



University of North Florida  
**UNF Digital Commons**

---

[UNF Graduate Theses and Dissertations](#)

[Student Scholarship](#)

---

1999

## Robust I-Sample Analysis of Means Type Randomization Tests for Variances

Anthony Joseph Bernard  
*University of North Florida*

Follow this and additional works at: <https://digitalcommons.unf.edu/etd>

 Part of the [Mathematics Commons](#)

---

### Suggested Citation

Bernard, Anthony Joseph, "Robust I-Sample Analysis of Means Type Randomization Tests for Variances" (1999). *UNF Graduate Theses and Dissertations*. 90.  
<https://digitalcommons.unf.edu/etd/90>

This Master's Thesis is brought to you for free and open access by the Student Scholarship at UNF Digital Commons. It has been accepted for inclusion in UNF Graduate Theses and Dissertations by an authorized administrator of UNF Digital Commons. For more information, please contact [Digital Projects](#).  
© 1999 All Rights Reserved



**Robust *I*-Sample Analysis of Means Type Randomization Tests for Variances**

**by**

**Anthony Joseph Bernard**

**A thesis submitted to the Department of Mathematics and Statistics in  
partial fulfillment of the requirements for the degree of**

**Master of Science in Mathematical Sciences**

**UNIVERSITY OF NORTH FLORIDA**

**COLLEGE OF ARTS AND SCIENCES**

**September, 1999**

**Unpublished work © Anthony Joseph Bernard**

The thesis of Anthony Joseph Bernard is approved:

(Date)

**Signature Deleted**

9/24/99

~~\_\_\_\_\_~~  
**Signature Deleted**

9-24-99

~~\_\_\_\_\_~~  
**Signature Deleted**

9-24-99

~~\_\_\_\_\_~~  
Committee Chairperson

Accepted for the Department:

**Signature Deleted**

9-24-99

\_\_\_\_\_  
Chairperson

Accepted for the College:

**Signature Deleted**

9-27-99

~~\_\_\_\_\_~~  
Dean

Accepted for the University:

**Signature Deleted**

10-4-99

~~\_\_\_\_\_~~  
Dean of Graduate Studies

## ACKNOWLEDGEMENTS

I would like to thank Dr. Peter Wludyka for his guidance and patience. I am grateful to Dr. Adel Boules and Dr. Ping Sa for serving on the thesis committee. In addition, thanks to David Cristol for helping with the computer programming. Finally, I would like to thank my family and friends for their love and support.

## TABLE OF CONTENTS

Chapter 1 Introduction.....	1
1.1 PROBLEM DEFINITION .....	1
1.2 HYPOTHESIS TESTING.....	2
1.2.1 STANDARD HYPOTHESIS TESTS.....	3
1.2.2 RANDOMIZED HYPOTHESIS TESTS.....	4
1.2.3 RANDOMIZATION TESTS .....	4
1.2.3.1 EXACT RANDOMIZATION TESTS.....	5
1.2.3.2 APPROXIMATE RANDOMIZATION TESTS.....	7
1.3 I-SAMPLE TESTS FOR VARIANCES .....	9
Chapter 2 ANOVA-F and ANOM Type Tests for Variances .....	10
2.1 DATA TRANSFORMATIONS.....	10
2.2 ANOVA-F TESTS FOR VARIANCES .....	11
2.2.1 LEVENE'S TEST .....	12
2.2.2 FLIGNER-KILLEEN TEST .....	12
2.3 ANOM TYPE TESTS FOR VARIANCES .....	12
2.3.1 ANOMV-LEV.....	13
2.3.2 ANOMV-TR .....	14
2.3.3 ANOMV-JK.....	14
Chapter 3 Analysis of Means Type Randomization Tests for Variances.....	17
3.1 OVERVIEW OF PROPOSED RANDOMIZATION TESTS .....	17
3.2 RANDANOMV-D.....	18
3.2.1 RANDANOMV-D1 EXAMPLE .....	25
3.2.2 RANDANOMV-D EXAMPLE .....	25
3.3 RANDANOMV-DD .....	27
3.3.1 RANDANOMV-DD EXAMPLE .....	29
3.4 RANDANOMV-R .....	29
3.4.1 RANDANOMV-R EXAMPLE.....	31
3.5 RANDANOMV-RD.....	32
3.5.1 RANDANOMV-RD EXAMPLE.....	34
3.6 COMPUTER PROGRAM FOR THE RANDOMIZATION TESTS .....	35
Chapter 4 Variance Configurations.....	46
Chapter 5 Monte Carlo Study.....	49
5.1 METHODOLOGY.....	49
5.2 VARIANCE CONFIGURATIONS STUDIED .....	50
5.3 TYPE I ERROR STABILITY.....	51
5.4 POWER.....	53
5.5 UNEQUAL MEANS.....	53
Chapter 6 Equivalence of Two Proposed Randomization Tests for Variances.....	62
Chapter 7 Summary, Conclusions and Future Research .....	65
7.1 SUMMARY AND CONCLUSIONS.....	65
7.2 FUTURE RESEARCH .....	67
APPENDIX.....	68
FORTRAN PROGRAM FOR THE RANDOMIZATION TESTS .....	68

BIBLIOGRAPHY .....	76
VITA.....	78

## LIST OF FIGURES

Figure 1: ANOMV-LEV Decision Chart.....	15
Figure 2: RANDANOMV-D Decision Chart.....	44
Figure 3: RANDANOMV-DD Decision Chart.....	44
Figure 4: RANDANOMV-R Decision Chart.....	45
Figure 5: RANDANOMV-RD Decision Chart.....	45

## LIST OF TABLES

Table 1: Raw Data – Test Examples .....	16
Table 2: Absolute Deviations from Sample Median – Test Examples .....	16
Table 3: Outline of Randomization Tests .....	43
Table 4: Empirical Type I Error Rates for $I = 5$ .....	57
Table 5: Empirical Type I Error Rates for $I = 10$ .....	57
Table 6: Empirical Rejection Rates – Configuration 1 for $I = 5$ , KUR(6).....	58
Table 7: Empirical Rejection Rates – Configuration 1 for $I = 10$ , KUR(6).....	58
Table 8: Empirical Rejection Rates – Configuration 2 for $I = 5$ , KUR(6).....	59
Table 9: Empirical Rejection Rates – Configuration 2 for $I = 10$ , KUR(6).....	59
Table 10: Empirical Rejection Rates – Configuration 2 for $I = 5$ , SKW(3). .....	60
Table 11: Empirical Rejection Rates – Configuration 2 for $I = 10$ , SKW(3). .....	60
Table 12: Empirical Rejection Rates – Configuration 2. ....	61
Table 13: Empirical Type I Error Rates – Unequal Means.....	61
Table 14: Empirical Rejection Rates – Unequal Means .....	61



University of North Florida

Abstract

Robust *I*-Sample Analysis of Means Type Randomization Tests for Variances

by Anthony Joseph Bernard

Chairperson of the Supervisory Committee: Professor Dr. Peter Wludyka

Department of Mathematics and Statistics

The advent of powerful computers has brought about the randomization technique for testing statistical hypotheses. Randomization tests are based on shuffles or rearrangements of the (combined) sample. Putting each of the  $I$  samples “in a bowl” forms the combined sample. Drawing samples “from the bowl” forms a shuffle. Shuffles can be made with or without replacement.

In this thesis, analysis of means type randomization tests will be presented to solve the homogeneity of variance problem. An advantage of these tests is that they allow the user to graphically present the results via a decision chart similar to a Shewhart control chart. The focus is on finding tests that are robust to departures from normality. The proposed tests will be compared against commonly used nonrandomization tests. The type I error stability across several nonnormal distributions and the power of each test will be studied via Monte Carlo simulation.

## Chapter 1

### Introduction

#### 1.1 PROBLEM DEFINITION

The global marketplace has become extremely competitive, and companies are always searching for cheaper and faster ways to produce their goods. As part of gaining process knowledge, companies must focus on process variability. When highly variable manufacturing methods are identified, alternative procedures may be implemented. These “new” processes must be assessed to determine how they impact process variability. Consider, for example, a battery manufacturer who is interested in reducing the variability in battery lifetime. Since consumers do not want batteries with significantly different lifetimes, it is desirable to make batteries that perform consistently. The manufacturer may evaluate several pumps used to supply anode to the battery. The claim that all anode pumps have the same variability will be referred to as the homogeneity of variance (HOV) hypothesis. The HOV hypothesis for a single factor experiment with  $I$  factor levels (different pumps) will be represented as

$$H_0 : \sigma_1^2 = \dots = \sigma_I^2 \quad (1.1)$$

where  $\sigma_i^2$  is the variance of the  $i^{\text{th}}$  population. The alternative hypothesis is  $H_A$  : not  $H_0$ . The interest here is the case where  $I > 2$ ; that is, where three or more populations

are being compared. The focus of this paper will be on one-way balanced designs, but the discussions extend to more complex designs.

## 1.2 HYPOTHESIS TESTING

Estimation of parameters and tests of statistical hypotheses are the two major areas treated by statistical theory. Parameter estimation uses information gathered from sample data to determine the value of a population parameter, and hypothesis testing uses sample data to determine which of two statements regarding a distribution is correct. The discussion that follows in the remainder of this section and sections 1.2.2 and 1.2.3 was summarized from Wludyka (1999).

Suppose a researcher is interested in  $I$  populations, each with distribution function  $F_i(x_i, \theta_i)$ , where  $\theta_i \in (a_i, b_i) \subseteq (-\infty, \infty)$  and  $x_i \in W_i \subseteq (-\infty, \infty)$ . Often  $(a_i, b_i) = (-\infty, \infty)$ . Let

$$\Omega = (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_I, b_I)$$

be the parameter space. Furthermore, partition the parameter space into subsets  $\Omega_0 \subset \Omega$  and  $\Omega_A \subset \Omega$ , where  $\Omega_0$  and  $\Omega_A$  are disjoint. Frequently  $\Omega = \Omega_0 \cup \Omega_A$ . Often there is interest in the case where  $\Omega_0 = \{(\theta_1, \dots, \theta_I) \mid \theta_1 = \dots = \theta_I\}$ . This leads to the definition of a null and alternative hypothesis.

**DEFINITION 1.** A null hypothesis is a statement of the form  $H_0 : (\theta_1, \dots, \theta_I) \in \Omega_0$ . The corresponding alternative hypothesis is a statement of the form  $H_A : (\theta_1, \dots, \theta_I) \in \Omega_A$ .

To determine if  $H_0$  is true, samples of size  $n_i$  are drawn from each of the  $I$  populations. In this paper the samples will be independent and random. Denote the observations by  $x_{ij}$ ,  $i=1,\dots,I$  and  $j=1,\dots,n_i$ . The set of all samples will be denoted by  $\Psi$ . After the samples are collected the researcher will use the data to make a decision concerning the hypotheses.

DEFINITION 2. *The Decision Space  $D$  is the set of all decisions the researcher can make. The points in the space will be denoted by  $\delta_k$ .*

Usually the decision space consists of the points  $\delta_1 = \text{ACCEPT } H_0$  and  $\delta_2 = \text{REJECT } H_0$ . In other cases, sequential tests for example, the decision space may be larger.

### 1.2.1 STANDARD HYPOTHESIS TESTS

DEFINITION 3. *A (standard) hypothesis test is a rule that, given  $H_0$  and  $H_A$ , associates with each sample  $X$  a point in the decision space. That is, a hypothesis test is a function that maps from the set of all samples to the decision space:*

$$T(X): \Psi \rightarrow D.$$

DEFINITION 4. *Two statistical tests,  $T_1$  and  $T_2$ , are equivalent if any given  $H_0$  and  $H_A$ , each sample  $X$  is mapped to the same point in the decision space. That is,*

$$T_1(X) = T_2(X) \quad \forall X \in \Psi.$$

DEFINITION 5. *The Power Function of a test is that function which yields the probability of correctly rejecting the null hypothesis.*

### 1.2.2 RANDOMIZED HYPOTHESIS TESTS

Randomized tests are used most frequently when the populations of interest are discrete. These tests allow the user to achieve an exact significance level by using the result of a supplemental independent random experiment. Hogg and Craig (1995) and Rinaman (1993) give examples of randomized hypothesis tests.

### 1.2.3 RANDOMIZATION TESTS

There are two types of randomization tests: exact randomization tests and approximate randomization tests. Randomization tests are based on shuffles (resamplings or rearrangements) of the (combined) sample. The combined sample is formed by putting each of the  $I$  samples “in a bowl.” “Drawing from the bowl” forms shuffles. The shuffles can be made with replacement (called a bootstrap shuffle) or without replacement (called a permutation shuffle), the latter being more commonly used in practice.

DEFINITION 6. *A shuffle consists of  $n_i$  assignments from the combined sample of sample values to each of  $I$  groups. Distinct shuffles can be labeled  $X^{(q)}$ .*

For example, consider the sample below (for which  $I = 3$  and  $n_i = 2$ ).

$$X = \begin{pmatrix} 5 & 4 & 9 \\ 7 & 3 & 2 \end{pmatrix}.$$

An example of a (bootstrap) shuffle is given by

$$X^{(1)} = \begin{pmatrix} 7 & 5 & 4 \\ 3 & 7 & 3 \end{pmatrix}.$$

An example of a (permutation) shuffle is given by

$$X^{(1)} = \begin{pmatrix} 2 & 9 & 7 \\ 4 & 5 & 3 \end{pmatrix}.$$

In an exact randomization test each of the  $N_E$  shuffles is found. That is, associated with each sample  $X$  is a collection of shuffles

$$S(X) = \{ X^{(q)} \mid \text{each } X^{(q)} \text{ is a distinct shuffle } X, q = 1, \dots, N_E \}.$$

### 1.2.3.1 EXACT RANDOMIZATION TESTS

**DEFINITION 7.** *An exact randomization test is a rule that, given  $H_0$  and  $H_A$ , associates with each sample  $X$  (and the set of shuffles  $S(X)$  associated with  $X$ ) a point in the decision space. That is, an exact randomization test is a function mapping from the set of all samples to the decision space:*

$$T(X, S(X)): \Psi \rightarrow D.$$

It should be noted that the above definition is the same as the definition of a (standard) hypothesis test in DEFINITION 3. The explicit inclusion of  $S(X)$  in the definition is there for emphasis and as a reminder that the exact randomization test itself is performed

in a somewhat different manner than other tests. Performance of a randomization test usually involves:

1. calculation of a test statistic for the initial sample  $C(X)$
2. calculation of the same test statistic for each of the shuffles  $C(X^{(g)})$
3. a decision based on the “unusualness” of  $C(X)$  in relation to the set of  $C(X^{(g)})$ 's.

For example, suppose one is testing the equality of  $I$  means based on independent samples of size  $n_i$ . At step one an ANOVA-F statistic is calculated based on the initial sample. At step two an ANOVA-F statistic is calculated for each possible shuffle. In step three the list  $N_E + 1$  F-statistics is ordered and the empirical quantile associated with the initial sample is calculated. If the empirical quantile is less than a prespecified level of significance then the equal means hypothesis is rejected. An advantage offered by this test is that no assumption regarding the distribution of the ANOVA-F statistic is required (Edgington, 1987).

The practical difficulty associated with an exact randomization test is that the number of shuffles in  $S(X)$  can become prohibitively large and hence creating  $S(X)$  can be too expensive. Thus, this paper will focus on approximate randomization tests.

### 1.2.3.2 APPROXIMATE RANDOMIZATION TESTS

In an approximate randomization test a random sample from  $S(X)$  is selected. That is,  $N_A$  shuffles are randomly selected from  $S(X)$ . Typically a computer is used to generate a sequence of (pseudo) random shuffled samples from the combined sample. Denote this random sample of shuffles by

$$SA(X) = \{ X^{(q)} \mid \text{each } X^{(q)} \text{ is a random shuffle } X, q = 1, \dots, N_A \}.$$

DEFINITION 8. *An approximate randomization test is a rule that, given  $H_0$  and  $H_A$ , associates with each sample  $X$  (and the random set of shuffles  $SA(X)$  associated with  $X$ ) a point in the decision space. That is, an approximate randomization test is a function  $RT$  mapping from the set of all samples augmented by the set of all shuffles for each  $X$  to the decision space*

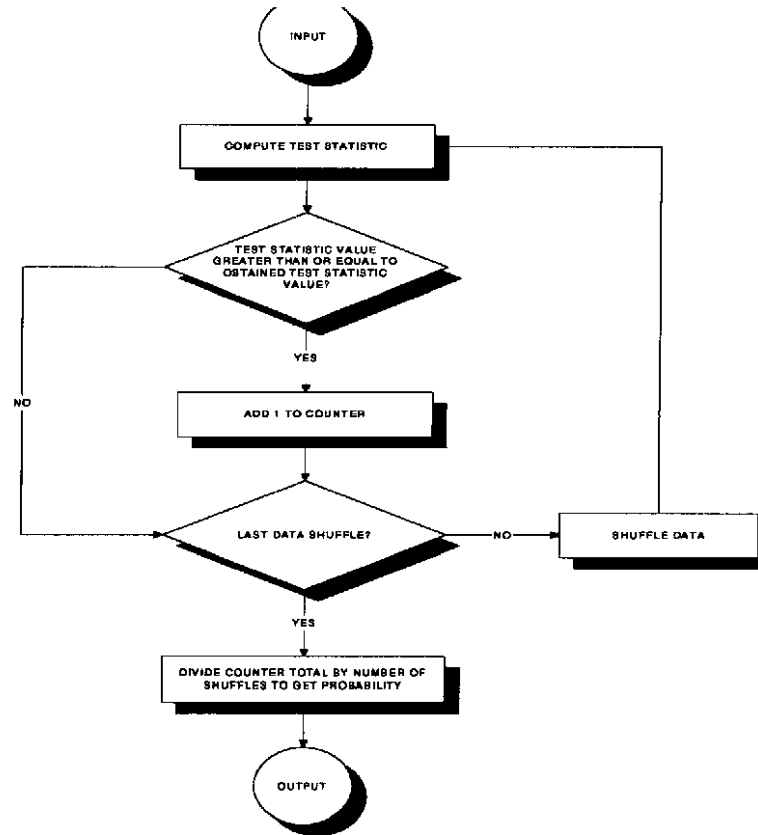
$$RT(X, SA(X)): \Psi, S(X) \rightarrow D.$$

DEFINITION 9. *Two randomization tests,  $RT_1$  and  $RT_2$ , are equivalent if for any given  $H_0$  and  $H_A$ , each sample  $X$  (and the same set of random shuffles  $SA(X)$  associated with  $X$ ) maps to the same point in the decision space. That is,*

$$RT_1(X, SA(X)) = RT_2(X, SA(X)).$$



The following flowchart, taken from Edgington (1987), will give the reader a better feel for the technique used for approximate randomization tests.



A key point is that the set of shuffles  $SA(X)$  is not unique to  $X$ . In practice there are a very large number of distinct  $SA(X)$ 's for any  $X$ . Hence, the decision is not uniquely determined by  $X$  and the test  $RT$ . For a test  $RT$  and a sample  $X$  there is associated with each point in the decision space a probability that the test will map to that decision. That is,

$$\Pr(\delta_i | X, N_A) = p_i.$$

In practice this probability will not be known. However, a heuristically sensible approximate randomization test should have the property that when the state of nature is such that decision  $\delta_i$  is the correct decision then for sufficiently large  $N_A$  the associated probability  $p_i$  should be close to one. When decision  $\delta_i$  is an incorrect decision then for sufficiently large  $N_A$  the associated probability  $p_i$  should be close to zero.

### 1.3 I-SAMPLE TESTS FOR VARIANCES

A user has many options in testing hypothesis (1.1). HOV tests proposed by Bartlett (1973), Hartley (1940 and 1950), Cochran (1941), Foster (1964) and Wludyka and Nelson (1997 A) may be employed when normality is a reasonable assumption. When the normality assumption is not valid the user may adopt an assortment of tests. See Conover, Johnson and Johnson (1981) for a good discussion. Robust analysis of means (ANOM) type tests have been proposed by Wludyka and Nelson (1997 B). The objective of this paper is to provide users with new HOV tests that are robust and compare these tests with some commonly used tests via a Monte Carlo study.

## Chapter 2

### ANOVA-F and ANOM Type Tests for Variances

#### 2.1 DATA TRANSFORMATIONS

In the previous chapter it was noted that a user has many options for testing hypothesis (1.1) when the normality assumption is not tenable. Transforming the original (location) measurements into scale measurements creates tests for variability by using the transformed measurements as input to standard location tests. Two standard location tests will be considered in this chapter: the ANOVA-F test and the ANOM test. Three standard transformations that have been used will be defined. In each of the transformations  $x_{ij}$  will be defined as the  $j^{\text{th}}$  observation from the  $i^{\text{th}}$  sample.

#### SQUARED DEVIATIONS FROM THE MEAN (SDM)

$$y_{ij} = (x_{ij} - \bar{x}_i)^2 \text{ where } \bar{x}_i \text{ is the mean of the } i^{\text{th}} \text{ sample} \quad (2.1)$$

#### ABSOLUTE DEVIATIONS FROM THE MEDIAN (ADM)

$$y_{ij} = |x_{ij} - x_i^{\text{med}}| \text{ where } x_i^{\text{med}} \text{ is the median of the } i^{\text{th}} \text{ sample} \quad (2.2)$$

Note that one of the absolute deviations from the median becomes zero when there are an odd number of observations in a sample. Wludyka and Nelson (1999) state that since

variability is being measured as the absolute deviation from the median, the median of the sample imparts no information about variability. Thus, for samples containing an odd number of observations the  $y_{ij}$  corresponding to the median should be deleted, and the sample size per group is reduced to  $n - 1$ .

### TRANSFORMATIONS OF RANKS (TR)

This transformation consists of three steps:

1.  $d_{ij} = |x_{ij} - x_i^{med}|$
2.  $r_{ij} = Rank(d_{ij})$ , the values of the combined sample are ranked from smallest to largest
3.  $y_{ij} = INV\left(\frac{0.5 + r_{ij}}{(2In + 1)}\right)$  where  $INV$  is the inverse normal score (2.3).

The inverse normal transformation is performed on a uniform random variable. This transformation produces a random variable that is approximately normal (Ross, 1997).

Either the ANOVA-F test or the analysis of means (ANOM) test can be applied to the transformed values (2.1) – (2.3).

### 2.2 ANOVA-F TESTS FOR VARIANCES

The ANOVA-F test for variances will be applied to scale transformed observations.

Denoting the transformed values by  $y_{ij}$ , the test statistic is

$$L = \frac{\sum_{i=1}^I n_i (\bar{y}_{i.} - \bar{y}_{..})^2 / (I-1)}{\sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 / (N-I)} \text{ where } \bar{y}_{i.} = \sum_{j=1}^{n_i} y_{ij} / n_i \text{ and } \bar{y}_{..} = \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij} / N. \quad (2.4)$$

HOV hypothesis (1.1) is rejected when  $L$  exceeds the  $100(1-\alpha)^{\text{th}}$  percentile of the F-distribution with  $(I-1)$  and  $(N-I)$  degrees of freedom.

### 2.2.1 LEVENE'S TEST

Various modifications of this test exist, but the version considered here is the one determined to be the best by Brown and Forsythe (1974) and Conover et al. (1981). Their version is simply the one-way ANOVA-F test on absolute deviations from the median, transformation (2.2).

### 2.2.2 FLIGNER-KILLEEN TEST

The version of the Fligner-Killeen (F-K) test presented here is one proposed by Conover et al. (1981). ANOVA-F test (2.4) uses the transformed ranks of the absolute deviations from the median (2.3) to perform the HOV test.

## 2.3 ANOM TYPE TESTS FOR VARIANCES

Ott (1967) was the first to introduce the ANOM. Wludyka and Nelson (1997 A) then developed an ANOM type test for variances (ANOMV). ANOM type tests are relatively simple to perform and they allow the user to assess practical and statistical significance by graphically displaying differences in a decision chart. The decision chart, similar to a Shewhart control chart, allows the user to view which populations differ from the overall

mean. The assumptions for ANOM are identical with those for the ANOVA-F test, and the two procedures have roughly the same power (Nelson, 1985). Three robust ANOM type HOV tests will be described: ANOMV-LEV, ANOMV-TR and ANOMV-JK.

### 2.3.1 ANOMV-LEV

ANOMV-LEV is the ANOM version of Levene's test. That is, the ANOM is applied to the absolute deviations from the median (2.2). The advantage of this new procedure is that the test can be presented via a decision chart that allows for graphical interpretations of the result.

The following example illustrates the procedure. The data in Table 1 consists of four

random samples of size 10. In Table 1,  $S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$ . Table 2 contains the

absolute deviations from the sample median. In Table 2,  $\bar{y}_i = \frac{1}{n} \sum_j (y_{ij})$  and

$S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2$ . Hypothesis (1.1) will be tested versus  $H_A$ : not the null. The

decision lines are constructed as follows:

$$UDL = \bar{\bar{Y}} + h_{\alpha} s \sqrt{(I-1)/(In)} = 0.893 + (2.59)(0.666)(0.2739) = 1.365$$

$$CL = \bar{\bar{Y}} = 0.893$$

$$LDL = \bar{\bar{Y}} - h_{\alpha} s \sqrt{(I-1)/(In)} = 0.893 - (2.59)(0.666)(0.2739) = 0.421$$

Where  $\bar{Y} = \sum \bar{y}_i / I$ ,  $s = \sqrt{\sum S_i^2 / I}$ , and critical value  $h_\alpha$  is obtained from Nelson (1983)

for  $I = 4$ ,  $n = 10$  and  $\alpha = 0.05$ , UDL is the upper decision line, CL is the center line and LDL is the lower decision line.

Since  $\bar{y}_4 = 1.464$  is above the UDL the HOV hypothesis (1.1) is rejected. The decision chart for the test is in Figure 1.

### 2.3.2 ANOMV-TR

ANOMV-TR is described in Wludyka and Nelson (1999). ANOMV-TR employs the same transformation as the Fligner-Killeen test. Instead of the ANOVA-F test, the ANOM is applied to the  $y_{ij}$  from (2.3).

### 2.3.3 ANOMV-JK

The third ANOM type test is ANOMV-JK where JK represents jackknifing. The technique is described in Wludyka and Nelson (1997 B). An overview of the test is presented below:

1. Replace each observation  $x_{ij}$  with a jackknifed variance
2. 
$$U_{ij} = n \ln(S_i^2) - (n-1) \ln(S_{i(j)}^2)$$
3. Perform ANOM on the  $U_{ij}$ .

Consider, for example, four samples containing five observations per sample. Each observation is replaced with a jackknifed variance. The jackknifed variance

$$S_{i(j)}^2 = \frac{1}{n-2} \sum_{k \neq j} (x_{ik} - \bar{x}_{i(k)})^2, \text{ where } \bar{x}_{i(j)} = \frac{1}{n-1} \sum_{k \neq j} x_{ik},$$

is the sample variance computed on the remaining four elements (the  $j^{\text{th}}$  observation is deleted). The “new” data set now consists of jackknifed variances. Wludyka and Nelson (1997 B) state that the resulting jackknifed variances are dependent and the ANOM may not be applied directly to them.

Thus, the transformation,  $U_{ij}$ , is applied (based on an idea of Tukey (1962)), and the ANOM procedure is applied directly to  $U_{ij}$ .

Figure 1: ANOMV-LEV Decision Chart

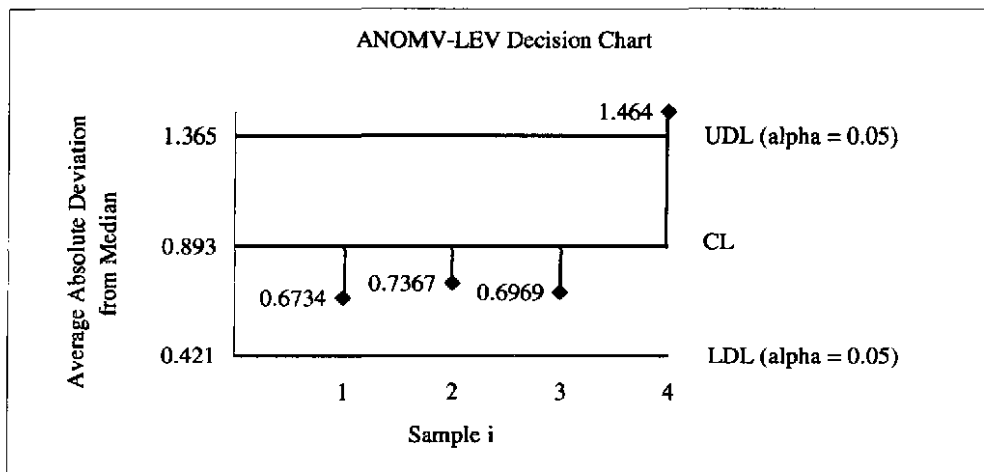




Table 1: Raw Data – Test Examples

	1	2	3	4
	0.5119	0.3756	-1.648	0.2821
	0.6137	0.1026	0.06582	-2.411
	-0.6311	-0.2989	0.5893	-4.359
	-0.5025	-0.5242	0.04181	-1.704
	1.577	-1.455	0.2225	0.9703
	-1.32	-1.481	-0.1234	-0.9169
	-0.3896	0.5361	-2.350	0.6221
	0.6738	0.6953	-0.5466	-3.017
	-0.09655	-1.243	1.269	-0.04799
	-0.6112	-1.253	0.1959	-2.235
$S_i^2$	0.7227	0.7384	1.120	3.071

Table 2: Absolute Deviations from Sample Median – Test Examples

	1	2	3	4
	0.7550	0.7872	1.702	1.593
	0.8568	0.5142	0.0120	1.101
	0.3880	0.1127	0.5355	3.049
	0.2594	0.1127	0.0120	0.3936
	1.820	1.043	0.1687	2.281
	1.077	1.069	0.1772	0.3934
	0.1465	0.9477	2.404	1.933
	0.9169	1.107	0.6004	1.707
	0.1465	0.8315	1.215	1.262
	0.3681	0.8415	0.1421	0.9246
$\bar{y}_i$	0.6734	0.7367	0.6969	1.464
$S_i^2$	0.2754	0.1375	0.6692	0.6920

## Chapter 3

### Analysis of Means Type Randomization Tests for Variances

#### 3.1 OVERVIEW OF PROPOSED RANDOMIZATION TESTS

Four ANOM type randomization tests for variances will be proposed. These tests can be performed using either permutation shuffles or bootstrapping shuffles. They are RANDANOMV-D, RANDANOMV-DD, RANDANOMV-R and RANDANOMV-RD.

These tests differ with respect to

1. the test statistics computed on the original data
2. the data that is shuffled
3. the test statistics computed on the shuffled data.

Table 3 outlines each randomization test with respect to the above items. In the table and

subsequently  $S_i^2 = \frac{1}{n-1} \sum (x_{ij} - \bar{x}_i)^2$  and  $\bar{S}^2 = \sum S_i^2 / I$ . From the table the reader may

see that the tests can be classified as either difference tests or ratio tests. The difference tests are RANDANOMV-D and RANDANOMV-DD, and the ratio tests are RANDANOMV-R and RANDANOMV-RD.

Two versions (one- and two-sided) of RANDANOMV-D will be presented. Only two-sided versions of the other tests will be presented. The two-sided version has an

advantage over the one-sided version in that a decision chart may be constructed. The decision chart enables its user to do two things: 1) assess practical as well as statistical significance and 2) determine which particular samples have more or less variability when compared to the other samples.

### 3.2 RANDANOMV-D

RANDANOMV-D is a randomization test that uses sample variances and a pooled estimate of the common variance (the average of the sample variances). Deviations of the sample variances from the common variance are computed, and extreme (maximum and minimum) deviations are used as test statistics on both the initial and shuffled data.

A one-sided (RANDANOMV-D1) and two-sided (RANDANOMV-D) version of this test will be investigated. As mentioned earlier, the two-sided version allows the user to construct decision lines and present a decision chart similar to ANOM.

The steps in RANDANOMV-D1 are as follows:

1. Calculate  $AD_0 = \max |S_i^2 - \overline{S^2}|$  on the initial sample where  $S_i^2$  is the sample variance and  $\overline{S^2}$  is the average sample variance
2. Randomly shuffle the original data some number of times,  $NS$
3. Calculate  $AD_j = \max |S_i^2 - \overline{S^2}|$  after each shuffle where  $S_i^2$  is the sample variance and  $\overline{S^2}$  is the average sample variance
4. If  $AD_j > AD_0$  then  $ge = ge + 1$

5. If  $p\text{-value} = \left( \frac{(ge + 1)}{(NS + 1)} \right) < \alpha$  then hypothesis (1.1) is rejected.

RANDANOMV-D is carried out by:

1. Calculate  $AD_{\max} = \max(S_i^2 - \overline{S^2})$  and  $AD_{\min} = \min(S_i^2 - \overline{S^2})$  on the initial sample where  $S_i^2$  is the sample variance and  $\overline{S^2}$  is the average sample variance
2. Randomly shuffle the original data some number of times,  $NS$
3. Calculate  $AD_{\max}^q = \max(S_i^2 - \overline{S^2})$  and  $AD_{\min}^q = \min(S_i^2 - \overline{S^2})$  after each shuffle where  $S_i^2$  is the sample variance and  $\overline{S^2}$  is the average sample variance
4. If  $AD_{\max}^q > AD_{\max}$  then  $ngmx = ngmx + 1$
5. If  $AD_{\min}^q < AD_{\min}$  then  $ngmn = ngmn + 1$
6. If  $p\text{-value-high} = \left( \frac{(ngmx + 1)}{(NS + 1)} \right) < \frac{\alpha}{2}$  or  $p\text{-value-low} = \left( \frac{(ngmn + 1)}{(NS + 1)} \right) < \frac{\alpha}{2}$  then hypothesis (1.1) is rejected.

This procedure will be called the p-value method.

A decision chart for RANDANOMV-D can be constructed. For level of significance  $\alpha$  the decision lines are

$$UDL = \overline{S^2} + AD_{\max}^{(1-\alpha/2)} \quad (3.1)$$

$$CL = \overline{S^2} \quad (3.2)$$

$$LDL = \overline{S^2} + AD_{\min}^{(\alpha/2)} \quad (3.3)$$

UDL is the upper decision line, CL is the center line, LDL is the lower decision line,

$AD_{\max}^{(1-\alpha/2)}$  is the (upper)  $\frac{\alpha}{2}$  quantile of  $AD_{\max}^q$   $q=1, \dots, NS$  and  $AD_{\min}^{(\alpha/2)}$  is the (lower)  $\frac{\alpha}{2}$

quantile of  $AD_{\min}^q$   $q=1, \dots, NS$ . Hypothesis (1.1) is rejected when at least one  $S_i^2$  plots

outside the decision lines.  $AD_{\max}^{(1-\alpha/2)}$  can be found by ordering the set

$A = \{AD_{\max}^q \mid q=1, \dots, NS\}$ . Then  $AD_{\max}^{(1-\alpha/2)}$  is the  $\left( NS - \left[ \left[ (NS+1)\frac{\alpha}{2} - 1 \right] \right] \right)^{th}$  largest value

in set A, where  $\lceil X \rceil$  is the greatest integer in  $X$ . Denote this as  $AD_{\max}^{\lceil \cdot \rceil}$ .  $AD_{\min}^{(\alpha/2)}$  can be

found by ordering the set  $B = \{AD_{\min}^q \mid q=1, \dots, NS\}$ . Then  $AD_{\min}^{(\alpha/2)}$  is the

$\left( NS - \left[ \left[ (NS+1)\left(1 - \frac{\alpha}{2}\right) - 1 \right] \right] \right)^{th}$  smallest value in set B, where  $\lfloor X \rfloor$  is the smallest

integer in  $X$ . Denote this as  $AD_{\min}^{\lfloor \cdot \rfloor}$ . This procedure will be called the decision chart

method.

**THEOREM** : The decision chart method is equivalent to the p-value method.

**PROOF** :

Notation:

$AD_{\max}^{(x)}$  = X order statistic; therefore,  $AD_{\max}^{(1000)}$  = largest order statistic.

$AD_{\min}^{(x)}$  = X order statistic; therefore,  $AD_{\min}^{(1)}$  = smallest order statistic.

$\lceil X \rceil$  is the greatest integer in  $X$ , and  $\lfloor X \rfloor$  is the smallest integer in  $X$ .

Assume the p-value method rejects  $H_0$  because p-value-high  $< \frac{\alpha}{2}$ .

$$\Rightarrow \left( \frac{(ngmx + 1)}{(NS + 1)} \right) < \frac{\alpha}{2} \text{ and } ngmx < (NS + 1) \left( \frac{\alpha}{2} \right) - 1.$$

$$\Rightarrow ngmx \leq \left\lceil \left[ (NS + 1) \frac{\alpha}{2} - 1 \right] \right\rceil$$

e.g.  $NS = 1,000$  and  $\alpha = 0.05$ ,  $ngmx \leq [24.025]$  or  $ngmx \leq 24$

There are 24 or fewer shuffles such that  $AD_{\max}^q > AD_{\max}$

$$\Rightarrow AD_{\max}^{(976)} < \max_i (S_i^2 - \overline{S^2}) \Rightarrow \exists i \ni S_i^2 - \overline{S^2} > AD_{\max}^{(976)}$$

$$\Rightarrow S_i^2 > \overline{S^2} + AD_{\max}^{(976)} \text{ for some } i$$

$\Rightarrow$  Reject  $H_0$  using the decision chart method.

Assume the p-value method rejects  $H_0$  because p-value-low  $< \frac{\alpha}{2}$ .

$$\Rightarrow \left( \frac{(ngmn + 1)}{(NS + 1)} \right) < \frac{\alpha}{2} \text{ and } ngmn < (NS + 1) \left( \frac{\alpha}{2} \right) - 1.$$

$$\Rightarrow ngmn \leq \left\lceil \left[ (NS + 1) \frac{\alpha}{2} - 1 \right] \right\rceil$$

e.g.  $NS = 1,000$  and  $\alpha = 0.05$ ,  $ngmn \leq [24.025]$  or  $ngmn \leq 24$

There are 24 or fewer shuffles such that  $AD_{\min}^q < AD_{\min}$

$$\Rightarrow AD_{\min}^{(25)} > \min_i (S_i^2 - \overline{S^2}) \Rightarrow \exists i \ni S_i^2 - \overline{S^2} < AD_{\min}^{(25)}$$

$$\Rightarrow S_i^2 < \overline{S^2} + AD_{\min}^{(25)} \text{ for some } i$$

⇒ Reject  $H_0$  using the decision chart method.

Assume the p-value method does not reject  $H_0$  because p-value-high  $\geq \frac{\alpha}{2}$ .

$$\Rightarrow \left( \frac{ngmx + 1}{NS + 1} \right) \geq \frac{\alpha}{2} \text{ and } ngmx \geq \left[ (NS + 1) \left( \frac{\alpha}{2} \right) - 1 \right] + 1.$$

e.g. There are 25 or more shuffles such that  $AD_{\max}^q > AD_{\max}$  when  $NS = 1,000$  and  $\alpha = 0.05$ ,  $ngmx \geq [24.025] + 1 = 25$ .

$$\Rightarrow AD_{\max}^{(976)} > \max(S_i^2 - \bar{S}^2) \Rightarrow S_i^2 < \bar{S}^2 + AD_{\max}^{(976)} \text{ for all } i$$

⇒ Do not reject  $H_0$  with the decision chart method.

Assume the p-value method does not reject  $H_0$  because p-value-low  $\geq \frac{\alpha}{2}$ .

$$\Rightarrow \left( \frac{ngmn + 1}{NS + 1} \right) \geq \frac{\alpha}{2} \text{ and } ngmn \geq \left[ (NS + 1) \left( \frac{\alpha}{2} \right) - 1 \right] + 1.$$

e.g. There are 25 or more shuffles such that  $AD_{\min}^q < AD_{\min}$  when  $NS = 1,000$  and  $\alpha = 0.05$ ,  $ngmn \geq [24.025] + 1 = 25$ .

$$\Rightarrow AD_{\min}^{(25)} < \min(S_i^2 - \bar{S}^2) \Rightarrow S_i^2 > \bar{S}^2 + AD_{\min}^{(25)} \text{ for all } i$$

⇒ Do not reject  $H_0$  with the decision chart rule.

Assume the decision chart procedure rejects  $H_0$  by exceeding the upper decision

line. Thus,  $\exists i \ni S_i^2 > \bar{S}^2 + AD_{\max}^{(976)}$

$$\Rightarrow S_i^2 - \overline{S^2} > AD_{\max}^{(976)} \text{ for some } i$$

$$\Rightarrow AD_{\max}^{(976)} < \max_i (S_i^2 - \overline{S^2}) = AD_{\max}$$

e.g. There are 24 or fewer  $AD_{\max}^q > AD_{\max}$  when  $NS = 1,000$  and  $\alpha = 0.05$ .

$$\Rightarrow ngmx \leq 24 \Rightarrow \text{p-value-high} = \frac{ngmx + 1}{NS + 1} \leq \frac{24 + 1}{1001} = .024975 < .025$$

$\Rightarrow$  Reject  $H_0$  using the p-value rule.

Assume the decision chart procedure rejects  $H_0$  by exceeding the lower decision

line. Thus,  $\exists i \ni S_i^2 < \overline{S^2} + AD_{\min}^{(25)}$

$$\Rightarrow S_i^2 - \overline{S^2} < AD_{\min}^{(25)} \text{ for some } i$$

$$\Rightarrow AD_{\min}^{(25)} > \min_i (S_i^2 - \overline{S^2}) = AD_{\min}$$

e.g. There are 24 or fewer  $AD_{\min}^q < AD_{\min}$  when  $NS = 1,000$  and  $\alpha = 0.05$ .

$$\Rightarrow ngmn \leq 24 \Rightarrow \text{p-value-low} = \frac{ngmn + 1}{NS + 1} \leq \frac{24 + 1}{1001} = .024975 < .025$$

$\Rightarrow$  Reject  $H_0$  using the p-value rule.

Assume the decision chart rule does not reject  $H_0$ .

$$\Rightarrow S_i^2 < \overline{S^2} + AD_{\max}^{(976)} \forall i \Rightarrow S_i^2 - \overline{S^2} < AD_{\max}^{(976)} \forall i$$

$$\Rightarrow \max_i (S_i^2 - \overline{S^2}) < AD_{\max}^{(976)} \Rightarrow AD_{\max} < AD_{\max}^{(976)}$$

e.g. There are 25 or more  $AD_{\max}^{(x)} > AD_{\max}$  when  $NS = 1,000$  and  $\alpha = 0.05$ .



$$\Rightarrow ngmx \geq 25 \Rightarrow \text{p-value-high} = \frac{ngmx + 1}{NS + 1} = \frac{25 + 1}{1001} = .02597 > .025$$

$\Rightarrow$  Do not reject  $H_0$  using the p-value rule.

Assume the decision chart rule does not reject  $H_0$ .

$$\Rightarrow S_i^2 > \bar{S}^2 + AD_{\min}^{(25)} \quad \forall i \Rightarrow S_i^2 - \bar{S}^2 > AD_{\min}^{(25)} \quad \forall i$$

$$\Rightarrow \min_i (S_i^2 - \bar{S}^2) > AD_{\min}^{(25)} \Rightarrow AD_{\min} > AD_{\min}^{(25)}$$

e.g. There are 25 or more  $AD_{\min}^{(x)} < AD_{\min}$  when  $NS = 1,000$  and  $\alpha = 0.05$ .

$$\Rightarrow ngmn \geq 25 \Rightarrow \text{p-value-low} = \frac{ngmn + 1}{NS + 1} = \frac{25 + 1}{1001} = .02597 > .025$$

$\Rightarrow$  Do not reject  $H_0$  using the p-value rule.

In general,

$$\left( \frac{(ngmx + 1)}{(NS + 1)} \right) < \frac{\alpha}{2}, \quad \left( \frac{(ngmn + 1)}{(NS + 1)} \right) < \frac{\alpha}{2}$$

$$\Rightarrow ngmx < (NS + 1) \left( \frac{\alpha}{2} \right) - 1, \quad ngmn < (NS + 1) \left( \frac{\alpha}{2} \right) - 1$$

$$\Rightarrow ngmx \leq \left\lceil (NS + 1) \frac{\alpha}{2} - 1 \right\rceil, \quad ngmn \leq \left\lceil (NS + 1) \frac{\alpha}{2} - 1 \right\rceil$$

So the upper and lower decision lines are

$$UDL = \bar{S}^2 + AD_{\max}^{NS - \left\lceil (NS + 1) \frac{\alpha}{2} - 1 \right\rceil} \quad \text{and} \quad LDL = \bar{S}^2 + AD_{\min}^{NS - \left\lceil (NS + 1) \left( 1 - \frac{\alpha}{2} \right) - 1 \right\rceil}.$$

e.g. When  $\alpha = 0.10$  and  $NS = 1,000$ :  $ngmx^* = 49$  and  $AD_{\max}^{951}$ ,  $ngmn^* = 49$  and  $AD_{\min}^{50}$ .

### 3.2.1 RANDANOMV-D1 EXAMPLE

The data in Table 1 will be used to illustrate the method.

$$\overline{S^2} = \frac{1}{4} \sum_{i=1}^4 S_i^2 = 1.413 \quad (3.4)$$

$$AD_0 = \max_i |S_i^2 - \overline{S^2}| = 1.658 \quad (3.5)$$

$$\text{p-value} = \frac{ge + 1}{NS + 1} = \frac{79 + 1}{1001} = 0.08 \quad (3.6)$$

The average sample variance was calculated to be 1.413. The test statistic computed on the original data ( $AD_0$ ) was 1.658. The data was shuffled 1,000 times and a significance level ( $\alpha$ ) of 0.05 was used. The p-value associated with RANDANOMV-D1 was determined to be 0.08. Therefore, the initial test statistic was exceeded 79 times out of the 1,000 data shuffles. Since  $0.08 > 0.05$  one may conclude that 1.658 was not unusually large, and hypothesis (1.1) was not rejected.

### 3.2.2 RANDANOMV-D EXAMPLE

The test statistics generated from the p-value method are below.

$$AD_{\max} = \max_i (S_i^2 - \overline{S^2}) = 1.658 \quad (3.7)$$

$$AD_{\min} = \min_i (S_i^2 - \overline{S^2}) = -0.690 \quad (3.8)$$

$$p\text{-high} = \frac{ngmx + 1}{NS + 1} = \frac{79 + 1}{1001} = 0.08 \quad (3.9)$$

$$p\text{-low} = \frac{ngmn + 1}{NS + 1} = \frac{635 + 1}{1001} = 0.635 \quad (3.10)$$

Test statistics on the original sample are given in (3.7) and (3.8). The data was shuffled 1,000 times and a significance level ( $\alpha$ ) of 0.05 was used. P-low was 0.635 and p-high was 0.080. Hypothesis (1.1) was not rejected since both p-high and p-low  $> 0.025$ .

The RANDANOMV-D decision chart for the example is in Figure 2. From 1,000 shuffles there were 1,000 values of  $AD_{\max}^q$  and 1,000 values of  $AD_{\min}^q$  generated. Since

$$\left( NS - \left[ \left[ (NS + 1) \frac{\alpha}{2} - 1 \right] \right] \right) = 1000 - \left[ \left[ (1001) \frac{0.05}{2} - 1 \right] \right] = 1000 - 24 = 976, \quad \text{the required}$$

$$\text{value was } AD_{\max}^{[976]} \text{ which } AD_{\max}^{(976)} = 1.852. \quad \text{Since } \left( NS - \left[ \left[ (NS + 1) \left( 1 - \frac{\alpha}{2} \right) - 1 \right] \right] \right) =$$

$$1000 - \left[ \left[ (1001) \left( 1 - \frac{0.05}{2} \right) - 1 \right] \right] = 1000 - 975 = 25, \text{ the required value was } AD_{\min}^{[25]} \text{ which}$$

$AD_{\min}^{(025)} = -1.223$ . The decision lines were determined to be

$$UDL = \bar{S}^2 + AD_{\max}^{(976)} = 1.413 + 1.852 = 3.265 \quad (3.11)$$

$$CL = \bar{S}^2 = 1.413 \quad (3.12)$$

$$LDL = \bar{S}^2 + AD_{\min}^{(025)} = 1.413 + (-1.223) = 0.190 \quad (3.13)$$

Notice each  $S_i^2$  was within the decision lines. That is, there were no differences in variability among the samples.

### 3.3 RANDANOMV-DD

Baker (1995) writes that many two-sample non-parametric tests of variability assume a common center of location. Good (1994) offers a solution (a two-sample test) that does not assume a common center of location. Good's test permutes squared deviations from the sample median. RANDANOMV-DD uses a similar modification where deviations from the mean are permuted, and a sample variance-type statistic is computed on the shuffled data. The reader may consult Table 3.

RANDANOMV-DD uses the sample variance and a pooled estimate of the sample variances (the average sample variance) to determine test statistics on the initial data. Deviations of the sample variance from the pooled value are computed. As in RANDANOMV-D, extreme deviations (maximum and minimum) are used as initial test statistics.

RANDANOMV-DD employs the following steps:

1. Calculate the mean of each sample,  $\bar{x}_i$
2. Calculate  $z_{ij} = x_{ij} - \bar{x}_i$

3. Calculate  $S_i^2 = \frac{\sum_{j=1}^{n_i} z_{ij}^2}{(n_i - 1)}$  and  $\bar{S}^2 = \frac{\sum_{i=1}^I S_i^2}{I}$

4. Calculate  $AD_{\max} = \max(S_i^2 - \bar{S}^2)$  and  $AD_{\min} = \min(S_i^2 - \bar{S}^2)$  on the initial sample
5. Randomly shuffle  $z_{ij}$  some number of times,  $NS$
6. Calculate  $P_i^2 = \frac{\sum_{j=1}^{n_i} z_{ij}^2}{(n_i - 1)}$  and  $\bar{P}^2$  after each shuffle
7. Calculate  $AD_{\max}^q = \max(P_i^2 - \bar{P}^2)$  and  $AD_{\min}^q = \min(P_i^2 - \bar{P}^2)$  after each shuffle
8. If  $AD_{\max}^q > AD_{\max}$  then  $ngmx = ngmx + 1$
9. If  $AD_{\min}^q < AD_{\min}$  then  $ngmn = ngmn + 1$
10. If  $p\text{-value-high} = \left( \frac{(ngmx + 1)}{(NS + 1)} \right) < \frac{\alpha}{2}$  or  $p\text{-value-low} = \left( \frac{(ngmn + 1)}{(NS + 1)} \right) < \frac{\alpha}{2}$  then hypothesis (1.1) is rejected.

As with RANDANOMV-D, this procedure will be called the p-value method.

Using (3.1) – (3.3) and a level of significance  $\alpha$ , a decision chart for RANDANOMV-DD can be constructed. Similar to RANDANOMV-D,  $AD_{\max}^{(1-\alpha/2)}$  is the (upper)  $\frac{\alpha}{2}$  quantile of  $AD_{\max}^q$   $q=1, \dots, NS$  and  $AD_{\min}^{(\alpha/2)}$  is the (lower)  $\frac{\alpha}{2}$  quantile of  $AD_{\min}^q$   $q=1, \dots, NS$ . Hypothesis (1.1) is rejected when at least one  $S_i^2$  plots outside the decision lines. This will be denoted the decision chart procedure.

### 3.3.1 RANDANOMV-DD EXAMPLE

The data in Table 1 will be used to illustrate the method. Test statistics (3.7) and (3.8) were the test statistics on the original sample for RANDANOMV-DD. The p-values from the test were

$$p\text{-high} = \frac{ngmx + 1}{NS + 1} = \frac{13 + 1}{1001} = 0.014 \quad (3.14)$$

$$p\text{-low} = \frac{ngmn + 1}{NS + 1} = \frac{327 + 1}{1001} = 0.328 \quad (3.15)$$

Since (3.14) < 0.025 hypothesis (1.1) was rejected.

The RANDANOMV-DD decision chart for the example is in Figure 3.  $AD_{\max}^{(.976)}$  and  $AD_{\min}^{(.025)}$  were computed in the same manner as was done for RANDANOMV-D. The decision lines were determined to be

$$UDL = \overline{S^2} + AD_{\max}^{(.976)} = 1.413 + 1.53 = 2.943 \quad (3.16)$$

$$CL = \overline{S^2} = 1.413 \quad (3.17)$$

$$LDL = \overline{S^2} + AD_{\min}^{(.025)} = 1.413 + (-1.008) = 0.405 \quad (3.18)$$

From Figure 3 one may see that  $S_4^2$  was above the upper decision line.

### 3.4 RANDANOMV-R

RANDANOMV-D and RANDANOMV-DD have test statistics that are based on extreme values of differences from some pooled estimate. RANDANOMV-R uses a ratio of the

sample variance to the sum of the sample variances. Extreme (largest and smallest) ratios are used as test statistics for the initial and shuffled data. RANDANOMV-R is a randomization version of the Analysis of Means for Variances presented by Wludyka and Nelson (1997 A).

RANDANOMV-R uses the following steps:

$$1. \quad \text{Calculate } AD_{\max} = \max \left( \frac{S_i^2}{\sum_{i=1}^I S_i^2} \right) \text{ and } AD_{\min} = \min \left( \frac{S_i^2}{\sum_{i=1}^I S_i^2} \right) \text{ on}$$

the initial sample where  $S_i^2$  is the sample variance

2. Randomly shuffle the original data some number of times,  $NS$

$$3. \quad \text{Calculate } AD_{\max}^q = \max \left( \frac{S_i^2}{\sum_{i=1}^I S_i^2} \right) \text{ and } AD_{\min}^q = \min \left( \frac{S_i^2}{\sum_{i=1}^I S_i^2} \right) \text{ after}$$

each shuffle where  $S_i^2$  is the sample variance

4. If  $AD_{\max}^q > AD_{\max}$  then  $ngmx = ngmx + 1$

5. If  $AD_{\min}^q < AD_{\min}$  then  $ngmn = ngmn + 1$

6. If  $p\text{-value-high} = \left( \frac{(ngmx + 1)}{(NS + 1)} \right) < \frac{\alpha}{2}$  or  $p\text{-value-low} =$

$\left( \frac{(ngmn + 1)}{(NS + 1)} \right) < \frac{\alpha}{2}$  then hypothesis (1.1) is rejected.

This procedure will be called the p-value method.

A decision chart for RANDANOMV-R can be constructed. However, since RANDANOMV-R is a ratio test, for a level of significance  $\alpha$  the decision lines are

$$UDL = \left( \sum_{i=1}^I S_i^2 \right) \left( AD_{\max}^{(1-\alpha/2)} \right) \quad (3.19)$$

$$CL = \overline{S^2} \quad (3.20)$$

$$LDL = \left( \sum_{i=1}^I S_i^2 \right) \left( AD_{\min}^{(\alpha/2)} \right) \quad (3.21)$$

$AD_{\max}^{(1-\alpha/2)}$  and  $AD_{\min}^{(\alpha/2)}$  represent the same values as presented in sections 3.2 and 3.3. This method will be denoted the decision chart procedure.

#### 3.4.1 RANDANOMV-R EXAMPLE

The data in Table 1 will be used to illustrate the method. Test statistics from the p-value method are below.

$$AD_{\max} = \max_i \left( \frac{S_i^2}{\sum_{i=1}^I S_i^2} \right) = 0.543 \quad (3.22)$$

$$AD_{\min} = \min_i \left( \frac{S_i^2}{\sum_{i=1}^I S_i^2} \right) = 0.128 \quad (3.23)$$

$$p\text{-high} = \frac{ngmx + 1}{NS + 1} = \frac{39 + 1}{1001} = 0.04 \quad (3.24)$$

$$p\text{-low} = \frac{ngmn + 1}{NS + 1} = \frac{528 + 1}{1001} = 0.528 \quad (3.25)$$



Test statistics on the original sample are given in (3.22) and (3.23). The data was shuffled 1,000 times and a significance level ( $\alpha$ ) of 0.05 was used. P-low was 0.528, and p-high was 0.040. Hypothesis (1.1) was not rejected since both p-low and p-high  $>$  0.025.

The RANDANOMV-R decision chart for the example is in Figure 4.  $AD_{\max}^{(.976)}$  and  $AD_{\min}^{(.025)}$  were computed in the same manner as was done for RANDANOMV-D and RANDANOMV-DD. The decision lines were determined to be

$$UDL = \left( \sum_{i=1}^I S_i^2 \right) \left( AD_{\max}^{(.976)} \right) = 5.653 \times 0.552 = 3.121 \quad (3.26)$$

$$CL = \overline{S^2} = 1.413 \quad (3.27)$$

$$LDL = \left( \sum_{i=1}^I S_i^2 \right) \left( AD_{\min}^{(.025)} \right) = 5.653 \times 0.049 = 0.278 \quad (3.28)$$

From Figure 4 one may see that each  $S_i^2$  was within the decision lines. That is, there were no differences in variability among the samples.

### 3.5 RANDANOMV-RD

RANDANOMV-RD is the ratio version of RANDANOMV-DD. The same issue presented in section 3.3 motivates RANDANOMV-RD. This test calculates initial test statistics using the ratio of each sample variance to the sum of the sample variances (similar to RANDANOMV-R), shuffles deviations from the mean (similar to RANDANOMV-DD) and calculates a ratio of sample variance-type statistics on the

shuffled data (similar to RANDANOMV-DD). Extreme ratios (maximum and minimum) are used as test statistics.

RANDANOMV-RD is carried out using the following steps:

1. Calculate the mean of each sample,  $\bar{x}_i$

2. Calculate  $z_{ij} = x_{ij} - \bar{x}_i$

3. Calculate  $S_i^2 = \frac{\sum_{j=1}^{n_i} z_{ij}^2}{(n_i - 1)}$

4. Calculate  $AD_{\max} = \max \left( \frac{S_i^2}{\sum_{i=1}^l S_i^2} \right)$  and  $AD_{\min} = \min \left( \frac{S_i^2}{\sum_{i=1}^l S_i^2} \right)$  on

the initial sample

5. Randomly shuffle  $z_{ij}$  some number of times,  $NS$

6. Calculate  $P_i^2 = \frac{\sum_{j=1}^{n_i} z_{ij}^2}{(n_i - 1)}$  after each shuffle

7. Calculate  $AD_{\max}^q = \max \left( \frac{P_i^2}{\sum_{i=1}^l P_i^2} \right)$  and  $AD_{\min}^q = \min \left( \frac{P_i^2}{\sum_{i=1}^l P_i^2} \right)$

8. If  $AD_{\max}^q > AD_{\max}$  then  $ngmx = ngmx + 1$

9. If  $AD_{\min}^q < AD_{\min}$  then  $ngmn = ngmn + 1$

$$10. \quad \text{If } p\text{-value-high} = \left( \frac{(ngmx + 1)}{(NS + 1)} \right) < \frac{\alpha}{2} \quad \text{or} \quad p\text{-value-low} = \left( \frac{(ngmn + 1)}{(NS + 1)} \right) < \frac{\alpha}{2} \text{ then hypothesis (1.1) is rejected.}$$

This will be referred to as the p-value method.

Using the same ideas presented for RANDANOMV-R, a decision chart can be constructed using (3.19) – (3.21) for a level of significance  $\alpha$ . This will be called the decision chart procedure.

### 3.5.1 RANDANOMV-RD EXAMPLE

The data in Table 1 will be used to illustrate the method. Test statistics (3.22) and (3.23) are the test statistics on the original sample for RANDANOMV-RD. The p-values from the test were

$$p\text{-high} = \frac{ngmx + 1}{NS + 1} = \frac{13 + 1}{1001} = 0.014 \quad (3.29)$$

$$p\text{-low} = \frac{ngmn + 1}{NS + 1} = \frac{327 + 1}{1001} = 0.328 \quad (3.30)$$

Since (3.29) < 0.025, hypothesis (1.1) was rejected.

The RANDANOMV-RD decision chart for the example is in Figure 5.  $AD_{\max}^{(976)}$  and  $AD_{\min}^{(025)}$  were computed in the same manner as was done for the other randomization tests. The decision lines were constructed in a similar manner to those for RANDANOMV-R. The decision lines were determined to be

$$UDL = \left( \sum_{i=1}^I S_i^2 \right) \left( AD_{\max}^{(.976)} \right) = 5.653 \times 0.521 = 2.943 \quad (3.31)$$

$$CL = \overline{S^2} = 1.413 \quad (3.32)$$

$$LDL = \left( \sum_{i=1}^I S_i^2 \right) \left( AD_{\min}^{(.025)} \right) = 5.653 \times 0.0716 = 0.405 \quad (3.33)$$

From Figure 5 one may see that  $S_4^2$  was above the upper decision line.

### 3.6 COMPUTER PROGRAM FOR THE RANDOMIZATION TESTS

The previous sections in this chapter outlined the steps to perform each approximate randomization test. A FORTRAN program that the reader may use to perform the tests will be discussed in this section. The program was used to determine the p-values and decision limits for the examples presented for each approximate randomization test. The program performs permutation shuffles, but it may be modified for bootstrapping shuffles.

```

integer numshuf,npops,nsamp,iseed
parameter (numshuf=1000, npops=4, nsamp=10, iseed=1579)
real alpha
parameter (alpha=0.05)
dimension semvar(npops), shufvar(npops), ddvarto(npops)
dimension rdratio(npops), ratiovar(npops)
dimension devvar(npops), devvarto(npops), sampavg(npops)
dimension randstol(numshuf), randstoh(numshuf)
dimension ddslo(numshuf), ddshi(numshuf)
dimension sampvar(npops), ratdev(npops*nsamp), dev(npops*nsamp)
dimension randsone(numshuf), ratioslo(numshuf), ratioshi(numshuf)
dimension rdslo(numshuf), rdshi(numshuf)
real e(npops,nsamp), devmean(npops,nsamp), c(npops,nsamp)
dimension pop(npops*nsamp)
data pop/.5119, .6137, -.6311, -.5025, 1.577, -1.32, -.3896,
+ .6738, -.09655, -.6112, .3756, .1026, -.2989, -.5242, -1.455,
+ -1.481, .5361, .6953, -1.243, -1.253, -1.648, .06582, .5893,
+ .04181, .2225, -.1234, -2.350, 1.269, -.5466, .1959, .2821, -2.411,
+ -4.359, -1.704, .9703, -.9169, .6221, -3.017, -.04799, -2.235/

```

```

npopsamp = nsamp*npops
halfalph = alpha/2.
ildl = numshuf*halfalph
iudl = numshuf - ildl + 1
donerjct = 0.
dtworjcl = 0.
dtworjch = 0.
ddrjcl = 0.
ddrjch = 0.
rrjcl = 0.
rrjch = 0.
rdrjcl = 0.
rdrjch = 0.

```

C      ASSIGNS THE DATA VALUES TO THEIR RESPECTIVE SAMPLE

```

jd = 1
do j = 1 , npops
  do k = 1 , nsamp
    dev(k) = pop(jd)
    jd = jd + 1
  enddo
enddo

```

C      FINDS MEAN AND VARIANCE FOR EACH SAMPLE

```

call ameanvar(dev,nsamp,savg,svar)
sampavg(j) = savg
sampvar(j) = svar

```

C      FINDS DEVIATION FROM MEAN FOR -DD AND -RD

```

do k = 1 , nsamp
  devmean(j,k) = dev(k) - sampavg(j)
enddo
enddo

```

```

avgvar = 0.
varsum = 0.

```

C      FINDS AVERAGE SAMPLE VARIANCE

```

do j = 1 , npops
  varsum = varsum + sampvar(j)
enddo
avgvar = varsum / float(npops)

```

C      FINDS DEVIATION FROM AVERAGE VARIANCE AND RATIO TO SUM

```

do j = 1 , npops
  devvar(j) = abs(sampvar(j) - avgvar)
  devvarcto(j) = sampvar(j) - avgvar
  ratiovar(j) = sampvar(j) / varsum
enddo

```

```

C      SORTS THE DEVIATIONS FROM AVERAGE VARIANCE AND RATIOS TO SUM
call  bsort(devvar,npops)
call  bsort(devvaro,npops)
call  bsort(ratiovar,npops)

C      ASSIGNS THE TEST STATISTICS FOR THE INITIAL DATA SET
randone = devvar(npops)
randtwo1 = devvaro(1)
randtwoh = devvaro(npops)
ratiolow = ratiovar(1)
ratiohi = ratiovar(npops)

C      PREPARES THE DEVIATIONS FROM MEAN SO THEY MAY BE SHUFFLED
md = 1
do j = 1 , npops
  do k = 1 , nsamp
    ratdev(md) = devmean(j,k)
    md = md + 1
  enddo
enddo

C      SHUFFLES THE ORIGINAL DATA AND THE DEVIATIONS FROM THE MEAN
C      SHUFFLES ARE PERFORMED WITHOUT REPLACEMENT
do i = 1 , numshuf
  do ij = 1 , npopsamp - 1
    rnn = float(npopsamp + 1 - ij)
    ih = int(rnn*ran1(iseed)) + ij
    tmp = pop(ij)
    tmpp = ratdev(ij)
    pop(ij) = pop(ih)
    ratdev(ij) = ratdev(ih)
    pop(ih) = tmp
    ratdev(ih) = tmpp
  enddo
enddo

C      ASSIGNS THE SHUFFLED DATA TO THEIR RESPECTIVE SHUFFLE SAMPLE
C      SQUARES THE DEVIATIONS FROM THE MEAN FOR -DD AND -RD
md = 1
do j = 1 , npops
  do k = 1 , nsamp
    devmean(j,k) = ratdev(md)
    c(j,k) = devmean(j,k)**2
    e(j,k) = pop(md)
    md = md + 1
  enddo
enddo

```

```

C     PERFORMING OPERATIONS ON SHUFFLED DATA
C     THIS IS IN PREPARATION FOR CALCULATING TEST STATISTICS
rn = float(nsamp)
do j = 1 , npops
  shsum = 0.
  shssum = 0.
  shsq = 0.
  do k = 1 , nsamp
    shsum = c(j,k) + shsum
    shssum = e(j,k) + shssum
    shsq = (e(j,k)**2) + shsq
  enddo
  semvar(j) = shsum / (rn - 1.)
  vars = shsq - (shssum*shssum/rn)
  shufvar(j) = vars / (rn - 1.)
enddo

savgvar = 0.
avssemvr = 0.
sumsem = 0.
sumshuf = 0.

C     FINDS AVERAGE VARIANCE AND AVERAGE SEMI-VARIANCE
do j = 1 , npops
  sumsem = semvar(j) + sumsem
  sumshuf = shufvar(j) + sumshuf
enddo
avssemvr = sumsem / float(npops)
savgvar = sumshuf / float(npops)

C     FINDS DEVIATIONS FROM AVERAGE AND RATIOS TO SUM
do j = 1 , npops
  devvar(j) = abs(shufvar(j) - savgvar)
  devvarto(j) = shufvar(j) - savgvar
  ddvarto(j) = semvar(j) - avssemvr
  ratiovar(j) = shufvar(j) / sumshuf
  rdratio(j) = semvar(j) / sumsem
enddo

C     SORTS THE DEVIATIONS FROM AVERAGE AND RATIOS TO SUM
call bubsort(devvar,npops)
call bubsort(devvarto,npops)
call bubsort(ddvarto,npops)
call bubsort(ratiovar,npops)
call bubsort(rdratio,npops)

C     ASSIGNS TEST STATISTICS FOR THE SHUFFLED DATA
randstone(i) = devvar(npops)
randstol(i) = devvarto(1)

```

```

    randstoh(i) = devvarto(npops)
    ddslo(i) = ddvarto(1)
    ddshi(i) = ddvarto(npops)
    ratioslo(i) = ratiovar(1)
    ratioshi(i) = ratiovar(npops)
    rdslo(i) = rdratio(1)
    rdshi(i) = rdratio(npops)

C     COMPARES TEST STATISTICS OF SHUFFLED DATA AND ORIGINAL DATA
C     ADDS ONE TO COUNTER IF ORIGINAL EXCEEDED BY SHUFFLED VALUE
    if (randstone(i).gt.randone) donerjct = donerjct + 1.
    if (randstol(i).lt.randtwol) dtworjcl = dtworjcl + 1.
    if (randstoh(i).gt.randtwoh) dtworjch = dtworjch + 1.
    if (ddslo(i).lt.randtwol) ddrjcl = ddrjcl + 1.
    if (ddshi(i).gt.randtwoh) ddrjch = ddrjch + 1.
    if (ratioslo(i).lt.ratiolow) rrjcl = rrjcl + 1.
    if (ratioshi(i).gt.ratiohi) rrjch = rrjch + 1.
    if (rdslo(i).lt.ratiolow) rdrjcl = rdrjcl + 1.
    if (rdshi(i).gt.ratiohi) rdrjch = rdrjch + 1.
enddo

C     SORTS SHUFFLED TEST STATISTICS AND FINDS DECISION LIMITS
call bubsort(randstol,numshuf)
call bubsort(randstoh,numshuf)
call bubsort(ddslo,numshuf)
call bubsort(ddshi,numshuf)
call bubsort(ratioslo,numshuf)
call bubsort(ratioshi,numshuf)
call bubsort(rdslo,numshuf)
call bubsort(rdshi,numshuf)

dldl = avgvar + randstol(ildl)
dudl = avgvar + randstoh(iudl)
ddldl = avgvar + ddslo(ildl)
ddudl = avgvar + ddshi(iudl)
rldl = varsum * ratioslo(ildl)
rudl = varsum * ratioshi(iudl)
rdldl = varsum * rdslo(ildl)
rdudl = varsum * rdshi(iudl)

C     CALCULATES P-VALUES
sn = float(numshuf) + 1.
pvone = (donerjct + 1.) / sn
pvtwol = (dtworjcl + 1.) / sn
pvtwoh = (dtworjch + 1.) / sn
pvddl = (ddrjcl + 1.) / sn
pvddh = (ddrjch + 1.) / sn
pvrl = (rrjcl + 1.) / sn
pvrrh = (rrjch + 1.) / sn
pvrdl = (rdrjcl + 1.) / sn
pvrrdh = (rdrjch + 1.) / sn

```



```

C      DETERMINES REJECT DECISION FOR EACH TEST
C      THE REJECT DECISION IS PRINTED AS OUTPUT ALONG WITH THE P-VALUE
C      THE DECISION LIMITS ARE ALSO PRINTED FOR THE APPLICABLE TESTS

```

```

C      THIS IS FOR RANDANOMV-D1
if (pvone.lt.alpha) then
  write(*,*) '-D1 rejects, p-value is' ,pvone
  print 5
5    format ('0')
  else
  write(*,*) '-D1 does not reject, p-value is' ,pvone
  print 15
15   format ('0')
endif

```

```

C      THIS IS FOR RANDANOMV-D
if (pvtwol.lt.halfalph) go to 100
if (pvtwoh.lt.halfalph) go to 100
go to 20
100 write(*,*) '-D rejects, p-values are' ,pvtwol,pvtwoh
write(*,*) 'LDL and UDL are' ,dldl,dudl
print 25
25   format ('0')
go to 30
20   write(*,*) '-D does not reject, p-values are' ,pvtwol,pvtwoh
write(*,*) 'LDL and UDL are' ,dldl,dudl
print 35
35   format ('0')

```

```

C      THIS IS FOR RANDANOMV-DD
30   if (pvddl.lt.halfalph) go to 200
if (pvddh.lt.halfalph) go to 200
go to 40
200 write(*,*) '-DD rejects, p-values are' ,pvddl,pvddh
write(*,*) 'LDL and UDL are' ,ddldl,ddudl
print 45
45   format ('0')
go to 50
40   write(*,*) '-DD does not reject, p-values are' ,pvddl,pvddh
write(*,*) 'LDL and UDL are' ,ddldl,ddudl
print 55
55   format ('0')

```

```

C      THIS IS FOR RANDANOMV-R
50   if (pvrl.lt.halfalph) go to 300
if (pvrrh.lt.halfalph) go to 300
go to 60
300 write(*,*) '-R rejects, p-values are' ,pvrl,pvrrh
write(*,*) 'LDL and UDL are' ,rldl,rudl
print 65

```

```

65  format ('0')
go to 70
60  write(*,*) '-R does not reject, p-values are' ,pvrl,pvrh
write(*,*) 'UDL and LDL are' ,rdldl,rudl
print 75
75  format ('0')

C    THIS IS FOR RANDANOMV-RD
70  if (pvrdl.lt.halfalph) go to 400
if (pvrdh.lt.halfalph) go to 400
go to 80
400 write(*,*) '-RD rejects, p-values are' ,pvrdl,pvrdh
write(*,*) 'LDL and UDL are' ,rdldl,rudl
print 85
85  format ('0')
go to 90
80  write(*,*) '-RD does not reject, p-values are' ,pvrdl,pvrdh
write(*,*) 'LDL and UDL are' ,rdldl,rudl

90  end

C    THIS SUBROUTINE CALCULATES THE SAMPLE MEAN AND VARIANCE
subroutine ameanvar(b,n,vmean,vvar)
dimension b(n)
sum = 0.
sq = 0.
do j = 1 , n
sum = sum + b(j)
sq = sq + (b(j)**2)
enddo
rn = float(n)
vmean = sum / rn
vvar = (sq - sum*sum / rn) / (rn - 1.)
return
end

C    THIS SUBROUTINE SORTS VALUES IN ASCENDING ORDER
subroutine bubsort(devs,n)
dimension devs(n)
do j = 1 , n-1
do k = 1 , n-j
if (devs(k).gt.devs(k+1)) then
tmp = devs(k)
devs(k) = devs(k+1)
devs(k+1) = tmp
endif
enddo
enddo
return
end

```

```

C      THIS FUNCTION PROVIDES RANDOM NUMBERS TO SHUFFLE DATA
C      THE FUNCTION WAS OBTAINED FROM NUMERICAL RECIPES
FUNCTION RAN1(iseed)
DIMENSION R(97)
PARAMETER (M1=259200, IA1=7141, IC1=54773, RM1=3.8580247E-6)
PARAMETER (M2=134456, IA2=8121, IC2=28411, RM2=7.4373773E-6)
PARAMETER (M3=243000, IA3=4561, IC3=51349)
DATA IFF /0/
IF (iseed.LT.0.OR.IFF.EQ.0) THEN
  IFF=1
  IX1=MOD(IC1-iseed,M1)
  IX1=MOD(IA1*IX1+IC1,M1)
  IX2=MOD(IX1,M2)
  IX1=MOD(IA1*IX1+IC1,M1)
  IX3=MOD(IX1,M3)
  DO 11 J=1,97
    IX1=MOD(IA1*IX1+IC1,M1)
    IX2=MOD(IA2*IX2+IC2,M2)
    R(J)=(FLOAT(IX1)+FLOAT(IX2)*RM2)*RM1
11  CONTINUE
  iseed=1
ENDIF
IX1=MOD(IA1*IX1+IC1,M1)
IX2=MOD(IA2*IX2+IC2,M2)
IX3=MOD(IA3*IX3+IC3,M3)
J=1+(97*IX3)/M3
IF (J.GT.97.OR.J.LT.1) PAUSE
RAN1=R(J)
R(J)=(FLOAT(IX1)+FLOAT(IX2)*RM2)*RM1
RETURN
END

```

The program allows the user to specify the number of shuffles (*numshuf*), the number of samples (*npops*), the number of elements in each sample (*nsamp*), the level of significance (*alpha*) and the seed (*iseed*) used to start the shuffle function. After these items have been declared, the user enters their data between the slashes (/). The data from Table 1 lies between the slashes. The program may then be compiled and executed. The program will output a reject/do not reject decision along with p-values. Upper and lower decision limits are produced so that a decision chart similar to those presented in sections 3.2 – 3.5 may be constructed.

Table 3: Outline of Randomization Tests

Test	Statistic (Initial)	What is shuffled?	Statistic (Shuffled)
-D1	$\max  S_i^2 - \bar{S}^2 $	Original data	$\max  S_i^2 - \bar{S}^2 $
-D	min and max $(S_i^2 - \bar{S}^2)$	Original data	min and max $(S_i^2 - \bar{S}^2)$
-DD	min and max $(S_i^2 - \bar{S}^2)$	Deviations from the sample mean, $z_{ij} = x_{ij} - \bar{x}_i$	min and max $(P_i^2 - \bar{P}^2)$ where $P_i^2 = \frac{\sum_{j=1}^{n_i} z_{ij}^2}{(n_i - 1)}$
-R	min and max $\left( \frac{S_i^2}{\sum_{i=1}^I S_i^2} \right)$	Original data	min and max $\left( \frac{S_i^2}{\sum_{i=1}^I S_i^2} \right)$
-RD	min and max $\left( \frac{S_i^2}{\sum_{i=1}^I S_i^2} \right)$	Deviations from the sample mean, $z_{ij} = x_{ij} - \bar{x}_i$	min and max $\left( \frac{P_i^2}{\sum_{i=1}^I P_i^2} \right)$ where $P_i^2 = \frac{\sum_{j=1}^{n_i} z_{ij}^2}{(n_i - 1)}$

Figure 2: RANDANOMV-D Decision Chart

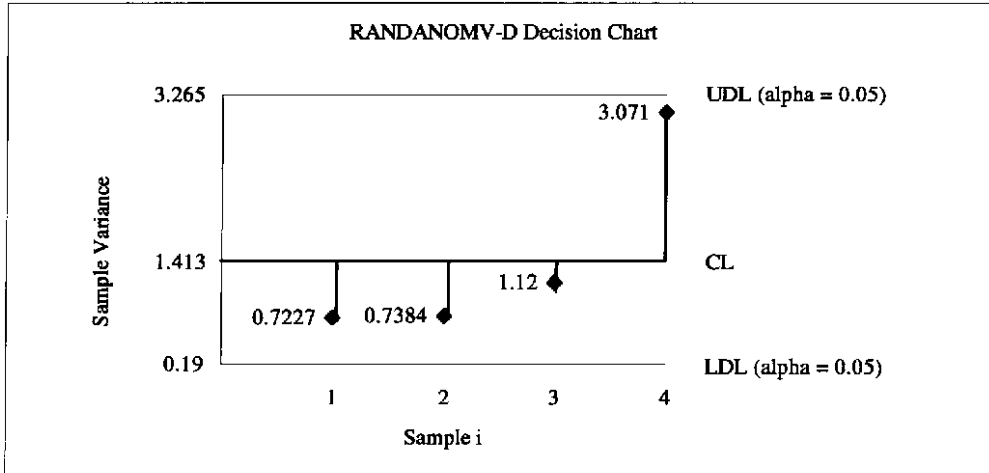


Figure 3: RANDANOMV-DD Decision Chart

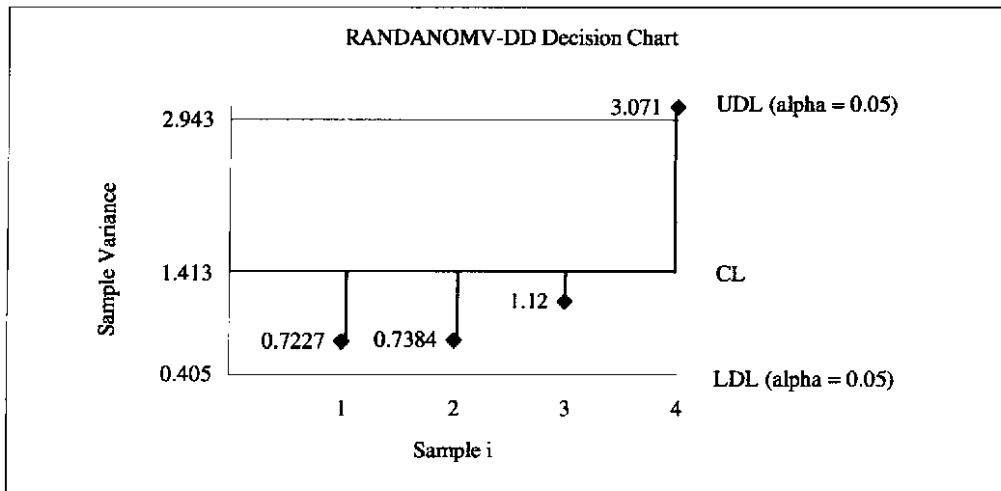


Figure 4: RANDANOMV-R Decision Chart

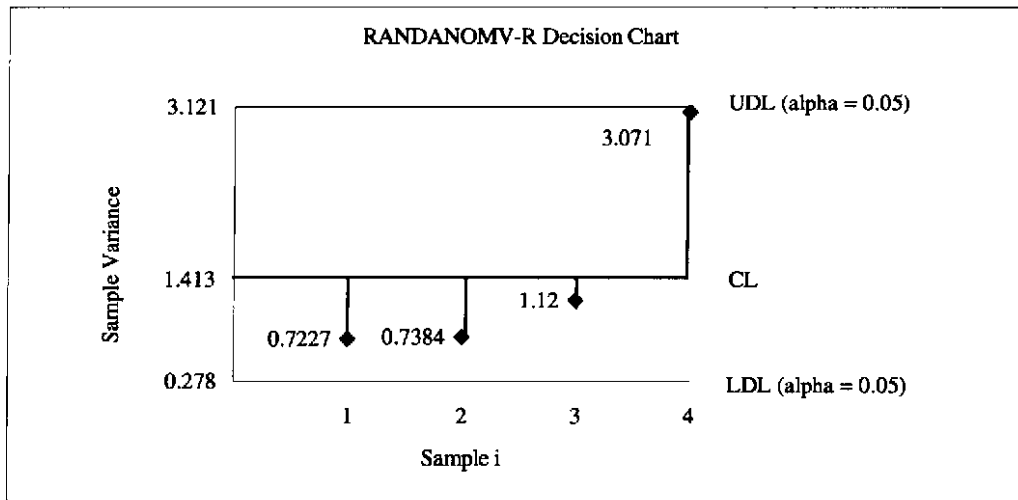
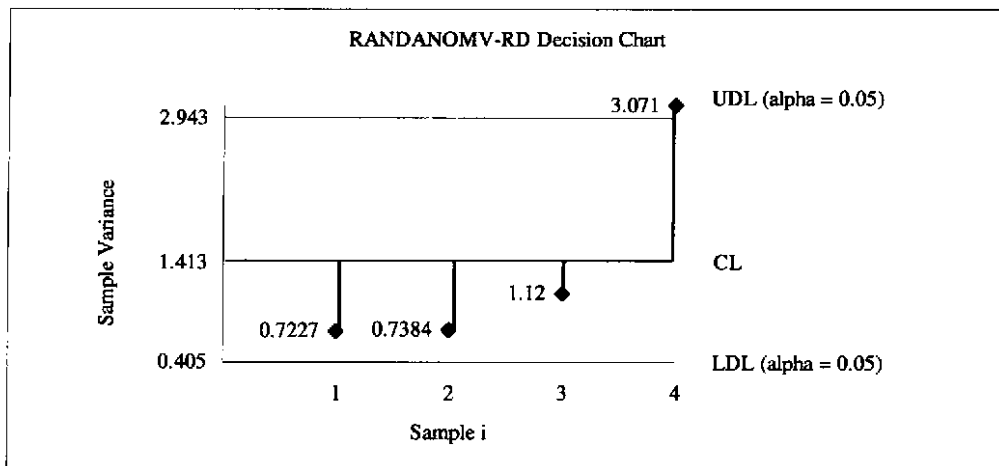


Figure 5: RANDANOMV-RD Decision Chart



Chapter 4  
Variance Configurations

Particularizing Section 1.2, suppose a researcher is interested in the variances of  $I$  populations, each with distribution function  $F_i(x_i, \sigma_i^2)$ . Let

$$\Omega = (0, \infty) \times (0, \infty) \times \dots \times (0, \infty)$$

be the parameter space for an HOV test. A variance configuration is a particular set of values for the  $I$  variances such that

$$\{\sigma_1, \dots, \sigma_I\} \in \Omega.$$

There are infinitely many variance configurations  $I$  populations may take. For that reason it is useful to partition the variance configuration into subspaces

$$\Omega_r = \left\{ (\sigma_1, \dots, \sigma_I) : \max_{i \neq j} \left( \frac{\sigma_i}{\sigma_j} \right) = r \right\}.$$

Note the HOV hypothesis (1.1) is true for all configurations  $(\sigma_1, \dots, \sigma_I)$  in  $\Omega_1$ .

**DEFINITION 10.** *The least favorable configuration (LFC<sub>r</sub>) for an HOV test is that configuration in  $\Omega_r$  with the lowest power.*

In this manner the LFC for the HOV tests is indexed by  $r$ . Different tests may have different LFCs for the same  $r$ .

Since ANOM and ANOVA are used in the nonrandomization tests presented in this paper, the LFC for these location tests will be presented. Let

$$M_{\Delta} = \left\{ \mu = \{\mu_1, \dots, \mu_I\} : \max_{i,j} |\mu_i - \mu_j| \geq \Delta\sigma \right\}$$

be subspaces of the parameter space of population means. Note that specifying a difference in terms of  $\Delta\sigma$  amounts to measuring differences of means in  $\sigma$  units. The LFC for the ANOVA-F test and ANOM (see Nelson (1998)) are of the form

$$\left( \frac{\Delta\sigma}{2}, -\frac{\Delta\sigma}{2}, 0, \dots, 0 \right) \quad (4.1)$$

That is, when  $I-2$  of the means are in the middle, and the other two means are equidistant above and below.

For the nonrandomization tests in this paper ANOVA and ANOM are applied to scale transformations  $y_{ij}$  to test the HOV hypothesis. These tests are actually comparing  $E(y_{ij})$  for the  $I$  populations. The LFC for these tests will be when the configuration of expected values  $(E(y_{1j}), \dots, E(y_{Ij}))$  is of the form of the ANOVA/ANOM LFC. That is, where one is large, one is small, and the rest are in the middle. Now a particular variance configuration  $(\sigma_1, \dots, \sigma_I)$  for the  $x_{ij}$ 's will induce a parameter configuration of  $(E(y_{1j}), \dots, E(y_{Ij}))$ . The configuration of  $\sigma$ 's that induces the LFC  $(E(y_{1j}), \dots, E(y_{Ij}))$  is the LFC for the variance test. This variance configuration should be of the form  $(1, m, \dots, m, r)$  since this will induce a configuration of means  $(E(y_{1j}), \dots, E(y_{Ij}))$ , which has the proper form. The value for  $m$  which produces the LFC likely depends on the underlying population (of  $x$ 's) and since that in general is not known the LFC is



indeterminate. There are complications involved since the transformed values do not in general meet the assumptions for ANOVA/ANOM tests and these assumptions are involved in the determination of the ANOVA/ANOM LFCs. However, the robustness of these tests is exploited for the HOV tests. Monte Carlo experimentation can be used to learn about HOV tests' LFCs for particular populations.

LFCs for the ANOM-type randomization tests presented in this paper again probably depend on the parent populations. Monte Carlo methods may be used to shed light on this problem. Intuition suggests these LFCs are of the form  $(1, m, \dots, m, r)$  since the randomization tests are modeled after the ANOM tests.

Chapter 5  
Monte Carlo Study

5.1 METHODOLOGY

A Monte Carlo study was carried out to evaluate each of the tests with respect to type I error stability and power. The type I error stability study assessed the relationship between observed rejection rates and the nominal rejection rates ( $\alpha$ ) when the HOV hypothesis (1.1) was true. The power study evaluated the ability of each test to detect differences among sample variances when the HOV hypothesis (1.1) was false.

FORTRAN programs were written to perform the Monte Carlo study. The program used to simulate the randomization tests is in the Appendix. The program had to be modified slightly to evaluate different numbers of populations and different sample sizes. The non-randomization tests used 10,000 replications, and the randomization tests used 2,000 replications with 1,000 shuffles per replication.

Since the focus of this paper was on robust tests, six of the seven distributions were nonnormal in the type I error stability study and power study. Three common and four special distributions were used in each study. The three common distributions were  $N(0,1)$ ,  $\chi^2(1)$  and an exponential distribution with  $\lambda = 1$ ,  $\text{Exp}(1)$ . The four special

distributions were (i) a symmetric distribution with kurtosis of 6 (generated using a method devised by Fleishman (1978) and employing tables from Barnes (1981)) to obtain

$$KUR(6) = 0.66268N(0,1) + 0.10189N^3(0,1) \quad (5.1)$$

(ii) a distribution with no kurtosis and skewness of 3 (generated using Fleishman (1978) and Barnes (1981)) to obtain

$$SKW(3) = -0.05134 - 2.91756N(0,1) + 0.05134N^2(0,1) + 0.87133N^3(0,1) \quad (5.2)$$

(iii) Gamma  $\left(\frac{4}{9}, 1\right)$  and (iv) a 50:50 mixture of two normals where one was  $N(-2,1)$ , and the other was  $N(2,1)$ .

The majority of the power study was conducted with the KUR(6) and SKW(3); however, the power of each test was evaluated for each of the distributions in the type I error stability study.

## 5.2 VARIANCE CONFIGURATIONS STUDIED

Two variance configurations were studied:

- Configuration 1, where  $I - 1$  variances were equal and the last was larger. This configuration was of the form  $(1, \dots, 1, r)$ .
- Configuration 2, where  $I - 2$  variances were equal, the first was smaller, and the last was larger. This configuration was of the form  $(1, m, \dots, m, r)$ . Two variations of Configuration 2 were studied. One arrangement was  $m = (0.5)r$ , and the other arrangement was  $m = (0.75)r$ . The latter variation of Configuration 2

was studied to see the performance of the tests on a configuration thought to produce lower power.

Configuration 1 was chosen because it was a common circumstance of interest, and it was a favorable LFC. Configuration 2 was selected because it was thought to be an unfavorable LFC.

### 5.3 TYPE I ERROR STABILITY

The type I error rates are given in the case where each of the  $I$  populations have the same variance. The ideal test will be robust across different underlying distributions. That is, a robust test will have a type I error rate that is consistent from distribution to distribution. Similar to Conover, Johnson and Johnson (1981), Wludyka and Nelson (1999) gave guidelines on assessing the degree of robustness. A test was deemed to have “good robustness” if its rejection rate was less than twice the nominal rate and “adequate robustness” if its rejection rate was less than three times the nominal rate. Tests may also be classified as either “conservative” or “liberal.” “Conservative” tests are those in which the empirical rejection rate is less than the nominal rate. “Liberal” tests are those in which the empirical rejection rate is greater than the nominal rate.

Tables 4 and 5 contain results from the Monte Carlo study performed to evaluate the empirical type I error stability. Table 4 holds information from the study conducted using five samples, and Table 5 holds similar information for 10 samples.

Levene and ANOMV-LEV were the only nonrandomization tests that displayed “good robustness” for each of the six nonnormal distributions, although ANOMV-LEV slightly exceeded the criterion for SKW(3). ANOMV-JK exhibited “good robustness” for the mixture, but the test was “inadequate” for SKW(3) and  $\chi^2(1)$ . ANOMV-TR demonstrated “good robustness” for all distributions except Exp(1) and  $\chi^2(1)$ . This test was “inadequate” for  $\chi^2(1)$ . F-K showed “good robustness” for parent distributions of KUR(6), SKW(3) and the mixture. The test was “inadequate” for  $\chi^2(1)$ . Levene’s test proved to be conservative for all parent distributions. None of the other nonrandomization tests were consistently conservative or liberal. ANOMV-LEV and ANOMV-JK leaned toward conservative and liberal, respectively. ANOMV-TR and ANOMV-FK jumped around the nominal rejection rate.

Two of the four randomization tests demonstrated “good robustness” for the nonnormal parent distributions. RANDANOMV-D and RANDANOMV-R were the two tests that were “good” for all distributions. RANDANOMV-DD and RANDANOMV-RD were “inadequate” for  $\chi^2(1)$ , Exp(1) and Gamma( $\frac{4}{9}, 1$ ) when permutation shuffling was used. RANDANOMV-DD exhibited “good” robustness when bootstrap shuffling was used. None of the four tests were consistently conservative or liberal; however, the tests were more conservative when bootstrap shuffling was used.

## 5.4 POWER

Tables 6 – 12 contain results from the power study. Tables 6 – 9 hold power information for KUR(6). Tables 10 and 11 contain power results for SKW(3), and Table 12 contains the results from a power study using one variance configuration for the other five distributions used in the type I error stability study.

Each of the tests had greater power for Configuration 1 than for variance Configuration 2. The power of each test increased with  $n$ . Among the randomization tests RANDANOMV-D and RANDANOMV-R appear to be the best. When looking at the nonrandomization tests, ANOMV-LEV was the best. There were no significant differences in power when comparing the nonrandomization group as a whole to the randomization group as a whole when permutation shuffling was used. Bootstrap shuffling had a negative impact on the power of the randomization tests. This type of shuffling had the greatest impact on RANDANOMV-D and RANDANOMV-DD while the other randomization tests were impacted minimally. There were instances where individual tests were inappropriate for the distribution under study.

## 5.5 UNEQUAL MEANS

It was noted in section 3.3 that many non-parametric tests of variability assume a common center of location (Baker 1995). For this reason RANDANOMV-DD was developed. The following example will illustrate the problem that arises with RANDANOMV-D when populations have unequal means.

Consider the sample below for which  $I = 3$ ,  $n_i = 2$ ,  $s_1^2 = 0.5$ ,  $s_2^2 = 0.5$ ,  $s_3^2 = 1250$  and  $\overline{s^2} = 417$ .

$$X = \begin{pmatrix} 2 & 20 & 100 \\ 3 & 21 & 150 \end{pmatrix}$$

From the initial sample  $AD_{\max} = 833$  and  $AD_{\min} = -416.5$ . There are six distinct permutations of the data that produce different test statistics. That is, the original sample produces the same test statistics as

$$X = \begin{pmatrix} 20 & 100 & 2 \\ 21 & 150 & 3 \end{pmatrix}.$$

Two such distinct permutations are

$$X^{(1)} = \begin{pmatrix} 2 & 3 & 100 \\ 20 & 21 & 150 \end{pmatrix} \text{ and } X^{(2)} = \begin{pmatrix} 2 & 20 & 21 \\ 3 & 100 & 150 \end{pmatrix}.$$

For the first permutation  $s_1^2 = 162$ ,  $s_2^2 = 162$ ,  $s_3^2 = 1250$ ,  $\overline{s^2} = 524.67$ ,  $AD_{\min}^1 = -362.67$  and  $AD_{\max}^1 = 725.33$ . The test statistics from the permuted data do not exceed the test statistics from the original sample. For the second permutation  $s_1^2 = 0.5$ ,  $s_2^2 = 3200$ ,  $s_3^2 = 8320.5$ ,  $\overline{s^2} = 3840.33$ ,  $AD_{\min}^2 = -3839.83$  and  $AD_{\max}^2 = 4480.17$ . The test statistics from the permuted data exceed the test statistics from the original sample.

This was done for the remaining four permutations. Four of the six distinct permutations produced test statistics that exceeded the initial test statistics. Thus, one would conclude that there was no difference in variability among the samples. This example illustrates a problem with RANDANOMV-D: this test may have difficulty detecting differences

among variances if the populations have very different means. Thus, caution should be exercised when using this test when the means are known to be very different.

A Monte Carlo study was carried out using an unequal means case. The study used a situation where the means were not “too different.” The type I error results are in Table 10. The results from the power study are in Table 11.

The example presented earlier in this section showed that large differences in the mean could lead to problems with RANDANOMV-D. The Monte Carlo results show that small differences in the mean may not have a big impact on RANDANOMV-D or any of the randomization tests.

From the Monte Carlo study it was seen that bootstrap shuffling is slightly more robust and less powerful than permutation shuffling. The small increase in robustness was offset by the decrease in power. It was also shown that small differences in the mean may not greatly impact the performance of the randomization tests, but care should be exercised when using these tests when the means greatly differ. Because of the previous two issues, the user should employ randomization tests with permutation shuffling when it is known that the means are not very different. When the randomization tests are used in these situations, RANDANOMV-D and RANDANOMV-R are the best, especially with larger sample sizes. These two tests control the type I error rate for all of the distributions in the study, and these tests are as robust as commonly used nonrandomization tests. With smaller sample sizes, RANDANOMV-D and



RANDANOMV-R are still robust tests, but the user sacrifices some power. The user should be hesitant about using RANDANOMV-DD and RANDANOMV-RD when the data is skewed as the Monte Carlo study shows inflated type I error rates for these tests.

Table 4: Empirical Type I Error Rates for  $I = 5, n = 10$  (20 in parentheses),  $\alpha = 0.05$  left of semi-colon,  $\alpha = 0.10$  right of semi-colon. \* Denotes Bootstrap Shuffling

Test	Distribution						
	N(0,1)	KUR(6)	SKW(3)	Chi-sq(1)	Exp(1)	Gamma(4/9,1)	50:50 Mixture
Levene	.029 (.033); .061 (.071)	.032 (.037); .073 (.083)	.015 (.017); .039 (.044)	.049 (.045); .095 (.094)	.046 (.049); .092 (.096)	.044 (.042); .088 (.082)	.018 (.014); .035 (.033)
ANOMV-LEV	.032 (.034); .069 (.074)	.044 (.047); .090 (.092)	.028 (.037); .066 (.083)	.063 (.055); .116 (.103)	.057 (.055); .103 (.100)	.056 (.046); .102 (.089)	.019 (.016); .042 (.036)
ANOMV-JK	.051 (.052); .088 (.091)	.099 (.100); .174 (.170)	.246 (.313); .344 (.430)	.136 (.122); .217 (.201)	.119 (.115); .195 (.185)	.121 (.110); .196 (.186)	.024 (.030); .043 (.060)
ANOMV-TR	.034 (.035); .072 (.073)	.041 (.039); .085 (.082)	.018 (.020); .044 (.048)	.154 (.180); .245 (.282)	.087 (.098); .145 (.171)	.081 (.087); .146 (.160)	.014 (.009); .032 (.025)
F-K	.031 (.033); .066 (.073)	.038 (.038); .080 (.081)	.017 (.018); .040 (.045)	.162 (.198); .254 (.305)	.087 (.106); .153 (.177)	.085 (.092); .149 (.164)	.012 (.009); .027 (.022)
RAND-D1	.057 (.052); .108 (.091)	.053 (.058); .099 (.091)	.046 (.066); .098 (.099)	.050 (.066); .090 (.110)	.054 (.048); .113 (.104)	.043 (.047); .096 (.089)	.050 (.046); .095 (.093)
RAND-D	.061 (.043); .108 (.083)	.059 (.046); .102 (.091)	.049 (.046); .089 (.113)	.058 (.064); .110 (.107)	.050 (.047); .101 (.093)	.048 (.050); .092 (.101)	.044 (.044); .090 (.090)
RAND-DD	.082 (.058); .145 (.099)	.114 (.072); .177 (.120)	.160 (.111); .231 (.165)	.410 (.272); .501 (.357)	.267 (.169); .367 (.235)	.258 (.207); .350 (.275)	.061 (.047); .107 (.082)
RAND-R	.060 (.043); .111 (.083)	.059 (.047); .107 (.090)	.047 (.051); .089 (.088)	.053 (.055); .102 (.102)	.054 (.042); .097 (.088)	.049 (.046); .096 (.103)	.044 (.046); .086 (.086)
RAND-RD	.082 (.058); .145 (.099)	.114 (.072); .177 (.120)	.160 (.111); .231 (.165)	.410 (.272); .501 (.357)	.267 (.169); .367 (.235)	.258 (.207); .350 (.275)	.061 (.047); .107 (.082)
*RAND-D1	.043; .094	.042; .103	.030; .094	.031; .105	.031; .105	.048; .108	.041; .091
*RAND-D	.026; .067	.018; .047	.015; .048	.012; .034	.012; .038	.018; .053	.030; .076
*RAND-DD	.027; .078	.021; .055	.025; .051	.038; .087	.045; .099	.050; .106	.046; .094
*RAND-R	.032; .071	.039; .074	.023; .065	.022; .058	.029; .064	.027; .072	.027; .055
*RAND-RD	.055; .112	.068; .147	.070; .160	.340; .442	.212; .322	.209; .302	.054; .098

57

Table 5: Empirical Type I Error Rates for  $I = 10, n = 10$  (20 in parentheses),  $\alpha = 0.05$  left of semi-colon,  $\alpha = 0.10$  right of semi-colon. \* Denotes Bootstrap Shuffling

Test	Distribution						
	N(0,1)	KUR(6)	SKW(3)	Chi-sq(1)	Exp(1)	Gamma(4/9,1)	50:50 Mixture
Levene	.022 (.034); .053 (.076)	.032 (.046); .071 (.093)	.013 (.023); .032 (.050)	.047 (.053); .094 (.105)	.050 (.051); .092 (.098)	.044 (.046); .084 (.090)	.009 (.007); .021 (.021)
ANOMV-LEV	.032 (.037); .061 (.075)	.074 (.070); .125 (.119)	.109 (.127); .187 (.202)	.096 (.091); .154 (.140)	.082 (.073); .134 (.123)	.080 (.071); .130 (.114)	.013 (.016); .030 (.033)
ANOMV-JK	.051 (.054); .086 (.101)	.128 (.137); .210 (.228)	.539 (.623); .627 (.718)	.157 (.155); .246 (.239)	.141 (.131); .227 (.214)	.142 (.135); .227 (.219)	.035 (.035); .050 (.065)
ANOMV-TR	.032 (.036); .061 (.075)	.042 (.043); .081 (.086)	.020 (.023); .042 (.050)	.180 (.236); .281 (.359)	.102 (.113); .167 (.196)	.092 (.105); .155 (.190)	.009 (.010); .023 (.025)
F-K	.024 (.034); .053 (.076)	.036 (.042); .074 (.088)	.011 (.014); .026 (.037)	.227 (.329); .338 (.454)	.104 (.136); .178 (.219)	.098 (.131); .168 (.217)	.005 (.005); .013 (.015)
RAND-D1	.031 (.046); .082 (.096)	.041 (.042); .086 (.103)	.063 (.046); .114 (.092)	.052 (.054); .113 (.101)	.051 (.068); .092 (.126)	.031 (.043); .076 (.087)	.051 (.044); .106 (.092)
RAND-D	.019 (.052); .062 (.103)	.028 (.048); .068 (.096)	.054 (.050); .110 (.094)	.030 (.055); .088 (.108)	.052 (.050); .092 (.093)	.042 (.044); .085 (.096)	.050 (.043); .106 (.091)
RAND-DD	.043 (.065); .092 (.122)	.072 (.084); .136 (.144)	.199 (.109); .263 (.173)	.500 (.380); .619 (.460)	.311 (.222); .423 (.306)	.301 (.242); .413 (.329)	.090 (.042); .135 (.093)
RAND-R	.024 (.058); .069 (.102)	.027 (.052); .070 (.101)	.058 (.049); .114 (.095)	.030 (.060); .076 (.106)	.046 (.056); .088 (.094)	.033 (.046); .074 (.093)	.052 (.044); .103 (.089)
RAND-RD	.043 (.065); .092 (.122)	.072 (.084); .136 (.144)	.199 (.109); .263 (.173)	.500 (.380); .619 (.460)	.311 (.222); .423 (.306)	.301 (.242); .413 (.329)	.090 (.042); .135 (.093)
*RAND-D1	.053; .113	.043; .107	.035; .108	.030; .115	.036; .106	.039; .110	.043; .096
*RAND-D	.029; .071	.018; .045	.013; .045	.004; .031	.014; .036	.014; .039	.039; .078
*RAND-DD	.037; .091	.023; .072	.018; .151	.025; .084	.031; .090	.033; .083	.052; .099
*RAND-R	.044; .096	.043; .087	.032; .075	.019; .060	.034; .077	.032; .070	.039; .076
*RAND-RD	.071; .135	.091; .180	.102; .213	.453; .569	.258; .373	.264; .385	.066; .105

Table 6: Empirical Rejection Rates – Configuration 1 for  $l = 5, n = 10$  (20 in parentheses),  $\alpha = 0.05$  left of semi-colon,  $\alpha = 0.10$  right of semi-colon, KUR(6) Distribution. \* Denotes Bootstrap Shuffling

Test	r = 1	r = 3	r = 4	r = 5	r = 6	r = 9	r = 16
Levene	.032 (.037); .073 (.083)	.149 (.363); .244 (.482)	.242 (.555); .360 (.674)	.334 (.702); .459 (.799)	.413 (.800); .544 (.874)	.596 (.938); .716 (.966)	.805 (.993); .885 (.997)
ANOMV-LEV	.044 (.047); .090 (.092)	.196 (.416); .288 (.518)	.313 (.619); .421 (.710)	.422 (.761); .525 (.829)	.508 (.849); .615 (.899)	.700 (.958); .780 (.975)	.886 (.997); .929 (.998)
ANOMV-JK	.099 (.100); .174 (.170)	.202 (.329); .299 (.435)	.265 (.457); .376 (.564)	.331 (.564); .444 (.661)	.390 (.645); .502 (.733)	.520 (.790); .629 (.851)	.698 (.917); .780 (.948)
ANOMV-TR	.041 (.039); .085 (.082)	.167 (.369); .252 (.478)	.261 (.567); .364 (.666)	.352 (.710); .461 (.792)	.432 (.807); .542 (.873)	.607 (.941); .712 (.966)	.813 (.993); .877 (.997)
F-K	.038 (.038); .080 (.081)	.151 (.333); .238 (.456)	.232 (.517); .341 (.638)	.311 (.658); .427 (.763)	.380 (.759); .501 (.845)	.546 (.914); .663 (.953)	.749 (.987); .839 (.995)
RAND-D1	.053 (.058); .099 (.091)	.183 (.389); .274 (.486)	.281 (.569); .373 (.676)	.365 (.697); .469 (.787)	.438 (.808); .551 (.850)	.605 (.903); .694 (.940)	.800 (.992); .864 (.997)
RAND-D	.059 (.046); .102 (.091)	.154 (.320); .245 (.428)	.231 (.502); .333 (.610)	.302 (.632); .410 (.723)	.369 (.746); .479 (.830)	.533 (.878); .632 (.915)	.724 (.983); .810 (.992)
RAND-DD	.114 (.072); .177 (.120)	.256 (.384); .353 (.500)	.353 (.577); .474 (.668)	.455 (.703); .555 (.791)	.522 (.798); .632 (.868)	.681 (.911); .767 (.950)	.859 (.994); .911 (.996)
RAND-R	.059 (.047); .107 (.090)	.166 (.323); .251 (.439)	.244 (.511); .357 (.620)	.328 (.653); .437 (.730)	.395 (.752); .516 (.837)	.579 (.887); .677 (.921)	.785 (.988); .860 (.994)
RAND-RD	.114 (.072); .177 (.120)	.256 (.384); .353 (.500)	.353 (.577); .474 (.668)	.455 (.703); .555 (.791)	.522 (.798); .632 (.868)	.681 (.911); .767 (.950)	.859 (.994); .911 (.996)
*RAND-D1	.042; .103	.160; .286	.235; .392	.313; .470	.372; .547	.494; .701	.654; .854
*RAND-D	.018; .047	.087; .167	.148; .240	.206; .320	.251; .377	.354; .498	.494; .657
*RAND-DD	.021; .055	.097; .177	.153; .250	.201; .316	.239; .377	.350; .498	.492; .638
*RAND-R	.039; .074	.131; .231	.216; .332	.287; .414	.353; .488	.513; .642	.728; .835
*RAND-RD	.068; .147	.214; .337	.307; .439	.384; .527	.458; .602	.613; .744	.814; .905

58

Table 7: Empirical Rejection Rates – Configuration 1 for  $l = 10, n = 10$  (20 in parentheses),  $\alpha = 0.05$  left of semi-colon,  $\alpha = 0.10$  right of semi-colon, KUR(6) Distribution. \* Denotes Bootstrap Shuffling

Test	r = 1	r = 3	r = 4	r = 5	r = 6	r = 9	r = 16
Levene	.032 (.046); .071 (.093)	.145 (.352); .222 (.458)	.238 (.555); .337 (.653)	.336 (.707); .444 (.785)	.425 (.809); .531 (.869)	.622 (.947); .718 (.967)	.841 (.996); .893 (.998)
ANOMV-LEV	.074 (.070); .125 (.119)	.252 (.457); .326 (.539)	.383 (.669); .459 (.735)	.497 (.806); .573 (.852)	.591 (.887); .665 (.918)	.779 (.976); .828 (.984)	.929 (.999); .950 (.999)
ANOMV-JK	.128 (.137); .210 (.228)	.200 (.308); .293 (.415)	.251 (.426); .352 (.538)	.306 (.531); .411 (.639)	.356 (.621); .463 (.722)	.488 (.792); .595 (.862)	.684 (.938); .770 (.965)
ANOMV-TR	.042 (.043); .081 (.086)	.170 (.385); .244 (.476)	.278 (.590); .359 (.678)	.373 (.738); .466 (.806)	.462 (.835); .555 (.885)	.657 (.956); .732 (.971)	.852 (.996); .892 (.998)
F-K	.036 (.042); .074 (.088)	.123 (.303); .202 (.411)	.196 (.485); .296 (.591)	.274 (.628); .383 (.724)	.348 (.733); .455 (.816)	.509 (.905); .626 (.941)	.735 (.985); .815 (.993)
RAND-D1	.041 (.042); .086 (.103)	.147 (.296); .227 (.387)	.255 (.477); .347 (.578)	.356 (.621); .474 (.697)	.451 (.715); .544 (.788)	.636 (.883); .708 (.926)	.804 (.980); .885 (.988)
RAND-D	.028 (.048); .068 (.096)	.121 (.263); .188 (.352)	.214 (.426); .297 (.524)	.292 (.573); .391 (.658)	.383 (.677); .472 (.745)	.555 (.854); .647 (.893)	.759 (.972); .810 (.983)
RAND-DD	.072 (.084); .136 (.144)	.189 (.322); .295 (.427)	.299 (.484); .413 (.597)	.404 (.631); .506 (.720)	.493 (.728); .599 (.802)	.672 (.894); .758 (.927)	.890 (.985); .920 (.990)
RAND-R	.027 (.052); .070 (.101)	.117 (.266); .186 (.356)	.227 (.426); .302 (.528)	.301 (.582); .400 (.666)	.393 (.687); .496 (.755)	.575 (.860); .679 (.900)	.812 (.976); .855 (.987)
RAND-RD	.072 (.084); .136 (.144)	.189 (.322); .295 (.427)	.299 (.484); .413 (.597)	.404 (.631); .506 (.720)	.493 (.728); .599 (.802)	.672 (.894); .758 (.927)	.890 (.985); .920 (.990)
*RAND-D1	.043; .107	.143; .247	.237; .359	.328; .457	.414; .533	.573; .703	.767; .865
*RAND-D	.018; .045	.079; .144	.144; .238	.211; .329	.281; .414	.418; .573	.590; .767
*RAND-DD	.023; .072	.087; .187	.143; .292	.218; .391	.283; .469	.410; .642	.573; .847
*RAND-R	.043; .087	.122; .205	.194; .296	.281; .397	.363; .477	.531; .646	.759; .835
*RAND-RD	.091; .180	.191; .311	.272; .412	.371; .516	.456; .595	.636; .749	.839; .910

Table 8: Empirical Rejection Rates – Configuration 2 for  $I = 5, n = 10$  (20 in parentheses),  $\alpha = 0.05$  left of semi-colon,  $\alpha = 0.10$  right of semi-colon, KUR(6) Distribution. \* Denotes Bootstrap Shuffling

Test	$m = 2.5, r = 5$	$m = 3.5, r = 7$	$m = 12, r = 16$	$m = 18.75, r = 25$	$m = 36.75, r = 49$
Levene	.132 (.346); .230 (.497)	.168 (.465); .289 (.628)	.265 (.788); .455 (.916)	.333 (.891); .545 (.969)	.429 (.961); .663 (.993)
ANOMV-LEV	.150 (.330); .243 (.468)	.175 (.439); .286 (.607)	.250 (.855); .463 (.951)	.337 (.945); .577 (.986)	.463 (.986); .727 (.998)
ANOMV-JK	.242 (.398); .349 (.516)	.319 (.544); .435 (.652)	.663 (.900); .755 (.932)	.778 (.953); .849 (.969)	.891 (.985); .928 (.989)
ANOMV-TR	.144 (.356); .246 (.501)	.192 (.528); .324 (.676)	.487 (.967); .686 (.987)	.670 (.996); .831 (.999)	.875 (1.00); .956 (1.00)
F-K	.150 (.372); .252 (.521)	.206 (.536); .331 (.684)	.447 (.940); .636 (.979)	.595 (.989); .777 (.997)	.796 (.999); .918 (1.00)
RAND-D1	.133 (.257); .210 (.376)	.150 (.281); .231 (.396)	.114 (.172); .200 (.365)	.117 (.192); .210 (.404)	.120 (.229); .217 (.442)
RAND-D	.154 (.329); .251 (.456)	.210 (.468); .316 (.609)	.508 (.927); .632 (.962)	.623 (.982); .740 (.998)	.741 (.997); .825 (.999)
RAND-DD	.285 (.386); .401 (.527)	.374 (.579); .492 (.693)	.773 (.969); .852 (.986)	.895 (.995); .938 (.998)	.977 (.998); .989 (1.00)
RAND-R	.170 (.329); .263 (.463)	.242 (.494); .351 (.623)	.652 (.946); .769 (.974)	.819 (.995); .884 (.995)	.948 (.998); .978 (.999)
RAND-RD	.285 (.386); .401 (.527)	.374 (.579); .492 (.693)	.773 (.969); .852 (.986)	.895 (.995); .938 (.998)	.977 (.998); .989 (1.00)
*RAND-D1	.119; .229	.129; .242	.101; .220	.104; .226	.105; .233
*RAND-D	.062; .131	.072; .149	.082; .193	.100; .223	.120; .259
*RAND-DD	.072; .143	.082; .163	.098; .236	.119; .284	.145; .328
*RAND-R	.130; .251	.200; .342	.573; .723	.767; .864	.926; .966
*RAND-RD	.227; .369	.321; .480	.716; .826	.857; .919	.965; .986

Table 9: Empirical Rejection Rates – Configuration 2 for  $I = 10, n = 10$  (20 in parentheses),  $\alpha = 0.05$  left of semi-colon,  $\alpha = 0.10$  right of semi-colon, KUR(6) Distribution. \* Denotes Bootstrap Shuffling

Test	$m = 2.5, r = 5$	$m = 3.5, r = 7$	$m = 12, r = 16$	$m = 18.75, r = 25$	$m = 36.75, r = 49$
Levene	.102 (.263); .178 (.385)	.120 (.341); .211 (.478)	.155 (.558); .281 (.739)	.185 (.685); .336 (.843)	.230 (.815); .400 (.931)
ANOMV-LEV	.162 (.270); .233 (.377)	.172 (.317); .249 (.461)	.151 (.755); .278 (.904)	.180 (.907); .347 (.975)	.245 (.983); .480 (.998)
ANOMV-JK	.208 (.334); .316 (.461)	.268 (.476); .384 (.594)	.640 (.907); .735 (.936)	.773 (.961); .841 (.973)	.903 (.989); .936 (.991)
ANOMV-TR	.111 (.255); .181 (.376)	.130 (.380); .219 (.529)	.333 (.953); .530 (.983)	.522 (.995); .727 (.999)	.804 (1.00); .925 (1.00)
F-K	.106 (.275); .179 (.400)	.135 (.395); .231 (.532)	.268 (.823); .434 (.916)	.360 (.940); .549 (.980)	.514 (.997); .717 (1.00)
RAND-D1	.083 (.139); .146 (.234)	.086 (.146); .133 (.247)	.068 (.083); .121 (.158)	.068 (.084); .123 (.160)	.070 (.085); .125 (.164)
RAND-D	.096 (.234); .161 (.342)	.140 (.395); .237 (.514)	.455 (.925); .574 (.958)	.604 (.984); .718 (.989)	.736 (.998); .852 (1.00)
RAND-DD	.173 (.314); .291 (.430)	.258 (.487); .379 (.600)	.739 (.958); .839 (.974)	.897 (.991); .937 (.994)	.987 (.999); .997 (1.00)
RAND-R	.104 (.241); .171 (.345)	.163 (.407); .243 (.527)	.590 (.944); .725 (.963)	.800 (.988); .886 (.991)	.963 (.999); .979 (1.00)
RAND-RD	.173 (.314); .291 (.430)	.258 (.487); .379 (.600)	.739 (.958); .839 (.974)	.897 (.991); .937 (.994)	.987 (.999); .997 (1.00)
*RAND-D1	.082; .181	.088; .187	.066; .144	.067; .145	.067; .146
*RAND-D	.041; .084	.042; .092	.032; .103	.039; .123	.043; .153
*RAND-DD	.056; .123	.058; .131	.053; .168	.064; .205	.082; .249
*RAND-R	.101; .181	.146; .254	.569; .684	.761; .852	.948; .974
*RAND-RD	.188; .311	.264; .394	.712; .808	.860; .925	.978; .988

Table 10: Empirical Rejection Rates – Configuration 2 for  $I = 5, n = 10$  (20 in parentheses),  $\alpha = 0.05$ , SKW(3) Distribution. \* Denotes Bootstrap Shuffling. \*\*\* Denotes Test is Inappropriate for SKW(3).

Test	m = 6.75, r = 9	m = 12, r = 16	m = 18.75, r = 25	m = 27, r = 36	m = 36.75, r = 49
Levene	.259 (.664)	.397 (.789)	.487 (.846)	.551 (.874)	.596 (.887)
ANOMV-LEV	.302 (.704)	.471 (.819)	.574 (.864)	.636 (.890)	.675 (.904)
ANOMV-JK	***	***	***	***	***
ANOMV-TR	.443 (.955)	.683 (.994)	.824 (.999)	.900 (.999)	.942 (1.00)
F-K	.379 (.926)	.590 (.987)	.727 (.997)	.824 (.999)	.881 (.999)
RAND-D1	.357 (.366)	.415 (.407)	.437 (.436)	.456 (.461)	.463 (.469)
RAND-D	.734 (.874)	.826 (.919)	.860 (.953)	.888 (.979)	.904 (.987)
RAND-DD	***	***	***	***	***
RAND-R	.833 (.880)	.903 (.928)	.934 (.962)	.954 (.980)	.970 (.990)
RAND-RD	***	***	***	***	***
*RAND-D1	.312 (.270)	.367 (.304)	.387 (.321)	.402 (.330)	.411 (.335)
*RAND-D	.370 (.363)	.429 (.411)	.455 (.433)	.471 (.451)	.486 (.463)
*RAND-DD	.381 (.373)	.439 (.420)	.469 (.442)	.484 (.458)	.497 (.471)
*RAND-R	.758 (.843)	.877 (.914)	.920 (.946)	.942 (.964)	.955 (.978)
*RAND-RD	.807 (.852)	.890 (.920)	.929 (.952)	.948 (.965)	.960 (.980)

Table 11: Empirical Rejection Rates – Configuration 2 for  $I = 10, n = 10$  (20 in parentheses),  $\alpha = 0.05$ , SKW(3) Distribution. \* Denotes Bootstrap Shuffling. \*\*\* Denotes Test is Inappropriate for SKW(3).

Test	m = 6.75, r = 9	m = 12, r = 16	m = 18.75, r = 25	m = 27, r = 36	m = 36.75, r = 49
Levene	.098 (.427)	.158 (.591)	.208 (.676)	.248 (.728)	.278 (.761)
ANOMV-LEV	.241 (.604)	.356 (.760)	.444 (.830)	.507 (.866)	.553 (.889)
ANOMV-JK	***	***	***	***	***
ANOMV-TR	.252 (.939)	.550 (.993)	.753 (.999)	.860 (.999)	.924 (1.00)
F-K	.155 (.807)	.270 (.948)	.368 (.985)	.459 (.997)	.535 (.999)
RAND-D1	.152 (.113)	.171 (.125)	.182 (.132)	.194 (.136)	.200 (.138)
RAND-D	.722 (.870)	.848 (.931)	.894 (.963)	.919 (.978)	.930 (.986)
RAND-DD	***	***	***	***	***
RAND-R	.837 (.877)	.908 (.933)	.944 (.967)	.955 (.977)	.962 (.987)
RAND-RD	***	***	***	***	***
*RAND-D1	.118 (.090)	.144 (.103)	.158 (.108)	.164 (.110)	.167 (.112)
*RAND-D	.207 (.264)	.262 (.308)	.287 (.330)	.300 (.345)	.315 (.352)
*RAND-DD	.240 (.278)	.288 (.322)	.319 (.348)	.333 (.369)	.344 (.375)
*RAND-R	.726 (.843)	.879 (.912)	.916 (.955)	.946 (.973)	.961 (.983)
*RAND-RD	.828 (.853)	.908 (.919)	.940 (.958)	.961 (.974)	.974 (.985)

Table 12: Empirical Rejection Rates – Configuration 2 for  $m = 27, r = 36, I = 5, n = 10$  (20 in parentheses),  $\alpha = 0.05$ . \* Denotes Bootstrap Shuffling. \*\*\* Denotes Test is Inappropriate for Distribution.

Test	N(0,1)	chi-sq(1)	exp(1)	Gamma(4/9,1)	50:50 Mixture
Levene	.703 (.100)	.192 (.504)	.318 (.845)	.319 (.840)	.789 (1.00)
ANOMV-LEV	.823 (1.00)	.178 (.526)	.307 (.914)	.316 (.914)	.910 (1.00)
ANOMV-JK	.993 (1.00)	***	***	***	.998 (1.00)
ANOMV-TR	.901 (1.00)	***	.802 (.999)	.796 (.999)	.921 (1.00)
F-K	.808 (1.00)	***	.752 (.996)	.750 (.998)	.831 (1.00)
RAND-D1	.206 (.890)	.069 (.122)	.065 (.122)	.060 (.094)	.885 (1.00)
RAND-D	.912 (1.00)	.139 (.630)	.186 (.869)	.171 (.839)	.993 (1.00)
RAND-DD	.997 (1.00)	***	***	***	1.00 (1.00)
RAND-R	.997 (1.00)	.465 (.887)	.826 (.997)	.840 (.995)	1.00 (1.00)
RAND-RD	.997 (1.00)	***	***	***	1.00 (1.00)
*RAND-D1	.213 (.858)	.036 (.051)	.040 (.066)	.045 (.075)	.882 (1.00)
*RAND-D	.535 (.997)	.014 (.036)	.017 (.118)	.021 (.107)	.990 (1.00)
*RAND-DD	.590 (.997)	.090 (.183)	.166 (.454)	.177 (.446)	1.00 (1.00)
*RAND-R	.998 (1.00)	.355 (.854)	.749 (.988)	.760 (.990)	1.00 (1.00)
*RAND-RD	.999 (1.00)	***	***	***	1.00 (1.00)

Table 13: Empirical Type I Error Rates – Unequal Means,  $I = 5, n = 10$  (20 in parentheses),  $\alpha = 0.05$ , Mean Configuration = (1,1.2,1.4,1.6,1.8), Permutation Shuffling

Test	Distribution		
	N(0,1)	KUR(6)	SKW(3)
RAND-D1	.040 (.040)	.050 (.050)	.045 (.045)
RAND-D	.032 (.027)	.037 (.035)	.043 (.041)
RAND-DD	.082 (.059)	.112 (.069)	.161 (.121)
RAND-R	.059 (.043)	.086 (.070)	.050 (.054)
RAND-RD	.082 (.059)	.112 (.069)	.161 (.121)

Table 14: Empirical Rejection Rates – Unequal Means,  $I = 5, n = 10$  (20 in parentheses),  $\alpha = 0.05$ , Mean Configuration = (1,1.2,1.4,1.6,1.8), KUR(6) Distribution, Permutation Shuffling

Test	Configuration		
	1, r=9	1, r=16	2, m=18.75 r=25
RAND-D1	.580 (.895)	.777 (.990)	.118 (.202)
RAND-D	.508 (.872)	.704 (.979)	.609 (.972)
RAND-DD	.681 (.911)	.859 (.994)	.891 (.995)
RAND-R	.604 (.901)	.806 (.989)	.823 (.994)
RAND-RD	.681 (.911)	.859 (.994)	.891 (.995)

## Chapter 6

### Equivalence of Two Proposed Randomization Tests for Variances

In the Monte Carlo study the randomization tests RANDANOMV-DD and RANDANOMV-RD exhibited identical type I error stability and power for permutation shuffles. The following lemma will be used to prove equivalence of RANDANOMV-DD and RANDANOMV-RD for permutation shuffles. DEFINITION 9 stated the equivalence of two randomization tests.

LEMMA 1. Consider the case where  $X_{ij}$  are  $I$  samples of size  $n$ . Let  $\bar{X}_i$  be the mean of the  $i^{\text{th}}$  group and  $\bar{X}$  be the grand mean. Suppose the observations are (randomly) shuffled.

Let  $\bar{X}_i^*$  and  $\bar{X}^*$  be the group and grand mean for the permuted data.

$\max(\bar{X}_i^*, i=1, \dots, I) - \bar{X}^* > \max(\bar{X}_i, i=1, \dots, I) - \bar{X}$  if and only if

$$\frac{\max(\bar{X}_i^*, i=1, \dots, I)}{\sum_{i=1}^I \bar{X}_i^*} > \frac{\max(\bar{X}_i, i=1, \dots, I)}{\sum_{i=1}^I \bar{X}_i}.$$

The lemma's proof is simplified since the observations are randomly permuted without replacement.

$$\text{Thus, } \overline{\overline{X}}^* = \overline{\overline{X}} \Rightarrow \frac{\sum_{i=1}^I \overline{X_i}^*}{I} = \frac{\sum \overline{X_i}}{I} \Rightarrow \sum_{i=1}^I \overline{X_i}^* = \sum_{i=1}^I \overline{X_i}.$$

$$\max(\overline{X_i}^*, i=1, \dots, I) - \overline{\overline{X}}^* > \max(\overline{X_i}, i=1, \dots, I) - \overline{\overline{X}}$$

$$\text{Since } \overline{\overline{X}}^* = \overline{\overline{X}} \Rightarrow \max(\overline{X_i}^*, i=1, \dots, I) > \max(\overline{X_i}, i=1, \dots, I).$$

$$\text{Since } \sum_{i=1}^I \overline{X_i}^* = \sum_{i=1}^I \overline{X_i} \Rightarrow \frac{\max(\overline{X_i}^*, i=1, \dots, I)}{\sum_{i=1}^I \overline{X_i}^*} > \frac{\max(\overline{X_i}, i=1, \dots, I)}{\sum_{i=1}^I \overline{X_i}}.$$

Proving the lemma in the other direction one gets

$$\frac{\max(\overline{X_i}^*, i=1, \dots, I)}{\sum_{i=1}^I \overline{X_i}^*} > \frac{\max(\overline{X_i}, i=1, \dots, I)}{\sum_{i=1}^I \overline{X_i}}.$$

$$\text{Since } \sum_{i=1}^I \overline{X_i}^* = \sum_{i=1}^I \overline{X_i} \Rightarrow \max(\overline{X_i}^*, i=1, \dots, I) > \max(\overline{X_i}, i=1, \dots, I).$$

$$\text{Since } \overline{\overline{X}}^* = \overline{\overline{X}} \Rightarrow \max(\overline{X_i}^*, i=1, \dots, I) - \overline{\overline{X}}^* > \max(\overline{X_i}, i=1, \dots, I) - \overline{\overline{X}}.$$

With this result the two tests may be shown equivalent.

$$\text{Let } z_{ij} = x_{ij} - \overline{x}_i, p_i^2 = \frac{\sum_{j=1}^{n_i} z_{ij}^2}{(n_i - 1)} \text{ and } \overline{p^2} = \frac{\sum_{i=1}^I p_i^2}{I}.$$

Since the  $z_{ij}$  are randomly permuted without replacement

$$\overline{p^2}^* = \overline{p^2} \Rightarrow \frac{\sum_{i=1}^I p_i^{2*}}{I} = \frac{\sum_{i=1}^I p_i^2}{I} \Rightarrow \sum_{i=1}^I p_i^{2*} = \sum_{i=1}^I p_i^2.$$

$$\max(p_i^{2*}, i=1, \dots, I) - \overline{p^2}^* > \max(p_i^2, i=1, \dots, I) - \overline{p^2}$$



Since  $\overline{p^{2^*}} = \overline{p^2} \Rightarrow \max(p_i^{2^*}, i=1, \dots, I) > \max(p_i^2, i=1, \dots, I)$ .

Since  $\sum_{i=1}^I p_i^{2^*} = \sum_{i=1}^I p_i^2 \Rightarrow \frac{\max(p_i^{2^*}, i=1, \dots, I)}{\sum_{i=1}^I p_i^{2^*}} > \frac{\max(p_i^2, i=1, \dots, I)}{\sum_{i=1}^I p_i^2}$ .

Using the lemma in the other direction one starts with

$\frac{\max(p_i^{2^*}, i=1, \dots, I)}{\sum_{i=1}^I p_i^{2^*}} > \frac{\max(p_i^2, i=1, \dots, I)}{\sum_{i=1}^I p_i^2}$ .

Since  $\sum_{i=1}^I p_i^{2^*} = \sum_{i=1}^I p_i^2 \Rightarrow \max(p_i^{2^*}, i=1, \dots, I) > \max(p_i^2, i=1, \dots, I)$ .

Since  $\overline{p^{2^*}} = \overline{p^2} \Rightarrow \max(p_i^{2^*}, i=1, \dots, I) - \overline{p^{2^*}} > \max(p_i^2, i=1, \dots, I) - \overline{p^2}$ . Thus, the

two tests are equivalent.

## Chapter 7

### Summary, Conclusions and Future Research

#### 7.1 SUMMARY AND CONCLUSIONS

Robust HOV tests are required when practitioners suspect that the populations being sampled are nonnormal. This is particularly true when populations are moderately skewed or kurtotic. In this thesis, randomization tests were proposed as alternatives to (some frequently used) HOV tests that in previous research have been shown to be robust to nonnormality.

Of the proposed randomization tests RANDANOMV-R performed well across all distributions and variance configurations. RANDANOMV-R was robust for all distributions examined (with both permutation shuffling and bootstrap shuffling) and displayed somewhat greater power than RANDANOMV-D. Power was somewhat higher for RANDANOMV-R with permutation shuffling, which in general produced greater power for all the randomization tests. RANDANOMV-R was much more powerful than Levene for Configuration 2, and it was nearly as good as (roughly equivalent to) Levene for Configuration 1. Since the former is likely near the LFC, this suggests that RANDANOMV-R has greater power than Levene's test at low power configurations. RANDANOMV-R showed power comparable to the best of the remaining nonrandomization tests for both configurations studied. Those nonrandomization tests were much less robust to extreme kurtosis than RANDANOMV-R.

Hence, RANDANOMV-R (with permutation shuffling) is a good all-purpose robust HOV test that outperforms other tests in circumstances in which the population means are not very different.

In the case where (it is suspected) the populations means are very different, RANDANOMV-DD/-RD (with permutation shuffling) or RANDANOMV-DD (with bootstrap shuffling) should be used instead of RANDANOMV-R. RANDANOMV-DD and -RD were shown to be equivalent in the case of permutation shuffling. A benefit associated with using RANDANOMV-DD/-RD (with permutation shuffling) is that it is more powerful than RANDANOMV-R. However, RANDANOMV-DD/-RD (with permutation shuffling) is not robust to situations where the distribution is extremely skewed or kurtotic. When that is suspected, RANDANOMV-DD (with bootstrap shuffling) should be used since it is robust in all cases, but this test has lower power than either RANDANOMV-DD/-RD (with permutation shuffling), RANDANOMV-R, or Levene's test.

The randomization tests that have been presented allow the user to construct a decision chart to assess practical as well as statistical significance. This offers an advantage to practitioners not offered by commonly used robust HOV tests such as those by Levene or Fligner and Killeen. The ANOM version of Levene's test (ANOMV-LEV) does offer this advantage. The other ANOM-type HOV tests (ANOMV-JK and ANOMV-TR) can be used to produce decision charts; however, the points plotted on these charts are not as easily interpreted as the

sample variances (standard deviations) plotted on the decision charts for Analysis of Means type randomization tests for variances.

## 7.2 FUTURE RESEARCH

While it appears the randomization tests presented in this paper provide viable alternatives to some commonly used HOV tests, there are areas that warrant further study. One such area relates to unequal means. It was presented that some randomization tests may not be useful when the means are “too different.” How different must the means be to render a test useless? Along similar lines it was shown that RANDANOMV-RD was not effective when the parent distribution was too kurtotic or too skewed. How kurtotic or skewed must a distribution be?

The last area for additional work centers on making the randomization tests more useful for practitioners. The programs in this paper were coded in FORTRAN, and this requires the use of a compiler and knowledge of a specific language. A macro could be written in a commonly used statistical analysis program. This could make the tests more available to potential users.

## APPENDIX

### FORTRAN PROGRAM FOR THE RANDOMIZATION TESTS

```
dimension semvar(5),shufvar(5),ddvarto(5)
dimension rdratio(5),ratiovar(5)
dimension cdev(1000),dev(1000)
dimension devvar(5),devvarto(5),var(5)
dimension sampavg(5),randstol(1000),randstoh(1000)
dimension ddslo(1000)
dimension sampvar(5),ratdev(1000)
dimension randsone(1000)
dimension ddshi(1000),ratioslo(1000),ratioshi(1000)
dimension rdslo(1000),rdshi(1000)
real e(5,10),devmean(5,10),c(5,10)
read (*,*) nsamp
read (*,*) npops
read (*,*) var
read (*,*) idist
read (*,*) alpha

nsampdub = nsamp*2
nreps = 2000
iseed = 1579
numshuf = 1000
npopsamp = nsamp*npops
halfalph = alpha/2.
rjdone = 0.
rjdtwo = 0.
rjdd = 0.
rjr = 0.
rjrd = 0.

do iz = 1 , nreps

donerjct = 0.
dtworjcl = 0.
dtworjch = 0.
ddrjcl = 0.
ddrjch = 0.
rrjcl = 0.
```

```
rrjch = 0.  
rdrjcl = 0.  
rdrjch = 0.
```

```
ak = -0.05134  
bk = -2.91756  
ck = 0.05134  
dk = 0.87133  
ee = 2.718281828  
ae = 0.637298719  
uone = 0.  
utwo = 0.
```

```
jd = 1  
do j = 1 , npops  
  if (idist.ne.7) then  
    do k = 1 , nsamp  
      if (idist.eq.1) then  
        x = gasdev(iseed)  
        dev(k) = x * sqrt(var(j))  
        cdev(jd) = dev(k)  
        jd = jd+1  
      endif  
  
      if (idist.eq.2) then  
        x = gasdev(iseed)  
        dev(k) = ((0.66268*x)+(0.10189*(x**3)))*sqrt(var(j))  
        cdev(jd) = dev(k)  
        jd = jd+1  
      endif  
  
      if (idist.eq.3) then  
        x = gasdev(iseed)  
        dev(k) = ak+(bk*x)+(ck*(x**2))+(dk*(x**3))  
        cdev(jd) = dev(k)  
        jd = jd+1  
      endif  
  
      if (idist.eq.4) then  
        x = gasdev(iseed)  
        dev(k) = x**2  
        cdev(jd) = dev(k)  
        jd = jd+1  
      endif  
  
      if (idist.eq.5) then  
        dev(k) = expdev(iseed)  
        cdev(jd) = dev(k)  
        jd = jd+1  
      endif  
  
      if (idist.eq.6) then  
        uone = ran1(iseed)  
        utwo = ran1(iseed)
```

1

```

yy = (-4./9.)*log10(uone)
zz = ee**((-5./9.)+(5.*yy/4.))
ww = (yy**(-5./9.))*ae*zz
  if (utwo.ge.ww) then
    go to 1
  else
    dev(k) = yy
    cdev(jd) = dev(k)
    jd = jd+1
  endif
endif
enddo
endif

if (idist.eq.7) then
do k = 1 , nsampdub
  x = gasdev(idum)
  dev(k) = x
enddo

do k = 1 , nsampdub-1, 2
  if (dev(k).gt.0) then
    anum = 2.
  else
    anum = -2.
  endif
  dev(k+1) = dev(k+1) + anum
enddo

do k = 2 , nsampdub, 2
  dev(k/2) = dev(k)
  cdev(jd) = dev(k/2)
  jd = jd + 1
enddo
endif

call ameanvar(dev,nsamp,savg,svar)
sampavg(j) = savg
sampvar(j) = svar

C      FINDS DEVIATION FROM MEAN FOR -DD AND -RD
do k = 1 , nsamp
  devmean(j,k) = dev(k) - sampavg(j)
enddo
enddo

avgvar = 0.
varsum = 0.

C      FINDING AVERAGE SAMPLE VARIANCE
do j = 1 , npops
  varsum = varsum + sampvar(j)
enddo
avgvar = varsum / float(npops)

```

```

C      FINDING TEST STATISTICS ON INITIAL DATA
do j = 1 , npops
  devvar(j) = abs(sampvar(j) - avgvar)
  devvarto(j) = sampvar(j) - avgvar
  ratiovar(j) = sampvar(j) / varsum
enddo

call bubsort(devvar,npops)
call bubsort(devvarto,npops)
call bubsort(ratiovar,npops)

randone = devvar(npops)
randtwo1 = devvarto(1)
randtwoh = devvarto(npops)
ratio1ow = ratiovar(1)
ratio1hi = ratiovar(npops)

C      SHUFFLES DATA
md = 1
do j = 1 , npops
  do k = 1 , nsamp
    ratdev(md) = devmean(j,k)
    md = md + 1
  enddo
enddo

do i = 1 , numshuf
  do ij = 1 , npopsamp - 1
    rnn = float(npopsamp + 1 - ij)
    ih = int(rnn*ran1(iseed)) + ij
    tmp = cdev(ij)
    tmpp = ratdev(ij)
    cdev(ij) = cdev(ih)
    ratdev(ij) = ratdev(ih)
    cdev(ih) = tmp
    ratdev(ih) = tmpp
  enddo

  md = 1
  do j = 1 , npops
    do k = 1 , nsamp
      devmean(j,k) = ratdev(md)
      c(j,k) = devmean(j,k)**2
      e(j,k) = cdev(md)
      md = md + 1
    enddo
  enddo

C      PERFORMING OPERATIONS ON SHUFFLED DATA
rn = float(nsamp)
do j = 1 , npops
  shsum = 0.
  shssum = 0.

```



```

shsq = 0.
do k = 1 , nsamp
  shsum = c(j,k) + shsum
  shssum = e(j,k) + shssum
  shsq = (e(j,k)**2) + shsq
enddo
semvar(j) = shsum / (rn - 1.)
vars = (shsq-shssum*shssum/rn)
shufvar(j) = vars / (rn - 1.)
enddo

savgvar = 0.
avssemvr = 0.
sumsem = 0.
sumshuf = 0.

do j = 1 , npops
  sumsem = semvar(j) + sumsem
  sumshuf = shufvar(j) + sumshuf
enddo
avssemvr = sumsem / float(npops)
savgvar = sumshuf / float(npops)

do j = 1 , npops
  devvar(j) = abs(shufvar(j) - savgvar)
  devvarto(j) = shufvar(j) - savgvar
  ddvarto(j) = semvar(j) - avssemvr
  ratiovar(j) = shufvar(j) / sumshuf
  rdratio(j) = semvar(j) / sumsem
enddo

call bubsort(devvar,npops)
call bubsort(devvarto,npops)
call bubsort(ddvarto,npops)
call bubsort(ratiovar,npops)
call bubsort(rdratio,npops)

```

C CALCULATES TEST STATISTICS ON SHUFFLED DATA

```

randstone(i) = devvar(npops)
randstol(i) = devvarto(1)
randstoh(i) = devvarto(npops)
ddslo(i) = ddvarto(1)
ddshi(i) = ddvarto(npops)
ratioslo(i) = ratiovar(1)
ratioshi(i) = ratiovar(npops)
rdslo(i) = rdratio(1)
rdshi(i) = rdratio(npops)

if (randstone(i).gt.randone) donerjct = donerjct + 1.
if (randstol(i).lt.randtwol) dtworjcl = dtworjcl + 1.
if (randstoh(i).gt.randtwoh) dtworjch = dtworjch + 1.
if (ddslo(i).lt.randtwol) ddrjcl = ddrjcl + 1.
if (ddshi(i).gt.randtwoh) ddrjch = ddrjch + 1.
if (ratioslo(i).lt.ratiolow) rrjcl = rrjcl + 1.

```

```

        if (ratioshi(i).gt.ratiohi) rrjch = rrjch + 1.
        if (rdslo(i).lt.ratiolow) rdrjcl = rdrjcl + 1.
        if (rdshi(i).gt.ratiohi) rdrjch = rdrjch + 1.

    enddo

    sn = float(numshuf) + 1.
    pvone = (donerjct + 1.) / sn
    pvtwol = (dtworjcl + 1.) / sn
    pvtwoh = (dtworjch + 1.) / sn
    pvddl = (ddrjcl + 1.) / sn
    pvddh = (ddrjch + 1.) / sn
    pvrl = (rrjcl + 1.) / sn
    pvrh = (rrjch + 1.) / sn
    pvrcl = (rdrjcl + 1.) / sn
    pvrch = (rdrjch + 1.) / sn

    if (pvone.lt.alpha) rjdone = rjdone + 1.
    if (pvtwol.lt.halfalph) go to 100
    if (pvtwoh.lt.halfalph) go to 100
    go to 20
100   rjdtwo = rjdtwo + 1.

    20   if (pvddl.lt.halfalph) go to 200
        if (pvddh.lt.halfalph) go to 200
        go to 30
200   rjdd = rjdd + 1.

    30   if (pvrl.lt.halfalph) go to 300
        if (pvrh.lt.halfalph) go to 300
        go to 40
300   rjr = rjr + 1.

    40   if (pvrcl.lt.halfalph) go to 400
        if (pvrch.lt.halfalph) go to 400
        go to 500
400   rjrd = rjrd + 1.

500   enddo

    pvaldone = rjdone / float(nreps)
    pvaldtwo = rjdtwo / float(nreps)
    pvaldd = rjdd / float(nreps)
    pvalr = rjr / float(nreps)
    pvalrd = rjrd / float(nreps)

    write (*,*) 'RANDANOMV-D1 rejected' , pvaldone
    write (*,*) 'RANDANOMV-D rejected' , pvaldtwo
    write (*,*) 'RANDANOMV-DD rejected' , pvaldd
    write (*,*) 'RANDANOMV-R rejected' , pvalr
    write (*,*) 'RANDANOMV-RD rejected' , pvalrd

end

```

```

subroutine ameanvar(b,n,vmean,vvar)
dimension b(n)
sum = 0.
sq = 0.
do j = 1 , n
    sum = sum + b(j)
    sq = sq + (b(j)**2)
enddo
rn = float(n)
vmean = sum / rn
vvar = (sq - sum*sum / rn) / (rn - 1.)
return
end

```

```

subroutine bubsort(devs,n)
dimension devs(n)
do j = 1 , n-1
    do k = 1 , n-j
        if (devs(k).gt.devs(k+1)) then
            tmp = devs(k)
            devs(k) = devs(k+1)
            devs(k+1) = tmp
        endif
    enddo
enddo
return
end

```

```

FUNCTION RAN1(iseed)
DIMENSION R(97)
PARAMETER (M1=259200,IA1=7141,IC1=54773,RM1=3.8580247E-6)
PARAMETER (M2=134456,IA2=8121,IC2=28411,RM2=7.4373773E-6)
PARAMETER (M3=243000,IA3=4561,IC3=51349)
DATA IFF /0/
IF (iseed.LT.0.OR.IFF.EQ.0) THEN
    IFF=1
    IX1=MOD(IC1-iseed,M1)
    IX1=MOD(IA1*IX1+IC1,M1)
    IX2=MOD(IX1,M2)
    IX1=MOD(IA1*IX1+IC1,M1)
    IX3=MOD(IX1,M3)
    DO 11 J=1,97
        IX1=MOD(IA1*IX1+IC1,M1)
        IX2=MOD(IA2*IX2+IC2,M2)
        R(J)=(FLOAT(IX1)+FLOAT(IX2)*RM2)*RM1
11    CONTINUE
    iseed=1
ENDIF
IX1=MOD(IA1*IX1+IC1,M1)
IX2=MOD(IA2*IX2+IC2,M2)
IX3=MOD(IA3*IX3+IC3,M3)
J=1+(97*IX3)/M3
IF (J.GT.97.OR.J.LT.1) PAUSE

```

11

```

RAN1=R(J)
R(J)=(FLOAT(IX1)+FLOAT(IX2)*RM2)*RM1
RETURN
END

FUNCTION GASDEV(iseed)
DATA ISET/0/
IF (ISET.EQ.0) THEN
1   V1=2.*RAN1(iseed)-1.
   V2=2.*RAN1(iseed)-1.
   R=V1**2+V2**2
   IF(R.GE.1.)GO TO 1
   FAC=SQRT(-2.*LOG(R)/R)
   GSET=V1*FAC
   GASDEV=V2*FAC
   ISET=1
ELSE
   GASDEV=GSET
   ISET=0
ENDIF
RETURN
END

FUNCTION EXPDEV(iseed)
   EXPDEV=-LOG(RAN1(iseed))
RETURN
END

```

## BIBLIOGRAPHY

- Baker, R. D. (1995). "Two Permutation Tests of Equality of Variances," *Statistics and Computing*, 5.
- Barnes, M. J. (1981). "The Effects of Kurtosis, Skewness and Sample Size on Type I Error Rates and Power of Tests of Homogeneity of Variance." Ph.D Thesis. Hofstra University, Hempstead, NY.
- Bartlett, M. S. (1973). "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society*, 160.
- Brown, M. B., and A. B. Forsythe, (1974). "Robust Tests for Equality of Variances," *Journal of the American Statistical Association*, 69.
- Cochran, W. G. (1941) "The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total," *Annals of Eugenics*, 11.
- Conover, W. J., Johnson, M. E., and M. M. Johnson (1981). "A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data," *Technometrics*, 23.
- Edgington, E. S. (1987). *Randomization Tests*. Marcel Dekker, Inc, city.
- Fleishman, A. I. (1978). "A Method of Simulating Non-normal Distributions" *Psychometrika*, 43.
- Foster, L. A. (1964). "Testing for Equality of Variances." PhD thesis. Purdue University, Lafayette, Indiana.
- Good, P. (1994). *Permutation Tests*. Springer-Verlag, New York.
- Hartley, H. O. (1940). "Testing the Homogeneity of a Set of Variances," *Biometrika*, 31.
- Hartley, H. O. (1950). "The Maximum F-Ratio as a Short Cut Test of Heterogeneity of Variances," *Biometrika*, 37.

- Hogg, R. V., and A. T. Craig (1995). *Introduction to Mathematical Statistics*. Prentice-Hall Inc., Englewood Cliffs, New Jersey.
- Nelson, L. S. (1983). "Exact Critical Values for Use with the Analysis of Means," *Journal of Quality Technology*, 15.
- Nelson, P. R. (1985). "Power Curves for the Analysis of Means," *Technometrics*, Vol. 27, No. 1.
- Nelson, P. R. (1998). "Notes on Design and Analysis of Experiments," Unpublished.
- Noreen, E. W. (1989). *Computer Intensive Methods for Testing Hypotheses An Introduction*. John Wiley & Sons, New York.
- Ott, E. R. (1967). "Analysis of Means – A Graphical Procedure," *Industrial Quality Control*, 24.
- Rinaman, W. C. (1993). *Foundations of Probability & Statistics*. Saunders College Publishing, Orlando.
- Ross, S. M. (1997). *Simulation*. Academic Press, Inc., San Diego, California.
- Tukey, J. W. (1962). "Data Analysis and Behavioral Science." Unpublished monograph.
- Wludyka, P. S., and P. R. Nelson. (1997 A) "Analysis of Means Type Tests for Variances from Normal Populations," *Technometrics*, 39.
- Wludyka, P. S., and P. R. Nelson. (1997 B) "Analysis of Means Type Tests for Variances Using Subsampling and Jackknifing," *American Journal of Mathematical and Management Sciences*, 17.
- Wludyka, P. S., and P. R. Nelson. (1999) "Two Nonparametric Analysis of Means Type Tests for Homogeneity of Variances," *Journal of Nonparametric Statistics*, Vol. 26, No. 2.
- Wludyka, P. S. (1999). "Notes on Hypothesis Tests," Unpublished.

VITA

**ANTHONY JOSEPH BERNARD**

**BIRTHDATE:**

**BIRTHPLACE:**

<b>EDUCATION:</b>	1995-1999	University of North Florida Master of Science in Mathematical Sciences
	1990-1993	Appalachian State University Bachelor of Science in Statistics
	1988-1990	University of Evansville
	1984-1988	Champion High School
<b>EMPLOYMENT:</b>	1995-1999	Quality Engineer Vistakon, Johnson & Johnson Vision Products, Inc.
	1993-1995	Quality Engineer Eveready Battery Company