	0.06	0.09	55.10	44.75

Table 72

Frequency (%) at Which Each Means Test is Used in the Asymmetric Distributions $n_1 = 10$ and $n_2 = 20$ Case

R	*t* test	Fprel Welch-S	MW	Welch-AS
0.25	0.08	17.82	3.48	78.62
0.50	1.83	14.99	9.11	74.14
0.67	4.85	11.24	17.18	66.73
1.00	5.74	7.98	24.13	62.16
1.50	2.92	7.37	18.58	71.13
2.00	1.45	5.67	14.29	78.59
4.00	0.34	1.98	10.13	87.55
R	*t* test	Lprel Welch-S	MW	Welch-AS
0.25	0.30	17.60	22.55	59.55
0.50	4.52	12.23	41.01	42.24
0.67	8.50	7.58	54.95	29.97
1.00	9.60	4.12	64.99	21.29
1.50	5.64	4.65	59.20	30.51
2.00	3.12	4.00	53.38	39.50
4.00	0.86	1.46	49.48	48.20

procedure incorrectly selected the MW test in 50% of the R = 0.67 cases and 61% of the R = 1.50 cases. The Welch-AS test was correctly selected for approximately 37% of the R = 0.67 and 1.50 cases when using the Lprel test procedure. The Fprel test procedure incorrectly selected the MW test for 15-17% of the R = 0.67 and 1.50 cases. The Welch-AS was correctly selected by the Fprel test procedure for the majority of the R = 0.67 and 1.50 cases (74% and 80%, respectively).

For the R = 0.67 and 1.50 cases with imbalanced sample sizes, results in Table 72 show the same trends as were seen for the equal sample size cases. The Lprel test procedure incorrectly used the MW test for more than half of the R = 0.67 and R = 1.50 cases (55% and 59%, respectively); and correctly selected the Welch-AS test for approximately 30% of the R = 0.67 and 1.50 cases. The Fprel test procedure correctly selected the Welch-AS test for more than two-thirds of the R = 0.67 and 1.50 cases (67% and 71%, respectively). The MW test was incorrectly chosen by the Fprel test procedure for approximately 18% of the R = 0.67 and 1.50 cases.

R = 0.25 and 0.50 (Includes the Imbalanced and Balanced Cases)

Results in Table 71 show for the balanced case that the Fprel test procedure correctly selected the Welch-AS test for approximately 80% of the R = 0.25 and 0.50 cases. The Welch-S test was incorrectly used for at least 13% of the R = 0.25 and 0.50 cases. The Lprel test procedure correctly selected the Welch-AS test for over 50% of the R = 0.50 cases and approximately two-thirds of the R = 0.25 cases. The MW test was incorrectly used for about 33% of the R = 0.50 cases, whereas the Welch-S test and the MW test were each incorrectly used for approximately 16% of the R = 0.25 cases.

Results in Table 72 for the unequal sample size case show that the Fprel test procedure correctly used the Welch-AS test for the R = 0.25 and 0.50 cases (79% and 74%, respectively). The Lprel test procedure correctly used the Welch-AS test for approximately 60% of the R = 0.25 cases, whereas the MW test and the Welch-AS test were each incorrectly selected for about 40% of the R = 0.50 cases.

<u>R = 2.0 and 4.0 (Includes the Imbalanced and Balanced Cases)</u>

Tables 71-72 show the Welch-AS test was correctly chosen by the Fprel test procedure for at least 78% of the R = 2.0 and 4.0 cases regardless of sample size configuration. The Lprel test procedure incorrectly used the MW test for about 50% of the R = 2.0 and 4.0 cases and the Welch-AS test was correctly used for approximately 45% of the R = 2.0 and 4.0 cases.

In summary, for the R = 1 cases regardless of the sample size configuration, the Lprel test procedure predominately used the MW test correctly, whereas the Fprel test procedure generally incorrectly used the Welch-AS test. For the R =0.67 and 0.50 cases, the Fprel test procedure used the Welch-AS test for at least 66% of the simulations, whereas the MW test was incorrectly selected for at least 50% of the simulations when using the Lprel test procedure. For the 0.25 and 0.50 cases, the Fprel and the Lprel test procedures generally correctly used the Welch-AS test; however, the Welch-AS test was selected more often when using the Fprel test procedure. The Fprel test procedure generally correctly selected the Welch-AS test for the R = 2.0 and 4.0 cases, whereas the Lprel test procedure selected the Welch-AS test correctly and the MW test incorrectly each for about 50% of the simulations.

CHAPTER VIII

CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

For the cases where variance homogeneity is unknown to the practicing statistician, the use of a preliminary test for H_o: $\sigma_1^2 = \sigma_2^2$ preceding the means test yielded improved results over using the *t* test or the Mann-Whitney-Wilcoxon test alone for all cases except the asymmetric unequal variances cases, where no method maintained the stated Type I error rate. Preliminary testing for variance homogeneity where $\alpha \geq 0.25$ yielded more robust tests than where $\alpha < 0.25$. Results using a preliminary test for variance homogeneity where $\alpha \geq 0.25$ were comparable to those when using the Welch test alone.

Results using the Triples test at a significance level of $\alpha = 0.25$ were more robust and powerful for testing H_o: symmetry than when using the Triples test at $\alpha = 0.05$ or the D'Agostino S_U test at $\alpha = 0.05$ and $\alpha = 0.25$ for the one sample cases and the two sample cases.

For the cases where variance homogeneity and symmetry each are unknown to the practicing statistician, an expert system using the Fprel test procedure and the Lprel test procedure yielded improved results with respect to robustness over using the *t* test or the Mann-Whitney-Wilcoxon test alone, except for the asymmetry unequal variance cases, where no method maintained the stated Type I error rate. Results using the Fprel test procedure were comparable to those when using the Welch test alone. Based on the results of the simulated rejection rates, the Fprel test procedure was generally more robust than the Lprel test procedure except for the asymmetric equal variance cases.

The performance of the expert system was also evaluated by the frequency at which the expert system selected the most appropriate test of means. For the symmetric equal variance cases, the Lprel test procedure correctly selected the *t* test for approximately 47% of the simulated cases compared to approximately 30% when using the Fprel test procedure. For the symmetric cases with severely unequal variances (R = 0.25, 0.50, 2.0, and 4.0), the frequency at which the Welch test was correctly selected ranged from approximately 58-61% when using the Fprel test procedure and ranged from approximately 52-61% when using the Lprel test procedure. Asymmetry was incorrectly concluded for approximately 39% of the simulated symmetric cases when using either the Fprel or the Lprel test procedure.

The frequency at which the Fprel and the Lprel test procedure correctly concluded asymmetry ranged from approximately 83-99% of the simulated cases for the families of asymmetric distributions examined. It is noted that the frequency at which asymmetry was concluded when using the Fprel and the Lprel test procedure increased as the *R* value increased. For the asymmetric equal variance cases, the Lprel test procedure correctly selected the Mann-Whitney-Wilcoxon test for approximately 67% of the simulated cases compared to 28% when using the Fprel test procedure. For the asymmetric cases with severely unequal variances, the frequency at which the Fprel test procedure correctly concluded asymmetry and variance heterogeneity ranged from approximately 77-90% of the simulated cases compared to a range of 43-63% of the simulated cases when using the Lprel test procedure.

Results showed that the expert system examined in this simulation study concluded asymmetry too often for the symmetric cases.

8.2 Future Work

Since the expert system examined in this simulation study concluded asymmetry too often, it would be of interest to examine the performance of an expert system using the Triples test for testing of symmetry at a lower significance level such as $\alpha = 0.05$. An alternative approach would be to examine the performance of an expert system which concludes asymmetry only if both samples were judged to be asymmetric at $\alpha = 0.25$.

Other interests for future work are:

1. Explore the appropriateness of using two classes of preliminary

tests (a test for variance homogeneity and a test for symmetry) preceding confidence interval estimation for $\mu_1 - \mu_2$.

2. Explore the appropriateness of testing the equality of k means ($k > 2$) using means tests for k independent samples based on the results of using the two classes of preliminary tests.

3. For the outlier models (symmetric and asymmetric contaminated normal distributions) examined, only the MW test was robust in both cases. Additional types of outlier models need to be investigated. If the simulation results continued to show that the MW test was robust, it would be of interest to evaluate the performance of an expert system which included a preliminary test for outlier models in addition to the two classes of preliminary tests explored in this dissertation work.

BIBLIOGRAPHY

Ahrens, J.H., & Dieter, U. (1974). Computer methods for sampling from gamma, beta, poisson, and binomial distributions. Computing, 12, 223-246.

Bancroft, T.A., & Han, C. (1983, December). A note on pooling variances. Journal of the American Statistical Association, 78 (384), 981-983.

Bartlett, M.S., & Kendall, D.G. (1946). The statistical analysis of variance heterogeneity and the logarithmic transformation. J. Royal Statist. Soc., Series B, 8, 128-138.

Box, G.E. (1953). Non-normality and tests on variance. Biometrika, 40, 318-335.

Box, G.E. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. Annals of Mathematical Statistics, 29, 290-302.

Brown, M.B., & Forsythe, A.B. (1974a, June). Robust tests for the equality of variances. Journal of the American Statistical Association, 69 (346), 364-367.

Brown, M.B., & Forsythe, A.B. (1974b). The small behavior of some statistics which test the equality of several means. Technometrics, 16, 129-132.

Campbell, M. (1990). Harvard graphics 2.3 made easy. Berkeley, CA: Osborne McGraw-Hill.

Carroll, R.J., & Schneider, H. (1985). A note on Levene's tests for equality of variances. Statistics & Probability Letters, 3, 191-194.

Cochran, W.G. (1964). Approximate significance levels of the Behrens-Fisher test. Biometrics, 20, 191-195.

Cochran, W.G., & Cox, G.M. (1957). Experimental designs. New York: John Wiley and Sons.

Conover, W.J., Johnson, M.E., & Johnson, M.M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. Technometrics, 23 (4), 351-361.

Crawley, J.W., & Miller, C.E. (1987). A structured approach to Fortran. Englewood Cliffs, NJ: Prentice-Hall, Inc.

D'Agostino, R.B. (1970, April). Transformation to normality of the null distribution of g_1. Miscellanea, 43 (2), 679-681.

D'Agostino, R.B., Belanger, A. & D'Agostino, R.B., Jr. (1990). A suggestion for using powerful and informative tests of normality. The American Statistician, 44 (4), 316-321.

D'Agostino, R.B., & Stephens, M.A. (1986). Goodness-of-fit techniques. New York and Basel: Marcel Dekker, Inc.

Daniel, W.W. (1978). Applied nonparametric statistics. Boston: Houghton Mifflin Company.

Didday, R., & Page, R. (1977). Fortran for humans. New York: West Publishing Co.

Fligner, M.A., & Killeen, T.J. (1976). Distribution-free two-sample tests for scale. Journal of American Statistical Association, 71, 210-213.

Gans, D.J. (1981). Use of a preliminary test in comparing two sample means. Communication in Statistics B, Simulation & Computation, B10 (2), 163-174.

Gupta, M.K. (1967). An asymptotically nonparametric test of symmetry. Annals of Mathematical Statistics, 38, 849-866.

Gurland, J., & McCullough, R.S. (1962). Testing equality of means after a preliminary test of equality of variances. Biometrika, 49 (3 & 4), 403-417.

Hasting, N.A.J., & Peacock, J.B. (1975). Statistical distributions. New York: John Wiley & Sons.

IMSL. (1989a, January). Stat/library user's manual (Version 1.1). Houston, TX: Author.

IMSL. (1989b, December). Math/library user's manual (Version 1.1). Houston, TX: Author.

James, G.S. (1951). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. Biometrika, 38, 19-43.

Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. Biometrika, 36, 149-176.

Johnson, N.L., & Kotz, S. (1970). Distributions in statistics: Continuous univariate distribution-I. New York: John Wiley & Sons.

Keselman, H.J., Games, P.A., & Clinch, J.J. (1979). Tests for homogeneity of variances. Communication in Statistics B, Simulation & Computation, B8 (2), 113-129.

Krutchkoff, R.G. (1988, April 14). One-way fixed effects analysis of variances when the error variances may be unequal. Journal of Statistical Computation & Simulation, 30, 259-271.

Lauer, G.N., & Han, C.P. (1974, November). Power of Cochran's test in the Behrens-Fisher problem. Technometrics, 16 (4), 545-549.

Layard, M.W.J. (1973, March). Robust large-sample tests for homogeneity of variances. Journal of the American Statistical Association, 68 (341), 195-198.

Lehmann, E.L. (1983). Theory of point estimation. New York: John Wiley & Sons.

Loh, W. (1987, May 21). Some modifications of Levene's test of variance homogeneity. Journal of Statistical Computation & Simulation, 28, 213-226.

Markowski, C.A., & Markowski, E.P. (1990, November). Conditions for the effectiveness of a preliminary test of variance. The American Statistician, 44 (4), 322-326.

McCullough, R.S., Gurland, J., & Rosenberg, L. (1960). Small sample behaviour of certain tests of the hypothesis of equal means under variance heterogeneity. Biometrika, 47, 231-442.

Milliken, G.A., & Johnson, D.E. (1984). Analysis of messy data--Volume I: Designed experiments. New York: Van Nostrand Reinhold Co.

Moser, B.K., & Stevens, G.R. (1992, February). Homogeneity of variances in the two-sample means test. American Statistical Association, 46 (1), 19-21.

Moser, B.K., Stevens, G.R., & Watts, C.L. (1989). The two-sample t-test versus Satterthwaite's approximate F test. Communication in Statistics A, Theory & Methodology, 18 (11), 3963-3975.

Murphy, B.P. (1976). Comparison of some two sample means tests by simulation. Communication in Statistics B, Simulation and Computation, B5 (1), 23-32.

O'Brien, R.G. (1979). A general ANOVA method for robust tests of additive models for variances. Journal of the American Statistical Association, 74, 877-880.

Olejnik, S. (1987). Conditional ANOVA for mean differences when population variances are unknown. Journal of Experimental Education, 55, 141-148.

Olejnik, S., & Algina, J. (1985). Power analysis of selected parametric and nonparametric tests of scale when sampled distributions are non-normal. Paper presented at the meetings of the American Educational Research Association, Chicago.

Overall, J.E., & Woodward, J.A. (1974). A simple test for homogeneity of variance in complex factorial designs. Psychometrika, 39, 311-318.

Paull, A.E. (1950). On a preliminary test for pooling mean squares in the analysis of variance. Annals of Mathematics & Statistics, 21, 539-556.

Pearson, E.S. & Hartley, H.O. (1966). Biometrika tables for statisticians (Vol. I, 3rd edition). New York: Cambridge University Press.

Ramsey, P.H. (1980). Exact Type I error rates for robustness of student's t test with unequal variances. Journal of Educational Statistics, 5, 337-350.

Randles, R.H., Fligner, M.A., Policello II, G.E., & Wolfe, D.A. (1980, March). An asymptotically distribution-free test for symmetry versus asymmetry. Journal of the American Statistical Association, 75 (369), 168-172.

Rogan, J.C., & Keselman, H.J. (1977, Fall). Is the ANOVA F-test robust to variance heterogeneity when sample sizes are equal?: An investigation via a coefficient of variation. American Educational Research Journal, 14 (4), 493-498.

Rothschild, V., & Logothetis, N. (1986). Probability distributions. New York: John Wiley & Sons.

Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. Biometrics Bulletin, 2, 110-114.

Scheffé, H. (1959). The analysis of variance. New York: John Wiley and Sons.

Schmeiser, B.W., & Lal, R. (1980). Squeeze methods for generating gamma variates. Journal of the American Statistical Association, 75, 679-682.

Schultz, B. (1983, February 16). On Levene's test and other statistics of variation. Evolutionary Theory, 6, 197-203.

Shapiro, S.S., Wilk, M.B., & Chen, H.J. (1968). A comparative study of various tests for normality. Journal of the American Statistical Association, 63, 1343-78.

Snedecor, G.W., & Cochran, W.G. (1967). Statistical methods. Ames, IA: The Iowa State University Press.

Welch, B.L. (1937). The significance of difference between two means when the population variances are unequal. Biometrika, 29, 350-362.

Welch, B.L. (1947). The generalization of Student's problem when several different population variances are unequal. Biometrika, 34, 28-35.

Winer, B.J. (1962). Statistical principles in experimental design. New York: McGraw-Hill.

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

Zimmerman, D.W., & Williams, R.H. (1989). Power comparisons of the student t-test and two approximations when variances and sample sizes are unequal. Journal of Indian Society Agricultural Statistics, 41 (2), 206-217.